Tutorial for Trypsin-Benzamidine complex molecular dynamics study.

Gromacs version 4.5.5

John E. Kerrigan, Ph.D.
Associate Director, Bioinformatics
The Cancer Institute of New Jersey
The Robert Wood Johnson Medical School
120 Albany Street
New Brunswick, NJ 08903

(732) 235-4473 <u>kerrigje@umdnj.edu</u> In this tutorial we will study the dynamics of the complex of benzamidine with trypsin, a serine protease related to the cascade of proteases responsible for regulation of blood coagulation. We will use the recently published x-ray structure (3ATL.pdb).[1] This tutorial requires Gromacs (http://www.gromacs.org) suite of molecular simulation programs,[2] the Amber Antechamber program (AmberTools from http://ambermd.org),[3, 4] UCSF Chimera (http://plato.cgl.ucsf.edu/chimera/), VMD (http://www.ks.uiuc.edu/Research/vmd/), and **ACPYPE** (http://code.google.com/p/acpype/).[5]

Benzamidinium ion

Starting from 3ATL.pdb, we need to extract out the protein from the ligand. Use the following commands in the unix shell.

grep 'ATOM ' 3ATL.pdb > protein.pdb
grep 'BEN A' 3ATL.pdb > ligand.pdb

Isolate the Ca²⁺ ion

grep 'HETATM 1631 CA' 3ATL.pdb > ca-ion.pdb

Combine protein and CA ion using the unix "cat" command.

cat protein.pdb ca-ion.pdb > protein2.pdb

Process protein2.pdb with pdb2gmx as follows ...

pdb2gmx -ff amber99sb -f protein2.pdb -o trp.pdb -p trp.top -water tip3p -ignh

Note we are using the Amber99SB force field[6] and the TIP3P water model.[7] We will process the ligand using UCSF Chimera to insure we build an amidinium ion as benzamidine is charged at physiological pH.

Open ligand.pdb in *UCSF Chimera*.
Go to Tools > Structure Editing > AddH
Follow with Tools > Structure Editing > Add Charge (Select AM1-BCC in dialogue box).

When finished the molecule should and overall formal charge of +1.00. Note we are using the general amber forcefield (gaff)[4] and AM1-BCC charges[8] for the small molecule.

Save the file in mol2 format with filename ben.mol2. Next we will run acpype using the partial atomic charges from the mol2 file.

```
Run ACPYPE ...
```

```
acpype -i ben.mol2 -c user
```

All output is dumped into a directory called "ben.acpype". Always check to make sure the atom types assigned make sense according to the *GAFF* atom type descriptions posted at http://ambermd.org/antechamber/gaff.html#atomtype and copy of same in the appendix of this document.

Combine protein with ligand coordinates ... **grep -h** ATOM trp.pdb ben.acpype/ben_NEW.pdb > complex.pdb

```
cp ben.acpype/ben_GMX.itp ben.itp
```

Using any text editor (e.g. vi) add #include "ben.itp" right after the forcefield in trp.top and add "BEN 1" to the [molecules] section at the end of trp.top.

```
; Include forcefield parameters
#include "amber99sb.ff/forcefield.itp"
#include "ben.itp"
...
[ molecules ]
```

```
; Compound #mols
Protein_chain_A 1
BEN 1
```

editconf -bt octahedron **-f** complex.pdb **-o** trp-b4solv.pdb **-d** 1.0 **genbox -cp** trp-b4solv.pdb **-cs** spc216.gro **-o** trp-b4ion.pdb **-p** trp.top

em.mdp

```
define
                        = -DFLEXIBLE
integrator
                        = cg ; steep
                        = 200
nsteps
constraints
                        = none
                        = 1000.0
emtol
                        = 10; do a steep every 10 steps of cg
nstcgsteep
                        = 0.01; used with steep
emstep
nstcomm
                        = PME
coulombtype
                        = grid
ns type
rlist
                        = 1.0
                        = 1.0
rcoulomb
```

