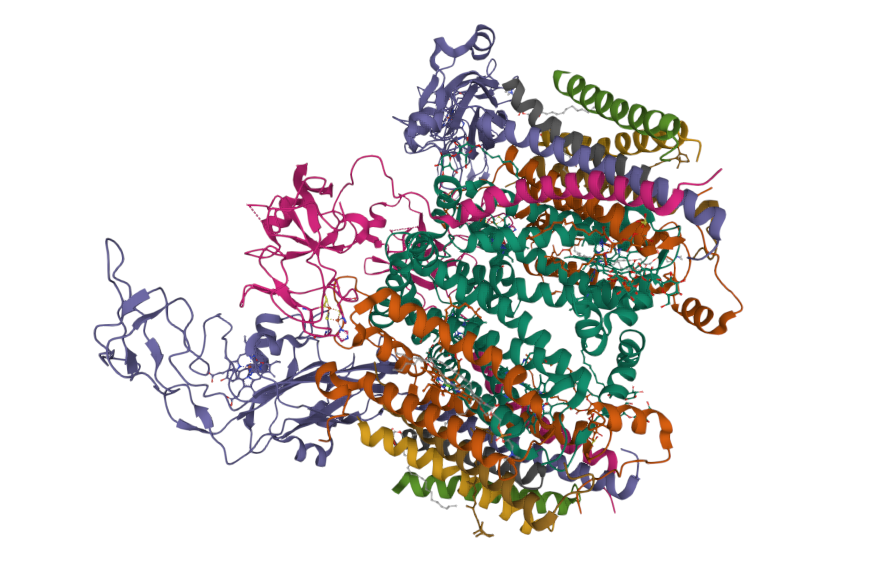
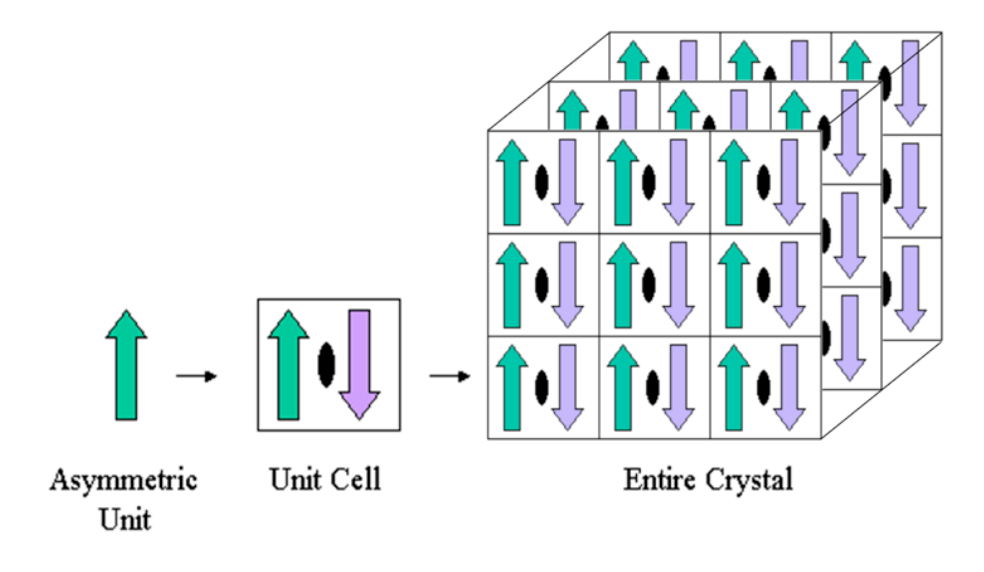
Vincent Zhang

