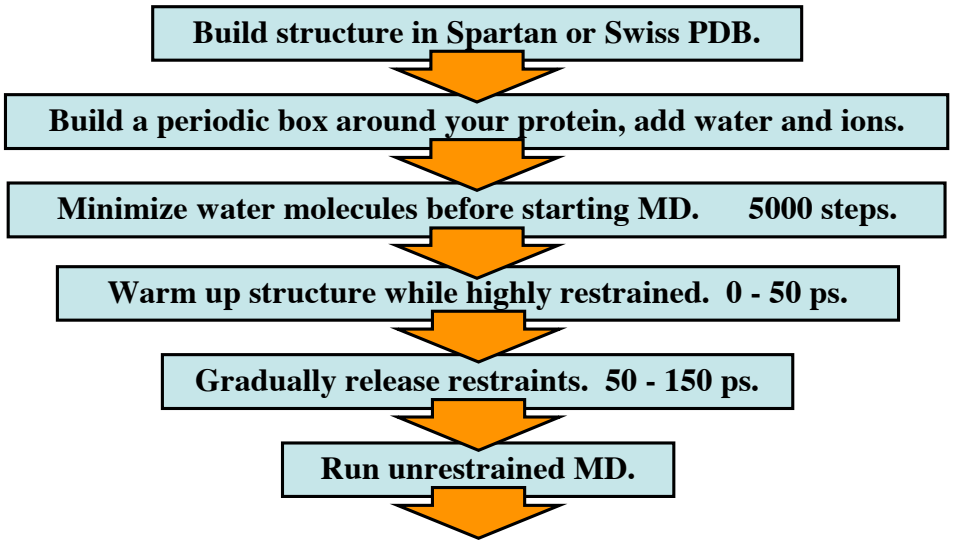
The Magic of Movies
A GROMACS Tutorial

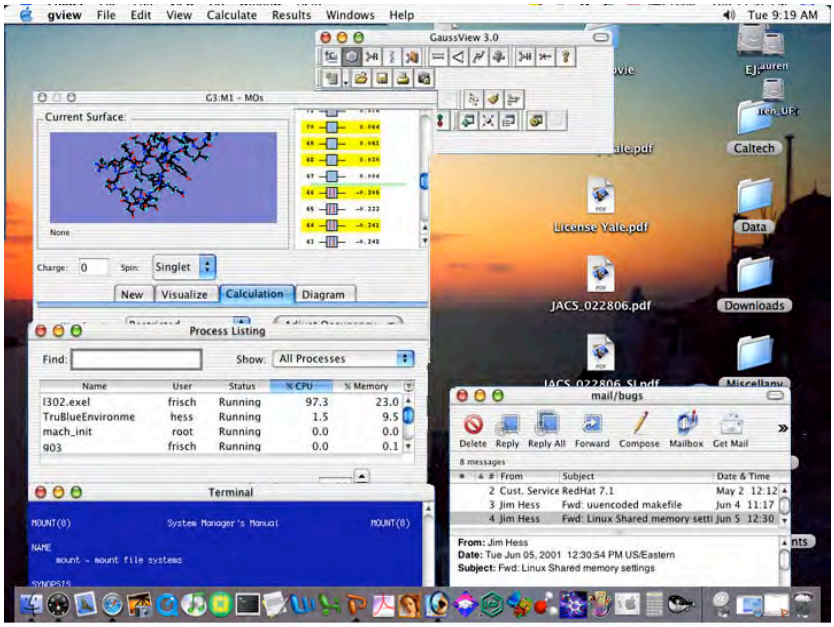
E. James Petersson
Schepartz Lab Meeting
03/03/06

GamePlan

aPP Y7A F20Y mutant (aPP_A7Y20) will be used as an example.

