Computational Biology Lab

BE21B037 Practical 5

Task 1. Generate topology file from the pdb file using charmm27 forcefield.

topol.top

```
;
;       File 'topol.top' was generated
;       By user: unknown (1000)
;       On host: LAPTOP-LE18RADI
;       At date: Wed Sep  4 15:10:21 2024
;
;       This is a standalone topology file
;
;       Created by:
;                  :-) GROMACS - gmx pdb2gmx, 2020.1-Ubuntu-2020.1-1 (-:
;
;       Executable:   /usr/bin/gmx
;       Data prefix:  /usr
;       Working dir:  /home/harsha/CBL/Practical_5
;       Command line:
;         gmx pdb2gmx -f 1ubq_cleaner.pdb -o 1ubq_processed.gro
;       Force field was read from the standard GROMACS share directory.
;

; Include forcefield parameters
#include "charmm27.ff/forcefield.itp"

[ moleculetype ]
; Name            nrexcl
Protein_chain_A     3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass  typeB    chargeB      massB
; residue   1 MET rtp MET  q +1.0
     1         NH3      1    MET      N      1       -0.3     14.007
     2          HC      1    MET     H1      2       0.33      1.008
     3          HC      1    MET     H2      3       0.33      1.008
     4          HC      1    MET     H3      4       0.33      1.008
     5         CT1      1    MET     CA      5       0.21     12.011
     6          HB      1    MET     HA      6        0.1      1.008
     7         CT2      1    MET     CB      7      -0.18     12.011
     8          HA      1    MET    HB1      8       0.09      1.008
     9          HA      1    MET    HB2      9       0.09      1.008
    10         CT2      1    MET     CG     10      -0.14     12.011
    11          HA      1    MET    HG1     11       0.09      1.008
    12          HA      1    MET    HG2     12       0.09      1.008
    13           S      1    MET     SD     13      -0.09      32.06
```

Ln 1, Col 1 | 3,42,006 characters | 100% | Unix (LF) | UTF-8

Q1a. Compare the gro and pdb files. Comment on the units of distance used in both the formats.

1ubq.pdb

```
REMARK 500    ARG A  54   NE  -  CZ  -  NH1 ANGL. DEV. =   5.5 DEGREES
REMARK 500
REMARK 500 REMARK: NULL
DBREF  1UBQ A    1    76  UNP    P62988   UBIQ_HUMAN       1     76
SEQRES   1 A   76  MET GLN ILE PHE VAL LYS THR LEU THR GLY LYS THR ILE
SEQRES   2 A   76  THR LEU GLU VAL GLU PRO SER ASP THR ILE GLU ASN VAL
SEQRES   3 A   76  LYS ALA LYS ILE GLN ASP LYS GLU GLY ILE PRO PRO ASP
SEQRES   4 A   76  GLN GLN ARG LEU ILE PHE ALA GLY LYS GLN LEU GLU ASP
SEQRES   5 A   76  GLY ARG THR LEU SER ASP TYR ASN ILE GLN LYS GLU SER
SEQRES   6 A   76  THR LEU HIS LEU VAL LEU ARG LEU ARG GLY GLY
FORMUL   2  HOH   *58(H2 O)
HELIX    1  H1 ILE A   23  GLU A   34  1                                  12
HELIX    2  H2 LEU A   56  TYR A   59  5                                   4
SHEET    1 BET 5 GLY A  10  VAL A  17  0
SHEET    2 BET 5 MET A   1  THR A   7 -1
SHEET    3 BET 5 GLU A  64  ARG A  72  1
SHEET    4 BET 5 GLN A  40  PHE A  45 -1
SHEET    5 BET 5 LYS A  48  LEU A  50 -1
CRYST1   50.840   42.770   28.950  90.00  90.00  90.00 P 21 21 21    4
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.019670  0.000000  0.000000        0.00000
SCALE2      0.000000  0.023381  0.000000        0.00000
SCALE3      0.000000  0.000000  0.034542        0.00000
ATOM      1  N   MET A   1      27.340  24.430   2.614  1.00  9.67           N
ATOM      2  CA  MET A   1      26.266  25.413   2.842  1.00 10.38           C
ATOM      3  C   MET A   1      26.913  26.639   3.531  1.00  9.62           C
ATOM      4  O   MET A   1      27.886  26.463   4.263  1.00  9.62           O
ATOM      5  CB  MET A   1      25.112  24.880   3.649  1.00 13.77           C
ATOM      6  CG  MET A   1      25.353  24.860   5.134  1.00 16.29           C
ATOM      7  SD  MET A   1      23.930  23.959   5.904  1.00 17.17           S
```