Bio Computer Assignment #6

**Part I: Protein Data Bank (PDB)**

1. What is the biological function of this protein (2ZT9)? From what species was this protein taken?
   1. This protein, called Cytochrome b6f Complex, is a membrane protein enzyme found in the chloroplasts of plants, and is responsible for mediating the transfer of electrons and of energy between two photosynthetic reaction center complexes (is directly related to photosynthesis).
   2. This protein was taken from the Nostoc species.
2. What is the reference paper for this structure? Briefly describe the findings of the paper
   1. The reference paper is called *Structure-Function, Stability, and Chemical Modification of the Cyanobacterial Cytochrome b6f Complex from Nostoc sp. PCC 7120* by Baniulius, et al.
   2. This paper talks about the crystal structure of the protein in question, and how the protein taken from the Nostoc has a stable dimeric structure, and strong electron transport activity. It also talks about its dominant conformation of heme b(p) structure, and the acetylation of PetC.
3. How many protein chains form the structure? How many amino acid residues are in each chain?
   1. There are 8 unique protein chains.
   2. 215 for the first, 160 for the second, 289 for the third, 179 for the fourth, 31 for the fifth, 34 for the sixth, 37 for the seventh, 29 for the eight.
4. How many ligands does this structure have? Name three of the ligands.
   1. There are 14 ligands
   2. Some of these ligands are called:
      1. [2ZT9\_HEM\_A\_303](https://www.rcsb.org/3d-view/2ZT9?preset=featureDensity&label_asym_id=K&auth_seq_id=303)
      2. [2ZT9\_HEM\_A\_301](https://www.rcsb.org/3d-view/2ZT9?preset=featureDensity&label_asym_id=I&auth_seq_id=301)
      3. [2ZT9\_HEM\_A\_302](https://www.rcsb.org/3d-view/2ZT9?preset=featureDensity&label_asym_id=J&auth_seq_id=302)
5. At the left corner of the window, you see a protein structure; the first structure displayed is called “Biological Assembly 1”. Under the picture click the tab “3D View: structure”.
   1. A new page should open with a structure of the protein in 3D that you can manipulate; play with the options to the right of the picture and insert a screenshot of the protein:
   2. Go back to the main page and repeat the above by choosing the “Asymmetric Unit'' option. What is the difference between the Biological Assembly and the Asymmetric Unit?
      1. An asymmetric unit as opposed to the whole biological assembly is the smallest portion of a crystal structure to which symmetry operates:The follow diagram helps illustrate how the asymmetric unit compares to the biological assembly it is a part of:
6. What was the experimental method used to obtain the 2ZT9 structure? What is the “resolution” of this structure? What is the meaning of the “resolution”? What protein structure has higher quality: One with “high resolution” or low “resolution”?
   1. The experimental method used to obtain the 2ZT9 structure is X-ray diffraction.
   2. The resolution of this structure 3.00Å.
   3. Resolution is a way of measuring how “high quality” we are able to observe the actual structure of the protein we are trying to anaylze. For example, when resolution is low (meaning high “resolution value”, such as the structure we observe right now), there is a lot of diffraction that does not scatter the same way because the structure is not uniform enough, leading to a lot of “blurriness” in the diffraction patterns we perceive. However, if the resolution is high, meaning low “resolution value”, then it means the crystal pattern of the protein is more uniform, making it easier to see the actual structure of the protein.
   4. Thus, a protein structure has a high quality and it has a lower “resolution value”.
7. Can you think of a reason (there are multiple valid answers) as to why the resolution-value increases as protein size increases? (Do not forget that the smaller the resolution-value, the better the power of resolving two atoms from each other).
   1. As explained in the pdb document on, when the crystal structure is more uniform, the diffracted light bounces in uniform ways that allow us to “see” the structure better. The larger the protein, the less uniform it becomes, and so the quality of the data we see decreases since the proteins are aligned less perfectly the larger it gets. Thus, the resolution value “increases”, which means it quality decreases.

**Part II: Introduction to VMD**

(remember: teal corresponds to carbon atoms, red to oxygen, blue to nitrogen, yellow to sulfur)

