

Molecular Dynamics Simulations using GROMACS

Software's required

1. GROMACS
2. VMD (Visual Molecular Dynamics)

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.

Methodology

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](https://www.rcsb.org/) 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

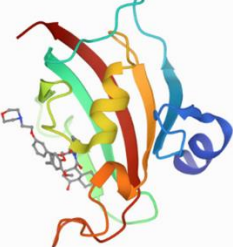
Crystal structure of the The Fk1 domain of FKBP51 in complex with (S)-(R)-3-(3,4-dimethoxyphenyl)-1-(3-(2-morpholinoethoxy)phenyl)propyl 1-((1R,4aR,8aR)-4-oxodecahydronaphthalene-1-carbonyl)piperidine-2-carboxylate.

Download pdb 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.

rcsb.org/structure/6SAF

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB

Biological Assembly 1



6SAF

The Fk1 domain of FKBP51 in complex with (S)-(R)-3-(3,4-dimorpholinoethoxy)phenylpropyl 1-((1R,4aR,8aR)-4-oxodecylidene-2-carboxylate)

PDB DOI: [10.2210/pdb6SAF/pdb](https://doi.org/10.2210/pdb6SAF/pdb)

Classification: **CHAPERONE**

Organism(s): Homo sapiens

Expression System: Escherichia coli BL21(DE3)

Mutation(s): Yes

Deposited: 2019-07-16 Released: 2019-12-18

Deposition Author(s): Feng, X., Sippel, C., Knaup, F., Bracher, A., Staibano, A.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.05 Å

R-Value Free: 0.273

R-Value Work: 0.190

R-Value Observed: 0.194

wwPDB Validation

Metric

Rfree

Clashscore

Ramachandran outliers

Sidechain outliers

RSRZ outliers

Validation Full PDF

Validation XML

Biological Assembly 1 (CIF - gz)

Biological Assembly 1 (PDB - gz)

fo-fc Map (DSN6)

2fo-fc Map (DSN6)

Map Coefficients (MTZ format)

Ligand Structure Quality

Worse 0 1 Better

Ligand structure goodness of fit to experimental data

Display Files Download Files

FASTA Sequence

PDB Format

PDB Format (gz)

PDBx/mmCIF Format

PDBx/mmCIF Format (gz)

PDBML/XML Format (gz)

Structure Factors (CIF)

Structure Factors (CIF - gz)

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

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

Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultExec Tunneling Packages Settings Help

Quick connect...

User sessions

paramshivay.iitbhu.ac.in (rajnish...)

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)
paramshivay.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>

```

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...

/home/rajnish.phe.iitbhu/

Last login: Wed May 11 16:33:16 2022 from 172.16.13.58
#####
Total number of compute nodes: 223
CPU nodes : 192
High Memory nodes : 20
GPU accelerated nodes : 11
#####
SLURM Related information:
#####
Four partions are available in PARAM-Shivay cluster:
standard* up 4-00:00:0 223 idle cn[001-192],gpu[001-011],hm[001-020]
cpu up 4-00:00:0 192 idle cn[001-192]
gpu up 4-00:00:0 11 idle gpu[001-011]
hm up 4-00:00:0 20 idle hm[001-020]
#
All compute nodes are connected over High-Speed IB network
#
srun - run a command on allocated compute nodes
squeue - show status of jobs in queue
scancel - delete a job
sinfo - show status of compute nodes
sbatch - submit a job script
salloc - allocate compute nodes for interactive use
Use #SBATCH -A account name in your script
#####
Module files: We strongly recommend users to use module file already available with the cluster #
Available module : module avail
To load module : module load <module name>
To unload module : module unload <module name>
To see loaded module: module list
#####
-----PARAM SHIVAY USER SUPPORT PORTAL-----
#
For any kind of Application/System support, please raise a ticket on
http://paramshivay.iitbhu.ac.in/support (log-in using the same as login credentials)
#####
# Important:- Request all users to take regular backup of their own data.
#####
Please donot submit jobs on login node
#####
[rajnish.phe.iitbhu@login2 ~]$