Ln 1, Col 1 | 78,570 characters | 100% | Unix (LF) | UTF-8

1ubq_processed.gro

```
UBIQUITIN
 1231
    1MET      N    1   2.734   2.443   0.261
    1MET     H1    2   2.696   2.362   0.217
    1MET     H2    3   2.804   2.483   0.203
    1MET     H3    4   2.775   2.417   0.349
    1MET     CA    5   2.627   2.541   0.284
    1MET     HA    6   2.587   2.562   0.195
    1MET     CB    7   2.511   2.488   0.365
    1MET    HB1    8   2.431   2.545   0.347
    1MET    HB2    9   2.493   2.394   0.335
    1MET     CG   10   2.535   2.486   0.513
    1MET    HG1   11   2.624   2.441   0.526
    1MET    HG2   12   2.542   2.582   0.541
    1MET     SD   13   2.393   2.396   0.590
    1MET     CE   14   2.445   2.398   0.762
    1MET    HE1   15   2.376   2.352   0.818
    1MET    HE2   16   2.532   2.351   0.771
    1MET    HE3   17   2.454   2.493   0.793
    1MET      C   18   2.691   2.664   0.353
    1MET      O   19   2.789   2.646   0.426
    2GLN      N   20   2.634   2.777   0.326
    2GLN     HN   21   2.556   2.780   0.263
    2GLN     CA   22   2.685   2.902   0.390
    2GLN     HA   23   2.782   2.895   0.413
    2GLN     CB   24   2.673   3.015   0.290
    2GLN    HB1   25   2.744   3.001   0.221
    2GLN    HB2   26   2.583   3.008   0.248
    2GLN     CG   27   2.688   3.155   0.341
    2GLN    HG1   28   2.616   3.172   0.408
    2GLN    HG2   29   2.777   3.163   0.386
```

Ln 1, Col 1 | 55,442 characters | 100% | Unix (LF) | UTF-8

The pdb and gro files differ in the following aspects

1) Coordinates in pdb are (x,y,z) and in gro files they are the velocities in each direction
2) The pdb file contains the headers, remarks and other metadata, and atoms are grouped by residues and each atom has a specific format, while the gro files contains only the list of atoms with velocities and at the end the box dimensions.

The distances in the PDB file are measured in Angstroms, while the gro file the distances are expressed in nanometers.

Q1b. What is the net charge on the protein?

