Molecular dynamics simulations of protein-ligand complexes using GROMACS 4.0.5

Walter Filgueira de Azevedo Jr

Porto Alegre – RS, Brazil.

2009

Walter F. de Azevedo Jr.

http://azevedolab.net

Faculdade de Biociências-PUCRS. Av. Ipiranga, 6681, CEP 90619-900,
Porto Alegre, Rio Grande do Sul, Brazil, e-mail: walter@azevedolab.net

Telephone:

+55 51 3353-4529

Index

Brief overview

GROMACS is a program developed for computational molecular dynamics simulation of macromolecules (van der Spoel et al., 2005). It is currently in its 4.0 version (June 2009). The present tutorial describes the use of the program GROMACS 4.0.5 for molecular dynamics simulation of a complex involving protein and ligand. This is a very useful simulation when one wants to analyze in details the dynamics of protein-ligand interactions (De Azevedo, 2008). In this tutorial we carried out MD simulations of the complex involving human cyclin-dependent kinase 2 (CDK2) in complex with roscovitine. Figure 1 shows the binary complex CDK2-roscovitine (PDB access code: 2A4L). Features related to changes in protein structure due to ligand binding can have a dynamic view from this simulation. Furthermore, the energetics of protein-ligand interaction can be calculated, generating results better than the direct application of empirical scoring functions (De Azevedo & Dias, 2008). This tutorial is available for download at http://azevedolab.net/md_75.html .

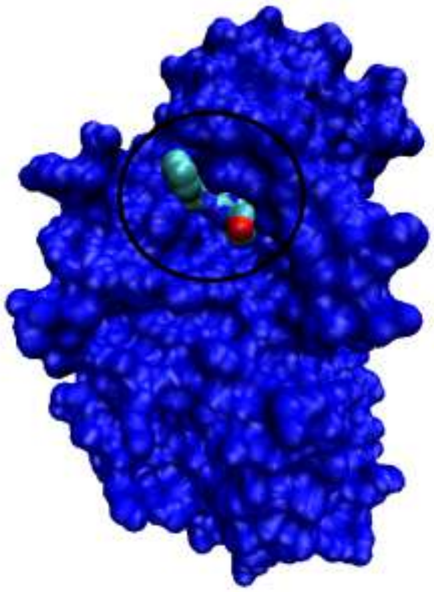

Figure 1. Molecular surface of CDK2 (blue) and the ATP-binding pocket with the structure of roscovitine (indicated by an ellipse).

Files needed to run GROMACS 4.0.5

You will need 6 files described in table 1 to run the present tutorial. The atomic coordinates for prot.pdb and lig.pdb were extracted from the structure of human CDK2 (Cyclin-Dependent Kinase 2) complexed with roscotvitine deposited at PDB (access code: 2A4L) (De Azevedo Jr. *et al*., 1997). These files were directly downloaded from PDB (http://www.rcsb.org) and edited with vi (http://www.vim.org/). The structures of complexes between CDK2 and several different ligands established the basis for understanding of the structural determinants for potency and specificity of CDK inhibitors (De Azevedo Jr. *et al*., 1996). To run this tutorial all these files should be in the same directory and can be downloaded at http://azevedolab.net/md_75.html . In addition to separate the atomic coordinates for ligand (lig.pdb) and protein (protein.pdb) we modeled missing regions in the structure of CDK2. There are two loops in the structure of human CDK2 that are hard to identify in the electron density maps and most of CDK2 structures do not present atomic coordinates for these regions, which need to generated before carrying out molecular dynamics simulations. We used the program MODELLER (Sali & Blundell, 1993) to model these missing loops. The atomic coordinates available in the file protein.pdb brings the complete CDK2 structure.

