Setting up MD simulations with GROMACS 2016.1

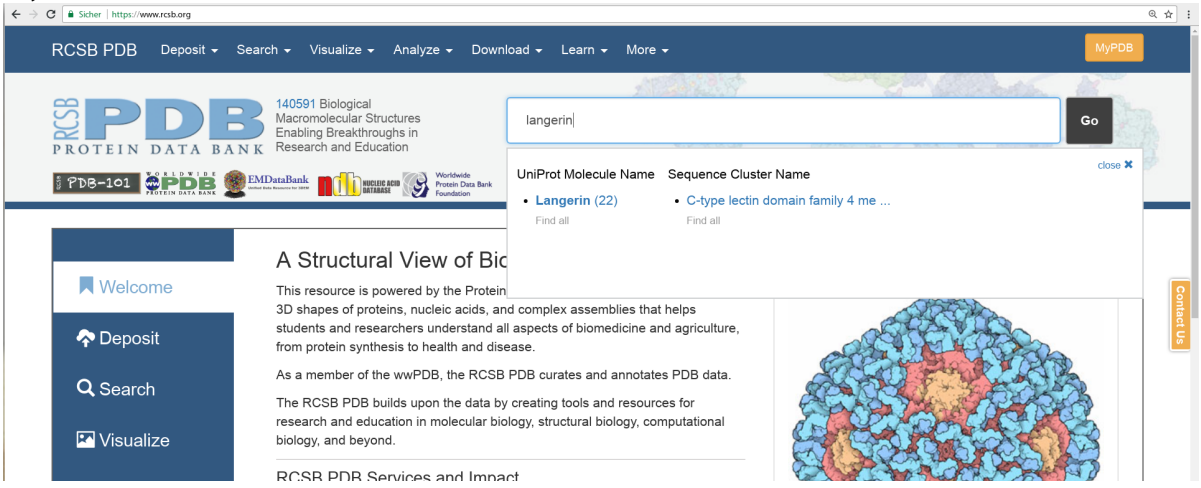
Jan Joswig, Freie Universität Berlin

v. May 24, 2018

For a useful tutorial see also:
http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/index.html

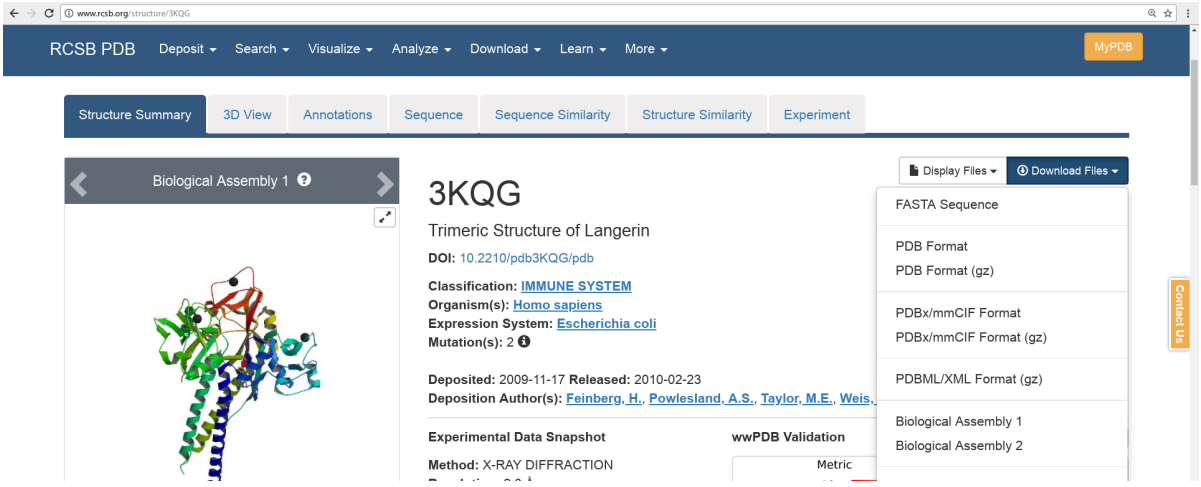
Starting structure

Crystal and NMR structures can be found in the RCSB Protein Data Bank (PDB):



