Biochem 4110
Pet Enzyme Project 1
Protein Structure

This project is to be done primarily on the Web, using publicly available programs. By the time you finish this assignment you should be more aware of the wealth of resources available for biochemists on the Web, and have a better concept of protein structure.

FIRST YOU NEED AN ENZYME:
Go to the course web site and pick one of the enzymes still available (edit the wiki by adding your group members last names next to the enzyme that you’ve chosen). After you have picked one that is available, you will need protein database (PDB) code.

1. On the list of enzymes, there is in a four-character code, or identifier. This is an identifier for the protein in the PDB. Go to the PDB at http://www.rcsb.org/pdb/. Here you will see a page that allows you to search the database using the 4 character code, or by keyword. You can do either, but I suggest trying a keyword search first, so you can get some idea of the number of variants\(^1\) of your sequence there are. Click on the pdb code of your enzyme. This takes you to the Structure Summary page for that enzyme. **Print this out.** It contains a variety of useful information. At the top of the page you will see a number of different tabs. Select “Sequence”. **Print out this page.** More useful information. Click on the “fasta” link in the Chain window to save this sequence in .txt format. (If your protein is composed of more than one polypeptide chain, only use one of the chains for this and the following two steps.) This will be useful for steps two and three. Using the tabs on the top, switch to the “Geometry” tab. Click on ‘MolProbity Ramachandran Plot’. This should result in a pdf download of several Ramachandran plots for your protein. **Print this out.**

2. Now the fun really begins. Find on the Web a program or programs that will take your amino acid sequence and predict secondary structure for you (yes, I know that you have the known secondary structure already printed out). With a little web surfing it is easy to find several sites that will do this. For these secondary structure prediction programs you will need to paste your copied sequence into a box on the web page. Most of these programs will use the sequence that you copied from PDB page.

Here are a few places you can try:


Another suite of free software, once you register. This may accept your copied FASTA sequence or it may not. If not, you will have to re-acquire your protein sequence. You can do this a number of ways. One is via a multiple database search from with Biology Workbench, search PDBFINDER database and again, pick the same identifier. Once it is imported, you can

---

\(^1\) Variant structures can be similar enzymes from different organisms, different fragments of one enzyme, mutant forms of one enzyme (differing in one or more amino acids), or even the exact same protein analyzed multiple times, perhaps by different research groups.

Copyright © 2016 Destin Heilman
run any of the protein analysis tools there, checking the box to the left of your protein name (above the operation box) before you hit the ‘Perform selected operation’ button.

Other lists of links can be found at:

http://sites.univ-provence.fr/wabim/pedro/rt_all.html

Note that these links are simply lists of links to other sites. At times the links become out of date, if the list is not well maintained. Finding a dead link is not uncommon, so don’t be too surprised (and don’t blame me!).

Use at least two different prediction programs (some sites will do a number of predictions simultaneously) and print out the results of both. Note: this is different from using two different sites. Protein secondary structure prediction is based on some algorithm, and there are several available. Use at least two different algorithms.

3. At the same time you should be looking for programs that will calculate the hydropathy (hydrophobicity or hydrophilicity) of your protein. These are harder to find, but also exist. Print out one hydrophobicity plot (or hydrophilicity plot).

4. While you are on the web, find one other program that uses your sequence to make some prediction about your protein. There are many of these. Take a look around at the various different programs that you can find and use at least one additional protein tool to either predict molecular weight, isoelectric point, known motifs or anything else. Print this out as well. Having spent much time cruising through the above pages, I know you shouldn’t have any trouble finding some additional program to analyze your sequence.

5. Look up your enzyme in the NCBI (National Center for Biotechnology Information) database. It is found at http://www.ncbi.nlm.nih.gov/. Once here you can search for protein or nucleotide sequences, structures, or literature. Select “Protein” from the drop down menu and type in the name of your enzyme. The search may yield a number of matches and will tell you how many. Some of these entries may be of mutant forms of the enzyme; some may be from different organisms. How many different entries are there for your enzyme? In order to acquire the sequence of the protein used in the structure go back and type in the PDB identifier code. This may give you a unique entry. Select this entry. If this does not work, and there are >30 entries, pick one that is from the same organism and that is not identified as a mutant or a fragment. If your protein has more than one chain, try typing in the pdb code followed by an “A”. This may get you the sequence of one of the chains. This page will provide a lot of information in addition to the sequence. At this time print out the screen. This includes the initial reference for the sequence and some information about the source of the gene and even some information about the enzyme itself.

6. Now, having looked at primary structure and secondary structure, let’s look at the protein’s 3D structure. On the front page of the PDB for your enzyme, there is a box on the right titled “Biological Assembly”. Select ‘Protein Workshop’ from the list under the picture (which may
take you to another screen where you have to click on “Launch RSCB protein workshop”). You should eventually, after everything gets loaded, see a structure on the left and a couple of boxes on the right. First play with the structure. You can use the mouse to rotate it and move it around. Look for some interesting angles by rotating the structure of your protein. See if you can find both (or all, if you have more than one chain) ends of your protein. If it has more than one chain, can you find an angle that clearly shows the region between the chains? If there are ligands (inhibitors or metal ions) can you find an angle that clearly shows these sites? Can you find an angle that clearly shows a cleft in the protein? Note that at any time, you can move the mouse over any portion of the structure, and in the bar below the structure, the identity of the corresponding part of the protein (if in ribbon diagram the secondary structure segment; if in atoms/bonds the atom and residue) will be listed. Then click on the ‘shortcuts’ button on the top at the right. Using the top box, you can change the color scheme of the protein. Try them all.