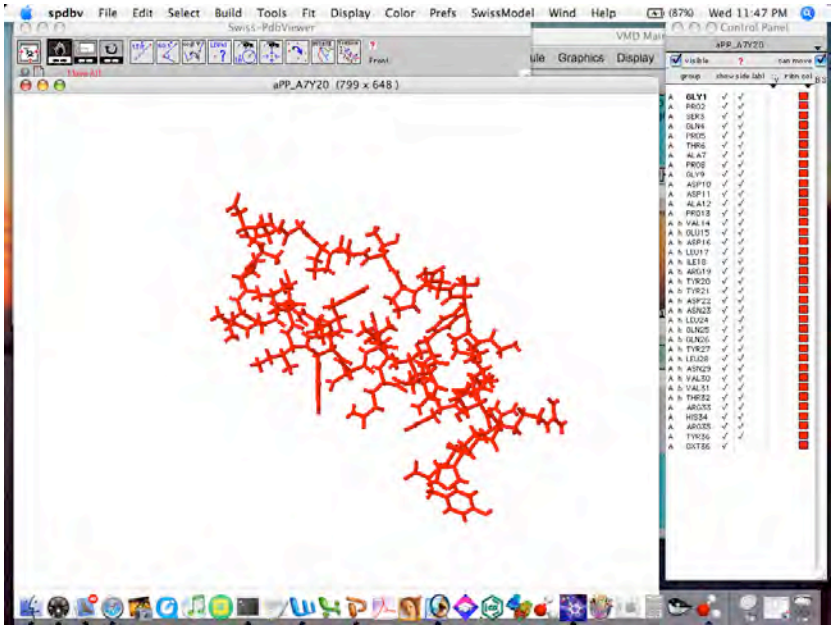


in silico Mutagenesis



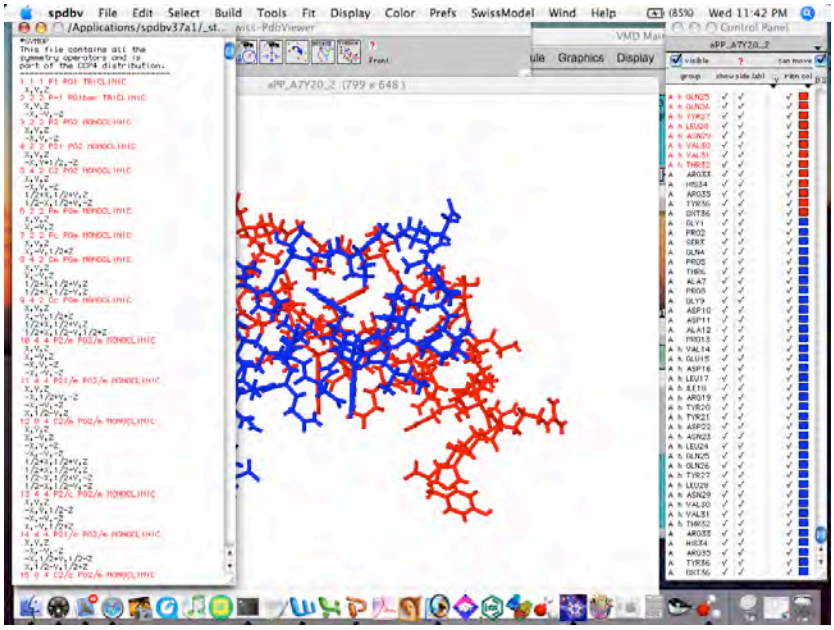
Gaussview, Spartan, Swiss PDB or PyMOL can be used to create a monomer structure.

Dimer Formation: Just the Two of Us



Dimer built from C2 symmetry operation in Swiss PDB Viewer.

Dimer Formation: Just the Two of Us



Dimer built from C2 symmetry operation in Swiss PDB Viewer.

Get into Gromacs

Imported this file (aPP_A7Y20_2.pdb) into Gromacs v. 3.1.4. Generated Gromacs and topology files:

```
%pdb2gmx -f aPP_A7Y20_2.pdb -o aPP_A7Y20_2.gro
-p aPP_A7Y20_2.top -i aPP_A7Y20_pr.itp
```

Select OPLS force field option.

File types:
aPP_A7Y20_2.gro - Gromacs coordinate file
aPP_A7Y20_2.top - topology file, specifies all simulation components
aPP_A7Y20_2_A(B).itp - parameter file for protein
aPP_A7Y20_A(B)_pr.itp - position restraint file

