Molecular Dynamics in practice with GROMACS

GROMACS is one of the wold's fastest software package for molecular dynamics simulations. One can find many helpful materials, manual as well as to download different version of GROMACS though the web. The official site of the package is www.gromacs.org.

The main steps to run an all atom explicit solvent molecular dynamics simulation are:

1. Getting the 3D structure of the molecule under investigation.
2. Assignment of the force field parameters to atoms and define molecular topology.
3. Definition of the simulation box - form and size.
4. Adding solvent molecules in the box.
5. Adding ions in the box if needed
6. Relaxation of the solvent around the molecule
7. Production molecular dynamics run.

The 3D structure of the molecules (mainly proteins, DNA, RNA and their complexes) can be found on RCSB Protein Data Bank (http://www.pdb.org). After obtaining the desired molecular structure one has to check its consistency and to add missing atoms if their are needed. There are many software packages suitable for molecular modeling and visualization of big molecules available for downloading. Some of them are free for academical use such as Coot (http://lmb.bioch.ox.ac.uk/coot/), PyMol (http://www.pymol.org/) and VMD (http://www.ks.uiuc.edu/Research/vmd/). These packages allow user to download molecular structures presented in PDB by means of specially developed modules which take only structure PDB identifier as input. Coot and PyMol are suitable for visualization and structure manipulation, while VMD is capable to load and visualize trajectories of millions of atoms and thousands of frames.

The protein under PDB ID 1L2Y is used in the following tutorial to demonstrate the MD protocol mentioned above. One can obtain the protein initial structure for MD simulation by starting VMD, loading 1L2Y and saving one of the PDB entry frames in file with name *1l2y_init.pdb* for example (tutorials on this subject are presented on http://www.ks.uiuc.edu/Research/vmd/).

Forces acting on each atom in MD simulations are calculated from predefined force constants, equilibrium parameters, partial atomic charges and van der Waals parameters usualy named force field. pdb2gmx is a tool from GROMACS software package which creates coordinate file and topology of the molecule according to the force field chosen by the user. The last could be done by running the command:

pdb2gmx -f 1l2y_init.pdb -o 1l2y_ff.pdb -p 1l2y_ff.top -ignh

A detailed description of GROMACS tools can be found on http://manual.gromacs.org. *-f* is an option of pdb2gmx which takes as an argument the file name with the initial molecule structure. The output structure and topology file names are give as arguments of -o and -p options respectively. The hydrogen atoms in the structure are ignored due to presence of *-ignh* option and new ones are added according to the selected force field. If option -ff is omitted the program will ask the user to select a force field for the molecules:

Select the Force Field:
From '/usr/share/gromacs/top':
 1: AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
 2: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)

3: AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
4: AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
5: AMBER99SB protein, nucleic AMBER94 (Hornak et al., Proteins 65, 712-725, 2006)
6: AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)
7: AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
8: CHARMM27 all-atom force field (with CMAP) - version 2.0
9: GROMOS96 43a1 force field
10: GROMOS96 43a2 force field (improved alkane dihedrals)
11: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
12: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
13: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
14: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
15: [DEPRECATED] Encad all-atom force field, using full solvent charges
16: [DEPRECATED] Encad all-atom force field, using scaled-down vacuum charges
17: [DEPRECATED] Gromacs force field (see manual)
18: [DEPRECATED] Gromacs force field with hydrogens for NMR

Let us use *CHARMM27* by entering 8 and pressing the "Enter" key. The next step is to choose a force field for water molecules yet to be added explicitly later on.

Select the Water Model:
1: TIP3P   TIP 3-point, recommended
2: TIP4P   TIP 4-point
3: TIPS3P CHARMM TIP 3-point with LJ on H's (note: twice as slow in GROMACS)
4: SPC   simple point charge
5: SPC/E   extended simple point charge
6: None

It should be noted that the water model has to be compatible with the selected force field for the protein (or/and RNA, DNA etc.) and if the user does not have any arguments to do so he should select the recommended list entry.