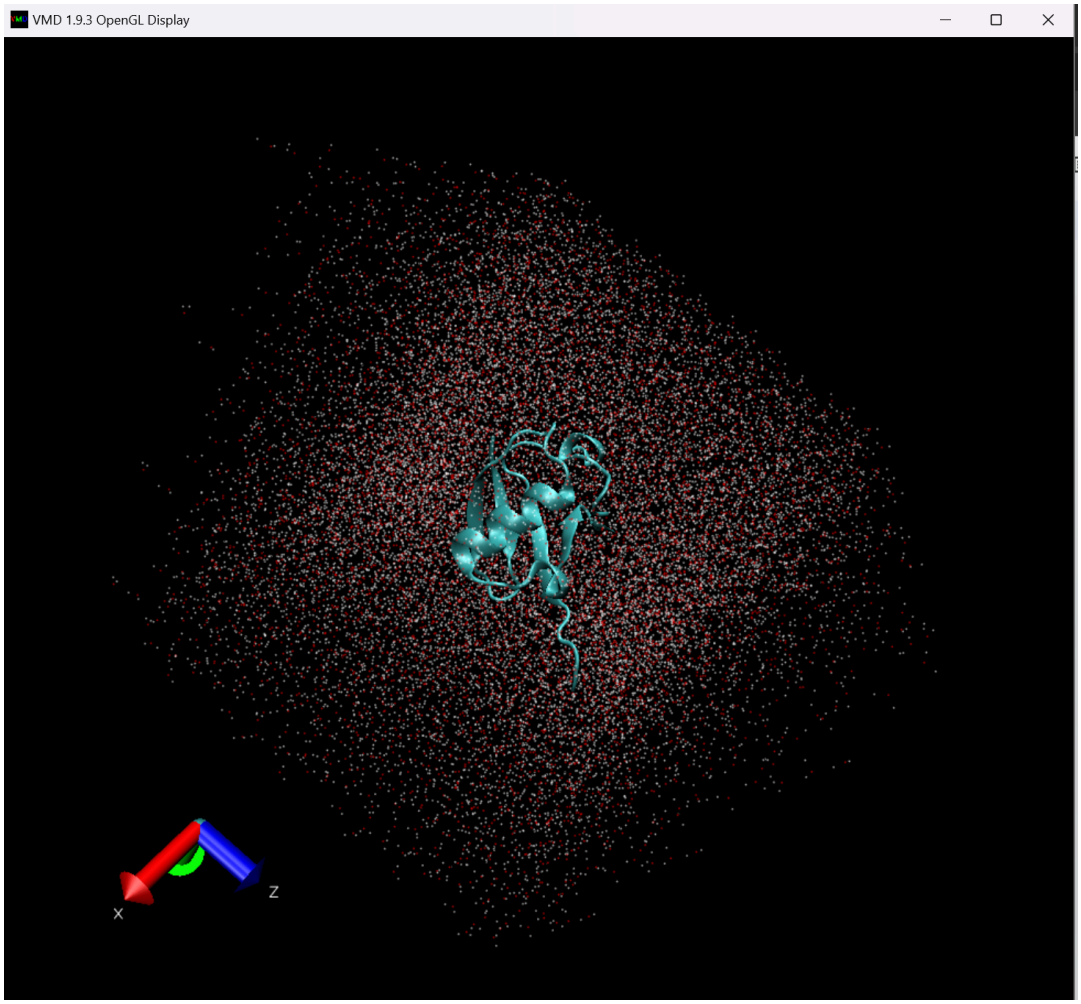
There is zero charge in the protein

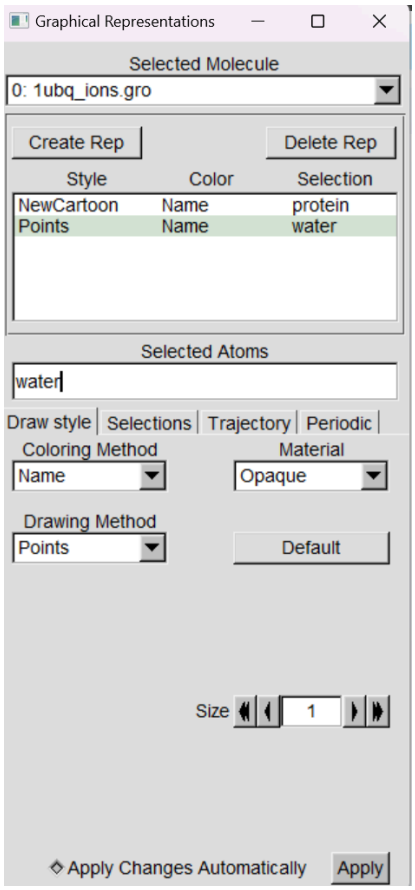
```
Total mass 1044.893 a.m.u.
Total charge 0.000 e
Including chain 1 in system: 1231 atoms 76 residues
Including chain 2 in system: 174 atoms 58 residues
Now there are 1405 atoms and 134 residues
Total mass in system 9609.761 a.m.u.
Total charge in system 0.000 e

Writing coordinate file...
                 --------- PLEASE NOTE ------------
You have successfully generated a topology from: 1ubq.pdb.
The Charmm27 force field and the tip3p water model are used.
                 --------- ETON ESAELP ------------

GROMACS reminds you: "Every Sperm is Sacred" (Monty Python)
```

Task 2. Create a cubic box around the protein and solvate it in water. Neutralize your system by adding ions to counter the charge.