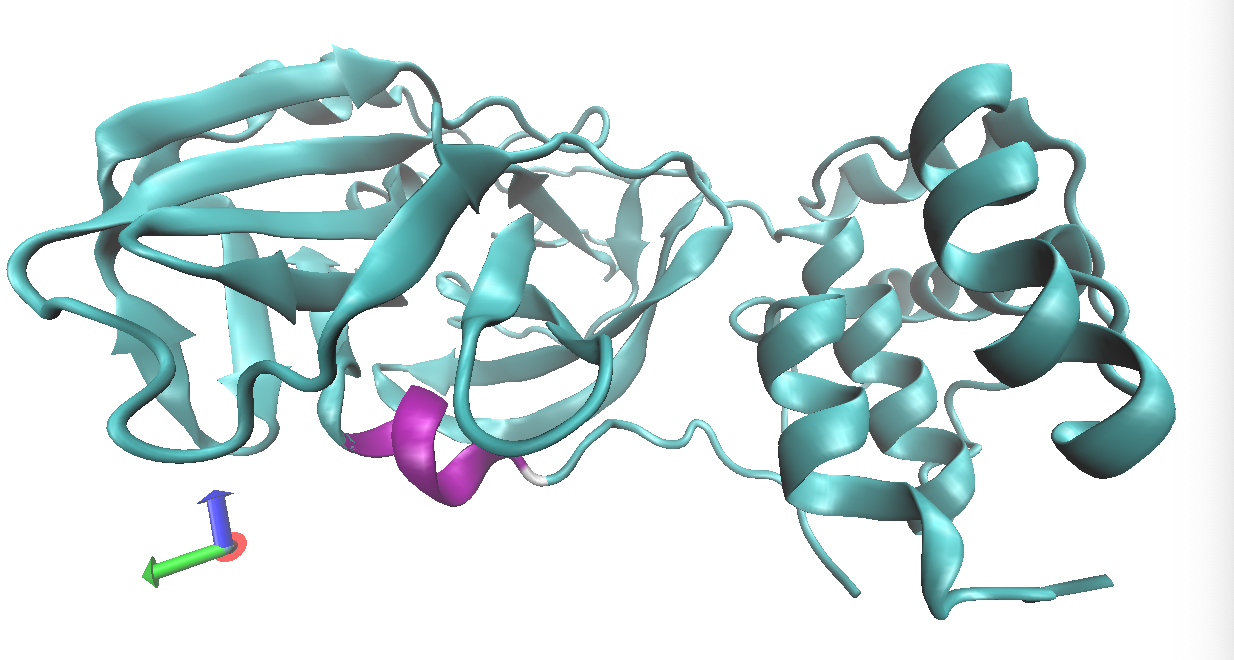
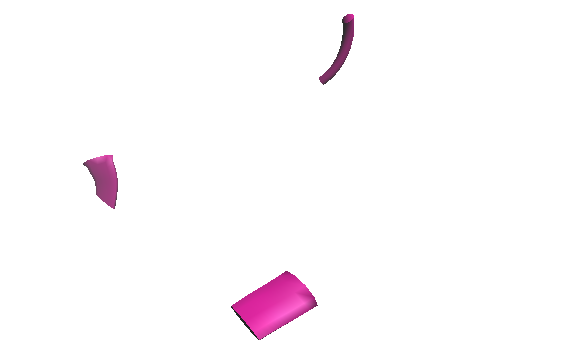
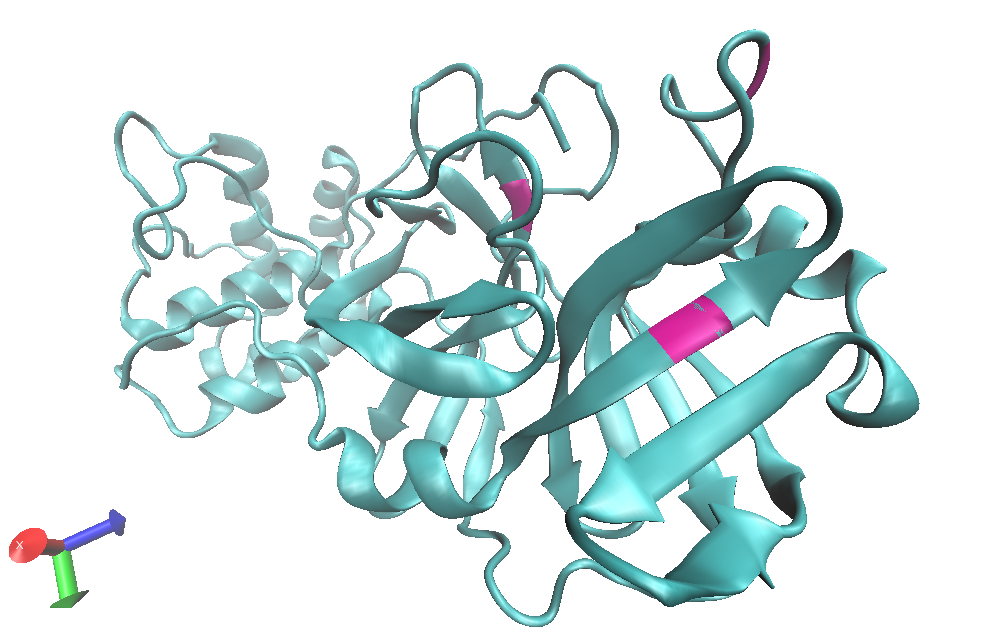
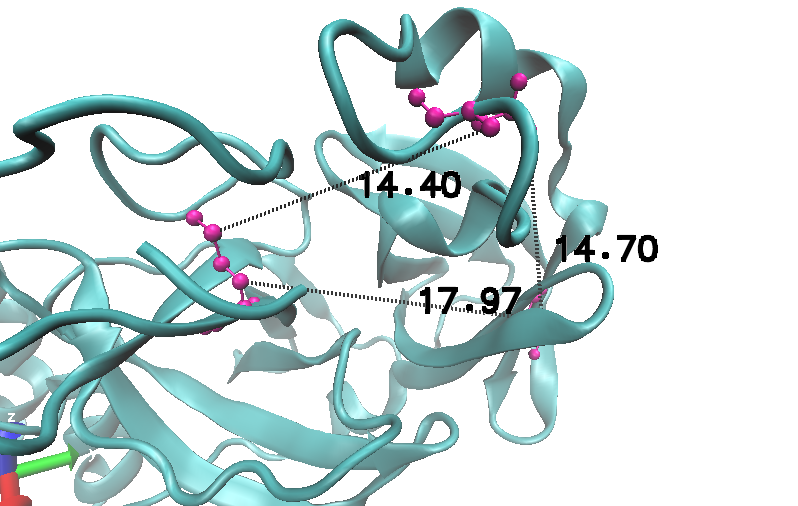
1. Fill in the blanks in the following table by finding the residue names of the next three residues. Note that the R-group may be a hydrogen atom, as in glycine.