The next step is to define the size and form of the simulation box by using *editconf* tool. GROMACS supports four box types namely *triclinic, cubic, dodecahedron* and *octahedron* which are acceptable as arguments of *-bt* option. The option *-d* specifies the minimum distance between protein molecule and the box in nanometers. The main consideration of the box type is the symmetry of the molecule. The aim is to minimize the number of water molecules and to have adequately chosen distance to the box.

> *editconf -f 1l2y_ff.pdb -o 1l2y_ff_cubic.pdb -bt cubic -d 1*

-f takes the name of the input file, *-o* output file.
   In order to have a backup copy of the topology file to be modified in the next step it is recommended to copy *1l2y_ff.top* to *1l2y_ff_cubic.top.*

> *cp 1l2y_ff.top 1l2y_ff_cubic.top*

The already defined simulation box could be filled up with water by means of the *genbox* tool.

> *genbox -cp 1l2y_ff_cubic.pdb -cs -p 1l2y_ff_cubic.top -o 1l2y_wb.pdb*

-cp takes the name of the coordinate file with defined box, *-cs* optionally takes a coordinate file
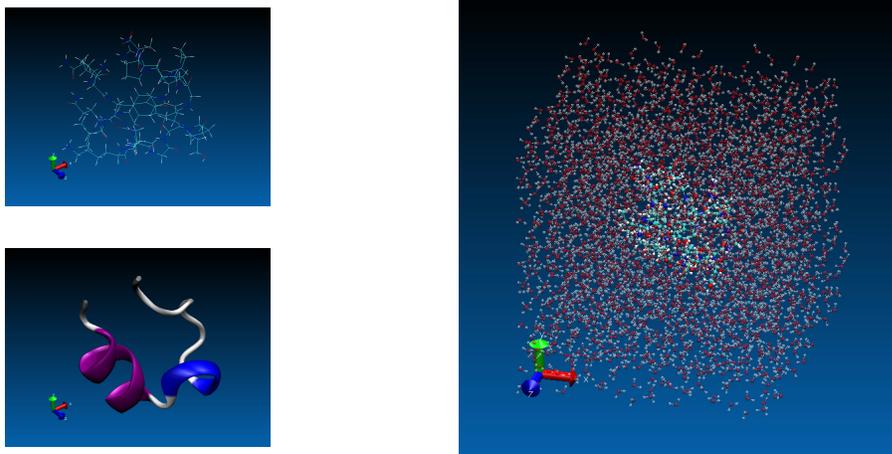
with solvent template and if used without argument the default template comming with GROMACS installation is used, *-p* takes the topology file name which is modified according to the number of added water molecules and with *-o* the name of the output system is given. It is better to have the topology file of the solvated system named with the same base name as coordinate one.

cp 1l2y_ff_cubic.top 1l2y_wb.top

In order to start proper integration of the equations of motion the system has be in state which is as close as possible to the state with minimal potential energy. GROMACS allows user to minimize the potential energy of the system using one of the three algorithms: steepest descent, conjugate gradient and the low-memory Broyden-Fletcher-Goldfarb-Shanno approach.

mdrun is the main program of the GROMACS software package which does energy minimization and supports different integrators of the equations of motion. It requires an input file wherein the system coordinates, box type, boudary conditions, topology, molecular dynamics parameters, temperature and pressure control etc. are describes. This file is prepared by grompp program which takes atom coordinates, topology file and file with MD parameters as input.

The simplest mdp (extension of the file where user can set desired MD parametes) for energy minimization looks like follows:

```
cpp            = /usr/bin/cpp
define         = -DFLEX_SPC
constraints     = none
integrator      = steep
nsteps         = 500
;
;       Energy minimizing stuff
;
emtol          = 500
emstep          = 0.01

nstcomm         = 1
ns_type         = grid
```

```
rlist        = 1
rcoulomb      = 1.0
rvdw         = 1.0
Tcoupl        = no
Pcoupl        = no
gen_vel        = no
```

