

# ECS 129: Winter 2024

## Assignment #1: Pymol

**Due:** Thursday, January 25

The objective of this assignment is double: firstly, to become more familiar with protein structures, and secondly, to become familiar with Pymol.

Pymol is an open-source program for displaying and manipulating macromolecular structures. It can be found at: <http://www.pymol.org>. Please download the software and install it on your computer (if you have not yet done it).

A simple tutorial on how to use pymol can be found on youtube at:

<https://www.youtube.com/watch?v=aDmOe1ZgTz0>

A (slightly) more complete tutorial can be found at:

<https://dasher.wustl.edu/bio5357/software/pymol/simple-tutorial.pdf>

In addition, please look at : A nice introduction to using scripts:

<https://www.compchems.com/scripting-in-pymol-a-comprehensive-guide-for-beginners/#why-scripting-in-pymol>

As well as:

[https://pymol.sourceforge.net/newman/user/S0210start\\_cmds.html](https://pymol.sourceforge.net/newman/user/S0210start_cmds.html)

Please check all!

### Installing Pymol:

- Information for free license for education: <https://pymol.org/edu/>
- If you fill in the form, you should receive an email from Schroedinger, with a username, a password, and a link from which you will be able to download the version of Pymol for your operating system. Do note that in the same link, they provide a license file that will allow you to run pymol. This is really useful, especially on Windows systems.
- Specific information for Windows:  
[https://pymolwiki.org/index.php/Windows\\_Install](https://pymolwiki.org/index.php/Windows_Install)  
This is really useful if you want to install the open-source version of pymol.
- Specific option for Mac OS X: Using brew to install pymol:  
<https://formulae.brew.sh/formula/pymol>.  
Do note that after running brew, pymol will be located at /usr/local/opt/pymol/bin/pymol !!
- Linux install of pymol: [https://pymolwiki.org/index.php/Linux\\_Install](https://pymolwiki.org/index.php/Linux_Install)

### Accessing and Analyzing Structures from the Protein Data Bank

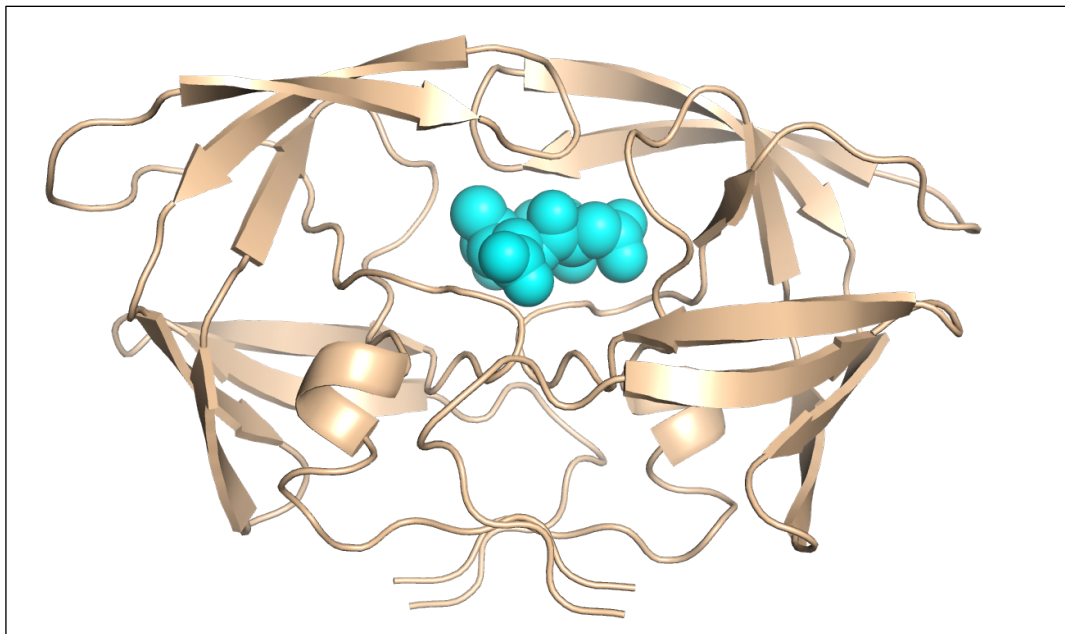
The Protein Data Bank, or PDB, is the best place to obtain 3-Dimensional structures of proteins, nucleic acids and other macromolecules. From the PDB site <http://www.rcsb.org>, you can locate proteins by keyword searching or by entering the PDB accession number for the structure file, like **5PTI**. Details on the molecule (how the

structure was determined, pertinent research articles, position of secondary structures, unusual amino acids, etc) can be found on the RCSB web site but also in the PDB file itself.

PDB files are just formatted text files, so you can open them in a text editor or even Word and read them. There is a wealth of information there! The molecular viewing programs use the ATOM records in the file, which contain residue name, residue number, atom name and atom number. Often, the structure files include other molecules besides the protein, such as water molecules, nucleic acid and bound ligands.

- 1) Find a protein that contains the unusual amino acid HYP. What is this amino acid?
- 2) We first look at the structure of an HIV protease. We choose the PDB file 1A30. Download this PDB file and display its content using PYMOL. This structure shows HIV protease as a dimer (chain A and B), bound to a small tripeptide (chain C, sequence EDL). Generate an image in which the ligand is shown in space filling mode (CPK), and the protein dimer is shown in cartoon mode. Save this picture and include it in your report.

Here is the type of picture I expect:



- 3) Now we look at the PDB file 1CWA, which contains the structure of the complex between two proteins. Find the protein names, sizes and secondary structure contents. Identify the non-standard amino acids in chain C.
  - a) Briefly describe the importance of this structure.
  - b) Generate an image containing a cartoon representation of chain A and a stick representation of chain C.

**Good Luck !**