Protein Modeling, Lecture I

LatEx, RCSB, VMD, Reduce, git, github, Entropy Maxima, CHARMM

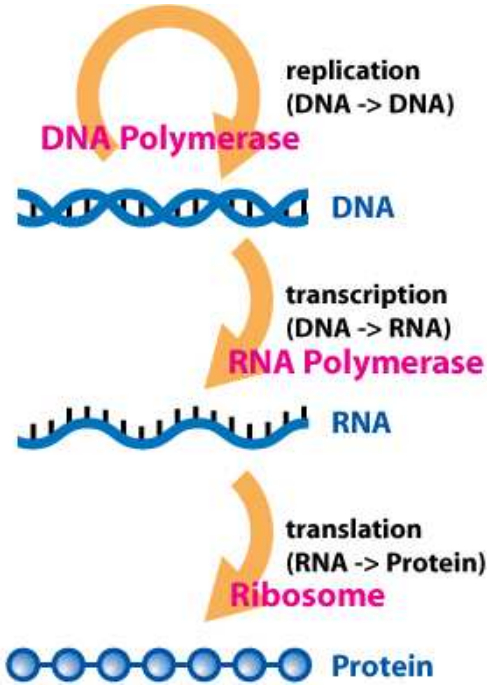
Noel Carrascal

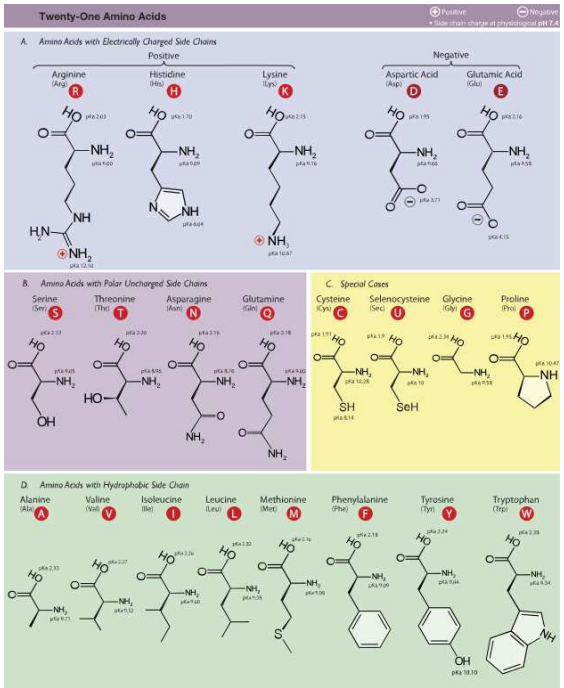
August 9, 2017

Outline

1. Introduction
2. Learn about the Protein Databank at rcsb.org
3. Download and install VMD. View Protein in VMD and learn some basic commands.
4. How Crystal structures are obtained. Hydroge can go unnoticed, need to be added.
5. Quick intro to git and github. Fork and pull EntropyMaxima and CCL_Lectures repositories.
6. Bash shell. Intro and simple use for automating multiple step procedures.
7. Prepare a protein structure.

