

Bioinformatics Exercises For Undergraduate Biochemistry Courses

Developed by Andrew Feig and Evelyn Jabri

**Indiana University
Department of Chemistry
Bloomington, IN 47405**

Phone: 812-856-5449

Fax: 812-855-8300

Email: afeig@indiana.edu

Last revised: 11/13/01

**Copyright, 2001
All rights reserved**

Table of Contents:**[Preface](#)****[Project 1](#)** – Identifying an unknown protein from short peptides sequences**[Project 2](#)** – Exploring the ExPASy and KEGG databases**[Project 3](#)** – Exploring the PDB, Protein Explorer and molecular visualization**[Project 4](#)** – Multiple Sequence Alignments – sequence conservation 1-D and 3-D spaces**[Project 5](#)** – Exploring enzymology in the BRENDA database**[Appendix 1](#)** – Table of unknowns for use in these exercises**Key web addresses:****Course web sites that use these exercises**

C484 (IU) **<http://chemlearn.chem.indiana.edu/c484/>****C483 (IU)** **<http://chemlearn.chem.indiana.edu/c483/>****Bioinformatics sites important for these exercises**

Blast: **<http://www.ncbi.nlm.nih.gov/blast/>****BRENDA:** **<http://www.brenda.uni-koeln.de/>****ExPASy:** **<http://www.expasy.ch/>****IUBMB** **<http://www.chem.qmw.ac.uk/iubmb/enzyme/>****Nomenclature****KEGG:** **<http://www.genome.ad.jp/kegg/kegg2.html>****MSA:** **http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_clustalw.html****OMIM** **<http://www.ncbi.nlm.nih.gov/Omim/>****PDB:** **<http://www.rcsb.org/pdb/>****Protein** **<http://proteinexplorer.org/>****Explorer****PubMed** **<http://www.pubmed.gov/>****Download free software for use in these exercises**

CHIME: **[http://www.mdli.com/cgi/dynamic/product.html?uid=\\$uid&key=\\$key&id=6](http://www.mdli.com/cgi/dynamic/product.html?uid=$uid&key=$key&id=6)****CN3D:** **<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dinstall.shtml>****Shockwave:** **<http://sdc.shockwave.com/shockwave/download/frameset.fhtml?>****VRML:** **<http://www.karmanaut.com/cosmo/player/>****Structure tutorials available on-line**

C483/C484 **http://php.indiana.edu/~afeig/c484/pdf_files/Unknowns.pdf**
tutorial

- **PE Quick** <http://molvis.sdsc.edu/protexpl/qtour.htm>
- **Tour**
- **CMU Chime** http://info.bio.cmu.edu/courses/03231//chime_tut/chime.html
- **Tutorial**

Preface to instructors who might want to use these assignments.

Preface to students interested in learning about bioinformatics.

Project 1 – Identifying an unknown protein from short peptide sequences.

Instructions

You are going to run a total of 5 BLAST searches during this exercise. Please follow the instructions carefully to insure that you get the necessary outputs. After you finish these exercises, you can get more information on the BLAST search algorithm by following the link to **BLAST course** or **BLAST Tutorial** in the left margin of the main BLAST web page. The questions here are simply for you to think about while you are doing the exercises and responses do not need to be written out. Question 6 on the problem set requires a formal response.

1. Go to the C483 web page (<http://chemlearn.chem.indiana.edu/c483>)
2. Click on **class links**.
3. Click on **Blast search** under bioinformatics links. This will take you to the NCBI web page. This is the National Center for Biotechnology Information. There are a lot of useful programs here, but today, we will be using one called **Blastp** that searches based on a protein sequence.
4. Under Protein-Blast, select **Search for short nearly exact matches**. Make sure the database is set to **nr** (stands for non-redundant and should be the default). This setting will search a collection of databases rather than just a single one. You could limit the search to just a single organism, if you wanted. Leave the rest of the settings in their default mode.
5. **1st BLAST SEARCH:** Enter the sequence of peptide 1 in the window labeled search. Be careful not to make typos. Then press **BLAST!**.
6. A new page will appear. Jot down the ID number. You can use this number to get at your results later if the search takes a long time because the server is busy. Press **format results** to continue in a new window.
7. The page of formatted results has four components:
 - a) graphic table at the top
 - b) summary of the search results
 - c) detailed set of pairwise alignments between your query (the peptide you entered) and the database subject
 - d) statistics on the search.

