Lab Report of BIM3019

MD Simulation

Name: HU Linfeng

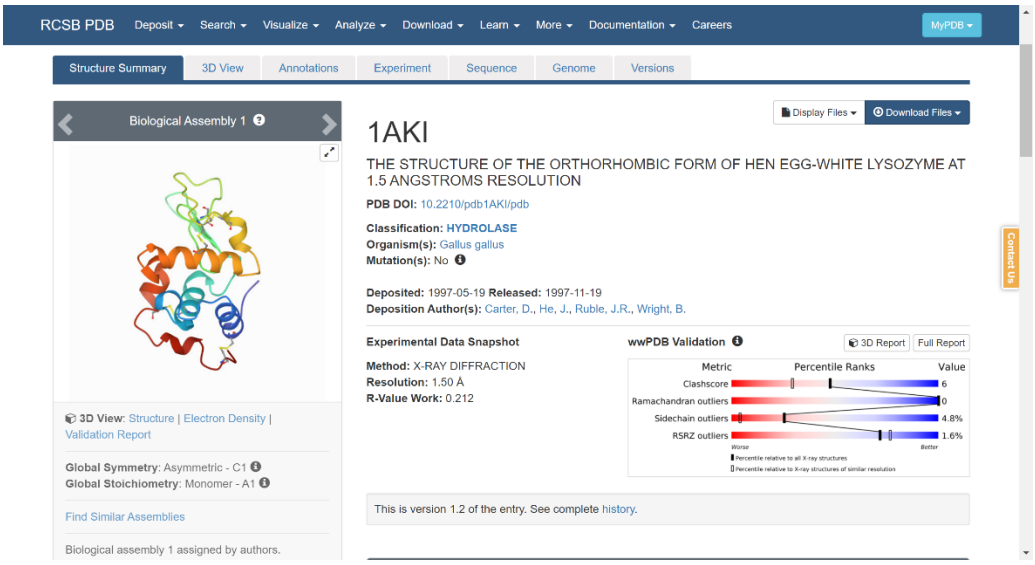
ID: 120090473

Start date: 4th Mar.

Question 1:You should now have a look of 1AKI on RCSB Protein Data Bank to know what's inside the 1aki.pdb. According to the PDB, what is the resolution of this structure? Does this protein exist as a monomer or dimer?

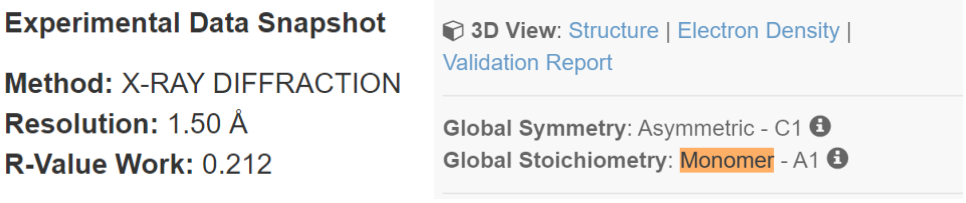
Answer:

The screen shot of the webpage of 1AKI in PCSB Protein Data Bank (https://www.rcsb.org/structure/1AKI) is as follows:



The title shows that 1AKI is the structure of the orthorhombic form of hen egg-white lysozyme at 1.5 angstrom resolution

In the webpage, we can also find:



Which also indicates **the resolution of this structure is 1.5 angstrom, the protein exist as monomer.**

Question 2: Read the gro file format on the webpage:

