CHEM 498Q / CHEM 630Q: Molecular Modelling of Proteins
TUTORIAL #1c: Protein visualization

INTRODUCTION

The goal of this short tutorial is to learn how to visually inspect a protein structure for its important features and how to produce publication-quality images of proteins.

Read carefully the instructions at the end of the document and make sure your figures are exactly as required.

REQUIRED PRE-LAB READING

http://www.pymolwiki.org/index.php/Practical_Pymol_for_Beginners

PRE-LAB REPORT

None.

ABOUT THE COMPUTER WORKSTATIONS

To start the Linux Virtual Machine

From now on, you will find it easier to work entirely from within the Linux Virtual Machine installed on the workstation. Start "Oracle VM VirtualBox" from the Windows XP environment, then start the "Molecular modelling" virtual machine by double-clicking on its icon.

You can enlarge the window or use the full-screen mode by selecting "View > Full-screen Mode" from the menu (or by pressing Host+F, which corresponds to Right-Ctrl+F).

To store files on the Linux VM

Create your own work folder on the desktop by right-clicking on the desktop background. Give it a name that can clearly be identified to you, containing no special characters and no spaces (example: "`work_guillaume`"). Put all your files in that folder. Do not expect this folder to be safe from one week to the other and back up everything at the end of every tutorial.

To back up your files on the Linux VM

Right-click on the folder icon, and use "Create Archive…" to create a ".tar.gz" file on the desktop (example: "`work_guillaume.tar.gz`"), which you can then email to yourself or upload to DropBox or Google Drive.

Alternatively, you can create the archive directly from the terminal. Start the Terminal Emulator, go to the Desktop directory by typing "`cd ~/Desktop`", and create the archive by typing the command "`tar -z -cvf work_guillaume.tar.gz work_guillaume/`".

To shut down the Linux VM

Select "Machine > ACPI Shutdown" from the Oracle VM VirtualBox menu. (In full-screen mode, this menu will be at the bottom of the screen.)

PROCEDURE

STEP 10: Examine the PDB file of your structural template

From the Protein Data Bank website (http://www.rcsb.org/pdb), search for the 4-character PDB ID of the protein structure you have chosen as a structure template for your query sequence.

✦ What is the primary citation for this structure?

✦ What experimental technique was used to solve the structure? What is the resolution?

View the PDB file by pulling down the "Display Files" menu.

✦ Is the protein binding any metal atom? (see "`REMARK 620`" lines) Are these atoms known to play a functional (or structural) role?

Examine the "`SEQRES`" lines and make sure the protein structure includes the amino acid sequence you are interested in.

✦ List all "hetero-residues" contained in the PDB file (see "`HET`", "`HETNAM`", and "`FORMUL`" lines) and comment on their significance. (Keep track of both their 3-character identifiers and their chemical names.)

✦ What is the secondary structure of the protein sequence? (see the "`HELIX`" and "`SHEET`" lines) How does it compare with the secondary structure predicted at STEP 6?

✦ Are there any disulfide bonds? (look for "`SSBOND`" lines)

STEP 11: Use PyMOL to inspect your structural template

Load the protein structure into PyMOL. You can either download the PDF file "`ABCD.pdb`" on your computer and open it from the "File > Open…" menu or type "`fetch abcd`" from the command-line interface:

```
PyMOL> fetch abcb
```

As soon as the file is read, two objects are created: "`all`" and "`abcd`". Change the rendering of the "`abcd`" object by clicking the "S" button and selecting "Show as cartoon".

Display the protein sequence (using the "Display > Sequence" menu) and highlight all the amino acids forming your template sequence. A new object called "`sele`" is created, that contains all selected residues.

✦ Make your template sequence a different color from the rest of the protein by clicking the "C" button of the "`sele`" object. Is the template sequence missing any residues?

Note that, at any time, the "`sele`" object contains the residues currently selected. An easy way to change the rendering of any part of your structure is to select the desired residues and use the "A", "S", "H", "L", or "C" buttons of the "`sele`" object.

STEP 12: Create an overall view of the protein

Create a general view of the protein similar to Figure (A) below. Show the entire structure in a "cartoon" representation.

Once you have a view that you like, raytrace it by typing "ray" in the PyMOL command-line interface:

```
PyMOL> ray
```

To save the image file as "`filename.png`", type:

```
PyMOL> png filename.png
```

STEP 13: Create a view of the active site of the protein

Locate the active site of the protein domain and create a selection containing all the residues known to be functionally important.

Create a "close-up" view of the active site similar to Figure (B) below, showing the important residues (with labels) and the ligand bound at the active site (if there is any).

Raytrace and save your image.

Note that you can save your PyMOL session with the "File > Save Session As..." menu. This will write a ".pse" file that you can re-open later.

HELP

Some useful PyMOL moves (using a 3-button mouse):

To rotate the molecule:	Left click + drag
To zoom in or out:	Right click + drag
To translate the molecule:	Middle click + drag
To clip the molecule:	Mouse wheel
To select a residue:	Click on one of its atoms (click again to deselect)
To select multiple residues:	CTRL + left click

Some useful PyMOL commands:

To create a new "selection" (a set of atoms that you can render differently from the rest):

```
PyMOL> select <name_of_object>, <selection>
```

Examples:

```
PyMOL> select metals, symbol mg+ca+fe+zn
PyMOL> select heme, resn hem
PyMOL> select helix, resi 1-42
PyMOL> select backbone, name c+o+n+ca
PyMOL> select nothydrogen, not symbol h
```

To "deselect" all atoms:

```
PyMOL> select none
```

To show disulfide bonds as sticks:

```
PyMOL> show sticks, (cys/c+ca+cb+sg) and byres (cys/sg and bound_to cys/sg)
```

For more information, see the (old) PyMOL manual at http://pymol.sourceforge.net/newman/user/toc.html.

INSTRUCTIONS FOR THE LAB REPORT

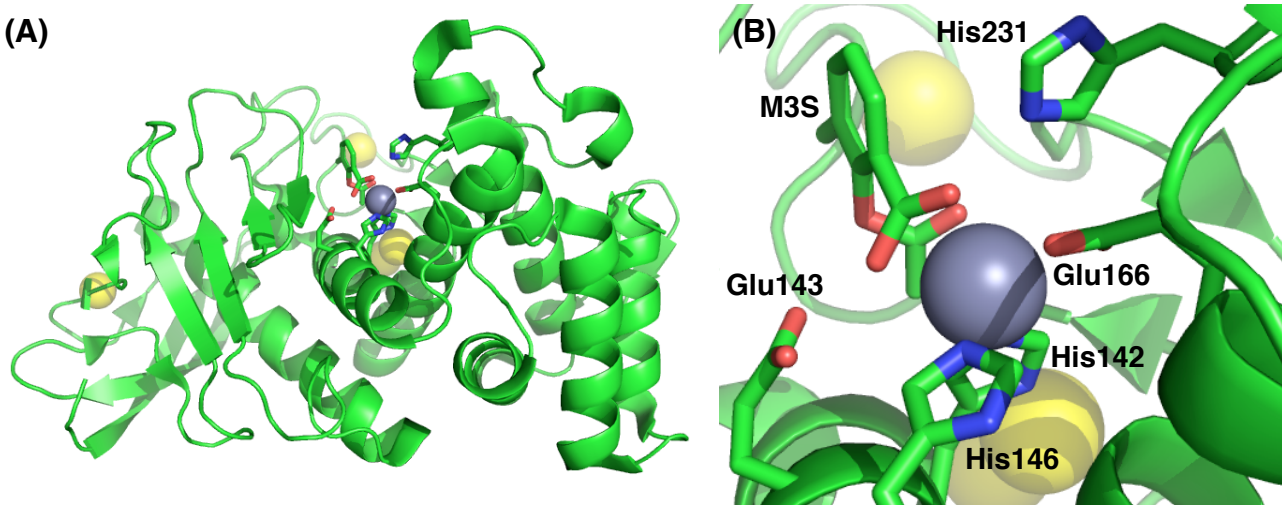
STEP 12 and STEP 13

Include the two images in your report, along with a detailed caption indicating the name of the protein and the PDB ID of the structure.

- Use a "cartoon" rendering for the overall protein sequence, a "sticks" rendering only for the important amino acids and the ligand molecules, and a "spheres" rendering for the metal atoms (if any).
- Remove all the hydrogen atoms using "Setting > Auto-Remove Hydrogens".
- Label an amino acid using its three-letter code, followed by the residue number (ex: "Ala143").

• If your structure contains multiple equivalent chains, show only one of them.
• All your molecular pictures should be on a white background (“Display > Background > White”).

Example:



Structure of thermolysin in complex with 2-(acetyloxy)-3-methylbenzoic acid (M3S) bound at the active site (PDB identifier 3FCQ). The active site contains a zinc atom, represented as a gray sphere. The yellow spheres are calcium ions. (A) Overall view of the protein. (B) Structure of the active site. The zinc atom is coordinated by residues His142, His146, Glu166 and the inhibitor M3S.

The ligands in the figure were selected using:

```
PyMOL> select ligands, (resi 142+146+166+143+231 and not name c+o+n)
       or resn m3s
```