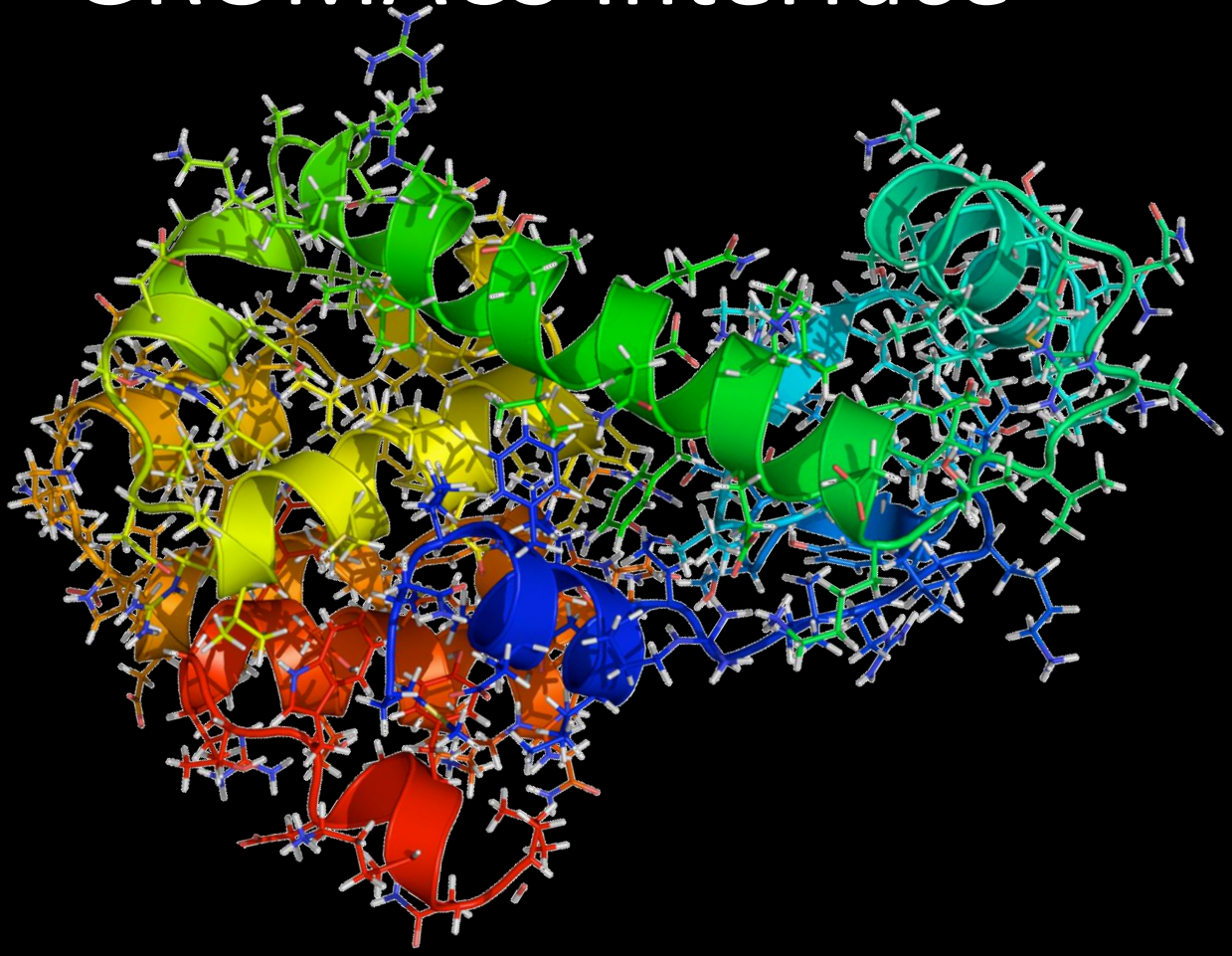
Performing MD Simulations: GROMACS Interface