Sections a, b and c rehash the same information in the same order, just in different formats and with progressively greater levels of detail. Save a copy of the output file to disk (for each search you do) or you might have to repeat the search later.

8. Look at the output. What do you see? Did the search find a match for your unknown protein? The E-value is the way the computer has scored your search. Is a good match a large or a small number? If it is not immediately obvious look at the pairwise alignments. Find two that differ in the number of identical amino acids and compare their E-values.

9. **2nd BLAST SEARCH:** Try peptide 2. Did it find a match this time?

10. **3rd and 4th BLAST SEARCHES:** For these two searches, use the **Standard protein-protein BLAST [blastp]** rather than the **Search for short nearly exact match** settings. What is different about the parameters used for these two types of searches? Enter both peptides together (enter them as Peptide 1 <space> Peptide 2 for search 3 and Peptide 2 <space> Peptide 1 for search 4). One or both of these searches should have found an unambiguous match to your unknown enzyme. Did these two searches provide identical results?

11. **5th BLAST SEARCH:** Now that you know what you are looking for when you have a good match, type in either peptide 1 or 2 (preferably one that provided some type of hit in your original BLAST searches) in reverse order and repeat the search. Did you find a match?

12. Write down the name of the protein. If it is cut off in the table, scroll down to the alignments. Also, try to find the Enzyme Classification number (format X.X.X.X) and the Swiss-Prot accession number (format: sp|PXXXXX|). You will want some of these numbers for the future exercises.

13. Use a hyperlink from the BLAST output file to find the full NCBI entry on your enzyme. At the bottom of the entry, you should find the full amino acid sequence of your protein. **Print out this entry to turn in with your problem set.**

Questions for exercise 1 (Problem Set 1 Question 5).

Follow the instructions on the handout to obtain 5 BLAST output files. Save these files. You will do more with these unknowns on PS3, but we will discuss the process of BLAST searching on Wed. and I want you to complete this part of the project in advance of that discussion.

a) What is your unknown number?

b) What is the name of your unknown protein?

c) Why are there multiple hits for what appears to be the same enzyme?

- d) Can you positively identify the organism from which this enzyme derived? If so, what organism is it from? If not, why not?
- e) What is its E.C. number and Swiss-Prot accession number?
- f) When you entered your peptide in backwards, did you get a significant match to anything in the database? What does this teach you about the importance of directionality and orientation in biological macromolecules?
- g) Compare the results from the two searches that used both peptide fragments (Searches 3 and 4). Did you get identical results? This comparison calls into play gaps and gap penalties, but the program has a maximum gap size of 40 amino acids. Explain your results from Searches 3 and 4 in relation to gaps and gap penalties.
- h) Find the peptide fragments you used to identify your unknown protein in the full protein sequence. Highlight them on the print out you are submitting with your problem set. Are the sequences contiguous to one another and does it matter? Please explain.
- i) Think of an example of how you might use a BLAST search for something other than simply identifying a protein for which you have partial sequence information.

Project 2 – Explore the ExPASy and KEGG databases.

Instructions

These instructions are going to walk you through finding an assortment of information regarding your protein and using some of the protein manipulation tools available on certain bioinformatics web sites. I encourage you to look around these sites beyond the explicit instructions. There are many useful tools for manipulating, comparing and otherwise probing protein sequences and structures on these sites. You will be able to use some of these sites to help you with other problems on your problem sets, even if I don't explicitly send you to the Internet to carry out the assignment.

Have handy the following information about your protein before you start: Name, E.C. number (format: #.#.#.#), Swiss-Prot Accession Number (i.e. P#####).