https://manual.gromacs.org/archive/5.0.3/online/gro.html

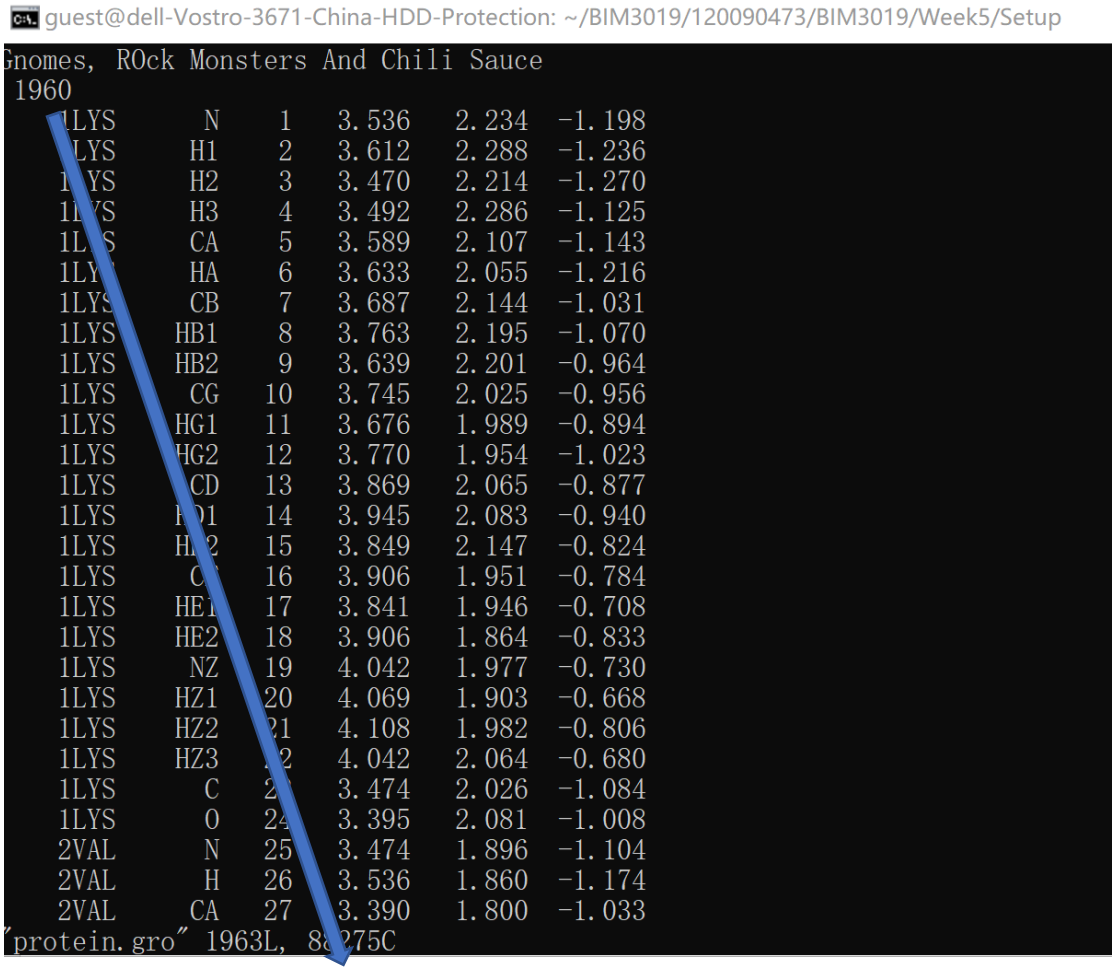
Answer the following questions:

(a) How many atoms are inside this protein?

Answer:

In the Setup directory, input command "vi protein.gro" .

We can see the content:



We can see the number 1960 on the second row, which indicates 1960 atoms are in this protein.

(b) What is the 2nd residue in this protein?

Answer:

```
    1LYS      O   24    3.395   2.081  -1.008
    2VAL      N   25    3.474   1.896  -1.104
    2VAL      H   26    3.536   1.860  -1.174
    2VAL     CA   27    3.390   1.800  -1.033
"protein.gro" 1963L, 88275C
```

We can see "2VAL" here, which means the second residue is **Valine (Val)**.

(c) Does the atom number start from 1 or from 0?

Answer:

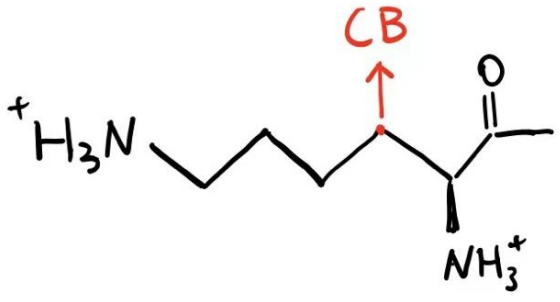
As the screenshot shown, the number **starts from 1**.

```
Gnomes, ROck Monsters And Chili Sauce
 1960
    1LYS      N    1    3.536   2.234  -1.198
    1LYS     H1    2    3.612   2.288  -1.236
    1LYS     H2    3    3.470   2.214  -1.270
    1LYS     H3    4    3.492   2.286  -1.125
    1LYS     CA    5    3.589   2.107  -1.143
    1LYS     HA    6    3.633   2.055  -1.216
    1LYS     CB    7    3.687   2.144  -1.031
```

(d) The 7th atom has the atom name CB. Draw the residue structure and indicate which atom is CB.

