

Bio314: Advanced Cell Biology

Working with Sequence databases and Protein Explorer Part I



Introduction

The Biology Workbench is a powerful web-based tool that has been available for free since 1996 [2]. It allows anyone with a web browser to easily use bioinformatics for research, teaching and learning. This computational interface contains a suite of sources (databases) and computational tools for working with molecular sequence and structure data that aids in understanding sequence relationships among proteins or nucleic acids. In this exercise, you will be using the Student Interface to the Biology Workbench to search for sequences from protein sequence databases for the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* and related proteins.

A protein sequence consists of a long chain of amino acid residues, which is the primary structure or backbone of a protein. The chemical interactions of the side chains of the amino acid residues cause the primary structure to fold into a stable three-dimensional (3-D) structure, described by three different levels of structure, secondary, tertiary, and quaternary [1]. Protein Explorer is a software program, available for free from the Internet, which allows anyone to easily visualize the 3-D structures of proteins, DNA or RNA macromolecules [5]. Just by clicking and rendering different color schemes and displays, you can visualize the backbone, secondary structure, distributions of hydrophobic vs. hydrophilic residues, noncovalent bonding interactions, amino acid or nucleotide sequences, sequence to structure mappings and locations of residues of interest and patterns of evolution and conservation.

To access either the Student Interface to the Biology Workbench or Protein Explorer,

which is also linked to Biology Workbench, see the following procedure.

Background

Proteins are made of long chains of amino acids. Each type of protein has its own unique sequence of amino acids that cause the long chain to fold into a stable 3-D structure (conformation) of lowest energy. The 3-D shapes of proteins have been adopted over time by natural selection, which stable proteins aid in the survival of cells. The evolution of a sequence family of proteins can be obtained by clustering sequences of proteins according to their similarity [4].

The function of a protein depends on its structure. Proteins that share similar sequences of amino acids have similar structure and function. In general conserved residues contribute to structure stability of the protein, essential for its function. Highly variable positions contribute very little to structure and function. In most cases positions containing substitutions are with an amino acid with similar properties. For example, valine can be occupied by leucine or isoleucine, hydrophobic residues.

The GFP is a fluorescent protein, which its function in coelenterates is to transduce the blue chemiluminescence of the protein aequorin into green fluorescent light by energy transfer [3]. The GFP gene has been successfully isolated and is used by biochemists for making chimeric proteins of GFP linked to other proteins where it functions as a fluorescent protein tag. In living cells it serves as a cell lineage tracer, reporter of gene expression or as a measure of protein-protein interactions. You will be

working with the GFP in later laboratory exercises.

Objectives

The objectives of this experiment are to

- Search and compare protein sequences using protein databases from the Student Interface to the Biology Workbench.
- Examine the 3-dimensional structure of the green fluorescent protein from *Aequorea victoria* using Protein Explorer.

Procedure

Before coming to lab:

Print a copy of the **Green Fluorescent Protein Activity** and **bring this to lab**, and **Register** into **Biology Student Workbench**.

To obtain a copy, follow these steps.

- Go to the Biology Workbench website at >><http://workbench.sdsc.edu>
 - Scroll down and click on [Biology Student Workbench](#)
 - click on lessons (top left)
 - Scroll down and click on [Proceed to all Labs >>](#)
 - Scroll down and find “[Green Fluorescent Protein Activity](#)” (GFP Activity)
 - click on [\[pdf\]](#) **Print** a copy and **bring to lab**.
- You can also access Biology Workbench through Dr. Easton’s website at >><http://faculty.buffalostate.edu/eastondp/dpeastonbiology>
- Under Links, click [To Other Sites](#)
 - Find and click on [Biology Workbench](#)

Optional but recommended

- Use the back button and go back to the LESSONS page
 - Scroll down to and click on [Proceed to all Tutorials>>](#)
 - Find “How to Use the Biology Workbench” (top of list)
 - Click on [\[pdf\]](#) and scan through. (can save it)
- This is a tutorial on how to use Biology Workbench, which will help you to become

familiar with the tools in Workbench. However, this tutorial is slightly different from the Student Interface. The Protein, Nucleic and Alignment tools will operate very similarly, Sessions will be replaced with Preferences and the Structure tools won’t be available to Student Interface.)