1. Go to the C483 web page (<http://chemlearn.chem.indiana.edu/c483/>) and from there click on class links. We will start our tour on the ExPASy web page so follow the link to ExPASy. This acronym stands for **Expert Protein Analysis System** and is one of the primary sites that annotates and cross-references the various databases in which protein information is stored. To make the best use of these sites, you need to understand their limitations. ExPASy, for instance, focuses solely on protein sequences and you will have to go elsewhere (NCBI) if you have nucleotide data.
2. Your first task is to find your protein in ExPASy. You will find it by several routes. The first is by using the name of the enzyme.
3. From the main ExPASy web page, click on **Swiss-Prot and TrEMBL**
4. Notice that at the top of this page you have a quick search box that will allow you to find specific pieces of information. On the lower part of your screen, you have several different search options for more advanced and refined searching. Enter the name of your enzyme into the quick search box.
5. Click on the enzyme from the appropriate organism from the subsequent list.
6. You are now at a page called **NiceProt** for this particular protein. Does the accession number match the one you were expecting? We will come back to this page in a moment if it was the correct enzyme.
7. Click on the back button twice to get back to the quick search window.
8. Enter your accession number into the quick search. It should take you directly to the **NiceProt** page. Print out a copy of the **NiceProt** page for future reference (Note: there is a button at the top labeled **PRINTER FRIENDLY VIEW**. You may want

to consider pressing it before you try to print.). Notice at the bottom of this page there is a full protein sequence. To the right is a button labeled **fasta format**. The fasta format is used by many programs, so remember how to access this file format. Also notice that there are a variety of tools available at the bottom of this page. These are some of the same tools available to you in step 13 below.

9. Click on the button in the menu bar to return to the **ExpPASy home page**.
10. Now, click on the link called **ENZYME**.
11. You are taken to a page where you have several additional search choices. Here you can enter either the E.C. number (you must use the tab key to go between the four boxes) or the enzyme name. Enter your E.C. number and press **Search**.
12. A new page called the *NiceZyme* page appears. Print out a copy of this page. How does the information on this page differ from that on the *NiceProt* page you looked at a few minutes ago? Why is there no sequence information on this page? There is a section at the bottom called Swiss-Prot. What do you see in this section? Click on one of the hyperlinks in this section. Where does it take you?
13. Return to the ExpPASy home page. Spend a few minutes exploring some of the proteomics tools accessible from the main ExpPASy web site (right column). Look in particular at the sites listed under **identification and characterization, primary sequence analysis, and secondary structure prediction**. Submit your protein to a few of these analyses by using the Swiss-Prot number and see what you learn about your protein.
14. Next, we will go to the KEGG database, listed under class links as **metabolic pathways**. You could get there directly from ExpPASy using the links from the *NiceZyme* page.
16. Press **Search object in pathway maps**.
14. In this box, you would be able to search for a variety of different things. Enter the E.C. number for your protein and press **exec**.
15. Take note of the pathway(s) in which your enzyme participates. Depending on the enzyme there may be only one listing or there may be several. Click on the hyperlink to bring up the actual pathway map. We will look at a number of these pathways later in the semester.
16. Your enzyme will be colored red on the pathway. All of the E.C. numbers and intermediary metabolites are hyperlinks. Press on the link to your enzyme. What information is here? Press the back button to return to the map.

17. Now, press on the circle that represents your starting material. A structure of the compound will appear. If this compound is used in several different metabolic reactions, you will be able to see that on the page that appears. Press the back button and return to the metabolic map.
18. You can superimpose a lot of information on the pathway by using the pull down menu at the top left. For instance, click and hold on the pull down menu that reads reference pathway. Scroll down almost to the bottom of the long list to where it says **3-D structures in the PDB**. Then press **Exec**. Some of the boxes are now shaded light blue. Is your enzyme shaded?
19. Now, select **Genetic Diseases in OMIM** in the same pull down menu and press **Exec**. OMIM is the on-line database of Mendelian Inheritance in Man. Is your enzyme colored? If yes, click on your enzyme. If no, answer no to 4 f on your problem set and follow the instructions below for another enzyme that is colored to see what is available in this database.
20. Click on your enzyme from the database. It will take you back to the page you saw in step 16.
21. Find the link labeled **Disease MIM: #####** and click on it.
22. Use this information to answer question 4f. Some of the proteins will have very extensive links while others will be pretty minimalistic. For anyone who wants to see how much information there is on a well-studied genetic defect, explore phenylalanine hydroxylase (EC: 1.14.16.1), the enzyme responsible for PKU.
23. Take notice of your enzyme throughout the course. You may be called on to tell us something interesting about your enzyme during our discussions.

Excerpt from Problem Set 2 – Problem 4

Follow instructions on the handout that lead you through an exploration of the ExPASy and KEGG databases. You will use the information from those sites to answer questions about your unknown protein. Please submit copies of the NiceZyme and NiceProt pages for your unknown protein with your problem set.

