

Molecular Biology Through Discovery

Problem Set 3: Protein structure and function¹

Submit reasoning and process, not merely answers!

- 3.1.** An enzyme has a molecular weight of 60,000 daltons. When it is exposed to detergent, the protein breaks up to identical inactive components with molecular weights of 20,000 daltons. If the detergent is removed by dialysis, the 60,000-dalton protein reforms and regains enzymatic activity. You have isolated two mutant proteins. Mutant 1 shows no enzymatic activity and has a molecular weight of 20,000 daltons whether or not detergent is present. Mutant 2 has a molecular weight of 60,000 without detergent and 20,000 with detergent but shows no enzymatic activity in either case.
- a. Draw a diagram that captures the important elements of what you've just read. **Provide the drawing, labeled.**
 - b. Suggest defects to explain the behavior of each of the mutant enzymes. **Provide an explanation tied to specifics of your diagram.**
 - c. A person is heterozygous for Mutant 2 (i.e., has 50% Mutant 2 polypeptide and 50% normal polypeptide). How would you explain an observation that the person has 87.5% of the enzymatic activity of a normal person? How would you explain an observation (in a parallel universe) of 12.5% activity? **Provide an explanation tied to your diagram and accounting for these specific numbers.**
 - d. Ascribe the terms "dominant" or "recessive" to the mutation leading to Mutant 2, according to the two situations presented in Problem 3.1c. **A sentence or two should do.**
- 3.2.** In Problem Set 2, problem 2c, you considered a patch of hydrophobic amino acids that appears on the surface of a single polypeptide chain of the lactate dehydrogenase (LDH). You may have come up with a hypothesis as to why that patch is present. Let's see if you're right, by looking at the experimentally determined structure of the enzyme
- a. Find a Protein Data Base ID for the structure of human LDH.² You'll find several candidate structures. Avoid those that have extra compounds or refer to domains or individual chains, because they will show only part of the protein. **What is the PDB ID you chose?**
 - b. Take the PDB ID you found above and use it to display within Chimera² the structure of human LDH (quick reminder: **File / Fetch Structure by ID / type PDB ID / Fetch**). How many polypeptide chains does LDH have? Recall the trick from the [Perutz companion](#) of using the preset ribbon structure. Here's another trick: Mousing over any portion of the protein will bring up a small box labeling the amino acid. The last letter in the label is the name of the chain. A third trick: **Select / Chain** will bring up the possible chains you can select. **Which chains are most closely associated with each other?**
 - c. Now to look for hydrophobic patches of a single polypeptide chain, let's say Chain A. First step is to display only the A chain (the procedure seems needlessly complicated to me, but I don't know how else to do it):
 - i. **Select / Select All.** A green line will encase all of the chains.
 - ii. **Actions / Ribbons / hide.** Everything will disappear except a few molecules. If you mouse over them, you'll see things like GOL (glycerol) and IPE (pentaethylene glycol). The latter was added during the protein crystallization process (for reasons that are beyond me) and can be ignored by...
 - iii. **Actions / Atoms/Bonds / Hide.** The screen should now be blank. Now to bring back just Chain A...
 - iv. **Select / Chain / A.** The screen will remain empty
 - v. **Actions / Atoms/Bonds / Show.** You should now see just the atoms of Chain A, encased in an eerie green line indicating selection.

¹ See the *Structure and Function of Proteins* topic page for links to possibly useful resources

² The [companion](#) to Perutz et al (1965) tells you how to do this.

- vi. **Actions / Atoms/Bonds / sphere.** This gives you Chain A with each atom represented by a space-filling sphere.
- d. There are a lot of red balls. What are they? Investigate by mousing over several of them, particularly those near the periphery. **How are they labeled?**
- e. You'll probably find lots of atoms labeled "HOH", i.e. water (H₂O). You may not have any water, depending on the version of LDH you chose from the Protein Data Base. If you found water, it from the protein by:
 - i. **Select / Residue / HOH.** You should now see just the green lines encasing some of the red balls.
 - ii. **Actions / Color / Cyano.** Now all the atoms from water should be a pretty blue-green.

Rotate the protein. **Is water (the blue-green balls) distributed randomly?**

- f. Now to color the amino acids by hydrophobicity. We'll use the convention:
 - Hydrophobic – dark gray
 - Negatively charged – red
 - Positively charged – green
 - Otherwise polar – yellow

The strategy will be to color everything dark gray, and then amino acid by amino acid change the colors as needed.

- i. **Select / Residue / standard amino acids.** The now familiar green line will outline the entire protein.
- ii. **Actions / Color / dark gray.** Everything will turn gray except for the water.
- iii. **Select / Residue / X** (where X is one of the negatively charged amino acids). You can look up which amino acids are which in your favorite amino acid chart, which might be the one linked to the [notes on protein](#).
- iv. **Actions / Color / red.** Several of the atoms will now appear as groups of red balls, the sites of the amino acid type you selected.
- v. Repeat Steps iii and iv with any other negatively charged amino acid.
- vi. Repeat Steps iii through v but selecting positively charged amino acids and coloring them green.
- vii. Repeat Steps iii through v but selectin polar, uncharged amino acids and coloring them yellow.

You should now have a very colorful Chain A. (**A cropped screenshot would be helpful**) The gray amino acids are hydrophobic, and the others are polar, possibly charged. **Do you detect any patterns of coloration?**

- g. Here's a strategy that might help. Make the B Chain visible so you can see where they interact, then make it invisible again so you can get a closer look at the interacting region. Here's how:
 - i. **Select / Chain / B.** You won't see any visible change.
 - ii. **Actions / Ribbon / Show.** Now you should see Chain B as a ribbon structure you can partially see through.
 - iii. Rotate the protein so that you can get a better idea where is the site of interaction. Then rotate it so that you're looking through the ribbon to the site of interaction behind it. Now to make Chain B disappear...
 - iv. **Actions / Ribbon / Hide.** You should now be looking at the site of interaction.