Table 1. Input files needed to run the present tutorial

File	Description
prot.pdb	Atomic coordinates for the protein, without any ligand. You may directly download a crystallographic structure at PDB and delete all lines except the ones with protein atomic coordinates. You may also use a structure obtained from homology modeling. This file was obtained from the complex CDK2-roscovitine (PDB access code: 2A4L).
lig.pdb	Atomic coordinates for the ligand, without any protein and solvent. You may directly download a crystallographic structure at PDB and delete all lines except the ones with ligand atomic coordinates. This file was obtained from the complex CDK2-roscovitine (PDB access code: 2A4L) and contains the atomic coordinates for roscovitine, as canonical inhibitor of human CDK2.
lig.itp	This file brings information about partial charge, mass and geometric parameters of each atom in the molecule. We may use PRODRG (Schuettelkopf & can Aalten, 2004) to generate this information.
em.mdp	Input file for energy minimization using GROMACS
pr.mdp	Input file for positional restraint minimization using GROMACS
md.mdp	Input file for molecular dynamics using GROMACS

The .mdp files are used for energy minimization (em.mdp), positional restraint minimization (pr.mdp) and molecular dynamics (md.mdp). Full details can be found at gromacs manual (www.gromacs.org).

em.mdp

```
title                   =  prot+lig
cpp                     =  /usr/bin/cpp
define                  =  -DFLEX_SPC
constraints             =  none
integrator              =  steep
dt                      =  0.002    ; ps !
nsteps                  =  2000
nstlist                 =  10
ns_type                 =  grid
rlist                   =  1.0
coulombtype             =  PME
rcoulomb                =  1.0
rvdw                    =  1.4
fourierspacing          =  0.12
fourier_nx              =  0
fourier_ny              =  0
fourier_nz              =  0
pme_order               =  6
ewald_rtol              =  1e-5
optimize_fft            =  yes
;
;       Energy minimizing stuff
;
emtol          =  1000.0
emstep           =  0.01
```

pr.mdp

```
title                  =  prot+lig MD
cpp                    =  /usr/bin/cpp
constraints            =  all-bonds
integrator             =  md
dt                     =  0.002          ; ps !
nsteps                 =  250000         ; total 500 ps.
nstcomm                =  1
nstxout                =  500
nstvout                =  0
nstfout                =  0
nstlist                =  10
ns_type                =  grid
rlist                  =  1.0
coulombtype            =  PME
rcoulomb               =  1.0
vdwtype                =  cut-off
rvdw                   =  1.4
fourierspacing         =  0.12
fourier_nx             =  0
fourier_ny             =  0
fourier_nz             =  0
pme_order              =  6
ewald_rtol             =  1e-5
optimize_fft           =  yes
; Berendsen temperature coupling is on
Tcoupl                 =  berendsen
tau_t                  =  0.1    0.1
tc_grps                =  protein        non-protein
ref_t                  =  300   300
; Use Energy group monitoring
energygrps             =  protein sol RRC
; Pressure coupling is on
Pcoupl                 =  Parrinello-Rahman
pcoupltype             =  isotropic
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               =  173529
```

md.mdp

```
title                   =  prot+lig
warnings                =  10
cpp                     =  /usr/bin/cpp
define                  =  -DPOSRES
constraints             =  all-bonds
integrator              =  md
dt                      =  0.002          ; ps !
nsteps                  =  1000000        ; total 2000 ps.
nstcomm                 =  1
nstxout                 =  250
nstvout                 =  1000
nstfout                 =  0
nstlog                  =  10
nstenergy               =  10
nstlist                 =  10
ns_type                 =  grid
rlist                   =  1.0
coulombtype             =  PME
rcoulomb                =  1.0
vdwtype                 =  cut-off
rvdw                    =  1.4
fourierspacing          =  0.12
fourier_nx              =  0
fourier_ny              =  0
fourier_nz              =  0
pme_order               =  6
ewald_rtol              =  1e-5
optimize_fft            =  yes
; Berendsen temperature coupling is on
Tcoupl                  =  berendsen
tau_t                   =  0.1    0.1
tc_grps                 =  protein        non-protein
ref_t                   =  300   300
; Pressure coupling is on
Pcoupl                  =  Parrinello-Rahman
pcoupltype              =  isotropic
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                =  173529
```

Commands needed to run GROMACS 4.0.5

To run GROMACS for the complex CDK2-roscovitine you need to run a dozen of commands. Figure 2 illustrates the main steps to run a molecular dynamics simulation for a binary complex.