Your Friend “command -h”

```
%pdb2gmx -h
```

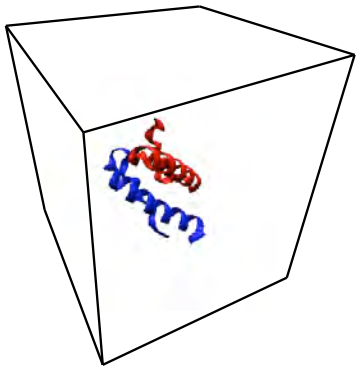
```
                :-)  G  R  O  M  A  C  S  (-:
          Good ROcking Metal Altar for Chronical Sinners

                :-)  VERSION 3.3  (-:


    Written by David van der Spoel, Erik Lindahl, Berk Hess, and others.
     Copyright (c) 1991-2000, University of Groningen, The Netherlands.
           Copyright (c) 2001-2004, 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.

                :-)  pdb2gmx  (-:

SCRIPTION:

is program reads a pdb file, reads some database files, adds hydrogens to the molecules and generates
ordinates in Gromacs (Gromos) format and a topology in Gromacs format. These files can subsequently be
ocessed to generate a run input file.

e force fields in the distribution are currently:

lsaa OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
3b1  GROMOS96 43b1 Vacuum Forcefield
3a1  GROMOS96 43a1 Forcefield
3a2  GROMOS96 43a2 Forcefield (improved alkane dihedrals)
5a3  GROMOS96 45a3 Forcefield
3a5  GROMOS96 53a5 Forcefield
3a6  GROMOS96 53a6 Forcefield
x    Gromacs Forcefield (a modified GROMOS87, see manual)
ads Encad all-atom force field, using scaled-down vacuum charges
adv Encad all-atom force field, using full solvent charges

e corresponding data files can be found in the library directory with names like ffXXXX.YYY. Check
apter 5 of the manual for more information about file formats. By default the forcefield selection is
teractive, but you can use the -ff option to specify one of the short names above on the command line
stead. In that case pdb2gmx just looks for the corresponding file.

te that a pdb file is nothing more than a file format, and it need not necessarily contain a protein
ructure. Every kind of molecule for which there is support in the database can be converted. If there
 support in the database, you can add it yourself.

e program has limited intelligence, it reads a number of database files, that allow it to make special
nds (Cys-Cys, Heme-His, etc.), if necessary this can be done manually. The program can prompt the user
 select which kind of LYS, ASP, GLU, CYS or HIS residue she wants. For LYS the choice is between LYS
wo protons on NZ) or LYSH (three protons, default), for ASP and
U unprotonated (default) or protonated, for HIS the proton can be either on ND1 (HISA), on NE2 (HISB) o
 both (HISH). By default these selections are done automatically. For His, this is based on an optimal
drogen bonding conformation. Hydrogen bonds are defined based on a simple geometric criterium, specifie
 the maximum hydrogen-donor-acceptor angle and donor-acceptor distance, which are set by -angle and -
st respectively.

tion -merge will ask if you want to merge consecutive chains into one molecule definition, this can be
eful for connecting chains with a disulfide brigde or intermolecular distance restraints.

b2gmx will also check the occupancy field of the pdb file. If any of the occupanccies are not one,
dicating that the atom is not resolved well in the structure, a warning message is issued. When a pdb
le does not originate from an X-Ray structure determination all occupancy fields may be zero. Either
y, it is up to the user to verify the correctness of the input
ta (read the article!).

ring processing the atoms will be reordered according to Gromacs conventions. With -n an index file ca
 generated that contains one group reordered in the same way. This allows you to convert a Gromos
ajectory and coordinate file to Gromos. There is one limitation: reordering is done after the hydrogen
s stripped from the input and before new hydrogens are added.
is means that you should not use -ignh.

s .gro and .g96 file formats do not support chain identifiers. Therefore it is useful to enter a pdb
le name at the -o option when you want to convert a multichain pdb file.

s option -vsite removes hydrogen and fast improper dihedral motions.  Angular and out-of-plane motions
n be removed by changing hydrogens into virtual sites and fixing angles, which fixes their position
lative to neighboring atoms. Additionally, all atoms in the aromatic rings of the standard amino acids
.e. PHE, TRP, TYR and HIS) can be converted into virtual sites, eliminating the fast improper dihedral
uctuations in these rings. Note that in this case all other hydrogen atoms are also converted to virtu
tes. The mass of all atoms that are converted into virtual sites, is added to the heavy atoms.

so slowing down of dihedral motion can be done with -heavyh done by increasing the hydrogen-mass by a
ctor of 4. This is also done for water hydrogens to slow down the rotational motion of water. The
crease in mass of the hydrogens is subtracted from the bonded (heavy) atom so that the total mass of t
stem remains the same.

tion     Filename  Type         Description
------------------------------------------------------------
-f       eiwit.pdb  Input        Generic structure: gro g96 pdb tpr tpb tpa
                                 xml
-o        conf.gro  Output       Generic structure: gro g96 pdb xml
-p       topol.top  Output       Topology file
-i       posre.itp  Output       Include file for topology
-n       clean.ndx  Output, Opt. Index file
-q       clean.pdb  Output, Opt. Generic structure: gro g96 pdb xml

    Option    Type  Value  Description
------------------------------------------------------
    -[no]h    bool    yes  Print help info and quit
    -[no]X    bool     no  Use dialog box GUI to edit command line options
     -nice     int      0  Set the nicelevel
-[no]merge    bool     no  Merge chains into one molecule definition
      -ff string select    Force field, interactive by default. Use -h for
                           information.
    -water    enum    spc  Water model to use: with GROMOS we recommend SPC,
                           with OPLS, TIP4P: spc, spce, tip3p, tip4p, tip5p
                           or f3c
-[no]inter    bool     no  Set the next 6 options to interactive
  -[no]ss     bool     no  Interactive SS bridge selection
 -[no]ter     bool     no  Interactive termini selection, iso charged
 -[no]lys     bool     no  Interactive Lysine selection, iso charged
 -[no]asp     bool     no  Interactive Aspartic Acid selection, iso charged
 -[no]glu     bool     no  Interactive Glutamic Acid selection, iso charged
 -[no]his     bool     no  Interactive Histidine selection, iso checking
                           H-bonds
   -angle     real    135  Minimum hydrogen-donor-acceptor angle for a
                           H-bond (degrees)
    -dist     real    0.3  Maximum donor-acceptor distance for a H-bond (nm)
 -[no]una     bool     no  Select aromatic rings with united CH atoms on
                           Phenylalanine, Tryptophane and Tyrosine
-[no]ignh     bool     no  Ignore hydrogen atoms that are in the pdb file
no]missing    bool     no  Continue when atoms are missing, dangerous
   -[no]v     bool     no  Be slightly more verbose in messages
  -posrefc    real   1000  Force constant for position restraints
   -vsite     enum   none  Convert atoms to virtual sites: none, hydrogens
                           or aromatics
[no]heavyh    bool     no  Make hydrogen atoms heavy
no]deuterate bool     no  Change the mass of hydrogens to 2 amu
```