Introduction

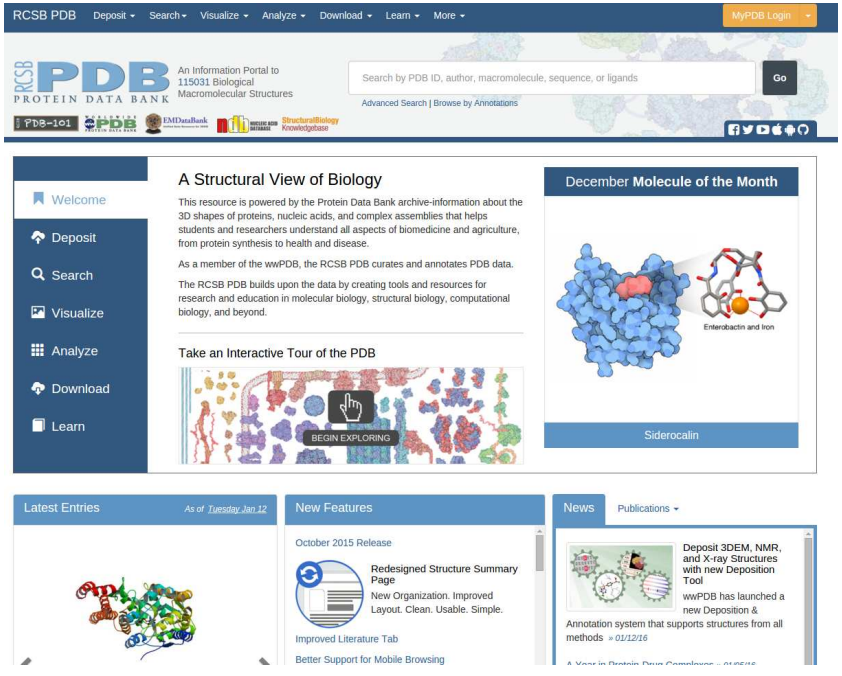


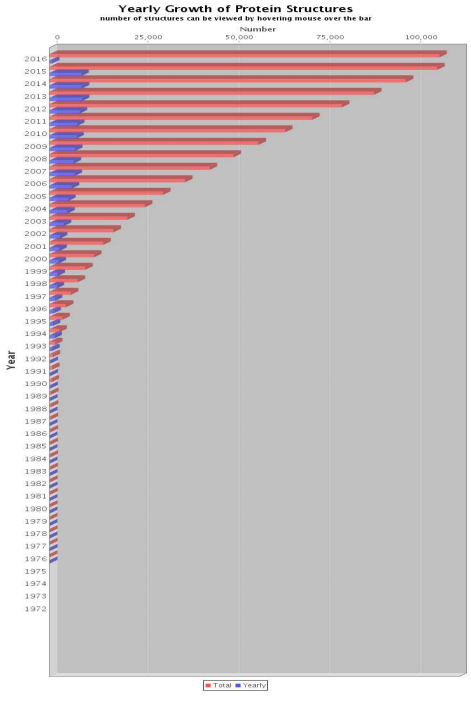


Licensed under CC BY-SA 3.0 via Commons

https://commons.wikimedia.org/wiki/File:Extended_Central_Dogma_with_Enzymes.jpg

Protein Data Bank (www.rcsb.org)

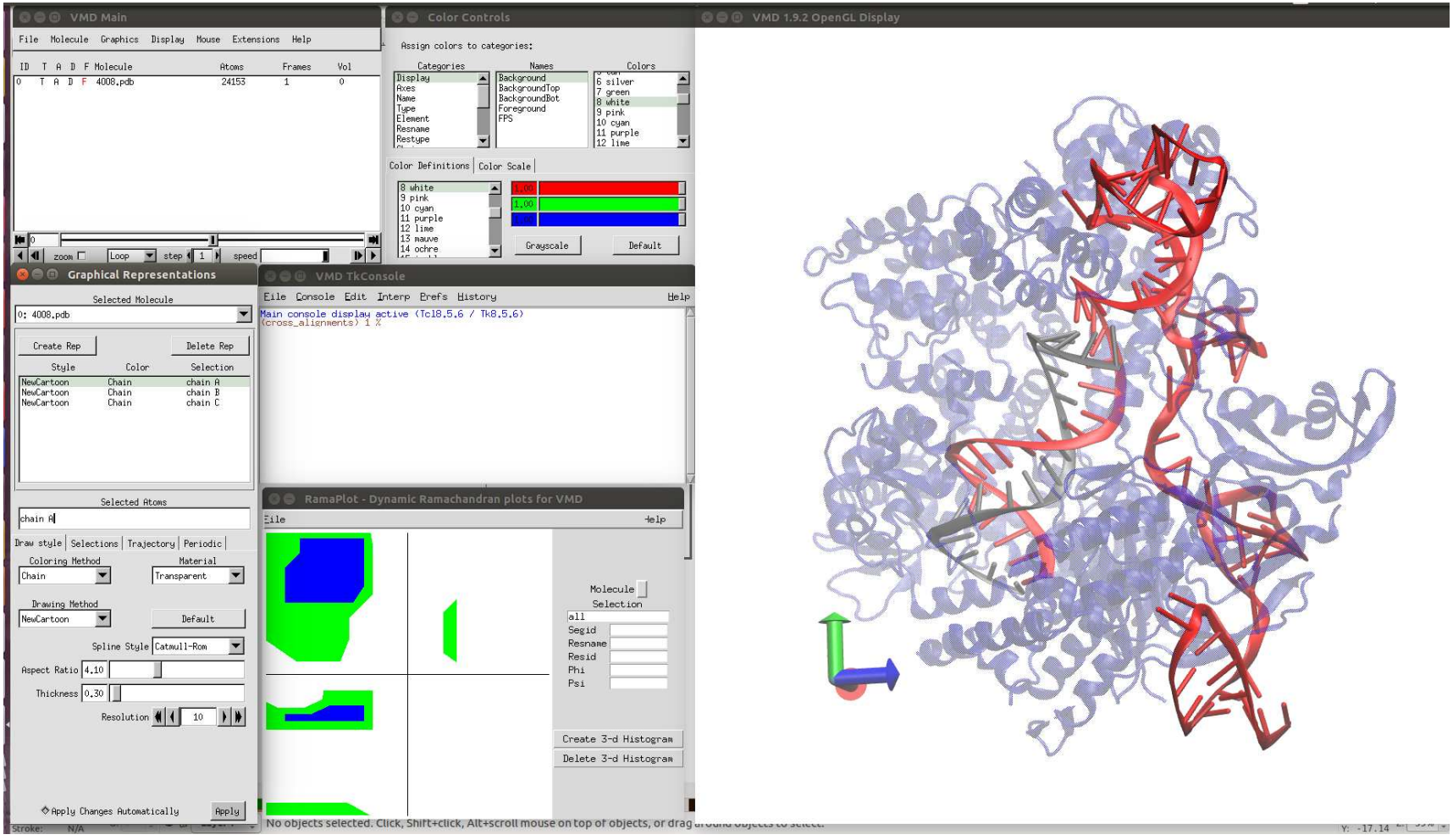




PDB Holdings Breakdown up to 2016

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-Ray	96036	1698	4850	4	102508
NMR	9854	1136	231	8	11229
Electron Density	673	29	230	0	932
Hybrid	83	3	2	1	89
Other	170	4	6	13	193
Total	106816	2870	5319	26	115031

Visualization with VMD



Using VMD I: Loading and Viewing files.

1. Download, install and open VMD from http://www.ks.uiuc.edu/Research/vmd/
2. Two windows open when VDM is launched. They are labeled 'VMD Main' and 'VMD 1.9.3 OpenGL Display'.
3. Download the protein structures 2hiu.pdb (insulin) from www.rcsb.org.
4. To Open 2hiu, in the 'VMD main' window go to the menu option File > New Molecule...
5. A window opens with three fields, make sure those fields have the right entries:
   - 5.1 In the "Load Files For:" Pull down menu, select "New Molecule"
   - 5.2 Click the "Browse" button to find the PDB file you want to open (Insulin or 2hiu)
   - 5.3 The "Determine File Type" field usually detects the right format for the file to upload, but be sure it is PDB in this case. Some structural file formats have the the same file name extension for different formats and VMD might pick the wrong one (e.g. The crd file name extension is used for both Amber and CHARMM).
   - 5.4 Click the "Load" button.
6. On the Display windows, the mouse pointer is an arrow. Hold down the right button in your mouse and move it. It rotates.
7. Hit the 'T' (pointer becomes a hand) letter in the keyboard and holding down the right button of the mouse translates the molecule. Hit the 'R' key to go back to rotate pointer is again an arrow.
8. Hit 'Ctrl A' or 'Ctrl Z' to zoom in and out respectively.
9. Hit 'Y' and the protein rotates by itself. Click anywhere in the display and move pointer while holding right button in the mouse to stop the rotation. Rotation, translation and zoom are possible while the protein rotates.
10. In the 'VMD main' window go to Display > Reset View to bring back the protein to the original position right after loading.

Using VMD II: Labels and Measurements.

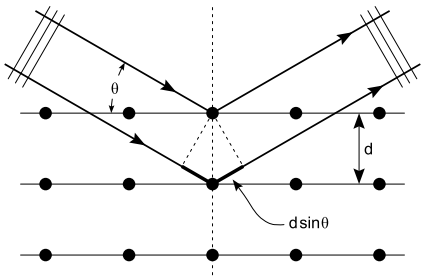
1. In the 'VMD Main" window go to menu option Graphics > Labels...
   - 1.1 Hit '1' on the keyboard, and the pointer in the display changes from an arrow to a cross. Click on an atom in the Display. You will see information about that atom both on the display and Labels window. Click again on the atom, what happens on the display and Label Window?
   - 1.2 Hit '2' on the key, the pointer is still a cross but this time it does something different. Click on one atom, and then click on a second atom. The distance appears on the Display. On the Labels Window, pull the drop down menu and go to bonds. What do you see?
   - 1.3 Hit '3', still a cross, click on three atoms, and drop down menu to Angles.
   - 1.4 Hit '4', still a cross, click on Four atoms, and drop down menu to Dihedrals.
2. Hit 'R' to avoid making an involuntary selection for measurement. The pointer goes from a cross back to an arrow.

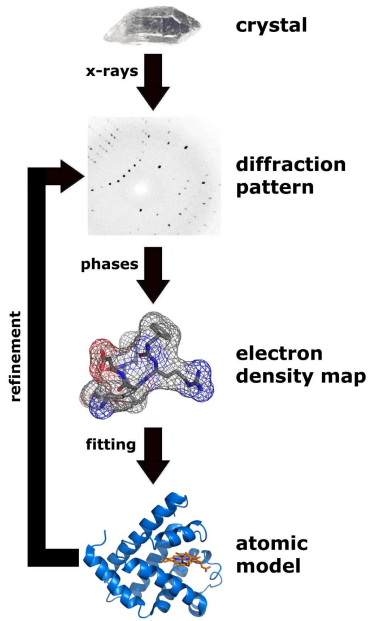
Using VMD III: Graphic Representations and save a visualization states.

1. In the 'VMD Main" window go to menu option Graphics > Representations...
2. The default Style, Color and selection is Lines, Name and All. Lets Play with the Coloring Method and Drawing Method Using the pull down menus in the 'Draw Style' tab below.
3. The 'Selected Atoms' Field is not obvious to understand. Go to the 'Selections' tab' For some options. I will go over a few only.
   - 3.1 Double click 'backbone', hit space, click the 'and' button, hit space, double click 'type', space, and double click CA. Hit 'Apply', and watch the display. If your selection is wrong you will get a message. If the selection is an empty set, you will see nothing.(of course). Play with other combination of selections.
   - 3.2 'resid 1', 'resid 1 and chain A', 'resid 1 to 10 and chain A'. Substitute 'resid 1' with 'resname CYS'.
   - 3.3 An useful one, not easy to figure out: 'all (within 5 of resid 1 and chain A)'.
4. You could spend a long time working on graphic representation. It is possible to save this visualization stat. Go to File > Save Visualization State, pick a name and directory. To reload, File > Load Visualization State.
5. VMD can do a lot more. Reading its manual or doing Google searches can give you more options.

X-Ray Crystallography and its Limitations

1. Hydrogen atoms have an electron density that X-rays might not detect, so they might not show up in crystal structures.
2. Regions of the protein with high thermal motion at cryogenic temperatures will not diffract x-rays in a way that can resolve their coordinates.





Some Powerful Tools Needed for Protein Modeling

1. Learn some basic Linux command line (see cheat sheet). Work with the Bash shell. It will empower you to do extremely complicated tasks you never thought you could.
2. You can use a text editor of your choice to create and save a list of bash commands. This will allow you to generate long lists of complicated procedures needed for protein modeling that you do not have to memorize.
3. I can help you with the Vi text editor for which I have included a cheat sheet, but the choice of editor is yours.
4. Learn the basics of git and open a github.com account. This is how you get my code a.k.a Entropy Maxima. This code works with a command line interface.

Get the code using Git from github.com

1. Create an account on github.com, and go to https://github.com/noelcjr/EntropyMaxima
2. Fork the repository on github.com https://guides.github.com/activities/forking/
3. For the time being you will only use git and github.com to get the code and use it. We will learn how to make changes to the code later. This first lecture is only a demo.
4. Now that you have a forked repository in your github account, let's download it to a directory in your computer and use the code in it to modify proteins.
5. git clone https://github.com/<url to your forked repository>/EntropyMaxima
6. get the files for this lecture: 'git clone https://github.com/noelcjr/CCL_Lectures' This repository has bash scripts, but do not run them in this folder, just copy the scripts to a temp folder.

Explore and Prepare a Protein Structure with Entropy Maxima

1. Create a folder called temp_lecture_1 in the same folder with EntropyMaxima and CCL_Lectures.
2. Type the bash commands: 'mkdir temp_Lecture_1' and 'cd temp_Lecture_1'
3. Copy the Lab_1_complete_structure.sh from the CCL_lectures repository to the folder just created.
4. Open Lab_1_complete_structure.sh and run each line at the time.

Setup Your Environment

1. Install the following programs to work from your computer:
   - 1.1 Install Linux, SSH, or use the bash shell in Macs or new Windows machines(?).
   - 1.2 Install Python 2.7.12 (We need to test program in newer versions.)
   - 1.3 Install git
   - 1.4 Install reduce from http://kinemage.biochem.duke.edu/software/index.php
   - 1.5 Install CHARMM from https://www.charmm.org/charmm/showcase/news/free-charmm/
   - 1.6 Install NAMD from http://www.ks.uiuc.edu/Research/namd/
2. Or, help us set up our servers to log into and run the tutorials and programs.

Thank you