- a) Compare and contrast the information available on the **NiceZyme** and **NiceProt** pages and explain when you might consult one rather than the other.
- b) Subject your protein to digestion with trypsin and chymotrypsin (high specificity) and cyanogen bromine. Use the tool PeptideCutter found on the ExPASy home page (ExPASy → proteomics tools → peptidecutter). Turn in the cleavage map with your problem set and perform the manipulations described in problem 5 on one peptide fragment from this map.

- c) What chemical reaction does your enzyme catalyze?
- d) Does your enzyme require a cofactor for catalytic activity?
- e) Interpret the E.C. number associated with your enzyme. Use the link to the Enzyme nomenclature page (<http://www.chem.qmw.ac.uk/iubmb/enzyme/>) to find a hyperlink table that will allow you to explore the way proteins are classified.
- f) In what metabolic pathway does your protein participate?
- g) Is there a disease associated with a mutation of your protein? If so, what is it? If there are multiple, describe one. Include information on what that disease is, what the mutation is that causes it and whether the disease is related to altered activity, altered expression levels or improper regulation of the activity.
- h) Tell me something interesting you discovered about your unknown protein in step 13 of the instructions for exploring ExPASy.

Project 3 – Exploring the PDB, Protein Explorer and Molecular Visualization

Instructions

After completing these simple exercises, you should be familiar with how to use many of the basic commands in Chime and Protein Explorer (PE). These programs will allow you to manipulate protein and nucleic acid structures to suit your needs. Chime is a plug-in module that provides extra functionality to web browsers. Many of the web pages we will be using incorporate Chime images. These images can be identified by the small MDL icon in the bottom of the frame and can be manipulated using the mouse. Protein Explorer is a more general program available through the PDB as a viewing program. It uses Chime, but also has a command line that allows you to type in more exact instructions for enhanced viewing and manipulations. Chime is available on any STC computer, or if you prefer, you can download it on to your computer. To do the latter, go to the class web page and access the Chime web page. Follow the directions to install the program. This tutorial will teach you to use PE. This knowledge will let you manipulate Chime images on other web pages in the same fashion.

Follow these steps to access the Protein Data Bank.

1. Open Netscape (Do NOT use Internet Explorer). Use of version 4.7x is highly recommended. Netscape 6.x has some problems with the way it supports Java and PE does not run consistently on 6.x.
2. Go to <http://www.rcsb.org/pdb/index.html>. This URL is the Protein Data Bank (PDB) web page. A link is available on the course links web page.
3. Use the Search tool to find entry **1IGD**.
4. A summary information window with a sidebar should appear.
5. Click on the View Structure button on the sidebar. Another window should pop up with a list of ways you can view the structure. Choose **Protein Explorer**.
6. Another window will appear from which you can start Protein Explorer. Follow the instructions on the next few pages until the structure appears (click on the button at the bottom of the page after setting the window size as you would like it. Use the Protein Explorer 2 beta version.)
7. You will have another startup window (this occurs due to the beta version we are using). Click on the **Advanced option** box and then **START SESSION**. Wait for the structure to appear and the green ready light to appear at the bottom of the left window.

8. Your window will have 3 frames
 - The Structure Window
 - First View Window
 - Message Window
9. Stop the molecule from spinning by pressing the **toggle spinning** button at the top right.
10. The large red balls are water molecules. You can remove them by pressing the **hide/show waters** button.
11. Press the link called **Explore More with Quick Views**.
12. There are several ways that you can adjust what you are looking at and the instructions below will walk you through some of the manipulations. Anytime you want to reset the image, press the **RESET VIEW** button and the original file will reappear on your screen.
13. Let's see what it is we are looking at. Press the **Mol Info** button. A new window will appear with links to information about the protein you are looking at. Press the link to the **sequence**. Here you can look at the primary structure of this small protein. Clicking on the boxes and pressing the apply will show you where in the sequence the given amino acids are located. Close this window when you are finished exploring.
14. Now, let's learn to rotate, zoom and translate the molecule in the structure window.

