Tutorial: Build and Simulate a Protein - Lipid System

This document assumes that you have basic working knowledge of CHARMM topology and parameter files and their generation and usage. One starting point to prepare is to review lectures 1 & 2 and carry out labs 0, I, II, and III in the Training course.

A few commands below require access to a Linux/Unix-like/OS X command line. To some extent these may be achieved on a Windows system using tools provided in the lab_0.pdf and the basic_windows_commands.pdf documents in the Training course. The commands below have **not** been tested on a Windows system.

Preamble

There are several force field subsets for CHARMM Mackerell Lab including the CHARMM generalized force field to build small molecules not contained in the canonical force field subsets (protein, nucleic acids, carbohydrates, lipids, etc.).

An introduction to CHARMM and it's usage can be found on the CHARMM Wiki Tutorial.

In this lab we will use an NMR structure of the HIV-1 MA protein in an orientation that is consistent with experimental data based on fitting neutron reflectivity data of the protein bound to a lipid membrane. The orientation and penetration of the protein into the membrane is consistent with experimental data. We will use these structural parameters to construct a protein-membrane system and run molecular dynamics simulation.

Simulation guided by experiment is a powerful tool for observing molecular interactions and mechanisms at a level of detail that is not observable by any experimental method.

In this lab you will:

1) Build a membrane and protein system
2) Orient the protein on the membrane based on experimental parameters
3) Delete overlapping atoms and solvate and ionize the system
4) Learn to equilibrate membrane protein simulations using external force constraints

---

Build Lipid Membrane Protein System with HIV-1 Matrix Protein and POPC Lipids

A) Clean up HIV-1 Matrix protein PDB file: Download hiv1_gag_matrix.pdb and use psfgen to build a PDB and a PSF file. In this lab you will use the following force field topology files: top_all36_prot.rtf, top_all36_lipid.rtf, and parameter files: par_all36_prot.prm, par_all36_lipid.prm, and toppar_water_ions_namd.str. Name the new files in your script gag_ma.pdb and gag_ma.psf. There will be several alias commands that you will have to employ and at least one non-standard patch.

Note that in this example all files created during the building process are saved in a folder called **output_building**. Files created during the dynamics runs are saved in a folder called **output**. Create these two folders now.

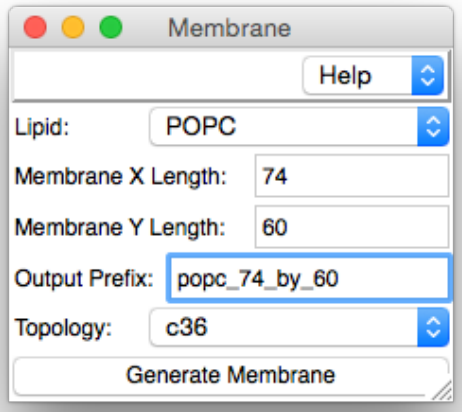
B) Build POPC Membrane:

- Load your newly created gag_ma.psf and gag_ma.pdb files into VMD and use the **measure minmax** command to determine the dimensions of the protein.

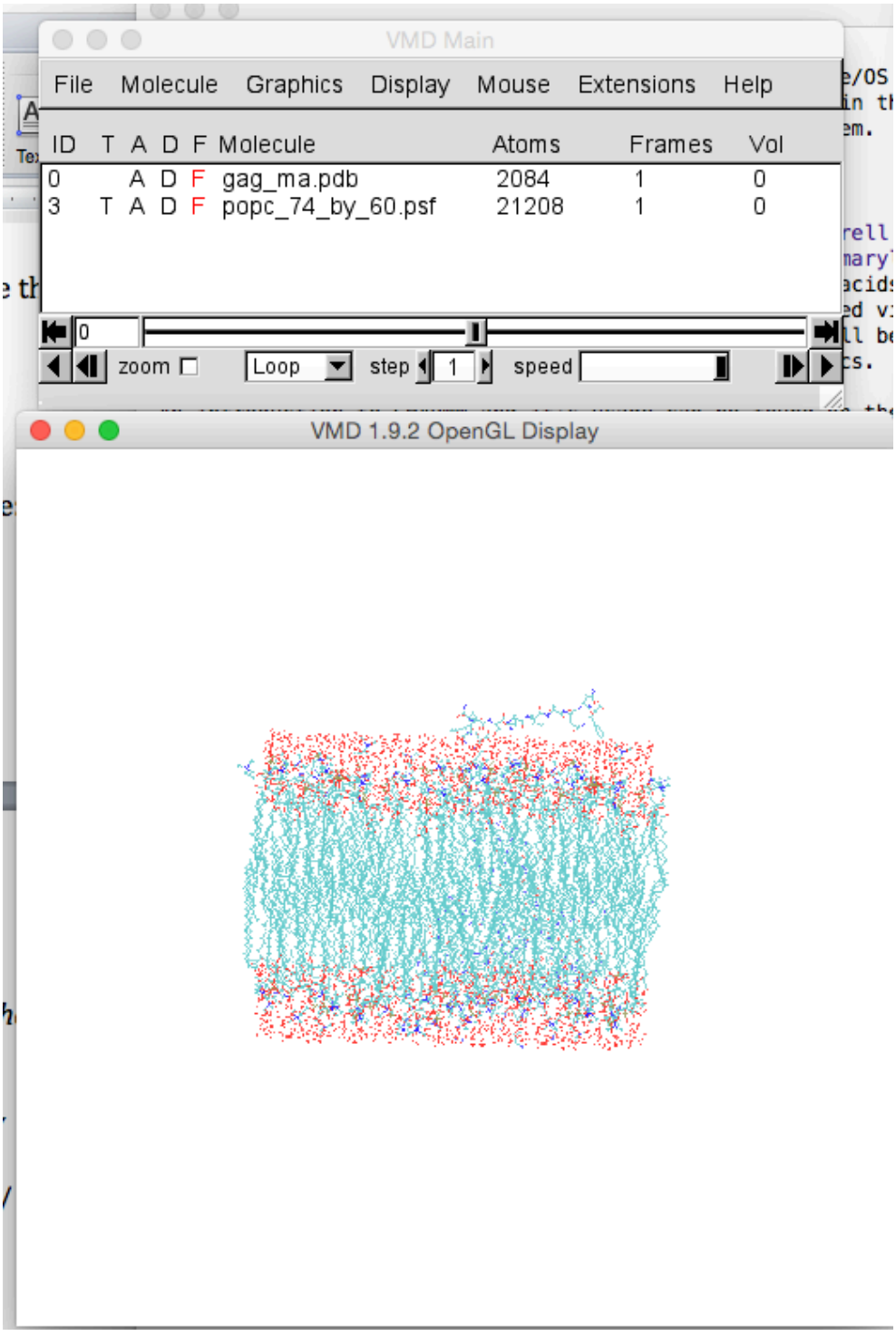
```
vmd> set protein [atomselect top "protein"]

vmd> measure minmax $protein
```

You will need this size to determine the size of the membrane patch that we are going to create. It should be roughly +/- 12 angstroms longer than the protein.

