**BIOS 10603 Lab 3**

**Water simulation, analysis, and Ubiquitin system minimization**

**NAME:**

Please insert all answers and figures into this Word document, and then save the completed lab as “Lastname\_Lab3.docx”. When you have completed the lab, submit it as an attachment to the Lab-3 assignment link on Canvas. The report is due by midnight the day after your lab period.

**Important points are colored in blue. Questions to be answered are in green.**

**Lab 3 Sections:**

1. Minimization and equilibration of water
2. Preparation and minimization of ubiquitin protein
3. Analysis of water simulation

**Objectives:** In this week’s lab, we will start our first round of molecular dynamics simulation to model water behavior and analyze the data. **The main question that we will try to answer is: “can we create a realistic model of water that has its basic, real-life properties?”.** We will then continue by building a system of the ubiquitin protein in water and submit it for minimization and equilibration. Building a protein simulation system is a little more complicated than our water modeling and requires more steps.

This ubiquitin simulation is very closely aligned with what you will be doing with your final project protein. Because this process will span multiple labs, we will include a summary table (like the one below), so you can keep track of the work you have done in previous weeks.

In order to keep track of the dozens of files through lab sessions, **please create a separate folder for ubiquitin**, just like what we did for water in the previous lab.

**Table below describes the steps that we will run in this lab:**

|  |  |  |  |
| --- | --- | --- | --- |
| **STEP** | | **DESCRIPTION** | **Lab** |
| 0: Preparation | Build PSF | Adding hydrogens to protein structure | Lab 3 |
| Solvation | Putting the protein in a water box |
| Ionization | Adding salt at a specified concentration |
| Restraining | Adding restraints to the protein backbone |
| 1: Energy Minimization | | Using NAMD to move the protein into an energetically favorable conformation before kinetic energy is introduced to the system. Then heating up the system to the desired temperature. |

**From this lab on, after each lab session, we will also perform simulation checks on the weekend to make sure everyone is on track. These will be worth 10 pts; further details about when/where these will occur will be provided.**

**Part 1. Minimization and equilibration of the water box**

There are three major components to an MD simulation:

1. **Energy Minimization**
2. **Temperature Adjustment and Equilibration**
3. **Production Simulation**

**What is energy minimization in MD?**

During energy minimization, the atoms in the system are moved **artificially** to minimize the potential energy as defined by the **force field** of our simulation. At a global minimum, we would expect the net force on each atom to completely vanish or become very small. Since the minimization is conducted without heat, **there is no kinetic energy involved.** Minimizing the system at the beginning allows us to start with a stable state that would prevent running into thermodynamically unlikely and energetically unfavorable situations when we later introduce kinetic energy into the system.

Interacting with the Midway supercomputer in this course is the same as in the previous BIOS10602 class. Please note that we will always be working in the **Midway3 scratch space**. If you do not remember how to interact with Midway, or if you are using an updated version of Windows 10 and would like to run a native Unix subsystem, review the **Midway and Unix Review** document at the end of this lab. Additionally, you may choose to interact with the Midway3 GUI via ThinLinc. Instructions for using ThinLinc can also be found at the end of the lab.

To begin your water simulation, you need the following files:

1. Your solvated water box PSF.
2. Your solvated and **centered** water box PDB.
3. A NAMD configuration file H2O-mini-heatup-equil-NPT1.conf.
4. An appropriate force field (we will be using [CHARMM36](https://pubmed.ncbi.nlm.nih.gov/23832629/)). In today’s lab, we need two files, par\_all36\_prot.prm and toppar\_water\_ions\_esmael.str.
5. A job-submit sbatch script, job-submit.sbatch.

Have all of the water box files you created last week saved in the same folder on your computer. Download the other needed files from Canvas and put them in the same folder.

