Running molecular dynamics simulations using GROMACS

Software's required

GROMACS

VMD (Visual Molecular Dynamics)

Jupyter Notebook

Introduction

Molecular dynamics (MD) is a method to simulate molecular motion by iterative application of Newton's laws of motion. It is often applied to large biomolecules such as proteins or nucleic acids.

Multiple packages exist for performing MD simulations. One of the most popular is the open-source GROMACS, which is the subject of this tutorial.

Process

Prior to performing simulations, a number of preparatory steps need to be executed.

The process can be divided into multiple stages:

1. Setup (loading data, solvation i.e. addition of water and ions)
2. Energy minimization of the protein
3. Equilibration of the solvent around the protein (with two ensembles, NVT and NPT)
4. Production simulation, which produces our trajectory
5. Trajectory Analysis.

Getting data

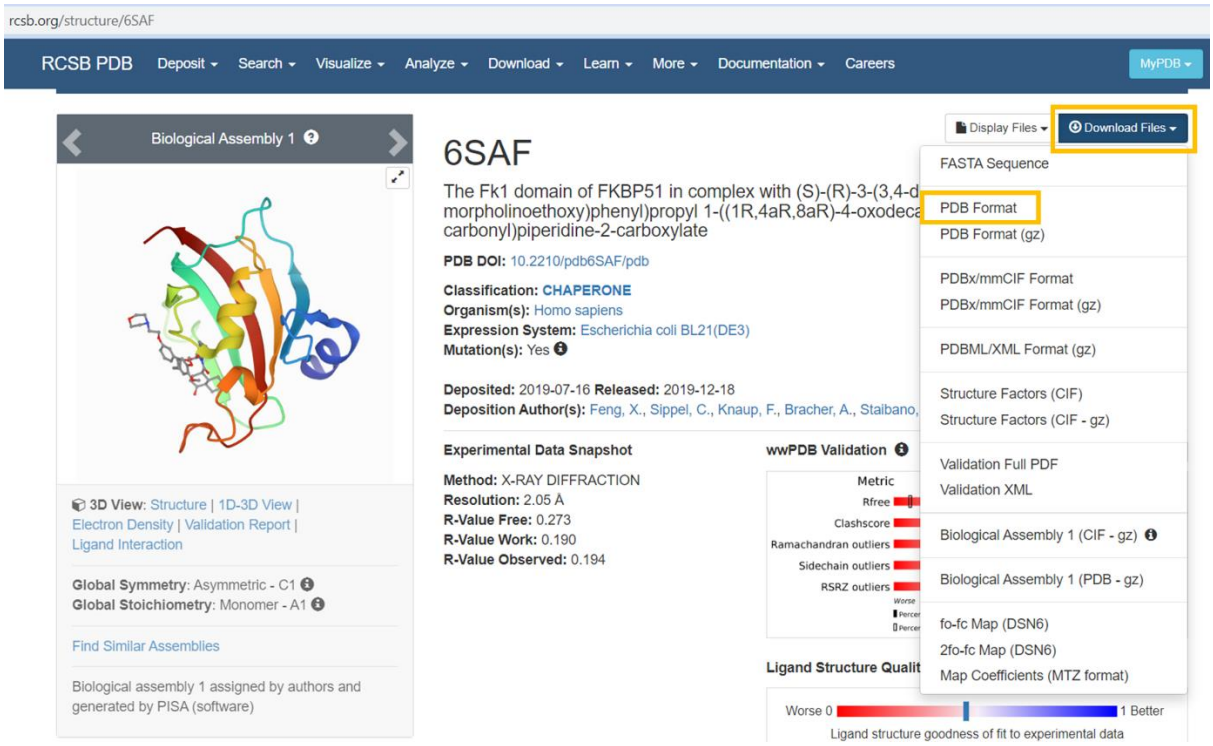
To perform simulation, an initial PDB file is required. This should be 'cleaned' of solvent and any other non-protein atoms. Solvent will be re-added in a subsequent step.

Download a PDB structure file from the Protein Data Bank and remove the unwanted atoms using the grep text processing tool. This simply removes the lines in the PDB file that refer to the unwanted atoms.

6SAF

Crystal structure of the The Fk1 domain of FKBP51 in complex with (S)-(R)-3-(3,4-dimethoxyphenyl)-1-(3-(2-morpholinoethoxy)phenyl)propyl 1-((1R,4aR,8aR)-4-oxodecahydronaphthalene-1-carbonyl)piperidine-2-carboxylate.

Download Receptor file from https://www.rcsb.org/structure/6SAF. On the RCSB website enzyme structure is given with 4 letter code. Click on "**Downloads Files**" and download the "**PDB Format**" file.