Thinking Inside the Box

Generated rectangular periodic boundary box with dimensions:

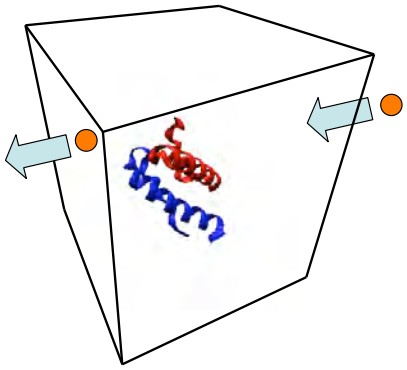
5.220 5.113 6.052

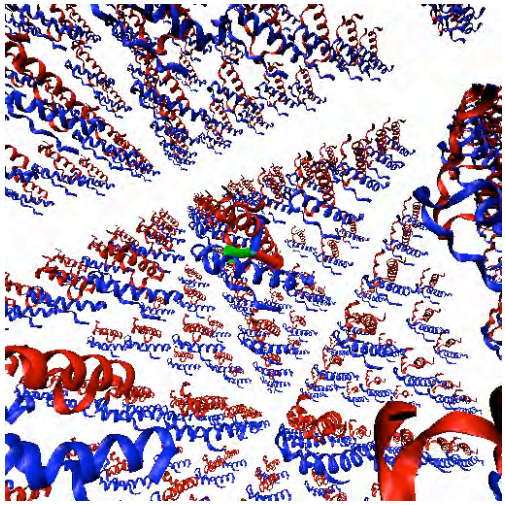
```
%editconf -f aPP_A7Y20_2.gro -o aPP_A7Y20_2box.gro -d 1
```



Thinking Outside the Box

Periodic boundary conditions can be thought of in two ways: like Asteroids or like a crystal.

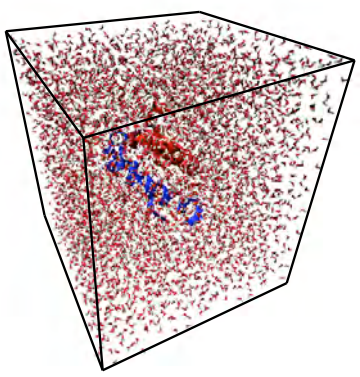




Going Swimming

Added waters to box.

```
%genbox -cp aPP_A7Y20_2box.gro -cs spc216.gro
-o aPP_A7Y20_2h2o.gro -p aPP_A7Y20_2.top
```



Getting Charged Up

Generated mdp file for genion run.

```
aPP_A7Y20_2genion.mdp
```

Generated start (.tpr) file for genion, looked for charge in output, -4.

```
%grompp -f aPP_A7Y20_2genion.mdp -c aPP_A7Y20_2h2o.gro
-p aPP_A7Y20_2.top -o aPP_A7Y20_2genion.tpr
```

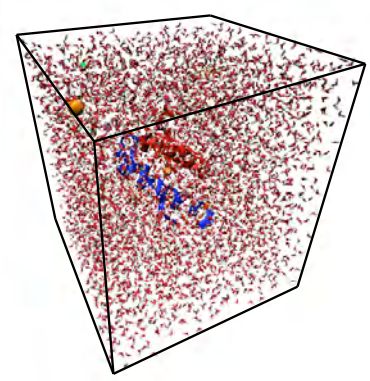
Made index file from aPP_A7Y20_2h2o.gro.

```
%make_ndx -f aPP_A7Y20_2h2o.gro -o aPP_A7Y20_2.ndx
```

Isn't it Ionic?

Generated ions. 7 ions = 50 mM for ~ 5nm box, more Na to compensate for -4 charge. Pick SOL index group to be replaced.

```
%genion -s aPP_A7Y20_2genion.tpr -o aPP_A7Y20_2ion.gro
-n aPP_A7Y20_2.ndx -g genion.log -np 7 -nn 3
```



Bookkeeping by Hand? In 2006?

Edited atom and molecule types as follows:
In .gro change all Na and Cl to NA+ or NA and CL- or Cl like this:

```
4973Na      Na15831   5.076   4.901   5.979
4974Cl      Cl15832   3.930   0.267   1.555
```

becomes

```
4973NA+     NA15831   5.076   4.901   5.979
4974CL-     CL15832   3.930   0.267   1.555
```

In .top replace 10 SOL molecules with 7 NA+ and 3 CL-.

```
[ molecules ]
; Compound        #mols
Protein_A             1
Protein_B             1
SOL                4894
NA+                   7
CL-                   3
```

Be Non-confrontational: Reduce Bad Contacts

min1 - Generated mdp file for minimization run 1.
define = -dPOSRES, freeze_grps = backbone.

```
aPP_A7Y20_2_min1.mdp
```

Made a new index file from aPP_A7Y20_2ion.gro.

```
%make_ndx -f aPP_A7Y20_2ion.gro -o aPP_A7Y20_2.ndx
```

Generated position restraints for protein

```
%genpr -f aPP_A7Y20_2ion.gro - n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2_A(B)_pr.itp -fc 1000 1000 1000
```

Pick PROTEIN group.

