

Biochemistry Lab II Informatics & Modeling Workshop



This laboratory workshop is designed to guide you through the basics of finding amino acid and nucleotide sequences, structural domains, conserved domains, functional domains. Then you will learn how to align the protein sequence to find important amino acids. Finally you will learn how to render/model a structure to identify the structure function for any known protein structure.

First a simple introduction is required. Go to the NCBI page at www.ncbi.nlm.nih.gov This is the national center for biotechnology information center hosted by the national library of medicine and the national institutes of health. Here you will find in the dark thin blue bar, a link for pubmed, and several "groups of programs"; Entrez, Blast and Structure. We will be working on each of these in this workshop/lab. Entrez is a set of search engines for many databases. Blast allows you to manipulate your sequences and search for related sequences, both nucleotide and amino acid. The structure link is the modeling and 3D programs. We will be using all three groups. If you get lost while clicking away, just find home by clicking on the NCBI icon and then moving into the group you desire.

The underlined questions and the questions at the end of this page are to be typed up and turned in for your assignment (50 pts). YOUR ASSIGNMENT IS A TYPED SET OF ANSWERS AND FIGURES FOR **EACH OF THE UNDERLINED QUESTIONS** BELOW AS WELL AS THE HOMEWORK ASSIGNMENT AT THE END OF THIS DOCUMENT.

And off we go...

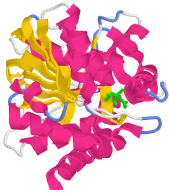
Step 1) Searching for protein vs DNA sequences (~2 hrs).

How does one find a DNA or protein sequence?

- Go to pub med and search for Phospholipase D. Look at links – what kind are they. Think about how many papers you would have to read to find the sequence.
- Record the reaction catalyzed by the enzyme and one or two specific biochemical characteristics of the enzyme.
- Go to the pull-down window next to the SEARCH box and find the nucleotide option database. Enter phospholipase D. Can you find it? Try searching for PLD. Sometimes it is hit and miss with an educated guess. (the "D" by itself broadens the search too much. Using a specific accession number found in a journal publication can help you narrow the search tremendously).
- Find the human PLD1. Are there different variants? If so, what are the differences? Do NOT click on the Gene page link!
- Note the blue number the upper left hand corner of the page – this is the nucleotide locus number also known as the **accession number**. Record this number.

Be careful when you click on different links. You may be taken to a different database such as "gene", "nucleotide" or "protein". The database for each page will be shown at the top and will have very different information.

- Click on the accession number.
- Notice the pubmed references the sequence references.
- Find the **gi. Number** (this is important for blast searches)
- Scan down and find the **CDS** number – The CDS is the "coding DNA sequences". This is where the DNA sequence start site is. – A common mistake is to assume that the first nucleotide is the first amino acid codon.
- Continue down to find both the aa translation and the nucleotide sequence.
- Look for the protein ID number. – click on that link. Where does it take you? Return to the nucleotide entrez. Click on Links – These are the options that are ready for this gene. Notice there is no structure for



Biochemistry Lab II Informatics & Modeling Workshop



this particular gene. Play around and look at the pub med link and a few of the others like related sequence link.

- Write down the nucleotide accession number for PLD and then click on the protein option in the black menu bar.
- Enter the accession number and describe what happens.
- Each gene has a number for the nucleotide, protein and structural database.
- Go back to the nucleotide database and either look at the entry for the protein number or use the links (protein) to get you to the right database.
- In the protein record you will find three additional links when in the protein database. BLink, Domains and Links. What are these links used for? We will come back to these later.

NOW – Go back to the **nucleotide** webpage where you first searched for PLD1. Click on one of the PLD links for the “This search in Gene shows...”

- Record the kind of information on this page and how it differs from the nucleotide page.
- Can you find the same info here as you found on the nucleotide page?
- Before you can continue, use the information here and in publications to find key amino acids for catalysis or regulation in PLD1. Record that information in your homework.

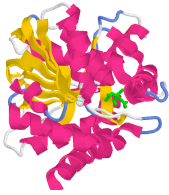
Step 2) Aligning aa sequences (1 hr) – Now that you have found a few important amino acids for one PLD, you need to find how those amino acids for PLD1 line up. This sequence is known but there is no structure for it.

- Click on the NCBI icon in the top left corner, and use the pull down menu to select the Protein: Sequence Database. Then type in human PLD1 and find the *Homo Sapiens* record. (Do not click on the gene info!)
- You will find the BLink link to the right. Click on it. This is a blast alignment of related proteins from highly conserved proteins at the top to disparate sequences at the bottom. What kind of information does this page have?
- Scan down the page. Notice the list of different genes that have some common amino acids with PLD1. Each matching gene receives a “score” number. What does that number mean and what information do you get when clicking on that number?
- What domain has the most hits? What is the function of this domain? What other information can you get about this domain? **YOU may have to search around and click on a few things to find the answers to these questions.**
- Go back to protein database page for PLD1. Click on the number under the “conserved domains” link. This page will show you some additional conserved domains. What is a conserved domain?
- The next link on the right is the protein record for the aligned PLD.
- The next link shows you that alignment search for that protein.

At this point you may find the protein/gene you are looking for. Then simply click and you are set up. This is good to find closely related proteins and find one using the “links” link to find which have a structure already in the database.

However, if you need to compare two sequences (say against a protein you have that doesn't have a structure. Or lets say you read a paper that indicates amino acid 212 was important for something. How would you directly line them up? One way is to manually slug it out by hand. No fun. Instead try the following:

- Use the GI of human PLD1 and research the nucleotide and an amino acid alignment against protein record number CAI24012 (PLD2 from mouse).
 - Then go to the NCBI webpage get to the blast index page by clicking on “blast” in the dark blue bar at the top of the page.