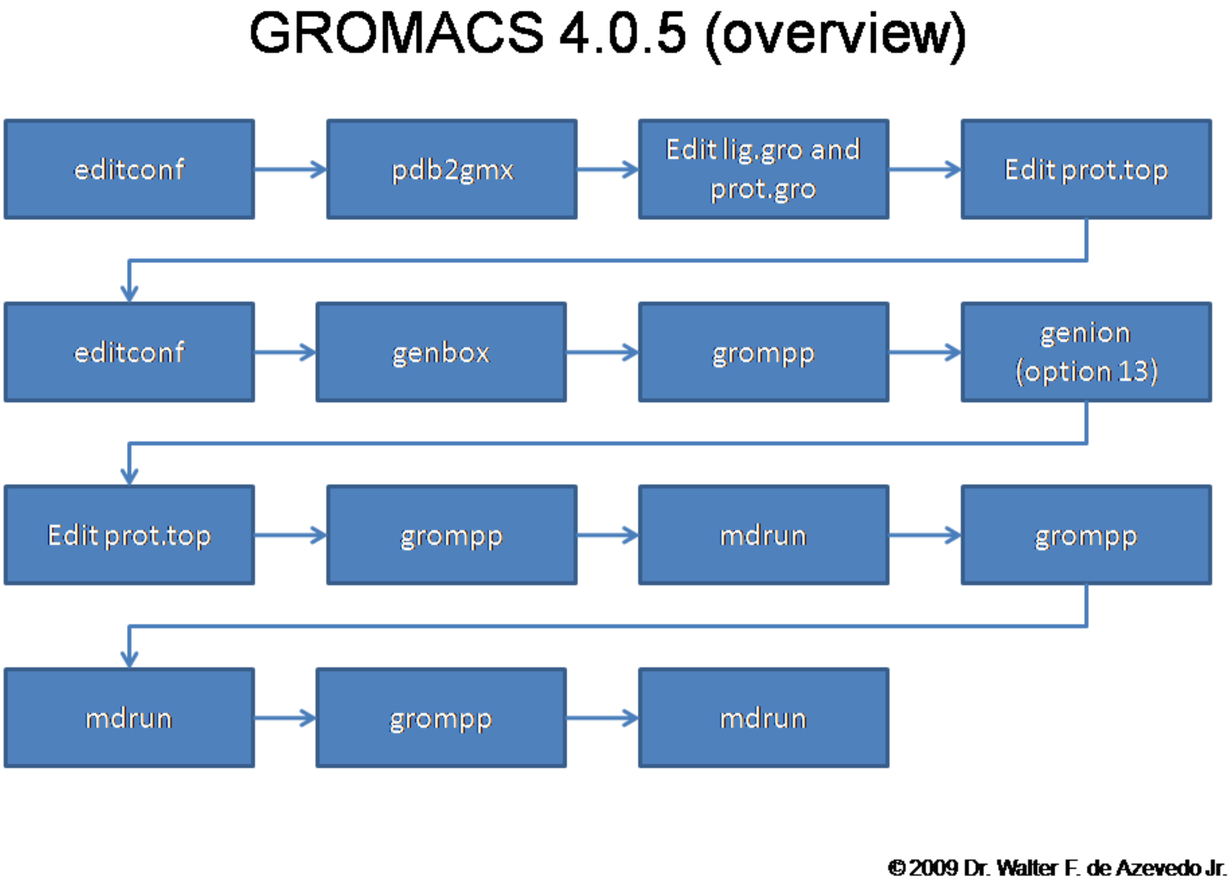

Figure 2. This scheme illustrates the main steps to run a molecular dynamics simulation.

In this tutorial all commands are shown in *italics* in the following lines. Lines started with ">" are command lines. A small comment here, you may generate lig.gro directly the PRODRG, we keep this step for didactic reasons.

```
>editconf -f lig.pdb -o lig.gro
>pdb2gmx -ignh -ff gmx -f prot.pdb -o prot.gro -p prot.top -water spce
```

Edit file *lig.gro* and copy atomic coordinates and paste in the file *prot.gro*. Add number of atoms to the second line of file *prot.gro*.

