Hands-On Fourteen
Protein Structure and Visualization

In this Hands-on Exercise, you will learn how to:

1. Search the Protein Data Bank (PDB).
2. Choose the best structure, when more than one is available.
3. Visualize a protein structure and features of interest.

You will work with Rhamnogalacturonan acetylesterase (RGAE) from the fungus *Aspergillus aculeatus*. You will first find the sequence of the enzyme, and then use the sequence to search the PDB.

1. Go to UniProt: http://www.uniprot.org/
2. Enter "rhamnogalacturonan acetylesterase" in the search field and click "Go". You should see several matches.
3. Click on the SwissProt entry (Q00017). What is the function of this protein?

4. If you scroll down to the "Features" you can learn a few things about this protein. Record the following information about this protein:
   - The signal peptide is from residue number _____ to _____.
   - The mature protein is from residue number _____ to _____, which means that the protein is _____ residues long.
   - The active site is made up of _____ residues. The residues are _____.
   - The protein is post-translationally modified, having two sites of N-glycosylation at _____ and _____.
   - There are two disulfide bridges in this protein. The first links two cysteines: Cys_____ and Cys _____.
5. Retrieve (copy) the FASTA sequence of the protein and go to PDB http://www.rcsb.org/ You can search the PDB immediately from the front page using a keyword or a PDB ID in the search field, or you can do a more advanced search using the buttons above the search field.
6. Click on the "Advanced" button at the PDB home page. Select "Sequence (BLAST/FASTA/PSI-BLAST)" from the "Choose a Query type" drop-down menu and paste the RGAE sequence into the sequence field. Set the "Mask Low Complexity" option to "No", set "E Cut Off" to some reasonable value, leave everything else as it is and click "Submit query" (lower right). Inspect your results.
7. Are all hits are relevant if you are looking for a representative structure of the sequence you entered? Which parameters should you look at to make this decision? [Hint: You may need to inspect the full alignment for each hit – simply press the "Display Full Alignment" option".]

8. You should find more than one structure, which represents RGAE. You only need one, so you will have to decide which one is the best to use. To create a table showing the parameters you wish to compare for selected structures, select "Customizable

Table" from the "Generate Reports:" drop-down menu. You now get a long list of parameters to include in a report. You should only choose the relevant ones, or your resulting table will be very large. Select the following:

- Ligand name
- Resolution
- R-free

9. Click "Create Report". Notice that if a PDB entry has more than one ligand, there will be one line for each ligand in the resulting table.
10. Click on the PDB ID of the structure you chose. This will take you to the page showing this entry in the PDB. Examine the information on the page.
11. If you click the "Display Files" drop-down menu of the web page and select "PDB File", you can see the contents of the PDB file in the text format. The first lines in this file are header lines and contain various pieces of information about the structure. Below the headers, you will find the 3-D coordinates (x,y,z) of each atom in the protein structure. These coordinates are found in the part of the PDB file in the lines that start with "ATOM" (or HETATM for non-protein atoms).
12. Open another browser window or tab and go to http://www.wwpdb.org/documentation/format33/v3.3.html to display the information about the format of the PDB files. What is the residue name for the sulfate ions? _____
13. You can visualize the structure directly through the PDB website using various Java-based viewers (buttons are found below the structure image), but we will use the viewer UCSF Chimera (http://www.cgl.ucsf.edu/chimera/download.html) for our purposes. Start UCSF Chimera and load the RGAE structure by going to "File", then "Fetch structure by ID" and enter "1K7C" in the "PDB" search box".
14. Select all water molecules by going to the "Select" drop-down menu, then "Residue" and then "HOH". Under "Actions" select "Atoms/Bonds" and then "Show".
15. Delete water molecules.
16. How many sulfate ions are there in the PDB file? _____
17. One of the sulfate ions is located near the enzyme active site. Can you identify the amino acids of the active site? [Hint: you can point with your mouse at the residues in the Chimera window]. Record the active site residues below:

    _____.

    Does this correspond to the information you wrote down earlier from the UniProt entry? Why/Why not? [Hint: display the sequence of the protein by going to "Tools", then "Sequence" and "Sequence".

These exercises have been adopted from Anne Mølgaard and Thomas Holberg Blicher