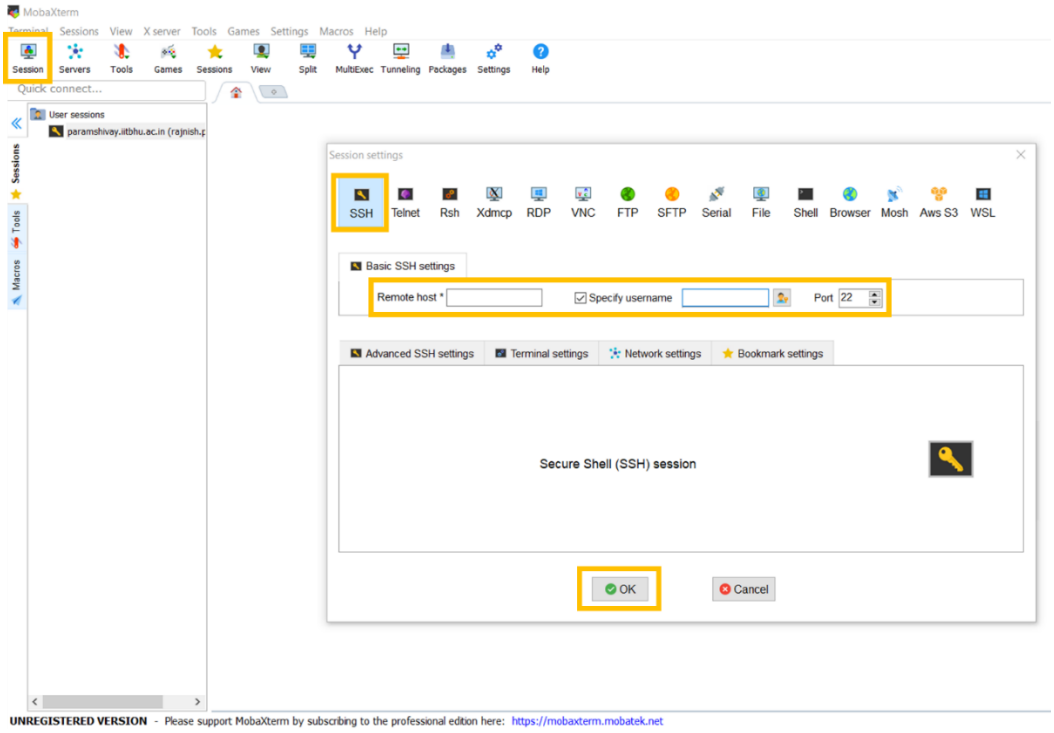


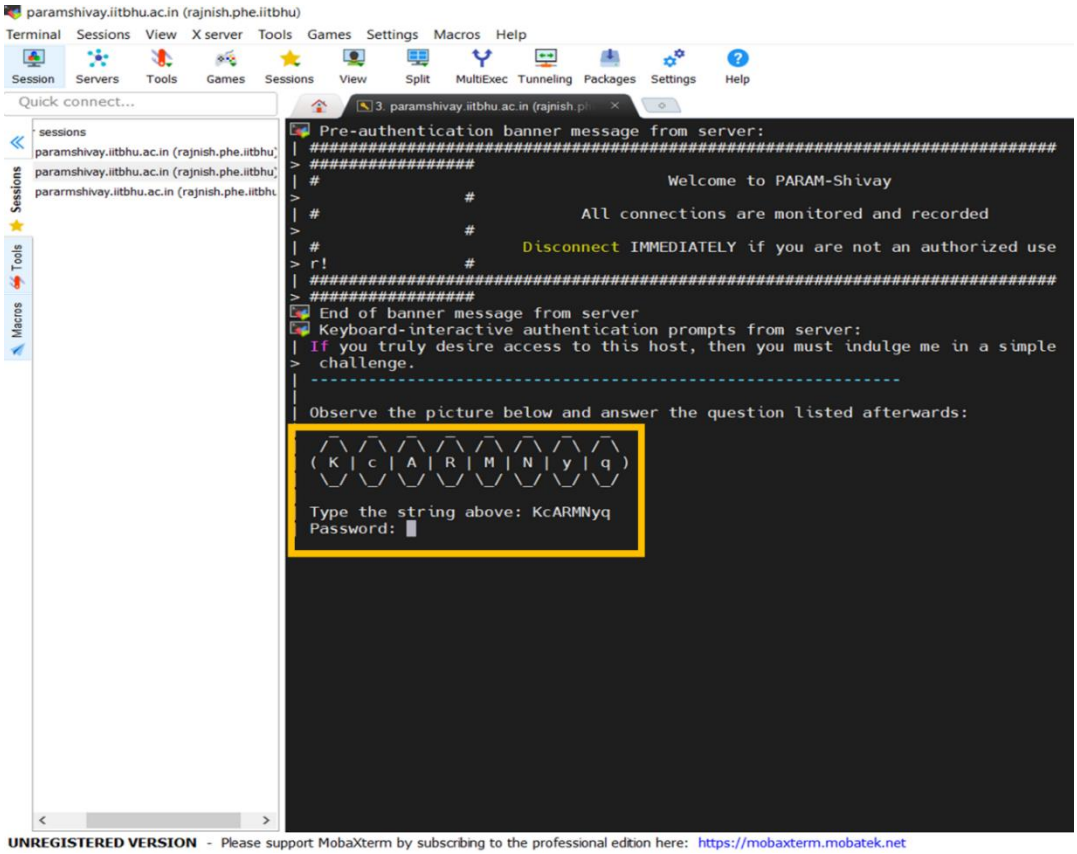
High Performance Computing (HPC)