MAC	PC
Zoom in and out by holding down the SHIFT key and mouse button at the same time, and dragging the mouse back and forth	Zoom in and out by holding down the SHIFT key and the left mouse button at the same time, and dragging the mouse back and forth
Rotate it by holding down the mouse button and moving the mouse in a circular motion	Rotate it by holding down the left mouse button and moving the mouse in a circular motion
Move the molecule up and down by holding down the APPLE key and the mouse and dragging the mouse up and down or side to side.	Move the molecule up and down by holding down the CONTROL key and the left mouse button and dragging the mouse up and down or side to side.

You can also zoom in and out using the button in the quick view window. Each time you press the button it will zoom an incremental amount.

15. There are 3 general ways to manipulate the image. They each have their merits so you will want to become familiar with each of them. These include:

- a) Using the quick view menus
- b) Typing into the message line – for this you will need to learn some of the syntax
- c) Using the menu located under the MDL icon in the lower right hand corner of the structure window.

16. Let's start with the MDL icon. Click on it and hold down the mouse button. A scroll bar will appear. You can manipulate the structure using this tool bar.

17. Convert the molecule to spacefilling mode.

MAC	PC
Scroll down the MDL menu to Display . When the submenu opens follow it to spacefill and then van der waals radii . Now let go of the mouse button. The command will act on any part of the molecule selected or if nothing is selected, the whole molecule.	Click on menu, release, and then click on the menu where it says Display . Repeat this process on Spacefill and again on van der waals radii .

This command will be written out as:

(MDL) Display → spacefill → van der waals radii

The MDL menu is present in all chime web pages and not just protein explorer. Many of the web pages used in the course include chime structures. This pull down menu is the only way to manipulate structures in the generic chime pages, so you will need to understand basic manipulations if you wish to get the most out of these web sites.

Use this tutorial to get more familiar with the program. Then answer the questions on the problem set by applying what you learned in the tutorial.

1IGD is a small protein with 61 amino acids. It is a domain of a larger protein involved in IgG binding. It has a simple compact structure that will make it a good place to learn to manipulate the images of macromolecules.

The first task is to remove the solvent molecules that make it hard to see the protein.

(MDL) select → hetero → solvent
(MDL) select → hide → hide selected

The same function can be carried out by using the water button in the upper right window that toggles the waters between visible and hidden. This command would be written as

(Quick View) water

The protein backbone is currently displayed in the spacefilling mode. Leave the waters toggled off. Let's change that to see it in cartoon format.

(Quick View) select → protein
(Quick View) display → cartoon

Rotate the molecule and look at the secondary structures present. This would be great view to use to sketch a topology diagram like you did on Problem Set 2.

There are several ways to select particular residues as you can see from the MDL select menu. Let's change the color of the helical region.

(MDL) select → protein → helix
(MDL) select → change color to → green

You can also use the command line for these manipulations. For instance, let's change the color of the some of the beta-sheet region using the command line. First we have to select the region of interest and then we have to change the color.

(Message) s sheet → return
(Message) color red → return

The selection is retained and you can split the commands between the command line and the MDL menu.

(MDL) display → sticks

reset the image to cartoon mode.

(MDL) display → cartoon

When using Protein Explorer, many of the manipulations will be easiest to perform in the Quick Views menus. Reset the view by pressing **Reset view** and return to the quick view menu after removing the waters and stopping the spinning.

(Quick View) Select → all
(Quick View) display → cartoon
(Quick View) select → helices
(Quick View) color → yellow
(Quick View) select → strands
(Quick View) color → violet
(Quick View) bkg

You will have to select different options to understand just how powerful the program is.

You can also pick out specific residues using the command line

(Message) s 47-51 → return
(Message) color cpk → return
(Message) st → return

You can use the mouse to determine the residue numbers by clicking on the appropriate part of the molecule.

(Message) s 28-42 → return
(Message) wf → return
(mouse) click on any atom

The residue number and atom number of the selection will appear in the message window. In a Chime web page where there is no message window, this information will appear at the very bottom of your browser window.

Find residues 29 and 40 in this manner, they are located at either end of the alpha helix, then reset the helix to cartoon mode using the quick views menu.

(Message) s 29 → return
(Message) color cpk → return
(MDL) display → sticks
(Message) s 40 → return
(Message) color cpk → return
(MDL) display → sticks

One thing to note is that there is currently no “UNDO” button. If you are not sure if you have selected the item of interest, simply change its color. That is easy to undo by changing it back to its original state.