Be Non-confrontational: Reduce Bad Contacts

min1 - Generated start file for minimization 1.

```
%grompp -f aPP_A7Y20_2_min1.mdp -c aPP_A7Y20_2ion.gro
-n aPP_A7Y20_2.ndx -p aPP_A7Y20_2.top  -o aPP_A7Y20_2_min1.tpr
```

Minimization run 1.

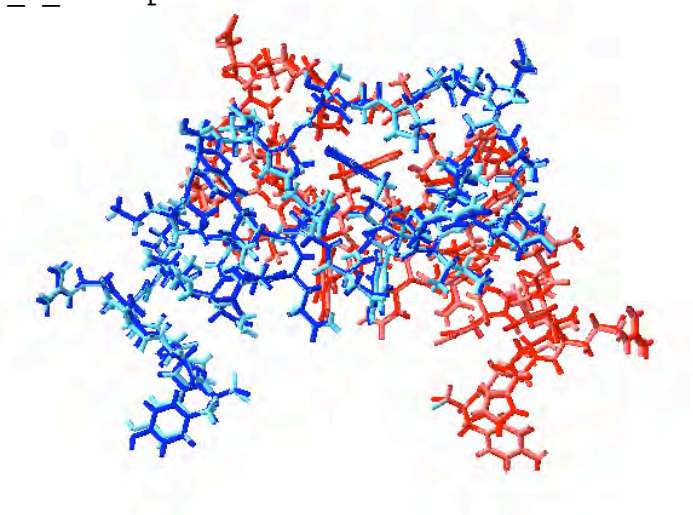
```
%mdrun - aPP_A7Y20_2_min1.tpr -o aPP_A7Y20_2_min1.trr
-x aPP_A7Y20_2_min1.xtc -c aPP_A7Y20_2_min1.gro
-e aPP_A7Y20_2_min1.edr -g aPP_A7Y20_2_min1.log &
```

Does not converge after 5000 steps. That's OK. With all these restraints, it's unlikely to converge.

Initial and Minimized Structures

min1 - Make sure the restraints worked and nothing happened to the protein. Generate pdb files, then align them with Swiss PDB.

```
%editconf -c aPP_A7Y20_2_min1.gro -n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2_min1.pdb
```



Starting MD: It's Getting Hotter in Here!

md1a - 50 ps. Annealing from 0 to 310 K over first 25 ps, held at 310 K for second 25 ps. Protein restrained with fc = 1000.
Generated aPP_A7Y20_2_md1a.mdp file for minimization run.

```
title                    = aPP_A7Y20_2_md1a
cpp                      = /lib/cpp
define                   = -dPOSRES
integrator               = md
dt                       = 0.002
tinit                    = 0.0
nsteps                   = 25000
nstxout                  = 5000
nstvout                  = 5000
nstlog                   = 250
nstenergy                = 500
nstxtcout                = 250
xtc_grps                 = Protein  SOL  NA  CL
energygrps               = Protein  SOL  NA  CL
nstlist                  = 10
ns_type                  = grid
rlist                    = 1.0
coulombtype              = PME
fourierspacing           = 0.10
pme_order                = 4
ewald_rtol               = 1e-5
optimize_fft             = yes
rcoulomb                 = 1.0
rvdw                     = 1.0
pbc                      = xyz
rcoulomb                 = 1.0
rvdw                     = 1.0
pbc                      = xyz
tcoupl                   = berendsen
tc-grps                  = Protein  SOL  NA  CL
tau_t                    = 0.1  0.1  0.1  0.1  0.1
ref_t                    = 310  310  310  310  310
annealing                = single  single  single  single
annealing_npoints        = 2  2  2  2  2
annealing_time           = 0  25   0  25   0  25   0  25
annealing_temp           = 0 310   0 310   0 310   0 310
Pcoupl                   = berendsen
pcoupltype               = anisotropic
tau_p                    = 1.0 1.0 1.0 1.0 1.0 1.0
compressibility          = 4.5e-5 4.5e-5 4.5e-5 0 0 0
ref_p                    = 1.0 1.0 1.0 1.0 1.0 1.0
E_z                      = 1 -0.05 1
gen_vel                  = no
gen_temp                 = 310
gen_seed                 = 173529
constraints              = all-bonds
constraint_algorithm     = lincs
unconstrained_start      = no
```

Starting MD: It's Getting Hotter in Here!

md1a - Generated position restraints for PROTEIN.

```
%genpr  -f aPP_A7Y20_2_md1a.gro  -n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2_A(B)_pr.itp  -fc 1000 1000 1000
```

Generated start file for MD run.

```
%grompp  -f aPP_A7Y20_2_md1a.mdp  -c aPP_A7Y20_2_min1.gro
-p aPP_A7Y20_2.top  -o aPP_A7Y20_2_md1a.tpr
```