We are going to edit the NAMD configuration file and job-submit script before copying these files to your midway account. **Before moving on, make sure you understand what each section of this configuration file is doing. These should have already been covered in the lecture and lab introduction video but ask clarifying questions now!** Using any text editor, change the following values to customize this file based on your own system:

* Under the **Input Files** section, replace inputfilename with your actual input files (solvated and centered!) for both **coordinates** and **structure**.
* Under **Parameter Files**, replace the two instances of the parameter file with the two parameter force field files, toppar\_water\_ions\_esmael.str and par\_all36\_prot.prm.
* Under **Periodic Boundary Conditions (PBC)**, replace XX.XX with the correct **xyz** dimensions of your solvated box. Please do this accurately to 1 decimal value. **Run the command pbc get in the Tk Console to confirm.**
* Under **Execution Script**, replace all text in the curly braces {} in the header line of the for loop with the proper syntax and values. A description of what we are trying to do (**not code**) is provided there for you. To see what the starting temperature is, go to the top of this file.

**Q1.** After making the above edits, paste a screenshot showing all **changes** to your configuration file below.

**Q2.** According to your knowledge from the lectures, Why are periodic boundaries important? How might changing the values of the pbc conditions in the configuration file alter your simulation?

Now edit the job-submit script with the following changes

* Change the job name **name\_me** to a name suitable to your job. **Do not use a long name, and do not include any spaces**
* Change the output name to something that represents the system. For example, for modeling a water box at 300K, the name waterbox-300K is sufficient.
* Change the time **00:00:00** to ask for 1 hour (real runtime should be about 30 minutes)
* Change the nodes **0** to ask for 4 nodes
* Make sure the account says **BIOS10603**
* Change the name of the config file to your **config** file name and the log file name to the prefix of your config file ending with **.log extension instead of .conf.**

**Q3.** Paste your job-submit script below.

Just like you have a folder for each simulation on your computer, we recommend that for each simulation, you have a folder in your **Midway3 scratch directory**. Save the two scripts file we just edited and transfer all 6 files from above into this folder. (If you forgot how to transfer files, recall the terminal command scp.) Then go to your Midway3 scratch water-box directory and submit the job. Useful commands to submit the job are **sbatch** (to submit) and **squeue** (-u CNET, to check status).

**Your script may contain carriage returns (Windows line endings). If when you submit the job, there is a DOS carriage return error, please use the following command to fix that:**

**tr -d ‘\r’ <infile> outfile**

Example: tr -d ‘\r’ <job-submit.sbatch> job-submit-2.sbatch

Note, the infile and outfile \*must\* be different names.

**If your job completes immediately, check your \*sh.e\* , \*sh.o\*, and \*log files for any potential errors. Please notify the TA if anything went wrong.**

IMPORTANT: If your job does not start to run immediately, please keep checking its status until tomorrow morning. If the job goes to completion, it should create a couple of files, including name.log, name.dcd, name.coor, name.xst, name.xsc. Also, you can check successful completion by opening the .log file using the following UNIX command: **tail filename.log**. You should see the line **End of program** printed at the very bottom, which indicates that your job script was successful. If your job is not completed in two days, please notify Dr. Haddadian or your TA.

**After the system minimization, we will incrementally adjust the temperature of our water box from 270K to 300K and then run a production simulation for 1ns**. We will then use the obtained data to see how well our water box replicates various properties of physical water.

**Part 2. Preparation and minimization of the ubiquitin protein system**

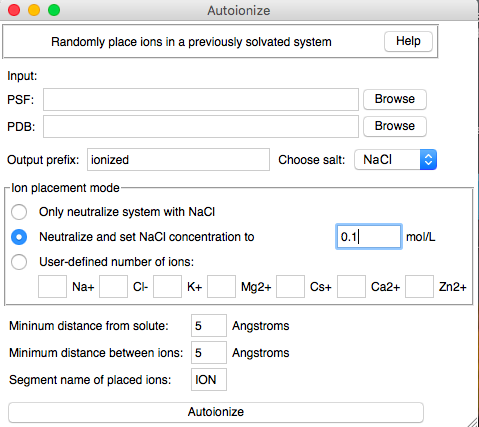
While we are waiting for the water simulation to finish, we will build our first protein system using ubiquitin (PDB code 1UBQ) and submit it for simulation. The preparation step involves **inserting this molecule into a water box, ionizing the water box to mimic physiological conditions, restraining our protein, and finally minimizing the energy of the whole system**. Create a folder called 1UBQ-system in your computer and change the VMD working directory to this folder (to ensure that all of the new files we create today are saved in this folder).

