Hands-on (Molecular Dynamics Simulations using GROMACS)

Dr. Nidhi Katyal

Research Associate
Translational Bioinformatics Group (Dr. Dinesh Gupta)
International Centre for Genetic Engineering and Biotechnology

Operating System

1) **Windows/Mac:**
   (a) Install MobaXterm (Portable edition)
   ssh -X username@ip

   (b) Install Putty
   Host Name: IP address
   Port: 22
   Connection type: ssh
2) **Linux:**
   ssh -X username@ip

Login Details

Username: sbw_as

Password: sbw@123

S.no.	IP address
1	192.168.5.95
2	192.168.5.89
3	192.168.5.87
4	192.168.5.88
5	192.168.15.78
6	192.168.15.79
7	192.168.15.69

Command: ssh -X username@ip

Modules

Molecular Dynamics Simulations using GROMACS software:

- Simulating protein in aqueous environment (protein_water)
- Simulating membrane protein (protein_membrane)

Folders inside the workshop folder:

- Module_protein_water
- Module_protein_membrane

Each Folder contains 2 subfolders: input and output

You will execute commands in input folder

output folder contains already executed files (for reference in case you miss anything)

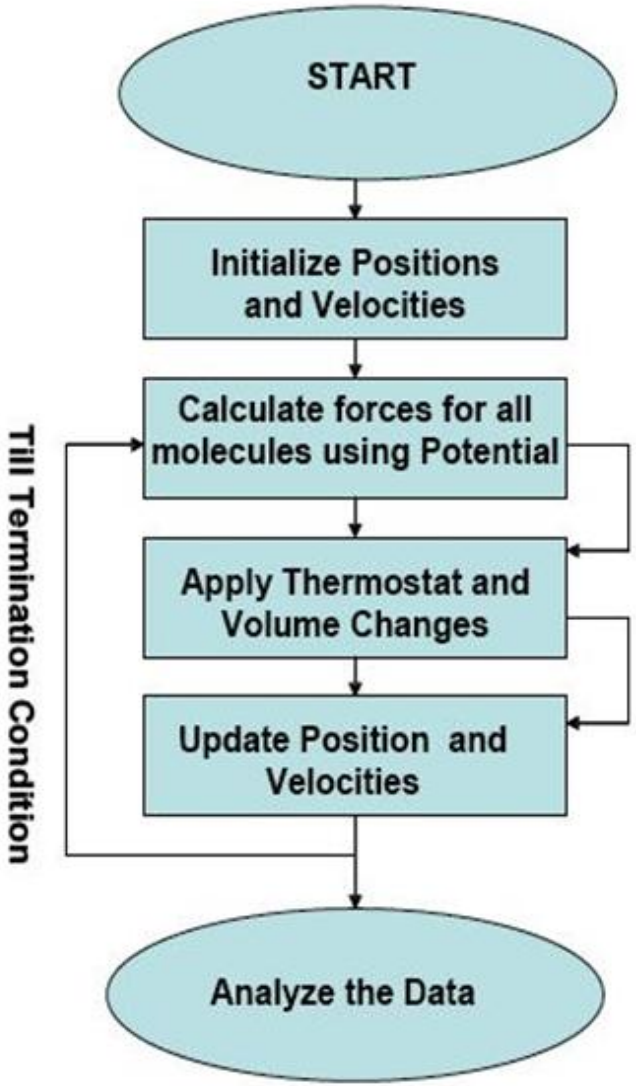
Commands

- ssh -X sbw_as@192.168.15.69
- ls
- cd workshop
- cd Module_protein_water
- cd input
- gedit commands.txt & (or vi commands.txt)
- Press control c

GROMACS files

- **Structure file** : *.gro
  (positions, velocities, & box vectors)
- **Topology file** : *.top; *.itp
  (bonded and non-bonded parameters based on atomtypes)
- **Trajectory file**: *.trr, *.xtc
  (positions, velocities, forces)
- **Parameter file**: *.mdp
  (time step, algorithms etc.)
- **Run input file**: *.tpr
  (system topology, parameters, coordinates and velocities (binary, portable))
- **Output file**: *.xvg
  (can be plotted using xmgrace, gnuplot, excel, origin etc.)

Workflow


START
Initialize Positions and Velocities
Calculate forces for all molecules using Potential
Apply Thermostat and Volume Changes
Update Position and Velocities
Till Termination Condition
Analyze the Data

Module 1

- Simulating protein in aqueous environment (protein_water)