SETS Summer Workshop

July 9-12, 2012, 201 Link Hall, Syracuse University

Objective

This tutorial is designed to give an overview for the typical steps used in practical simulations using GROMACS software

We will perform a very short sample simulation

You will also have access to a slightly longer pre-calculated trajectory for analysis

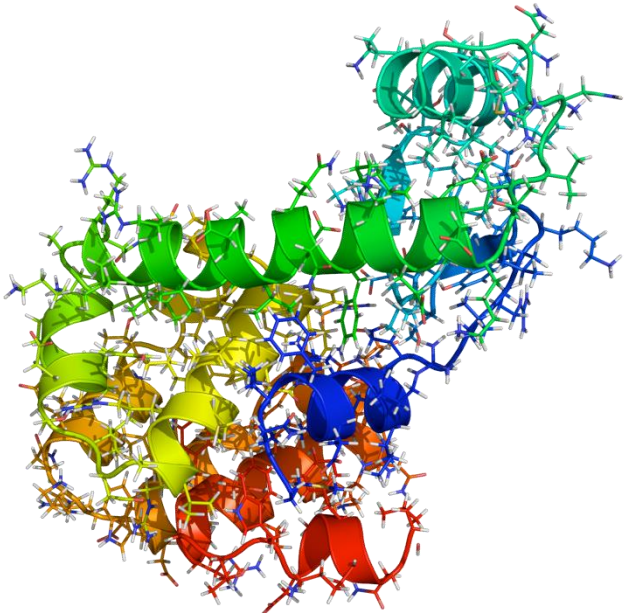
Lysozyme

Lysozyme is a fascinating enzyme that has ability to kill bacteria (kind of the body's own antibiotic), and is present e.g. in tears, saliva, and egg white.

It was discovered by Alexander Fleming in 1922, and one of the first protein X-ray structures to be determined (David Phillips, 1965).

It is fairly small by protein standards

- 164 amino acid residues
- consists of 2890 atoms



Tutorial steps

1. Look at the structure
2. Prepare the input files necessary for simulation
3. Solvate the structure in water
4. Minimize
5. Equilibrate
6. Perform a short production simulation
7. Analysis of results

1. Look at the structure

We will look for the (Protein Data bank) PDB file of Lysozyme

A. Go to the Protein Data Bank at http://www.pdb.org
   Search for “Lysozyme”

B. There will be lots of hits for the search
   Use the structure 1LYD.pdb

C. Load the file in VMD and visualize the Lysozyme structure

Files needed to run GROMACS

Essential files

1. gro file: list all particles (atoms or beads) their x, y, z coordinates, (may or may not) v_x, v_y, v_z velocities
2. top file: Describes the bonding parameters of molecules
3. itp file: Include topology file
4. mdp file: contains run parameters

Output/run files

1. ndx file : index file
2. trr file: trajectory file (coordinates, velocities, forces)
3. xtc file: x, y, z coordinates for trajectory
4. log file: contains details of the run
5. …
6. …
7. …

Example of .gro file

http://manual.gromacs.org/current/online/gro.html

Files with the gro file extension contain a molecular structure in Gromos87 format. gro files can be used as trajectory by simply concatenating files. An attempt will be made to read a time value from the title st

A sample piece is included below:

```
MD of 2 waters, t= 0.0
    6
    1WATER  OW1    1   0.126   1.624   1.679  0.1227 -0.0580  0.0434
    1WATER  HW2    2   0.190   1.661   1.747  0.8085  0.3191 -0.7791
    1WATER  HW3    3   0.177   1.568   1.613 -0.9045 -2.6469  1.3180
    2WATER  OW1    4   1.275   0.053   0.622  0.2519  0.3140 -0.1734
    2WATER  HW2    5   1.337   0.002   0.680 -1.0641 -1.1349  0.0257
    2WATER  HW3    6   1.326   0.120   0.568  1.9427 -0.8216 -0.0244
   1.82060   1.82060   1.82060
```