```

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=1 --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 1aki.pdb > 6saf_clean.pdb
```

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

In summary, the initial setup tool will:

1. create a 'topology' file
2. convert a PDB protein structure into a GRO file, with the structure centered in a simulation box (unit cell)
3. 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', centred 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.

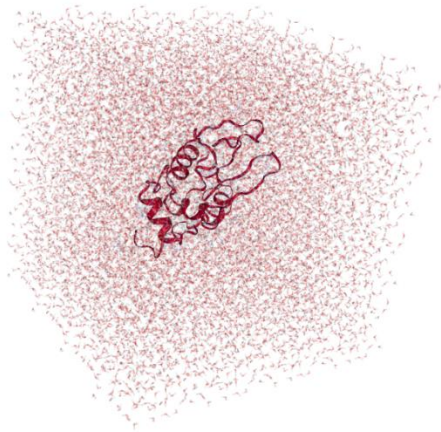
```
gmx solvate -cp 6saf_newbox.gro -cs spc216.gro -o 6saf_solv.gro -p topol.top
```

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 **number** of particles, **volume** and **temperature**.

```

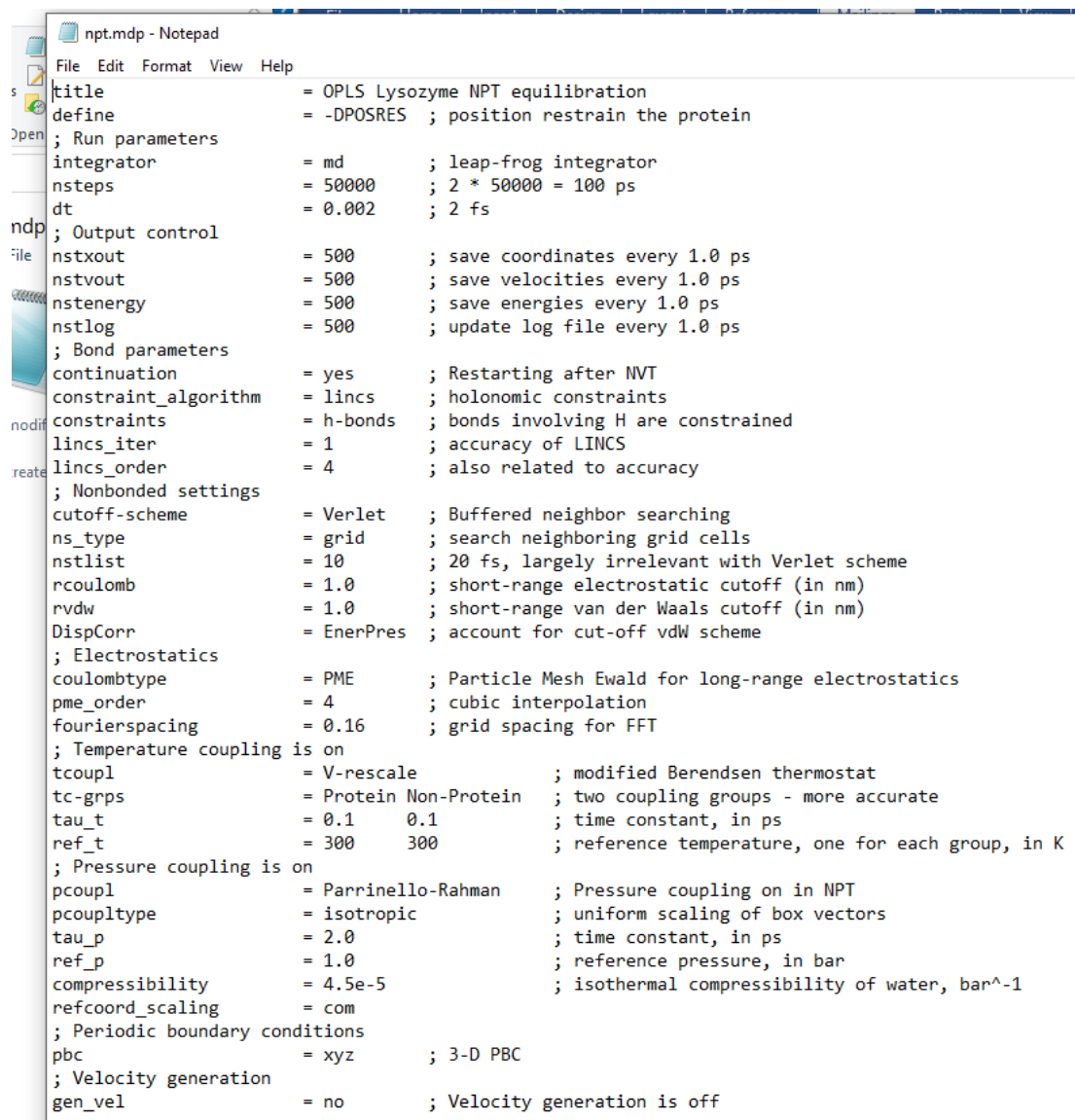
nvt.mdp - Notepad
File Edit Format View Help
title = Protein-ligand complex NVT equilibration
define = -DPOSRES ; position restrain the protein and ligand
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
nstxout-compressed = 500 ; save coordinates every 1.0 ps
; Bond parameters
continuation = no ; first dynamics run
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdw
cutoff-scheme = Verlet
ns_type = grid ; search neighboring grid cells
nstlist = 20 ; largely irrelevant with Verlet
rlist = 1.2
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
rcoulomb = 1.2 ; short-range electrostatic cutoff (in nm)
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
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 = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl = no ; no pressure coupling in NVT
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr = no
; Velocity generation
gen_vel = yes ; assign velocities from Maxwell distribution
gen_temp = 300 ; temperature for Maxwell distribution
gen_seed = -1 ; generate a random seed

```

`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 **number** of particles, **pressure** and **temperature**.



