MD Sims Guide for Petrache Lab

Ronan Tiu and Daniel Nykamp

August 2024?

1 Setting Up CHARMM-GUI and Big Red 200 Accounts

To generate our lipid bilayer, we use an online interface called Chemistry at Harvard Molecular Mechanics (CHARMM-GUI). The first thing you will want to do is go to charmm-gui.org and register an account.

Eventually you will also need accounts with Big Red 200, the supercomputer that your simulations will run on. Work with Dr. Petrache to get yourselves registered. You must sign up to use Big Red 200 online, and then Dr. Petrache can grant you permission to use the supercomputer by officially adding your email addresses to the project on his end.

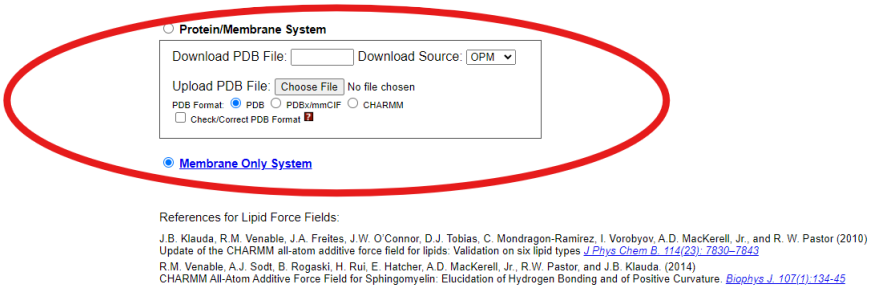
It will take a few days before your accounts are approved so try and get this set up right away.

2 Generating a Lipid Bilayer Simulation

Go back to the charmm-gui.org. On the menu to the left select "Input Generator" (you may have to login first). On the left menu hover over "Membrane Builder" and click on "Bilayer Builder".

Step 1

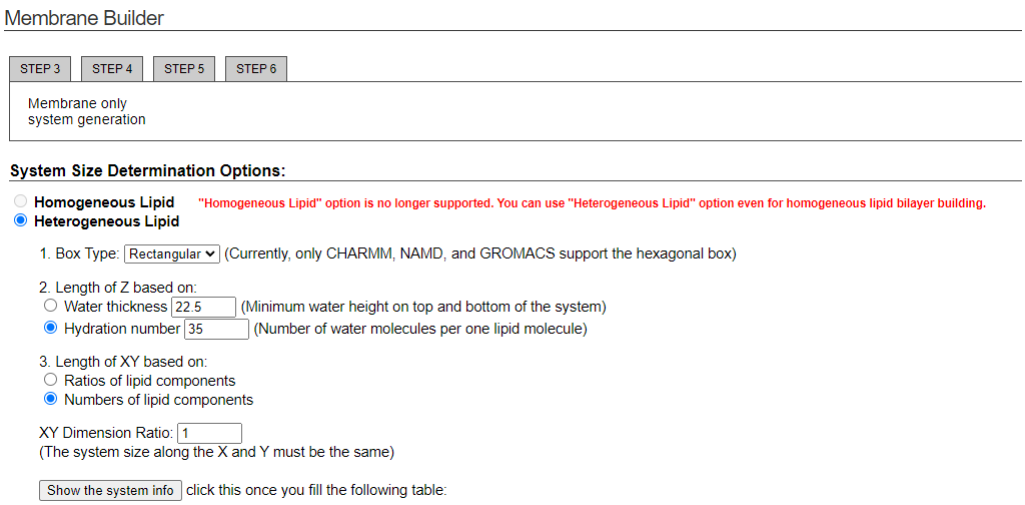
Scroll down until you see the following options:



Select "Membrane Only System" and click "Next Step" on the bottom right hand corner.

Step 2

Change your size determination settings to the following:



These settings will generate a heterogeneous, rectangylar lipid bilayer 35 water molecules for every lipid molecule (this is called the "hydration number"). The z length is based on this number and the xy lengths are based on the number of lipid components, with the x and y dimensions being the same. Scroll down to where you select your lipids.

Use the dropdown menus to select which lipid group you want. Then enter the

number of lipids you want on the upper and lower leaflets of the membrane. We typically do 100 total lipids on each leaflet[1]. In this tutorial we select 100 POPC Lipids.

