Übung: Molecular Dynamics Simulation of a Protein

SS 2014 Übung zu Struktur und Simulation

01-Jul-2014

A brief report should be handed in by 15-Jul-2014.

1. Introduction

It is assumed that

* you remember how to get a file of coordinates from the protein databank (`www.rcsb.org`).

* you can produce a simple *x*-*y* plot

* you remember the program chimera, but see the notes on chimera at the end

The aim is to have different people run simulations at different temperatures, to see when a protein falls apart. We work on two different proteins and they will behave differently. This means we need the results from different people to write the report.

There is much room for making typing mistakes and everybody makes typing misakes. The various programs will print an error message if something did not work. Please check the output carefully. If you see very strange results, it almost certainly means you have typed something wrong. Some people will see strange results because of the temperatures they are simulating at.

1.1. Calculations

The simulations are rather realistic, but we will make a drastic simplification so as to save computer time. We will run on proteins *in vacuo*. There will be no solvent, explicit or implicit. This saves more than an order of magnitude CPU time. The model uses united atoms (hydrogens merged into their carbons), but polar hydrogens are treated explicitly. There are lots of parameters and unfamiliar terms in the documentation. Either ignore them or ask what they mean.

The MD calculations will use a time step of 0.002 ps (2×10^{-15}s). You have to know this when labelling the *x*-axis on the plots. Each 500 steps = 1 ps.

The steps will be

* convert files to gromacs format

* run an energy minimisation

* run a very short MD simulation for equilibration and not analyse anything

* run a longer simulation

* do some analysis of the runs and exchange results with people who worked at different temperatures

2. Choosing your protein and temperature

To choose which protein you will simulate, add together the first letters of your first name (TOM = 20 + 15 + 13 = 48). If the number is even, use protein 2a3d, else use protein 1omb. To get the temperature you should simulate at, add together the letters of your last name like above and take that number modulo 7. If your number is n, then choose the n-th (start to count at 0) temperature from this list: 200 K, 250 K, 300 K, 350 K, 400 K, 500 K, 600 K.

3. Gromacs calculations

You are responsible for giving your input and output files sensible names. The ones given here are suggestions. If you are working on protein 1omb, you would want to replace each `2a3d` below with `1omb`.

Start by finding the coordinates for your protein and save them in a filename like 2a3d.pdb or 1omb.pdb.

First one removes unwanted hydrogens and adds polar hydrogens if necessary. If your protein is 2a3d, then :

```
~matthies/gromacs/bin/pdb2gmx \
            -ignh -ff G43b1 -f 2a3d.pdb -o 2a3d_start.pdb -p 2a3d
```

This will put starting coordinates in `2a3d_start.pdb`. The "G43b1" tells gromacs to use a united-atom force field with parameters suited to *in vacuo* simulations. `2a3d.top` contains description (topology) of the molecule.

Next, one has to make input files for the simulation or minimization programs. For this, one uses "grompp". This is a pre-processor which checks if the input is sensible, writes a description of the molecule and then a long input file for the next step.

Copy the files

`~matthies/uebung_gromacs/*.mdp` to your working directory. Open `em.mdp` in an editor and

* set `nsteps = 50000` (the maximum number of steps of steepest descent minimisation)

* set `emtol = 20` (set the convergence criterion)

Then

```
~matthies/gromacs/bin/grompp \
     -f em.mdp -c 2a3d_start.pdb \
     -p 2a3d.top -po 2a3d_em -o 2a3d_em
```

The switches tell gromacs to run an energy minimising calculation using the steepest descents method for a maximum of 50000 steps.

Now you should have `2a3d_em.tpr` and `2a3d_em.mdp` which contain enough information for an energy minimisation calculation. Finally run the energy minimisation calculation:

```
~matthies/gromacs/bin/mdrun \
    -s 2a3d_em -o 2a3d_em -c 2a3d_em.pdb -g 2a3d_em -e 2a3d_em
```

What should happen is that the coordinates from `2a3d_start.pdb will be read, minimised and the new, lower energy coordinates written to 2a3d_em.pdb`. The information about energies should be in the log file, `2a3d_em.log`.