Protein -> Lysozyme

Water model -> spce

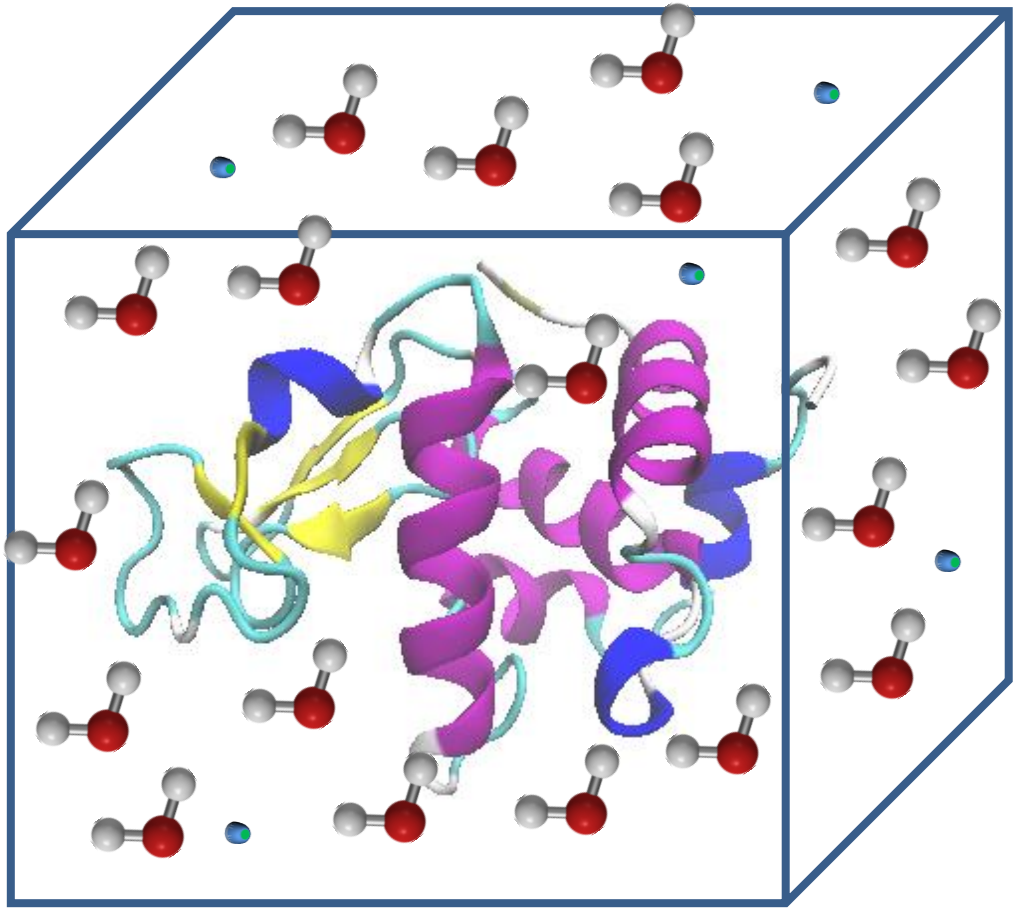
Force field -> OPLS-AA

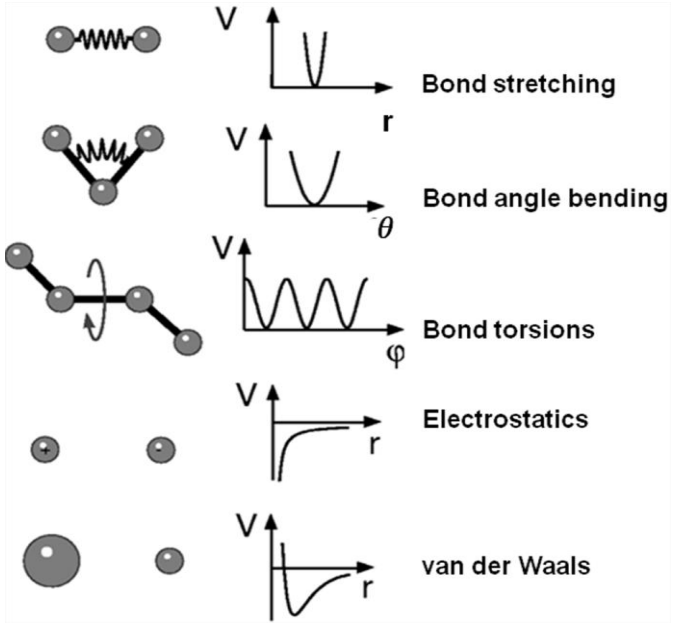
Steps

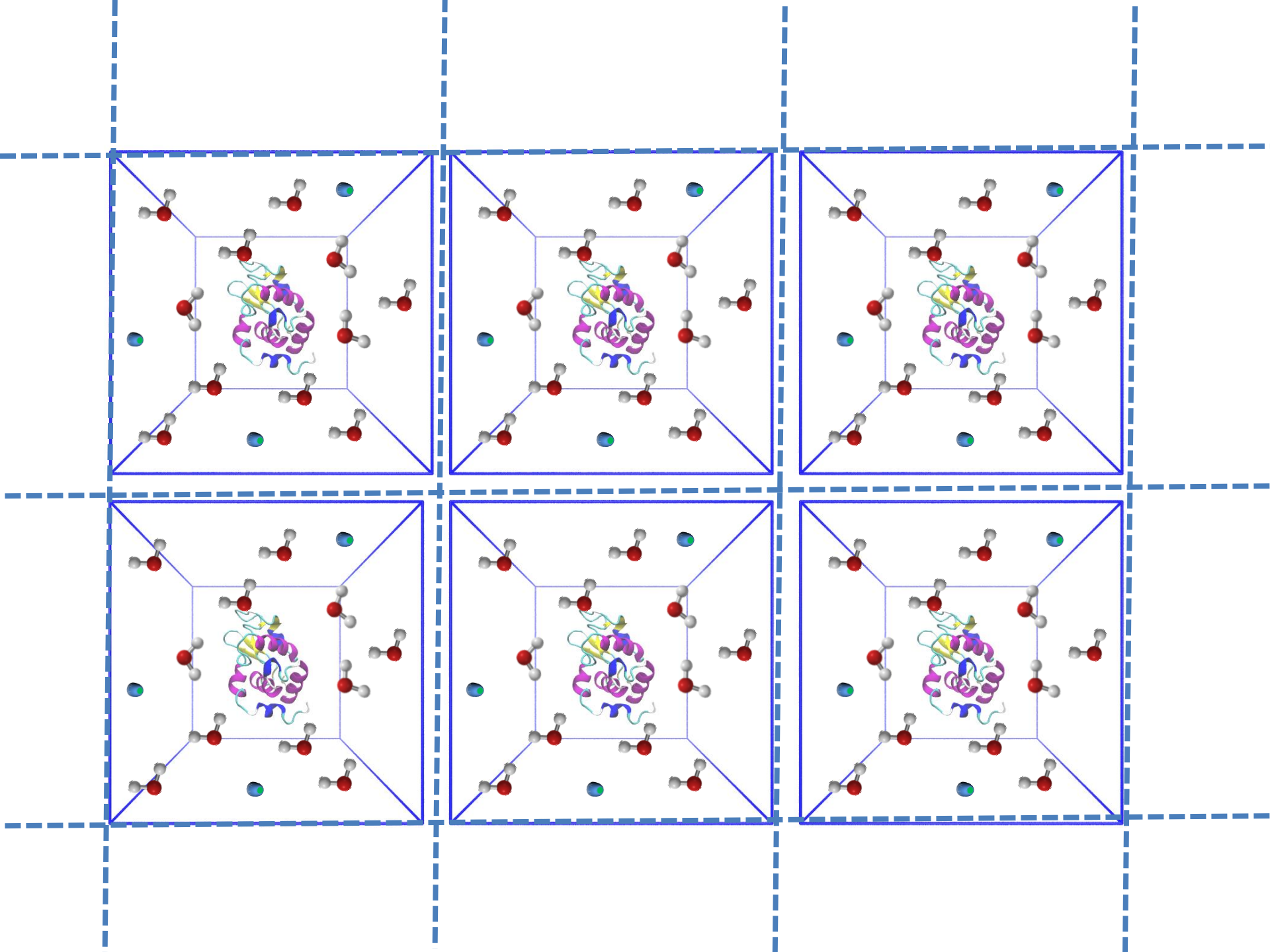
- Pre-processing
- Choosing force field and creating topology
- Defining box
- Solvation
- Adding ions to neutralize system
- Energy Minimization
- Equilibration
- Production MD
- Analysis
- Visualization

Choosing force field and creating topology
Adding ions to neutralize system
Solvation
Pre-processing
Defining box

$$E_{total} = E_{bonds} + E_{angles} + E_{dihedral} + E_{improper} + E_{vdw} + E_{elec}$$







Steps

- Pre-processing
- Choosing force field and creating topology
- Defining box
- Solvation
- Adding ions to neutralize system
- Energy Minimization
- Equilibration
- Production MD
- Analysis

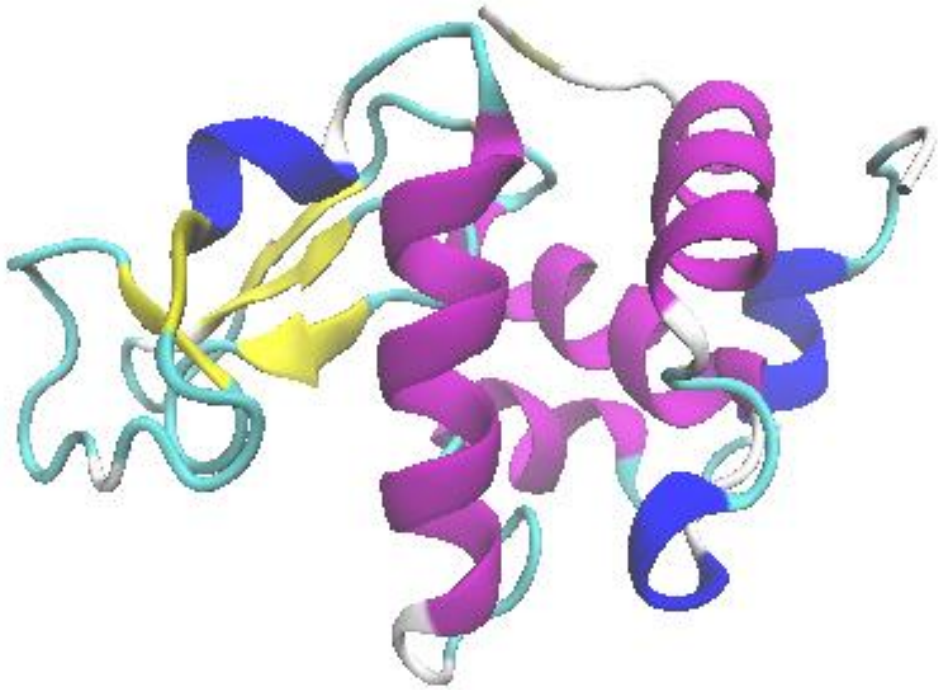
Pre-processing

- Hen Egg White Lysozyme -> RCSB -> 1AKI.pdb
- Remove crystal water, if not required

  grep -v HOH 1aki.pdb > 1AKI_clean.pdb

  (-v option inverts the sense of matching)
- Missing atoms ?
- Capping of terminals ?