Lines begin with ";" are comments. "*integrator = steep*" means that steepest descent method will be used for energy minimization, "*nsteps*" sets the maximum number of iterations, the maximum step size is "*emstep*" [nm] and the tolerance is "*emtol*" [kJ mol^{-1} nm^{-1}]. If the parameters are saved in file named *em.mdp*, for example, the next command line will produce *em.tpr* file for energy minimization:

> *grompp -f em.mdp -c 1l2y_wb.pdb -p 1l2y_wb.top -o em*

The energy minimization run command line:

> *mdrun -deffnm em -v -c em.pdb*

The output of the mdrun program in this case includes files em.trr, em.edr, em.log and em.pdb. *-deffnm* means that the base name of the all input and output file is *em*. The trajectory is written in em.trr file and different energy terms, temperature, pressure etc. - in *em.edr*. The user is able to extract the information from *.edr* files using the *g_energy* GROMACS tool.
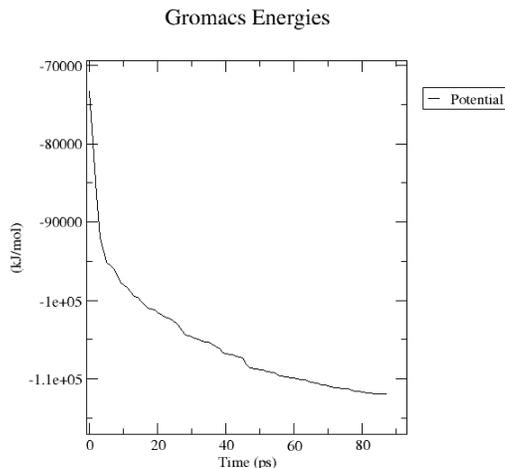
The aim of the next step is relaxation of solvent molecules around the protein and a possible way to do this with GROMACS is position restrain of the protein atoms done by putting protein atoms in very narrow and deep potential traps. The user can enable the feature setting "define   = -DPOSRES" in the following pr.mdp example file:

```
title        = pr run
cpp         = /usr/bin/cpp
define        = -DPOSRES
constraints     = all-bonds
integrator      = md
dt          = 0.002   ; ps !
nsteps       = 32000
```

```
nstcomm         = 1
nstxout         = 1000
nstvout         = 0
nstfout         = 0
nstlog          = 10
nstenergy       = 10
nstlist         = 10
ns_type         = grid
rlist           = 1.0
rcoulomb        = 1.0
rvdw            = 1.0

Tcoupl          = berendsen
tc-grps         = Protein  non-protein
tau_t           = 0.1     0.1
ref_t           = 300     300
refcoord-scaling    = all
; Energy monitoring
energygrps      = Protein  SOL

Pcoupl          = berendsen
tau_p           = 0.5
compressibility     = 4.5e-5
ref_p           = 1.0
; Generate velocites is on at 300 K.
gen_vel         = yes
gen_temp        = 300.0
gen_seed        = -1
```