Edit the file *prot.top* and add the following line “#include lig.itp”, after the “*forcefield*” include information about the ligand in the last line. The information to be included in the last line is the following:

```
 RRC     1
```

And then type the following commands:

```
>editconf -bt cubic -f prot.gro -o prot.gro -d 1.0
>genbox -cp prot.gro -cs spc216.gro -o prot_b4ion.gro -p prot.top
>grompp -f em.mdp -c prot_b4ion.gro -p prot.top -o prot_b4ion.tpr
```

Here we will get the message “Note: system has non-zero charge”, and also indication of charge. This information will be employed in the following step “genion”.

```
>genion -s prot_b4ion.tpr -o prot_b4em.gro -nname Cl -nn 4 -g prot_ion.log
```

Pick option 13 (It is your lucky number)

Edit the file *prot.top* and subtract the number of charges from the number of solvent molecules

The line with information about the solvent was:

```
SOL             18725
```

And it will be

```
SOL             18721
```

Now we add a line with ion, for instance Cl, as follows:

```
Cl              4
```

We run energy minimization step, this will take several minutes (21 minutes in an Intel® Core™ 2 Duo CPU T 8100 @ 2.1 GHz), we type the following commands:

```
>grompp -f em.mdp -c prot_b4em.gro -p prot.top -o em.tpr
>mdrun -v -s em.tpr -e em -o em -c after_em -g emlog >& em.job &
```

Next step “Position Restraint” (PR)

```
>grompp -f pr.mdp -c after_em.gro -p prot.top -o pr.tpr
>mdrun -v -s pr -e pr -o pr -c pr -g prlog >& pr.job &
```

This will take about 90 minutes in an Intel® Core™ 2 Duo CPU T 8100 @ 2.1 GHz

Next step is a MD for 1000 ps (48 hours in the same CPU)

```
>grompp -f md.mdp -c pr.gro -p prot.top -o md.tpr
```

```
>mdrun -v -s md.tpr -e md -o md -c md -g mdlog>& md.job &
```

Use “g_energy” to analyze the potential energy output (the md.edr file), as follows

```
>g_energy –f md.edr –o pe.xvg
```