Biochemistry Lab II Informatics & Modeling Workshop



- o Then under the special heading click on “align two sequences (bl2seq)”
- o Now use the protein database gi number for PLD1 and PLD2 and find which amino acid in PLD1 corresponds to aa 212 in PLD2 from mouse. You will need to make certain you change the program settings from blastn to blastp (n = nucleotide, p = protein) in the pull-down window.
- o Finally, go to the protein record number and find out which triplet bases codes for the PLD1 amino acid. Don't forget that the first nucleotide is not the start site. See above for the CDS number! This will be important when comparing critical amino acids and domains between different isoforms of MDH!
- o Print out the results from your alignment!

Step 3) Structural Data Base and Rendering (3 hrs) Next you will be working with Pymol. This program is used to display structures (called rendering), focus in on specific amino acid sequences, lay two structures on top of the other to compare shape of the protein, and even make mutations to see what would happen, and to create publication quality images of structures or domains of a protein.

- 1) Go to the biochem lab II website to see several links for the Pymol tutorials. You must do EACH tutorial. There are links to support this.

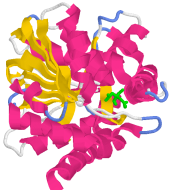
Step 4) Researching MDH Publication (2 hrs)

Go back to the pubmed website and start to research Malate Dehydrogenase (MDH).

- 1) Look for structural, kinetic and other papers on MDH. There are several very good reviews and critical publications on MDH. You might start your search using google and other web-based pages, but the bulk of your search and understanding should come from peer-reviewed publications found on pubmed.
- 2) Look for the sequences DNA and amino acid, that have been published. Use the domain function combined with key papers you found above to identify important domains of the protein.
- 3) Search the database to find solved structures of MDH.

Final Homework Assignment: *ALL work must be done independently!*

- 1) Finish all of the underlined questions from the questions above.
 - a. Record the reaction catalyzed by the enzyme and one or two specific biochemical characteristics of the enzyme
 - b. Are there different variants? If so, what are the differences?
 - c. Record the number.
 - d. Where does this take you?
 - e. Write down the nucleotide accession number for PLD and then click on the protein option in the black menu bar.
 - f. Enter the accession number and describe what happens.
 - g. What are these links used for?
 - h. Record the kind of information on this page and how it differs from the nucleotide page.
 - i. Can you find the same info here as you found on the nucleotide page?
 - j. Before you can continue, use the information here and in publications to find key amino acids for catalysis or regulation in PLD1. Record that information in your homework.
 - k. What kind of information does this page have?
 - l. What does that number mean and what information do you get when clicking on that



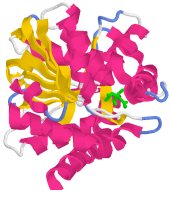
Biochemistry Lab II Informatics & Modeling Workshop



- number?
- What domain has the most hits? What is the function of this domain? What other information can you get about this domain?
 - What is a conserved domain?
 - Print out the results from your alignment!

For the rest of this exercise, use the gene/protein, Malate Dehydrogenase pdb 1SMK, 1SMKF, and the nucleotide sequence M33148 for your problems. Use the databases and the papers you worked with in steps 1 through 3 above to answer the questions

- Search the protein, nucleotide and structure database for Malate Dehydrogenase. How many hits do you find?
- Align the DNA and amino acid sequences for one of the MDH records vs. the MDH listed above. Printout or do a screen shot of the result. Repeat the same for the amino acids.
- Using the results in question 2, describe the differences in terms of length, gaps, percent differences, domains between each MDH.
- Pick the 50th nucleotide for the reference MDH and report the corresponding nucleotide in the second sequence. Do the same thing with the amino acids.
- Does your protein have any disulfide bonds? Are there any disulfide bonds between chains?
- Briefly summarize the secondary structure of your protein. Is the structure all alpha helix, all beta strands, or a mixture of both? Does your protein appear to include multiple domains?
- Does your protein have an "active site"? If so, what secondary structures are included in the active site?
- Do you think your protein is neutral, or has negative or positive net charge at pH 7? If it has a net charge, how does this support its function? What is the isoelectric point (pI) of your protein? Use protein explorer or the program you used last semester to calculate the pI for MGH.
- What are the important biological functions of your protein?
 - How does its structure support its functions?
 - What is the primary literature reference for this structure?
 - Look for thermal stability, protein dimerization, Km and Vmax, substrate (NADH and OAA) binding sites, the differences between NADPH and NADH binding. Hint - there are two very good papers on the active site and how LDH and MDH are very similar.
 - Ask yourself can you convert MDH to another enzyme (one that uses lactate or NADPH) or one that has a different temperature tolerance.
 - One with a different Km or Vmax. Are there other unidentified amino acids that should be looked at?
 - List each of the key domains and key amino acids with a description of what they do.
- Does your molecule form specific oligomers necessary for function?
- What structures other of MDH have been reported? List the record number, species/organism and any other pertinent information. What is the difference between e.coli MDH, and the MDH structure, 1SEV?



Biochemistry Lab II

Informatics & Modeling Workshop



- 13) Focus on the amino acids in the active site. Zoom in on those amino acids, label each of the active site amino acids and hide the others. Save the image and export for your homework. Then change one of the amino acids in the active site to a different non-conserved amino acid. Report the differences and save that image. Insert the two images in a file (powerpoint or Microsoft word) and indicate what you have created.
- 14) Create an overlay of the structure of the two different MDHs. Save the image as above.
- 15) Create a publication quality image for export of one of the two MDH. What part of the molecule is up to you. Be creative and have fun.

This assignment is to be TYPED and submitted to the chemistry office by the deadline shown on the laboratory website.