- In VMD go to "Extensions -> Modeling -> Membrane Builder" and set the following parameters then "Generate Membrane"



- Which results in the following VMD Display. Use VMD to view the membrane by using segname. Notice the imperfect packing and voids along the edges. This will be fixed in a subsequent section.



C) Build Protein - Membrane System: Use the newly created PDB files (gag_ma.pdb/gag_ma.psf and popc_74_by_60.pdb/popc_74_by_60.psf) and use psfgen to build new PDB and PSF files for the complete

system. Since you have complete PDB/PSF pairs that have no missing atoms and all patches have been applied you can easily create a new combined PDB/PSF pair. Since this is a new way to make a combined PDB/PSF pair we'll give you the script here which is shown in detail below. Note that the final files are named **gag_ma_popc.pdb** and **gag_ma_popc.psf**.

```
set path output_building
resetpsf
readpsf $path/gag_ma.psf
coordpdb $path/gag_ma.pdb
readpsf $path/popc_74_by_60.psf
coordpdb $path/popc_74_by_60.pdb
writepdb $path/gag_ma_popc.pdb
writepsf $path/gag_ma_popc.psf
```

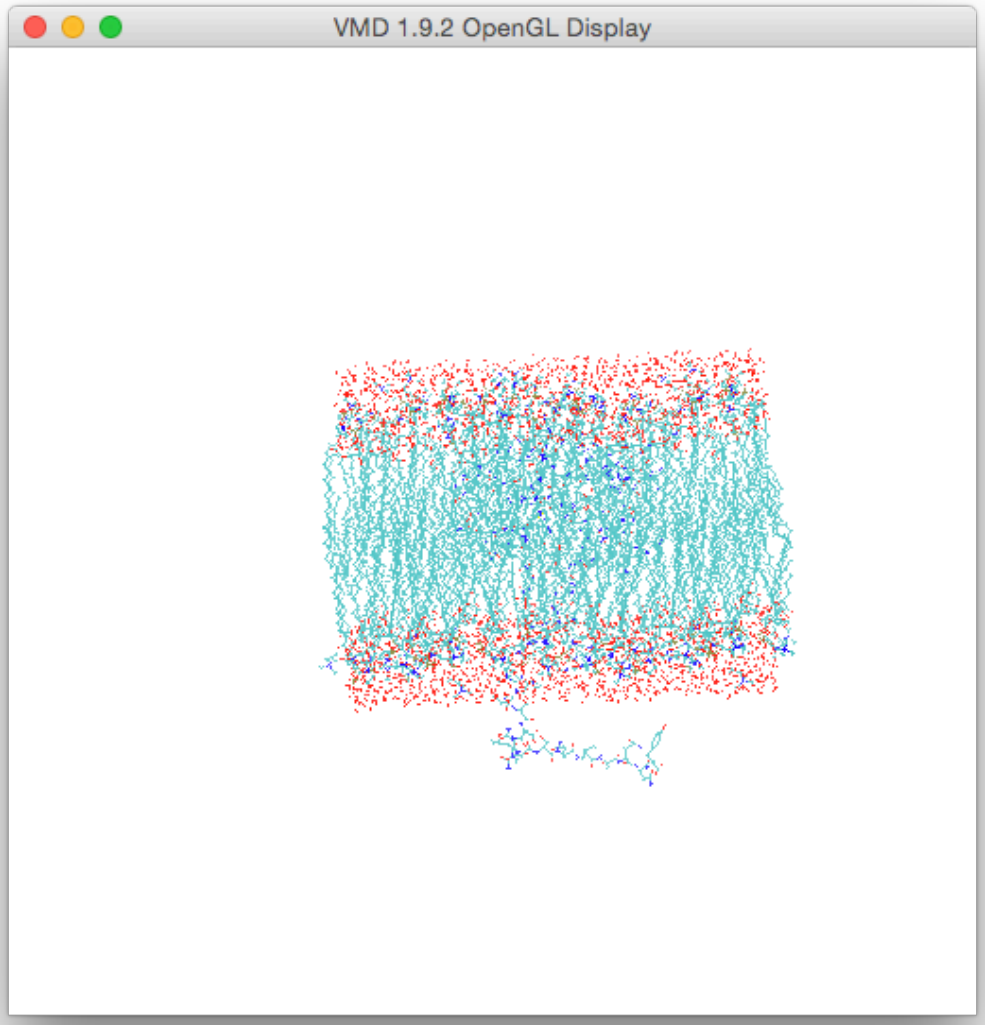
Build Lipid Membrane Protein System with HIV-1 Matrix Protein and POPC Lipids

In this step you will center the system, rotate and translate the protein to match experimentally determined orientation and insertion depth.

A) Center system: Start a new VMD session and load gag_ma_popc.psf and gag_ma_popc.pdb. Then to center the protein type the following into the TkConsole

```
vmd > set protein [atomselect top "protein"]
vmd > $protein moveby [vecinvert [measure center $protein]]
```

Then repeat this process for the membrane plus water together by making the appropriate selection and using similar syntax as used above. This should give you an representation as shown below.



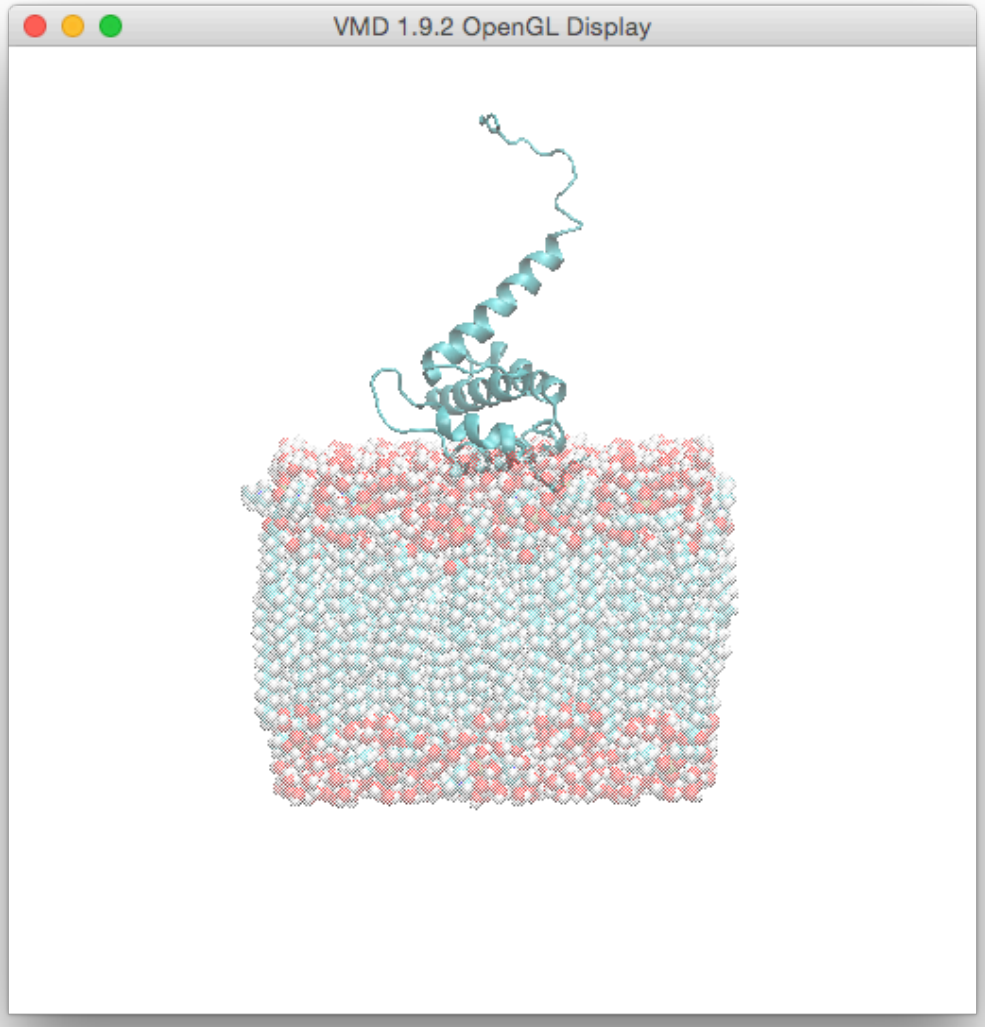
B) Orient protein: Modeling of neutron reflectivity data indicates that the optimal orientation of HIV-1 Gag Matrix protein is with the Euler angles (theta,phi) = (20,-120). Therefore the commands are

```
vmd> $protein move [transaxis z -120]
vmd> $protein move [transaxis y 20]
```