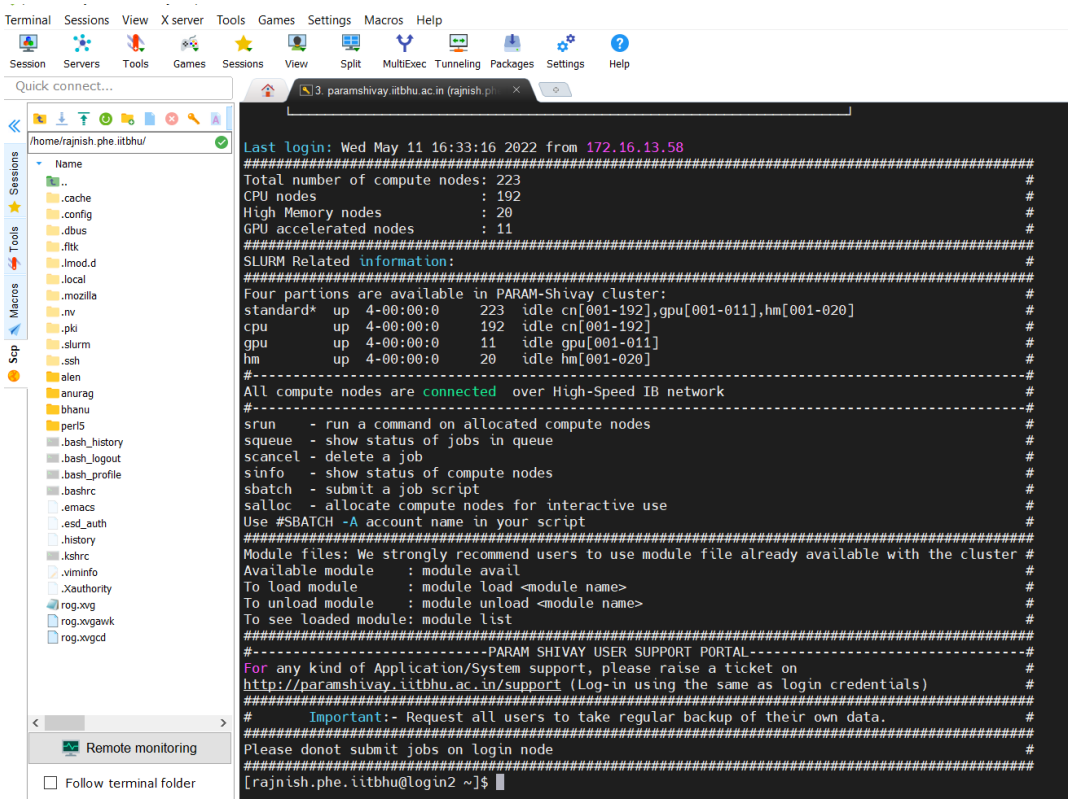
PARAM Shivay

The supercomputer of 837 TFLOPS capacity, built at the cost of Rs 32.5 crore under the National Super Computing Mission at Indian Institute of Technology (IIT), Banaras Hindu University (BHU).