| Residue Number | Amino Acid |
| --- | --- |
| 1 | SER |
| 2 | GLY |
| 3 | PHE |
| 4 | ARG |

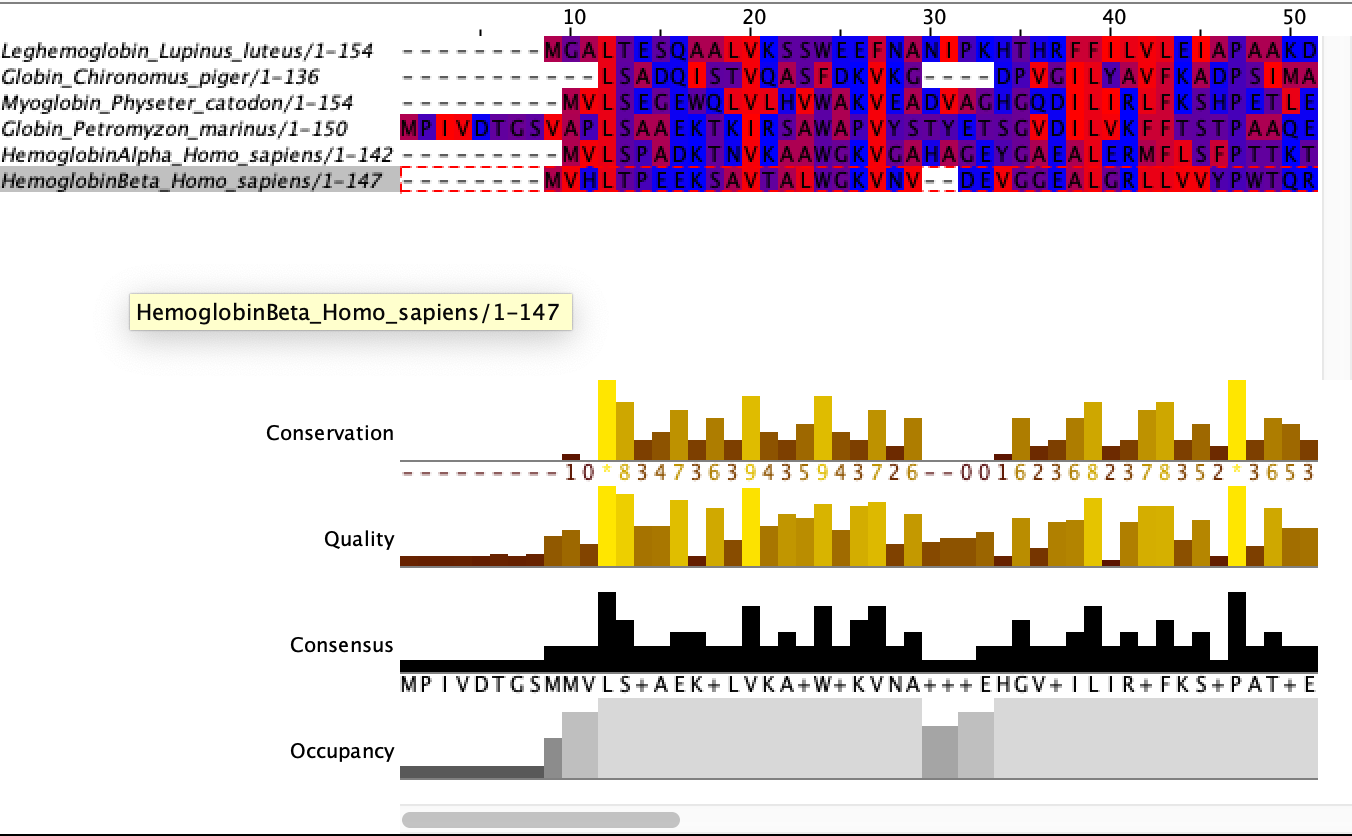
1. Use the Sequence Viewer to find the residue numbers and names of the last five amino acids in the COVID protease (NOT the inhibitor).

