**Cell Biology Laboratory**

Project Introduction, Exploration Sessions and Documentation

Table of Contents

[Lab Project Objectives\* 1](#_Toc51606818)

[Unnamed Proteins and Sequence Analysis 2](#_Toc51606819)

[TASK OVERVIEW (Project Exploration 1) 4](#_Toc51606820)

[TASK 1.1 (Project Exploration 1) 5](#_Toc51606821)

[TASK 1.2 (Project Exploration 1) 5](#_Toc51606822)

[TASK 1.3 (Project Practice 2) 5](#_Toc51606823)

[TASK 1.4 (Project Practice 3) 5](#_Toc51606824)

[TASK 1.5 (Project Exploration 2) 5](#_Toc51606825)

[TASK 1.6 (Project Exploration 2) 5](#_Toc51606826)

[TASK 1.7 (Project Documentation Draft) 6](#_Toc51606827)

[TASK 1.8 (Project Documentation Final) 7](#_Toc51606828)

# Lab Project Objectives\*

**Sequence Analysis of Novel Unnamed Protein**

1. Emulate the process that cell biologists use to research the cellular location of novel unnamed proteins to ascertain putative function.
2. Use free online resources to gather location clues for a published unnamed protein
3. Interpret data generated from the unnamed protein’s sequence to develop a hypothesis about its location and possible function.

# Unnamed Proteins and Sequence Analysis

**Purpose**

The purpose of this lab project is to 1) analyze an unnamed protein’s sequence of amino acids with various cell biology relevant online resources and 2) utilize generated *in silico* observations to develop a valid hypothesis about the unnamed protein’s putative location within the cell.

**Background**

The development of methods that reveal gene and protein sequences without associated experimental data has created a backlog of uncharacterized protein sequences. Some of the protein sequences are hypothetical, being transcribed *in silico* from DNA sequences. Other protein sequences are **unnamed protein** products, being experimentally translated in the lab from the DNA code.

Unfortunately, the high-throughput production of protein sequences has outpaced cell biologists’ ability to have meaningful insight into these proteins. A bunch of letters in a string doesn’t really increase our understanding of the how the protein plays a role in the cell’s life.

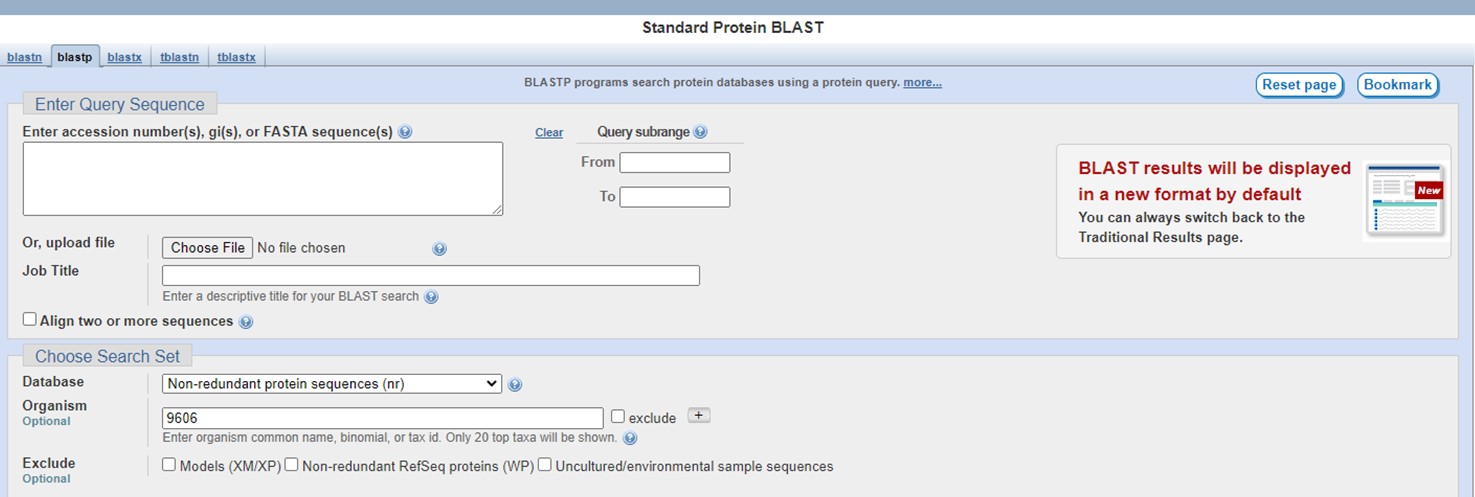
Location….Location….Location. A cell biologist can glean a lot of information about a protein’s function from its location within the cell. And, a putative location can be gleaned from the cells’ primary peptide sequence. Example, an unnamed protein X without a transmembrane domain would probably not be an integral membrane protein. From this kind of information, we can make suppositions like - unnamed protein X is not an ion channel.