Lysozyme Structure



Choosing force field & creating topology

gmx pdb2gmx -f 1AKI_clean.pdb -o 1AKI_processed.gro -water spce

Type 15 (OPLS ff) & press enter

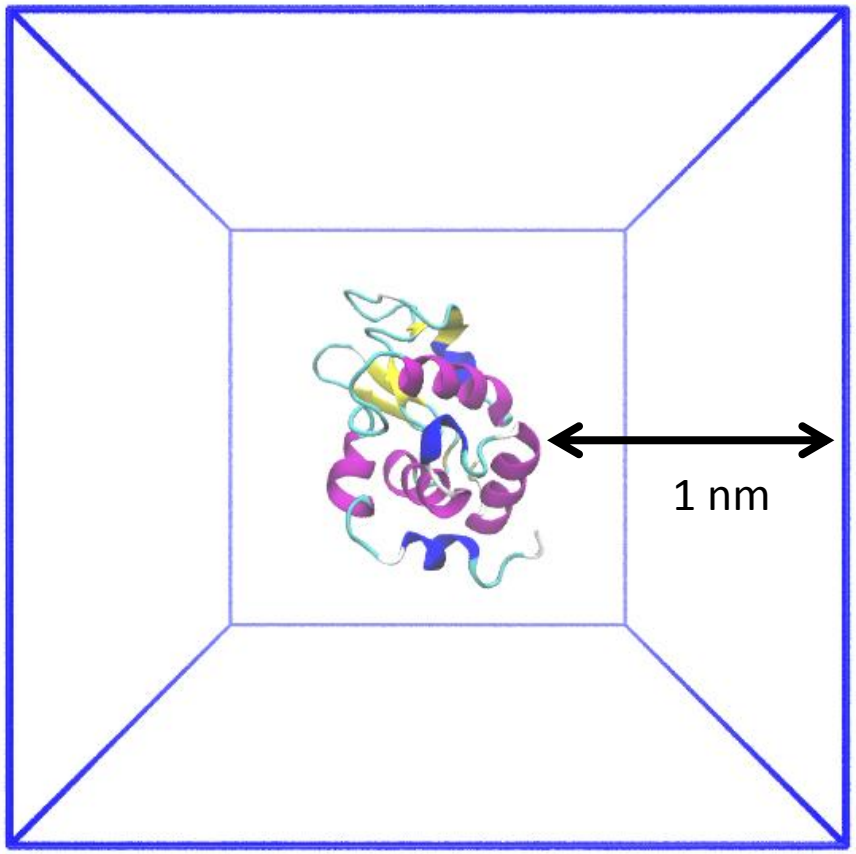
It generates 3 files:

a) Topology of molecule
b) Position restraint file
c) Post-processed structure file

Note the net charge on protein.

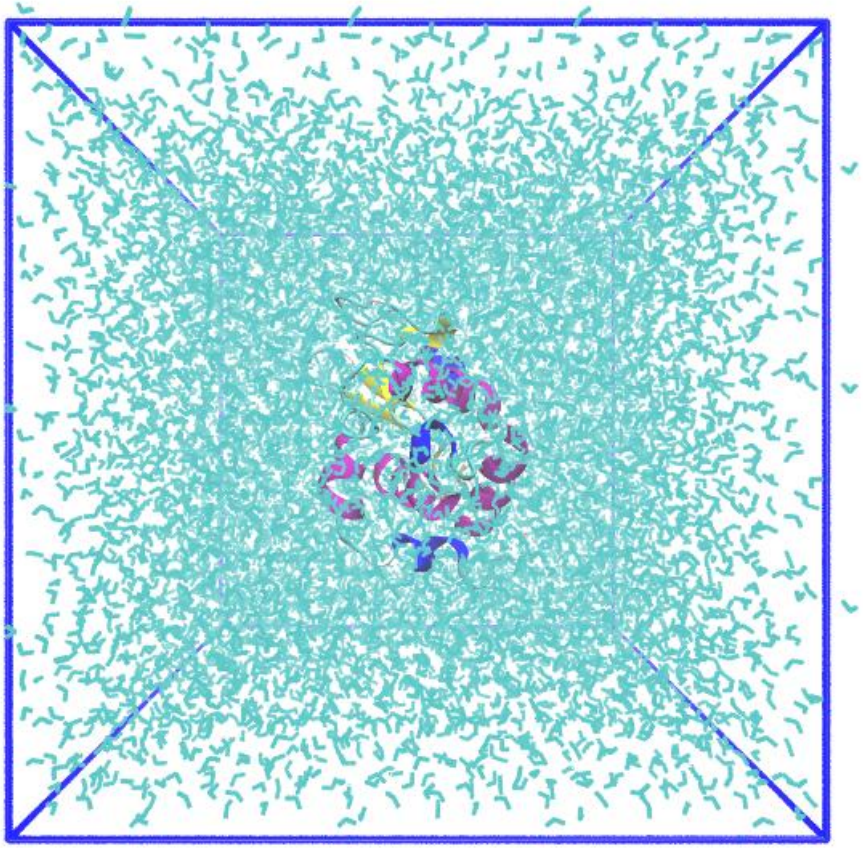
Defining box

gmx editconf -f 1AKI_processed.gro -o 1AKI_newbox.gro -c -d 1.0 -bt cubic



Solvation

gmx solvate -cp 1AKI_newbox.gro -cs spc216.gro -o 1AKI_solv.gro -p topol.top



Adding ions to neutralize system

Charge on system:

- Checked using pdb2gmx (+8e)

gmx grompp -f ions.mdp -c 1AKI_solv.gro -p topol.top -o ions.tpr

gmx genion -s ions.tpr -o 1AKI_solv_ions.gro -p topol.top -pname NA -nname CL –neutral

Type 13 (SOL)

ions.mdp

```
; ions.mdp - used as input into grompp to generate ions.tpr
; 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 < 1000.0 kJ/mol/nm
emstep      = 0.01          ; Minimization 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    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
rlist           = 1.2       ; Cut-off for making neighbor list (short range forces)
coulombtype     = cutoff    ; Treatment of long range electrostatic interactions
rcoulomb        = 1.2       ; Short-range electrostatic cut-off
rvdw            = 1.2       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

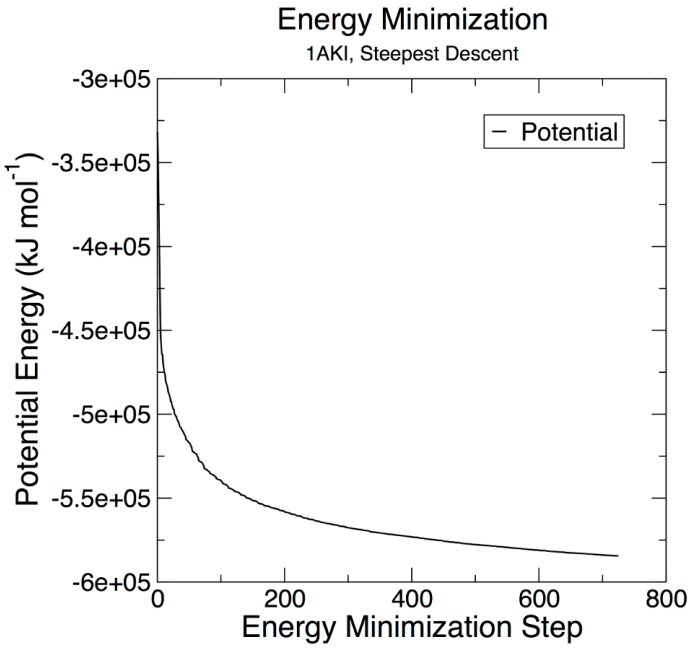
Energy Minimization

gmx grompp -f minim.mdp -c 1AKI_solv_ions.gro -p topol.top -o em.tpr

gmx mdrun -v -deffnm em

gmx energy -f em.edr -o potential.xvg

- Check that potential energy is negative
- F_{max} < emtol



Equilibration

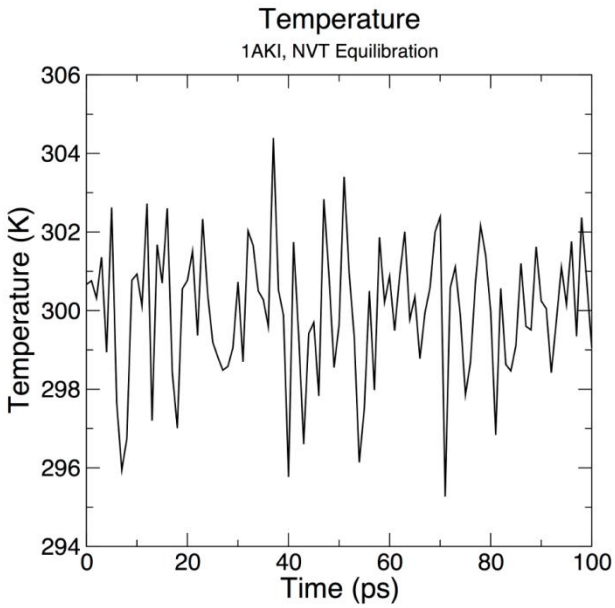
- Establish correct temperature and pressure
- Equilibrate solvent & ions around protein by applying position restraining force on the heavy atoms of the protein
- 2 phases:
  - NVT
  - NPT

NVT Equilibration

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

gmx mdrun –v -deffnm nvt

gmx energy -f nvt.edr -o temperature.xvg



nvt.mdp