We’re going to start with the same steps we did with the water box. As a refresher, the steps are:

1. Delete all of the water files and load the protein into VMD
2. Create a PSF of the protein: Extensions > Modeling > Automatic PSF Builder
3. Put protein in a solvation box: Extensions > Modeling > Add Solvation Box.
   1. Load the PSF and corresponding PDB file you just created (load the PSF into VMD first by using the File --> New Molecule command, then add the PDB file by choosing the Load Data into Molecule option.
   2. Unselect the “water box only” option
   3. Choose “rotate to minimize volume” and “use molecule dimensions”
   4. Pad protein with 10Å of water on each side (x, y, and z)
   5. Change the output name to something recognizable (1UBQ-solvate)

**Q4.** What does the “rotate to minimize volume” option do? How would you expect the output to change if this option is not selected?

**Q5.**

1. Is padding with 5Å of water on each side sufficient? Why or why not? (Hint: Think about periodic images).
2. Why should we not pad the protein with more water? (Hint: try padding by 20Å on each side and see how many water molecules you would get).
3. Ionize the solvation box: It is important to ionize protein boxes because, inside a cell, proteins are surrounded by ions. Also, NAMD calculates electrostatic interactions between atoms using a grid and requires that the net charge of the system be zero for an accurate calculation. Try the following steps
   1. Extensions > Modeling > Add Ions
   2. Input PSF and PDB from the water solvation above
   3. Choose “Neutralize and set NaCl concentration to” and enter 0.1 mol/L
   4. Click “Autoionize”

**Q6.** The above process created a new PSF and PDB file. Check the end of the PDB file.

1. How many Cl and Na atoms are added?
2. How does VMD’s autoionization affect the charge of the overall system?
3. What was the total number of atoms in the system before and after ionizing?
4. Given this, how does VMD’s autoionization tool work?
5. What is the total charge of the ubiquitin protein?
6. **Center your solvated and ionized box.** Remember, the commands are:

set sel [atomselect top all]

set coords [measure center $sel]

$sel moveby [vecinvert $coords]

$sel writepdb 1UBQ-solvated-ionized-centered.pdb

1. **Restrain the protein backbone:** During minimization, we want the water and ions to move and adjust their positions with respect to the protein with minimal changes to the protein backbone; therefore, we restrain the protein backbone using simple harmonic springs. This can be done with the following Tcl commands.

Delete all of the structures in VMD and just load the 1UBQ-solvated-ionized-centered.pdb structure and then type:

set sel [atomselect top "protein and backbone"]

$sel set beta 10

set sel [atomselect top all]

$sel writepdb 1UBQ-restrain.pdb

**Q7.** Open the 1UBQ-restrain.pdb file using a text editor and check the second column from the right.

**A.** Explain the results of the above commands.

**B.** What does the value of 10 stand for in our situation?

**C.** What is the unit of the restraints used in NAMD (Hint: use google).

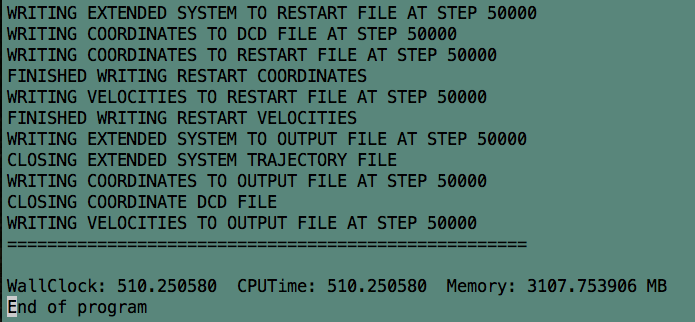
**D.** What differences would you expect if we set the restraints to 100 as opposed to 10?

1. Measure the size of your centered system; using either **measure** **minmax** or **pbc get** commands. Then use these values to prepare the 1UBQ-mini.conf file as you did for the water box (the file is available in the Lab-3 folder on Canvas). In the configuration file there is a new section where you should put the name of your restrained PDB file. **There are a total of 9 lines to be changed.**
2. Create a new directory in your midway3 scratch for the **1ubq simulation** and transfer your files there. Make sure you transfer:
   1. 1UBQ-solvated-ionized-centered.pdb
   2. 1UBQ-restrain.pdb
   3. Your solvated, ionized PSF file (ionized.psf)
   4. The NAMD configuration file
   5. Appropriate force fields (par\_all36\_prot.prm and toppar\_water\_ions\_esmael.str); you can copy this from your directory on Midway that contains your water box simulation if you don’t feel like transferring them again.
   6. A job submit script; You can copy this from your directory on Midway that contains your water simulation and change it accordingly. Remember to change the configuration file name to the correct .conf file for this simulation. Ask for 1 hour and 2 nodes.
3. After transferring the files, submit your protein minimization job

**As before, if your job completes immediately, and you can’t figure out why, ask Prof. Haddadian or your TA ASAP.**

**Q8.** What is energy minimization in the context of molecular dynamics simulations? What might happen if the energy were not minimized?

**Part 3. Analysis of water simulation**

Once your water simulation completes, you will see a summary at the bottom of the log file like the one shown below. The command **taillogfile\_name.log** can be used to display only the end of the log file.

Although your WallClock, CPUTime, and Memory values may be different, you should see the line “End of program” printed, which indicates that your simulation was successful.

If you do not see this line, please let Prof. Haddadian or your TA know, as you will have to resubmit the simulation.

Looking at the output from the job, you should see a variety of files with different suffixes. The contents of the files are explained below. Feel free to open any of them in emacs or vim to better understand their contents. The file that ends in .dcd is in “binary’ format.

|  |  |
| --- | --- |
| **File Extension** | **Explanation of File Contents** |
| .dcd | Contains (in binary format) the coordinates for every atom in your system during each timestep over the course of the minimization trajectory. |
| .log | Contains information about the progress of the simulation (that is written to as the simulation proceeds in a text format). Useful for simulation analysis. |
| .coor | Final coordinates of the system’s atoms at the end of simulation |
| .vel | Final velocities of the system’s atoms at the end of simulation |
| .xst, .xsc | Contains a record of the periodic cell parameters over the course of the simulation |
| .restart.coor (.vel, .xsc) | Files that can be used to restart simulations if the simulation crashes |
| .old | Old copies of files that can be used to restart simulations if the simulation crashes |

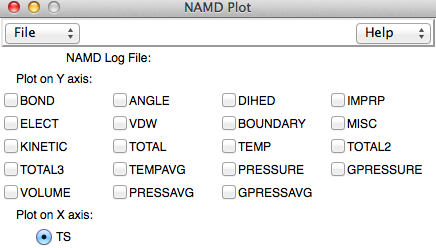
If everything is in order, use the **scp** command to copy the following files to your computer:

1. Your logfile (ends in **.log**)
2. Your trajectory file (ends in **.dcd**)

To view the trajectory:

1. **load the water box PSF file into VMD**
2. **load the DCD** **file** **into** the PSF file. (Select the PSF file from VMD Main and go to File --> Load Data into Molecule).

The system trajectory will appear in the graphics window – **Call your TA if it does not!**

We are now going to examine some properties of our trajectory using **NAMD** **Plot**. Go to Extensions > Analysis > NAMD Plot to open a window like the one on the right. These NAMD Plots extract information from the .log files of the trajectory. Load the correct .log file for the water simulation into the NAMD Plot window via File > NAMD logfile.

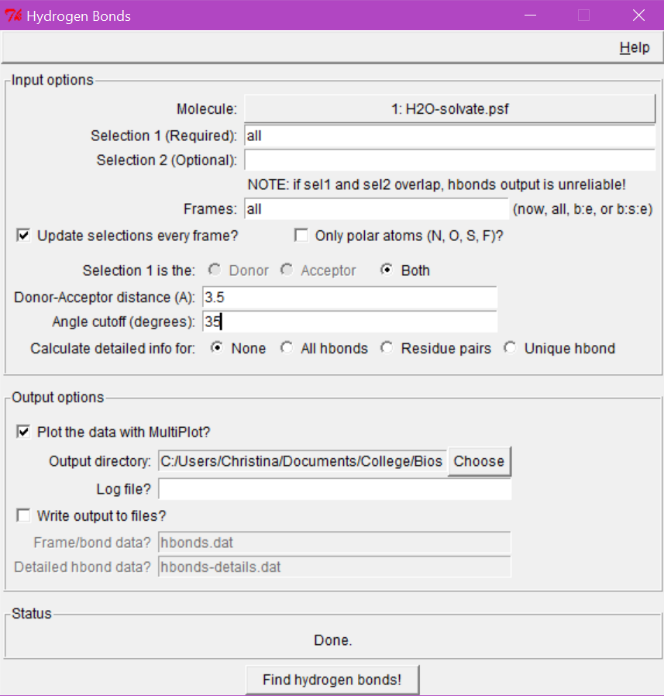
For all plots that you submit, you are required to provide an explanation of the shape of the graph in the context of its simulation. Explain **all sections** of the plot and **why** they may be increasing, decreasing, or constant. **Explain how different plots relate to one another (if needed**). The best way to describe the plots is to follow the trends by watching the simulations. If the graph generated by VMD is not clear; you can download the data and use R to plot (see below).

**Q9.** Produce the following NAMD Plots **separately**:

1. **Electrostatic energy**
2. **Kinetic Energy**
3. **Bond Energy**

Attach screenshots of the three plots and explain the features of each. NAMD plot is not a good plotting software, and if the data range is large, it might not show on the graph. Therefore, you will write out the data after you have generated the NAMD Plot by going to File > Export to ASCII matrix, and then use R to generate the graph with this data file as input.

**Q10.** If you recall last week’s lab, we found that the H-O-H angle was not consistent with the TIP3P model. **In the last frame of your water box simulation**, choose **5** water molecules and measure their H-O-H angles. Report the residues, the angles, and the average of the four angles. **Is this value more consistent with the TIP3P water model? Explain.** Note: the best way to measure the angle is to go to the end frame of the trajectory **(move the playing button on the VMD Main window all the way to the right)** and pick one water molecule to show at a time. This can be achieved by typing random water residue numbers in the VMD “representation” window, e.g., resid 100.

**Part 4. Analysis of ice melting simulation**

We can also plot the number of hydrogen bonds in the system as a function of time by going to Extensions > Analysis > Hydrogen Bonds. A window will appear like the one on the right. Make sure you have the right molecule selected. Use Selection 1 to choose which atoms to find hydrogen bonds between. Since all of our molecules are water, type “all”. Select “all” frames; then **change Donor-Acceptor distance to 3.5 and Angle cutoff to 35**. Make sure plot data is selected. You will also have the option to save the data to a separate file, which is essential for you to generate your plot in R.