Lines contain the following information (top to bottom):

- title string (free format string, optional time in ps after 't=')
- number of atoms (free format integer)
- one line for each atom (fixed format, see below)
- box vectors (free format, space separated reals), values: v1(x) v2(y) v3(z) v1(y) v1(z) v2(x) v2(z) v3(x) v3(y), the last 6 values may be omitted (they will be set to zero). Gromacs only supports boxes

This format is fixed, ie. all columns are in a fixed position. Optionally (for now only yet with trjconv) you can write gro files with any number of decimal places, the format will then be n+5 positions with n deci be inferred from the distance between the decimal points (which will be n+5). Columns contain the following information (from left to right):

- residue number (5 positions, integer)
- residue name (5 characters)
- atom name (5 characters)
- atom number (5 positions, integer)
- position (in nm, x y z in 3 columns, each 8 positions with 3 decimal places)
- velocity (in nm/ps (or km/s), x y z in 3 columns, each 8 positions with 4 decimal places)

mdp file format

Main Table of Contents
GROMACS homepage

Follow this link for a detailed description of the options.

Below is a sample mdp file. The ordering of the items is not important, but if you enter the same thing twice, the

The values of the options are reasonable values for a 1 nanosecond MD run of a protein in a box of water.

```
title                    = Yo
cpp                      = /lib/cpp
include                  = -I../top
define                   =
integrator               = md
dt                       = 0.002
nsteps                   = 500000
nstxout                  = 5000
nstvout                  = 5000
nstlog                   = 5000
nstenergy                = 250
nstxtcout                = 250
xtc_grps                 = Protein
energygrps               = Protein   SOL
nstlist                  = 10
ns_type                  = grid
rlist                    = 0.8
coulombtype              = cut-off
rcoulomb                 = 1.4
rvdw                     = 0.8
tcoupl                   = Berendsen
tc-grps                  = Protein       SOL
tau_t                    = 0.1   0.1
ref_t                    = 300   300
Pcoupl                   = Berendsen
tau_p                    = 1.0
compressibility          = 4.5e-5
ref_p                    = 1.0
gen_vel                  = yes
gen_temp                 = 300
gen_seed                 = 173529
constraints              = all-bonds
```