Answer:

From BIM2005, we learned the basic structural of amino acid. We label the chiral center as CA (also, from the content in the protein.gro, CA, CB, CG, CD is Cα, C β, Cγ, Cδ correspondingly). Since we can find HZ1, HZ2, HZ3 and H1, H2, H3, the two amino group are all protonated, thus, the residue structure should be:



residue structure of 1LYS

CB is the first heavy atom on the side chain, **which is labelled in red on the structure**.

(e) What are the coordinates of the Cα atom of the second residue?

Answer:

1LYS	O	24	3.395	2.081	-1.008
2VAL	N	25	3.474	1.896	-1.104
2VAL	H	26	3.536	1.860	-1.174
2VAL	CA	27	3.390	1.800	-1.033
2VAL	HA	28	3.317	1.852	-0.990
2VAL	CB	29	3.314	1.703	-1.123

The coordinates (unit: nm, in xyz form) of Cα is **(3.390, 1.800, -1.033)**.

Question 3: Are all atoms restrained? Compare the atom number in posre.itp and in topol.top to see which atoms are excluded.

Answer:

Knowing that the posre.itp is the position restraints file, the topol.top is the topology file.

Screenshots of partial content of posre.itp and topol.top (reserved as #topol.top.1# after the tutorial) are as follows:

```
; In this topology include file, you will find position restraint
; entries for all the heavy atoms in your original pdb file.
; This means that all the protons which were added by pdb2gmx are
; not restrained.

[ position_restraints ]
; atom  type      fx      fy      fz
     1     1  1000  1000  1000
     5     1  1000  1000  1000
     7     1  1000  1000  1000
    10     1  1000  1000  1000
    13     1  1000  1000  1000
    16     1  1000  1000  1000
    19     1  1000  1000  1000
    23     1  1000  1000  1000
    24     1  1000  1000  1000
    25     1  1000  1000  1000
    27     1  1000  1000  1000
    29     1  1000  1000  1000
    31     1  1000  1000  1000
    35     1  1000  1000  1000
    39     1  1000  1000  1000
    40     1  1000  1000  1000
    41     1  1000  1000  1000
    43     1  1000  1000  1000
```

Figure Q3.A

```
[ moleculetype ]
; Name            nrexcl
Protein_chain_A     3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass  typeB
chargeB      massB
; residue   1 LYS rtp LYSH q +2.0
     1   opls_287      1     LYS     N      1       -0.3   14.0027
     2   opls_290      1     LYS     H1     1       0.33      1.008
     3   opls_290      1     LYS     H2     1       0.33      1.008
     4   opls_290      1     LYS     H3     1       0.33      1.008
     5  opls_293B      1     LYS     CA     1       0.25     12.011
     6   opls_140      1     LYS     HA     1       0.06      1.008
     7   opls_136      1     LYS     CB     2      -0.12     12.011
     8   opls_140      1     LYS     HB1    2       0.06      1.008
     9   opls_140      1     LYS     HB2    2       0.06      1.008
    10   opls_136      1     LYS     CG     3      -0.12     12.011
    11   opls_140      1     LYS     HG1    3       0.06      1.008
    12   opls_140      1     LYS     HG2    3       0.06      1.008
    13   opls_136      1     LYS     CD     4      -0.12     12.011
    14   opls_140      1     LYS     HD1    4       0.06      1.008
    15   opls_140      1     LYS     HD2    4       0.06      1.008
    16   opls_292      1     LYS     CE     5       0.19     12.011
    17   opls_140      1     LYS     HE1    5       0.06      1.008
    18   opls_140      1     LYS     HE2    5       0.06      1.008
    19   opls_287      1     LYS     NZ     6       -0.3   14.0067
    20   opls_290      1     LYS     HZ1    6       0.33      1.008
```

Figure Q3.B

Where figure Q3.A is partial screenshot of posre.itp, Figure Q3.B is partial screenshot of topol.top (reserved as #topol.top.1#). We can tell that in the position restraints file (posre.itp), the number of atoms are listed as 1, 5, 7, 10… which indicates **only some of the atoms are restrained**. In the topology file, the first part of content is atoms, where contain all the information of 1960 atoms. **Thus, not all atoms are restrained, only some of the atoms are restrained**.