The protein had no charge, and we solvated using TIP3P water molecules.

Q2a. Open the system in VMD. Provide a snapshot of the protein present in the water box with ions (protein: in New Cartoon representation, water: in Points representation, ions: in VDW).


VMD 1.9.3 OpenGL Display
X
Z


Graphical Representations
Selected Molecule
0: 1ubq_ions.gro
Create Rep
Delete Rep
Style
Color
Selection
NewCartoon
Name
protein
Points
Name
water
Selected Atoms
water
Draw style
Selections
Trajectory
Periodic
Coloring Method
Name
Material
Opaque
Drawing Method
Points
Default
Size
1
Apply Changes Automatically
Apply

Since there were no ions present we did not represent them

Q2b. What is the volume of the box generated? How many water molecules were added at this step?

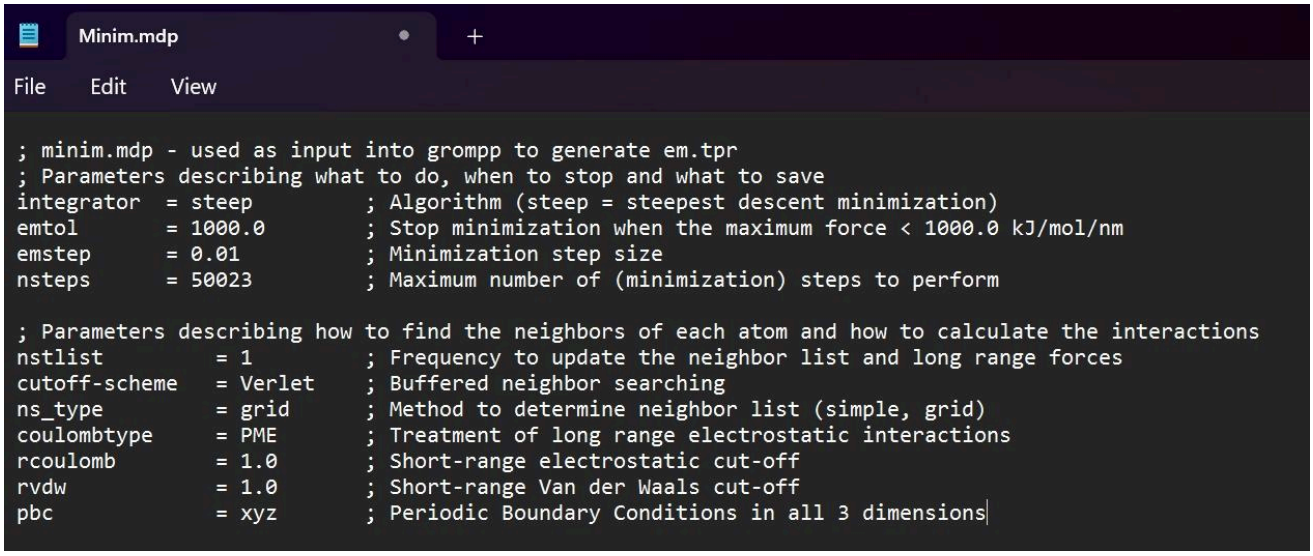
Volume of box = 62.95 nm^3, No of water molecules added = 8472

```
Note that major changes are planned in future for editconf, to
Read 1405 atoms
Volume: 62.9497 nm^3, corresponds to roughly 28300 electrons
No velocities found
    system size :  3.168  3.487  3.857 (nm)
    diameter    :  4.491               (nm)
    center      :  3.025  2.891  1.503 (nm)
    box vectors :  5.084  4.277  2.895 (nm)
    box angles  :  90.00  90.00  90.00 (degrees)
    box volume  :  62.95               (nm^3)
    shift       :  0.221  0.355  1.743 (nm)
new center      :  3.246  3.246  3.246 (nm)
new box vectors :  6.491  6.491  6.491 (nm)
new box angles  :  90.00  90.00  90.00 (degrees)
new box volume  : 273.54               (nm^3)

GROMACS reminds you: "There is no place like ~" (Anonymous)
```