C) Translate protein to insertion depth: HIV-1 Gag Matrix protein only inserts ~4.8 angstroms into the lipid

layer. Therefore we should translate the protein 39.7 angstroms in the z-dimension. After entering the following commands save the coordinates of the full system to a new PDB file named **gag_ma_popc_orientation_and_shift.pdb**.

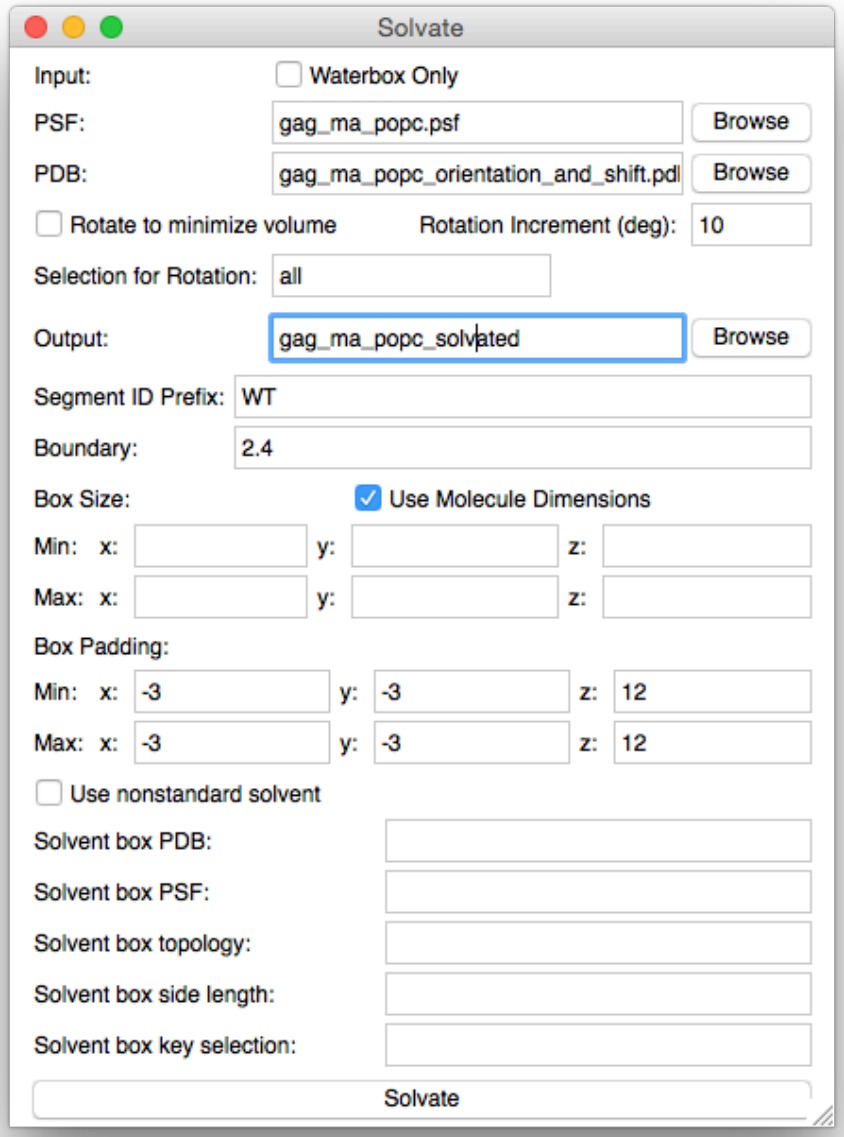
```
vmd > $protein moveby {0 0 39.7}
```



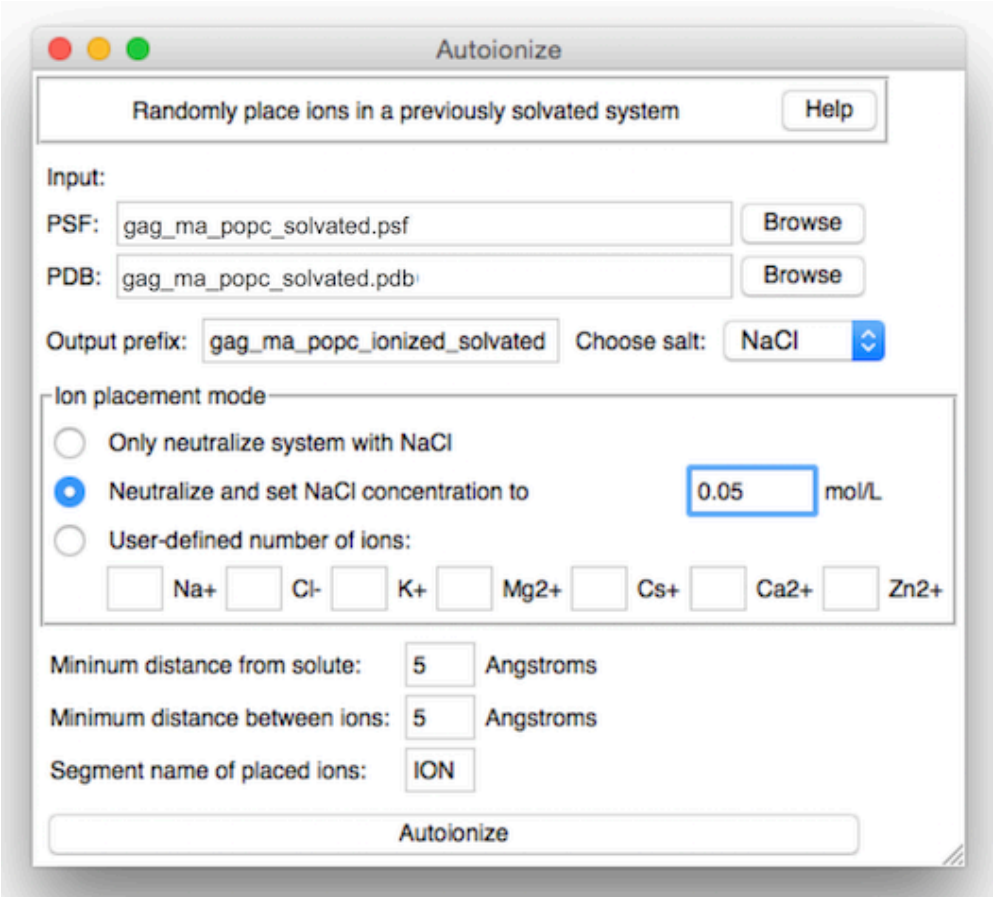
Solvate, Ionize and Delete Overlapping Waters