MD run started.

```
%mdrun  -s aPP_A7Y20_2_md1a.tpr  -o aPP_A7Y20_2_md1a.trr
-c aPP_A7Y20_2_md1a.gro  -x aPP_A7Y20_2_md1a.xtc
-e aPP_A7Y20_2_md1a.edr  -g aPP_A7Y20_2_md1a.log &
```

Showing Some Restraint

md1b - 50 ps. Backbone restrained with fc = 1000. Generated aPP_A7Y20_2_md1b.mdp file for minimization run. Similar to md1a.mdp, no annealing, start time = 50.0.

```
title                    = aPP_A7Y20_2_md1b
cpp                      = /lib/cpp
define                   = -dPOSRES
integrator               = md
dt                       = 0.002
tinit                    = 50.0
nsteps                   = 25000
nstxout                  = 5000
nstvout                  = 5000
nstlog                   = 250
nstenergy                = 500
nstxtcout                = 250
xtc_grps                 = Protein  SOL  NA  CL
energygrps               = Protein  SOL  NA  CL
nstlist                  = 10
ns_type                  = grid
rlist                    = 1.0
coulombtype              = PME
fourierspacing           = 0.10
pme_order                = 4
ewald_rtol               = 1e-5
optimize_fft             = yes
rcoulomb                 = 1.0
rvdw                     = 1.0
pbc                      = xyz
rcoulomb                 = 1.0
rvdw                     = 1.0
pbc                      = xyz
tcoupl                   = berendsen
tc-grps                  = Protein  SOL  NA  CL
tau_t                    = 0.1  0.1  0.1  0.1  0.1
ref_t                    = 310  310  310  310  310
Pcoupl                   = berendsen
pcoupltype               = anisotropic
tau_p                    = 1.0 1.0 1.0 1.0 1.0 1.0
compressibility          = 4.5e-5 4.5e-5 4.5e-5 0 0 0
ref_p                    = 1.0 1.0 1.0 1.0 1.0 1.0
E_z                      = 1 -0.05 1
gen_vel                  = no
gen_temp                 = 310
gen_seed                 = 173529
constraints              = all-bonds
constraint_algorithm     = lincs
unconstrained_start      = no
```

Showing Some Restraint

md1b - Generated position restraints for BACKBONE.

```
%genpr  -f aPP_A7Y20_2_md1a.gro  -n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2_A(B)_pr.itp  -fc 1000 1000 1000
```

Generated start file. -t and -e read in the exact final velocities and pressure coupling statistics to restart the MD properly.

```
%grompp  -f aPP_A7Y20_2_md1b.mdp  -c aPP_A7Y20_2_md1a.gro
-p aPP_A7Y20_2.top  -o aPP_A7Y20_2_md1b.tpr
-t aPP_A7Y20_2_md1a.trr -e aPP_A7Y20_2_md1a.edr
```

MD run started.

```
%mdrun  -s aPP_A7Y20_2_md1b.tpr  -o aPP_A7Y20_2_md1b.trr
-c aPP_A7Y20_2_md1b.gro  -x aPP_A7Y20_2_md1b.xtc
-e aPP_A7Y20_2_md1b.edr  -g aPP_A7Y20_2_md1b.log &
```

Showing Less Restraint

md1c - 50 ps. Backbone restrained with fc = 500. aPP_A7Y20_2_md1c.mdp identical to aPP_A7Y20_2_md1b.mdp except for start time, tinit = 100.0.

Generated weaker position restraints for BACKBONE.

```
%genpr  -f aPP_A7Y20_2_md1b.gro  -n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2_A(B)_pr.itp  -fc 500 500 500
```

Generated start file for MD run. Again, -t and -e.

```
%grompp  -f aPP_A7Y20_2_md1c.mdp  -c aPP_A7Y20_2_md1b.gro
-p aPP_A7Y20_2.top  -o aPP_A7Y20_2_md1c.tpr
-t aPP_A7Y20_2_md1b.trr -e aPP_A7Y20_2_md1b.edr
```

MD run started.

```
%mdrun  -s aPP_A7Y20_2_md1c.tpr  -o aPP_A7Y20_2_md1c.trr
-c aPP_A7Y20_2_md1c.gro  -x aPP_A7Y20_2_md1c.xtc
-e aPP_A7Y20_2_md1c.edr  -g aPP_A7Y20_2_md1c.log &
```

Run Free Little Molecules, Run Free!

md1 - 850 ps. Generated aPP_A7Y20_2_md1.mdp file for minimization run, identical to aPP_A7Y20_2_md1c.mdp except that tinit = 150.0, nsteps = 425000 and “define = -dPOSRES” has been deleted.
Generated start file for MD run.