```
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                    = 300     300           ; 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                 = 300       ; temperature for Maxwell distribution
```

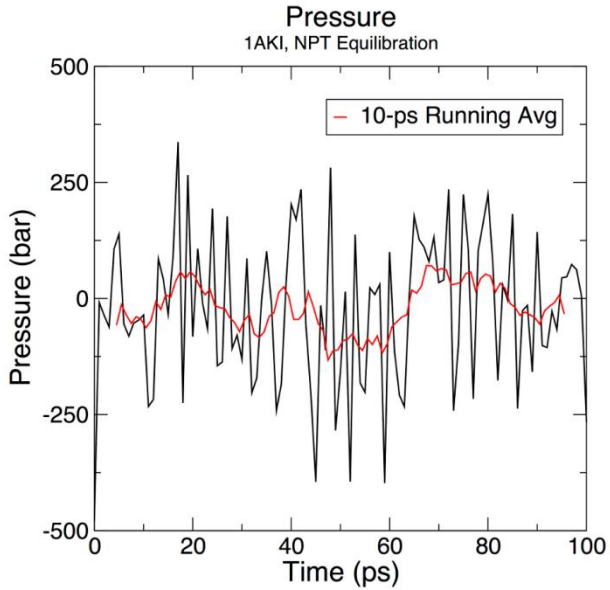
NPT Equilibration

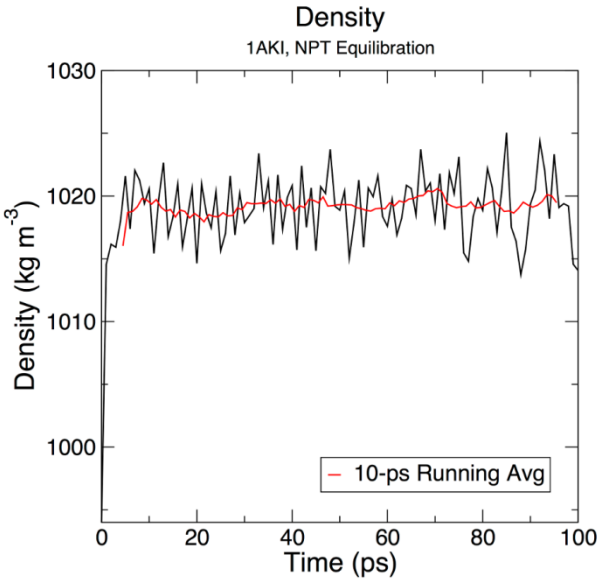
gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr

gmx mdrun –v -deffnm npt

gmx energy -f npt.edr -o pressure.xvg

gmx energy -f npt.edr -o density.xvg





npt.mdp

```
define                  = -DPOSRES  ; position restrain the protein
; Run parameters
integrator              = md        ; leap-frog integrator
nsteps                  = 25000     ; 2 * 25000 = 50 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                   = 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
```

Production MD

- Release position restraints

gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr

gmx mdrun -v -deffnm md_0_1

md.mdp

```
title                    = OPLS Lysozyme NPT equilibration
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 500000    ; 2 * 500000 = 1000 ps
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 0         ; suppress bulky .trr file by specifying
nstvout                  = 0         ; 0 for output frequency of nstxout,
nstfout                  = 0         ; nstvout, and nstfout
nstenergy                = 5000      ; save energies every 10.0 ps
nstlog                   = 5000      ; update log file every 10.0 ps
nstxout-compressed       = 5000      ; save compressed coordinates every 10.0 ps
compressed-x-grps        = System    ; save the whole system
; Bond parameters
continuation             = yes       ; Restarting after NPT
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
; Neighborsearching
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)
; 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
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
; Dispersion correction
DispCorr                 = EnerPres  ; account for cut-off vdW scheme
; Velocity generation
gen_vel                  = no        ; Velocity generation is off
```

Analysis

gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -pbc mol –center

gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns

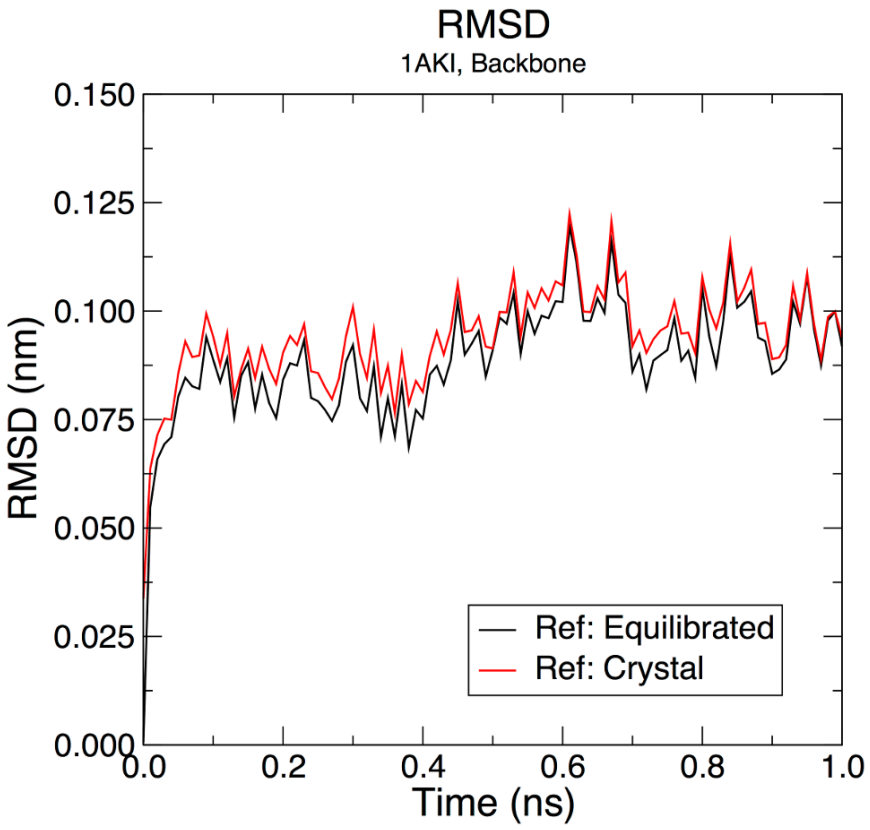
gmx rms -s em.tpr -f md_0_1_noPBC.xtc -o rmsd_xtal.xvg -tu ns

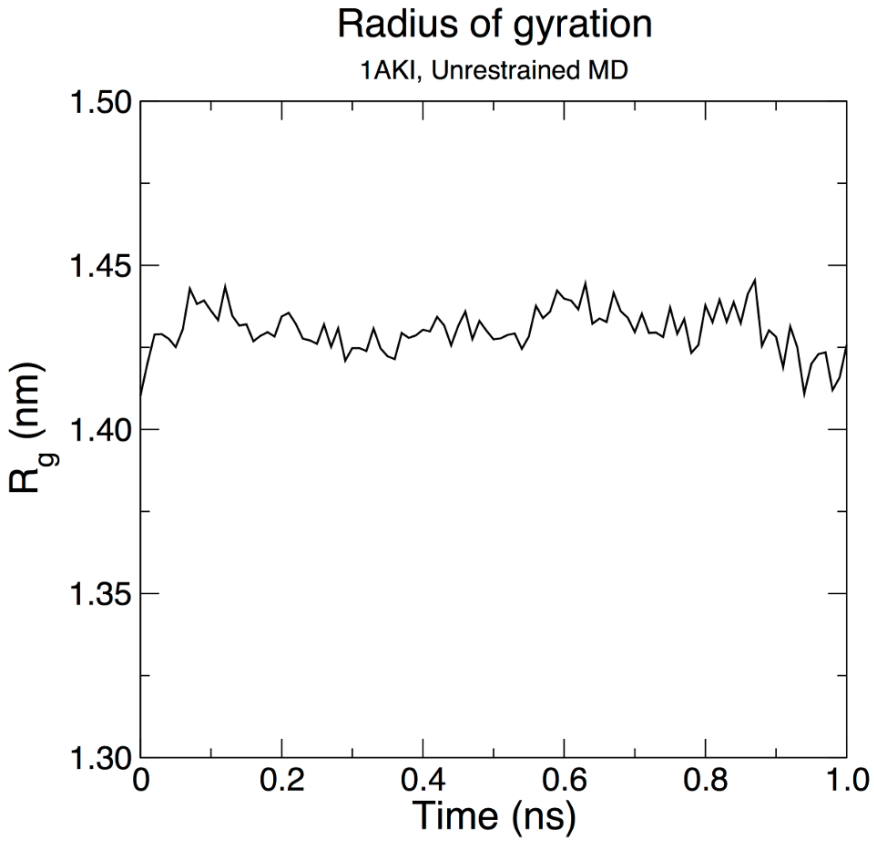
gmx gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -o gyrate.xvg

gmx sasa -s md_0_1.tpr -f md_0_1_noPBC.xtc -o sasa.xvg

Gnuplot for plotting

gnuplot
> set datafile commentschars "#@&"
> plot "rmsd.xvg" using 1:2 with lines



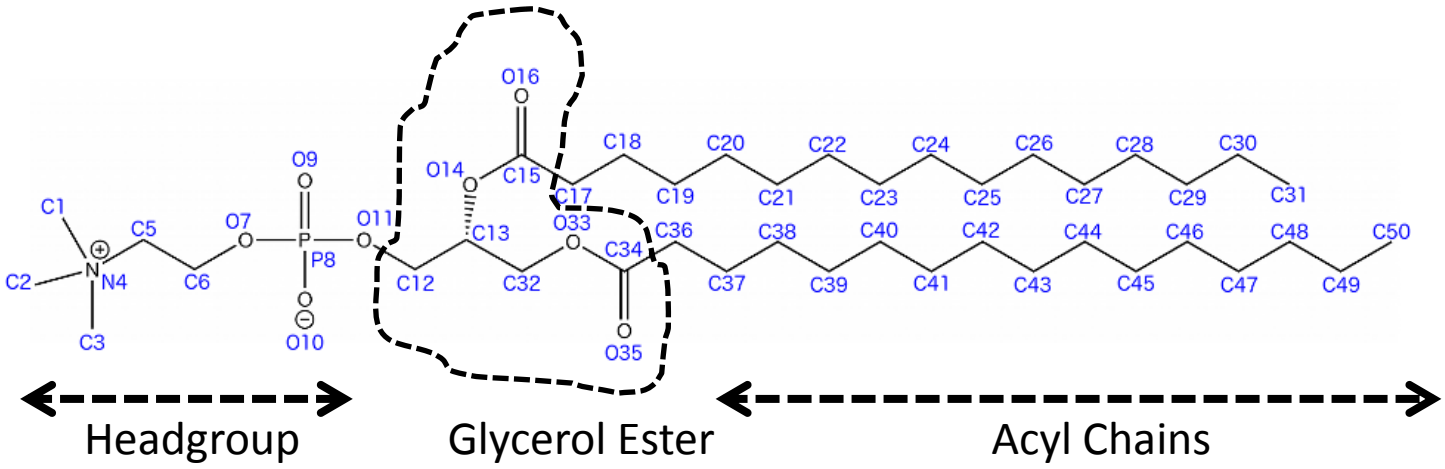


Module 2

- **Membrane-protein simulations**

Protein -> Lysozyme

Membrane -> DPPC
(dipalmitoylphosphatidylcholine)



Steps

- Pre-processing
- Choosing force field and creating topology
- Membrane packing
- Solvation
- Adding ions to neutralize system
- Energy Minimization
- Equilibration
- Production MD (slight variation in parameters)
- Analysis

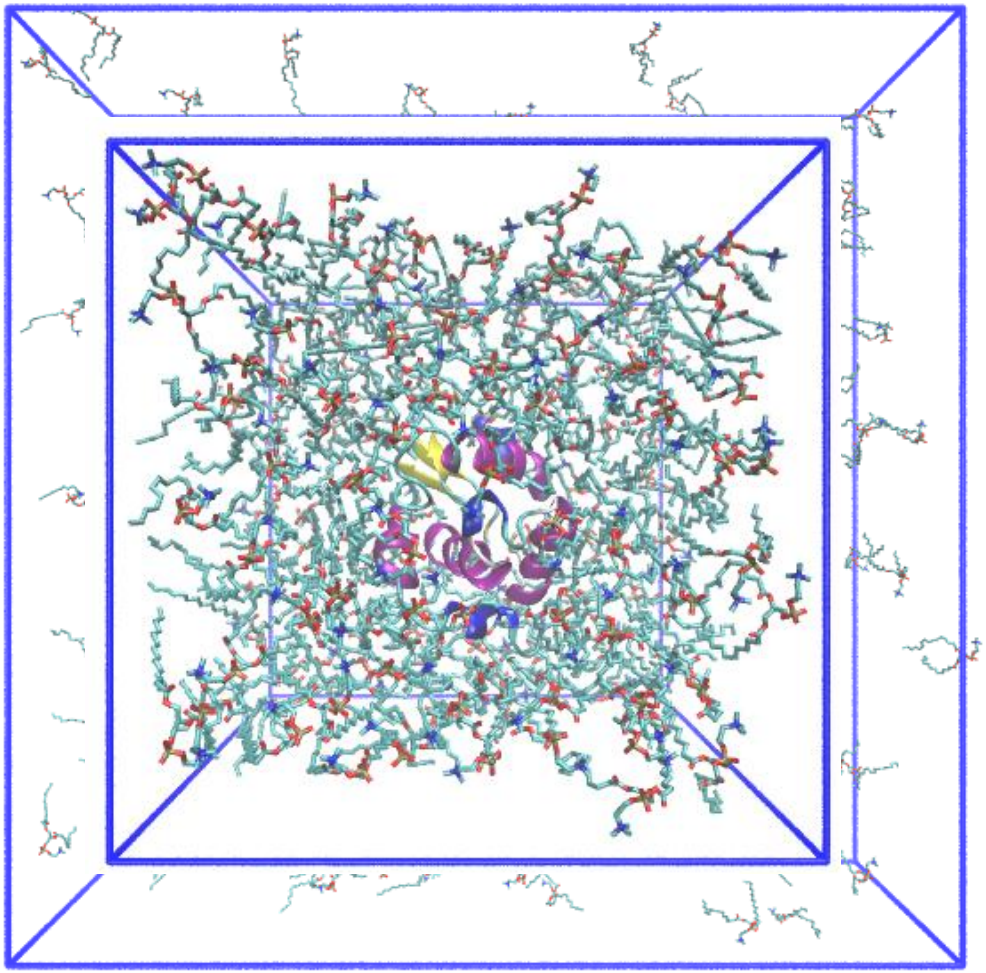
Steps

Creating unit cell of membrane followed by energy minimization

Inflation of the lipid bilayer followed by energy minimization

Deflation of membrane followed by energy minimization

Alignment of protein on the membrane followed by energy minimization



Pre-processing

grep -v HOH 1aki.pdb > 1AKI_clean.pdb

Choosing force field & creating topology

gmx pdb2gmx -f 1AKI_clean.pdb -o 1AKI_processed.gro -ignh -water spc

Select GROMOS96 53a6 force field

Modify topology to incorporate membrane parameters

- Add non bonded parameters of lipid (gromos53a6_lipid.ff; For steps look at commands_creating_gromos53a6_lipid_folder.txt)

  cd workshop/Module_protein_membrane/input_2

  ls

  ... gromos53a6_lipid.ff...

- Include topology parameters of lipid

Include topology parameters of lipid

topol.top file:

```
;         gmx pdb2gmx -f 1AKI_clean.pdb -o 1AKI_processed.gro -ignh -water spc
;       Force field was read from the standard GROMACS share directory.
;

; Include forcefield parameters
#include "gromos53a6_lipid.ff/forcefield.itp"