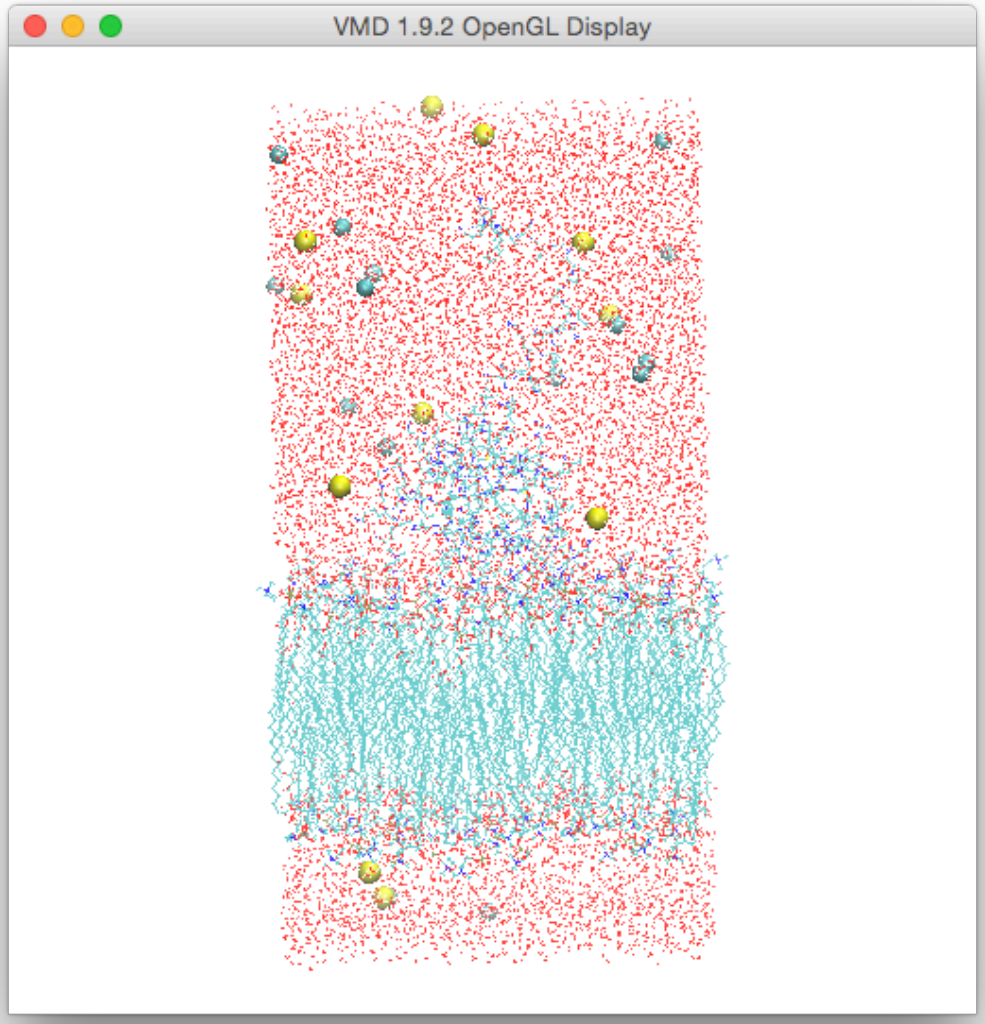
Using the same general method used to build complete systems with systems in previous labs.

A) Solvate system: Start a new VMD session and load gag_ma_popc.psf and gag_ma_popc_orientation_and_shift.pdb. Use the "Extension -> Modeling -> Add Solvation Box" tool with **Box Padding** set to xmin=xmax=ymin=ymax=-3 and zmin=zmax=12 and save the new file as **gag_ma_popc_solvated** (VMD will add the correct suffixes: .pdb & .psf).



B) Add Ions: Use the "Extension -> Modeling -> Add Ions" tool with the concentration set to 0.05 mol/L, neutralize check box selected, segname to "ION" and the output prefix as **gag_ma_popc_ionized_solvated** (VMD will add the correct suffixes: .pdb & .psf). The input and ionized / solvated system should look similar to the following.


Autoionize
Randomly place ions in a previously solvated system
Help
Input:
PSF:
gag_ma_popc_solvated.psf
Browse
PDB:
gag_ma_popc_solvated.pdb
Browse
Output prefix:
gag_ma_popc_ionized_solvated
Choose salt:
NaCl
Ion placement mode
Only neutralize system with NaCl
Neutralize and set NaCl concentration to
0.05
mol/L
User-defined number of ions:
Na+
Cl-
K+
Mg2+
Cs+
Ca2+
Zn2+
Mininum distance from solute:
5
Angstroms
Minimum distance between ions:
5
Angstroms
Segment name of placed ions:
ION
Autoionize