**Online Resources**

That being said, free online resources for **sequence analysis** are bountiful. Using these resources, we can subject unnamed protein sequences to a range of cell specific analyses to appreciate the unnamed protein’s general cellular location.

**BLAST**

* To begin a sequence analysis, a protein’s sequence of amino acids can be identified using the US National Library of Medicine’s “**Basic Local Alignment Search Tool**” (BLAST) accessed through the “National Center for Biotechnology Information” (NCBI) sponsored by the US National Institute of Health (NIH). In short, we call it performing a BLAST search. This software aligns the protein of interest to all the proteins published in the database and provides the most likely match.

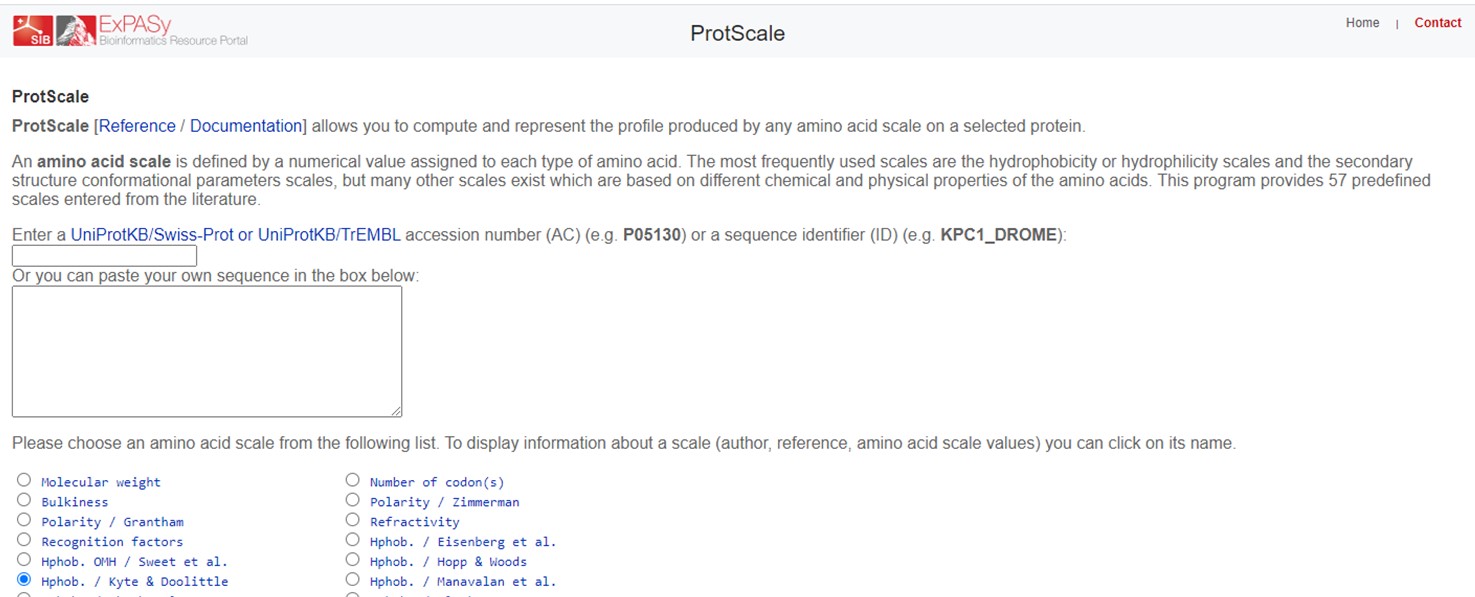


A **BLAST** search on an unnamed protein delivers either no result or the unnamed protein itself. However, the listing for an unnamed protein usually has useful pieces of information, such as the tissue/organ that was the source of the protein. It is always useful to begin a protein sequence analysis with a cursory BLAST search.