Now scroll back up and hit "Show the system info". A box should appear:

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	0
Lipid Area	6830	6830
# of Lipids	100	100
Total Area	6830	6830
Average Area	6830.00	
A	82.64	
B	82.64	

Check your numbers to see if you have the right lipid numbers. The total lipid area should be around the same as ours.

Step 3

You can look at the "Determined System Size" section to reconfirm the lipid counts and that the system is reasonably rectangular:

Determined System Size:

Box Type	Rectangle		
Crystal Type	TETRAGONAL		
System Size	A	82.643814	Dimension along the A (X) axis
	B	82.643814	Dimension along the B (Y) axis
	C	70.685335	Dimension along the C (Z) axis
Crystal Angle	Alpha	90.0	Angle between the axis B and C
	Beta	90.0	Angle between the axis A and C
	Gamma	90.0	Angle between the axis A and B
# of Lipids	on Top	100	
	on Bottom	100	
Z Center	0		Center of the system along the Z axis

For the "System Building Options" select "Replacement method". For the "Component Building Options" do not include ions for now [2].

[1] "Leaflet" means a single layer of the bilayer.
[2] See Section 4.1 for more on ions

Step 4

Click "next step" to generate the waterbox. Once this is finished, click "next step" again to assemble the components.

Step 5

Recheck your determined system size. Your numbers should be similar to these:

Determined System Size:

# of Atoms	47800		
Crystal Type	TETRAGONAL		
System Size	A	82.643814	Dimension along the A (X) axis
	B	82.643814	Dimension along the B (Y) axis
	C	76.74	Dimension along the C (Z) axis
Crystal Angle	Alpha	90.0	Angle between the axis B and C
	Beta	90.0	Angle between the axis A and C
	Gamma	90.0	Angle between the axis A and B
# of Lipids	on Top	100	
	on Bottom	100	
# of Water	7000		
Z Center	0.0		Center of the system along the Z axis

Skip the "Force Field Options". Under "Input Generation Options", select "More CHARMM minimization during input generation" and "NAMD". The "Equilibration options" can be skipped.

Step 6

Expect a bit of buffering before you get to step 6. Take note of the "Equilibration Input Notes" and click the "download.tgz" button to download to the computer. Extract the folder in your file manager app. In this tutorial, I rename this folder "POPC100", because our lipid bilayer is made up of 100% POPC. I recommend making a new folder called "MDSims" and putting POPC100 (and all future simulation setups) in here. Now that you have a simulation setup ready, you can copy the files over to Big Red.

3 Using Big Red 200

3.1 Standard File Transfer Protocol

Standard File Transfer Protocol (SFTP) allows you to transfer files between your local computer to Big Red. I use a LINUX/UNIX terminal. For those of you with LINUX computers you can follow along easily in your default terminal. Those with Windows systems would use the windows subsystem for Linux

(WSL) in order to follow along easily[3]. Commands might be slightly different for MACOS users.

3.1.1 Copying Files to Big Red

To begin SFTP, enter the following command into your terminal:

```
$ sftp [USER]@bigred200.uits.iu.edu
```

where USER is your username with Big Red. You should be prompted for your password and two factor authentication. Once successfully connected, your terminal should say "Connected to bigred200.uits.iu.edu." and the beginning of your line should appear as

```
sftp>
```

From here you can navigate through two computers: Big Red and your local computer. For navigating through your local computer, you will put an "l" in front of the commands to indicate that you want the command to execute locally ("l" stands for "local"). For example, you can navigate to your local MDSims directory with the following commands:

```
sftp> lcd [PATH_TO_POPC100]
sftp> lls
```

Now you want to navigate to your user directory within Big Red:

```
sftp> cd /N/Scratch/[USER]
```

For organization's sake, you make want to make a new directory for your simulations:

```
sftp> mkdir SimulationDir
sftp> cd SimulationDir
```

To copy your file from your local machine onto BigRed use the "put" command:

```
sftp> put -r [PATH_TO_POPC100] ./
```

3.1.2 Copying Files to Local Computer

You will need to get files from Big Red once simulation has finished[4]. Begin SFTP using the same login process specified in Section 3.1.1.

```
$ sftp [USER]@bigred200.uits.iu.edu
```

Now make a target directory in your local computer:

```
sftp> lmkdir CompletedSims
sftp> lcd CompletedSims
```

[3]Installation tutorial for WSL: https://youtu.be/HrAsmXy1-78

[4]Skip this part of the tutorial for now until your simulations are done running.