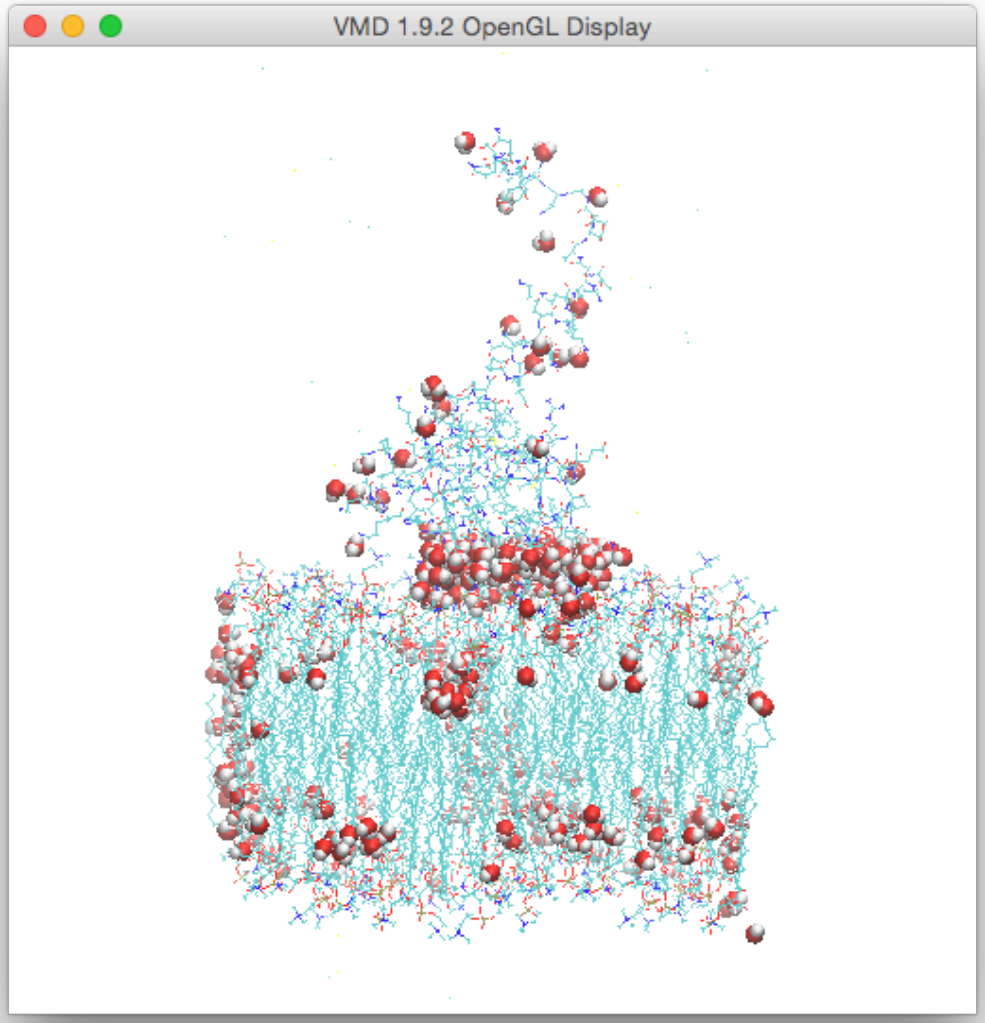


C) Remove Waters: Some waters that originally solvated the membrane now overlap with the protein. Furthermore because POPC lipids are not yet packed properly there are gaps and voids in the hydrophobic region that have been filled with water. By using "Graphical Representations" we can visualize waters as a function of distance from various parts of the system.

- Create the following representations:

```
not water
water and same residue as within 2.5 of protein
(segname WT1 to WT99) and same residue as abs(z) < 25
```

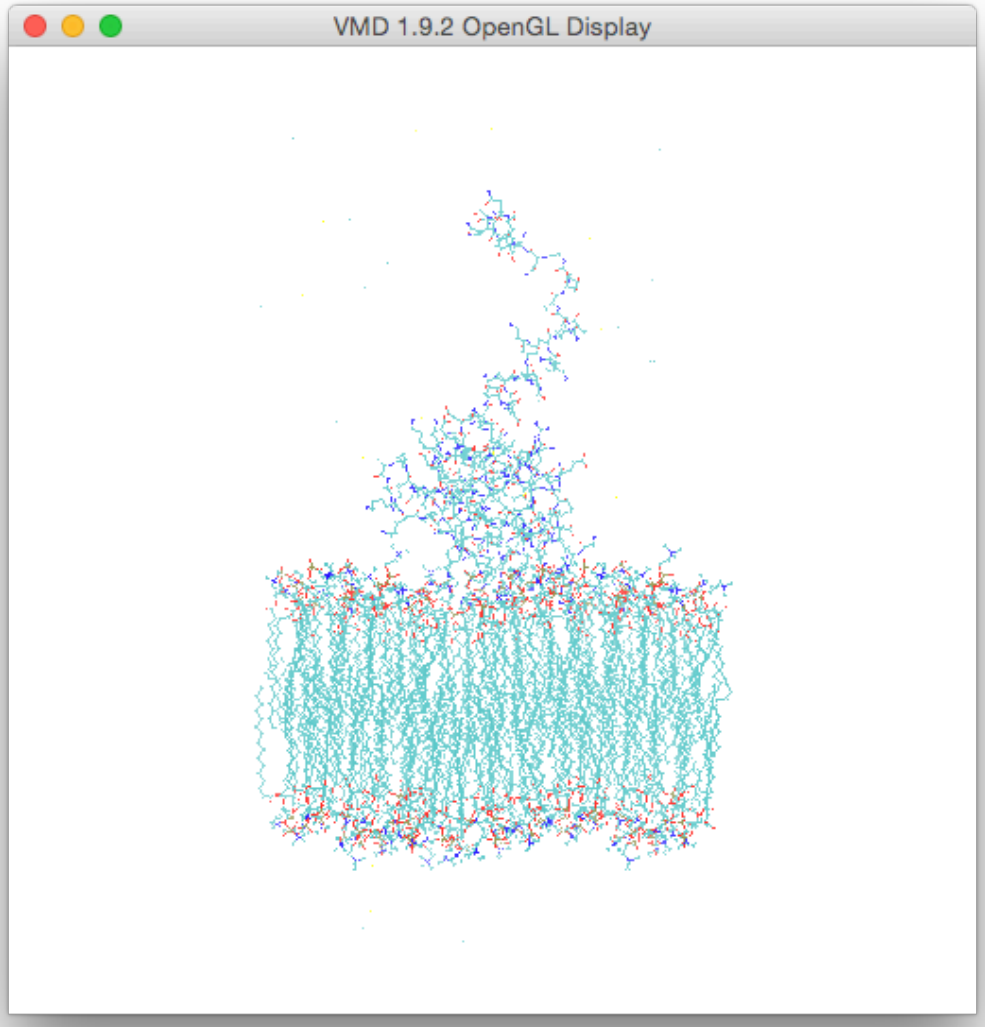
This will show both the waters that are too close to the protein and the waters that were inserted into the hydrophobic region of the lipid membrane core. Your image will resemble the following.



To remove the waters download remove_water.tcl open VMD and type the following in the TkConsole:

```
vmd > source remove_water.tcl
```

This script will remove the bad waters. Using the same representations used above one can see that the inappropriate waters have been removed. Note that in the picture, due to our selections, the remaining water molecules are not shown.



The script will save the cleaned up system in the folder output_building a files called **gag_ma_popc_ionized_solvated_clean.pdb** and **gag_ma_popc_ionized_solvated_clean.psf**. These files

will be used in the subsequent section to energy minimize and run molecular dynamics simulation.

Equilibration and Dynamics

As was done in lab II, we will perform a series of energy minimization and dynamics runs with parts of the system either held fixed or with external harmonic constraints. Unlike the previous building examples, this system has protein inserted into the membrane (consistent with experimental results) and the lipid chains themselves are not equilibrated, thus extra care must be taken to avoid distorting the structure and energy minimization runs alone will not adequately relax the system. Therefore we will also perform constrained dynamics runs.