```

npt.mdp - Notepad
File Edit Format View Help
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
nstfout = 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 = 300 300 ; 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
; Velocity generation
gen_vel = no ; Velocity generation is off

```

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

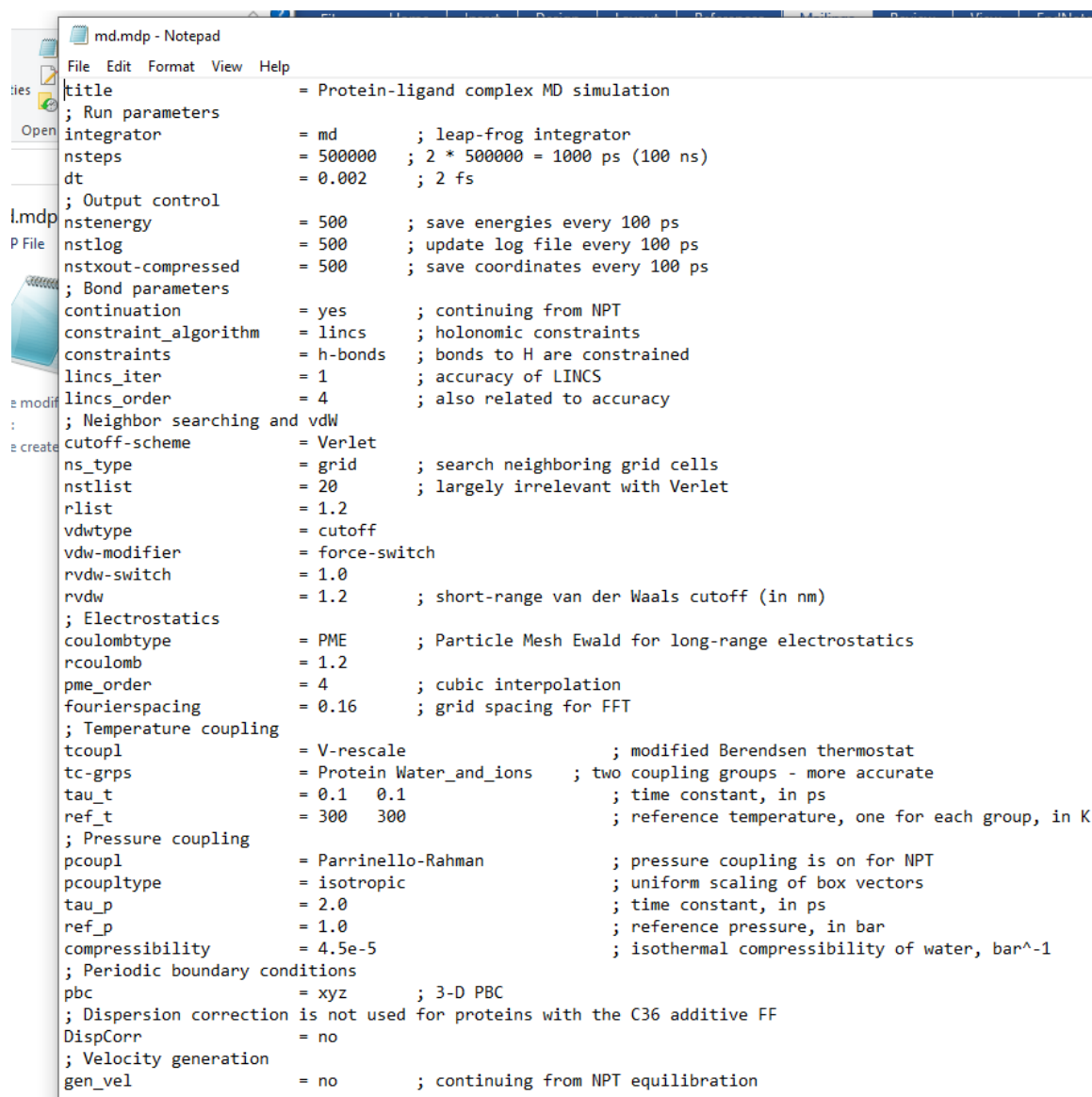
```
gmj mdrun -deffnm npt
```

(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.