To copy the files from Big Red to your local computer:

```
sftp> get -r [PATH_TO_COMPLETED_SIMULATION_FOLDER] ./
```

3.2 Secure Shell

Secure Shell (SSH) is the terminal that allows you to access and execute simulations on Big Red. To SSH into Big Red:

```
$ ssh [USER]@bigred200.uits.iu.edu
```

Enter your password and two factor authentication. Once successfully logged in your terminal line should begin with

```
[USER]@login2:~>
```

Navigate to the namd folder within your simulation folder.

```
[USER]@login2:~> cd /N/scratch/[USER]/SimulationDir/POPC100/namd
[USER]@login2:~> ls
```

Remember, POPC100 is the folder that we copied into Big Red. The namd folder should already created because of Step 5 of generating a lipid bilayer simulation. The next section will cover making a script that Big Red will run.

3.2.1 Writing NAMD Script[5]

Once you are in the namd directory (located within the simulation directory), create a file that will contain the script that will submit the job:

```
vim script_to_run.sh
```

Here I use vim but you can use whichever file editor you prefer. Once you are editing the file copy the following script:

```
#!/bin/bash

#SBATCH -J [SIMULATION_NAME]
#SBATCH -p general
#SBATCH -o [SIMULATION_NAME]_%j.txt
#SBATCH -e [SIMULATION_NAME]_%j.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=[EMAIL]@iu.edu
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --time=24:00:00
#SBATCH --mem=16G
#SBATCH -A [PROJECT_ID]

#Load any module that your program needs
module load namd

#Run your program
srun namd2 [INPUT_FILE].inp > output.txt
```

[5]NAMD documentation: https://kb.iu.edu/d/awrz

SIMULATION_NAME can be whatever you would like. PROJECT_ID must be provided by Dr. Petrache. INPUT_FILE must be the name of the .inp file that you will run. Note that line 11 specifies the time that the simulation will run (currently set to 24 hours), which you can modify as desired.

3.2.2 Running simulations

For the simulation we created with CHARMM-GUI, there are 6 equilibration steps that must run before the actual simulation begins. The equilibration steps make sure that the lipid box you generate have physiologically accurate levels of pressure, volume, temperature, and energy. You run the equilibration steps by repeating line 19, replacing INPUT_FILE with the different names of your equilibration input files. Repeat line 19 once more with INPUT_FILE as the name of the step 7 production file. Each step will run sequentially, allowing each job to finish before starting another one. Lines 18 onward of your final script should resemble the following:

```
#Run your program
srun namd2 step6.1_equilibration.inp > output_Equilibration_1.txt
srun namd2 step6.2_equilibration.inp > output_Equilibration_2.txt
srun namd2 step6.3_equilibration.inp > output_Equilibration_3.txt
srun namd2 step6.4_equilibration.inp > output_Equilibration_4.txt
srun namd2 step6.5_equilibration.inp > output_Equilibration_5.txt
srun namd2 step6.6_equilibration.inp > output_Equilibration_6.txt
srun namd2 step7_production.inp > output_Production.txt
```

Once you are done making your script, go back to the Big Red 200 terminal. To run the script:

```
[USER]@login2:~> scratch script_to_run.sh
```

where script_to_run.sh is the name of the script you created. The terminal should say "submitted batch job" followed by the job ID. Otherwise you likely have some error in your script.

To see information about the jobs you are running type the following:

```
[USER]@login2:~> squeue -u [USER]
```

You can also do

```
[USER]@login2:~> squeue -j [JOB_ID]
```

to search by job ID. To cancel a job, use

```
[USER]@login2:~> scancel [JOB_ID]
```

You will receive an email when your jobs begin/end. If a job fails, be sure to take a look at the output file to try a troubleshoot why the simulation has failed[6].

[6] Section 4 covers some troubleshooting methods.