Starting structure

A commonly used file format is .pdb:



PDB file

```
HEADER      SUGAR BINDING PROTEIN 08-OCT-10     XXXX
[...]
SITE        1 BC3 6 GLU D 285 ASN D 287 GLU D 293 ASN D 307
SITE        2 BC3 6 ASP D 308 FUC D 402
CRYST1      79.860     79.860     90.170     90.00     90.00     90.00 P 42         16
ORIGX1          1.000000   0.000000   0.000000         0.00000
ORIGX2          0.000000   1.000000   0.000000         0.00000
ORIGX3          0.000000   0.000000   1.000000         0.00000
SCALE1          0.012522   0.000000   0.000000         0.00000
SCALE2          0.000000   0.012522   0.000000         0.00000
SCALE3          0.000000   0.000000   0.011090         0.00000
MODEL          1
ATOM        1  N   GLY A 198      1.923   33.617   6.889   1.00 29.44        N
ATOM        2  CA  GLY A 198      2.927   34.067   5.942   1.00 33.42        C
ATOM        3  C   GLY A 198      4.064   33.079   5.747   1.00 21.30        C
ATOM        4  O   GLY A 198      5.016   33.355   5.022   1.00 26.76        O
ATOM        5  N   TRP A 199      3.964   31.919   6.391   1.00 19.08        N
[...]
HETATM   1287  O   HOH A 813     20.169   44.865   0.963   1.00 30.98        O
HETATM   1288  O   HOH A 814     18.723   22.257   6.152   1.00 38.17        O
ENDMDL
MASTER         0    0    0   20   40    0   28    6 5094    4    0   44
END
```

Lines beginning with keywords HEADER, AUTHOR, ... can be deleted for MD setup. We are only interested in atomic coordinates

```
HEADER      SUGAR BINDING PROTEIN 08-OCT-10     XXXX
[...]
SITE        1 BC3 6 GLU D 285 ASN D 287 GLU D 293 ASN D 307
SITE        2 BC3 6 ASP D 308 FUC D 402
CRYST1      79.860     79.860     90.170     90.00     90.00     90.00 P 42         16
ORIGX1          1.000000   0.000000   0.000000         0.00000
ORIGX2          0.000000   1.000000   0.000000         0.00000
ORIGX3          0.000000   0.000000   1.000000         0.00000
SCALE1          0.012522   0.000000   0.000000         0.00000
SCALE2          0.000000   0.012522   0.000000         0.00000
SCALE3          0.000000   0.000000   0.011090         0.00000
MODEL           1
ATOM        1  N   GLY A 198      1.923   33.617   6.889   1.00 29.44        N
ATOM        2  CA  GLY A 198      2.927   34.067   5.942   1.00 33.42        C
ATOM        3  C   GLY A 198      4.064   33.079   5.747   1.00 21.30        C
ATOM        4  O   GLY A 198      5.016   33.355   5.022   1.00 26.76        O
ATOM        5  N   TRP A 199      3.964   31.919   6.391   1.00 19.08        N
[...]
HETATM   1287  O   HOH A 813     20.169   44.865   0.963   1.00 30.98        O
HETATM   1288  O   HOH A 814     18.723   22.257   6.152   1.00 38.17        O
ENDMDL
MASTER        0    0    0   20   40    0   28    6 5094    4    0   44
END
```