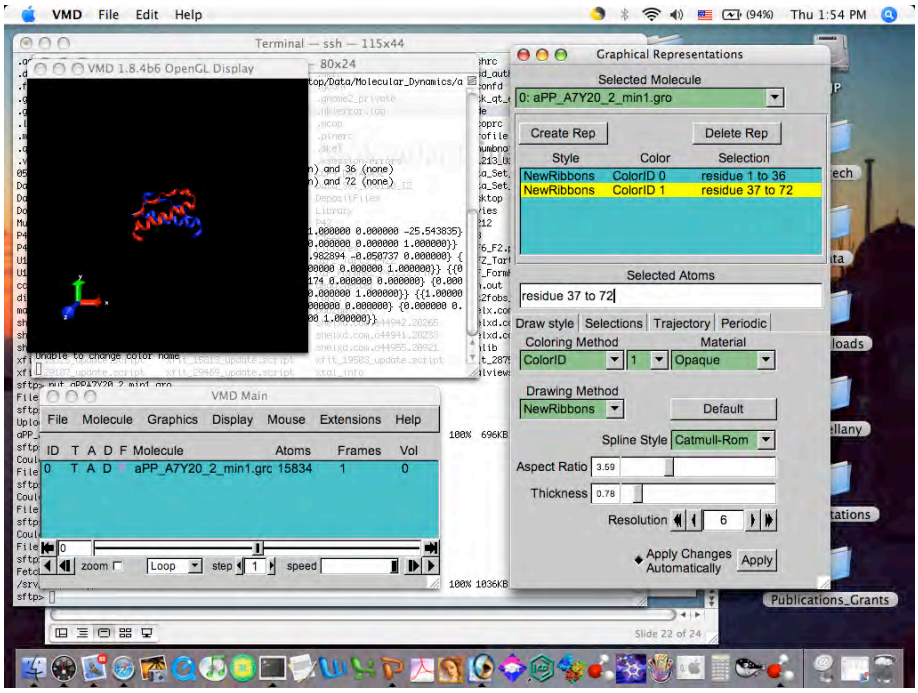
```
%grompp  -f aPP_A7Y20_2_md1.mdp  -c aPP_A7Y20_2_md1c.gro
-p aPP_A7Y20_2.top  -o aPP_A7Y20_2_md1.tpr
-t aPP_A7Y20_2_md1c.trr -e aPP_A7Y20_2_md1c.edr
```

MD run started.

```
%mdrun  -s aPP_A7Y20_2_md1.tpr  -o aPP_A7Y20_2_md1.trr
-c aPP_A7Y20_2_md1.gro  -x aPP_A7Y20_2_md1.xtc
-e aPP_A7Y20_2_md1.edr  -g aPP_A7Y20_2_md1.log &
```

All subsequent runs mirror md1.

VMD: The Molecular Cineplex

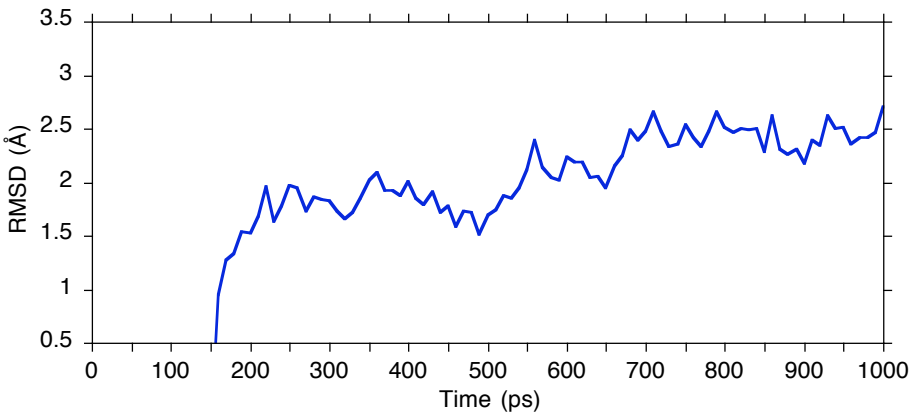


Getting Numbers, Not Just Movies

To find the RMSD of the protein atoms as a function of time:

```
%g_rms  -s aPP_A7Y20_2_md1.tpr  -f aPP_A7Y20_2_md1.trr
-n aPP_A7Y20_2.ndx  -o aPP_A7Y20_2_md1_rmsd.xvg
```

Select PROTEIN.

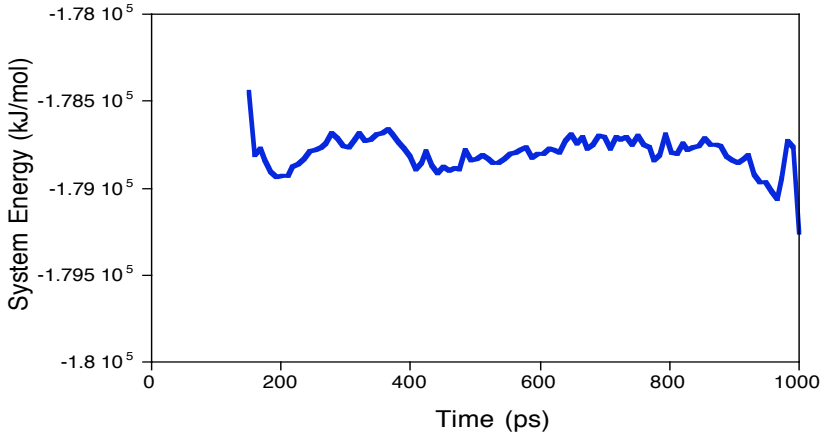


Getting Numbers, Not Just Movies

To find the energy of the system as a function of time:

```
%g_energy  -f aPP_A7Y20_2_md1.edr  -o aPP_A7Y20_2_md1_enrg.xvg
```

Select 11 = Total Energy.