Login


MobaXterm
Session
Servers
Tools
Games
Sessions
View
Split
MultiExec
Tunneling
Packages
Settings
Help
Quick connect...
User sessions
paramshivay.iitbhu.ac.in (rajnish.p
Session settings
SSH
Telnet
Rsh
Xdmcp
RDP
VNC
FTP
SFTP
Serial
File
Shell
Browser
Mosh
Aws S3
WSL
Basic SSH settings
Remote host *
Specify username
Port 22
Advanced SSH settings
Terminal settings
Network settings
Bookmark settings
Secure Shell (SSH) session
OK
Cancel
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net


paramshivay.iitbhu.ac.in (rajnish.phe.iitbhu)
Terminal
Sessions
View
X server
Tools
Games
Settings
Macros
Help
Session
Servers
Tools
Games
Sessions
View
Split
MultiExec
Tunneling
Packages
Settings
Help
Quick connect...
sessions
paramshivay.iitbhu.ac.in (rajnish.phe.iitbhu)
paramshivay.iitbhu.ac.in (rajnish.phe.iitbhu)
pararmshivay.iitbhu.ac.in (rajnish.phe.iitbhu)
Pre-authentication banner message from server:
Welcome to PARAM-Shivay
All connections are monitored and recorded
Disconnect IMMEDIATELY if you are not an authorized use
End of banner message from server
Keyboard-interactive authentication prompts from server:
If you truly desire access to this host, then you must indulge me in a simple challenge.
Observe the picture below and answer the question listed afterwards:
(K | c | A | R | M | N | y | q)
Type the string above: KcARMNyq
Password:
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net



Running Interactive Jobs

In general, the jobs can be run in an interactive manner or in batch mode.

You can run an interactive job as follows:

The following command asks for a single core on one hour with default amount of memory.

```
$ srun --nodes=1 --ntasks-per-node=40 --time=01:00:00 --pty bash -i
```

Setup

The GROMACS initial setup tool uses the PDB input to create three files which will be required for MD simulation.

```
grep -v HOH 6saf.pdb > 6saf_clean.pdb
```

```
gmx pdb2gmx -f 6saf_clean.pdb -o 6saf_processed.gro -water spce -ignh
```

In summary, the initial setup tool will:

- create a 'topology' file
- convert a PDB protein structure into a GRO file, with the structure centered in a simulation box (unit cell)
- create a position restraint file

After these files have been generated, a further step is required to define a simulation box (unit cell) in which the simulation can take place.

This can be done with the **GROMACS structure configuration** tool. It also defines the unit cell 'box', centered on the structure.

```
gmx editconf -f 6saf_processed.gro -o 6saf_newbox.gro -c -d 1.0 -bt cubic
```

Options include box dimensions and shape; here, while a cuboidal box may be most intuitive, rhombic dodecahedron is the most efficient option, as it can contain the protein using the smallest volume, thus reducing the simulation resources devoted to the solvent.

Solvation

The next stage is protein solvation, performed using **GROMACS solvation and adding ions** tool.

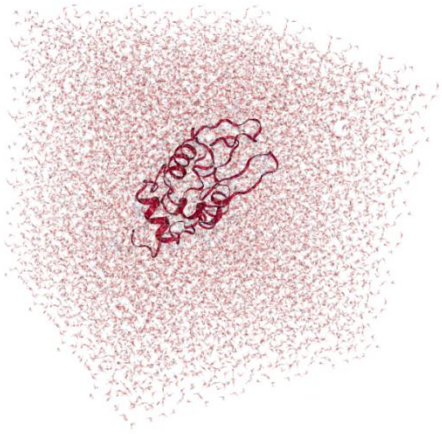
```
gmx solvate -cp 6saf_newbox.gro -cs spc216.gro -o 6saf_solv.gro -p topol.top
```

Water molecules are added to the structure and topology files to fill the unit cell.

At this stage sodium or chloride ions are also automatically added to neutralize the charge of the system.

```
gmx grompp -f ions.mdp -c 6saf_solv.gro -p topol.top -o ions.tpr
```

gmx genion -s ions.tpr -o 6saf_solv_ions.gro -p topol.top -pname NA -nname CL -neutral



Energy minimization

To remove any steric clashes or unusual geometry which would artificially raise the energy of the system, we must relax the structure by running an energy minimization (EM) algorithm.

em.mdp - Notepad

File Edit Format View Help

```
; LINES STARTING WITH ';' ARE COMMENTS
title                = Minimization      ; Title of run

; Parameters describing what to do, when to stop and what to save
integrator           = steep             ; Algorithm (steep = steepest descent minimization)
emtol                = 1000.0    ; Stop minimization when the maximum force < 10.0 kJ/mol
emstep          = 0.01      ; Energy step size
nsteps               = 50000             ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist              = 1                 ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet
ns_type              = grid                  ; Method to determine neighbor list (simple, grid)
rlist                = 1.2                   ; Cut-off for making neighbor list (short range forces)
coulombtype          = PME                   ; Treatment of long range electrostatic interactions
rcoulomb             = 1.2                   ; long range electrostatic cut-off
vdwtype         = cutoff
vdw-modifier    = force-switch
rvdw-switch     = 1.0
rvdw                 = 1.2                   ; long range Van der Waals cut-off
pbc             = xyz               ; Periodic Boundary Conditions
DispCorr        = no
```