Help under http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html

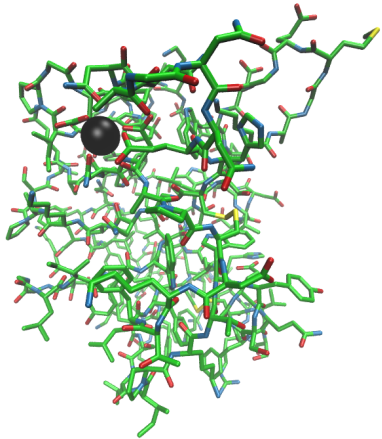
keyword, atom ID, atom name, residue name, chain ID, residue ID, x-, y-, z-coordinate, occupancy, B-factor, element

```
ATOM        1  N   GLY A 198      1.923   33.617   6.889   1.00 29.44        N
ATOM        2  CA  GLY A 198      2.927   34.067   5.942   1.00 33.42        C
ATOM        3  C   GLY A 198      4.064   33.079   5.747   1.00 21.30        C
ATOM        4  O   GLY A 198      5.016   33.355   5.022   1.00 26.76        O
ATOM        5  N   TRP A 199      3.964   31.919   6.391   1.00 19.08        N
[...]
HETATM   1287  O   HOH A 813     20.169   44.865   0.963   1.00 30.98        O
HETATM   1288  O   HOH A 814     18.723   22.257   6.152   1.00 38.17        O
END
```

Hydrogen atoms are usually not included in X-ray data. Crystal waters, ligands and other heteroatoms can be deleted as needed.

Visualisation

The structure can be visualised in your favourite viewer (for example VMD, http://www.ks.uiuc.edu/Research/vmd/)



GROMACS

GROMACS (http://www.gromacs.org) can be used for structure preparation, simulation and analysis. It has no graphical user interface. Every operation is executed via a terminal command of the form:

```
gmx command -flag (optional) keyword
```

Try for example `gmx -version` to see which GROMACS you are currently using. Use http://manual.gromacs.org/documentation/2016-current/index.html, if you get lost.

You can use version 2016.1 installed on the PCs in the PC-pool, Fabeckstr. 36a, 304. Connect from your own PC via:

`ssh -X user@login.bcp.fu-berlin.de` and than
`ssh -X user@pool<XY>.bcp.fu-berlin.de`,

where <XY> is a number between 02 and 14. Leave the PCs on after your tutorial, so you can connect to them later.

Structure preparation

A standard PDB file can be prepared as starting structure for a simulation with `pdb2gmx`. This will generate the necessary files:

A processed coordinate file .gro (or .pdb).

A topology file .top – This contains information about how to treat the molecule represented in the structure during the simulation. It holds parameters for bonded and non-bonded interactions (or links to corresponding force field files).

Supplementary topology files .itp – Topology files for different parts of the system. These files can be also optionally used during a simulation to fix the positions of certain atoms.

Try:

```
gmx pdb2gmx -f protein.pdb -o protein.gro -p protein.top -water tip3p -ignh
-ff AMBER99SB-ILDN -his
```

The flags translate to:

`-f` primary input file, `-o` primary output file, `-p` topology output, `-water` water model, `-ff` force field, `-his` choose HIS protonation interactively, `-ignh` ignore hydrogens in the input file.

Structure file .gro

```
C-type lectin domain family 4 member K
2016
  198GLY    N      1    0.192  3.362  0.689
  198GLY    H1     2    0.121  3.431  0.697
  198GLY    H2     3    0.152  3.276  0.657
  198GLY    H3     4    0.235  3.347  0.778
  198GLY    CA     5    0.293  3.407  0.594
  198GLY    HA1    6    0.329  3.493  0.628
  198GLY    HA2    7    0.247  3.422  0.507
  198GLY    C      8    0.406  3.308  0.575
  198GLY    O      9    0.502  3.336  0.502
  199TRP    [...]
```

Topology file .top

```
; Include forcefield parameters
#include "amber99sb-ildn.ff/forcefield.itp"

; Include chain topologies
#include "topol_Protein_chain_A.itp"
#include "topol_Ion_chain_A2.itp"

; Include water topology
#include "amber99sb-ildn.ff/tip3p.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
; i funct fcx fcy fcz
    1 1 1000 1000 1000
#endif

; Include topology for ions
#include "amber99sb-ildn.ff/ions.itp"

[ system ]
; Name
C-type lectin domain family 4 member K

[ molecules ]
; Compound     #mols
Protein_chain_A   1
Ion_chain_A2   1
```

lines starting with ; are just comments
Where to find force field parameters
lines starting with # are GROMACS internal (compiler-like) directives