Not important for our purpose

Time step 2 fs

Number of steps

Output parameters

Cut-off parameters

Temperature coupling

Pressure coupling

Overall simulation = 2fs* 500000 = 1 ns

top file format

Main Table of Contents
GROMACS homepage

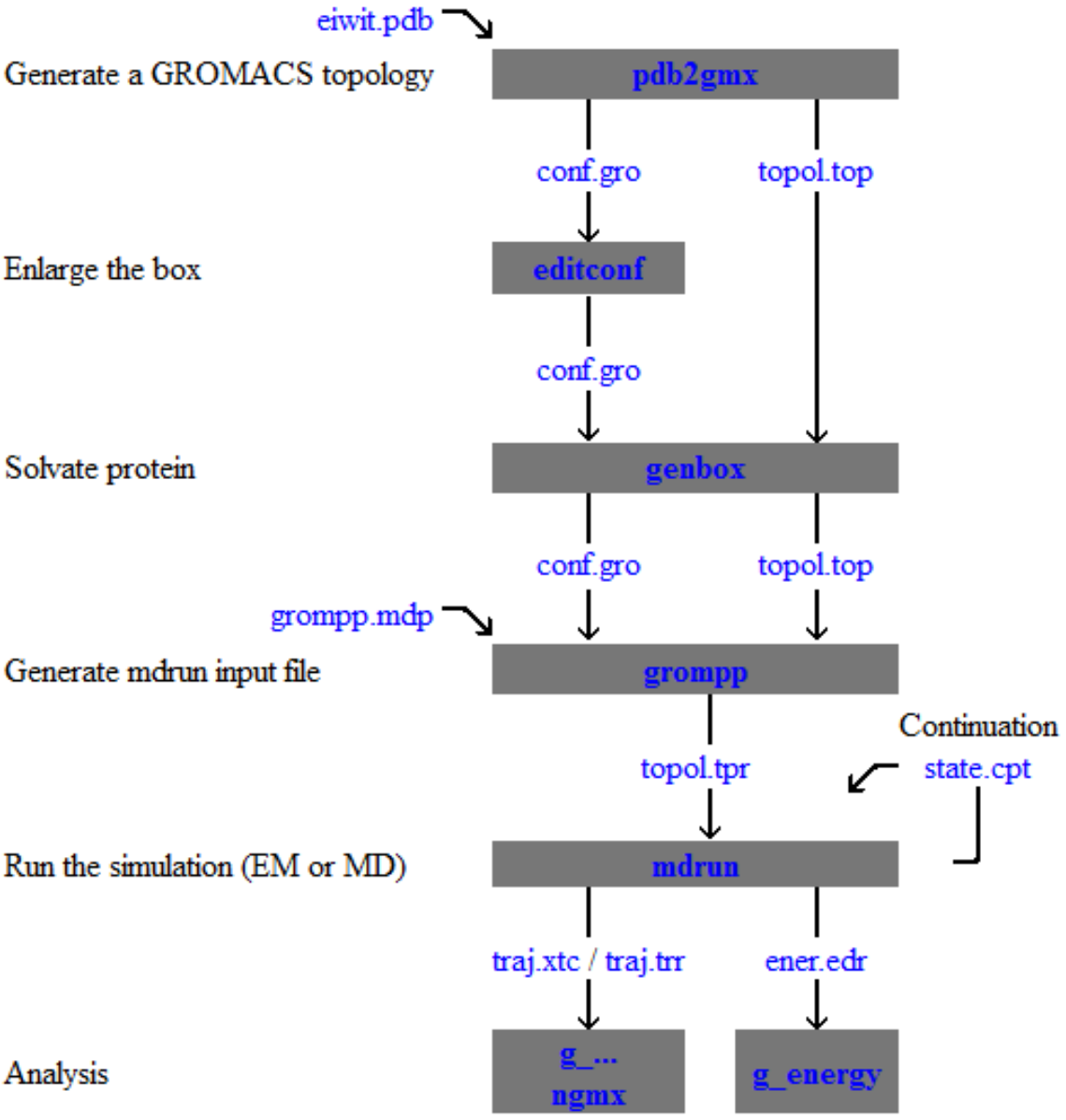
http://manual.gromacs.org/online/top.html

Description

The top file extension stands for topology. It is an ascii file which is read by grompp which processes it and creates a binary topology (tpb-file).
A sample file is included below:

```
;
;       Example topology file
;
[ defaults ]
; nbfunc        comb-rule       gen-pairs       fudgeLJ fudgeQQ
  1             1               no              1.0     1.0

; The force field files to be included
#include "rt41c5.itp"

[ moleculetype ]
; name  nrexcl
Urea         3

[ atoms ]
;   nr    type   resnr  residu    atom    cgnr  charge
     1       C       1    UREA      C1       1   0.683
     2       O       1    UREA      O2       1  -0.683
     3      NT       1    UREA      N3       2  -0.622
     4       H       1    UREA      H4       2   0.346
     5       H       1    UREA      H5       2   0.276
     6      NT       1    UREA      N6       3  -0.622
     7       H       1    UREA      H7       3   0.346
     8       H       1    UREA      H8       3   0.276

[ bonds ]
;  ai    aj funct           c0           c1
    3     4     1 1.000000e-01 3.744680e+05
    3     5     1 1.000000e-01 3.744680e+05
    6     7     1 1.000000e-01 3.744680e+05
    6     8     1 1.000000e-01 3.744680e+05
    1     2     1 1.230000e-01 5.020800e+05
    1     3     1 1.330000e-01 3.765600e+05
    1     6     1 1.330000e-01 3.765600e+05

[ pairs ]
;  ai    aj funct           c0           c1
    2     4     1 0.000000e+00 0.000000e+00
    2     5     1 0.000000e+00 0.000000e+00
    2     7     1 0.000000e+00 0.000000e+00
    2     8     1 0.000000e+00 0.000000e+00
```