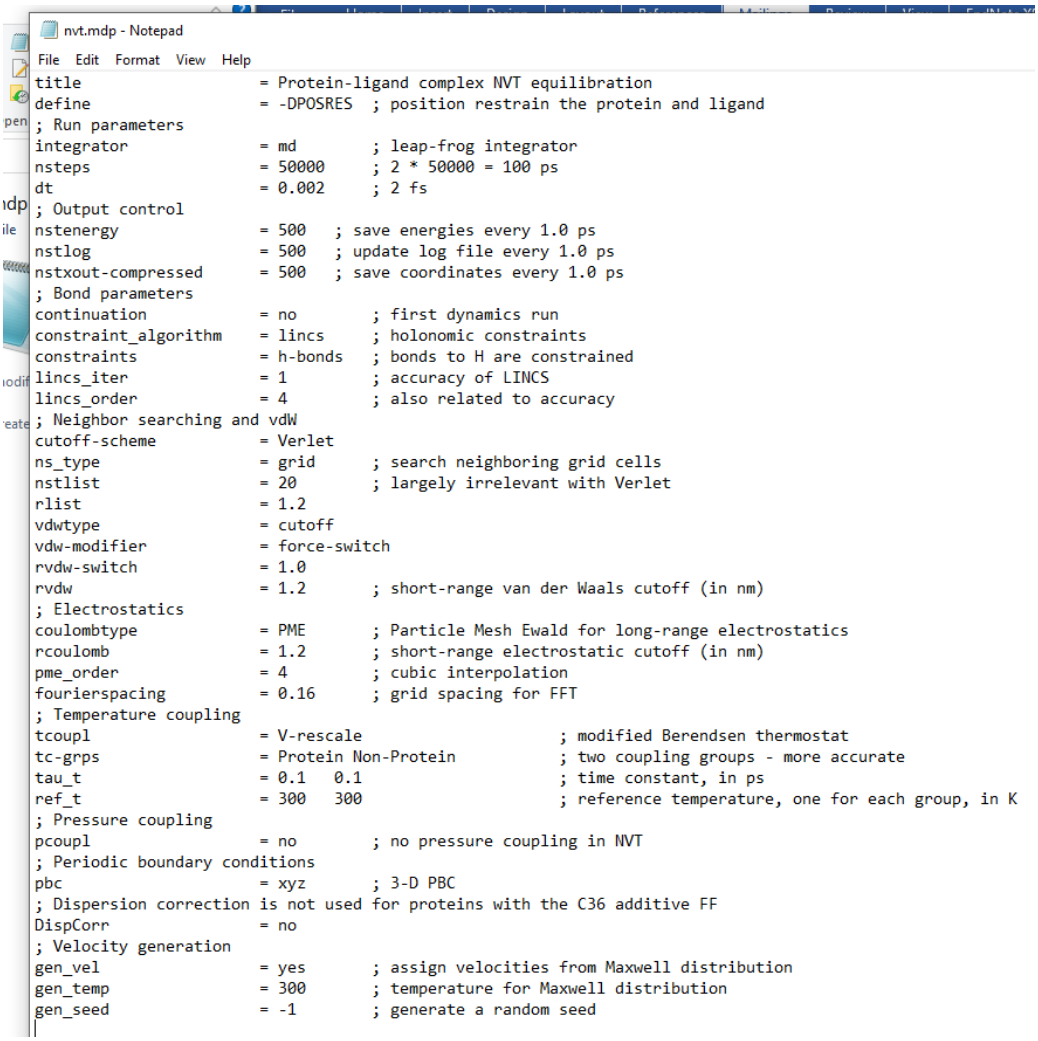
gmx grompp -f em.mdp -c 6saf_solv_ions.gro -p topol.top -o em.tpr

gmx mdrun -v -deffnm em

Equilibration

At this point equilibration of the solvent around the solute (i.e. the protein) is necessary. This is performed in two stages: equilibration under an NVT ensemble, followed by an NPT ensemble.