Provide a cropped screenshot, with the site of interaction circled. Now that Chain B has directed your attention, **anything peculiar about that site relative to the rest of Chain A?** Are there any other regions of Chain A that is peculiar in a similar way? If so, **provide a cropped screenshot, with the peculiar site circled.**

- h. What is a region of interaction between Chain A and Chain D? **Provide a cropped screenshot, with the site of interaction circled.**

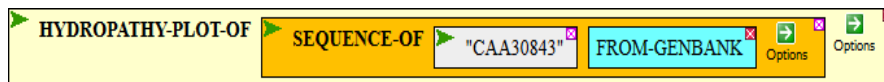
- i. With all this in mind, answer Problem Set 2, number 2c again: **Why does a single polypeptide chain of LDH possess one or more hydrophobic patches on its surface?** Of course you will make use of all the evidence you can, based on the work you've just accomplished.

3.3. It is currently very cheap and easy to synthesize DNA according to any sequence you specify. It is also easy to express into protein and suitable DNA you synthesize. Since we know exactly how proteins are determined by DNA, it is possible to manufacture any conceivable protein. In principle, this ability should revolutionize organic chemistry. Just make an enzyme that will catalyze whatever organic reaction you want. Imagine the cost of synthesizing pharmaceuticals dropping to near zero!

Unfortunately this is all fantasy, because a key unmentioned step – predicting the structure and function of a protein from its sequence -- remains beyond us. You'd think at least part of the problem shouldn't be so difficult. Look at Figure 5 of the [notes on protein](#), showing the passage of the protein glycoporphin through a membrane. This seems utterly predictable. The 20 amino acids in the transmembrane region are on average highly hydrophobic, as you'd expect. We should be able to look at the amino acid sequence of a protein, find regions of comparable hydrophobicity, and predict on that basis that the region lies in the membrane. Let's try it.

3.3a. Predict the membrane placement of glycoporphin: BioBIKE provides a program that will calculate the average hydrophobicity of protein regions and display the calculations graphically, on what is called a hydropathy plot.

- Go to CyanoBIKE
- From the black **File** menu, click **User Contributed Stuff** and click the link to bring in the function HYDROPATHY-PLOT-OF
- Go to the blue **Functions** menu and click HYDROPATHY-PLOT-OF, to bring the function into your workspace. This function calculates the average hydrophobicity of 15 (default) amino acid-regions along the length of a given protein.
- Give the function the protein sequence of glycoporphin (step 1). Do this by clicking *protein-sequence* and then from the green **Genes-Proteins** menu, click SEQUENCE-OF.
- Give the function the protein sequence of glycoporphin (step 2). Click the *entity* box of SEQUENCE-OF and enter "CAA30843" (in quotes). Press **enter** as usual to close the box. "CAA30843" is the accession number for glycoporphin in the [National Center for Biotechnology Information \(NCBI\)](#) GenBank protein database. You can visit the page and learn more about the glycoporphin sequence by going to NCBI and entering the accession number in the search box.
- Give the function the protein sequence of glycoporphin (step 3). From the SEQUENCE-OF options menu, select **FROM-GENBANK**, and click **Apply**. Then execute the function from the green wedge action menu. The function should look like this:



- You should now have a hydropathy plot, ... what is it? There's obviously something special happening around amino acid position 100. Compare the graph to the display of glycoporphin in Figure 5 of the [notes on protein](#). Are there hydrophobic amino acids around position 100? **Describe (appealing to specific coordinates) how the figure and the graph are or are not in accord.**

3.3b. Predict the membrane placement of glycoporphin (troubleshooting): What's going on? Maybe the sequence from GenBank isn't the same as the one in the figure? You can see the sequence in the figure. Get the sequence from GenBank by executing the SEQUENCE-OF function (click **Execute** from the green wedge action menu of SEQUENCE-OF, not HYDROPATHY-PLOT-OF). **Are the sequences the same? Is the beginning the same? Is the end the same? Is the number of amino acids the same? Give specifics.**

- 3.3c.** Predict the membrane placement of glycoporphin (troubleshooting, part 2): What's going on? Why should the two sequences differ? Focus on the exact region that's different and if you can't think of an explanation, seek outside help: the [notes on protein](#), the glycoporphin GenBank page, other resources, people, anything, but find an explanation. **How do you explain why the sequence shown in Figure 5 differs from the one obtained from GenBank?**
- 3.3d.** Predict the transmembrane regions placement of glycoporphin (bottom line): From the sequence of glycoporphin and the hydrophobicity plot, predict which amino acids span the membrane. **Provide the plot and a narrative that appeals to specific amino acid coordinates. Give the identity of the amino acids you predict span the membrane.**

Extra – in case you have the time and inclination

- 3.4.** Plants and photosynthetic bacteria have been enjoyed enormous success in part because of their ability to harness sunlight to power the reduction of CO₂ to sugar. Organisms that can use sunlight as to drive the reduction of N₂ to biologically useful nitrogen compounds are far more rare. One reason for this is that the enzyme that catalyzes the nitrogen reduction is extremely sensitive to O₂, greatly limiting the environments in which nitrogen fixation can take place. Some have said that this is just the way it is – it is not possible for an enzyme to fix N₂ without also being killed by O₂. Why? Well if such an enzyme could exist, it would have arisen sometime over the last 4 billion years of organismal evolution and organisms with this capability would have taken over the world!

Let's examine this reasoning. Have all possible proteins been tried sometime during the lifetime of the earth? If not, then what is a reasonable estimate for an upper limit on what fraction has arisen?