**HYDROPATHY PLOT**

* Our next analysis determines if the protein is soluble or embedded in a membrane. Information of this kind narrows the field of possible locations. A membrane-bound protein will not be found in aqueous solution. So, we can rule out most intracellular proteins involved in glycolysis, citric acid cycle, cytoskeletal polymerization, vesicle coat proteins, etc.

**Hydropathy Plot** is a graph which shows the hydrophobicity of each amino acid versus where it is located on the polypeptide. It is used to find clusters of hydrophobic amino acids that could pass through the lipid bilayer to indicate that the polypeptide in question has a transmembrane domain and is integral. A hydropathy plot can be generated by a **hydropathy plot generating software tool** accessed through the “Expert Protein Analysis System” (ExPASy) sponsored by the Swiss Institute of Bioinformatics (SIB).

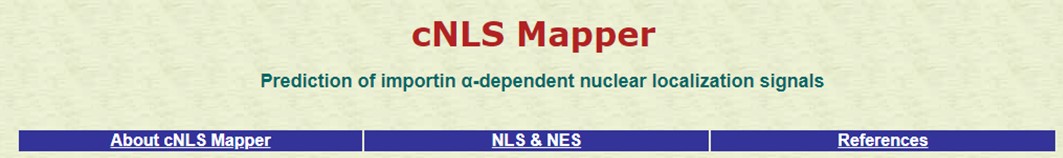


**cNLS Mapper and 3of5 Website**

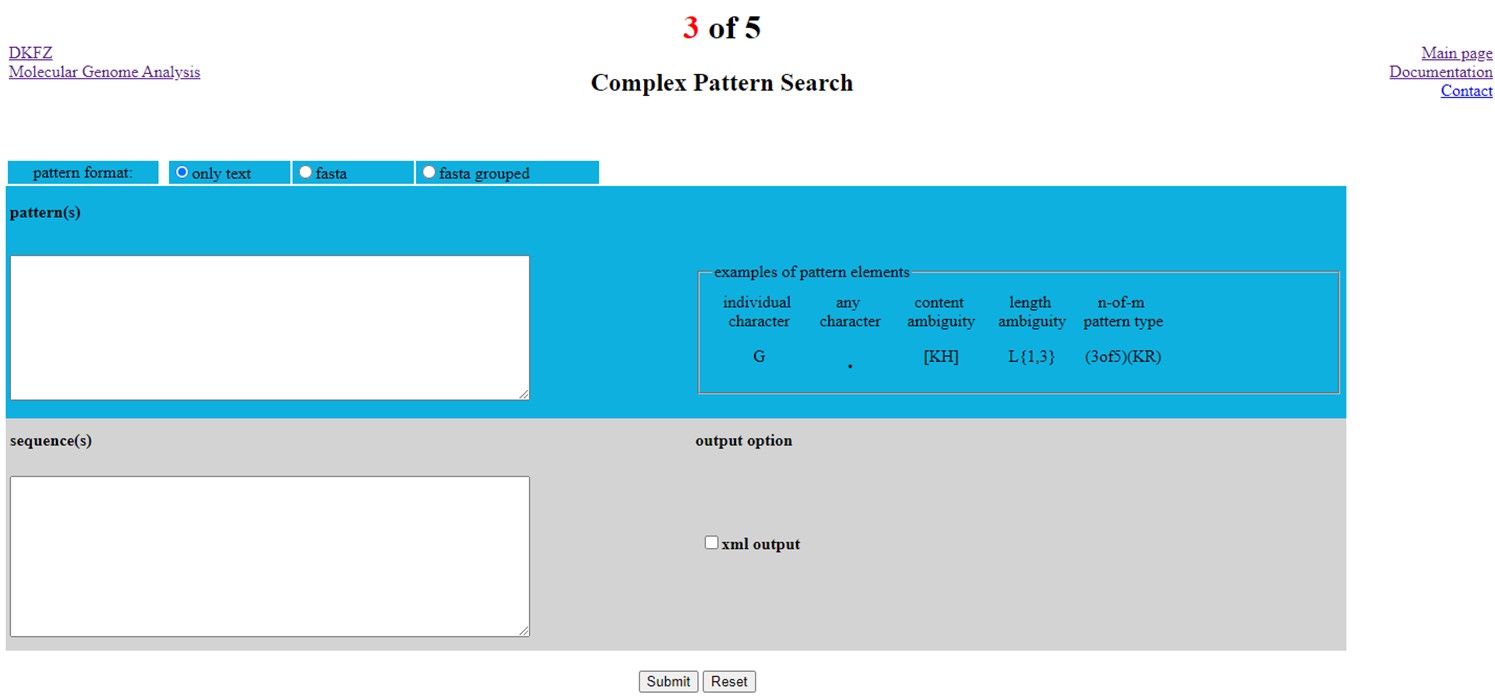
* We continue our sequence analysis by assessing in which cellular compartment the protein would most likely be housed. Proteins destined to be imported/exported into/from a particular cellular compartment contain ‘address codes’, which are a sequence of specific amino acids that determines the protein’s location. This concept is called the **signal hypothesis**.