Berendsen thermostat and Berensen barostat control the temperature during the simulation - Tcoupl and Pcoupl parameters respectively. GROMACS allows temperature control by coupling of different thermostats to different groups of molecular system - in this example protein atoms and water atoms - parameter tc-grps. Berendsen thermostat is the best choice during equilibration. "constraints = all-bonds", "integrator = md", "dt = 0.002", "nsteps =32000" means than all bonds are kept with constant length during the integration with leap-frog algorith with time step 2 fs for 32000 steps - total simulation time equals to 64 *ps*. "*gen_vel = yes*", "*gen_temp = 300.0*", "*gen_seed = -1*" - the initial values of the atoms velocities with distribution corresponding to 300 K will be generated and the random generator seed will be set to the pid of the process.
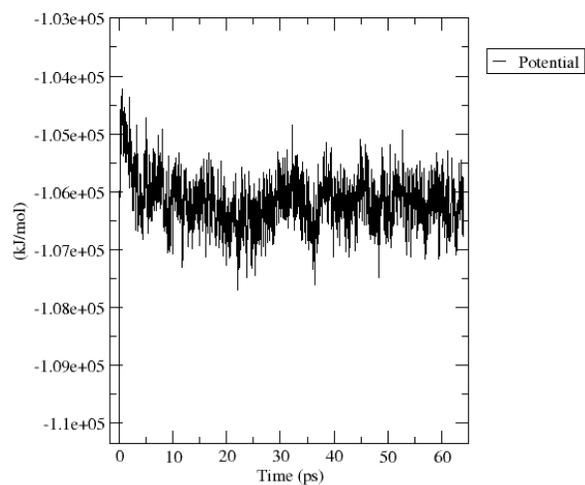
 After produsing pr.tpr file for position restrain run:
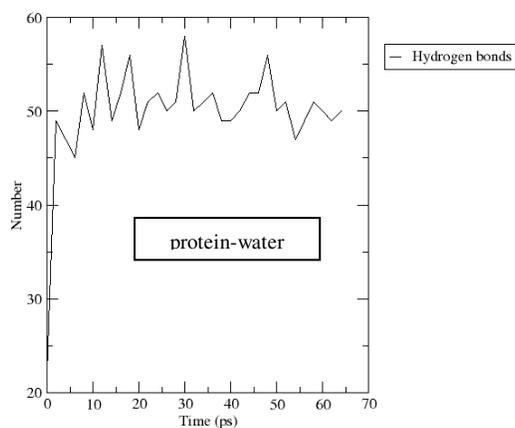
   grompp -f pr.mdp -c em.pdb -p 1l2y_wb.top -o pr

if can be strarted by the command line:

   mdrun -deffnm pr -v -c pr.pdb

Gromacs Energies

Hydrogen Bonds

The coordinates after the equilibration run are saved in pr.pdb file and they are the initial values of atomic positions for molecular dynamics production run. Again simple example of mdp file for md run is:

title = MD run 1l2y
cpp = /lib/cpp
constraints = all-bonds
integrator = md
dt = 0.002 ; ps !
nsteps = 1000000
nstcomm = 1
nstxout = 5000
nstvout = 0
nstfout = 0
nstlist = 10

ns_type = grid
rlist = 0.9
coulombtype = PME
rcoulomb = 0.9
rvdw = 0.9

Tcoupl = nose-hoover
tau_t = 0.4 0.4
tc-grps = protein Non-Protein
*ref_t = 300 300*