You can use the mouse to activate a variety of other functions as well.

(MDL) select → mouse click action → toggle atom label
(Mouse) click on one of the side chain oxygens from Glu 29
(Mouse) find on Asn 40 and click on it

If you click on the atom again, the label will disappear

(MDL) select → mouse click action → toggle distance monitor
(Mouse) Find the side chain oxygen of Asn 40 and click on it
oxygen is red in cpk coloring
(Mouse) click on the side chain oxygen from Glu 29
(Quick View) bkg

Orient the molecule so that you can clearly see the distance label and print out a copy to turn in with your problem set.

If your molecule had more than one polypeptide chain, you could use slightly different syntax to pick the correct amino acid(s). The syntax would be:

Command	action
<i>(Message) s (1-20:B)</i>	select amino acids 1-20 on chain B
<i>(Message) s (*:A)</i>	select all of chain A
<i>(Message) s 36,48</i>	select residues 36 and 48
<i>(MDL) select → chain → C</i>	select chain C

You will have to use the program more to learn all of its nuances. But this quick run through should get you started. If you have specific questions on how to get Chime/Protein Explorer to do something, email either Steve, Tara or Prof. Feig or bring the questions to office hours. Two additional tutorials are available on-line. One is the chime tutorial on the course links web page, the other is the 1 hour quick tour of Protein Explorer. They take awhile to complete thoroughly but will run you through some addition practice with the program. If you need information on how to enter something into the command line, press **Show Aliases**. A new window will appear with all of the commands.

Many of the structures of your unknown proteins have ligands (substrates) bound to them. A very powerful command in PE is the show contacts. To take advantage of this feature when exploring the structure of your unknown protein

(quick view) select ligand
(quick view) display contact

The resulting image will have the ligand inside an opaque surface and only the parts of the protein within 5 Å of the ligand shown. Everything else will be hidden. Turn off the surface hiding your ligand by

(MDL) select → display → toggle visibility

You will then be able to explore how the ligand is bound to your protein with a lot of the other residues removed for greater clarity.

Excerpt from Problem Set 3 – Problems 5 and 6.

5) Access the PDB file on your unknown protein. You may already have the PDB record number from your visit to the nice-zyme page on ExPASy. If you do not have it, type in the name of your protein in the line labeled “**search the archive.**” There may be multiple

structures. You may pick any of the relevant structures but questions 5 and 6 may be easier or harder depending on which one you select.

a) Complete the worksheet, “Record of 3D Macromolecular Structure Observations for Protein Explorer” based on observations of your unknown protein.

b) Print out representative views of your protein in as a space filling model. Print out an *equivalent view* in cartoon mode. Turn in those print outs with this problem set.

6) Write a short paragraph (200 words or less – typed please) that describes the juxtaposition of secondary structure elements in your protein. Illustrate your discussion with an appropriate figure of your protein (made in Protein Explorer) as if it were to be included in a research paper. An example of a structure description can be found posted on the class web page. Please see a writing tutor if you have difficulty with this assignment. You will be graded on content and clarity as well as organizational, stylistic and grammatical considerations. (**NOTE:** This question will constitute a large portion of the total grade on this problem set. Please spend an appropriate amount of effort on this paragraph showing us that you have learned to look at and think about protein structure.)

Project 4 – Multiple Sequence Alignments – Sequence Conservation 1-D and 3-D Spaces

Instructions

This project is called a multiple sequence alignment (MSA). We will use a program called Clustalw to perform the alignment. So far, we have seen alignments in our Blast output, where small fragments of the input peptide were mapped against the intact protein sequences. In this exercise, you will be looking at the whole sequence of several proteins compared against each other.

