Molecular Modeling 2013:

Exercise 2: Molecular Dynamics Simulation of Lysozyme using GROMACS

The aim of this exercise is to get familiar with Molecular Dynamics (MD) simulations using a program called GROMACS. There are many different methods of running dynamics on a protein and many different software packages available for these. Here we focus on MD with GROMACS.

Crystal structures of proteins provide static images of protein systems. However to understand various aspects of protein function, ligand binding, domain movements, etc we need to be able to study the dynamic behaviour of proteins. For this purpose we use MD simulations where all atoms of the system are treated explicitly using classical force fields. MD simulations are rigorous and computationally expensive because of the explicit treatment of every atom, but are still feasible to study large systems like proteins because of the classical treatment of bonds and angles (as opposed to a quantum mechanical treatment).

In this lab we will study Lysozyme by simulating it with GROMACS. Lysozyme is a fascinating enzyme that is used to kill bacteria and is found in tears, saliva, and egg white for example. It was one of the first proteins to be crystallized in 1965, consists of 164 residues and has molecular weight of 14.4 kDa.

There are 5 Steps we will have to complete during the lab.

1. **Creating a topology:** A topology tells the program everything it needs to know to run a simulation. Basically it is a file containing atom co-ordinates (which is obtained from the PDB) and the corresponding force field specific bonded and non-bonded parameters for each atom.
2. **Minimizing the Structure:** We then have to minimize the system before performing producing the MD trajectory. This is mainly because the crystal structure itself is not at a global energy minimum. In addition, steps like hydrogen addition, water addition (solvation), which are performed while generating the topology, can produce non-optimal solutions. Minimization allows the relaxation of the entire system hence reducing the total potential energy of the system.
3. **Restrained equilibration:** To prevent strange behaviour of the protein in solvent, we have to allow the waters around the protein to equilibrate around the protein. Restraining the protein atoms while allowing the waters to move and find their preferred positions enables us to do this. In this step we also slowly raise the temperature to our simulation temperature of 298K
4. **Production run:** Finally your system is ready for producing the simulation using MD. We run MD for a given amount of time and get what is known as the "trajectory" as output for carrying out further analysis.
5. **Analysis:** This is the "fun" step. We actually get to analyze our outputs and learn about the dynamics of our protein.

So now we actually go ahead and carry out these five steps to analyze our protein.

Before starting you will need to run a BASH script. (This allows you to use all the modules required to complete this lab)

First you have to copy the script over to the home directory. You do this by going to the home directory and then doing

*cp **/afs/pdc.kth.se/home/c/cschwaig/vol05_Teaching/gromacs-4.5.3-x86-static/bin/GMXRC.bash .***

*Do not forget the '.' Sign in the end. This specifies the destination of the copy command.

Then you need to run this script by typing
source GMXRC.bash

Your environment is now set with all the GROMACS and Xmgrace (an analysis tool) modules available to you.

A small note all text that follows a "#" sign is a comment, i.e. it gives you the explanation of the commands which you execute.

Step1: Topology generation

1. Download PDB code 1LYD (text file) and save it in your working directory.
2. Analyze the PDB first using PyMol. Does it look like a protein? Does it have any missing residues? Are there waters in the crystal structure? Do you see anything you cannot explain?
3. After this the next step is to convert the PDB into a file that is readable by GROMACS. You do this by running

pdb2gmx –f 1LYD.pdb –water tip3p #tip3p is a specification for water parameters (you will add waters later)

4. When this command is run the program should ask you for a force field we choose OPLS-AA/L and this is pre-chosen (the corresponding number at this step when it asks will be number 14). You will get a lot of output lines on your screen look for ERRORS and WARNINGS.
5. You will have 3 important output files called conf.gro (the co-ordinate file in GROMACS format), topol.top (the topology file which was described earlier) and posre.itp (which will be used in step 3 to restrain atoms)
6. We then need to put the system in a box, solvate the system with water and check the output. This is done with the following command line instructions.

editconf –f conf.gro –bt dodecahedron -c –d 0.5 –o box.gro # here we generate a do-decahedron box with a distance of at least 0.5 nm for any

protein atom to the edge of the box. (In reality the distance would be greater but since time is a limitation we make a small system). The "-c" flag is to centre your protein in the box.

genbox –cp box.gro –cs spc216.gro –p topol.top –o solvated.gro #we then solvate our box i.e. fill it water – spc216 is a type of water system. We use the box.gro file from the previous step and we add the waters to the topology file. Our output it solvated.gro

grompp -f steep.mdp -c solvated.gro -p topol.top -o steep.tpr

trjconv_d -s steep.tpr -f solvated.gro -pbc mol -center -o solvated_new.pdb

trjconv_d -s steep.tpr -f solvated_new.pdb -pbc mol –center –ur compact -o solvated_new.pdb

In this step we convert the gro file to a pdb file which can be analyzed with pymol. Steep.mdp is an input file that has been provided via email.