To register follow these steps.

- Click the back button to go back to [Biology Student Workbench](#).
- Click USE STUDENT WORKBENCH (under [STUDENTS](#))
- Click on REGISTRATION
- Enter information (remember it)
- Click REGISTER
- Close out Biology Workbench for now until lab.

Now Go to the Protein Explorer website at

- >><http://www.proteinexplorer.org>
- Click on [Protein Explorer in Chime \(current version\)](#)
- Scroll down to and click on Proceed to the [Protein Explorer FrontDoor](#)
- Click on 1. Watch the Protein Explorer Demo. You only need to go through the first 3 categories of the demo, which will take ~15 minutes. **Stop here.**

Continue from here in lab.

Utilize this handout and GFP activity

Assignment

Answer questions 1-13 in this handout.

Working in Student Interface

- Open into the Student Interface as you did before lab and log in. You should be in the page Preferences Tools with Current; Default Session
- Follow [steps 3 & 4](#) exactly in the GFP Activity
 - Ndjinn is pronounced “engine” and is used to find sequences of proteins for comparison.
 - To select 2 databases, highlight the first, press Ctrl and U at the same time, hold them until you find and highlight the 2nd database, then release. Scroll to make sure that both are highlighted then click on the Ndjinn button.

You should now have a list of ~approximately 162 protein sequences to choose from.

-Under step 5

- Make sure you select the three listed in the GFP activity.
- Make sure you select mutants or synthetic constructs, and ignore cloning vectors.

-Follow step 6,

- The CLUSTALW tool aligns protein sequences for comparison by finding the best ways to make the different amino acid residue positions line up with each other. The residue positions of the sequences are then color coded to differentiate the highly conserved and semi-conserved regions from the non-conserved regions.

-Save & print the sequence alignment.

- Highlight only from fasta label to end of sequences. Copy, paste and save into word. Then print from this document.

1. What protein sequences are most similar?

Provide evolutionary evidence for why you chose those sequences. (Describe regions)

2. What protein sequences are most different?

Provide evolutionary evidence for why you chose those sequences. (Describe regions)

-Follow step 7,

- Unrooted phylogenetic trees or denograms are generated to show how closely related different protein or nucleotide sequences are to a particular protein or nucleotide sequence whose function is actually known from biological or biochemical information. Related meaning sharing a common ancestry (related by descent), which a sequence is sufficiently similar to the sequence of interest.

-Save & print the phylogenetic tree.

- Highlight the phylogenetic tree of right click. Copy, paste and save into word.

3. Determine which invertebrate protein sequences are most closely related and which are least related.

Explain why you formed these conclusions.

-Follow step 8 a, b, & c. New window opens

d) click on proteineexplorer.org

WAIT!

Working in Protein Explorer [4]

-Step 9a will probably take you directly to the First View page, so you can skip 9b

-In the Control Panel on top, read from **Rotate to Identify** and click on underlined entries. (For the term Identify, click on the chains in different locations.)

-At step 9c, click on the 'Hide/Show Water' button a couple of times and the 'Toggle Spinning' button to see what these do. For easier viewing, leave the spin off and the water off.

- Rotate the molecule with the mouse, to observe the different views.

-Follow step 9d exactly. Enter the commands where it says # Commands May Be Entered Here. (Did you find the fluorophore and where is it located in the molecule?)

-Rotate the protein and determine a nice view then save it to print later.

Click on PE Site Map. Click on Reset Session. Answer the following questions.

