

Chemistry 481 - Biochemistry

Bioinformatics Project: Visualization of Conserved Regions of Proteins

Goal: The purpose of this project is to introduce biochemistry students to some easily accessible protein data base tools and allow the students to explore these databases so they get a glimpse of what is available.

Overview

One of the most powerful bioinformatics tools for the study of proteins is the ability to search databases for similar sequences and to make multiple alignments of those sequences. Increased computing power in the last few years has resulted in these techniques being readily available to anyone with a PC and internet access. Study of these alignments for conserved regions can reveal much about proteins and also the relationship between the organisms (to be addressed in another bioinformatics exercise). In this exercise the protein myoglobin will be explored as this, along with hemoglobin, has been discussed in detail in the CHEM 481 lectures. Multiple myoglobin sequences will be lined up against human myoglobin to determine which residues are conserved across all species. These conserved residues will then be located on the human myoglobin structure.

Procedure

Part 1 Alignment

1. Log onto San Diego Supercomputer Center Biology Workbench by typing in the URL: workbench.sdsc.edu.
2. Set up a free account by clicking the register button. Simply follow the instructions, be sure to remember your user name and password.
3. Once logged into Workbench. Enter "session tools" and choose "Start New Session" from the dropdown menu. All the work you do will be organized in this session; you can keep the work on multiple topics separate. Select "run". You will be asked for a session description (name) then "start new session".
4. Select "protein tools" from the menu bar across the top of the screen. A list of tools appears. Scroll down and select "Ndjinn" and "run". This accesses a multi data base search.
5. Enter "myoglobin" as the chosen molecule and scroll down the list of databases to find "PDBFINDER" (it is in green just below a blue highlighted database). Return to the top of the page and "search".
6. A list of results will be displayed. Select the pdb for human myoglobin 2MM1 then click on "import sequence".
7. You will be returned to the Protein Tools screen. Select your sequence by checking its box then select "BLASTP" in the tools list. "Run".

BLAST (Basic Local Alignment Search Tool) searches through the data base for similar sequences. Use the default setting “Blosum62”. Chose the database “SwissProt” and change the expectation value from the default of 10 to 0.1 (this reduces the number of poor matches). The other settings are left at the default values. “Submit” (at the bottom of the page) will run the search.

8. The program will display a list of sequences identified by the BLAST search; the first sequence is the protein whose structure was used (the pdb file you selected in (6) above).

9. Select a number of the sequences from the search and “import sequence(s)”. I chose 35 sequences for my alignment and that gave me some nice data. Choose them at random from the list or go down and pick your favorite critters bearing in mind that you want to align a good cross section of myoglobin molecules.

10. You will be returned to the “protein tools” screen, “select all sequences” “run”. Scroll through the list of tools to find “CLUSTALW” and “run”. CLUSTALW is a multi-alignment tool; each sequence is aligned with every other sequence to give a best fit.

11. The screen will switch to allow a choice of options; the default settings work fine “submit” to run the alignment.

12. It may take a short while for the alignment to run especially if the server is busy. The alignment will be displayed with color coding. Blue is reserved for completely conserved residues, green is for residues that are strongly conserved ie the chemical properties are similar, dark blue shows weak conservation, black is for residues that show no consensus. Beware the mid blue and dark blue can appear very similar at least to my eyes fortunately there are asterisks at the bottom of the completely conserved residues.

13 Save the alignment by “import alignment” (at the very bottom of that long page). You will be in the alignment tools mode, select your alignment by checking the box on the left hand side of your sequences. Then scroll down the selection box to “texshade” “run” using the default settings, “submit” (bottom of page). Keep this window open or print this out on a color printer so you can refer back to it as you move onto the next step of the exercise. Note if the file is big you may have to run “boxshade” instead of “texshade”. The texshade/boxshade file has the same information as on the file in (12) above but can be easier to see as the residue columns are shaded rather than each letter being colored.

Note: if any of the sequences you chose are fragments you will need to remove these from the alignment as the program will not show consensus in the regions where there are no amino acids; I ran into trouble with a giant panda sequence that was not complete so it messed up the alignment in the missing areas.

Now you are ready to move on to the second part of this exercise where you will examine the human myoglobin structure and look to see where the completely conserved residues are found on this structure.

Part 2 Protein Structure Visualization

1. Enter the Protein data Bank by going to www.rcsb.org. Take a couple of minutes to explore the site; there is usually a “molecule of the month” on the front page.
2. Enter the PDB ID of the molecule of interest (2MM1) for human myoglobin into the search box at the top of the page. You will see details on the structure including a citation for the original crystal structure.
3. First go into KiNG viewer. It may be necessary to download an applet so that you can view the molecule; if so you will be guided through in a stepwise process. In the KiNG viewer window rotate the molecule and check out the position of the heme group; you can display or remove the heme group by selecting the “hets” check box on the right hand side.
4. Close out the KiNG viewer and open up WebMol. The drop down selection menu at the top on the right hand side allows you to change the view. Select “backb” to show just the backbone of the molecule.
5. Click the “select” button, a second window will open up. Change the method from select to SC vs BB. The select window shows a number of things; the sequence is displayed and to the left of this the helical regions are shown (H).
6. Choose one of your conserved residues and find where it is sited on the molecule; to do this make sure you have the residue number from the human myoglobin. Change color from current to “yellow” (this shows up well but you can choose any color that contrasts well.) Highlight the residue name and “apply”. The residue should appear on the molecule. If you now change the top right hand drop down to “AllAt” the side chain should be seen. Changing the linewidth to “3” in the select menu may improve the clarity of the image.
7. Play with your conserved residues; the figure might get too busy if you add them all but put in a number to see where they lie. You should be able to produce a nice final image with a number of them shown; hopefully you can find the proximal and distal histidines and place these on your molecule. Manipulate the image to best show off the conserved residues. If you prefer prepare a couple of final images, one with eg the histidines and a second one with other conserved residues shown.

Report

Please turn in the following items:

1. Your sequence alignment as a texshade or boxshade.
2. A list of the conserved residues from your alignment.
3. One or more images showing the positions of the conserved residues on the human myoglobin molecule.
4. Are the conserved residues clustered in any specific areas of the protein?