```

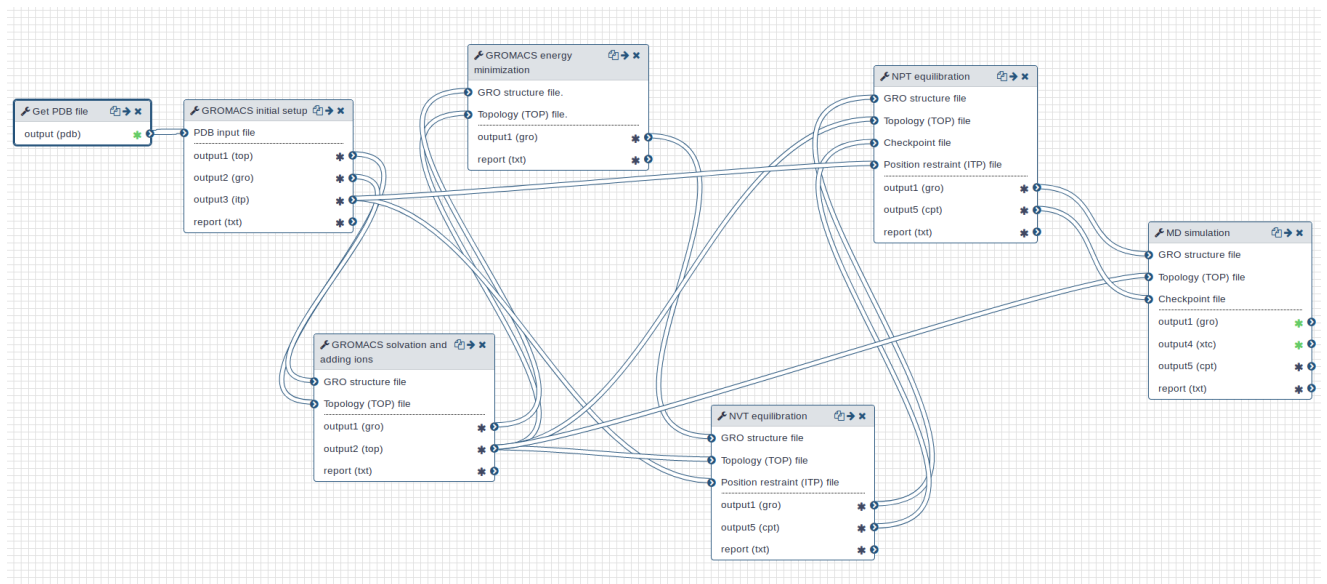
md.mdp - Notepad
File Edit Format View Help
title = Protein-ligand complex MD simulation
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 500000 ; 2 * 500000 = 1000 ps (100 ns)
dt = 0.002 ; 2 fs
; Output control
nstenergy = 500 ; save energies every 100 ps
nstlog = 500 ; update log file every 100 ps
nstxout-compressed = 500 ; save coordinates every 100 ps
; Bond parameters
continuation = yes ; continuing from NPT
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
ns_type = grid ; search neighboring grid cells
nstlist = 20 ; largely irrelevant with Verlet
rlist = 1.2
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
rcoulomb = 1.2
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps = Protein Water_and_ions ; two coupling groups - more accurate
tau_t = 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl = Parrinello-Rahman ; pressure coupling is on for 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
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr = no
; Velocity generation
gen_vel = no ; continuing from NPT equilibration

```

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.

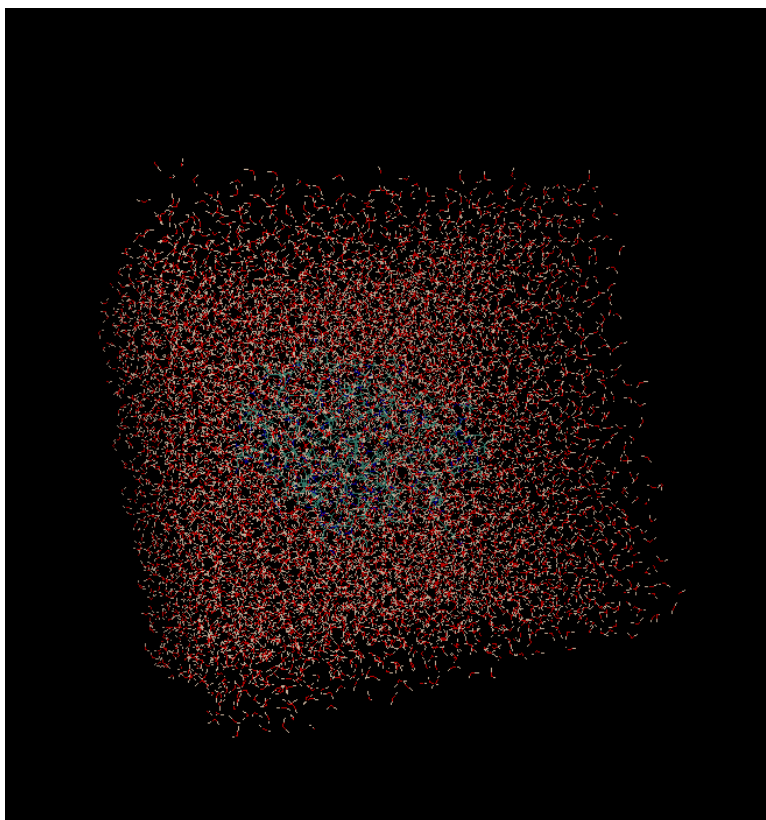


Recentering and Rewrapping Coordinates

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

Analysis will be carried out for **RMSD, RMSF, RoG, and PCA analysis** of the whole MD simulation trajectory will be carried out using MDAnalysis python library using google colab.