A) Download and use the following script to generate a PDB file with beta values set to 1.00 for the lipid head groups: set_beta_1_popc_fixed_lipid_head_groups.tcl. Read the script to understand what it is doing. Alternatively, you can execute the commands in the script in the TkConsole. This will create a PDB file called **gag_ma_popc_ionized_solvated_clean_fixed_lipid_head_groups.pdb**) to use to set constraints in the NAMD run. To execute the script, in a TkConsole type:

```
vmd> source set_beta_1_popc_fixed_lipid_head_groups.tcl
```

B) Download and edit the NAMD input script dyn0.inp.

```
* Use the minmax command in VMD to determine the size of your system.
* Edit the CellBasisVector values so that the dimensions of the box are 2 angstroms bigger
than your measured values.
```

Our values are shown below for "all" atoms and for "water" atoms

```
vmd> set everyone [atomselect top "all"]
atomselect0
vmd> measure minmax $everyone
{-39.21500015258789 -32.042999267578125 -40.36800003051758} {38.229000091552734
31.586999893188477 96.68399810791016}
```

X + pad ~ 79.4 : Y + pad ~ 65.6 : Z + pad ~ 139.1

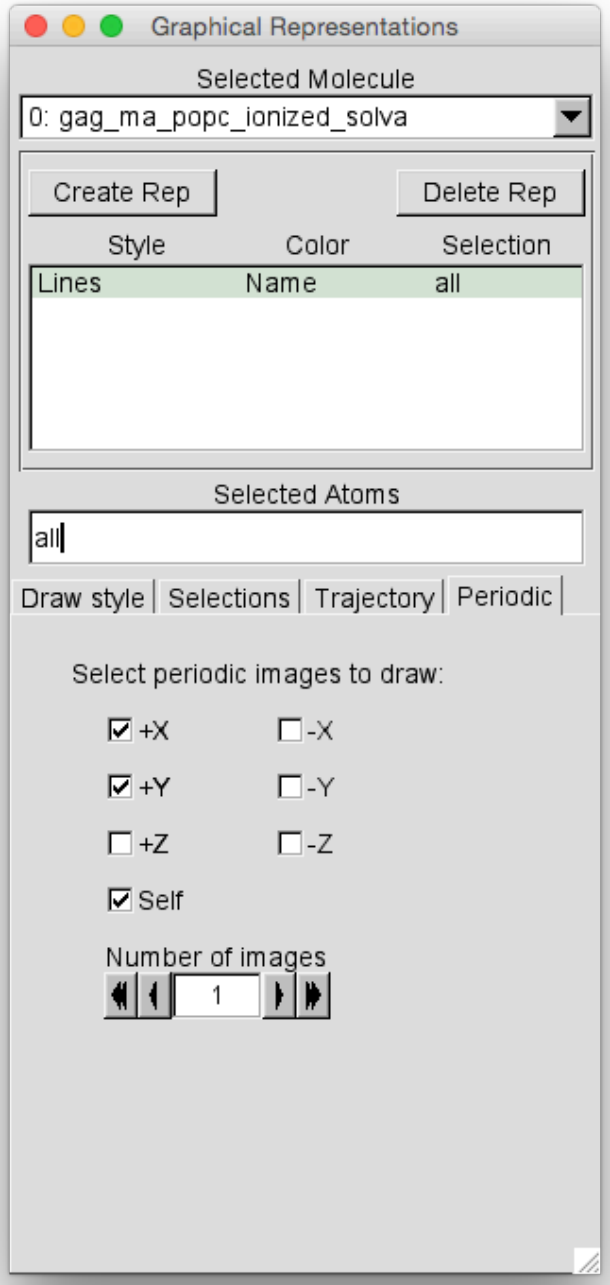
```
vmd> set water [atomselect top "water"]
atomselect1
vmd> measure minmax $water
{-36.2130012512207 -29.040000915527344 -40.36800003051758} {35.224998474121094
28.56800079345703 96.68399810791016}
```

X + pad ~ 73.5 : Y + pad ~ 59.6 : Z + pad ~ 139.0

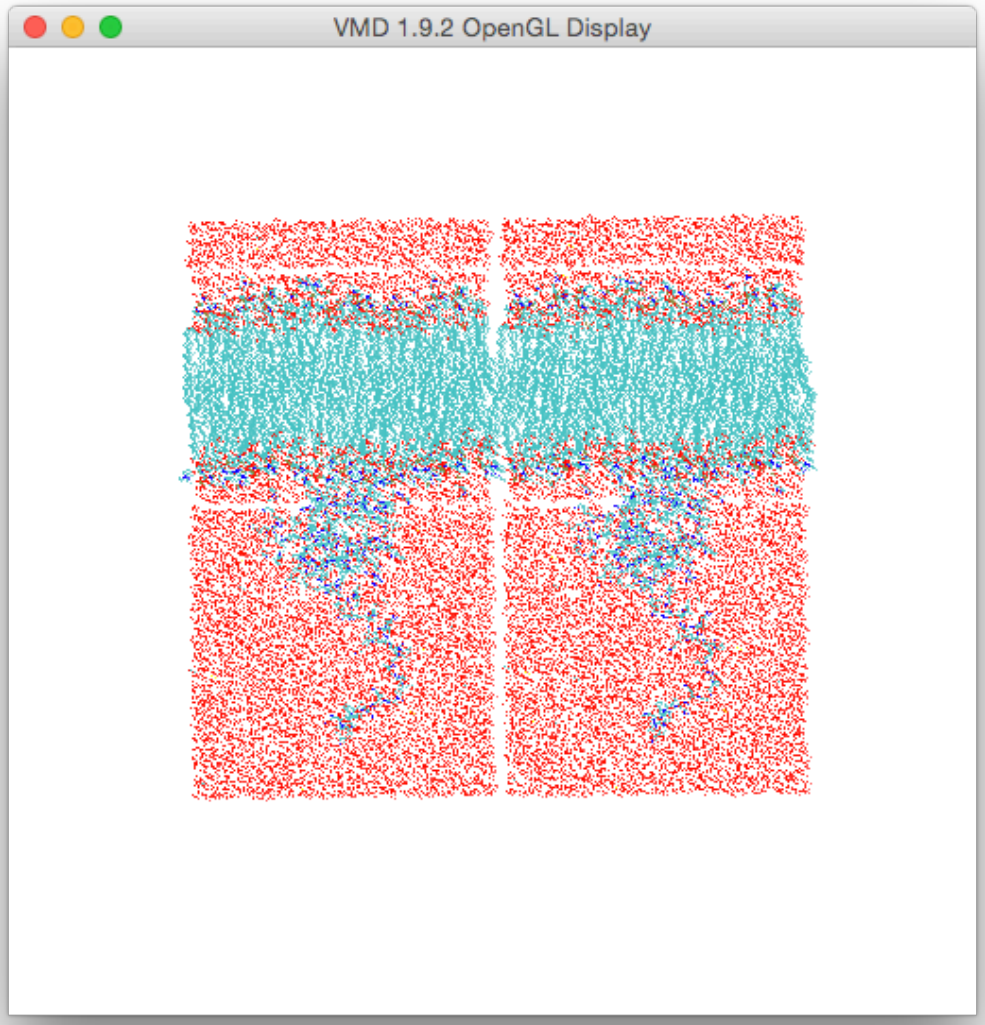
By looking at the Z dimension we can conclude that some of the lipids extend outside the edges (X & Y). To visualize the size of the box using the result of the water measurements above type the following commands in the TkConsole (replacing with your measurements if they are different)

```
vmd> molinfo top set a 74
vmd> molinfo top set b 60
vmd> molinfo top set c 139
```

Then select periodic images as shown below.