Now for the more advanced work. Go back to the ‘tools’ page. In box 1, click on ‘visibility’; in box 2 click on atoms and bonds; in box 4 click on the chain (some of you will have more than 1 chain. Select both). This should cause the image to include a ball-and–stick structure of the protein on top of the ribbon diagram. Now we need to get rid of the ribbon diagram. In box 2 select ‘ribbons’; in box 4 again click on the chain(s). Now only the ball-and-stick image should remain. You can use these sets of commands to make appear and disappear any heteroatoms or waters visible in the structure. If you now go to the shortcuts page, you can visualize the protein a number of different ways. Again, select some method to color the backbone (and hit the enact button). Make sure that you select from box 2 (Recolor the atoms/bonds by..) the ‘Corresponding backbone color’ radio button. Again, try the various color schemes. You can get a space-filling structure by going back to the ‘tools’ page. Select the ‘styles’ button in box 1, the ‘atoms and bonds’ button in box 2. In box 3 the ‘radius of atoms’ scroll down to CPK, and again in box 4 select the chain. You can also view the overall molecular surface structure (based on electron density mapping) by sliding the “surfaces” slider at the bottom right.

I want you to print out at least two images (these do not have to be in color) of your protein viewed from different angles and in different styles (spacefill and ribbons). To print:

1. Click on “Options” at upper right toolbar
2. Click on ‘Save Image’, select a destination, name the file and press enter to save.
3. Print the resulting JPG file.

To get even more fancy, go to the ‘options’ page and hit the reset button. Then, go back to the tools page. Click on the visibility button, then the atoms/bonds button. Now in box 4, click on the little arrow beside the chain. This will expand the list to include ALL the amino acids in the chain in linear order. Select one or two. On the structure diagram, you should now see these amino acids in ball and stick form on the structure. You can change these to CPK by the same methods as above, except that again in box 4, you should select the same set of amino acids. Do this for several different amino acids in the structure. Try to determine whether the side chain would be exposed on the surface or buried in the interior of the protein. Print out at least one more image of your protein that has at least one amino acid whose side chain is on the surface in CPK. Identify that amino acid by residue number and amino acid identity.
WRITTEN PORTION OF ASSIGNMENT. This must use correct English, be word-processed, spell checked and in general as professional as you can make it. All the technical know-how in the world is useless if you cannot communicate it accurately and clearly. Writing is, therefore, an essential, integral part of both science and engineering. So, I DO care about grammar and organization and your grade will reflect the care that you put into this portion of the assignment.

1. Is the 3-D structure you are using the structure of the entire protein or just a portion of the protein? (Use Chain color ramp to help you determine this) Why might the researchers have tried to crystallize only a portion of the protein? Why, even if the entire protein was crystallized, might the structure only be displayed for a part of the protein? Why might researchers choose to work with a mutant form of the enzyme?

2. Does the hydropathy plot that you obtained for your protein indicate that your protein is more hydrophobic or hydrophilic? Are the amino acids on the surface of the protein more hydrophobic or hydrophilic? Given this data, would you expect your protein to be cytoplasmic or membrane localized? Why?

3. Compare the secondary structure predictions to the secondary structure you obtained under sequence details on the PDB pages. How closely did prediction match reality? How much trust will you place in secondary structure predictions of proteins for which the 3D structure is not known?

4. What percentage of the amino acids were in the favored or allowed regions of the Ramachandran plots? Are there any outliers? If so, which amino acids? Are you surprised by this? Why or why not. If your structure had no outliers, which amino acids might you expect to have disallowed dihedral angles? Why?

5. What have you learned about your enzyme so far? Does it have more than one chain? Does it contain a metal ion? Can you, just by looking at the various pieces of information, give an educated guess as to what part of the enzyme carries out the chemistry? Tell me as much as you can about your enzyme. You may use information discovered in the information lines of the various programs that you have used, but use only the information that you have gathered for the completion of this assignment. Don’t go anywhere else.

6. What don’t you know about your enzyme?

What must be handed in:
1. Printouts of your Protein Data Bank (Summary Information, Sequence details, MolProbity) and Entrez Genpept reports.
2. Printouts of at least two different secondary structure predictions using different algorithms, each with the address of the site used.
3. A printout of the hydrophobicity plot with the address of the program used to generate it.
4. A printout of whatever other information you obtained about your enzyme from web-based tools.
5. At least two printouts of 3D images of your enzyme taken from different angles and in different styles (these need not be in color).
6. A printout of a 3D image of your enzyme with a surface residue distinguished somehow, and identified as to amino acid and number (you can write on the printout).
   **That makes a total of 11 different printouts.**
7. Answers to the above questions.

**Project Grading**

*** Print out the last page of this document, fill in your group members last names and your enzyme name and include it as the final page of your report. ***

It denotes the point values for each part. **I will use this for grading and return it to you.** If I am sufficiently impressed with your report, there could be a few bonus points.

Don’t forget to hand in your peer evaluations with your report! Missing evaluations will result in point deductions for that member. You may submit these individually or as a group if you are able to reach a consensus.
Pet Enzyme Report I (40 points possible)

Biochemistry I

Grading Sheet for Group members:

Enzyme Name: ____________________________________

Printouts:

_____ PDB Summary Information (1 point)

_____ PDB Sequence information (1 point)

_____ PDB MolProbity (1 point)

_____ Entrez Genpept Reports (1 point)

_____ Secondary structure prediction I (2 points)

_____ Secondary structure prediction II (2 points)

_____ Hydrophobicity Plot (2 points)

_____ Web tools information (2 points)

_____ Structural printout I (2 points)

_____ Structural printout II (2 points)

_____ Structural printout with labeled amino acid (2 points)

Questions:

_____ Question 1 (5 points)

_____ Question 2 (3 points)

_____ Question 3 (3 points)

_____ Question 4 (3 points)

_____ Question 5 (5 points)

_____ Question 6 (3 points)