File names of position restraints

Where to find water parameters

What to do, if position restraints are activated for water

Put large energy penalty on translation in every direction

Where to find ion parameters

How many molecules of which kind are in the system

Additional topology files .itp

```
[atoms ] ; nr      type      resnr      residue      atom      cgnr      charge      mass      typeB      chargeB
; residue 198 GLY rtp NGLY q +1.0
  1      N3      198      GLY      N      1      0.2943      14.01      ; qtot 0.2943
  2      H       198      GLY      H1     2      0.1642      1.008      ; qtot 0.4585

[ bonds ] ; ai      aj      funct      c0      c1      c2      c3
  1      2      1
  1      3      1
  1      4      1

[ angles ]
  [...]

[ dihedrals ]
  [...]
```

Putting the structure in a simulation box

The next step is to define a simulation box around the structure:

```
gmx editconf -f protein.gro -o protein_box.gro -c -bt cubic -d 1
```

`-c` puts the molecule in the centre of the box, `-bt` type of box, `-d` minimum distance of molecule to the borders of the box (important for pbc)

Minimisation in vacuum

Removes steric clashes or problematic geometric arrangements in the starting structure (when necessary). GROMACS uses a preprocessor (grompp) to prepare any minimisation or simulation runs. This produces a run input file .tpr from coordinates, topology and run instructions passed in a .mdp file.

```
gmx grompp -f steep.mdp -c protein_box.gro -o protein_vmin.tpr -p topol.top
```

`-f` simulation parameters, `-c` structure file

The actual run is then started by:

```
gmx mdrun -v -deffnm protein_vmin
```

`-v` verbose, `-deffnm` job name

Simulation parameters - Steepest decent minimisation

```
; Parameters describing what to do, when to stop and what to save
integrator      = steep         ; Algorithm (steep = steepest descent minimisation)
emtol           = 1000.0        ; Stop minimization when the maximum force < 1000.0 kJ/(mol nm)
emstep          = 0.01          ; Energy step size
nsteps          = 10000         ; Maximum number of (minimisation) steps to perform

; Parameters describing how to find the neighbours of each atom and how to calculate the interactions
nstlist         = 10            ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet
ns_type         = grid          ; Method to determine neighbor list (simple, grid)
rlist           = 1.0           ; Cut-off for making neighbor list (short range forces)
coulombtype     = PME           ; Treatment of long range electrostatic interactions
vdwtype         = Cut-off       ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0           ; Short-range Van der Waals cut-off
rvdw            = 1.0           ; Short-range Van der Waals cut-off
pbc             = xyz           ; Periodic Boundary Conditions
```

Simulation output

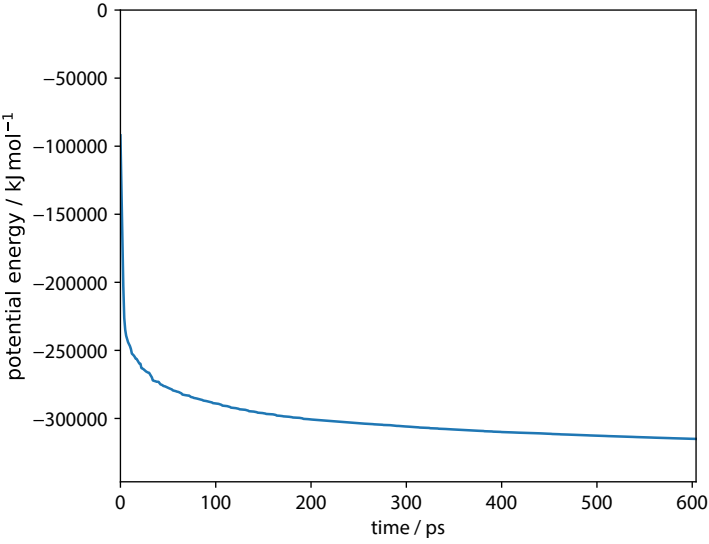
A GROMACS mdrun call produces several files:

- A trajectory file, either full-precision .trr or compressed .xtc (if specified in .mdp)
- A log file .log
- A binary energy file .edr
- A final structure file .gro

From the .edr file a set of useful measures can be extracted:

```
echo "10 2" | gmx energy -f protein_vmin.edr -o energy.xvg
```

Minimisation converged?



Solvation

The molecule in the box can be solvated. We specified earlier to use TIP3P waters.

```
gmx solvate -cp protein_box -cs spc216.gro -o protein_water.gro -p
topol.top
```

`-cp` protein structure file, `-cs` water structure file (spc216 is a generic one for 3-point water that is part of the GROMACS installation), `-p` make sure to specify the right topology to be updated (water molecules will be added under [molecules])

Another minimisation run should be done (Check for convergence, -10^5 to -10^6 KJ/mol is a reasonable value).

Charge neutralisation

When the considered molecule has non-zero total charge, the system should be neutralised by addition of corresponding counterions (Ca^{2+}, Cl^-). This can be done be another call of grompp:

```
gmx grompp -f steep.mdp -c protein_water.gro -o ions.tpr -p topol.top
```

Then, solvent molecules can be uniformly replaced by ions (four chlorides in this case):

```
gmx genion -s ions.tpr -o protein_ions.gro -p topol.top -nname CL -nn 4
```

Equilibration in the *NVT* ensemble

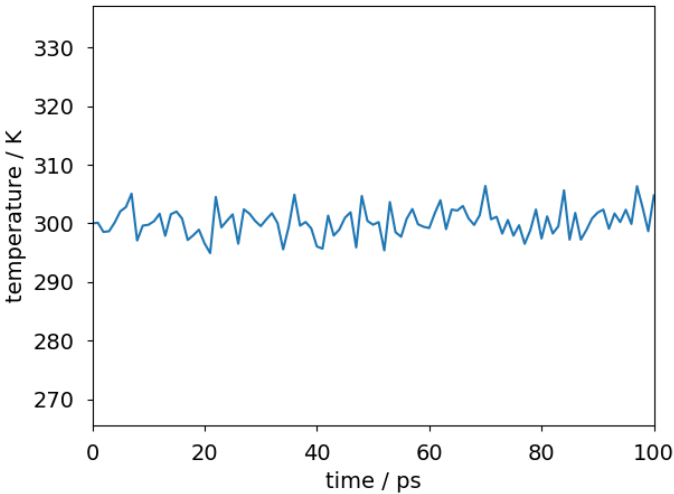
Before the minimised structure in water can be used for an actual production run, the systems temperature should be equilibrated. This is normally done in a short simulation at constant atom number *N*, volume *V* and *T*. The temperature is held constant by coupling the system to a thermostat.

If the system becomes unstable during this run, position restraints can be applied to the protein beforehand, to allow the equilibration of the solvent around the fixated solute first.

Simulation parameter - *NVT*

```
define                  = -DPOSRES              ; position restrain the protein (if necessary)
; Run parameters
integrator              = md                    ; leap-frog integrator
nsteps                  = 50000                 ; 2 fs * 50000 = 200 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             = all-bonds             ; all bonds (even heavy atom-H bonds) constrained
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                    ; largely irrelevant with Verlet
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               = 6                     ; 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                   ; pbc
; Dispersion correction
DispCorr                = EnerPres              ; account for cut-off VdW scheme
; Velocity generation
gen_vel                 = yes                   ; assign velocities from Maxwell distribution
gen_temp                = 300                   ; temperature for Maxwell distribution
gen_seed                = -1                    ; generate a random seed
```

Temperature equilibrated?