Read the log file and note

* how many steps were run
* the initial energy
* the final energy

} look at potential energy, there is no kinetic energy in an energy minimisation

You might also want to use chimera to compare the starting and finishing coordinates. They are probably rather similar.

3.1. First MD run

The energy minimisation calculation will have removed some high energy problems such as bad contacts (atoms too close to each other). Now, the molecule can be relaxed further by a short MD run. We will not be doing much analysis of this step. The aim is to relax the coordinates.

If you copied the files `*.mdp` as described above, you should find a file called `md1.mdp`. Edit this file. Set the reference temperature `ref_t` and initial temperature `gen_temp` to 250 K. Set the number of steps `nsteps` to 10 000. As before, we need two commands to run the simulation.

```
~matthies/gromacs/bin/grompp \
      -c 2a3d_em.pdb -f md1.mdp -p 2a3d –o 2a3d_md1
```

then

```
~matthies/gromacs/bin/mdrun \
  -s 2a3d_md1 -o 2a3d_md1 -e 2a3d_md1 -c 2a3d_md1.pdb  \
  -g 2a3d_md1
```

which should produce

* `2a3d_md1.edr` (energies)
* `2a3d_md1.log` (log file to be read)
* `2a3d_md1.pdb` (final coordinates)
* `2a3d_md1.trr` (trajectory)

Now, you have to stop and collect some information necessary for the report.

- start chimera and look at the initial PDB file, the coordinates after energy minimising and the coordinates after the short MD run. It should be easy to superimpose the files and see if the structure has changed much at each stage. Your trajectory is very short, but you may also like to have a look at the trajectory reading features in chimera.

- from the log file, note down the initial and final energies (total, potential and kinetic) as well as the average and the root mean square (rms) fluctuations

3.2. Second MD run

The system (protein) may not be truly equilibrated, but it should roughly fit into the force field. Now, we run a second, longer simulation. The inputs will be different for different people. We will run the two steps grompp and mdrun, so edit md2.mdp and set the temperature `ref_t` to the temperature that was given to you. Then edit `md2.mdp` and set `nsteps` to a very small number like 20.

```
~matthies/gromacs/bin/grompp  \
     -c 2a3d_md1 -f md2.mdp -p 2a3d -o 2a3d_md2
~matthies/gromacs/bin/mdrun \
   -s 2a3d_md2 -o 2a3d_md2 -c 2a3d_md2.pdb -e 2a3d_md2 \
```

```
-g 2a3d_md2
```

Check that the commands have worked. For example, do you have a `2a3d_md2.tpr` and the corresponding log file ? If so, edit `md2.mdp` and set `nsteps` to 250 000 and run the simulation.

4. Simple analysis

4.1. Energies

From your calculations, you should have files like

`xxx_md2.trr` (trajectory)

`xxx_md2.edr` (energies)

`xxxx_md2.log` (log file)

where xxx will depend on your protein and the names you used.

Read the log file. Note down the start, finish, average and root mean square fluctuations of the total, potential and kinetic energies. In chimera, try to look at the trajectory file.

First look at the energies from the calculations. There is a program which will write these in a format for xmgrace. Try

```
~matthies/gromacs/bin/g_energy -f xxx.edr  -o xxx.xvg
```

then

```
xmgrace xxx.xvg
```

You do not have to use xmgrace, but it should be convenient. The program g_energy will ask which energies you want. Try

Potential
Kinetic-En.
Total-Energy

Giving each answer on its own line.

4.2. Fluctuations

When looking at the properties of a trajectory, the gromacs programs will often ask which atoms you want to look at. In each case, we will choose the backbone atoms only.

It is not impossible that some people remember the formula for root mean square differences of coordinates given last semester. In an MD simulation, it is useful to compare each snapshot with the starting coordinates, to see how far the structure has moved. The formula is still

$rmsd = \left(\frac{1}{n} \sum_{i=1}^{n} |\vec{r}_{1,i} - \vec{r}_{2,i}|^2 \right)^{\frac{1}{2}}$ where $\vec{r}_{1,i}$ is the starting configuration and $\vec{r}_{2,i}$ a snapshot from a trajectory. To see this from a gromacs trajectory,

```
~matthies/gromacs/bin/g_rms -f  xxx.trr  -s xxx.tpr
                          -o xxx_rms.xvg
```

Your results will not be the same as anybody else's. The "`xxx`" will refer to your protein and your temperature.