GROMACS flowchart



2. Prepare the necessary input files

A. Creating a Gromacs topology from the PDB file using the tool “pdb2gmx”
   - Use a command:

     pdb2gmx -f 1LYD.pdb -water tip3p –o conf.gro

B. The program will ask you for a force field
   - Use Option 5 “OPLS-AA/L”
   - Press “Enter”
   - The program write out lots of information. The important thing to look for is if there were any errors or warnings (it will say at the end).

C. Use (ls command) to check for these files

   conf.gro

   topol.top

   posre.itp

3. Adding solvent water around the protein

A. To further reduce the volume of the box we'll use a rhombic dodecahedron box:

editconf –f conf.gro –bt dodecahedron –d 0.5 –o box.gro

B. In GROMACS, solvate the new box using

genbox –cp box.gro –cs spc216.gro –p topol.top –o solvated.gro

C. Transfer solvated.gro to your desktop computer and visualize it in VMD

4. Energy minimization

To minimize we will use the GROMACS em.mdp file

Type vi em.mdp to see the contents of the em.mdp file

```
integrator = steep
nsteps = 200
nstlist = 10
rlist = 1.0
coulombtype = pme
rcoulomb = 1.0
vdw-type = cut-off
rvdw = 1.0
nstenergy = 10
```

Use commands on the to run the energy simulation

1. grompp –f em.mdp –p topol.top –c solvated.gro –o em.tpr

   Press enter

2. mdrun –v –deffnm em

   Press enter

5. Equilibrate

Perform an equilibration run where all heavy protein atoms are restrained to their starting positions while the water is relaxing around the structure.

Temperature and pressure are maintained at desired values

To equilibrate we will use the GROMACS pr.mdp file

Use “vi” to see the contents of the pr.mdp file

1. grompp –f pr.mdp –p topol.top –c saved_em.gro –o pr.tpr
   Press Enter
2. mdrun –v –deffnm pr
   Press Enter


```
------pr.mdp------
integrator = md
nsteps = 2500
dt = 0.002
nstlist = 10
rlist = 1.0
coulombtype = pme
rcoulomb = 1.0
vdw-type = cut-off
rvdw = 1.0
tcoupl = Berendsen
tc-grps = protein non-protein
tau-t = 0.1 0.1
ref-t = 298 298
Pcoupl = Berendsen
tau-p = 1.0
compressibility = 5e-5 5e-5 5e-5 0 0 0
ref-p = 1.0
nstenergy = 100
define = -DPOSRES
------------------
```

5. Production simulation Run

Use the GROMACS run.mdp file

Use “vi” to see the contents of the run.mdp file

1. grompp –f run.mdp –p topol.top –c saved_pr.gro –o run.tpr
   Press Enter
2. mdrun –v –deffnm run
   Press Enter


```
------run.mdp------
integrator = md
nsteps = 5000
dt = 0.002
nstlist = 10
rlist = 1.0
coulombtype = pme
rcoulomb = 1.0
vdw-type = cut-off
rvdw = 1.0
tcoupl = Berendsen
tc-grps = protein non-protein
tau-t = 0.1 0.1
ref-t = 298 298
nstxout = 1000
nstvout = 1000
nstxtcout = 100
nstenergy = 100
-----------------
```

6. Analysis of results

A. Making a movie of the simulation

Using Xshell transfer saved_run.gro and saved_run.xtc on the windows desktop

Load these files in vmd

Choose the display settings

Play movie