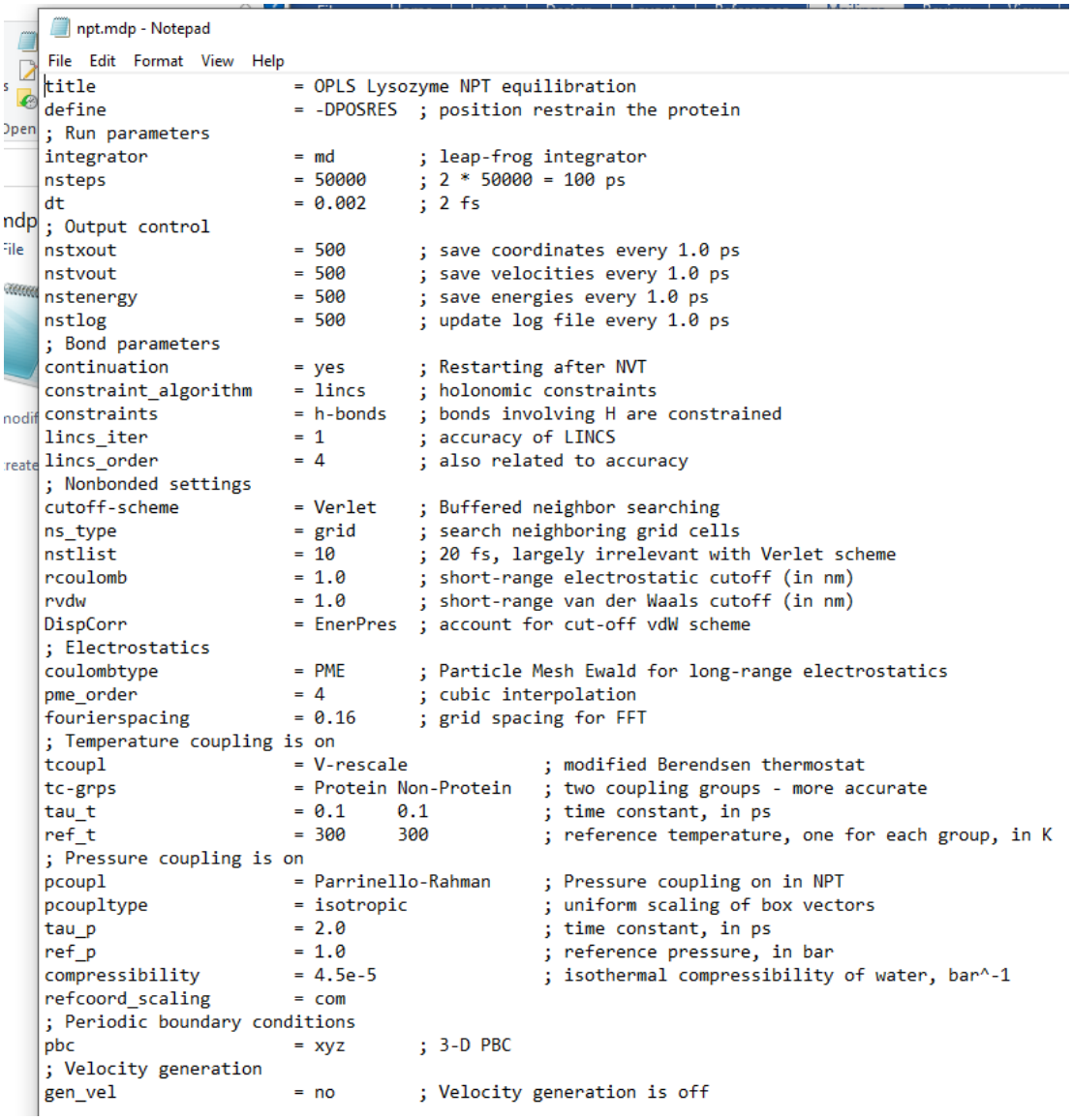
Use of the **NVT** ensemble entails maintaining constant **n**umber of particles, **v**olume and **t**emperature.



gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr

gmx mdrun -deffnm nvt

While the **NPT** ensemble maintains constant **n**umber of particles, **p**ressure and **t**emperature.



gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr

gmx mdrun -deffnm npt

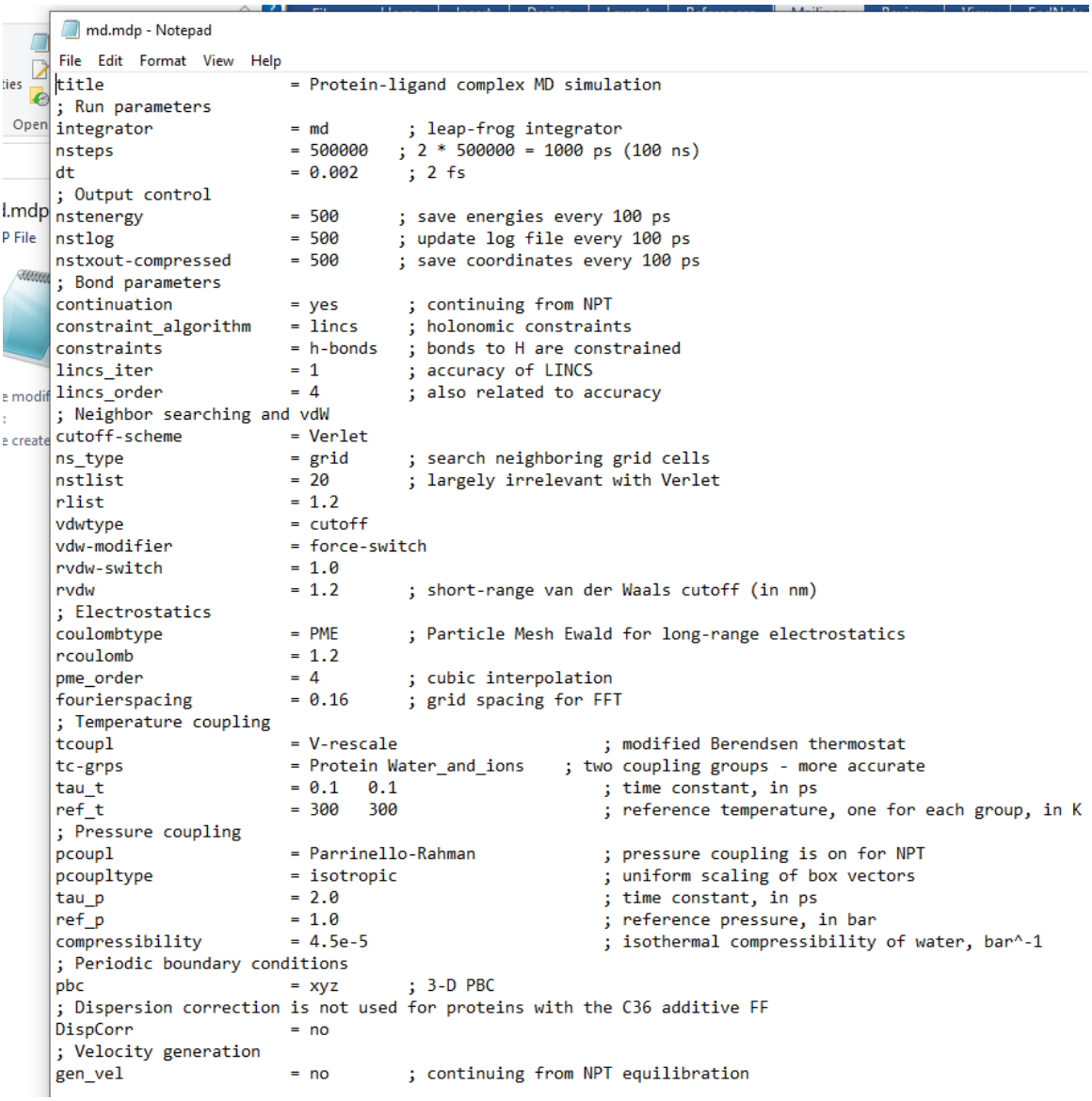
(The NVT ensemble is also known as the isothermal-isochoric ensemble, while the NPT ensemble is also known as the isothermal-isobaric ensemble).

During the first equilibration step (NVT), the protein must be held in place while the solvent is allowed to move freely around it. This is achieved using the position

restraint file we created in system setup. When we specify this restraint, protein movement is not totally forbidden, but is energetically punished. During the second NPT step, we remove the restraints.