We will add NaCl to concentration of 0.15 M in the periodic box and make an overall neutral system in terms of formal charge.

```
grompp -f em.mdp -c trp-b4ion.pdb -p trp.top -o ion.tpr
genion -s ion.tpr -o trp-b4em.pdb -neutral -conc 0.15 -p trp.top -g ion.log
Select Group 19 (SOL)
```

grompp -**f** em.mdp -**c** trp-b4em.pdb -**p** trp.top -**o** em.tpr

mdrun -v -deffnm em

Our run finished with the following result ...

```
Polak-Ribiere Conjugate Gradients converged to Fmax < 1000 in 63 steps Potential Energy = -4.4418656e+05 Maximum force = 9.5519733e+02 on atom 2472 Norm of force = 4.4277103e+01
```

Position restrained run – Here we will perform 100 ps of NVT (constant volume and temperature) ensemble dynamics to relax the water in the system while applying restraints to the protein.

pr.mdp

```
= -DPOSRES
define
integrator
                        = md
                        = 50000
nsteps
                        = 0.002
dt
                        = all-bonds
constraints
nstcomm
                        = 10
ns type
                        = grid
rlist
                        = 1.2
rcoulomb
                        = 1.0
rcoulomb-switch
                        = 0.9
                        = 1.0
rvdw
                        = shift
vdwtype
rvdw-switch
                        = 0.9
                        = PME-Switch
coulombtype
Tcoupl
                        = v-rescale
tau t
                        = 0.1 0.1
tc-grps
                       = protein non-protein
ref t
                        = 300 300
```

```
Pcoupl
                     = no
Pcoupltype
                     = isotropic
                     = 0.5
                     = 4.5e-5
compressibility
                      = 1.0
ref p
gen vel
                     = yes
gen temp
                     = 300
                     = -1
gen seed
                     = 250 ; write coords every # step
nstxout
                     = 100
nstenergy
energygrps
                     = Protein BEN SOL ; energy group monitoring
lincs-iter
DispCorr
                      = EnerPres
optimize fft
                      = yes
```

grompp –**f** pr.mdp –**c** em.gro –**p** trp.top –**o** pr.tpr

nohup mdrun -deffnm pr &

Production run – Run at constant pressure and temperature (NPT).

md.mdp

```
integrator
                      = md
                     = 250000
nsteps
                     = 0.002
constraints
                      = all-bonds
nstcomm
                      = 10
                     = grid
ns type
rlist
                     = 1.2
rcoulomb
                     = 1.0
rcoulomb-switch
                    = 0.9
                     = 1.0
rvdw
                     = shift
vdwtype
                     = 0.9
rvdw-switch
                     = PME-Switch
coulombtype
Tcoupl
                     = v-rescale
tau t
                     = 0.1 0.1
tc-grps
                     = protein non-protein
                     = 300 300
ref t
                      = parrinello-rahman
Pcoupl
Pcoupltype
                     = isotropic
                      = 0.5
tau p
compressibility
                     = 4.5e-5
ref p
                      = 1.0
                     = yes
gen vel
gen_temp
                      = 300
                     = -1
gen seed
                    = 500 ; write coords every # step
nstxout
                     = 100
nstenergy
                    = Protein BEN SOL ; energy group monitoring
energygrps
                     = 2
lincs-iter
                      = EnerPres
DispCorr
optimize fft
                      = yes
```

grompp -**f** md.mdp -**c** pr.gro -**p** trp.top -**o** md.tpr

nohup mdrun -deffnm md &

Our 500 ps run completed in 1 hour and 24 minutes using all 8 cores of a dual quad core Intel Xeon 2.8 GHz cpu workstation.