Getting Numbers, Not Just Movies

To find the distance between the centers of mass of the two aPP molecules as a function of time:

First you need to edit the index file to define the sets of residues that make up aPP molecule A and aPP molecule B. You need to look in the .gro file to figure this out.

```
%make_ndx -f aPP_A7Y20_2_md1.gro -n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2.ndx
```

When prompted, select the appropriate residues, then rename this group aPPA.

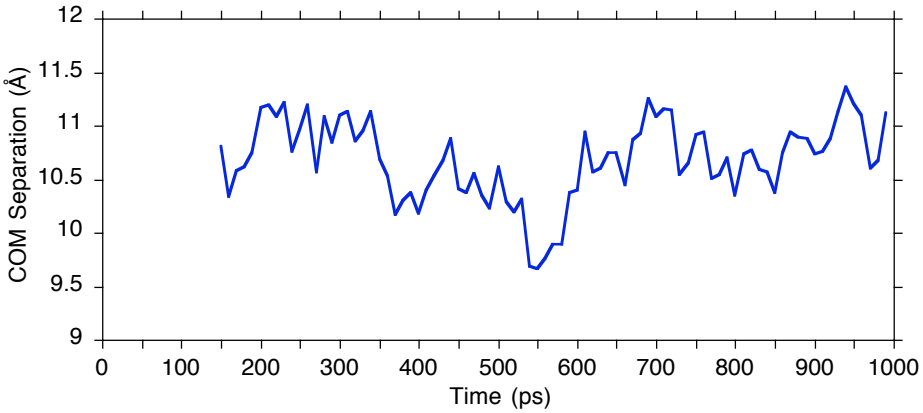
```
> r1 | r2 |  r3 | . . . . | r35 | r36
> name 17 aPPA
```

Repeat this for residues 37 to 72, name this group aPPB.

Getting Numbers, Not Just Movies

Use g_dist. Pick index groups aPPA and aPPB when asked.

```
%g_dist  -f aPP_A7Y20_2_md1.trr  -n aPP_A7Y20_2.ndx
-s aPP_A7Y20_2_md1.tpr -o aPP_A7Y20_2_comdist.xvg
```



Getting Numbers, Not Just Movies

To find the plane angle between two Tyrs as a function of time:

First you need to edit the index file to sets of three atoms to define the planes of each Tyr sidechain. Here, we use CG, CE1, and CE2 (atoms 271, 276, 278 and atoms 842, 847, 849).

```
%make_ndx -f aPP_A7Y20_2_md1.gro -n aPP_A7Y20_2.ndx
-o aPP_A7Y20_2.ndx
```

When prompted, select the appropriate atoms, then rename this group aPPA_Y20.

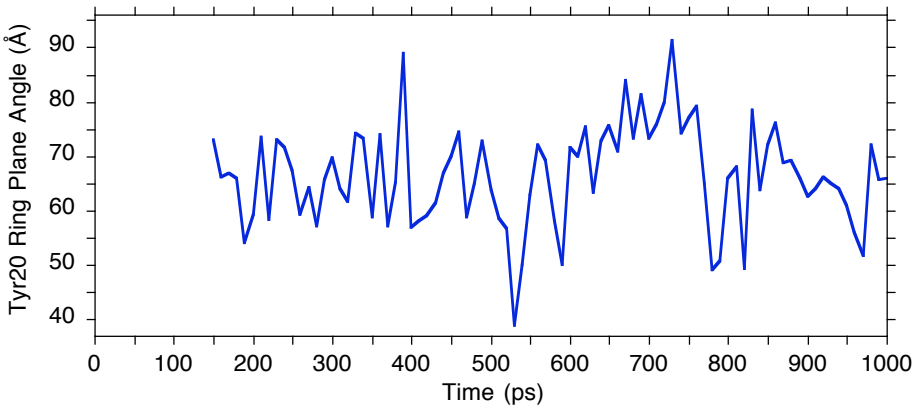
```
> a271 | a276 | a278
> name 18 aPPA_Y20
```

Repeat this for atoms 842, 847, 849, name this group aPPB_Y20.

Getting Numbers, Not Just Movies

Use g_sgangle. Pick the index groups aPPA_Tyr20 and aPPB_Tyr20 when asked.

```
%g_sgangle  -f aPP_A7Y20_2_md1.trr  -n aPP_A7Y20_2.ndx
-s aPP_A7Y20_2_md1.tpr -oa aPP_A7Y20_2_Y20ang.xvg
```



References

Swiss PDB Viewer website. Decent userguide.

http://us.expasy.org/spdbv/

GROMACS website. Good downloadable PDF manual and very helpful listserver.

http://www.gromacs.org

VMD website. Online or PDF manual.

http://www.ks.uiuc.edu/Research/vmd/