Production simulation

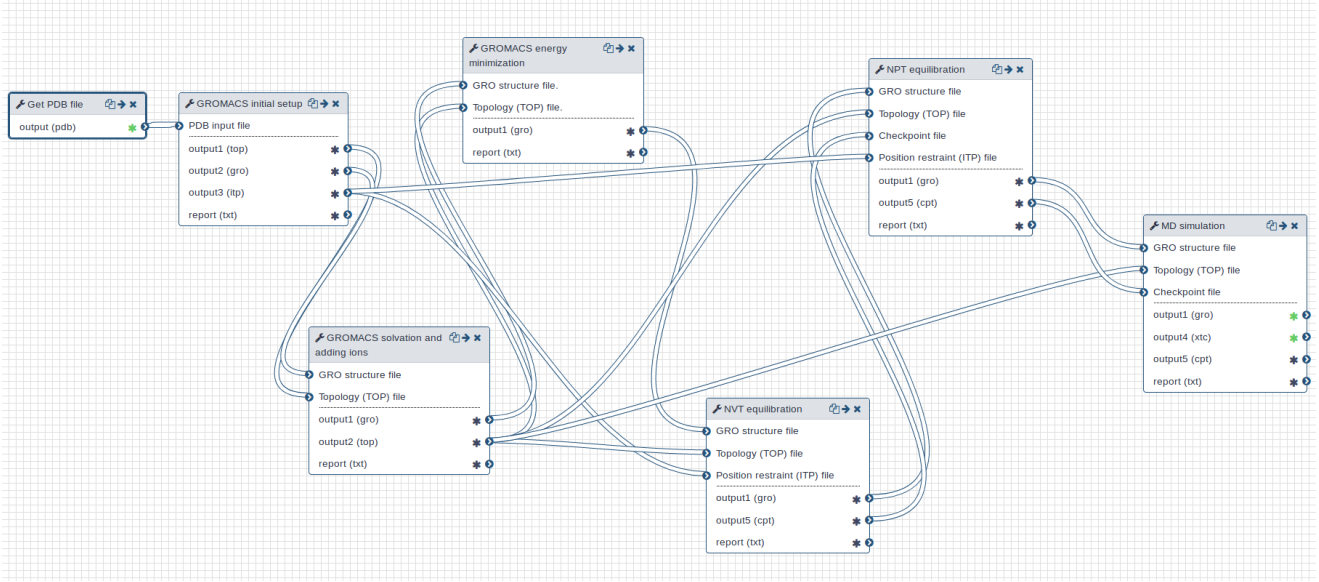
Now that equilibration is complete, we can release the position restraints. We are now finally ready to perform a production MD simulation.




```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
```

```
gmx mdrun -deffnm md
```

After completing the steps, or running the workflow, we have successfully produced a trajectory (the xtc file) which describes the atomic motion of the system. This can be viewed using molecular visualization software or analysed further.

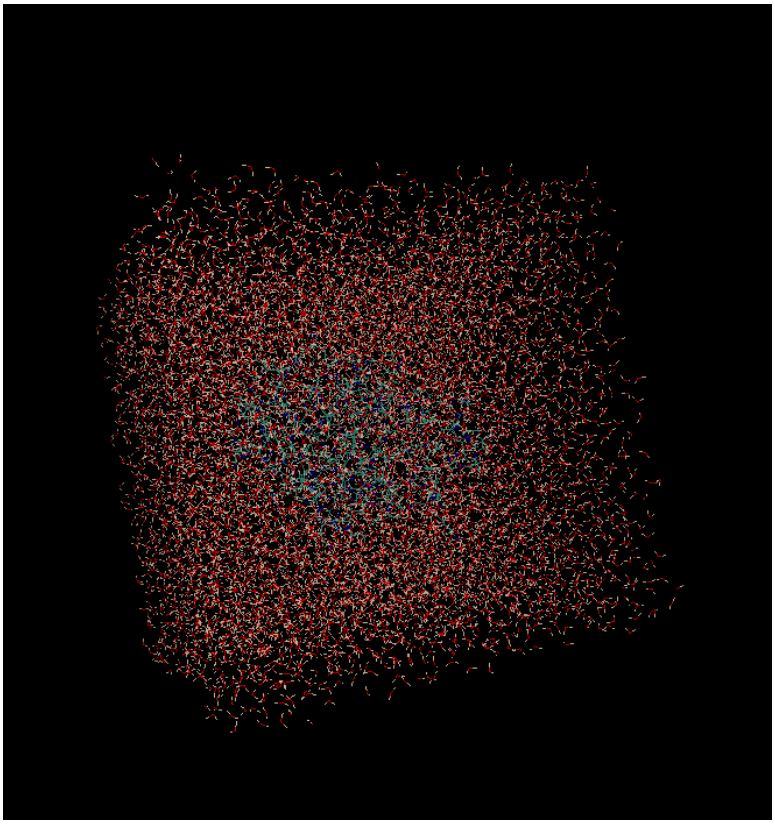


Recentering and Rewrapping Coordinates

As in any simulation conducted with periodic boundary conditions, molecules may appear "broken" or may "jump" back and forth across the box.

gmx trjconv -s md.tpr -f md.xtc -o md_center.xtc -center -pbc nojump -ur compact

gmx trjconv -s md.tpr -f md_center.xtc -o start.pdb -dump 0



Analysing the trajectory

After getting the trajectory file, we will analyse it using MDAnalysis tool.