3.2.3 Modifying Input Files

You can change aspects of the simulation by modifying the input file. Once you are in the namd folder, begin modifying the desired input file

```
[USER]@login2:~> vim [INPUT_FILE].inp
```

You should see a file that resembles the following:

```
structure               step5_input.psf
coordinates             step5_input.pdb

set temp                303.15;
outputName              step7_production; # base name for output from this run
                                          # NAMD writes two files at the end, final coord and vel
                                          # in the format of first-dyn.coor and first-dyn.vel


set inputname           step6.6_equilibration;
binCoordinates          $inputname.coor;    # coordinates from last run (binary)
binVelocities           $inputname.vel;     # velocities from last run (binary)
extendedSystem          $inputname.xsc;     # cell dimensions from last run (binary)

dcdfreq                 50000;
dcdUnitCell             yes;                # the file will contain unit cell info in the style of
                                            # charmm dcd files. if yes, the dcd files will contain
                                            # unit cell information in the style of charmm DCD files.
xstFreq                 5000;               # XSTFreq: control how often the extended systen configuration
                                            # will be appended to the XST file
outputEnergies          5000;               # 5000 steps = every 10ps
                                            # The number of timesteps between each energy output of NAMD
outputTiming            5000;               # The number of timesteps between each timing output shows
                                            # time per step and time to completion
restartfreq             5000;               # 5000 steps = every 10ps
```

The structure and coordinate files are usually the .psf and .pdb files created in step 5. Here you can change how often you can change how often restart files[7] are created by changing "restartfreq". Note that the units used are number of simulation steps. You can also change the "dcdfreq", which is how often a snapshot will be taken of the simulation, which is used during simulation playback. Other file output frequencies can be modified as described the code's comments.

At the bottom of the file you will find code that looks like the following:

```
# Constant Pressure Control (variable volume)
useGroupPressure        yes;              # use a hydrogen-group based pseudo-molecular viral to calcualte pressure and
                                          # has less fluctuation, is needed for rigid bonds (rigidBonds/SHAKE)
useFlexibleCell         yes;              # yes for anisotropic system like membrane
useConstantRatio        yes;              # keeps the ratio of the unit cell in the x-y plane constant A=B

langevinPiston          on;               # Nose-Hoover Langevin piston pressure control
langevinPistonTarget    1.01325;          # target pressure in bar 1atm = 1.01325bar
langevinPistonPeriod    50.0;             # oscillation period in fs. correspond to pgamma T=50fs=0.05ps
                                          # f=1/T=20.0(pgamma)
langevinPistonDecay     25.0;             # oscillation decay time. smaller value correspons to larger random
                                          # forces and increased coupling to the Langevin temp bath.
                                          # Equall or smaller than piston period
langevinPistonTemp      $temp;            # coupled to heat bath

# Constant Temperature Control
langevin                on;               # langevin dynamics
langevinDamping         1.0;              # damping coefficient of 1/ps (keep low)
langevinTemp            $temp;            # random noise at this level
langevinHydrogen        off;              # don't couple bath to hydrogens
# run
numsteps                500000;           # run stops when this step is reached
run                     500000;           # 1ns
```

[7] See section 3.2.4 for more on restart files.

"numsteps" controls how many simulation steps it takes for you to run. I am unsure of what "run" does **Daniel or Dr. Petrache?**. The other parameters shown can be modified to help keep your simulations from failing as well[8].

3.2.4 Restart Files

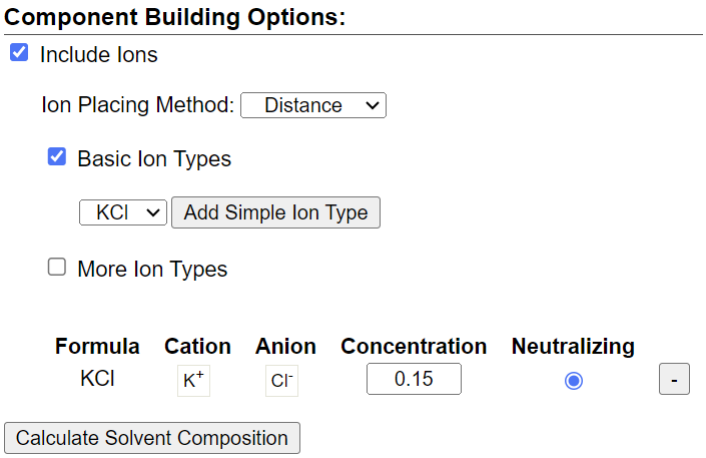
As the simulation runs, restart files will be created at certain intervals that will save your simulation progress if your simulation begins to fail at any point. When a restart file is created, there are three files that are created: .coor, .vel, and .xsc. To restart a simulation with these files, you must modify the input file (see previous section) by changing "inputname" to the name of the restart files that were created.