Analysis

Inspect the trajectory visually. First convert the trajectory to compressed format and remove pbc artifacts.

trjconv -**s** md.tpr -**f** md.trr -**o** md.xtc -**pbc** nojump

Select 0 (System). Use a molecular viewing program like VMD (http://www.ks.uiuc.edu/Research/vmd/) to view the trajectory (md.xtc). Using VMD load pr.gro and load md.xtc into pr.gro. Display residue ID 58, 189 and 195 (resid 58 189 195) in addition to benzamidine (resname BEN).

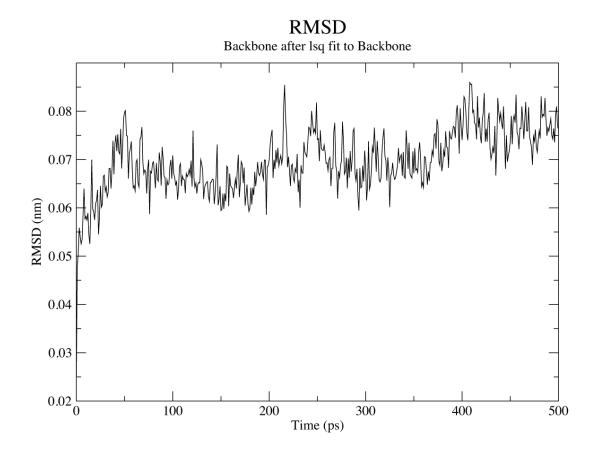
Does benzamidine remain in the active site pocket during the course of the 500 ps simulation or does it begin to slip out?

What residue(s) does benzamidine interact with during the course of the 500 ps simulation? (*Hint*: In VMD use "protein within 8 of resname BEN" to view protein atoms within 8 Å of benzamidine.)

Check protein backbone RMSD reference to the crystal structure ...

g_rms -**s** em.tpr -**f** md.trr -**o** md-bkbone.xvg

Select 4 (Backbone) both times.

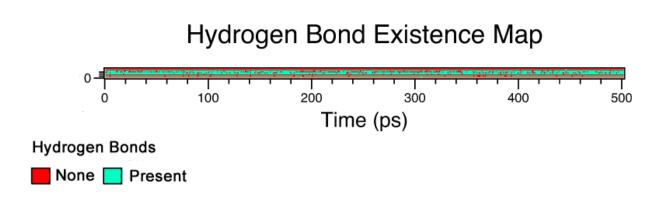


Evaluate the number of hydrogen bonds between benzamidine and the protein over the course of the trajectory.

g_hbond -**s** md.tpr -**f** md.xtc -**num** -**hbm**

Select BEN and Protein

xpm2ps -**f** hbmap.xpm -**o** hbplot.eps -**rainbow** red



We had about 4 hydrogen bonds in existence on average between BEN and the Protein during the course of the simulation. See the Gromacs manual for more **g_hbond** tool options.

Energy Analysis. Obtain the short range Lennard-Jones and short range Coulomb energies of the BEN-Protein interaction using the Gromacs **g_energy** energy analysis tool.

g_energy -s md.tpr -f md.edr -o ben-trp-LISR-CoulSR.xvg

Statistics over 250001 steps [0.0000 through 500.0000 ps], 2 data sets All statistics are over 25001 points

Energy	Average	Err.Est.	RMSD	Tot-Drift	
Coul-SR:Protein-BEN	-190.533	0.88	15.0135	-3.43307	(kJ/mol)
LJ-SR:Protein-BEN	-68.6689	1.3	11.0709	-6.82649	(kJ/mol)

$$E_{\text{int}} = \langle E_{\text{LJ}} \rangle + \langle E_{\text{Coul}} \rangle = -68.7 + -190.5 = -259.2 \text{ kJ/mol} = -62.0 \text{ kcal/mol}$$

The $E_{\rm int}$ is a crude interaction energy based on short-range energy components. This is not a binding energy. This number is a crude qualitative estimate of the stability of the Trypsin-benzamidine complex.