Question 4: Load the protein.gro generated by Gromacs and answer the following questions:

(a) How many disulfide bond do you find? List the cysteine pairs that can form disulfide bonds.

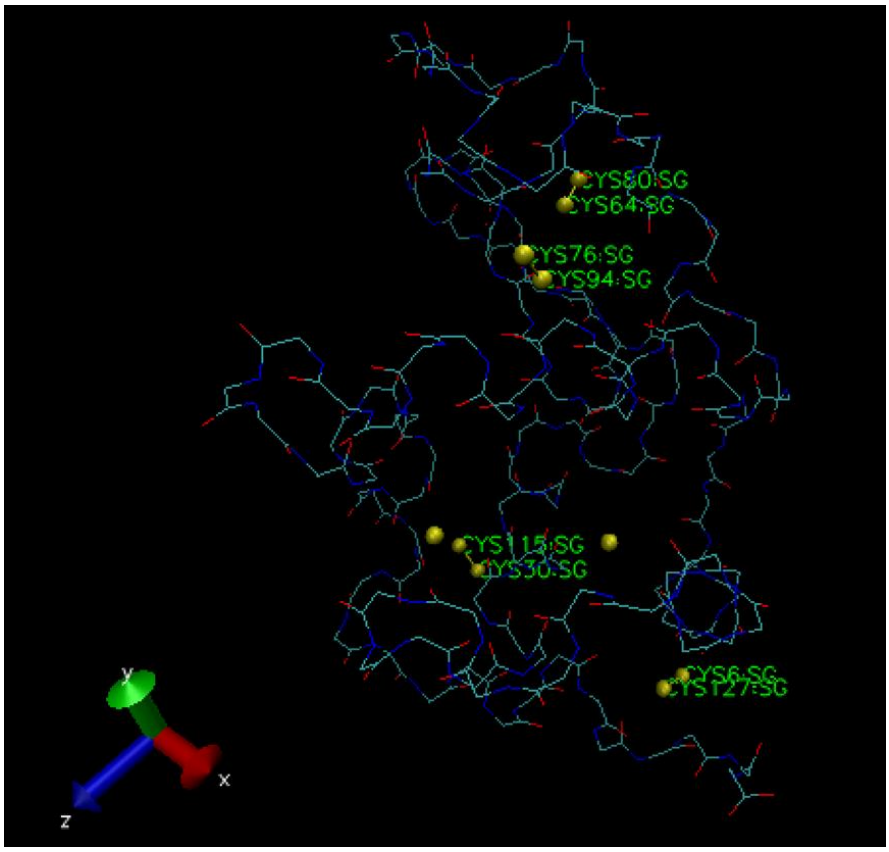
Answer:

4 disulfide bond can be found. The systein pairs are: **CYS80-CYS64, CYS76-CYS94, CYS30-CYS115 and CYS6-CYS127.**

(b) Show a figure of the protein with a clear secondary structure and all disulfide bond. Describe how you made the figure.

Answer:

The figure can be shown as:



Step:

1. Load protein.gro in VMD. Select Graphic→Representation.
2. Type in “backbone”, show it by Lines; type in “sulfur”, show it by CPK.
3. In main VMD, choose “Mouse→Label→Atom.” Then left click the yellow

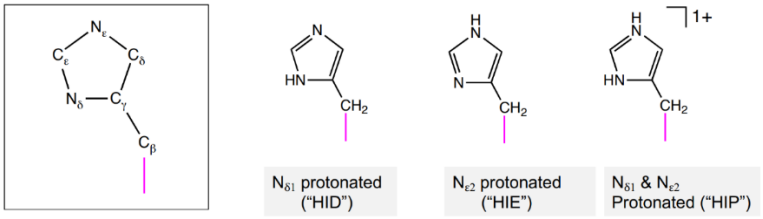
balls (sulfur atoms) to label the Cys residues.