4 Lipid Mixtures & Simulation Troubleshooting

Our simulations focused on mixtures of POPC, POPS, and POPE lipids. POPC was the "control" lipid, which had no trouble equilibrating. However, certain mixtures of these lipids will fail after certain equilibration steps.

4.1 Mixtures with POPS

POPS has a net charge, and Na+ ions must be included to be balanced out. This must be done when making the simulation with CHARMM-GUI. After step 3 of the simulation, you will see the include ions screen:



Leave the boxes checked as shown in the above image. Now click on the K^+ box in the "Cation" column. Select Alkalai Metals and then select Na^+. Now in the

[8]For instance, modifying the "langevin" parameters might loosen the constraints allow the system to equilibrate. See section 4.

"concentration" column set to 0. Click "Calculate Solvent Composition", and you should see that the Na^+ ion count matches the POPS lipid count.

4.2 Mixtures with POPE

We found that only system with low concentrations of POPE can equilibrate. We think this occurs because high concentrations of POPE will have tendencies towards inverse hexagonal phase, causing the simulation pressure to be too high.

4.3 Other equilibration strategies

It is important to know some strategies for when trying to get simultion setups to equilibrate. One strategy is to run equilibration steps for longer by modifying the input file at the the desired equilibration step[9]. This may be particularly helpful for simulations with ions, as the randomly placed ions will move closer to the POPS lipids they should be paired with.

You can also modify the physiological constants in the input file to try and reach a system that will equilibrate. Be careful to not create simulations that are physiologically unrealistic, but this troubleshooting method is worth considering.

5 Analysis of Results

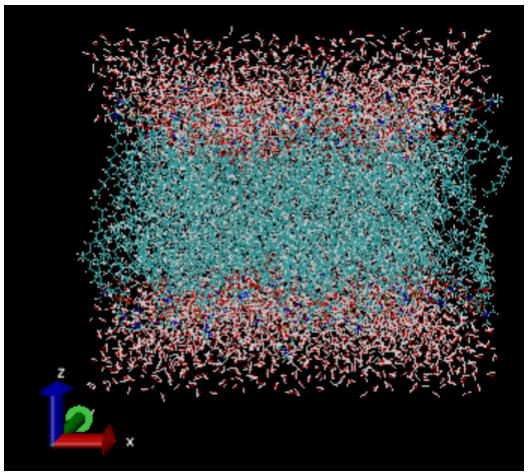
5.1 Basics of Visual Molecular Dynamics

Visual molecular dynamic (VMD) is the software you can use to view and analyze your lipids. Register, download, and install to your computer.

Open the application. You can load in a new molecule by clicking "File" $\rightarrow$ "New Molecule" in the main window. In the pop up window, click "Browse". In your desired simulation folder, go into your namd folder and click on a .pdb file (step 5) to look at a simulation setup. If you already ran a simulation, load the .psf (step 5) and .dcd (step 7) files instead which will display the snapshots of a completed simulation[10]. After loading your files your hydrated bilayer should now be shown in the display window.

[9] see section 3.2.3 for modifying input files.

[10] see section 3.1.2 for copying files from Big Red 200 onto your local computer.



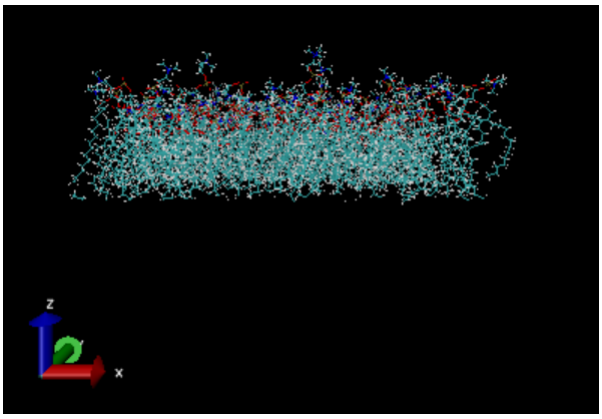
In the main window you can play/pause button to run the simulation snapshots (only if you loaded the .psf/.dcd files). In the "Display" opton on the toolbar window I recommend changing from "perspective" view to "orthographic".