4. How many protein chains are in your molecule? Are there any nucleic acid chains? (Click on backbone trace.)

5. Does your protein have any disulfide bonds? Are there any disulfide bonds between chains?

(Click on Disulfide bonds.)

6. Are there any ligands in your protein? If so, what are they? (Click on More elements. Also, click on any of the colored letters.)

Scroll down and click on Quick views. A new screen appears with an additional screen in the middle or Help Frame that you will need to refer to each time a change is made to the protein. It will contain explanations, color keys and help messages for viewing molecules.

Click a couple of times on the following buttons to see what happens; Spin, Zoom +, -, Bkg, Water, and Ligand (nothing happens, why?).

You may want to leave the black background on for better viewing.

Click the 2° button and answer the following question.

7. Briefly describe the secondary structure of your protein. (Use the Help Frame). **How many helices and/or strands are there?** (Click on Show counts under the Message Box at the bottom. Then read inside the Message Box.)

Now SELECT Protein, DISPLAY Spacefill, and COLOR Polarity 2, and answer the question. (Use the slab button to see inside.)

8. a. Is the surface of your protein largely hydrophilic or hydrophobic?
b. Based on the surface of your protein, do you think it is soluble in water, or is it a transmembrane domain?

Now leave SELECT & DISPLAY, and change COLOR to Polarity 5, and answer the questions.

- To answer questions 9-11, use your book and click on the charge at a given pH, or the isoelectric point link.
- To calculate the isoelectric point (pI) follow the directions given in this link.
- When highlighting the sequence, make sure you only highlight the sequence and exclude the comment lines, and highlight only one chain for GFP.
- Shrink the Help/Index/Glossary page to the taskbar so that you can refer to it if necessary. Also so that you can go to the EMBL web page.

9. a. What is the isoelectric point (pI) of any protein?

b. Calculate the pI for both the GFP

10. What is the quantitative net charge at pH 7 of GFP?

11. What are the important biological functions of your protein? How does its structure support its functions?

View conserved regions, follow these steps

- Go to the PE Site Map
- Click External Resources

- Click Conserved Regions (ConSurf)

A new screen appears.

- Enter the PDB ID code 1GFL into the box if it is not already entered, and enter A in the Chain Identifier box.

- Click Submit button.

Wait! Screen will flash.

- When Go to the results appears, click on it.

- Click on View ConSurf Results with Protein Explorer.

- Observe the different colored residues, which in the key starts at 1-variable-green and goes to 9-conserved-dark magenta.
- If you click on backbone then number 1 and number 9 you can easily compare the most variable residues to the most conserved.

12. What regions of your molecule are variable and conserved in evolution?

- Click on backbone again.

- Click ConSurf Seq3-D.

Find and Click on SER65, by scrolling through the colored sequence. SER65 will show up in the text box when you find it. After clicking on it, it will show up as a spacefill in Chain A. Find and Click on TYR66 and GLY67 the same.

13. Which of these residues are conserved and which are not?

Literature Cited & Resources

1. Alberts, B., et al, 2002. Molecular Biology of The Cell, 4th Ed. NY, Garland Science.
2. <http://bsw-uiuc.net/tutorials.php> Accessed; 10-14-07 How to use Biology Workbench v.3.2
3. <http://dwb.unl.edu/Teacher/NSF/CO8/CO8Links/pps99.cryst.bbk.ac.uk/projects/gmocz/GFP.htm> Accessed; 8-26-05 Protein Fluorophores
4. <http://faculty.buffalostate.edu/eastondp/Dpeastonbiology> Accessed 8-11-05 Protein Sequence Alignment tutorial
5. http://www.umass.edu/microbio/chime/pe_beta/pe/protexpl/whatispe.htm Accessed 10-14-07. What is Protein Explorer
6. http://www.umass.edu/microbio/chime/pe_beta/pe/protexpl/qtour.htm Accessed; 10-14-07 1-Hour Tour for Protein Explorer