A SIMPLIFIED METHODOLOGY FOR CREATION OF COMPLETE ENVIRONMENT FOR THE MOLECULAR DYNAMICS SIMULATION OF PROTEINS

Manish Kumar Suthar
Scientist
ICAR-Directorate of Medicinal and Aromatic Plants Research,
Anand, Gujarat 387310, India

ABSTRACT

Molecular dynamics simulations (MDS) have emerged as an effective tool to understand the structural-functional relationships of various biomolecules like proteins, RNA, DNA, membrane lipids, etc. The MDS can show the positions, velocities, and changes in orientations of biomolecules in model liquids. Further, MDS allows to analyse the interactions between proteins and their respective ligands. The behavior of biomolecules like drug targets or other molecules within lipid bilayer can also be simulated. These studies are important for the analysis of interactions between drug candidates and various cell receptors or drug targets. In the present methodology, a simplified step by step protocol starting from downloading the GROMACS and its accessories to installing on our system is presented which is otherwise a very tedious and a technical issue. Finally, the simplified methodology for the MDS of a protein in water is presented. Various commands to run MDS for a protein structure are presented in a simplified manner. This methodology is very useful for the researchers and students engaged in the field of bioinformatics.

Requirements:

Operating system: Linux (Ubuntu)

Tools: GROMACS, g++, CMake, fftw, Xmgrace, pymol, and VMD.

Preparatory steps:

Download the g++, fftw and CMake on Ubuntu and install them on system using following commands: (In terminal)

sudo apt-get install g++

Go to fftw website and download fftw tar.gz file

Go to cmake website and download cmake tar.gz file

Cd download to go to download folder

ls for listing the downloaded tools

Three components should be there: cmake, fftw and GROMACS

Unlocking the tar file: tar xzvf cmake + tab button for version

ls followed by cd to in cmake and ls to see files within the cmake folder

./configure enter

make -jX (X is number of threads in system)

sudo make install -jx

sudo nautilus

Go to computer

Search for cmake

Copy cmake binary file/logo

Go to computer

Go to usr

Go to bin

Paste the binary file here

To go back type cd .. and go to download folder

Tar xvzf fftw

ls followed by cd fftw tab and ls

./configure

make -jX

make install –jx

cd .. to go back go to download folder

tar xzvf GROMACS tab

Go to gromacs folder and ls

Make an empty directory: mkdir build

cd build and ls (nothing should be there)

type command "cmake" and drag the entire folder of GROMACS enter

-DGMX_BUILD_OWN_FFTW=ON

make -jx

sudo make install -jx

source /usr/local/gromacs/bin/GMXRC

Now our GROMACS has installed and to check it type

pdb2gmx

Upon enter all details of gromacs should be there like force fields etc.

After successfully installing GROMACS on Ubuntu now we can run MD simulation of proteins.

For simulations some molecular dynamics parameter files are required which are comprises the main parameters like temperature, time duration of simulation etc. These files as ".mdp" format should be present along with pdb file. To start the simulation make a folder and name it. Keep the protein structure file in .pdb format along with .mdp files. The .mdp files can be modified according to need of experiment.

Commands for the MD simulation (protein in water model):

Protein in Water:

Place protein.pdb file (XXX.pdb), md.mdp, ions.mdp, npt.mdp and nvt.mdp in a folder. Protein .pdb file should not contains any water molecule. Go within the folder having above file and open terminal of Ubuntu. Now, follow the commands shown below:

1. **Activating GROMACS within folder**

source /usr/local/gromacs/bin/GMXRC

2. **Topology generation**

gmx pdb2gmx -f 1XXX.pdb -o 1XXX_processed.gro -water spce

type 9 (this is the type of force field: if GROMOS96 43a1 force field is to be selected type 9, similarly 1 to 15 different force fields can be used)

3. **Define box and solvation**

gmx editconf -f 1XXX_processed.gro -o 1XXX_newbox.gro -c -d 1.0 -bt cubic

gmx solvate -cp 1XXX_newbox.gro -cs spc216.gro -o 1XXX_solv.gro -p topol.top

4. **Adding ions to neutralise the system**

Go to topol.top file and find the overall charges and finally neutralise with Na+ or Cl- accordingly.