The following information will be prompted:

```
                        :-)  G  R  O  M  A  C  S  (-:
               GRoups of Organic Molecules in ACtion for Science
                           :-)  VERSION 3.3.1  (-:
     Written by David van der Spoel, Erik Lindahl, Berk Hess, and others.
      Copyright (c) 1991-2000, University of Groningen, The Netherlands.
            Copyright (c) 2001-2006, The GROMACS development team,
           check out http://www.gromacs.org for more information.
        This program is free software; you can redistribute it and/or
         modify it under the terms of the GNU General Public License
        as published by the Free Software Foundation; either version 2
            of the License, or (at your option) any later version.
                             :-)  g_energy  (-:

Option     Filename  Type         Description
------------------------------------------------------------
  -f         md.edr  Input        Generic energy: edr ene
 -f2       ener.edr  Input, Opt.  Generic energy: edr ene
  -s      topol.tpr  Input, Opt.  Generic run input: tpr tpb tpa xml
  -o         pe.xvg  Output       xvgr/xmgr file
-viol  violaver.xvg  Output, Opt. xvgr/xmgr file
-pairs    pairs.xvg  Output, Opt. xvgr/xmgr file
-ora    orienta.xvg  Output, Opt. xvgr/xmgr file
-ort    orientt.xvg  Output, Opt. xvgr/xmgr file
-oda    orideva.xvg  Output, Opt. xvgr/xmgr file
-odr    oridevr.xvg  Output, Opt. xvgr/xmgr file
-odt    oridevt.xvg  Output, Opt. xvgr/xmgr file
-oten    oriten.xvg  Output, Opt. xvgr/xmgr file
-corr   enecorr.xvg  Output, Opt. xvgr/xmgr file
-vis      visco.xvg  Output, Opt. xvgr/xmgr file
-ravg  runavgdf.xvg  Output, Opt. xvgr/xmgr file

      Option   Type  Value  Description
------------------------------------------------------
      -[no]h   bool     no  Print help info and quit
       -nice    int     19  Set the nicelevel
          -b   time      0  First frame (ps) to read from trajectory
          -e   time      0  Last frame (ps) to read from trajectory
      -[no]w   bool     no  View output xvg, xpm, eps and pdb files
   -[no]xvgr   bool    yes  Add specific codes (legends etc.) in the output
                            xvg files for the xmgrace program
   -[no]fee   bool     no  Do a free energy estimate
    -fetemp   real    300  Reference temperature for free energy calculation
      -zero   real      0  Subtract a zero-point energy
   -[no]sum   bool     no  Sum the energy terms selected rather than display
                           them all
    -[no]dp   bool     no  Print energies in high precision
 -[no]mutot   bool     no  Compute the total dipole moment from the
                           components
      -skip    int      0  Skip number of frames between data points
  -[no]aver   bool     no  Print also the X1,t and sigma1,t, only if only 1
                           energy is requested
      -nmol    int      1  Number of molecules in your sample: the energies
                           are divided by this number
       -ndf    int      3  Number of degrees of freedom per molecule.
                           Necessary for calculating the heat capacity
  -[no]fluc   bool     no  Calculate autocorrelation of energy fluctuations
                           rather than energy itself
-[no]orinst   bool     no  Analyse instantaneous orientation data
  -[no]ovec   bool     no  Also plot the eigenvectors with -oten
    -acflen    int     -1  Length of the ACF, default is half the number of
                           frames
-[no]normalize bool    yes  Normalize ACF
         -P   enum      0  Order of Legendre polynomial for ACF (0 indicates
                           none): 0, 1, 2 or 3
     -fitfn   enum   none  Fit function: none, exp, aexp, exp_exp, vac,
                           exp5, exp7 or exp9
    -ncskip    int      0  Skip N points in the output file of correlation
                           functions
  -beginfit   real      0  Time where to begin the exponential fit of the
                           correlation function
    -endfit   real     -1  Time where to end the exponential fit of the
                           correlation function, -1 is till the end

Opened md.edr as single precision energy file

 Select the terms you want from the following list
--------------------------------------------------------
Angle            Proper-Dih.      Improper-Dih.
LJ-14            Coulomb-14       LJ-(SR)
LJ-(LR)          Coulomb-(SR)     Coul.-recip.
Potential        Kinetic-En.      Total-Energy
Temperature      Pressure-(bar)   Box-X
Box-Y            Box-Z            Volume
Density-(SI)     pV               Vir-XX
Vir-XY           Vir-XZ           Vir-YX
Vir-YY           Vir-YZ           Vir-ZX
Vir-ZY           Vir-ZZ           Pres-XX-(bar)
Pres-XY-(bar)    Pres-XZ-(bar)    Pres-YX-(bar)
```

```
Pres-YY-(bar)          Pres-YZ-(bar)          Pres-ZX-(bar)
Pres-ZY-(bar)          Pres-ZZ-(bar)          #Surf*SurfTen
Box-Vel-XX             Box-Vel-YY             Box-Vel-ZZ
Mu-X                   Mu-Y                   Mu-Z
Coul-SR:Protein-Protein LJ-SR:Protein-Protein   LJ-LR:Protein-Protein
Coul-14:Protein-Protein LJ-14:Protein-Protein   Coul-SR:Protein-SOL
LJ-SR:Protein-SOL      LJ-LR:Protein-SOL      Coul-14:Protein-SOL
LJ-14:Protein-SOL      Coul-SR:Protein-RRC     LJ-SR:Protein-RRC
LJ-LR:Protein-RRC      Coul-14:Protein-RRC     LJ-14:Protein-RRC
Coul-SR:Protein-rest    LJ-SR:Protein-rest      LJ-LR:Protein-rest
Coul-14:Protein-rest    LJ-14:Protein-rest      Coul-SR:SOL-SOL
LJ-SR:SOL-SOL          LJ-LR:SOL-SOL          Coul-14:SOL-SOL
LJ-14:SOL-SOL          Coul-SR:SOL-RRC        LJ-SR:SOL-RRC
LJ-LR:SOL-RRC          Coul-14:SOL-RRC        LJ-14:SOL-RRC
Coul-SR:SOL-rest       LJ-SR:SOL-rest         LJ-LR:SOL-rest
Coul-14:SOL-rest       LJ-14:SOL-rest         Coul-SR:RRC-RRC
LJ-SR:RRC-RRC          LJ-LR:RRC-RRC          Coul-14:RRC-RRC
LJ-14:RRC-RRC          Coul-SR:RRC-rest       LJ-SR:RRC-rest
LJ-LR:RRC-rest         Coul-14:RRC-rest       LJ-14:RRC-rest
Coul-SR:rest-rest      LJ-SR:rest-rest        LJ-LR:rest-rest
Coul-14:rest-rest      LJ-14:rest-rest        T-Protein
T-Non-Protein          Lamb-Protein           Lamb-Non-Protein

Potential
Last frame read 10000 time  2000.000
Statistics over 1000001 steps [ 0.0000 thru 2000.0000 ps ], 1 data sets
Energy                 Average      RMSD    Fluct.     Drift  Tot-Drift
-------------------------------------------------------------------------------
Potential              -912739     756.723    731.316   -0.336777   -673.555
```