You may also obtain other information from the energy file (.edr) like Pressure, Temperature, Kinetic Energy, etc.

References:

- 1. Yamane, J., et al., *In-crystal affinity ranking of fragment hit compounds reveals a relationship with their inhibitory activities.* J. Appl. Crystallogr., 2011. **44**: p. 798.
- 2. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* J. Chem. Theor. Comp., 2008. **4**(3): p. 435-447.
- 3. Wang, J., et al., *Antechamber, an accessory software package for molecular mechanics calculations.* J. Mol. Graphics, 2006. **25**: p. 247-260.
- 4. Wang, J., et al., *Development and testing of a general amber force field.* J. Comput. Chem., 2004. **25**(9): p. 1157-1174.
- 5. Sousa Da Silva, A.W., W.F. Vranken, and E.D. Laue, *AnteChamber PYthon Parser interfacE.*, 2010.
- 6. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.
- 7. Jorgensen, W., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* J. Chem. Phys., 1983. **79**: p. 926-935.
- 8. Jakalian, A., D.B. Jack, and C.I. Bayly, *Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation.* J. Comput. Chem., 2002. **23**(16): p. 1623-41.

Appendix

The following table was obtained from $\underline{\text{http://ambermd.org/antechamber/gaff.html\#atomtype}}$. Table I lists the basic (a) and special (b) atom types in GAFF.

Table I (a). Basic Atom Types in GAFF

Atom type	Description	Atom type	Description
С	sp2 C in C=O, C=S	0	sp2 O in C=O, COO-
c1	sp1 C	oh	sp3 O in hydroxyl group
c2	sp2 C, aliphatic	os	sp3 O in ether and ester
c3	sp3 C		
ca	sp2 C, aromatic		
n	sp2 N in amide	s2	sp2 S (p=S, C=S etc)
n1	sp1 N	sh	sp3 S in thiol group
n2	sp2 N with 2 subst.	SS	sp3 S in -SR and SS
n3	readl double bond	s4	hypervalent S, 3 subst.
n4	sp3 N with 3 subst.	s6	hypervalent S, 4 subst.
na	sp3 N with 4 subst.	hc	H on aliphatic C
nh	sp2 N with 3 subst	ha	H on aromatic C
no	amine N connected	hn	H on N
	to the aromatic rings	ho	H on O
	N in nitro group	hs	H on S
		hp	H on P
f	any F	p2	sp2 P (C=P etc)
cl	any Cl	p3	sp3 P, 3 subst.
br	any Br	p4	hypervalent P, 3 subst.
i	any I	p5	hypervalent P, 4 subst.

Table I (b). Special Atom Types in GAFF

Atom type	Description	Atom type	Description
h1	H on aliphatic C with 1 EW group;	cc(cd)	inner sp2 C in conj. ring systems
h2	H on aliphatic C with 2 EW group;	ce(cf)	inner sp2 C in conj. chain systems
h3	H on aliphatic C with 3 EW group;	cp(cq)	bridge aromatic C
h4	H on aromatic C with 4 EW group;	cu	sp2 C in three-memberred rings
h5	H on aromatic C with 5 EW group;	cv	sp2 C in four-memberred rings
		cx	sp3 C in three-memberred rings
		cy	sp3 C in four-memberred rings
n	aromatic nitrogen	pb	aromatic phosphorus
nb	inner sp2 N in conj. ring systems	pc(pd)	inner sp2 P in conj. ring systems
nc(nd)	inner sp2 N in conj. chain systems	pe(pf)	inner sp2 P in conj. chain systems
SX	conj. S, 3 subst.	px	conj. P, 3 subst.
sy	conj. S, 4 subst.	py	conj. P, 4 subst.

EW: electron-withdraw group

4. Wang, J., R. Wolf, J. Caldwell, P. Kollman, and D.A. Case, *Development and testing of a general amber force field.* J. Comput. Chem., 2004. **25**(9): p. 1157-1174.