Problem Set 5   (**Due Sept 30ᵗʰ**)

1. Draw the Hawthorn projection (ring form) of Mannose and its C2 epimer. What is the common name of this epimer?

2. Examine the sugar shown here.
    a. Is this an aldose or ketose? Pentose or hexose?
    b. Label each carbon appropriately (C1, C2, etc.)
    c. From the image, how can you be sure that it's an epimer of glucose?
    d. Which sugar is it?
    e. Draw the sugar using a Hawthorn projection and a Fisher projection.

3. Sucrose is a common disaccharide. What is the full name of this sugar?

4. What is meant by the term reducing sugar? Which of the common disaccharides discussed in class are not reducing sugars? Justify your answer. What experiment can you do to determine if a disaccharide is reducing?

5. Using Chimera or another 3D molecule viewer, open pdbID 4B7I (the last letter is eye, not el).
    a. Verify that this is an N-linked oligosaccharide. What is the residue number of the glycosylated Asn?
    b. Does the protein follow the correct amino acid pattern for glycosyl transferase specificity? Be specific – you book discusses three invariable features of N-linked sugars.
    c. Using a sketch like seen in Figure 8.19, map the sugars. For each glyosidic bond, determine the linkage (e.g. $\alpha(1\rightarrow4)$). Hint: To determine the α/β orientation, you can set the focus to each sugar and observe if the anomeric carbon orientation is axial (α) or equatorial (β).
    d. Are there any H-bonds between the oligosaccharide and amino acid side chains?
    e. Draw this oligosaccharide using a Hawthorn projection. Make sure to show the linkage to Asn.

6. Discuss the differences between starch and cellulose. Make sure to consider the 3D structure of the polymers and how each interacts with other polymers.

7. Bacterial cell walls are made from repeating patterns of two modified sugars. Draw these sugars as they would look in a cell wall. Discuss how linear polymers are connected in a cell wall.

8. Draw a lipid containing stearic acid, linoleic acid and phosphatidylcholine. Recall that there are common positions on the glycerol backbone for each of these classes of molecules. Identify which regions of this molecule will occur on the surface of a lipid bilayer.

9. Why are triglycerides not common components of biological membranes? What is their function?

10. Describe two trends in membrane fluidity. How does cholesterol influence membrane fluidity?

11. The structure of the photosynthetic reaction center from *Rps. viridis* has been determined (pdbID 1PRC). Please download this structure and use an appropriate program to show how this protein is embedded in a membrane. A good place to start would be to color all polar and ionic amino acids one color and the non-polar amino acids another color. You can select types by Select→Residue → amino acid category.

12. This exercise will guide you through the last (and most useful) method to predict protein structure and function. There are A LOT of protein structures known; biochemists use this to their advantage. If we're trying to predict the structure of an unknown protein, we can commonly find the structure of a
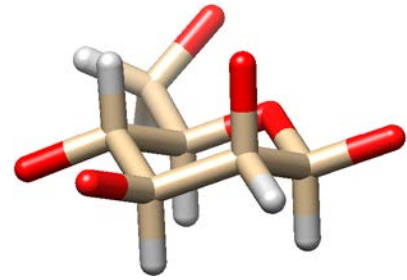
similar protein and use that as a guide to calculate a homology model. I stress that this is only a model and not a real structure, but nonetheless, it can be incredibly useful.

On the last page of this problem set, you'll find two FASTA sequences: one for Chain A of a known crystal structure and another from an E. coli protein that has unknown function but has a high degree of similarity with the known protein.

The first step to developing a homology model is to align the sequences. When you do a blast search, this is exactly what is happening behind the scenes; a computer is taking your protein (or DNA sequence) and aligning it with all known protein to find the best matches. We're going to do this manually for these two proteins.

- From the ExPASy page, select genomics then sequence alignment. As you can see, there are a lot of options – I like to use the ClustalW2 server, but you can try any of them.
- On the ClustalW2 site, input the sequences seen below.
- On the step 3 box, click on *more options* and select Pearson/FASTA as the output format.
- Submit this project and wait for the alignment to complete. This should be very quick.
- In this example, the sequences align very nicely, so there are not any gaps in the core region. Do note, however, that there are some dashes at the beginning and end of the unknown protein – this means that there are not any amino acids that align with the crystal structure at this point.
- Download the alignment file and paste it into a word document. Save this document as at .txt file (this is really important – Chimera won't recognize the .doc or .docx format as an alignment file)

Open this file in Chimera (make sure to select aligned FASTA as the format).

Ask Chimera to load any structures that are available. (Structures → Load structures). I would recommend that you hide chain B and set the focus on Chain A (select all of Chain A and then Action → focus).

Use Chimera to predict the secondary structural elements for the unknown protein (Structure → Secondary Structure). **Does the structure prediction match up with the actual structure of 3N73?**

Use Chimera to determine the % Identity. This can be found under the Info menu. You may have to manually open the reply log to see the value. This is available under the Favorites menu.

Now we're actually going to build a homology model of the unknown protein using 3N73 Chain A as a template. Go back to the sequence alignment tab.

- Select Structure → Modeller tools.
- The target will be the unknown protein and the template will be 3N73_A. Make sure these are selected.
- Now select ok. If you are asked to input a code for Modeller, use MODELIRANJE
- Wait for the sequence to complete.

Chimera will build and display multiple models. When we show several structures on top on each other, this is called a structure ensemble. Carefully inspect the ensemble – **what regions of the protein appear to be the most similar to the template structure? What regions tend to deviate? Does this make sense based on what we've learned about protein structure?**

There is one loop region on the exterior of the protein that seems to have the most deviation from the template. Justify this based on your sequence alignment.

Please provide an image of the homology model. You can either take a screen shot or use the save image command in the file menu.

>3N73:A|PDBID|CHAIN|SEQUENCE
MTNIQKRFYKGRVALNVLANNIENAKDIFEAAEGYVVVGVLSKDYPTVEEAVTAMKAYGKEIEDAVSIGLGAGDNRQAAVVAEIAKHYP
GSHINQVFPSVGATRANLGGKDSWINSLVSPTGKVGYVNISTGPISAAGEEKAIVPIKTAIALVRDMGGNSLKYFPMKGLAHEEEYRAV
AKACAEEGFALEPTGGIDKENFETIVRIALEANVERVIPHVYSSIIDKETGNTKVEDVRELLAVVKKLVDQYA

>gi|571159785|emb|CDL57015.1| putative periplasmic protein [Escherichia coli ISC56]
MKLTPNFYRDRVCLNVLAGSKDNAREIYDAAEGHVLVGVLSKNYPDVASAVADMRDYAKLIDNALSVGLGAGDPNQSAMVSEISRQVQP
QHVNQVFTGVATSRALLGQNETVVNGLVSPTGTPGMVKISTGPLSSGAADGIVPLETAIALLKDMGGSSIKYFPMGGLKHRAEFEAVAK
ACAAHDFWLEPTGGIDLENYSEILKIALDAGVSKIIPHIYSSIIDKASGNTRPADVRQLLEMTKQLVK