We entered “Potential” followed by <enter> to end selection. All energies are expressed in kJ/mol. Use the g_energy command to obtain other components (e.g. kinetic energy, etc.).
To visualize we use the xmgrace as follows:

>xmgrace pe.xvg

Figure 3 shows the energy plot.

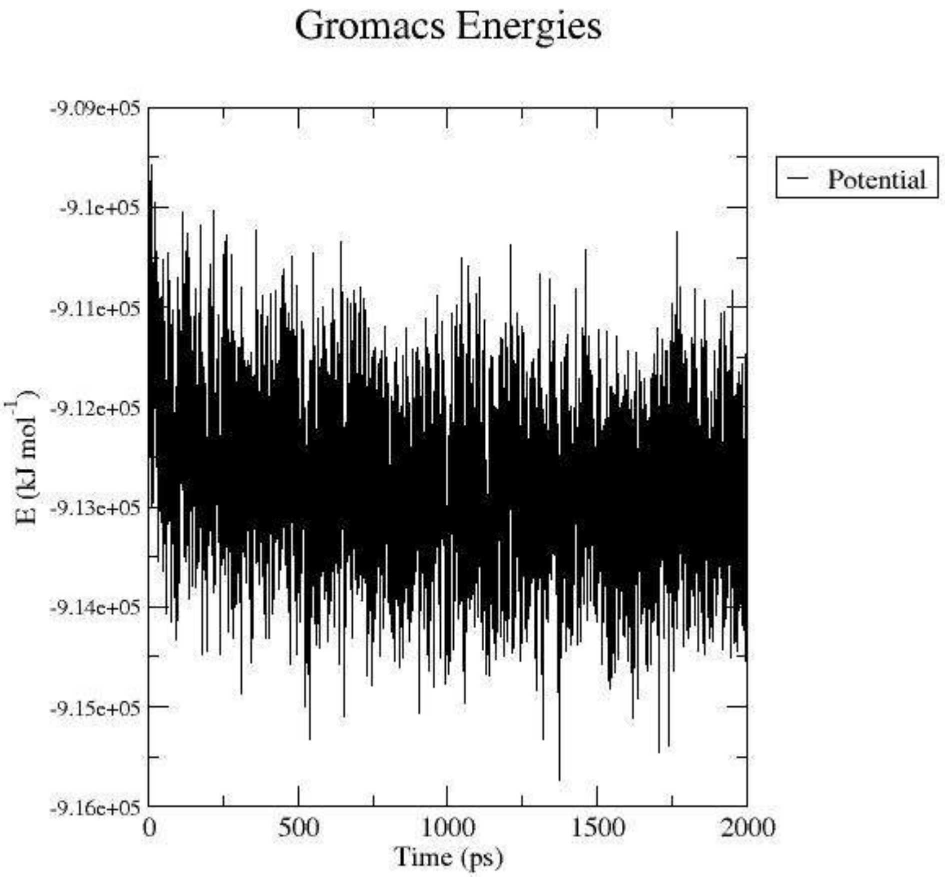

Figure 3. Potential energy of the complex during the 500 ps trajectory.