4. Adjust the angle of view by mouse until find a suitable angle.
5. Select “File→Render→Tachyon” and we can have a bmp file in the VMD folder.

(c) Histidine has 3 possible protonation states. What are they?

Answer:

From BIM2005, we learned that the N on the side chain of Histidine can be protonated:



As the picture shown here, when Nδ1 is protonated, it is called HID; when Nδ2 is protonated, it is called HIE; when both Nδ1 and Nδ2 are protonated, it is called HIP.

(d) Is there any histidine? What is its protonation state?

Answer:

By checking the content of protein.gro, there is a Histidine residue by checking the sequence:

```
15HIS      N  229   3.968   2.325   0.546
15HIS      H  230   3.914   2.249   0.510
15HIS     CA  231   3.908   2.459   0.553
15HIS     HA  232   3.988   2.519   0.554
15HIS     CB  233   3.822   2.492   0.428
15HIS    HB1  234   3.773   2.410   0.398
15HIS    HB2  235   3.757   2.564   0.450
15HIS     CG  236   3.908   2.539   0.317
15HIS    ND1  237   3.946   2.464   0.209
15HIS    CD2  238   3.974   2.657   0.299
15HIS    HD2  239   3.966   2.739   0.356
15HIS    CE1  240   4.021   2.531   0.126
15HIS    HE1  241   4.049   2.504   0.034
15HIS    NE2  242   4.052   2.642   0.189
15HIS    HE2  243   4.124   2.706   0.160
15HIS      C  244   3.831   2.486   0.681
15HIS      O  245   3.761   2.588   0.695
```

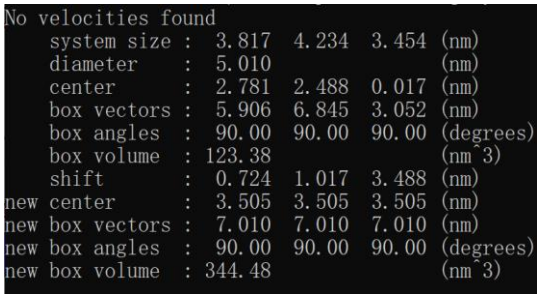
We can see that on Nδ1 there is no H atom, on Nε2 there is HE2. Thus we can conclude that the histidine residue is in Nε2 protonated state (HIE).

Question 5: Answer the following 2 questions:

(a) In the editconf command, we use the flag -d to assign the distance from solute to box to be 1.0. What is the unit of this 1.0?

Answer:

From the screenshot of the process, we can tell that the unit should be **nanometer**.



(b) Why do we use -d 1.0 instead of -d 0.2? (Hint: Consider what will happen if you use -d 0.2 with the periodic boundary conditions.) Give your answer along this line.

Answer:

We need to create a box such that we can have an equivalent surrounding environment of atoms as if it were in the bulk. If we use -d 0.2, the periodic boundary condition may cause one atom interact with atoms in the adjacent block (i.e. it calculates the forces on itself and its replicas, which is not a possible case), since the cutoff radius should be large enough to avoid the possibility of the above situation.

Question 6: Compare the topol.top before and after adding solvent. (The old topol.top is renamed to #topol.top.1#) Do you fine anything different in the topol.top file after solvation?

Answer:

After the whole MD simulation, the topol.top in the question becomes '#topol.top.2#'. Thus, in Linux, we use "cd 120090473/BIM3019/Week5/Setup". We can find all

topology files are in this directory.

Use command diff '#topol.top.1#' '#topol.top.2#' to differ, we have:

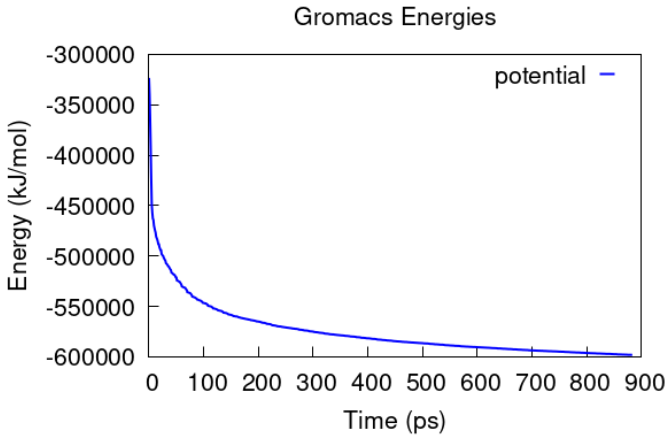
```
guest@dell-Vostro-3671-China-HDD-Protection:~/BIM3019/120090473/BIM3019/Week5/Setup$
 diff '#topol.top.1#' '#topol.top.2#'
18404c18404
< Protein
---
> Protein in water
18408a18409
> SOL             10644
```

We can tell that the later topology file set the condition: protein in water. After the solvation, the condition of the environment is set to be "protein in water", with solvent around the protein.