You can change the way that molecules are represented by selecting "Graphics" → "Representations". In the new pop-up window you can hit "Create Rep"/"Delete Rep" to add/delete ways to select atoms in the display. By default, the changes you will be making are applied to all of the molecules, as you can see in the box underneath "selected atoms" indicating "all". Filtering different criteria in this box allows you to apply changes to select molecules[11]. For example, typing

```
resname POPC and z > 0
```

will filter for all of the POPC molecules in the top leaflet of the bilayer. The display should look something like this:

[11] For more on the atom selection language: www.ks.uiuc.edu/Research/vmd/vmd-1.3/ug/node132.html



You can also find a single lipid molecule with "resid":

```
resname POPC and resid 1
```

After you filter for the desired molecules you can change use the "Drawing Method" box to change how the molecules are drawn in the display screen. I recommend using "Bonds" or "CPK" for an intuitive presentation of the bilayer. You can also have one selection show the entire bilayer and another select a single POPC molecule and make it stand out by selecting a different representation (try using "Beads"). If you want the display window to look more full, the "Periodic" tab allows you to show the periodic boundaries of the simulation.

Going back to the main window, I recommend exploring the "Graphics", "Display", and "Mouse" settings to gain a better understanding for your options for using the display window.

5.2 Tracking Coordintes with VMD

Once you have your .psf and .dcd files loaded into VMD, you can a) qualitatively assess your simulation progress by viewing the movie and b) quantitatively look at the trajectories of the molecules. This section covers how to get started on the latter.

In your terminal, go to your simulation directory make a new file which will contain our script analysis:

```
vim coordinateScript.tcl
```

Now copy in the following code:

```
set outfile [open coord.dat w]
set numPhosphorus 100
set numFrames [molinfo top get numframes]

for {set i 0} {$i < $numFrames} {incr i} {
    for {set j 1} {$j <= $numPhosphorus} {incr j} {
```

```
        set sel [atomselect top "resname POPC and name P and z > 0
    and resid $j"]
        $sel frame $i
        set coords "$i $j [join [$sel get {x y z}] " " ]"
        puts $outfile $coords
    }
}

close $outfile
```

This script uses a language called tcl[12]. This script in particular selects the phosphorus molecules in the top leaflet POPC lipids and obtains their x, y, z coordinates for every frame. The output file (named "coord.dat") is formatted as

Frame Resid x-coordinate y-coordinate z-coordinate

with no delimiter. Now in your VMD main window, click "Extensions"→ "TK-Console". A new window should pop up. To source/run your script type the following:

```
(VMD)1% source [PATH_TO_coordinateScript.tcl]
```

A new file should be created called "coord.dat", which contains the coordinate data. You can plot this data using your plotting software of choice.

5.3 Using Gnuplot to Analyze Output Files

Gnuplot[13] is a plotting software that is convinient to use when working directly in your terminal. To install Gnuplot:

```
$ sudo apt install gnuplot
```

Now you should be able to enter "gnuplot" in your terminal and the terminal prompt should change to the following:

```
gnuplot>
```

Make sure to have your finished simulation files copied onto your local computer[14]. You can copy certain sections of the output file you created into smaller files. For instance, if you only want to look at the energies you can create a file that looks like this:

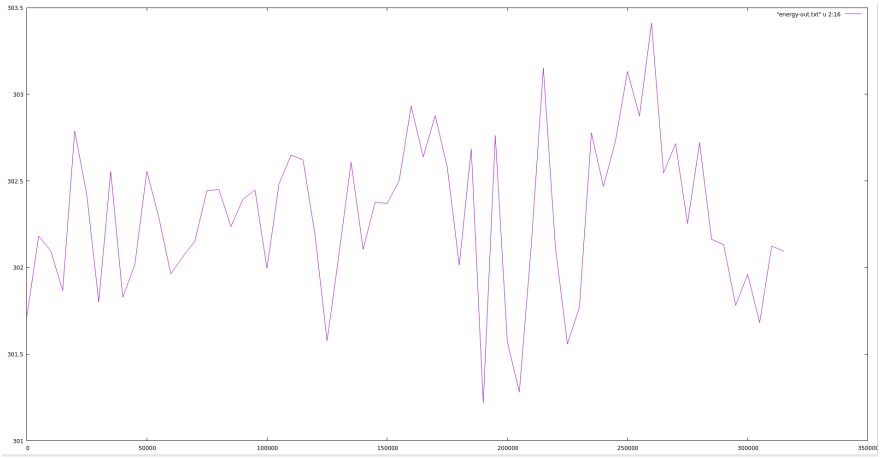
[12] For more about tcl: www.tcl.tk/about/language.html
[13] Gnuplot website: www.gnuplot.info
[14] see section 3.1.2 for copying files from Big Red 200 onto your local computer.

```
Info: ENERGY OUTPUT STEPS    5000
Info: CROSSTERM ENERGY INCLUDED IN DIHEDRAL
Info: INCONSISTENCY IN FAST TABLE ENERGY VS FORCE: 0.000325096 AT 11.9556
Info: INCONSISTENCY IN SCOR TABLE ENERGY VS FORCE: 0.000324844 AT 11.9556
Info: INCONSISTENCY IN VDWA TABLE ENERGY VS FORCE: 0.0040507 AT 0.251946
Info: INCONSISTENCY IN VDWB TABLE ENERGY VS FORCE: 0.00150189 AT 0.251946
ENERGY:       0      2783.5064     16091.6262     10478.0803       116.8978
 0.0000     31776.3228         -36798.5608       301.7089    -68574.8836    -36532.
3136    458184.7739        62.9719        62.3136
ENERGY:    5000      2887.2786     16135.1221     10462.6063       111.2157
 0.0000     31926.0642         -36636.9374       303.1307    -68563.0016    -36377.
8897    454903.9255         2.8159         2.8159
ENERGY:   10000      2860.7359     15912.1664     10429.9324       116.0769
 0.0000     31745.9583         -37160.7679       301.4206    -68906.7262    -36900.
1481    454643.2097         0.8792         0.9254
ENERGY:   15000      2841.9576     16064.7984     10459.0943       121.7596
 0.0000     32099.8101         -36690.9236       304.7804    -68790.7337    -36424.
0872    454886.2043        -0.7910        -0.7901
ENERGY:   20000      2868.5179     16074.7559     10582.9698       114.3388
```

Name this file "energy_out.txt". This contains columns of energy data that we can plot in gnuplot. Open gnuplot (make sure you are in the folder that contains "energy_out.txt"). To plot the energies:

```
plot "energy_out.txt" u 2:16 w l
```

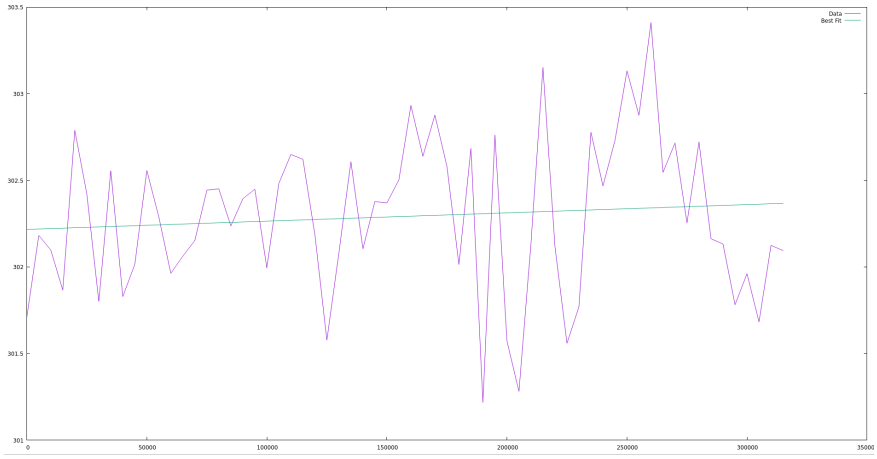
This plots the 2nd (steps) and 16th (energy) columns for the x and y axes, respectively. "u" stands for "using". "w l" stands for "with lines", indicating to draw a line graph.



You can also create a line of best fit.

```
    f(x) = m*x + b
    fit f(x) 'energy_out.txt' u 2:16 via m,b
    plot 'energy_out.txt' u 2:16 w l title 'Data',  f(x) w l title
    'Best Fit'
```

Line 1 defines a linear function, and line 2 does the fitting. Line 3 plots the original data and best fit lines with the appropriate labels:



5.4 SIMtoEXP

Coming soon to a lab near you