Which will look like the following:



These results are typical and the acutal padding that you should use depends on your system. As we progress through the equlibration process we will eventually perform simulation in the NPT ensemble which will allow the system to equilibrate to the correct box size and density.

C) Run constrained minimization and short (20 ps) constrained dynamics run.

While it may be possible to energy minimize and run dynamics on these starting coordinates with fixed lipid head groups, it is likely that there are some steric clashes in the system that need to be relaxed. Therefore, the approach is to apply harmonic constraints to both the protein and water molecules while concurrently

holding the lipid head groups fixed.

There are two things needed to enable harmonic constraints. The first is a PDB file with values written into the "beta" column to indicate which atoms are being constrained. Download and run this script harmonic_protein_and_water.tcl to create a PDB file **gag_ma_popc_ionized_solvated_clean_constrained_water_and_protein.pdb** that contains the constraint information. Then to enable the restraint the following was added to the NAMD input file, before the minimize & run commands

```
constraints on
consexp 2
consref output_building/gag_ma_popc_ionized_solvated_clean.pdb
conskfile gag_ma_popc_ionized_solvated_clean_constrained_water_and_protein.pdb
conskcol B
```

The power of the equation used to enable constraints is defined by the "consexp" value, in this case the square is used so the equation is Energy = k*(r-ro)2. The spring constant is the value put in the "beta" column (in kcal/mol/Ang).

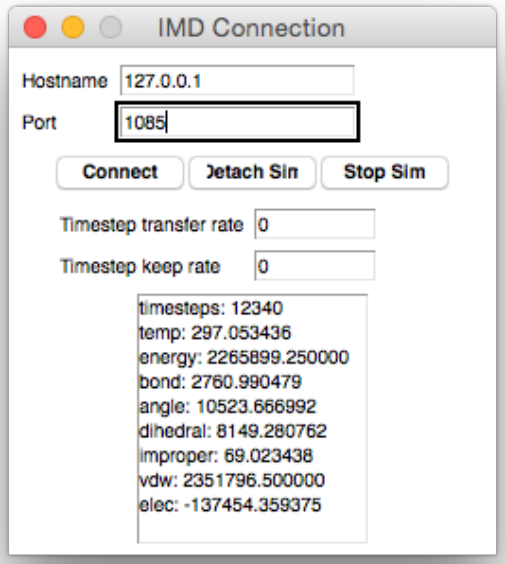
Using the supplied input file dyn0.inp with the edits you provided to set the box size, carry out the constrained simulation. **Make sure you have created the /output directory before starting the simulation!** Note that on 12 CPUs this run took 1.5 hours. During the run you can type:

```
>> tail -90f dyn0.out
```

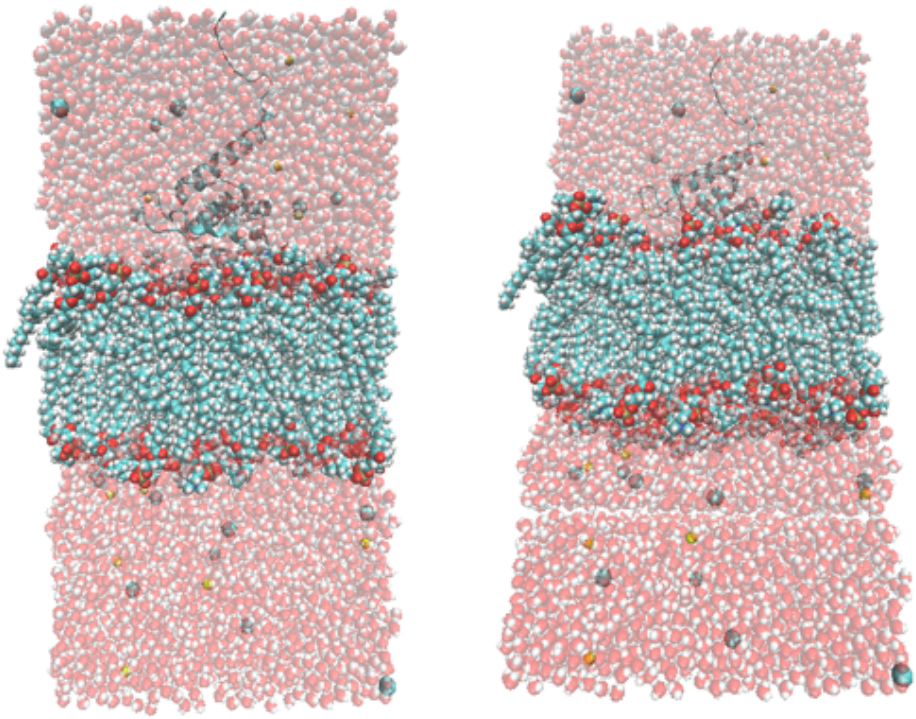
to see the energies and timings, or you can type

```
>> tail -90f dyn0.out | grep emain
```

to track the time remaining. You can also use IMD to watch the run in progress via VMD as shown in this graphic



At the end of the simulation inspect the VDW energies as they could still be high. Unfavorable contacts because of the fixed atoms still prevent some parts of the lipid chains from relaxing properly. Look at the trajectory in VMD by loading the DCD output file. The lipids chains melted to some degree. Check to make sure that no chains have flipped out into the water region. Note that the waters were under a harmonic constraint and therefore gaps remain just above the lipid head groups in the top leaflet and the section and the middle of the bottom water section. The resultant dyn0.coor file for the structure shown below can be downloaded to compare to your result. Since your simulation may take several hours depending on how many processors you have available and the speed of your computer, this file can also be used to move on to step D) below without waiting for your simulation to complete.



D) Relax lipids and waters using TCL Forces.