| Residue Number | Amino Acid |
| --- | --- |
| 302 | GLY |
| 303 | VAL |
| 304 | THR |
| 305 | PHE |
| 306 | GLN |

* 1. I am not too sure what the lab is talking about between the inhibitor and the protease, but I assumed that the inhibitor would be the last three amino acids with residue numbers 2, 3, and 4.

1. The first non-coil secondary structure is an alpha helix
   1. Which residue number does the helix start at, and what amino acid is it?
      1. It starts at residue number 10, and its SER.
   2. What is the ending residue number, and what amino acid is that residue?
      1. It ends at residue number 15, and its GLY.
   3. Select the alpha helix the Sequence Viewer. Amend the representation of this alpha helix by using “**Secondary Structure**” as the coloring method and “**NewCartoon**” as the drawing method. Include a screenshot of this alpha helix along with the rest of the protease protein
2. Try selecting protein residues 21, 48, and 165 at once using the Sequence Viewer. Find the representation created by the Sequence Viewer for these residues
   1. What is typed into the “**Selected Atoms**” box to give you these residues?
      1. (chain A and resid 21 48 165)
   2. What is the drawing method that the Sequence Viewer is using?
      1. New Cartoon
   3. Include a screenshot of the *protein* shown as **new cartoon** with these selected residues shown clearly. Note that, as mentioned before, there is also the *N3 inhibitor* present. Therefore, in this same screenshot, the N3 inhibitor should separately have the style DynamicBonds. Also, set the inhibitor’s coloring method to ColorID, and then pick a color that contrasts clearly against the protease.
3. Measure the distance between any two atoms
   1. Insert a picture of the measurement that clearly shows the distance. You might need to modify the color for this label. Refer to prior instructions on how to do it. We recommend using Black as your new color.
   2. Drawing on your previous knowledge, what unit do you think these measurements are in? If you are not sure, this information might be found by looking at the PDB page for 6LU7 (think about how the structure was obtained).
      1. These units are in Å, or the same units as when when we did the resolution stuff. 1Å represents 100p, and the reason why I think this make sense is that the average length of a bond between two molecules is 1-2Å, and since these molecules seem to be pretty far from each other, I would say that it makes sense for them to be 10-20Å away from each other.
4. Again, find the first Serine residue on the molecule (**type protein and "resid 1 2" in the VMD selection window; then push the equal sign to center these residues in your screen; remember to hide any other representations to reduce clutter**). From this residue, follow the molecule with your mouse (from N-terminal to C-terminal) until you reach the atom to which an oxygen molecule is bonded. This is the carbonyl carbon, and the preceding carbon is the alpha carbon. Note that in the labels, the carbonyl carbon has “C” after the colon, while the alpha carbon has “CA”. You may find the “lines” or “CPK” drawing methods helpful. Measure the following bond distances (No picture is required but units are):
   1. Amide N to alpha C: 1.46Å
   2. Carbonyl C to alpha C: 1.52Å
   3. Carbonyl C to N in the peptide bond: 1.33Å
5. Repeat this exercise for the same bonds in residue 25 (hint: in the VMD screen just show residue 24, 25, and 26; no picture is required):
   1. Amide N to alpha C: 1.46Å
   2. Carbonyl C to alpha C: 1.52Å
   3. Carbonyl C to N in the peptide bond: 1.33Å
6. Of the three types of bonds we measured, which type of bond appears to be the shortest? Can you think of the reason why this is the case?
   1. The peptide bond between the carbonyl C and the N seems to be the smallest. I think one possible answer for this is that the functional group can often be large, and when they’re large their elections repel more and caused the bond length to be bigger closer to the functional group, but the peptide bond between two amino acids is the most “loose”, where there is more available space, allowing for the bonds to be closer.
7. Describe the function of the COVID-19 main protease and briefly explain the purpose of the synthetic N3 inhibitor. (Hint: check the PDB101 page; you can google the answer!).
   1. The function of the COVID-19 main protease is to help its own maturation. The purpose of the synthetic N3 inhibitor is to form a complex with the main covid protese and control its infection.
8. Go to the PDB page for the 6LU7 protein, at left corner of the window you see a protein structure; the first structure displayed is called "Biological Assembly 1". If you click on the arrow above this picture you will see another picture called "Asymmetric Unit". What is the difference between the Biological Assembly and the Asymmetric Unit?
   1. An asymmetric unit as opposed to the whole biological assembly is the smallest portion of a crystal structure to which symmetry operates

