PREPARE A PROTEIN STRUCTURE FROM EXPERIMENTAL DATA

Jordi Faraudo [last updated 3-April-2024]

The objective is to prepare the pdb and psf data files (coordinates, structural information) needed to perform a MD simulation of a hydrated protein starting from the coordinates available at the Protein Data Bank. The employed forcefield for the simulations will be CHARMM.

Download from the Protein Data Bank https://www.rcsb.org/ the atomic coordinates in pdb format of the protein with code 1UBQ. Create an appropiate folder in your computer and save the pdb file with the name "1ubq.pdb".

1. From the data at protein data bank locate which protein is, the experimental method employed for obtaining the structure and the resolution of the atomic coordinates.
2. Check the presence of atoms that are not protein atoms in the structure (make a selection with only the protein and a selection with all that is not protein). What are these „nonprotein" atoms?
3. Look now at the protein itself. Try different visualizations to get an idea about how is the protein. How is the secondary structure? Is this protein globular? Do you think that the structure is complete?

Go to the File > Save coordinates option and save the coordinates of **only**the protein in a pdb file, use the file name "protein.pdb".

The pdb file contains only the atomic coordinates of the molecule, but essential information for the MD simulation is not there (bonds between atoms, electric charge of each atom). This information can be generated by VMD from the forcefield data and it will be stored in a structure psf file.

You will find the files corresponding to the the version of CHARMM force field that we are using for proteins at https://github.com/jfaraudo/Running_NAMD/tree/main/CHARMM_forcefield_files

4. Using a text editor, open the file top_all36_prot.inp. Which specific information is contained in this file? Now open the file par_all36m_prot.inp. Which information contains this file? What is the difference between them?

Now we will create the protein structure file in psf format.
Open with VMD the protein.pdb file that you created before and remove all other open molecules.

If you are using Windows, you need to tell VMD to create the new files in your selected folder. To do that, open the Extensions -> Tk console menu in order to open a terminal. In this terminal you can check the working directory of VMD by typing the command *pwd*. Usually it will be something like *C:/Program Files (x86)/University of Illinois/VMD* as shown in Figure 1. Change it to your working directory by using the command *cd* in the VMD Tk console, for example:

- cd C:/Users/Jordi\ Faraudo/Documents/Curs_Grau_nano/2020/mdprotein/exercici/build/

Check again that you are in the correct folder by typing *pwd*. **Do not close** the Tk terminal, since VMD will output many messages from here.

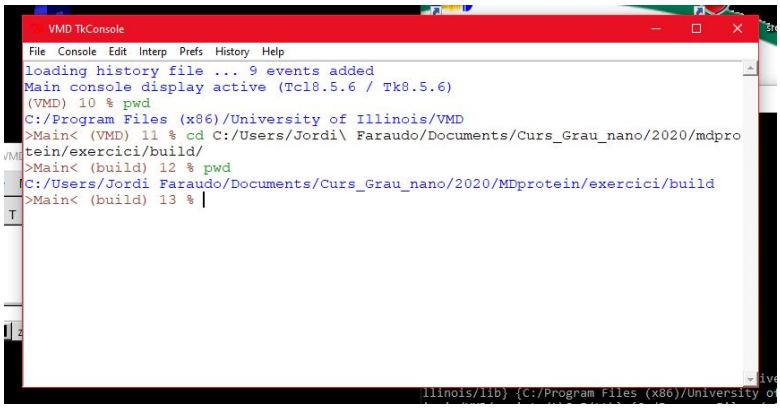

Figure 1: Changing working folder in Tk console

Open the psf file generator located at:

Extensions > Modelling > Automatic psf builder.

This will open a window as the one shown in Figure 2. Go to the window "topology files" (see red circle in Figure 2), select all files and delete them using the „delete“ button. Use the „add button“ to add the file top_all36_prot.inp and press "load input files" (see arrow with “1“).

Press the buttons indicated by arrows 2 and 3 in the figure to complete the structure and “accept“ when asked. The program will create a psf file and a new pdb file named protein_autopsf that are open automatically. Also check that you have new pdb, psf and log files saved in your build folder. In the Tk terminal you will see a list of the actions done by VMD (addition of missing atoms, building bonds, angles and dihedrals, and many more information).

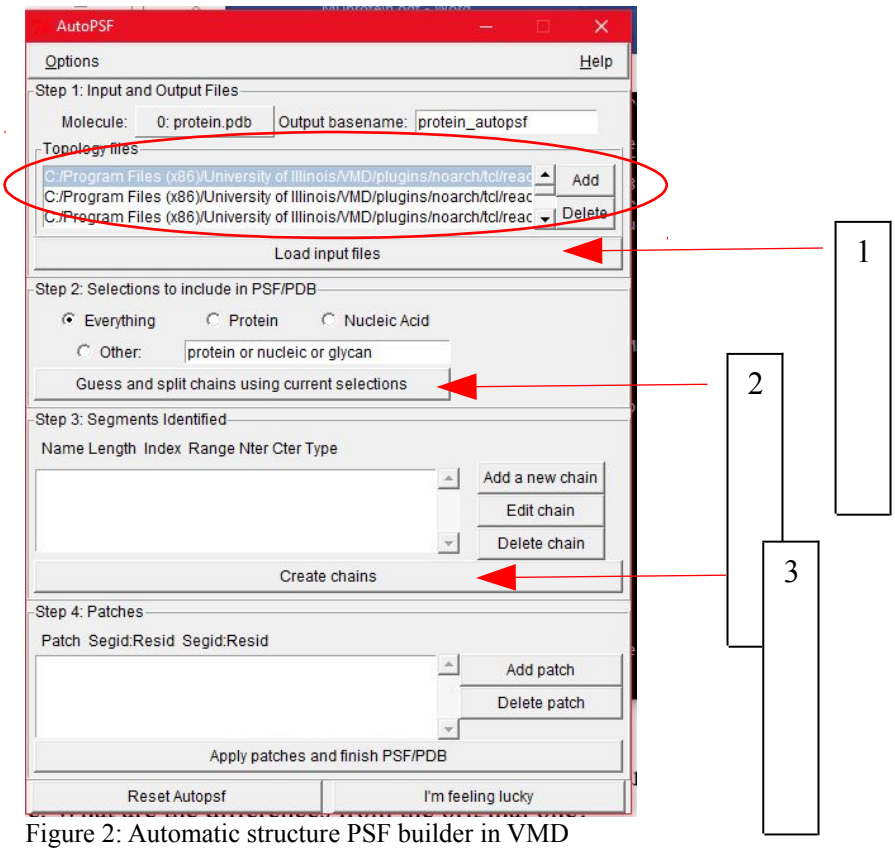

Figure 2: Automatic structure PSF builder in VMD

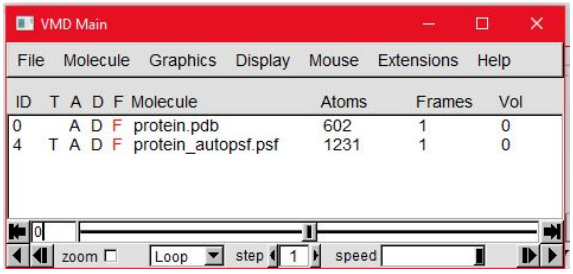

Figure 3: VMD main menu with the two structures

Now you have two molecules open in your VMD (protein and protein_autopsf, see Figure 3). Compare these two molecules. What changes do you see? What are the new atoms? Could you make an image to show the difference?

Adding solvation water

Proteins are always solvated so we will add hydration water using VMD. Delete the molecule „protein.pdb“ in VMD (remember: click over the name of the molecule in the main menu, press right mouse button and select „delete molecule“). Now you have only the „protein_autopsf.psf“ molecule. To add solvation water to this molecule, execute the "solvation" applet, located at Extensions > Modelling > Add solvation Box.

A box as in Figure 4 will open. The default is to add only a layer of water around the molecule (see the tick in the Use Molecule dimensions option), which is OK for us now. Please write a name for the new result (please use „system“) and click the "solvate" button. This will add solvation water to the protein. In your folder you will have new system.pdb, system.psf files with the new coordinates and structures and a system.log file.

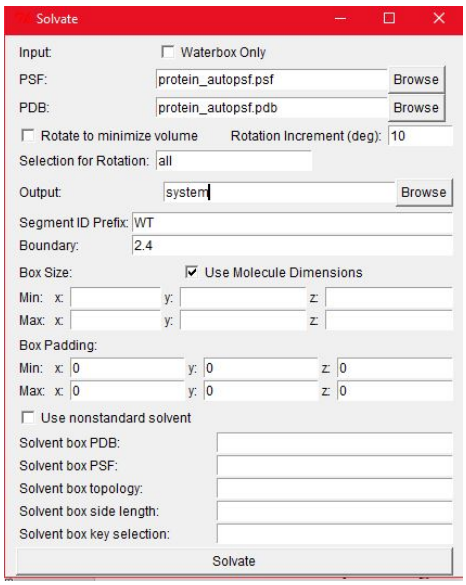

Figure 4: VMD solvate plugin

In the VMD main menu, delete the protein_autopsf molecule and leave only "system". Try different representations. In the Graphical Representations, write "water" at "Selected Atoms", you will see only water. Now click on "Create Rep" to create a new representation and write "protein" at "Selected Atoms". Try different drawing methods for the protein and water in order to visualize both of them and make an image as example (in the image, add a scale bar to help undertand the size of the system).

Centering the structure

One practical problem with the protein coordinates obtained from experimental repositories is that they are not centered. Use the Tk console to manipulate coordenates and move all the system so that the center will be in (0,0,0). Using the Tk console check now the minimum and maximum values of the coordinates of the system. **You may need these values for running the simulation.**

Now save the new coordinates in pdb format by using „Save coordinates“ in the VMD main menu. Save it as system.pdb (if you wish, make a copy of the „old“ system.pdb file before overwriting it).