[ moleculetype ]
; Name            nrexcl
Protein_chain_A     3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass  typeB    chargeB      mass
; residue   1 LYS rtp LYSH q +2.0
```

```
; Include Position restraint file
#ifdef POSRES
#include "posre.itp"
#endif

; Include DPPC chain topology
#include "dppc.itp"

; Include water topology
#include "gromos53a6.ff/spc.itp"
```

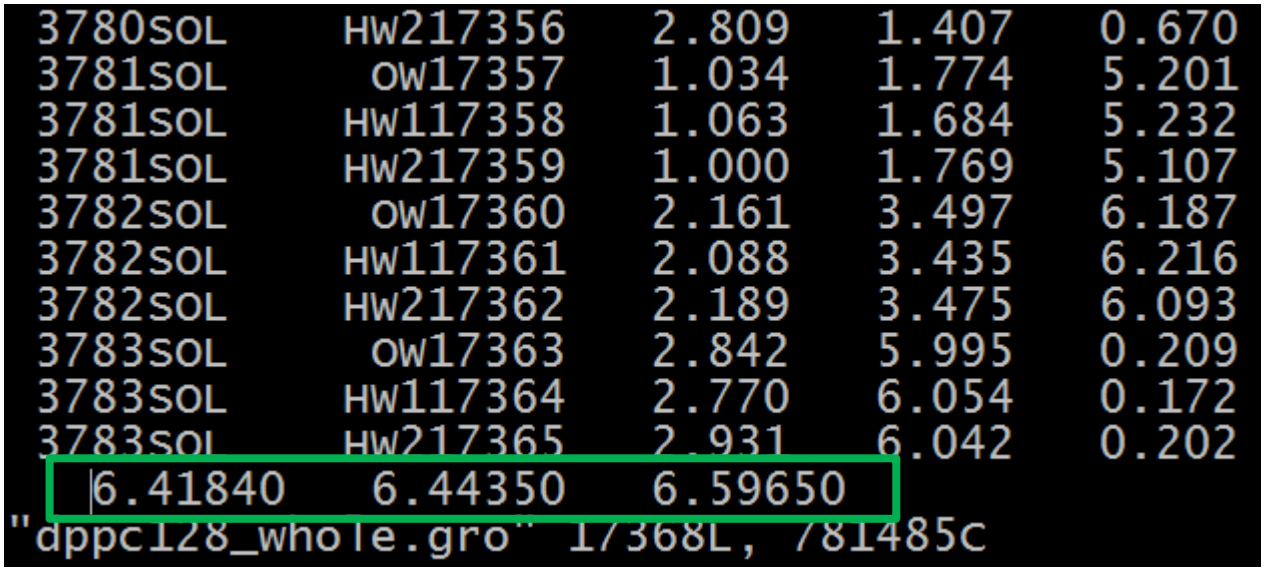
Building Unit Cell

gmx grompp -f minim.mdp -c dppc128.pdb -p topol_dppc.top -o dppc.tpr

gmx trjconv -s dppc.tpr -f dppc128.pdb -o dppc128_whole.gro -pbc mol -ur compact

select "0" system

dppc128_whole.gro file:



Defining box & orienting protein and membrane in same coordinate frame

gmx editconf -f 1AKI_processed.gro -o 1AKI_newbox.gro -c -box 6.41840 6.44350 6.59650

cat 1AKI_newbox.gro dppc128_whole.gro > system.gro

system.gro

```
  129LEU      C 1321   4.671   1.963   4.015
  129LEU     O1 1322   4.632   1.953   4.134
  129LEU     O2 1323   4.782   1.993   3.969
   6.41840   6.44350   6.59650
128-Lipid DPPC Bilayer
17365
    1DPPC     C1    1   1.577   5.265   0.920
    1DPPC     C2    2   1.675   5.295   1.135
    1DPPC     C3    3   1.648   5.482   0.985
```

Delete these lines

18688
(17365+1323)

```
LYSOZYME
1323
    1LYS       N    1   3.995   2.959   2.080
    1LYS      H1    2   4.071   3.013   2.042
    1LYS      H2    3   3.929   2.939   2.008
    1LYS      H3    4   3.951   3.011   2.153
    1LYS      CA    5   4.048   2.832   2.135
```

Pack lipids around protein & minimize energy of system

topol.top

```
; Include Position restraint file
#ifdef POSRES
#include "posre.itp"
#endif

; Strong position restraints for InflateGRO
#ifdef STRONG_POSRES
#include "strong_posre.itp"
#endif

; Include DPPC chain topology
#include "dppc.itp"
```

gmx genrestr -f 1AKI_newbox.gro -o strong_posre.itp -fc 100000 100000 100000

Continuation...

perl inflategro.pl system.gro 4 DPPC 14 system_inflated.gro 5 area.dat

(syntax: inflategro.pl bilayer.gro scaling_factor lipid_residue_name cutoff inflated_bilayer.gro gridsize areaperlipid.dat (protein))

There are 4 lipids within cut-off range...

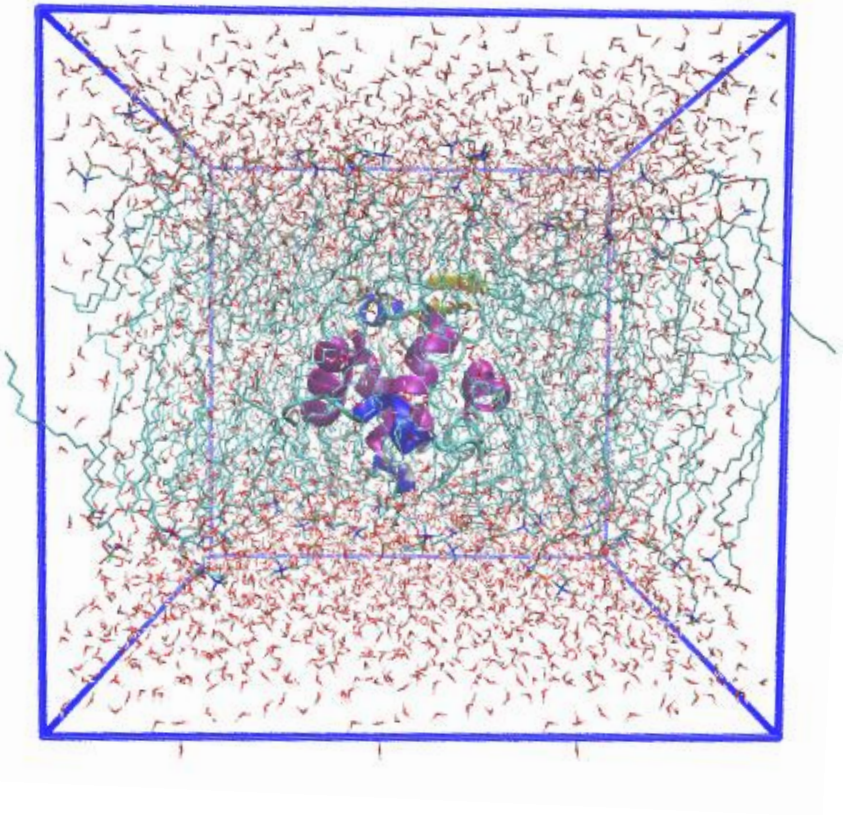
128-4=124

```
; Include topology for ions
#include "gromos53a6.ff/ions.itp"