To generate information about intermolecular hydrogen bonds we employ the following command:

```
>g_hbond –f md.trr –s md.tpr –num prot_lig_h.xvg
```

We select group “1” for protein and <enter> and then group “12” for RRC and <enter>. This will generate the graph in the format .xvg
To visualize we use the xmgrace as follows:

```
>xmgrace prot_lig_h.xvg
```

Figure 4 shows the intermolecular hydrogen bonds.

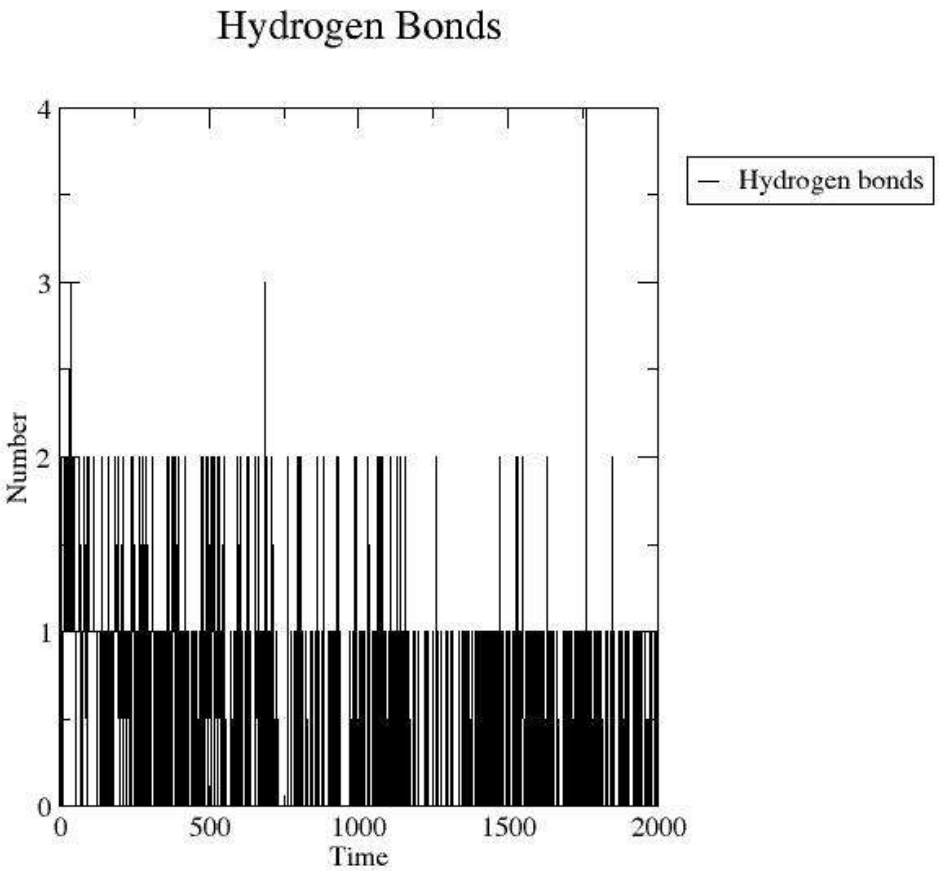

Figure 4. Intermolecular hydrogen bonds during 500 ps trajectory.

We may use g_rms to analyze RMSD fluctuation during molecular dynamics, as follows:

```
>g_rms –s md.tpr –f md.trr –o backbone_rmsd.xvg
```

Select “1” then hit <enter> and “4” for backbone atoms and then hit <enter> P
We can see clearly (Figure 5) that the structure reached the equilibrium.