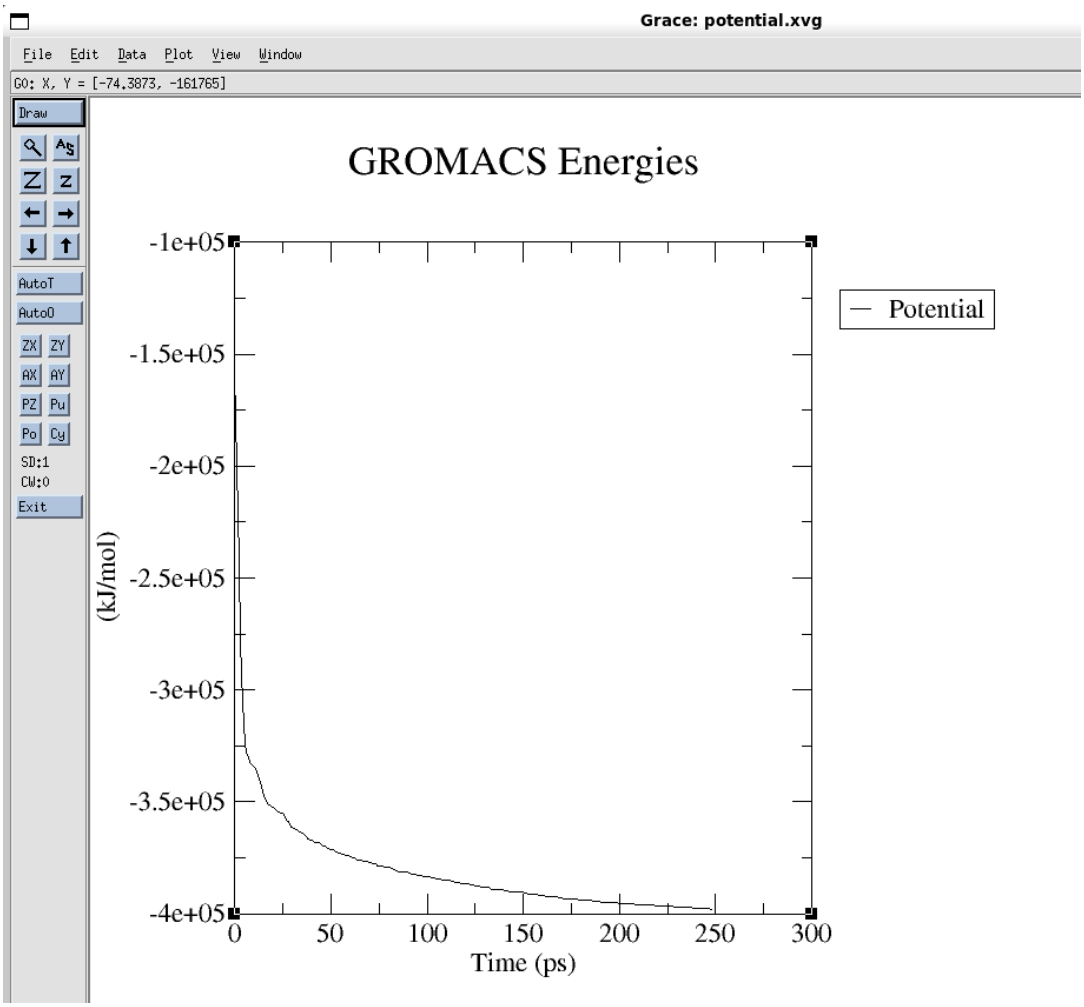
```
Writing generated configuration to 1ubq_solv.gro

Output configuration contains 26821 atoms in 8606 residues
Volume                 :     273.543 (nm^3)
Density                :     988.617 (g/l)
Number of solvent molecules:   8472

Processing topology
Adding line for 8472 solvent molecules with resname (SOL) to topology file (topol.top)

Back Off! I just backed up topol.top to ./#topol.top.1#
```

Task 3. Edit the mdp file and perform energy minimization for 50023 steps.