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

gmx genion -s ions.tpr -o 1XXX_solv_ions.gro -p topol.top -pname NA -nname CL -nn 3
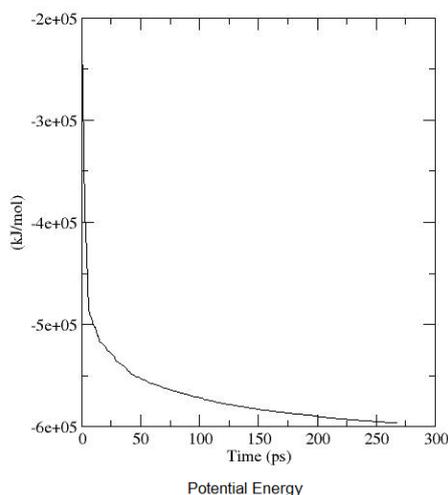
13 "SOL"

5. **Energy minimization**

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

gmx mdrun -v -deffnm em

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

At the prompt, type "10 0" to select Potential (10); zero (0) terminates input.

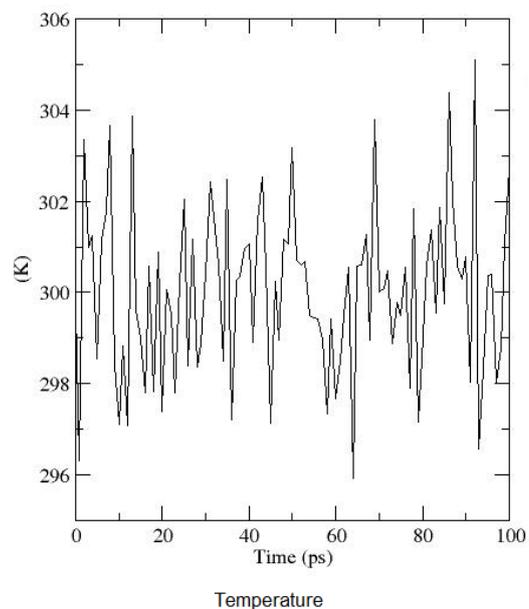At this point the potential.xvg file can be open and analysed using Xmgrace.

Potential Energy

6. Equilibration

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

gmx mdrun -v -deffnm nvt -nt 4
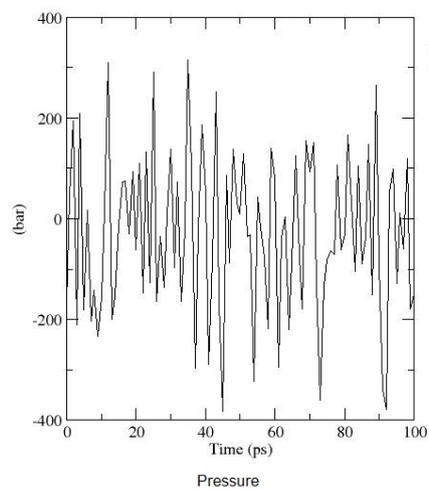
gmx energy -f nvt.edr -o temperature.xvg

Temperature

Type "15 0" at the prompt to select the temperature of the system and exit.

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

gmx mdrun -v -deffnm npt -nt 4

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

Pressure

4

Type "16 0" at the prompt to select the pressure of the system and exit.

Let's take a look at density as well, this time using energy and entering "22 0" at the prompt.