Next we should look at the radius of gyration of the protein. This is a measure of the size of a protein. First one calculates the centre of mass and then the root mean square difference of each atom to this point. The gromacs command is

```
~matthies/gromacs/bin/g_gyrate -f  xxx.trr -s xxx.tpr
                                              -o xxx_gyr.xvg
```

and the file, xxx_gyr.xvg can be viewed and printed with xmgrace.

5. Assignment

All graphs must have labels on the axes with units. Any numbers in tables should have units. Please do not quote more significant digits than is sensible. If you have a quantity such as 2355.55 ±323 kJ mol^{-1} (rms fluctuation), the best way to quote the energy is 2.4×10^3 kJ mol^{-1}.

1. Write the most important details of your protein:

* its name

* number of residues

* how many disulfide bridges does it have

* very crude description of structure (mostly α-helical, mixed α/β, much irregular structure..)

* how the structure was solved (X-ray, NMR)

* write the temperature you did your simulation at

2.

* Build a table for each protein like:

				Temperature (K)		
	initial	after EM	MD1	200	250	…
	kJ mol^{-1} ± rms fluctuations					
E_{tot}						
E_{pot}						
E_{kin}	X	X				

E_{tot}, E_{pot}, E_{kin} are the total, potential and kinetic energies respectively. There is no kinetic energy from an energy minimisation. Rms fluctuation refers to the root mean square fluctuations. All these quantities are in the xxx.log files from the calculations. For all MD runs, quote the average energy values.

* For your protein and your temperature, produce an *x*-*y* plot from the second MD simulation of total, potential and kinetic energies as a function of time. Label all axes and put in units.

During this longer simulation,

* Is there a trend in total energy ?

* Does the system seem to be equilibrated ?

* Are the fluctuations of potential and kinetic energy correlated ?

3. In chimera, compare the structure

(a) from the PDB

(b) after energy minimising

(c) after the first MD run

(d) after the second MD run

* Does the structure change much in any single step ? The answer to this question will be different for every person in the class.

* If you view the trajectory in chimera, is there any trend ? Does the structure open or close ? Do the secondary structure units remain intact ?

* Does the plot of radius of gyration or root mean square displacement (rmsd) support what you say ?

4. In some of the simulations, there will be trend in the total energy. It will have decreased. This seems to violate energy conservation. Where has the energy gone to ? This question requires the lecture material on temperature baths.

5. Imagine that after simulating at 300 K, the average coordinates have definitely moved away from the starting conformation, but the energy does not seem to have increased. Someone claims that the native conformation from the protein data bank should be the lowest energy. What is a plausible explanation ?

6. Background and reference notes

6.1. Gromacs and the commands

Gromacs is one of the popular MD codes used for research projects. It has been installed under `~matthies/gromacs`. There is no single program called "gromacs". Instead, there is a collection of programs in `~matthies/gromacs/bin`.

Gromacs is free and can be found at `www.gromacs.org`. This web page also has the full documentation. We have the text for the man pages sitting in the distribution directory. Try a command like

```
cd ~matthies/gromacs/man/man1
man ./grompp.1
man ./mdrun.1
```

We will also be using the commands `g_energy` and `g_rms`.

6.2 Chimera: displaying structures and movies

Everyone has used "chimera" last semester. It can display trajectories from simulations from a program like Gromacs. It is installed as `~matthies/bin/chimera`.

If you have a trajectory of a molecule like 1omp, you will need the trajectory file `xxx.trr` and the atomic description file, `xxx.tpr`. Look for the menu entry, "tools", then "MD/Ensemble analysis". Specify the format as gromacs.

6.3 Plotting with xmgrace

To choose which columns to plot from the command line, use the command

```
xmgrace -block foo.xvg -bxy 1:2 -bxy 1:3
```

The option `-bxy 1:2` means to plot column 1 against column 2. You can use one or more `-bxy` options, depending on how many graphs should be in the plot. Take a look at the .xvg file with a text editor to see how many columns of data are in the file and what each column means. To save a plot to a file with xmgrace, choose the "Print Setup" option from the menu and select a file type (e.g. PNG) and a file name. Then, choose the "Print" menu item to save the file.