[ system ]
; Name
LYSOZYME in water

[ molecules ]
; Compound        #mols
Protein_chain_A     1
DPPC                124
```

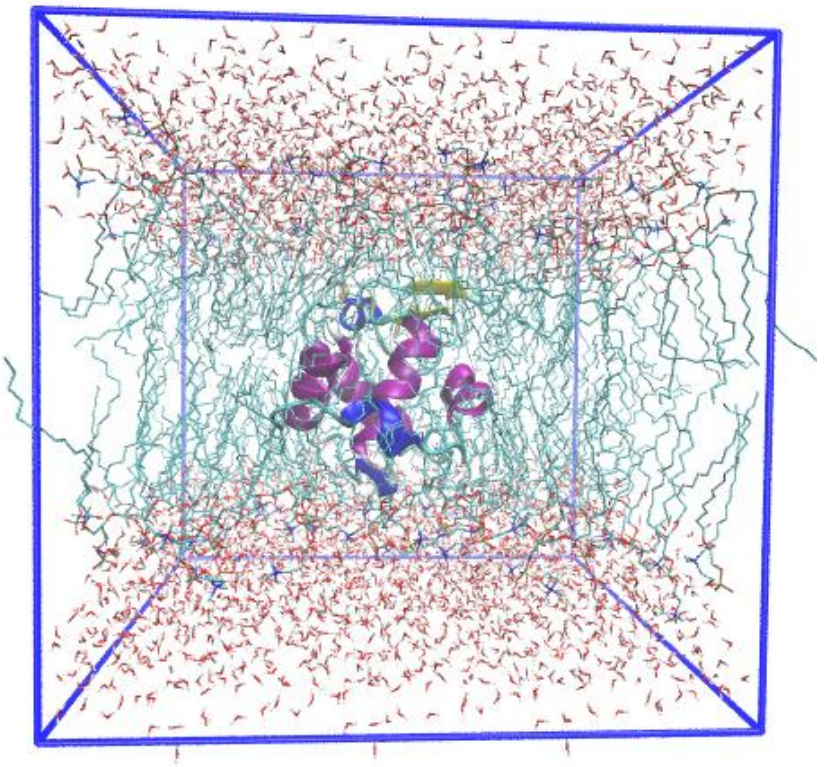
```
gmx grompp -f minim_inflategro.mdp -c system_inflated.gro -p
    topol.top -r system_inflated.gro -o system_inflated_em.tpr

gmx mdrun -deffnm system_inflated_em

gmx trjconv -s system_inflated_em.tpr -f
    system_inflated_em.gro -o tmp.gro -pbc mol
select 0 "system"

mv tmp.gro system_inflated_em.gro

perl inflategro.pl system_inflated_em.gro 0.95 DPPC 0
    system_shrink1.gro 5 area_shrink1.dat

bash run_inflategro.sh
```

Solvation

(a) Solvate with water



(b) Delete water inside lipid acyl chain



Solvation

- Solvate with water as usual (using gmx solvate)

gmx solvate -cp system_shrink26_em.gro -cs spc216.gro -o system_solv.gro -p topol.top

- Gaps in the lipid acyl chains would also be filled by water molecules. Delete them (water_deletor.pl)

perl water_deletor.pl -in system_solv.gro -out system_solv_fix.gro -ref O33 -middle C50 -nwater 3

Look for this line!!

1409 water molecules have been deleted.

4182 water molecules remain. Update your topology!

topol.top file:

```
[ system ]
; Name
LYSOZYME in water

[ molecules ]
; Compound        #mols
Protein_chain_A     1
DPPC                124
SOL              4182
CL                  8
"topol.top" 8421L, 308370C
```

Further Steps

- Similar as in Module 1
- bash further_steps.sh
- cp ../output/md_0_1.tpr .
- cp ../output/md_0_1.xtc .
- bash analysis.sh

Comparison with standard MD

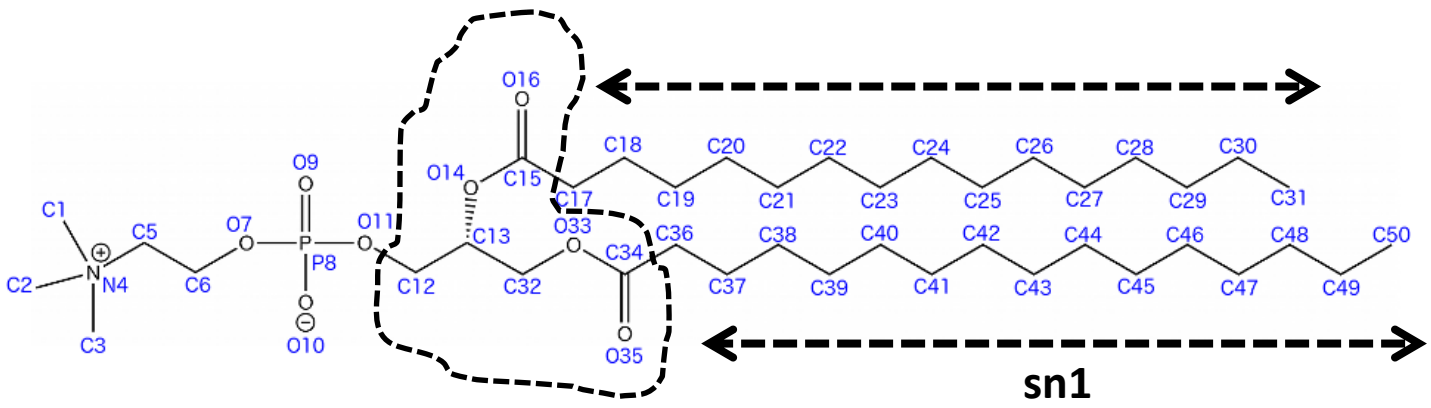
Property	Non-membrane Simulations	Membrane Simulations
Equilibration time	~100 ps	~1 ns
Temperature	Any	Must be above phase transition of lipid
Pressure coupling	Isotropic	Semi-isotropic
Center of mass motion removal groups	Protein, Non-protein	Protein_DPPC, Water_Ions

Analysis

- **Deuterium order parameter**

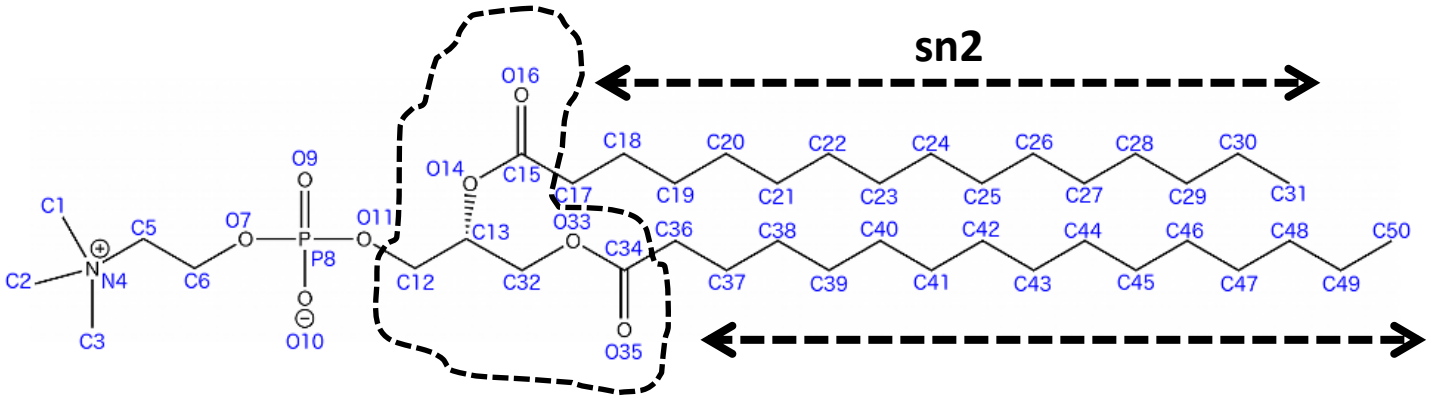
gmx make_ndx -f md_0_1.tpr -o sn1.ndx

```
> a C34
> a C36
> a C37
> a C38
> a C39
> a C40
> a C41
> a C42
> a C43
> a C44
> a C45
> a C46
> a C47
> a C48
> a C49
> a C50
> q
```

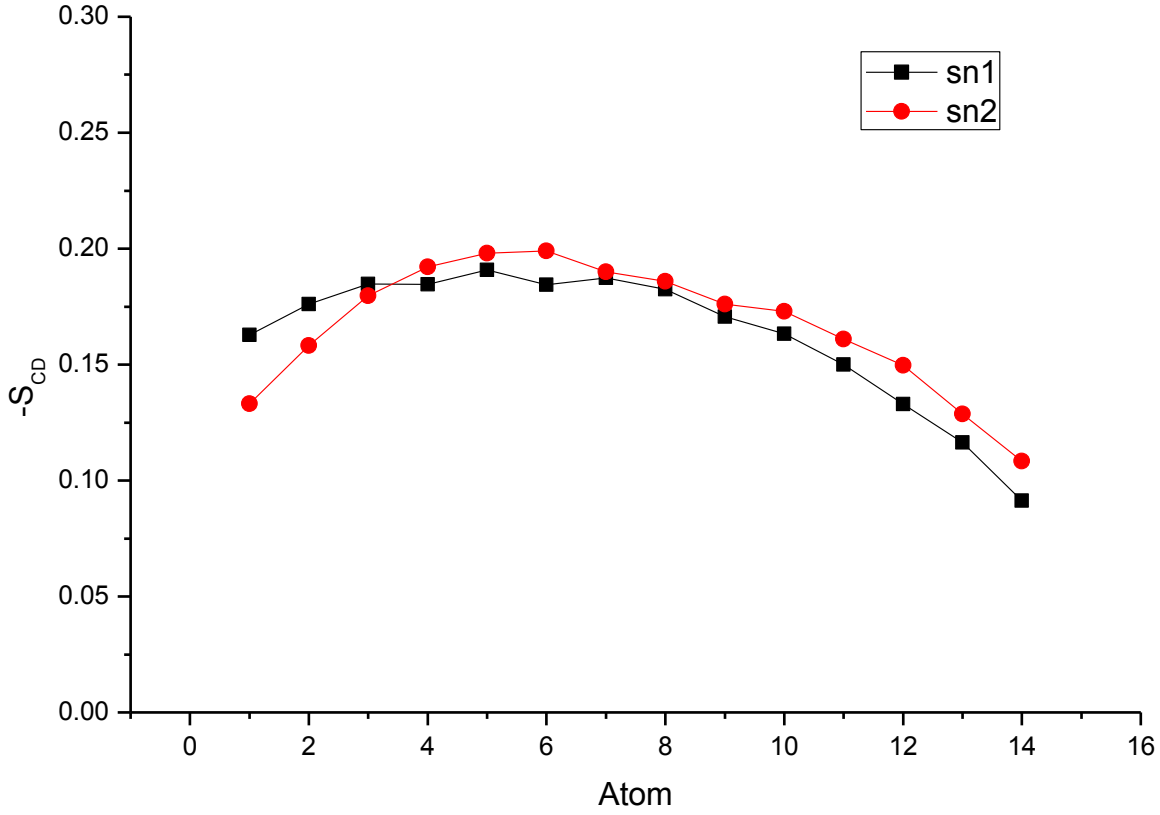


- **gmx make_ndx -f md_0_1.tpr -o sn2.ndx**

- > a C15
- > a C17
- > a C18
- > a C19
- > a C20
- > a C21
- > a C22
- > a C23
- > a C24
- > a C25
- > a C26
- > a C27
- > a C28
- > a C29
- > a C30
- > a C31
- > q



- gmx order -s md_0_1.tpr -f md_0_1.xtc -n sn1.ndx -d z -od deuter_sn1.xvg
- gmx order -s md_0_1.tpr -f md_0_1.xtc -n sn2.ndx -d z -od deuter_sn2.xvg

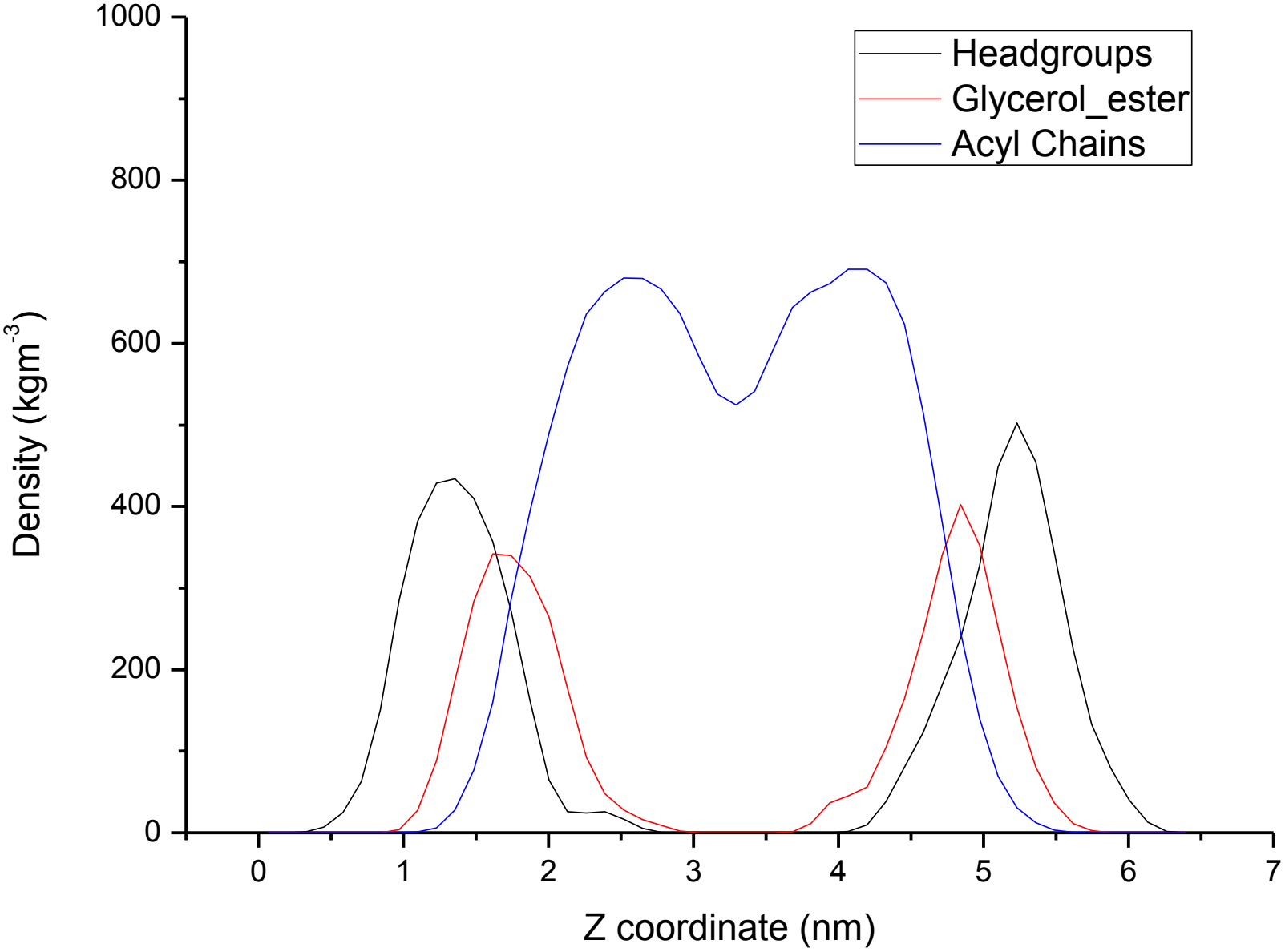


Density of membrane

- **gmx make_ndx -f md_0_1.tpr -o density_groups.ndx**

- 13 & a C1 | a C2 | a C3 | a N4 | a C5 | a C6 | a O7 | a P8 | a O9 | a O10 | a O11
- name 22 headgroups
- 13 & a C12 | a C13 | a O14 | a C15 | a O16 | a C32 | a O33 | a C34 | a O35
- name 23 Glycerol_Ester
- 13 & ! 22 & ! 23
- name 24 Acyl_Chains

- **echo 22 | gmx density -s md_0_1.tpr -f md_0_1.xtc -n density_groups.ndx -o dens_headgroups.xvg -d Z**
- **echo 23 | gmx density -s md_0_1.tpr -f md_0_1.xtc -n density_groups.ndx -o dens_glycerol_ester.xvg -d Z**
- **echo 24 | gmx density -s md_0_1.tpr -f md_0_1.xtc -n density_groups.ndx -o dens_acyl_chains.xvg -d Z**


Headgroups
Glycerol_ester
Acyl Chains
Density (kgm-3)
Z coordinate (nm)
0
200
400
600
800
1000
0
1
2
3
4
5
6
7

Lateral Diffusion of lipids

- **gmx make_ndx -f md_0_1.tpr -o p8.ndx**
- ...
- > a P8
- > q

- **gmx msd -s md_0_1.tpr -f md_0_1.xtc -n p8.ndx -lateral z**

- Fitting from 100 to 900 ps

- D[P8] 0.0301 (+/- 0.0250) 1e-5 cm^2/s


Thank you!