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

7. Production MD

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 -nt 4
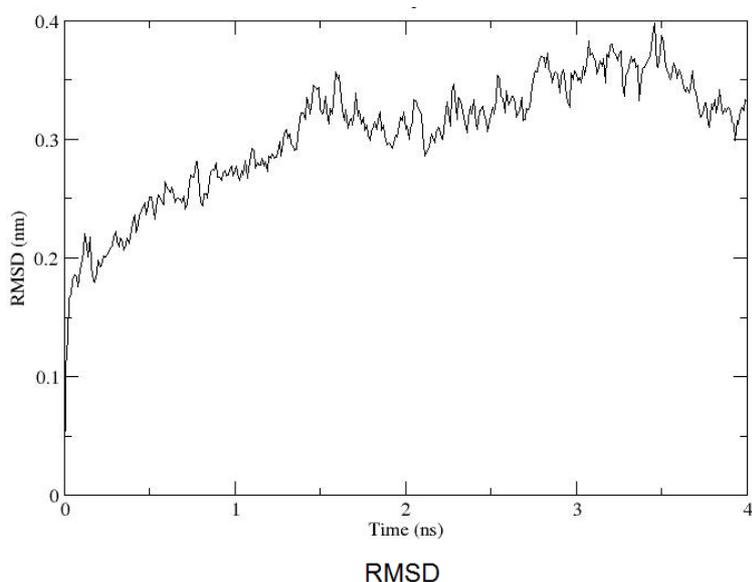
It will take days to week time to complete production MD.

8. Analysis

gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -pbc mol -ur compact

Select 0 ("System") for output.

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

Choose 4 ("Backbone") for both the least squares fit and the group for RMSD calculation.

RMSD

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

g_rmsf -f md_0_1.xtc -s md_0_1.tpr -o rmsf.xvg -res

group 1

The RMSD, RMSF and gyration can be analysed by using Xmgrace and respective .xvg file as follow:

xmgrace rmsf.xvg

Secondary structure analysis: For secondary analysis the program DSSP should be downloaded and installed. For analysis of secondary structure during simulation course DSSP's path must be given as follow:

export DSSP=/usr/bin/dssp

do_dssp -f md_0_1.xtc -s md_0_1.tpr -sc scount.xvg -o ss.xpm

group 1

xmgrace -nxy scount.xvg

xpm2ps -f ss.xpm -di ps.m2p -o ubq_ss.eps

gv ubq_ss.eps

gmx trjconv -s topol.tpr -f md_0_1.xtc -dt 100 -o trj.pdb

Now all simulations can be analysed using VMD tool.

…………………….

Examples of .mdp files:

ion.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 calcu-
late 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)
coulombtype     = cutoff    ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0       ; Short-range electrostatic cut-off
rvdw            = 1.0       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

minim.mdp

```
; minim.mdp - used as input into grompp to generate em.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 calcu-
late 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)
coulombtype     = PME        ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0        ; Short-range electrostatic cut-off
rvdw            = 1.0        ; Short-range Van der Waals cut-off
pbc             = xyz        ; Periodic Boundary Conditions in all 3 dimensions
```

nvt.mdp

```
title                   = 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 elec-
trostatics
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 ac-
curate
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 distribu-
tion
gen_temp                = 300       ; temperature for Maxwell distribution
gen_seed                = -1        ; generate a random seed
```

npt.mdp

```
title                   = 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 elec-
trostatics
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 ac-
curate
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
```

md.mdp

```
title           = MD simulation
; Run parameters
integrator      = md             ; leap-frog integrator
nsteps          = 25000000       ; 2 * 1000000 = 2000 ps (50 ns)
dt              = 0.002          ; 2 fs
; Output control
nstxout         = 5000           ; save coordinates every 10.0 ps
nstvout         = 5000           ; save velocities every 10.0 ps
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
```

```
                             ; nstxout-compressed replaces nstxtcout
compressed-x-grps  = System      ; replaces xtc-grps
; Bond parameters
continuation       = yes          ; Restarting after NPT
constraint_algorithm  = lincs    ; holonomic constraints
constraints        = all-bonds  ; all bonds (even heavy atom-H bonds) con-
strained
lincs_iter         = 1     ; accuracy of LINCS
lincs_order        = 4     ; also related to accuracy
; Neighborsearching
cutoff-scheme       = Verlet
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 electrostat-
ics
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
```

References
1. http://www.gromacs.org/
2. http://www.fftw.org/
3. https://cmake.org/
4. https://www.ks.uiuc.edu/Research/vmd/
5. https://pymol.org/2/