Q3a. Provide a screenshot of the edited mdp file.



Q3b. Plot potential vs time for your system in XM Grace. In how many steps did the system converge to its least energy state?




```
writing lowest energy coordinates.

Steepest Descents converged to Fmax < 1000 in 249 steps
Potential Energy  = -3.9795725e+05
Maximum force     =  9.6786945e+02 on atom 913
Norm of force     =  3.8827768e+01

GROMACS reminds you: "I'm Not Gonna Die Here !" (Sphere)
```

Steps: 249

Task 4. Perform NVT equilibration for the system at temperature 310 K.

NVT.mdp

File Edit View

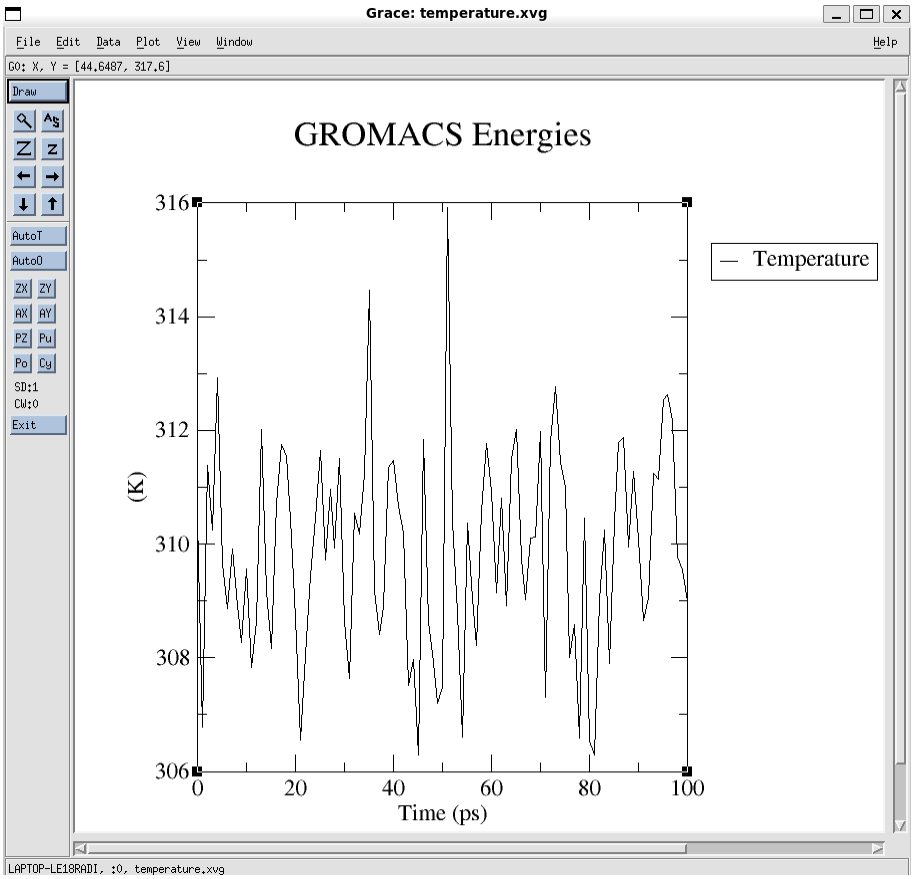
```
title                    = OPLS Lysozyme NVT equilibration
define                   = -DPOSRES  ; position restrain the protein
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 50000     ; 2 * 50000 = 100 ps
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 500       ; save coordinates every 1.0 ps
nstvout                  = 500       ; save velocities every 1.0 ps
nstenergy                = 500       ; save energies every 1.0 ps
nstlog                   = 500       ; update log file every 1.0 ps
; Bond parameters
continuation             = no        ; first dynamics run
constraint_algorithm     = lincs     ; holonomic constraints
constraints              = h-bonds   ; bonds involving H are constrained
lincs_iter               = 1         ; accuracy of LINCS
lincs_order              = 4         ; also related to accuracy
; Nonbonded settings
cutoff-scheme            = Verlet    ; Buffered neighbor searching
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 10        ; 20 fs, largely irrelevant with Verlet
rcoulomb                 = 1.0       ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0       ; short-range van der Waals cutoff (in nm)
DispCorr                 = EnerPres  ; account for cut-off vdW scheme
; Electrostatics
coulombtype              = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling is on
tcoupl                   = V-rescale             ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                    = 0.1     0.1           ; time constant, in ps
ref_t                    = 310     310           ; reference temperature, one for each group, in K
; Pressure coupling is off
pcoupl                   = no        ; no pressure coupling in NVT
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
; Velocity generation
gen_vel                  = yes       ; assign velocities from Maxwell distribution
gen_temp                 = 310       ; temperature for Maxwell distribution
gen_seed                 = -1        ; generate a random seed
```