**Part III: Multiple Sequence Alignment with ClustalOmega**

1. There are three symbols that appear below alignment columns, which represent three different qualities of the column: sites that are entirely conserved as a single amino acid, sites in which all sequences have amino acids with strongly similar biochemical properties, and sites in which all sequences have amino acids with weakly similar biochemical properties.
   1. Which symbol corresponds to which class?
      1. \_\*\_ corresponds to amino acids that are entirely conserved
      2. \_:\_ corresponds to amino acids having strong similar biochemical prop.
      3. \_.\_ corresponds to amino acids that have weakly similar biochemical prop.
   2. How many sites are completely conserved across the alignment?
      1. Only 5 sites are completely conserved across the alignment
   3. What does this imply about the functional importance of these sites?
      1. That implies that these particular sites are functionally extremely important, as they are shared between all of the globins, and thus must be something essential to making the globins what they are, or at the very least essential to their survival/function.
2. Consult the [UniProt](https://www.uniprot.org/) page for human hemoglobin subunit beta (the UniProt accession number is P68871). Under the "Sites" annotation, this page lists two amino acids that participate in the coordination of an iron ion and binding oxygen (designated "Metal binding"). Determine the residue numbers of these amino acids and find their corresponding columns in the alignment. What are the identities of these amino acids in hemoglobin beta and the other globins in this alignment? What might this imply about functional conservation in this family?
   1. The residue numbers of both of these amino acids are simply 1
   2. Their corresponding columns in the alignment are 64 and 93 respectively
   3. The identity of of these amino acids for the iron (heme ligand) at 64, for example is V,S,E,Q,L,A, and at 93 it is G - - - - -. What this suggests is that since all of the have different identities in general, it may not be that important for the functional conservation in this family.
   4. What this might imply about the functional conservation in this family of globlins is that these iron ligands are core into the function of the heme binding of them.
3. Scroll along the entire alignment and look at how well the hydrophobic/hydrophilic character of an amino acid position is maintained within an alignment column. What might this imply about the 3-dimensional structures of these six proteins? (Remember that generally speaking, hydrophobic amino acids are buried in the interior of a protein structure while hydrophilic residues are found on the surface.)
   1. Since hydrophilic residues are found on the surface, and hydrophobic residues are found mostly in the interior of a protein, the fact that the colors along columns mostly align (meaning that if a residue is hydrophobic/hydrophilic for a globin, then it is more likely to also be hydrophobic/hydrophilic in other globins and follow the same pattern) suggests that the 3D structure of these six proteins are pretty similar.
4. Do you see any qualitative correlation between hydrophobicity and conservation? What does this say about where amino acid substitutions are best tolerated in a protein?
   1. I see a very subtle general trend in which the for columns that are very hydrophobic, they are also conserved very well. This suggests to me that amino acid substitutions are best tolerated in the interior of the protein, since hydrophobic residues are located in the interior.
5. Repeat this process for the other helices noted on the UniProt page. Paste a screen shot of the Jalview window spanning as much of the alignment as you can.
   1. 
   2. 