**Now delete your water simulation and load the .psf and .dcd files for the ice melting simulation that are provided on Canvas: Ice-melt-100-grad2-01.psf and Ice-melt-100-grad2-01.dcd.**

**Q11.** Using the play button at the bottom of VMD Main, view the trajectory a few times. You can alter the step or speed to make the trajectory go faster or slower. Describe what you see happening (Hint: use licorice as presentation and rotate the box in a way that the y-axis points out of the screen toward you; this way, you can see the ice’s hexagonal lattice).

**Q12.** Plot the number of hydrogen bonds in your simulation and in the melting simulation (using R).

**A.** How do you expect the number of hydrogen bonds change in the process of melting ice? Why is this the case?

**B.** Does the number of hydrogen bonds in your simulation correspond to your expectation? Explain.

**C.** Insert images of these plots.

**Q13.** Make a plot (using R) of the volumes of both systems (separately) by exporting the data from NAMD Plot and explain their differences. Insert images of both plots (Note: make sure that you use the correct log file for each plotting).

The article on Canvas (filename “supercooling-water”) discusses the phenomena of cooling water below freezing temperatures without crystal lattice formation at length. A good summary of this paper can also be found on the Smithsonian [website](https://www.smithsonianmag.com/science-nature/at-what-temperature-does-water-freeze-1120813/).

More information can also be found from

http://www.ks.uiuc.edu/Training/CaseStudies/pdfs/water-1.pdf

**Q14.** Do some research on the process in which water turns into ice. Also consider the number of interacting sites required for the crystallization of ice.

**A.** Report your findings.

**B. You can look at the melting simulation in reverse to get a sense for the water freezing process.** However, this is not a good simulation approach to observe the water freezing. What are some general MD limitations that we might encounter if we try to model the water freezing process? (Hint: Consider the number of interaction sites and probabilities and the subsequent system sizes).

A picture containing text

Description automatically generated**RCC Thinlinc**

The RCC allows Midway users to log to a virtual machine for Midway that runs with the Scientific Linux distribution. To use that, you can access Midway via

https://midway3.rcc.uchicago.edu

There you can login with your Midway account (just as you do to login to my.uchicago). After login you will see a window similar to one on the right opens on your desktop:

In the bottom left corner, you can open the terminal (either by clicking the icon with a terminal screen, or going to Applications -->Terminal), which has the exactly the same directory hierarchy you use when you *ssh* to midway.

In the terminal, you can access the VMD GUI by using the following commands:

module load vmd/1.9.4 #This loads the vmd module to your session

vmd #This starts vmd

On the flipside, if you ever want to use the Tk console directly on Midway (without the GUI), say, when you ssh to Midway, you can use the following commands:

module load vmd #Same as before

vmd -dispdev none #Dispdev stands for display device; you are telling VMD to open # without a display device

This way you can load your protein without any graphics representation and run analysis scripts directly on Midway. This would greatly expedite the analysis!

### Programs for Windows Users

**This lab requires access to a Unix command prompt. Mac users may access the ‘Terminal’ application under Applications > Utilities > Terminal. Windows 10 users may install a Linux subsystem, see** <https://docs.microsoft.com/en-us/windows/wsl/install-win10>**.**

Instead of connecting to Midway through the ssh command in Terminal, Windows users must install the program PuTTY (<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>). After connecting to Midway through PuTTY, you will be presented with the same Unix command prompt that you can access from Mac computers. Information on connecting from Windows is also provided on the RCC website (<https://rcc.uchicago.edu/docs/connecting/index.html#connecting>).

PuTTY allows you to log onto Midway to submit analyses and modify files that have already been uploaded. However, to transfer files to and from Midway (scp from lab), you will need to download a secure file transfer client program WinSCP (<http://winscp.net/eng/index.php>). Once again, the RCC website provides details about connecting to Midway through WinSCP (<https://rcc.uchicago.edu/docs/data-transfer/index.html#winscp-gui-for-windows-clients>).

Alternatively, as a more permanent solution, you could download the program Cygwin (<https://cygwin.com/install.html>), which gives you a Unix command prompt on your local Windows machine.