Question 7: Show the EM.png you got. Is the energy minimization working for you? How can you tell from the plot?

Answer:

The EM.png I obtain is:

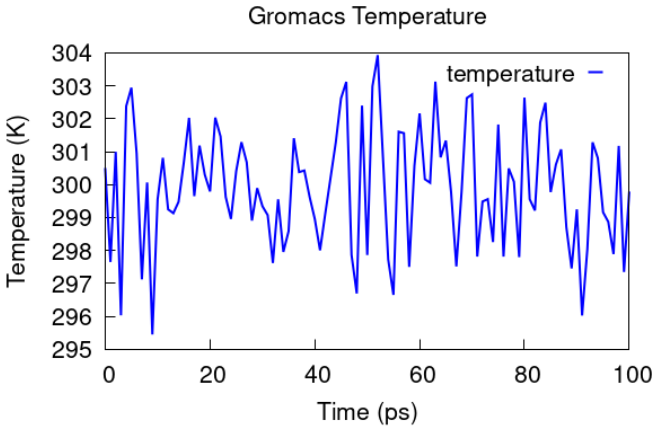


The lowest potential energy approach to -600000 kJ/mol. It works well. The potential energy drops sharply at first and then slowly levels off at -600000 kJ/mol.

Question 8: Copy the plot.gpl and modify it to plot the temperature.xvg. Show your temperature.png file. **How does the system temperature change over time? What's the temperature value?**

Answer:

The temperature.png I obtain is:



During the whole MD process, **the temperature fluctuates irregularly between 295.5K and 304K. The temperature value is between 295.5K (approximately) and 304K.**

Question 9: Look at the nvt.mdp file:

(a) Is the POSRES flag defined? What does it do?

Answer:

```
title                    = OPLS Lysozyme NVT equilibration
define                   = -DPOSRES  ; position restrain the protein
```

From the content in nvt.mdp file, we can tell that **POSRES flag is defined**. Its function is to **set the position condition that restrain the protein**.

(b) What was the integrator used?

Answer:

```
integrator               = md        ; leap-frog integrator
```

We run the MD simulation by **using the leap-frog integrator**.

(c) What was the step size (dt) ?

Answer:

```
nsteps                   = 50000     ; 2 * 50000 = 100 ps
dt                       = 0.002     ; 2 fs
```

The **step size (dt) here is 2 fs**.

(d) How many steps did we use?

Answer:

```
nsteps                   = 50000      ; 2 * 50000 = 100 ps
dt                       = 0.002      ; 2 fs
```

The **step here is 50000 steps**.

(e) How long did we propagate the system (in ps)?

Answer:

```
nsteps                   = 50000      ; 2 * 50000 = 100 ps
dt                       = 0.002      ; 2 fs
```

The time is given by **steps multiplies step size**, from the content of the file, we propagate the system **in 100ps**.

Question 10: Use the plot.gpl to plot pressure.xvg and density.xvg. Show your Figures. What should the average pressure and the average density be? Do the results agree with your expectation?

Answer:

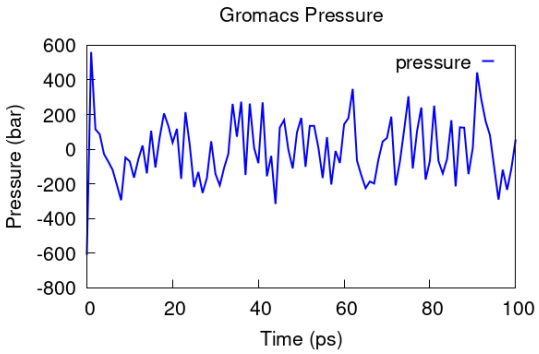


Figure Q10-1

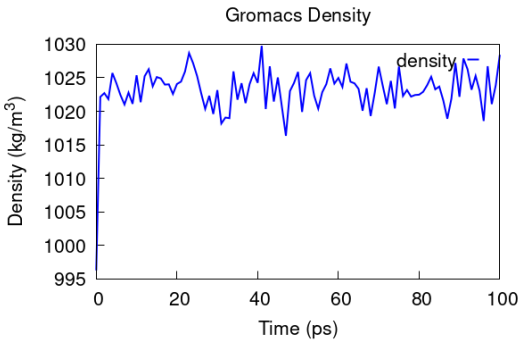


Figure Q10-2

We can tell that Figure Q10-1 is the figure about the pressure, Q10-2 is the figure about the Density during the MD simulation.

```
<====== ############### ==>
<==== A V E R A G E S ====>
<== ############### ======>

Statistics over 50001 steps using 501 frames

Energies (kJ/mol)
      Bond        Angle    Proper Dih. Ryckaert-Bell.       LJ-14
 1.45562e+03   3.72252e+03   2.48148e+02   1.82835e+03   2.67655e+03
  Coulomb-14        LJ (SR)  Disper. corr.   Coulomb (SR)   Coul. recip.
 7.74597e+03   9.41782e+04  -4.57856e+03  -6.49365e+05   3.04201e+03
Position Rest.      Potential    Kinetic En.   Total Energy  Conserved En.
 4.67347e+02  -5.38579e+05   8.55849e+04  -4.52994e+05  -4.52618e+05
 Temperature Pres. DC (bar) Pressure (bar)   Constr. rmsd
 2.99962e+02  -2.26964e+02  -4.21817e+00   0.00000e+00
```

By checking the npt.log, we can tell that the average pressure is -4.21817 bar.

Edit the plot.gpl file:

```
##############################################################################################
# This is a gnuplot script. More can be found on http://gnuplot.sourceforge.net/demo_5.4.
##############################################################################################
set terminal pngcairo size 600,400 enhanced font "Helvetica,16"
set output "density.png"
set title "Gromacs Density"
set xlabel "Time (ps)"
set ylabel "Density (kg/m^3)"

# Graph setting here
set tics nomirror font "Helvetica,16"
set key samplen 0.3 spacing 0.8 font "Helvetica,16"

# mean value
mean(x)=avg
fit mean(x) "density.xvg" u 1:2 via avg

# Plot data here
plot "density.xvg" u 1:2 w l lc rgb "blue" lw 2.0 title "density"
```

By using "# mean value" part we can have the average value of the density during the MD simulation process.

Then use the "gnuplot plot.gpl" to see the output:

```
guest@dell-Vostro-3671-China-HDD-Protection:~/BIM3019/120090473/BIM3019/NPT$ vi plot.gpl
guest@dell-Vostro-3671-China-HDD-Protection:~/BIM3019/120090473/BIM3019/NPT$ gnuplot plot.gpl
iter      chisq       delta/lim  lambda   avg
   0 1.0552371826e+08   0.00e+00  1.00e+00    1.000000e+00
   1 1.1488510116e+04  -9.18e+08  1.00e-01    1.013122e+03
   2 1.3460292246e+03  -7.54e+05  1.00e-02    1.023142e+03
   3 1.3460291252e+03  -7.39e-03  1.00e-03    1.023143e+03
iter      chisq       delta/lim  lambda   avg

After 3 iterations the fit converged.
final sum of squares of residuals : 1346.03
rel. change during last iteration : -7.38518e-08

degrees of freedom    (FIT_NDF)                        : 100
rms of residuals      (FIT_STDFIT) = sqrt(WSSR/ndf)    : 3.66883
variance of residuals (reduced chisquare) = WSSR/ndf   : 13.4603

Final set of parameters            Asymptotic Standard Error
=======================            ==========================
avg             = 1023.14          +/- 0.3651       (0.03568%)
```

Thus the average value of density should be 1023.14 (kg/m^3).

Basically, the results agree with my expectations.

Resource availability:

All the content of the report can be found in Github website:

(https://github.com/HULinfengHideki/BIM3019_Gromacs_MD_Tutorial).

All of the content is obtained by WinSCP, which can be found in computer terminal in school. A exception is "npt.trr" is so large that it cannot be update normally.