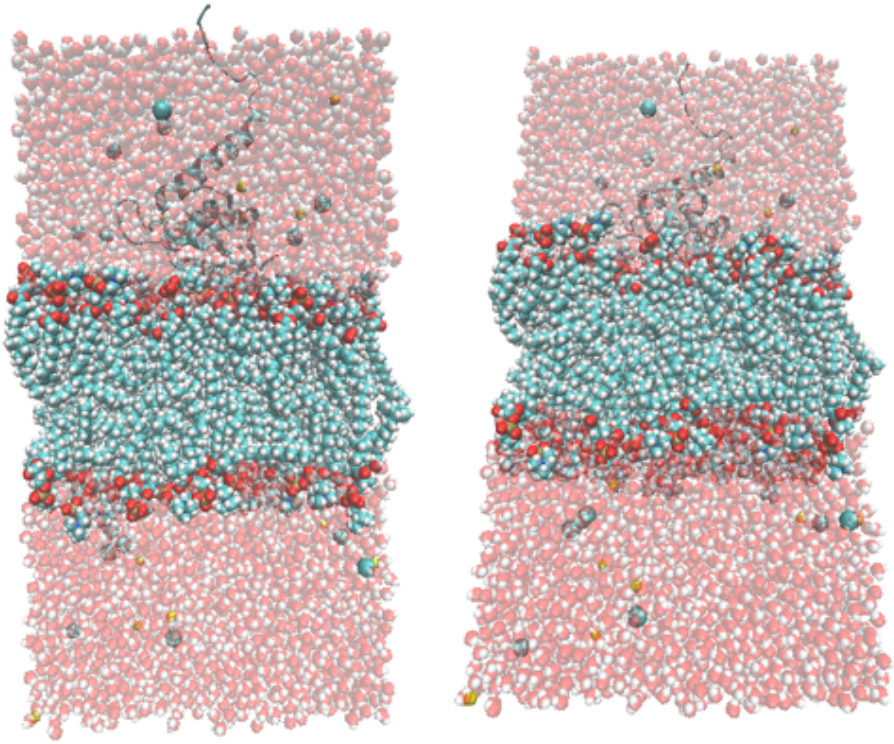
The next step in the relaxation process is to remove the constraints on the lipid head groups and include a bias in the force field to prevent waters from entering the partially relaxed lipid membrane. The harmonic constraints will remain on the protein to prevent it from denaturing in any way due to external forces from the rest of the system. In addition, the simulation will be done in the NPT ensemble so that the volume will begin to equilibrate. The sample input file dyn1.inp has several modifications. Compare this file to dyn0.inp to determine if you recognize and understand the changes. Only the protein will remain constrained so create this file **gag_ma_popc_ionized_solvated_clean_constrained_protein.pdb** by editing the **harmonic_protein_and_water.tcl** file to only constrain the protein to tell NAMD which atoms to be constrained in the harmonic potential.

```
constraints on
consexp 2
consref output/dyn0.coor
conskfile gag_ma_popc_ionized_solvated_clean_constrained_protein.pdb
conskcol B
```

Since the system has not been equilibrated in the NPT ensemble the volume is far from equilibrium. Therefore it is necessary to apply a force bias on the water molecules to prevent them from entering the lipid bilayer. Download the following script keep_water_out.tcl and note the addition of the following lines near the bottom of the dyn1.inp script. This will keep waters from entering the hydrophobic core of the membrane.

```
tclforces on
set waterCheckFreq 100
set lipidCheckFreq 100
set allatompdb output/dyn0.coor
tclForcesScript keep_water_out.tcl
```

As before, this run took ~1.5 hours using 12 CPUs. At the end of the run you should notice that the gaps in between the water segments are gone. The resultant dyn1.coor file for the structure shown below can be downloaded to compare to your result. Since your simulation may take several hours depending on how many processors you have available and the speed of your computer, this file can also be used, along with the velocity dyn1.vel and extended system dyn1.xsc files, to move on to step E) below without waiting for your simulation to complete.

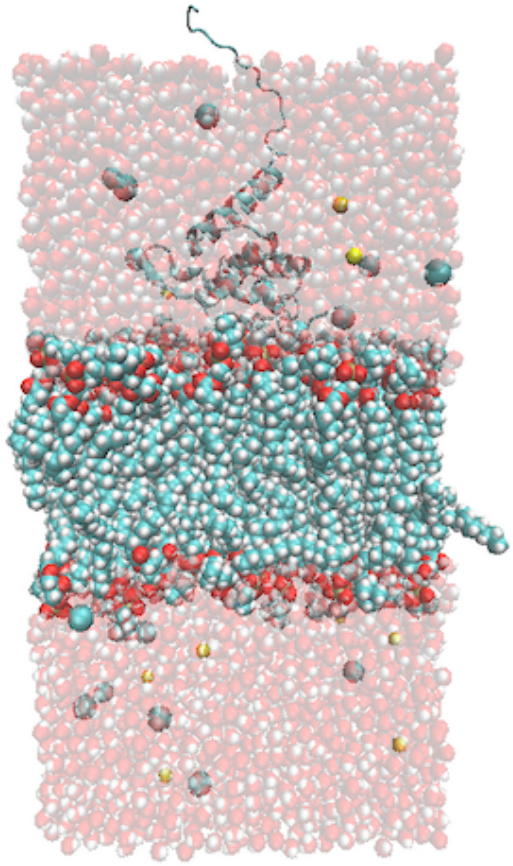


E) NPT Equilibration of Entire System.

The purpose of this step will be to remove all constraints and bias forces and begin equilibrating the system. This will involve reading in the previous coordinates, velocities, and extended system files that set the coordinates, temperature and box size respectively. The temperature variable, PBC (cellbasisVector lines), constraint definitions, and the minimize statement need to be commented out.

Run using the same number of run steps as in the previous run. This run also took ~1.5 hours using 12 CPUs. The resultant dyn2.coor file for the structure shown below can be downloaded to compare to your result. Since your simulation may take several hours depending on how many processors you have available and the speed of your computer, this file can also be used, along with the velocity dyn2.vel and extended system

dyn2.xsc files, to move on to step F) below without waiting for your simulation to complete.



F) NPT Production Run.

Set up a new input file to continue from the previous equilibration run. You will need to update the coordinate, velocity, and extended system input files, the names of the output files, and set the number of steps to 1000000 (2 ns).

At the end of the run, using the procedure from lab III, make a plot of the energies, temperature, and volume. In addition, make a RMSD plot of the MA protein from this simulation. Does your analysis indicate that 2 ns was long enough to equilibrate the system?

This simulation will take awhile (~15 hours using 12 CPUs, or more if using fewer CPUs). The results of this NPT production run and the tcl scripts needed for the analysis are provided as a zip file finished_dynamics.zip so that you can perform the analysis without waiting for your simulation to complete. You can compare the results from your simulation once it is finished.

finished_dynamics/ dyn3.out
rmsd.tcl
rmsd_no_correction.tcl
residue_rmsd.tcl

namdstats.tcl

finished_dynamics/output/
dyn3.coor
dyn3.dcd
dyn3.rest.pdb.coor
dyn3.rest.pdb.vel
dyn3.rest.pdb.xsc
dyn3.vel
dyn3.xsc

For your own membrane simulations, it is important to compare properties of your lipid membrane to known structural parameters. These commonly include: area/lipid, headgroup to headgroup spacing, electron or neutron density profiles and for multi-layer systems one can compare directly to experimental structure factors.

Reference(s) and Citations

1. Scalable molecular dynamics with NAMD James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. J. Comput. Chem. 26, 1781-1802 (2005). BIBTeX EndNote Plain Text
2. NAMD User's Guide
3. VMD User's Guide
4. Electrostatic interactions and binding orientation of HIV-1 matrix, studied by neutron reflectivity H. Nanda, S. A. K. Datta, F. Heinrich, A. Rein, S. Krueger, J. E. Curtis, Biophys. J. 99, 2516-2524 (2010). BIBTex EndNote Plain Text
5. SLDMOL: A tool for the structural characterization of thermally disordered membrane proteins J. E. Curtis, H. Zhang, H. Nanda, Comp. Phys. Comm. 185, 3010-3015 (2014). BIBTex EndNote Plain Text

---

Go to top

Supported via CCP-SAS a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant