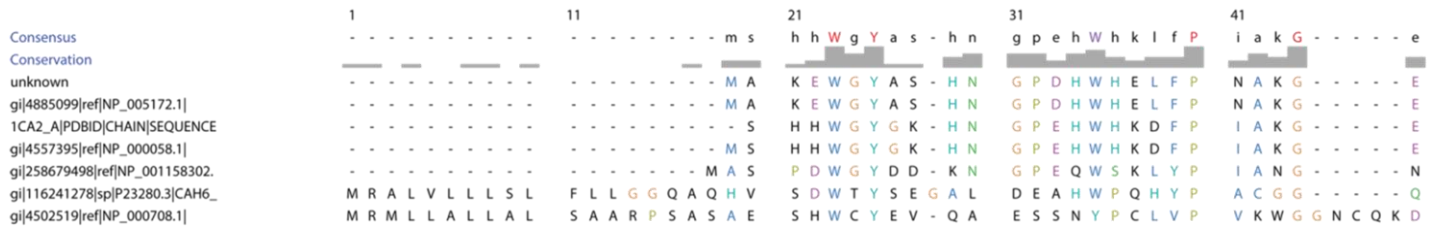
Problem Set 4a - Chimera

You think you have identified a previously unknown protein from your pet bonobo.

1. Using the sequence in the CHIMERA.txt file available online, determine if there is a homologue in *homo sapiens*
    a. This can be accomplished using a Blast Search. It takes the amino acid sequence and searches through other organisms to see if there are any similar proteins. The NCBI website is a great place to start for any bioinformatics related needs.
    b. http://www.ncbi.nlm.nih.gov/guide/
    c. Find the Blast link on the menu bar on the right side of the page.
    d. Select the appropriate Basic Blast link (remember this is an amino acid sequence)
    e. Paste the Bonobo protein sequence (including the >Bonobo) into the big textbox. This is the amino acid sequence that the server will use to try to find a similar protein.
    f. Make sure the Database is the Non-redundant protein sequences and enter Homo Sapiens (taxid:9606) into the Organism box.
    g. Now click blast.
    h. When the search is complete, you'll see several lines in the cartoon output. The color of these lines tells you about the degree of similarity – black is not similar at all, red is extremely similar. The scale is given to you at the top of the white box.
    **i. Please take a screen shot and comment on what class of protein this is. Do humans have a protein that likely has the same function?**

2. Import this sequence into Chimera as an aligned FASTA. (File → Open → select your file → aligned FASTA)
    a. Use Chimera to predict secondary structural elements (Structure → secondary structure → predict). The output indicates β sheets as green boxes and α helices as yellow/brown boxes.
    b. Save this image as an .eps file. This can be imported into Microsoft Word. Comment on the predicted structural elements.
3. The next step in a bioinformatics approach to learning about this protein is to find several proteins that should be similar. In this case, we want to find several homologues in humans. We can then use one of numerous common websites to find sequence similarities – this is called a sequence alignment. I've taken the liberty of doing this for you – Alignment.clustal. Import this file into Chimera (make sure to select Clustal when asked the file type).




A great deal of information can be learned from this image. A capital letter in the **Consensus** row tells you that this amino acid is present in ALL sequences while a lowercase letter tells you that this is the most common

amino acid, but it is not present in all cases. The **Conservation** bars allow you to quickly determine regions of high conservation.

4. Using the first 100 amino acids, comment on the similarities between these 7 proteins. Are any of them more dissimilar than the rest?

5. Predict the secondary structural elements for each of these proteins. Export this as an .eps file and import it into a word document as above. Comment on predicted regions of secondary structural similarities. Again, comment on any outliers (are any more dissimilar than the rest?)

6. After examining this format, you can now ask Chimera to load any structures that are available. Fortunately, this sequence alignment was done with Chain A from Human Carbonic Anhydrase II (PDBid 1CA2). (Structures → Load structures).

By reading the primary literature about this family of Carbonic Anhydrase (alpha), you learn that four histidine residues play a critical role in the enzyme. Three are used to coordinate to a $Zn^{2+}$ ion that facilitates the catalytic cycle.

7. What residue number do these 3 histidines have? _____ _____ _____

8. The 4$^{th}$ important histidine is located on the other side of the active site. It is His-64 in the 1CA2 structure.
    a. To select this residue, open the command line. Sel :residuenumber.chain
    b. Show the residue (Action→Atom-bond→show)
    c. Now select the appropriate nitrogen on the imidazole ring (you can ctrl-click on image to select) AND the $Zn^{2+}$ ion (with the nitrogen still selected, ctrl-shift-click the Zn)
    d. You can learn a lot about this structure under Tools → Structure Analysis

    How far from the $Zn^{2+}$ ion is this residue? _____

9. This histidine plays an important role by deprotonating a water molecule that is coordinated to the $Zn^{2+}$. Is this residue close enough to do this chemistry directly? Assume that it needs to be within H-bonding distance with the water molecule, which is shown as a red dot bound to the Zn. A standard H-bond (X−H···Y) has distance values of X−H ~110 pm, and H···Y distance is ~160 to 200 pm.

    What is the distance between the histidine and the water? _____

10. So we need to find a pathway to facilitate this interaction. Select all atoms within a 4 angstrom zone of the $Zn^{2+}$ coordinated water. Select this water, go to the select menu → zone, change the value to 4 and OK. Now show those atoms. You should see two more solvent molecules appear. Our pathway is beginning to form. Select the solvent molecule that is in a closest to His-64 and repeat.

A standard H-bond (X−H···Y) has distance values of X−H ~110 pm, and H···Y distance is ~160 to 200 pm. Given these guidelines, verify that a hydrogen bonding network is available between His-64 and the water that is coordinated to the $Zn^{2+}$. Please provide an image that verifies this. The slab mode (side panel) would be very helpful for this as well as changing the color of the distance labels (under display options).

11. Does the bonobo protein have a histidine in the same position as His-64?

12. If not, use the mutate tool to verify that this other amino acid can participate in a H-bond network similar to the histidine. (Tools → Structure Editing → Rotamers → select the desired mutation in the drop down box). You'll get a list of possible conformations of this mutated side chain. Identify a rotamer that is in the appropriate orientation. What are the Chi1, Chi2, Chi3 and Chi4 values? Please provide an image of this mutation with distances to the closest water.

13. Now we're actually going to build a homology model of the bonobo protein using 1CA2 as a template. Go back to the sequence alignment tab.
    a. Select Structure → Modeller tools.
    b. The target will be the unknown protein and the template will be 1CA2. Make sure these are selected.
    c. Now select ok. If you are asked to input a code for Modeller, use MODELIRANJE
    d. Wait for the sequence to complete.

14. Multiple possible structures will appear – please hide all structures except the first modeled one. Zoom in on the active site. Does it appear that the Bonobo Carbonic Anhydrase has all the appropriate tools to catalyze the same reaction? Show an image to justify your answer.