Q4a. What is the average temperature at the end of the run?

```
Statistics over 50001 steps [ 0.0000 through 100.0000 ps ], 1 data sets
All statistics are over 501 points

Energy                      Average   Err.Est.       RMSD  Tot-Drift
-------------------------------------------------------------------------------
Temperature                 309.713       0.16    2.86559   0.901137  (K)

GROMACS reminds you: "I Was Born to Have Adventure" (F. Zappa)
```

Average Temperature: 309.713K

Q4b. Provide a plot of temperature vs time.



Task 5. Perform NPT equilibration at temperature 310 K and 1 bar pressure.

NPT.mdp

File Edit View

```
title                    = OPLS Lysozyme NPT equilibration
define                   = -DPOSRES  ; position restrain the protein
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 50000     ; 2 * 50000 = 100 ps
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 500       ; save coordinates every 1.0 ps
nstvout                  = 500       ; save velocities every 1.0 ps
nstenergy                = 500       ; save energies every 1.0 ps
nstlog                   = 500       ; update log file every 1.0 ps
; Bond parameters
continuation             = yes       ; Restarting after NVT
constraint_algorithm     = lincs     ; holonomic constraints
constraints              = h-bonds   ; bonds involving H are constrained
lincs_iter               = 1         ; accuracy of LINCS
lincs_order              = 4         ; also related to accuracy
; Nonbonded settings
cutoff-scheme            = Verlet    ; Buffered neighbor searching
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb                 = 1.0       ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0       ; short-range van der Waals cutoff (in nm)
DispCorr                 = EnerPres  ; account for cut-off vdW scheme
; Electrostatics
coulombtype              = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling is on
tcoupl                   = V-rescale             ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                    = 0.1     0.1           ; time constant, in ps
ref_t                    = 310     310           ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                   = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype               = isotropic             ; uniform scaling of box vectors
tau_p                    = 2.0                   ; time constant, in ps
ref_p                    = 1.0                   ; reference pressure, in bar
compressibility          = 4.5e-5                ; isothermal compressibility of water, bar^-1
refcoord_scaling         = com
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
```

Q5a. What is the average pressure at the end of the run?

```
Statistics over 50001 steps [ 0.0000 through 100.0000 ps ], 1 data sets
All statistics are over 501 points

Energy                      Average   Err.Est.       RMSD  Tot-Drift
-------------------------------------------------------------------------------
------
Pressure                   -3.32408          4    144.294  -0.499317  (ba
r)

GROMACS reminds you: "Only entropy comes easy." (Anton Chekov)
```

Average Pressure : -3.32408 Bar

Q5b. Provide a plot of pressure vs time.