7. GROMACS will ask you to choose a group (from a list of 13) in this step choose 0 i.e. the entire system
8. Look at the solvated.pdb file in Pymol. Observe the difference in size with the original PDB.
9. We now have our topology file

Step 2: Energy minimization

1. Each of you will receive 3 files before the lab via e-mail. Each line in these files will be explained in detail during the lab session. Download the em.mdp file to the working directory.
2. GROMACS has a pre-processor called grompp which collates the required information into a single file. This is done by

   grompp –f em.mdp –p topol.top –c solvated.gro –o em.tpr #em.tpr is the collated output

3. Check for errors and then proceed
4. We now run minimization

   mdrun –v –deffnm em #we ask it to consider em as the default name. This step takes a few minutes to complete.

Step 3: Restrained equilibration

1. In this step we use the posre.itp to restrain the protein atoms to their initial positions with a force constant. Copy the pr.mdp file from your mail to the working directory.
2. We then run grompp and mdrun like the previous step

   grompp –f pr.mdp –p topol.top –c em.gro –o pr.tpr

mdrun –v –deffnm pr # this takes around 10-15 minutes to complete.

Step 4: Production run

1. Copy the run.mdp to your working directory
2. We again run grompp and mdrun

   grompp –f run.mdp –p topol.top –c pr.gro –o run.tpr

   mdrun –v –deffnm run # Note the time taken for this simulation to complete.

Step 5: Analysis

1. We can now analyze the trajectories. The runs performed in this lab due to time constraints are too short. You will be provided with 3 files run.gro run.tpr and run.xtc from a much larger simulation to analyze. The first fun thing to do is make a movie. After all we did "produce" a trajectory. This is done by

   cp/afs/pdc.kth.se/home/c/cschwaig/vol05_Teaching/Molecular_Modeling_course/Lab2_gromacs/run.* .

   trjconv –s run.gro –f run.xtc –e 2500 –o movie.pdb #this movie is with the first 2.5 ns of the longer md run.

2. The second interesting analysis is to get the Root Mean Square Deviation (RMSD) from the run. This compares movement of the protein backbone between the starting structure and during the course of MD. We do this by

   g_rms –s run.tpr –f run.xtc –xvg none –o rmsd.xvg

   NOTE: You have to choose group 4 (backbone) while running this.

3. A third task would be to follow a hydrogen bond. We do this using an index file. We choose Glu22 and Arg 137 with "r 22" and "r 137" respectively at the prompt, followed by del 0-14 (removes all other groups) and q to save the index file. The command to use is

   make_ndx –f run.gro

4. We then generate the distances with

   g_mindist –s run.tpr –f run.xtc –n index.ndx –xvg none –o dist.xvg
   g_hbond –s run.tpr –f run.xtc –n index.ndx –xvg none –num hbnum.xvg

5. Plot the outputs with any plotting software of your choice. For this lab we have provided 3 plot (.plt) files for using GNUPLOT a command line based plotting software. Copy these over to your working folder.

6. To use GNUPLOT – type ***gnuplot*** in the terminal. You will enter a gnuplot terminal. There you have to load the plot (.plt) file using the command

load "./graph_RMSD.plt"
load "./graph_dist.plt"
load "./graph_hb.plt"

7. You will have 3 outputs called RMSD_out.ps, DIST_out.ps and HB_out.ps from GNUPLOT which you can open by typing ***open*** followed by the filename and check the output plots.

Your Report must include!

1. A short introduction to MD (0.5 page)
2. A pymol figure for the initial PDB, and the solvate.pdb.
3. The plots for RMSD, distances and hydrogen bond count using your favourite plotting tool.

Questions that need to be answered in the report

1. MD is based on classical mechanics and force-fields. Can you think of one example of a phenomenon for which MD cannot be used?
2. Comment on the crystallographic structure? How many helices, sheets, and waters do you see with pymol?
3. We mention that we slowly raise the temperature of the system to 298K. What is the initial temperature of the system?
4. What would happen if we run a protein in vacuum without solvent?
5. How long did the production run of MD take?
6. What pattern do we observe in the RMSD plot (max deviation, time required for stabilization etc). Try and provide an explanation for what we see.
7. Are the hydrogen bonds we follow conserved throughout the trajectory?
8. Pick two input paameters each from the em.mdp and run.mdp and explain their significance. The parameters picked should be different for each file. What would the effect of a change in value be?

Hand in the report no later than on 4th January 2014! If you pass the first time, you get one bonus point to the exam.