For this project, you will need to first collect the sequences of your protein and 4 homologs, evolutionarily related versions of your protein from different organisms. You will also need the sequence of the protein from the PDB that you have been looking at to analyze your structure. To do this:

- First collect the sequence of the protein from the X-ray or NMR structure by going to the PDB.
- Call up the structure you have been using by entering the PDB record number.
- Press the “**sequence details**” button on the left menu bar.
- Press “**Download all chains in FASTA format**”
- Copy the sequence to a Microsoft Word file (including the line that starts with “>”). You can replace the information after the “>” with the PDB number if you wish.
- Remove all carriage returns (<CR>) from the sequence (making it one long word), but leaving the <CR> at the end of the line starting with “>”

- Now go to the ExPASy web page and then to the Nice-Zyme page for your unknown enzyme.
- At the bottom of that page is a list of organisms that have homologs of your protein. You will be able to identify some of the species (i.e. drome = drosophila melanagastor which is the fruit fly). Others may be more obscure.
- Click on the links for one of the sequences you want to collect, you will arrive at the Nice-Prot page for the enzyme. At the bottom of that page is the sequence and the link to the fasta formatted sequence.
- Click on the **fasta** link
- Copy the sequence (including the line that starts with “>” into the same Microsoft Word file for later use. Edit the sequence as discussed above. You may replace the information after the “>” with a short descriptive name if you wish (i.e. the name of the organism).
- Repeat this procedure until you have collected several sequences one after the next all in the same file. Leave one or two <CR> between each sequence.

- Go to the Clustalw web site at PBIL via the link on the C484 course links web page.
- Paste the sequences from the MS Word file into the sequence window.
- Press the “**Submit**” button. All of the default parameters in the program can be left as they are.
- Look at the alignment statistics. If they are between 5-15% identity, you should proceed to the next part of the assignment. If it is below 5%, remove one of the sequences (not the one from the PDB file) and repeat the alignment. If they are significantly more than 15% identical, add another sequence to your alignment, or replace a closely related species with one that is farther away evolutionarily speaking (i.e. humans are close to rats both being mammals but are distant from *E. coli*).
- **Print out the alignment results and hand them in with the PS.**
- Map the residues that are 100% conserved onto the 3-D structure. Use Protein Explorer to do this by selecting the residue(s), then changing them to CPK coloring and spacefilling mode. The rest of the structure can be in backbone mode. **Print out a copy of this figure to hand in with your problem set.**

NOTE: if you wish to enter more than one selection at a time, you may do so by stringing them together with commas in the command line (i.e. s 1-5:a, 7:a, 29:b). If you need a list of commands for use in PE, click on aliases for a complete list, If you need help with specific commands, feel free to email your questions to Prof. Feig or Peter.

Excerpt from problem set 4 – Problem 5

- a) What does the “*” indicate on the MSA alignment?
- b) What does an “:” indicate?
- c) When you look at the primary sequence and the amino acid conservation across organisms, do you see any patterns or organization?
- d) Is there a relationship between the sequence conservation and the overall 3-D structure?
- e) The exercise above looked at the MSA of a protein. How would you use MSA on a nucleic acid (such as a tRNA) and what information might you obtain from that exercise? Will you always look simply for conservation or are there other factors that might come into play? To answer this question, think about the structure of tRNA you work on above.

Project 5 – Exploring Enzymology in the BRENDA database

Instructions

In this exercise, you will simply have to use a database called BRENDA to look up some of the specific kinetic properties of your unknown protein. On the BRENDA page, numbers with the format: #5# indicate that the data refers to the enzyme isolated from organism 5. The list of organisms at the top for any given enzyme is found at the top of the BRENDA entry for the protein. Numbers with the format <5> indicate that the data derives from reference 5. References are found at the end of the entry.

To get to the BRENDA database:

- Go to the **course links** web page
- Click on the link to **ExPASy**
- Click on the link to the **ENZYME** program
- Under related tools and databases, click on **BRENDA**
- Enter the E.C. number of your enzyme in the search window and press **query**
- One or more E.C. listings should appear. To enter the database, click on the E.C. number of the correct enzyme in the query table.

You can also access this page from the BRENDA link on the Nice-Zyme page within ExPASy.

Excerpt from Problem Set 5 – Problem 5

- a) What are the ranges of specific activity that have been measured for this enzyme?
- b) Are their different specific activities reported for the enzyme from the same organism?
- c) Why would the specific activity vary in this case?
- d) Why might you expect the specific activity of the same enzyme isolated from different organisms to be different?
- e) Go to the section on K_M . Pick a substrate that has been used in several studies. What are the ranges of K_M 's that have been measured for that substrate?
- f) What is the pH optimum of your enzyme?

- g) Identify two competitive inhibitors for the enzyme from the table of inhibitors. Remember that competitive inhibitors usually look like the natural substrate.