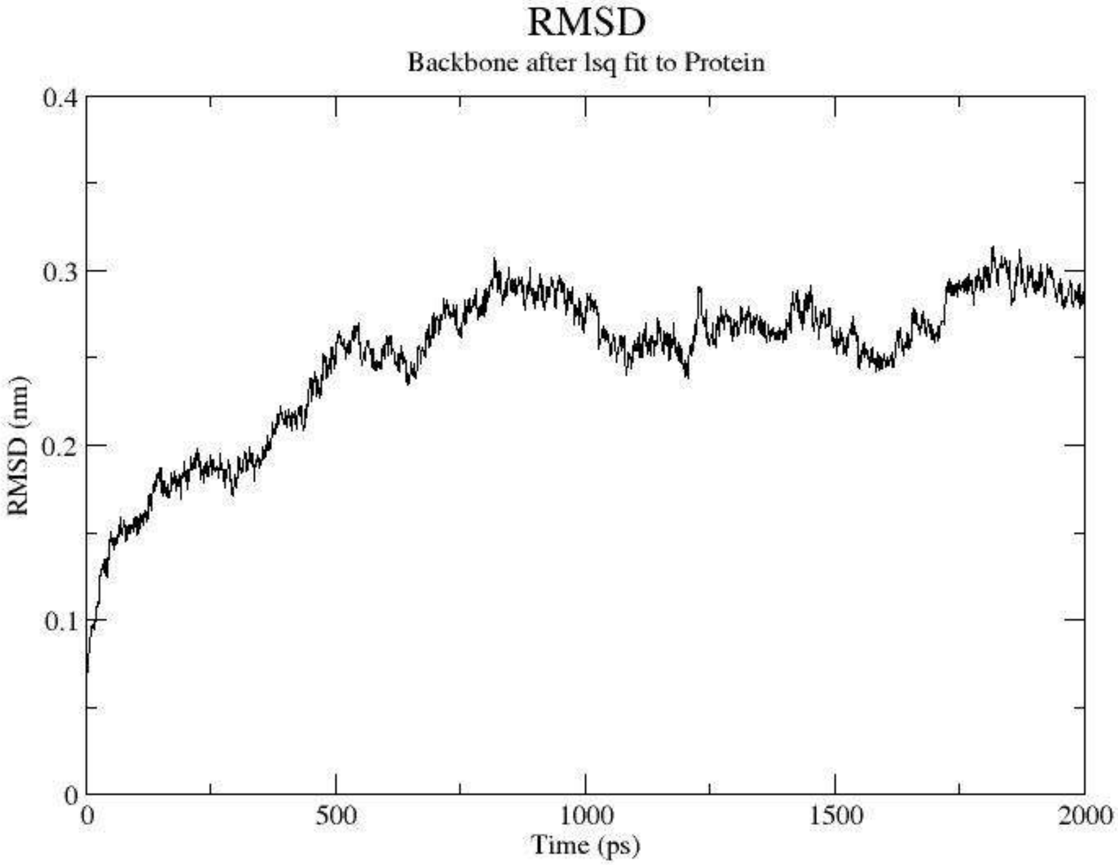

Figure 5. RMSD during 500 ps trajectory.

References

De Azevedo Jr., W. F. *Curr Drug Targets*, 2008, **9,** 1030-1030.

De Azevedo WF Jr, Dias R. *Bioorg Med Chem.* 2008, **16,** 9378-9383, 2008

De Azevedo, W. F.; Mueller-Dieckmann, H. J.; Schulze-Gahmen, U.; Worland, P. J.; Sausville, E.; Kim, S-H. In *Proc. Natl. Acad. Sci.* 1996, **93**, 2735.

De Azevedo, W. F.; Leclerc, S. J.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S. H. *European Journal of Biochemistry***. 1997, 243**, 518.

Sali A, Blundell TL. J Mol Biol. 1993,234(3):779-815.

van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) J. Comp. Chem. **26** p. 1701-1718 [http://www.GROMACS.org]