Instituto Politécnico Nacional

Escuela Superior de Cómputo

Bioinformatics

Practice 5 - NAMD

Group: 3CV9

Student:
Ramos Diaz Enrique

Professor:
Rosas Trigueros Jorge Luis

Development Date:
October 28th 2020

Due Date:
November 11th 2020

1 Theoretical Framework

NAMD, recipient of a 2002 Gordon Bell Award and a 2012 Sidney Fernbach Award, is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. Based on Charm++ parallel objects, NAMD scales to hundreds of cores for typical simulations and beyond 500,000 cores for the largest simulations. NAMD uses the popular molecular graphics program VMD for simulation setup and trajectory analysis, but is also file-compatible with AMBER, CHARMM, and X-PLOR. NAMD is distributed free of charge with source code [1].

A force field is a mathematical expression of the potential which atoms in the system experience. CHARMM, X-PLOR, AMBER, and GROMACS are four types of force fields, and NAMD is able to use all of them. The parameter file defines bond strengths, equilibrium lengths [2].

In order to make a psf file for a system, a force field topology is necessary. This file contains information on atom types, charges, and how the atoms are connected in a molecule. A pdb file contains only coordinates, but not the connectivity information [2]. Figure 1 shows a detailed flowchart representing the types of files required by NAMD to perform a simulation.

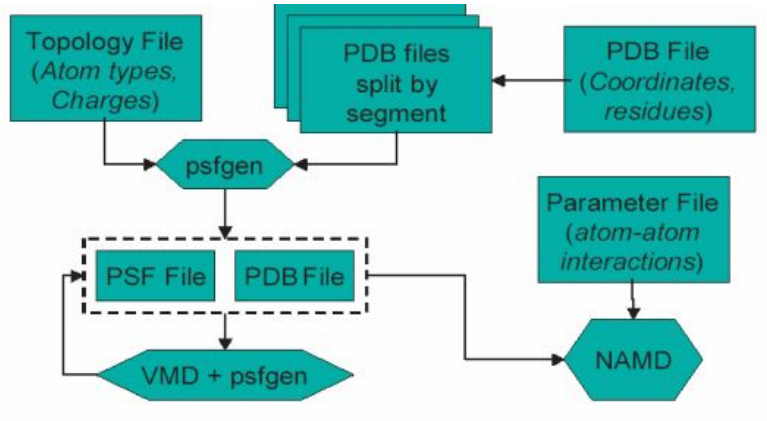


Figure 1: Flowchart indicating the role of files as used by VMD, NAMD, and psfgen. Image taken from [2].

1.1 Minimization and equilibration

Minimization and equilibration differ from each other by the nature in which they implement the molecular dynamics force field. Energy minimization involves searching the energy landscape of the molecule for a local minimum, i.e., place in which the molecule is relaxed, by systematically varying the positions of atoms and calculating the energy. Equilibration involves molecular dynamics, whereby Newton Second Law is solved for each atom in the system to dictate its trajectory. Achievement of equilibrium is judged by how well velocities, pressure, etc. are distributed in the system over a given amount of time [2].

2 Material and Equipment

- VMD - Visual Molecular Dynamics Software.
- NAMD Software (version 2.14 Win64-CUDA) [3].
- NAMD Tutorial Windows Version [2].
- NAMD Tutorial's Files [4]:
  - **Ubiquitin PDB file: 1-1-build/1UBQ.pdb**
  - **NAMD Simulation's Configuration File: 1-3-box/ubq_wb_eq.conf**
- Additive Force Field Parameter Files version 3.6 [5]:
  - **toppar_c36_jul20/par_all36_prot.prm**
  - **toppar_c36_jul20/param19.inp**
- MD Analyses Tcl scripts:
  - **Root Mean Square Desviation: rmsd.tcl**
  - **Radius of Gyration: rg.tcl**
  - **Solvent Accessible Surface Area: sasa.tcl**
  - **Root Mean Square Fluctuation: rmsf.tcl**
- Text editor

3 Practice Development

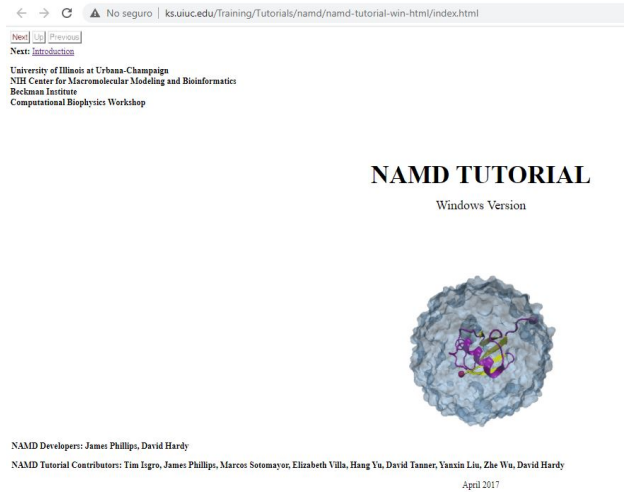
3.1 Preparation

According to the NAMD Tutorial, it's needed to download their files first, available in [4]. For this and the next practices, the Windows Version [2] will be used (see Figure 2a).

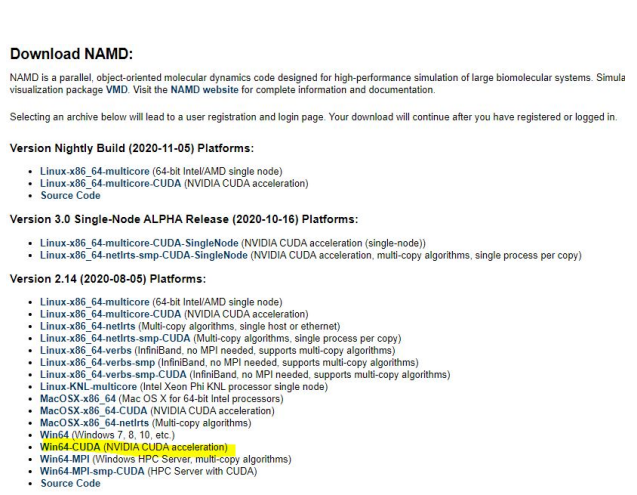
A work directory must be created, where all the downloaded and generated files will be stored. There are only two necessary tutorial's files for this practice, enlisted in the Material and Equipment section.

Then, NAMD has to be downloaded through the link provided by [3] and extracted. Figure 2a shows the available versions of NAMD, being the selected one for this and the next practices the 2.14 version for Windows 64 bits with CUDA acceleration (download this only if your computer has CUDA installed and a compatible GPU).

NAMD is just an executable, so it shouldn't be installed. The unzipping extracts a folder with all the necessary files for execution, that will be used later.



(a) NAMD Tutorial - Windows Version cover. Image taken from [2].

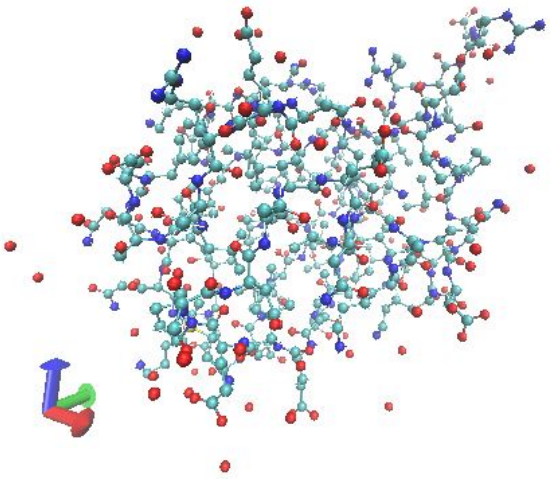


(b) NAMD available downloads from it's official webpage [3].

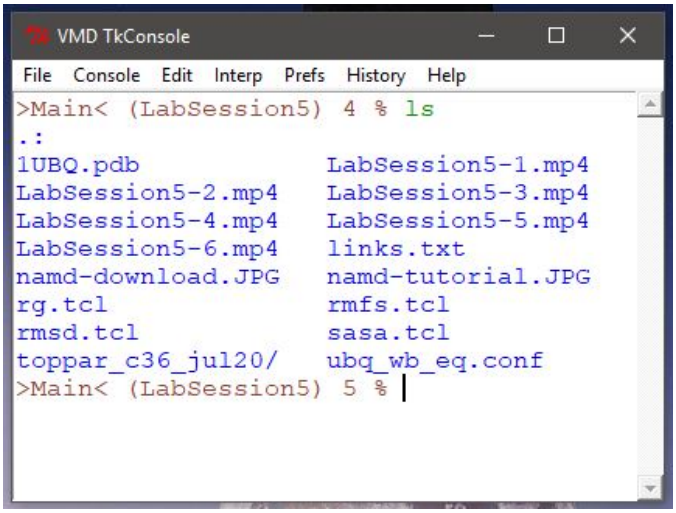
Figure 2: NAMD resources.

3.2 Generating a Protein Structure File (PSF)

Start VMD. Load the **1UBQ.pdb** molecule provided by [4] (see Material and Equipment section) either though the graphical interface using the Molecular File Browsing window (*File* > *New Molecule*), or through the TkConsole by typing the *mol new* command. Figure 3a shows the molecule loaded. Despite which method was used to load the Ubiquitin molecule, open the TkConsole choosing *Extensions* > *Tk Console*, and go to the work directory previously created or set, so in that way all the generated and required files stay together for better organization. Figure 3b shows the files on the work directory through the TkConsole.



(a) Loaded 1UBQ model with CPK style.



(b) Work directory's files enlisted in the TkConsole.

Figure 3: Files required.

After these, in the VMD Main Window select *Extensions* > *Modeling* > *Automatic PSF Builder* (see Figure 4).

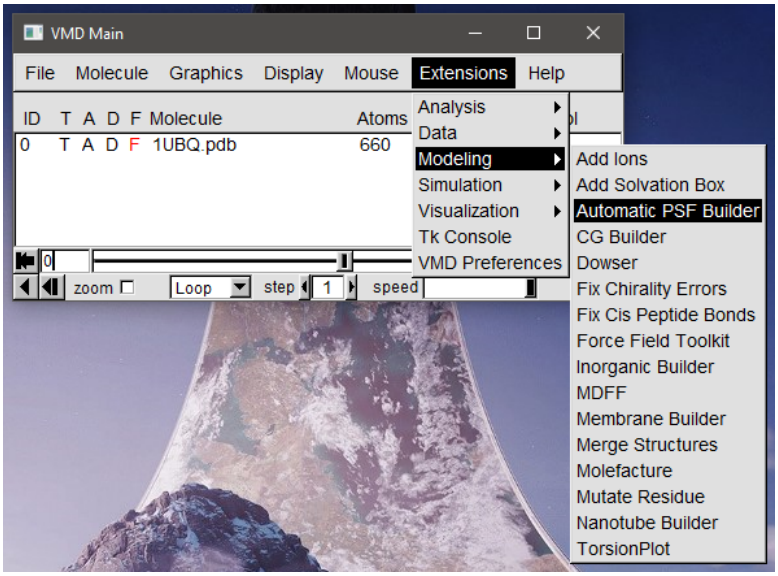


Figure 4: Open Automatic PSF Builder window.

The molecule's atoms and bonds in the PDB files are just for graphical representations, they aren't actually there, so for example if a simulation wants to be applied to this molecule, it won't work. For this purpose, a PSF file of the molecule its needed.

When the AutoPSF window pop-up, the first step is to select input and output files: the input molecule is the 1UBQ (coordinates file), a topology input file must be selected, and this will be the first option of the list: top_all36_prot.rtf, and let the output file by default. Figure 5 shows how the step 1 section should look.

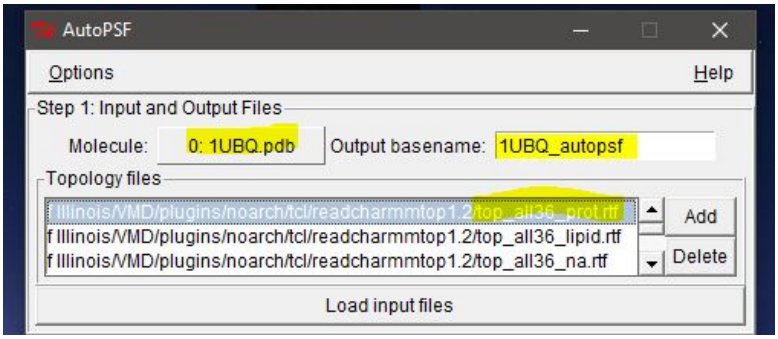


Figure 5: AutoPSF Step 1: Select a molecule and a topology.

Click on the *Load Input Files*. Then, on the second step of the AutoPSF window select the Protein option (see Figure 6) and click on the *Guess and split chains using current selections*.

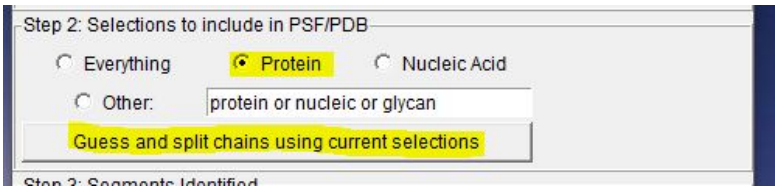


Figure 6: AutoPSF Step 2: Set a 1UBQ's selection.

Figure 7 shows the result of the protein's guessed chains on the third step of the AutoPSF window. Click on the *Create chains* button to continue.

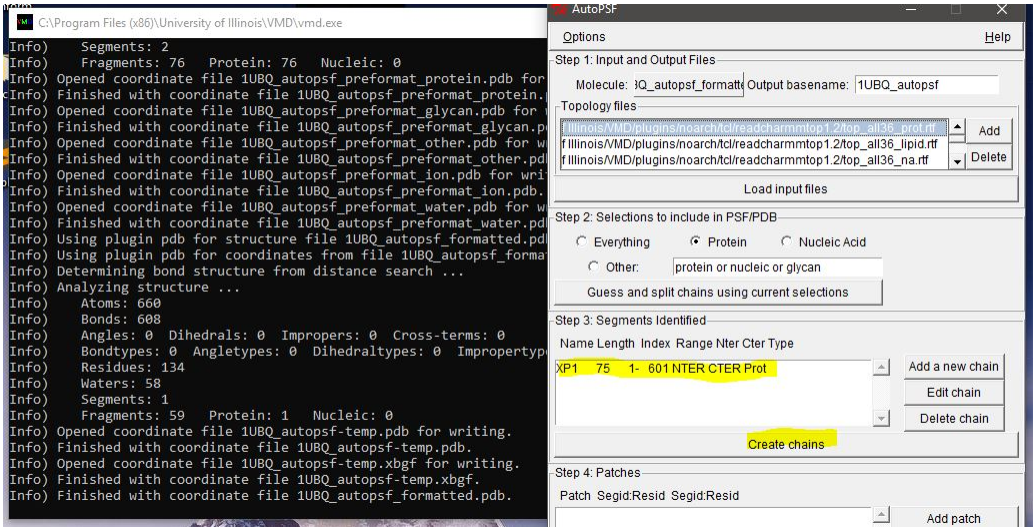


Figure 7: AutoPSF Step 3: 1UBQ guessed chains.

If everything goes well, a window like the shown in Figure 8a should be displayed. Click on *Accept* button and another window like the shown in Figure 8b will appear, showing that two new files were generated: **1UBQ_autopsf.pdb** and **1UBQ_autopsf.psf** (Figure 9 shows these).

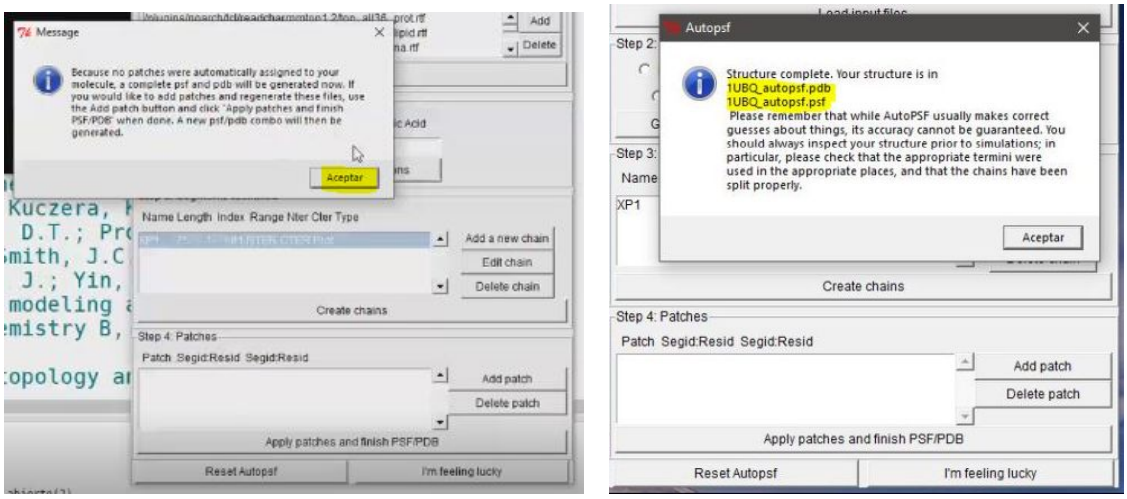


Figure 8: Alerts telling that the PSF and PDB files were generated successfully.

Figure 9: Generated PSF and PDB Files by AutoPSF on the work directory.

The VMD Main window now should looks like Figure 10, and a comparative between the PDB 1UBQ and the new PSF generated 1UBQ is shown in Figure 11. As it can be seen, now the hydrogen atoms appears and there was only the protein selection (water was removed). Now, this can be called a "complete" 1UBQ protein model.

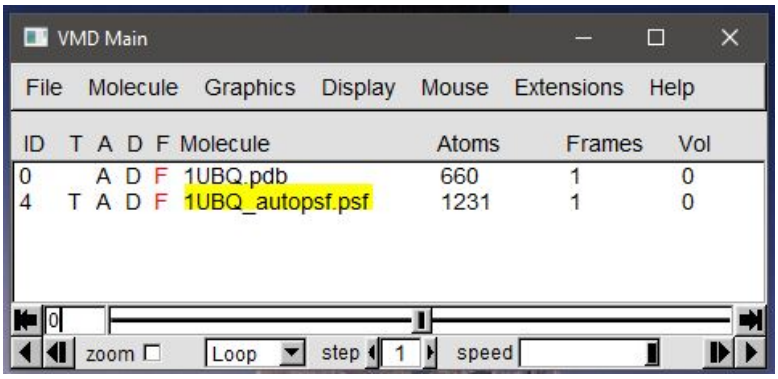
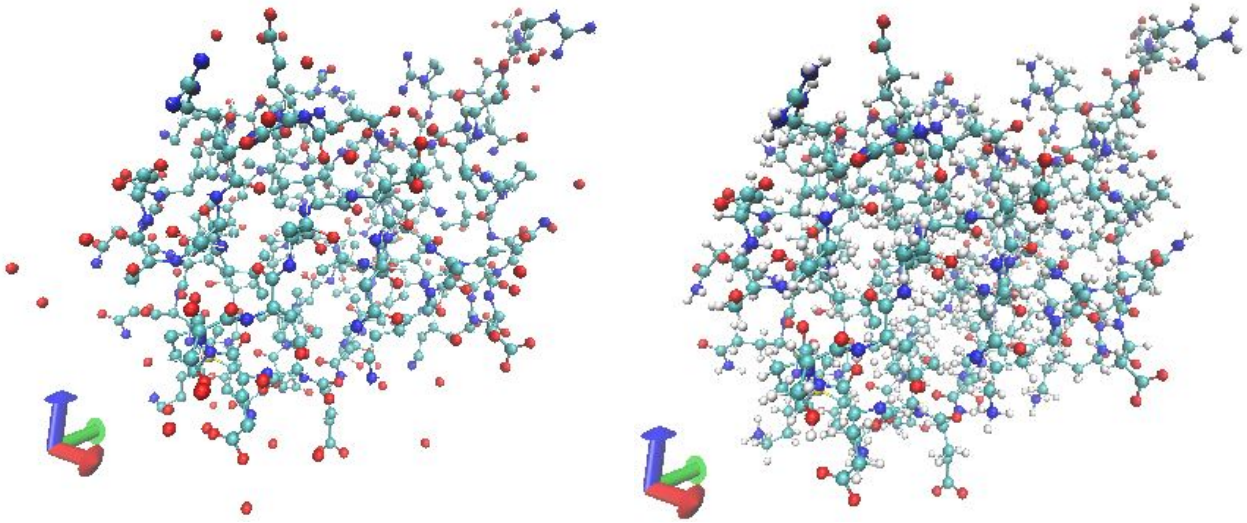


Figure 10: Generated PSF Files loaded in the VMD Main window.

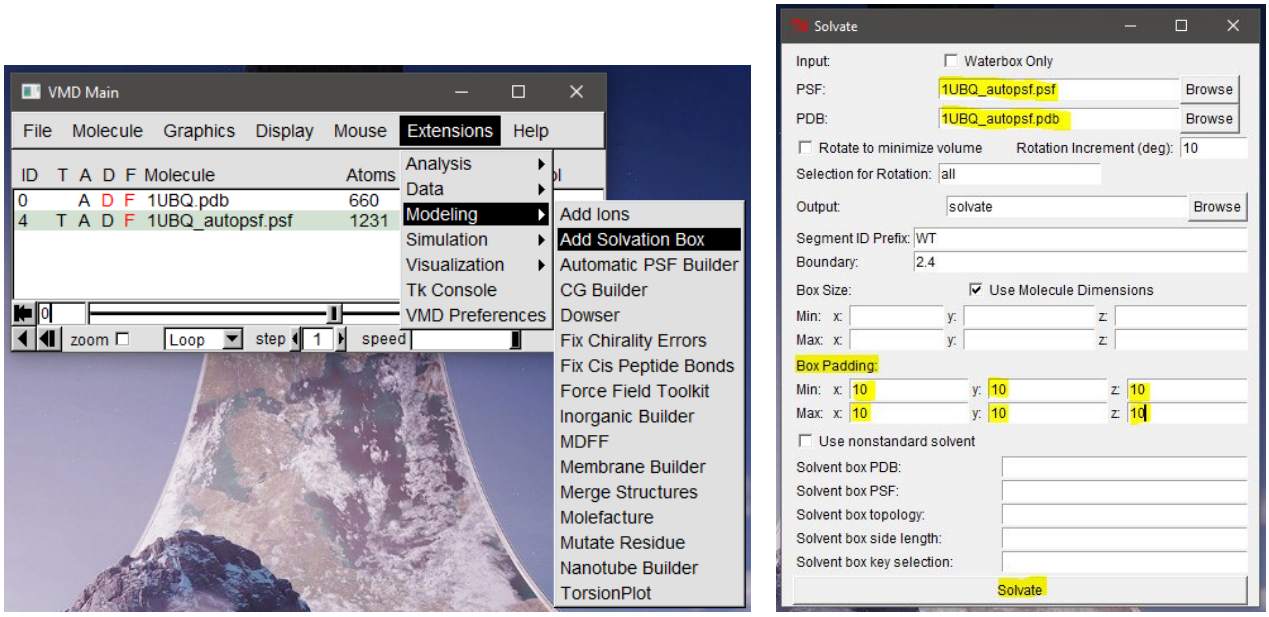


(a) Original 1UBQ PDB model. (b) New 1UBQ PSF generated model.

Figure 11: Comparative between the 1UBQ PDB model and the generated PSF model.

3.3 Solvating a Protein: Ubiquitin in a Water Box

For the required NAMD simulation of this practice, the 1UBQ should be put inside a water box. To accomplish this, in the VMD Main window select *Extensions* > *Modeling* > *Add Solvation Box* (see Figure 12a).



(a) Open Solvate window.

(b) Adding a water box in the Solvate window.

Figure 12: Putting the 1UBQ inside a water box.

A Solvate window will be displayed. Let everything by default and just add a box padding of 10 armstrongs on each axis, as shown in Figure 12b. Click the *Solvate* button to generate a two files: **solvate.pdb** and **solvate.psf**, with the complete 1UBQ inside a water box (see Figure 13). Figure 14 shows how the model looks like.

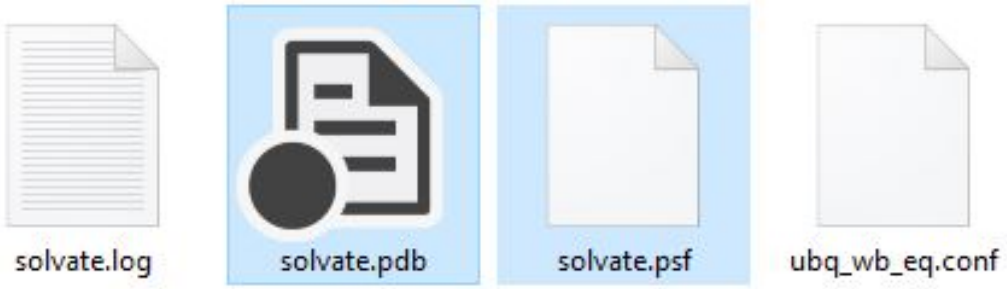


Figure 13: Generated PSF and PDB Files by Solvate on the work directory.

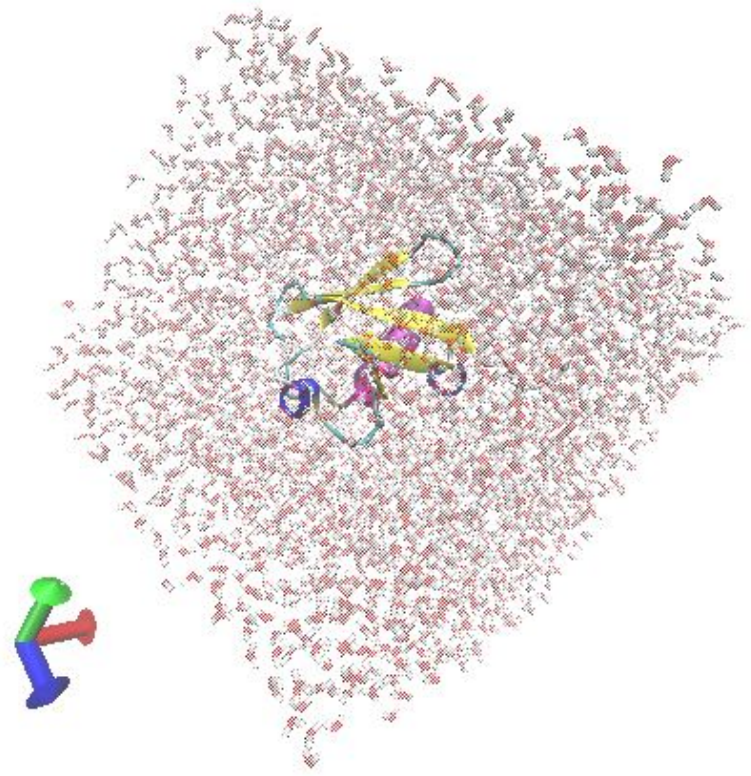


Figure 14: "Complete" 1UBQ inside a water box. Protein has NewCartoon style and Secondary Structure color, while the water box has transparent Licorice style.

3.4 Setting parameters on a NAMD Simulation's Configuration File

The 1UBQ inside a water box PSF file generated in the Solvating a Protein: Ubiquitin in a Water Box subsection is the one that will be simulated on NAMD, in order to examine a minimization and equilibration process in it.

But before that, a configuration file has to be created. Fortunately, the NAMD Tutorial's Files [4] provide one, so it's not necessary to create one from scratch.

However, some modifications must be applied to this configuration file. On the work directory (created on the Preparation subsection), locate the file called **ubq_wb_eq.conf** and open it with any text editor (see Material and Equipment section).

Go to the **ADJUSTABLE PARAMETERS** section and change the *structure* and *coordinates* parameters for the water box files created in the previous section, like shown below:

```
1 ###################################################################
2 ## ADJUSTABLE PARAMETERS                                         ##
3 ###################################################################
4 structure          solvate.psf
5 coordinates        solvate.pdb
```

Second, go to the **SIMULATION PARAMETERS** section, then to the **Input** subsection and here add both Additive Force Field Parameter files [5] (see Material and Equipment section) as *parameters* like shown below:

```
#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################
# Input
paraTypeCharmm       on
parameters          ./toppar_c36_jul20/par_all36_prot.prm
parameters          ./toppar_c36_jul20/param19.inp
temperature         $temperature
```

Finally, on the same section, go to the **Periodic Boundary Conditions** where the *cellBasisVector1*, *cellBasisVector2*, *cellBasisVector3* and *cellOrigin* parameters have to be changed.

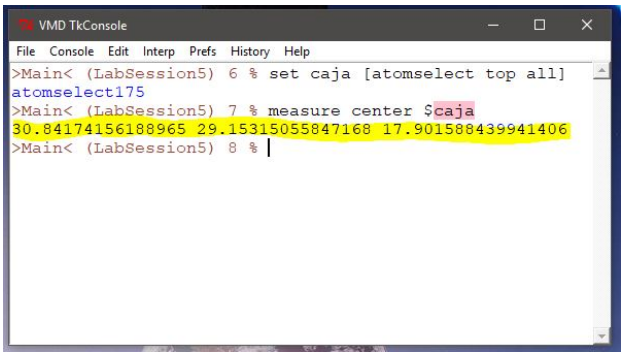
To get the new correct values of these, in VMD open the TkConsole and type the next commands (be sure that the **solvate.psf** molecule is on top):

```
set caja [atomselect top all]
measure center $caja
measure minmax $caja
```

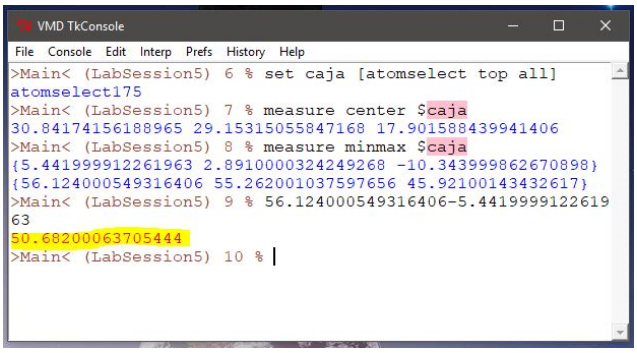
The value of the *center* command will be the new value for the *cellOrigin* parameter. For the other ones it's needed to do the difference between the x, y and z minimum and maximum values (obtained by typing the *minmax* command). Figure 15 shows the calculated values of interest in the TkConsole after typing their respective commands.

The **Periodic Boundary Conditions** subsection on the configuration file should look like this after modifying the previously mentioned parameters:

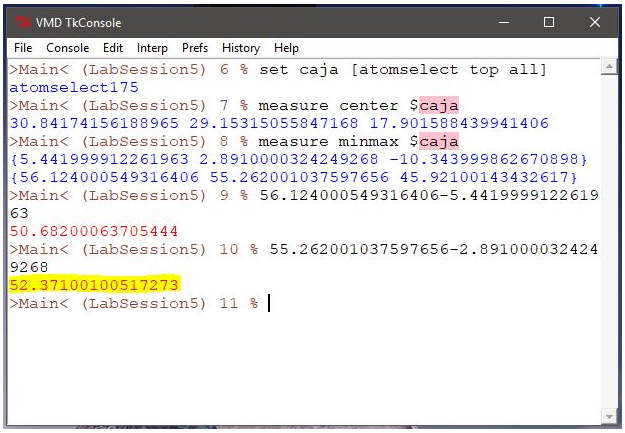
```
# Periodic Boundary Conditions
cellBasisVector1    50.68200063705444    0.    0.0
cellBasisVector2     0.0  52.37100100517273   0.0
cellBasisVector3     0.0    0   56.26500129699707
cellOrigin          30.84174156188965   29.15315055847168
 ↪  17.901588439941406
```



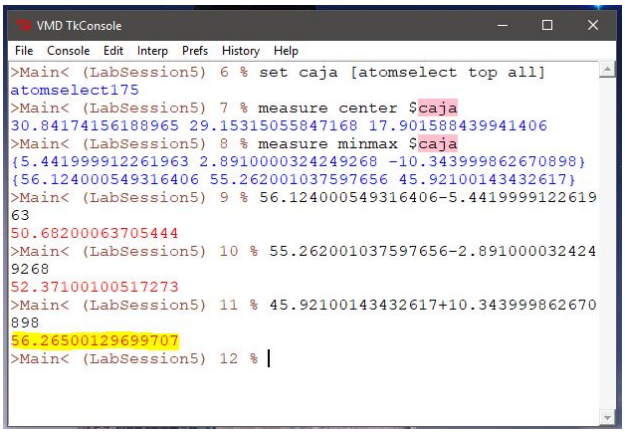
(a) cellOrigin parameter value.



(b) cellBasisVector1 parameter value.



(c) cellBasisVector2 parameter value.



(d) cellBasisVector3 parameter value.

Figure 15: Calculated values on the TkConsole required for the Periodic Boundary Conditions parameters in the configuration file.

With these modifications on the configuration file **ubq_wb_eq.conf**, the simulation en NAMD now can be run.

3.5 Using NAMD to perform a Simulation

Everything ready to run the simulation on NAMD. Copy the **namd2.exe** and **tcl85t.dll** (this one only if you are using Windows) files from the NAMD extracted directory (see Preparation subsection) to the work directory (created on the Generating a Protein Structure File (PSF) subsection).

Open a terminal or a CMD window on the work directory and type the following command to start the simulation of minimization and equilibration on a 1UBQ protein solved inside a water box:

```
1       namd2.exe ubq_wb_eq.conf > ubq_wb_eq6nov2020.log
```

3.6 Simulation Results

Onche finished the simulation, many files were generated (see them in Figure 16). The important ones are two: **ubq_wb_eq6nov2020.log**, which is a logger where all the simulation process and messages were written (Figures 17 and 18 shows some extracts from this file), and the DCD file **ubq_wb_eq.dcd**, which is the trajectory file for this simulation.

Figure 16: Generated files after the simulation.

```
Charm++> No provisioning arguments specified. Running with a single PE.
         Use +auto-provision to fully subscribe resources or +p1 to silence this message.
Charm++: standalone mode (not using charmrun)
Charm++> Running in Multicore mode: 1 threads (PEs)
Charm++> Using recursive bisection (scheme 3) for topology aware partitions
Charm++ warning> fences and atomic operations not available in native assembly
Converse/Charm++ Commit ID: v6.10.2-0-g7bf00fa-namd-charm-6.10.2-build-2020-Aug-05-556
Charm++> Disabling isomalloc because mmap() does not work.
CharmLB> Load balancer assumes all CPUs are same.
Charm++> Running on 1 hosts (1 sockets x 4 cores x 1 PUs = 4-way SMP)
Charm++> cpu topology info is gathered in 0.009 seconds.
Info: Built with CUDA version 10010
Did not find +devices i,j,k,... argument, using all
Pe 0 physical rank 0 binding to CUDA device 0 on OUTERHEAVEN: 'GeForce GTX 1060 6GB'  Mem: 6144MB  Rev: 6.1  PCI: 0:1:0
Info: NAMD 2.14 for Win64-multicore-CUDA
Info:
Info: Please visit http://www.ks.uiuc.edu/Research/namd/
Info: for updates, documentation, and support information.
Info:
Info: Please cite Phillips et al., J. Chem. Phys. 153:044130 (2020) doi:10.1063/5.0014475
Info: in all publications reporting results obtained with NAMD.
Info:
Info: Based on Charm++/Converse 61002 for multicore-win64
Info: Built Wed, Aug 05, 2020 9:54:41 PM by jim on europa
Info: Running on 1 processors, 1 nodes, 1 physical nodes.
Info: CPU topology information available.
Info: Charm++/Converse parallel runtime startup completed at 0.7 s
CkLoopLib is used in SMP with simple dynamic scheduling (converse-level notification)
Info: 257.68 MB of memory in use based on GetProcessMemoryInfo
Info: Configuration file is ubq_wb_eq.conf
Info: Working in the current directory C:\Users\YaKerTaker\Desktop\Bioinformatics\LabSession5
TCL: Suspending until startup complete.
Info: SIMULATION PARAMETERS:
Info: TIMESTEP               2
Info: NUMBER OF STEPS        0
Info: STEPS PER CYCLE        10
Info: PERIODIC CELL BASIS 1  50.682 0 0
Info: PERIODIC CELL BASIS 2  0 52.371 0
Info: PERIODIC CELL BASIS 3  0 0 56.265
Info: PERIODIC CELL CENTER   30.8417 29.1532 17.9016
Info: WRAPPING ALL CLUSTERS AROUND PERIODIC BOUNDARIES ON OUTPUT.
Info: LOAD BALANCER  Centralized
Info: LOAD BALANCING STRATEGY  New Load Balancers -- DEFAULT
```

Figure 17: Start of the simulation log file.

```
GPRESSURE: 2300 -42.3499 197.317 181.596 119.677 -258.618 108.439 333.664 1.8947 -285.526
PRESSAVG: 2300 102.562 19.8736 66.1194 19.8733 79.299 122.813 66.1191 122.813 -81.3109
GPRESSAVG: 2300 99.4544 11.0932 61.5569 21.49 76.7095 123.559 55.3279 124.673 -74.5552
ENERGY:    2300       253.6070       696.3911       747.2045        47.1750        -47847.0832      3419.7627        0.0000        0.0000      8389.0047        -34293.9383
302.3809    -42682.9430    -34269.2456       304.0342          -230.0042      -195.4981    132649.7439        33.5169        33.8696

PRESSURE: 2400 5.91656 -206.692 232.781 -206.69 -578.941 -3.83694 232.778 -3.83895 -749.809
GPRESSURE: 2400 86.1454 -168.288 85.8631 -86.5544 -563.028 -5.53155 232.798 159.495 -758.458
PRESSAVG: 2400 30.3149 89.5369 16.1813 89.5372 56.1991 104.668 16.1811 104.668 -57.6068
GPRESSAVG: 2400 39.6077 94.2987 22.6402 89.0538 55.8862 101.481 24.6555 103.424 -65.8316
ENERGY:    2400       252.0474       682.6858       727.9416        48.3942        -47860.9041      3392.7876        0.0000        0.0000      8408.7372        -34348.3103
303.0921    -42757.0475    -34324.1139       301.9347          -440.9444      -411.7800    132989.1309        9.6357         9.8874

PRESSURE: 2500 -88.548 277.421 -123.562 277.421 293.15 42.6881 -123.561 42.6868 -598.337
GPRESSURE: 2500 5.95338 8.94384 -359.025 320.005 331.246 76.5103 -19.571 173.484 -631.543
PRESSAVG: 2500 -1.48043 -8.57936 -10.6307 -8.57946 -7.14153 169.106 -10.6306 169.106 -45.3038
GPRESSAVG: 2500 -1.63862 -2.30798 -13.1133 -13.1265 -11.1607 177.128 -10.0902 165.627 -40.4623
ENERGY:    2500       232.6461       680.2688       726.4188        49.6896        -47857.3741      3400.7882        0.0000        0.0000      8448.1453        -34319.4173
304.5126    -42767.5626    -34293.4130       302.0770          -131.2451       -98.1146    133014.1341       -17.9753       -17.7539

WRITING EXTENDED SYSTEM TO RESTART FILE AT STEP 2500
WRITING COORDINATES TO DCD FILE ubq_wb_eq.dcd AT STEP 2500
WRITING COORDINATES TO RESTART FILE AT STEP 2500
FINISHED WRITING RESTART COORDINATES
The last position output (seq=2500) takes 0.004 seconds, 435.176 MB of memory in use
WRITING VELOCITIES TO RESTART FILE AT STEP 2500
FINISHED WRITING RESTART VELOCITIES
The last velocity output (seq=2500) takes 0.002 seconds, 435.176 MB of memory in use
PRESSURE: 2600 309.899 19.9346 -57.0707 19.9406 175.736 -92.5038 -57.0688 -92.5041 391.655
GPRESSURE: 2600 371.085 226.213 -26.4502 108.83 111.129 -116.635 -33.0768 -69.1099 495.13
PRESSAVG: 2600 164.453 96.8392 -198.169 96.8393 9.02847 -44.4903 -198.17 -44.4903 -151.433
GPRESSAVG: 2600 158.223 89.5697 -188.301 98.0751 14.0851 -52.1284 -204.218 -40.6644 -149.322
ENERGY:    2600       209.9507       687.0159       727.3568        42.9439        -47962.6619      3586.3583        0.0000        0.0000      8440.4761        -34268.5601
304.2361    -42709.0362    -34246.0524       303.2616          292.4297       325.7815    133126.4132         7.3497         7.6620

WRITING EXTENDED SYSTEM TO OUTPUT FILE AT STEP 2600
CLOSING EXTENDED SYSTEM TRAJECTORY FILE
WRITING COORDINATES TO OUTPUT FILE AT STEP 2600
CLOSING COORDINATE DCD FILE ubq_wb_eq.dcd
WRITING VELOCITIES TO OUTPUT FILE AT STEP 2600
=====================================================

WallClock: 38.554001  CPUTime: 38.555000  Memory: 435.175781 MB
```

Figure 18: End of the simulation log file.

To watch the resulted trajectory file of the simulation, on the VMD Main window select *File* > *New Molecule* and load the **ubq_wb_eq.dcd** file into the **solvate.psf** (as shown in Figure 19).

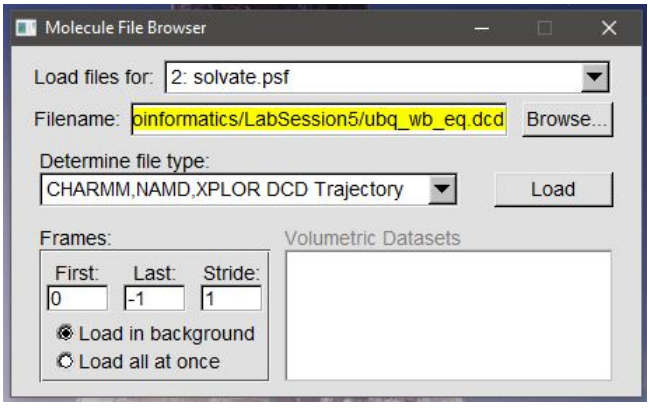
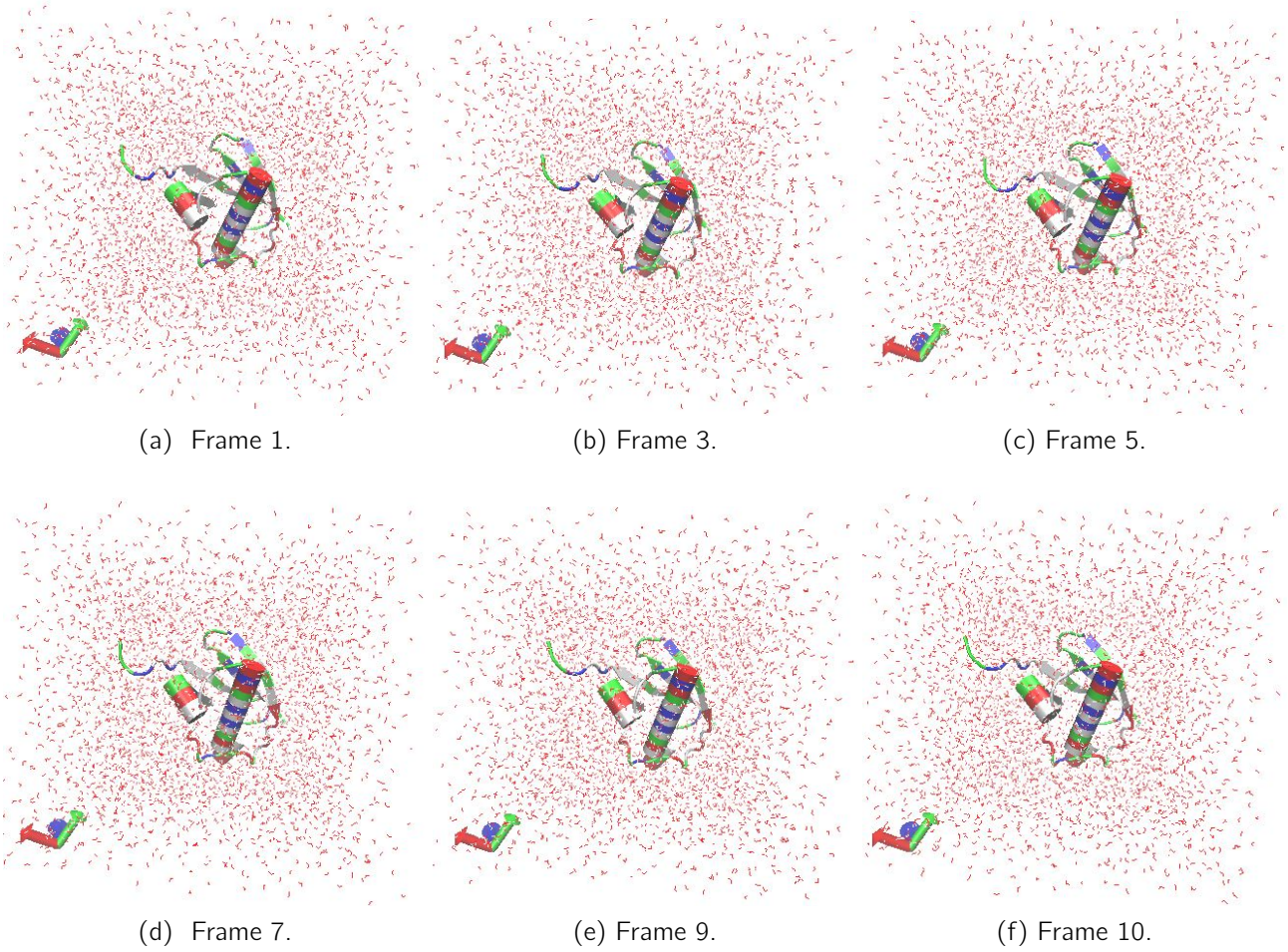


Figure 19: Load simulation's trajectory file into the 1UBQ inside a water box.

Figure 20 shows the behavior of the protein in different times of the simulation's resulted trajectory (DCD file), using the slider bar on the Animation Tools in the VMD's main menu.



(a) Frame 1. (b) Frame 3. (c) Frame 5.

(d) Frame 7. (e) Frame 9. (f) Frame 10.

Figure 20: Every second frame (10 in total) of the trajectory file as result of the minimization and equilibration simulation.

3.7 Molecular Dynamics Analyses

Now let's perform four MD Analyses on the generated simulation and watch their respective results:

- **Root Mean Square Desviation calculation**
- **Radius of Gyration calculation**
- **Solvent Accessible Surface Area calculation**
- **Root Mean Square Fluctuation calculation**

This analyses' files will generate another .dat files with the results of each one. Be sure that the **solvate.psf** molecule (1UBQ inside a water box) with the **ubq_wb_eq.dcd** file loaded is on top in VMD, and to put this Tcl analyses files on the work directory for better access and organization.

3.7.1 Root Mean Square Desviation - RMSD

Use the Tcl file called **rmsd.tcl**, whose content is the following:

```
# script for Root Mean Square Desviation (RMSD) calculation
set outfile [open rmsd.dat w]
set sel [atomselect top "protein"]
set frame0 [atomselect top "protein" frame 0]
set nf [molinfo top get numframes]

for { set i 1 } { $i <= $nf } { incr i } {
   $sel frame $i
   $sel move [measure fit $sel $frame0]
   puts $outfile "[measure rmsd $sel $frame0]"
}
close $outfile
```

Open a TkConsole (be sure to stay on the work directory) and run the Tcl calculation file typing the next command:

```
source rmsd.tcl
```

The content of the generated **rmsd.dat** file with the Rg calculation result of the present simulation is the following. See that there's a value that keeps growing until it stabilizes into a constant one:

```
0.938098669052124
1.1234450340270996
1.141625165939331
1.1098982095718384
1.1507549285888672
1.133998155593872
1.1411759853363037
1.1435587406158447
1.1523386240005493
1.1739864349365234
1.1739864349365234
```

3.7.2 Radius of Gyration - Rg

Use the Tcl file called **rg.tcl**, whose content is the following:

```
# script for Radius of gyration (Rg) calculation
set outfile [open rg.dat w]
```

```
set sel [atomselect top "protein"]
set frame0 [atomselect top "protein" frame 0]
set nf [molinfo top get numframes]

for { set i 1 } { $i <= $nf } { incr i } {
   $sel frame $i
   $sel move [measure fit $sel $frame0]
   puts $outfile "[measure rgyr $sel]"
}
close $outfile
```

In the TkConsole run the Tcl file typing:

```
source rg.tcl
```

The content of the generated **rg.dat** file with the RMSD calculation result of the present simulation is the following. The result should be constant, meaning that this molecule doesn't spin so much:

```
11.676355361938477
11.658084869384766
11.603926658630371
11.577113151550293
11.568012237548828
11.6662015914917
11.614256858825684
11.584607124328613
11.663848876953125
11.651816368103027
11.651816368103027
```

3.7.3 Solvent Accessible Square Fluctuation - SASA

Use the Tcl file called **sasa.tcl**, whose content is the following:

```
# script for Solvent Accessible Surface Area (SASA) calculation
set outfile [open sasa.dat w]
set sel [atomselect top "protein"]
set frame0 [atomselect top "protein" frame 0]
set nf [molinfo top get numframes]

for { set i 1 } { $i <= $nf } { incr i } {
   $sel frame $i
   $sel move [measure fit $sel $frame0]
```

```
10    puts $outfile "[measure sasa 1.4 $sel]"
11 }
12 close $outfile
```

In the TkConsole run the Tcl file typing:

```
1       source sasa.tcl
```

The content of the generated **sasa.dat** file with the SASA calculation result of the present simulation is the following:

```
1  4918.37890625
2  4914.15625
3  4841.53857421875
4  4829.70068359375
5  4816.4287109375
6  4835.0341796875
7  4809.95703125
8  4790.064453125
9  4856.11474609375
10 4848.04248046875
11 4848.04248046875
```

3.7.4 Root Mean Square Fluctuation - RMSF

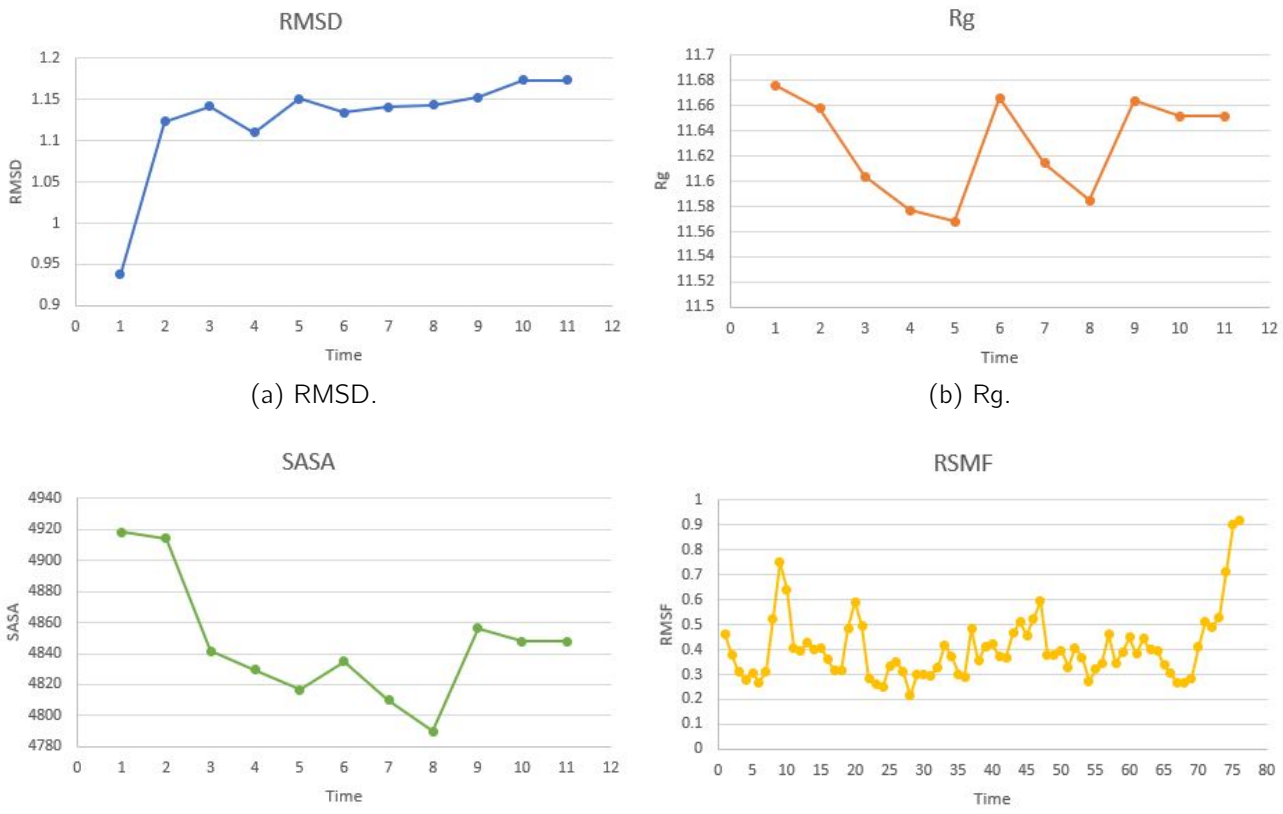
Use the Tcl file called **rmsf.tcl**, whose content is the following:

```
1  # script for Root mean square fluctuation (RMSF) calculation
2  set outfile [open rmsf.dat w]
3  set sel [atomselect top "name CA"]
4  set rmsf "[measure rmsf $sel first 0 last -1 step 1]"
5
6  for {set i 0} {$i < [$sel num]} {incr i} {
7     puts $outfile "[expr {$i+1}] [lindex $rmsf $i]"
8  }
9  close $outfile
```

In the TkConsole run the Tcl file typing:

```
1       source rmsf.tcl
```

Due the fact that the generated **rmsf.dat** file is very large, instead of putting it here, a graph of the resulted values will be shown next. The graphs of each MD analyses results of the present simulation can be seen in Figure 21.



(a) RMSD. (b) Rg.

(c) SASA. (d) RMSF.

Figure 21: Graphs from the performed MD Analyses.

3.8 Protein Molecular Dynamics Article: Typhoid [6]

The main topic in [6] describes the use of de novo design approach to create single-chain fragment variable (scFv) for Salmonella enterica subspecies enterica serovar Typhi TolC protein (typhoid fever).

For the **Molecular Dynamics Engine**, all simulations (including the minimization, equilibration, and production runs) were performed with the **Gromacs software version 4.5.4** [6].

For the **simulation time** the minimization run took 200 ps, the equilibration 1 ns and the production 20 ns. Summing up these makes **21.2 ns** in total, where only the last 15 ns were used for analyses [6].

The **analyses performed** for this article, which results are shown in Figures 22 and 23, include the **RMSD**-based clustering, **GROMOS** algorithm and **RMSF** calculation [6].

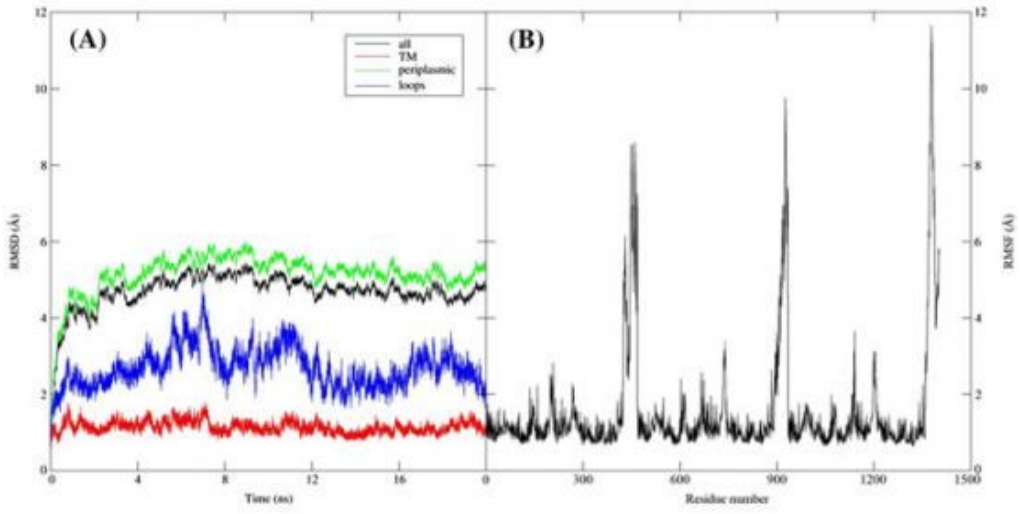


Figure 22: MD analyses for the simulation of the S. Typhi TolC protein. A) RMSD graph. B) RMSF graph. Image taken from [6].

Cluster	No. of Conformations	Central Conformation, ns
1	7262	13.12
2	2687	16.68
3	2474	5.97
4	704	6.23
5	479	19.92
6	432	10.83

Figure 23: Clusters found by the GROMOS method. Table taken from [6].

As **conclusion**, the authors of [6] identified the most dominant structure of TolC protein. Also, through de novo hotspot-based design, novel scFv designs against S. Typhi TolC protein were generated.

They presented 5 scFv designs from different scaffolds with predicted high binding affinity, shape complementarity, and a network of favorable interactions for TolC protein, that are shown in Figure 24. These desings have comparable packing and buried surface area in the interface with that of antibody-protein crystal structures, and could be useful in future development for possible application in S. Typhi TolC protein detection [6].

4 Conclusions and recommendations

Simulations on NAMD are a very important topic for this subject, because they let us understand better the behaviour and the chemical and biological reactions of a protein. It also provide the possibility to generate very useful files, such as the PSF file, that fix PDB files, that only contains the atoms coordinates, into a connectivity one. as well as the DCD file, a trajectory

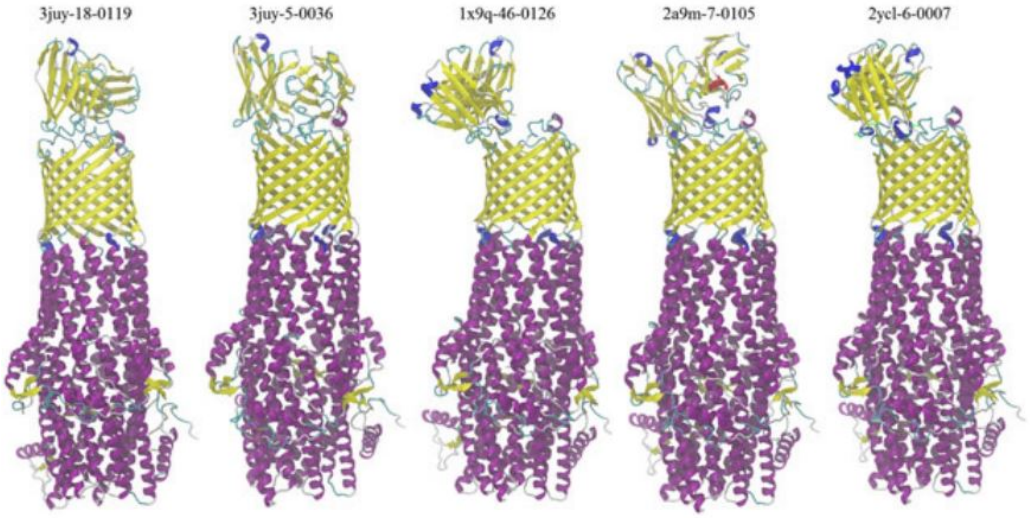


Figure 24: The 5 selected scFv designs and their orientations with respect to S. Typhi TolC protein (ribbon representation). Image taken from [6].

file that models the sequence through time of a process applied in any molecule. On the other hand I would like to make emphasis on the importance of the molecular dynamics analyses, that measures the physics on the molecule are on the simulation.

5 References

[1] Theoretical and Computational Biophysics Group, "NAMD," https://www.ks.uiuc.edu/Research/namd/, [Online; last access November 8, 2020].

[2] Isgro, Tim et al., "NAMD Tutorial. Windows Version," http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-win-html/index.html, 2017, [Online; last access November 4, 2020].

[3] Theoretical and Computational Biophysics Group, "Download NAMD," https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=NAMD, [Online; last access November 4, 2020].

[4] ——, "NAMD Tutorial Files," http://www.ks.uiuc.edu/Training/Tutorials/namd-index.html, [Online; last access November 4, 2020].

[5] M. L. S. of Pharmacy. University of Maryland., "CHARMM Force Field Files," http://mackerell.umaryland.edu/charmm_ff.shtml, [Online; last access November 4, 2020].

[6] Leong, Siew Wen and Lim, Theam Soon and Ismail, Asma and Choong, Yee Siew, "Integration of molecular dynamics simulation and hotspot residues grafting for de novo scFv design against Salmonella Typhi TolC protein," *Journal of Molecular Recognition*, vol. 31, no. 5, p. e2695, 2018.