Pcoupl = Parrinello-Rahman
pcoupltype = isotropic
tau_p = 0.5
compressibility = 4.5e-5 ; at 300 K
ref_p = 1.0
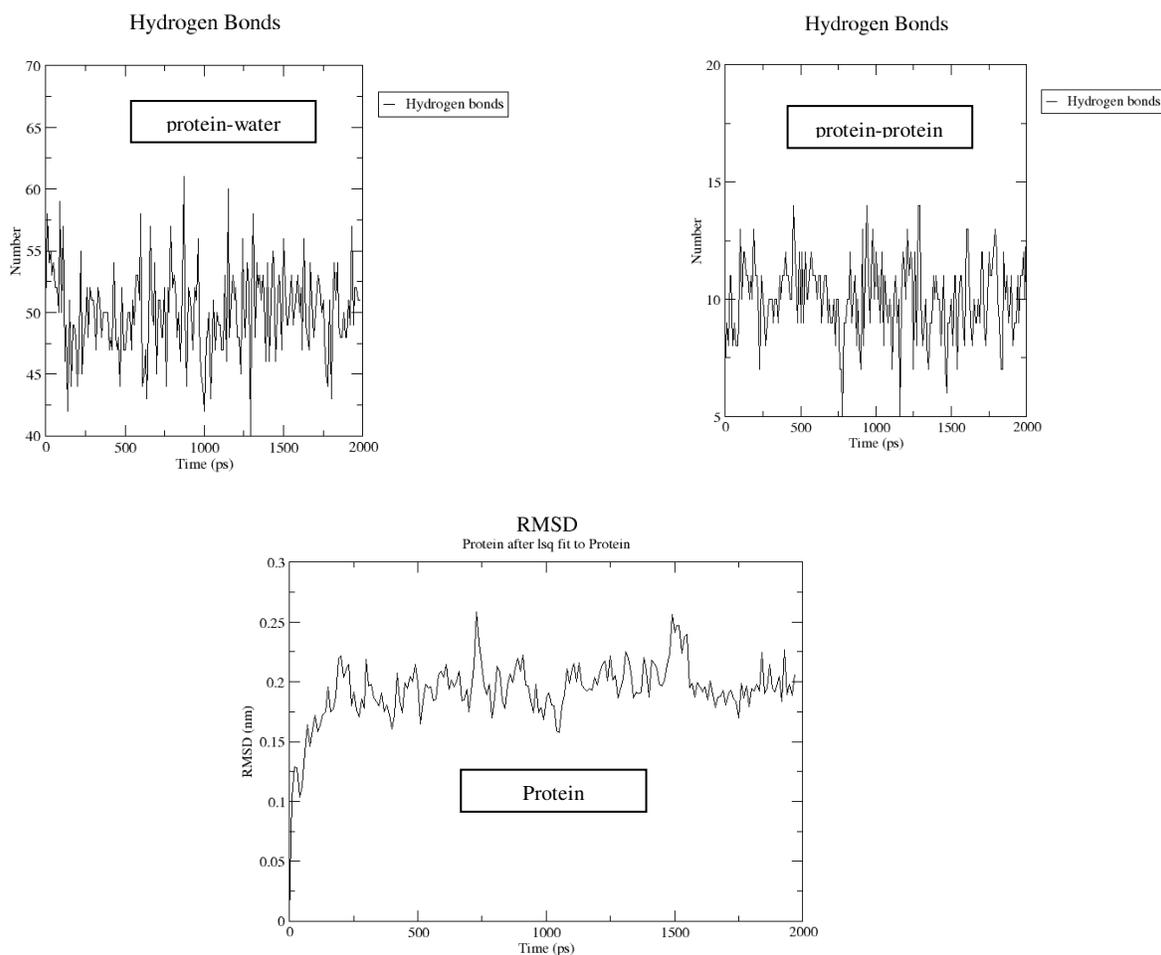
gen_vel = yes
gen_temp = 300.0
gen_seed = -1

One can mention different from equilibration run termostat and barostat. According to the GROMACS manual these ones preproduse better ensemble. For better treatment of the long range electrostatics Particle Mesh Ewald (PME) is used.

grompp -f md.mdp -c pr.pdb -p 1l2y_wb.top -o md

mdrun -deffnm md -v -c md.pdb
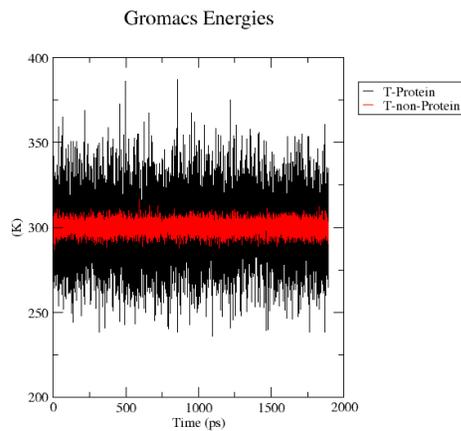
Some quantities like temperature of the T-coupled groups, potential energy, RMSD, number of hydrogen bonds during the simulation are useful to check if there is equilibrium in the system. As one can see in the plots they reach plateau and fluctuate around certain value. The plots are made by g_hbond, g_energy and g_rms tools with md.trr, md.edr and md.tpr files as input and then figures are produced by xmgrace. The messages that GROMACS tools print on the screen are very helpful and user friendly.

Hydrogen Bonds

protein-water

Hydrogen Bonds

protein-protein

RMSD
Protein after lsq fit to Protein

Protein

The example files like those included in this tutorial can be found on www.gromacs.org and we advice reader to have a look at the GROMACS manual for more information.