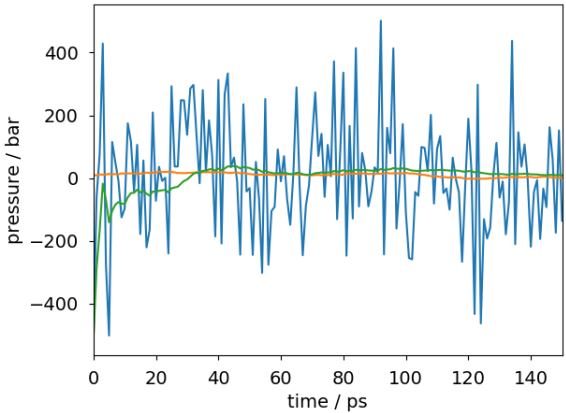
Equilibration in the *NPT* ensemble

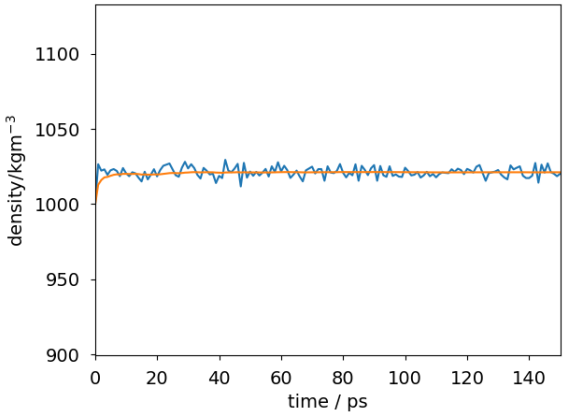
After the temperature has reached a stable plateau, coupling to a barostat can be enabled to equilibrate pressure and density of the system.

Simulation Parameter – *NPT*

```
define                  = -DPOSRES              ; position restrain the protein (if necessary)
; Run parameters
integrator              = md                    ; leap-frog integrator
nsteps                  = 50000                 ; 2 fs * 50000 = 200 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                   ; restart after NVT !!!
constraint_algorithm    = lincs                 ; holonomic constraints
constraints             = all-bonds             ; all bonds (even heavy atom-H bonds) constrained
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                    ; largely irrelevant with Verlet
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               = 6                     ; 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                  = Parinello-Rahman      ; Pressure coupling in NPT !!!
pcoupltype              = isotropic             ; uniform scaling of box vectors !!!
tau_p                   = 2                     ; time for coupling rate in ps !!!
ref_p                   = 1                     reference pressure in bar !!!
compressibility         = 4.5e-5                ; isothermal compressibility of water, bar⁻¹
refcoord_scaling        =                       com
; Periodic boundary conditions
pbc                     = xyz                   ; pbc
; Dispersion correction
DispCorr                = EnerPres              ; account for cut-off VdW scheme
; Velocity generation
gen_vel                 = no                    ; assign velocities from Maxwell distribution !!!
```

Pressure equilibrated? Density correct?





Production simulation

The production run is often also carried out in the *NPT* ensemble, because it most closely resembles experimental conditions.

Position restraints usually have to be removed. To save disk space and minimize I/O load, options can be set to restrict the output to protein coordinates only (without solvent):

```
nstxout-compressed = 500
compressed-x-grps  = Protein
```

Analysis

For most analyses and for visualisation of the simulated trajectories it is necessary to process the raw output data by fixing the periodic boundary conditions and removing translational and rotational degrees of freedom:

```
gmx trjconv -f traj.xtc -o traj_f.xtc -s traj.tpr -pbc mol
gmx trjconv -f traj_f.xtc -o traj_f.xtc -s traj.gro -fit rot+trans
```

A trajectory file can be inspected for its length and its basic content by:

```
gmx check -f traj.xtc
```

Run input files can be checked for more detailed information by:

```
gmx dump -s traj.tpr | more
```