For example, proteins that completed being translated on a free ribosome, but are intended for the nucleus should contain a nuclear localization signal (NLS). Additionally, a protein that travels to the Golgi from the ER should contain an ER export signal. Analyses could also be performed to determine if a protein has a mitochondrial targeting signal or a signal sequence for ER import or other sequence signals.

The “**cNLS mapper**” was developed by Dr. Kosugi and is accessed through the “Institute for Advance Bioscience” sponsored by Keio University in Japan. This algorithm accurately predicts nuclear localization signals (NLSs) specific to the importin pathway [Kosugi *et al*., (2009), *PNAS* 106:10171].



The identification of complex and highly variable patterns, like the ER export signal, is a particular challenge. So, the “**3of5 website**” was developed by Dr. Seiler at the Division of Molecular Genome Analysis sponsored by the German Cancer Research Center (Deutsches Krebsforschungszentrum). This website allows the user to define their own search strings to predict variable sequences such as the ER export signal. [Seiler *et al*., (2006), *BMC Bioinformatics* 7:14].



## TASK OVERVIEW (Project Exploration 1)

There are many more free, online resources for **sequence analysis**. Because a cell biologist is resourceful, flexible and creative, they are able to find and use these resources to deduce the general cellular location of a protein.

Example, MitoProt predicts mitochondrial targeting sequence probabilities. It was developed by Dr. Claros at the Institute of Human Genetics sponsored by the Technical University of Muich at <https://ihg.helmholtz-muenchen.de/ihg/mitoprot.html> [Claros, et al., 1996, Eur J Biochem 241:779].

**Now, let’s perform a sequence analysis on the primary sequence for an unnamed protein. After receiving the sequence, perform a series of *in silico* analyses: 1) BLAST, 2) hydropathy plot generator, 3) NLS mapper, 4) 3of5 ER export, 5) your own.**

**Document information from sequence analysis on the project documentation report (provided under separate cover), and 2) record a synopsis of the project predicting the cellular location of the protein and hypothesizing its function.**

## TASK 1.1 (Project Exploration 1)

Use a BLAST search to determine how your unnamed protein aligns with the known database of proteins. In the report, take note of details that you find interesting, such as type of organ/tissue/cell that was the source of the protein or the author who posted it or if the unnamed protein is published (if so where).

## TASK 1.2 (Project Exploration 1)

Use a hydropathy plot generating program to generate a hydropathy plot for the unnamed protein. Save the hydropathy plot and paste it into the report.

## TASK 1.3 (Project Practice 2)

Use nuclear localization signal (NLS) mapper program to determine if the unnamed protein contains an NLS. In the report, note the NLS found, include details such as mono- or bipartite and score of the NLS.

## TASK 1.4 (Project Practice 3)

Use the 3of5 website to determine if the unnamed protein contains an ER export signal. In the report, note the location of the ER export signal.

## TASK 1.5 (Project Exploration 2)

Based on the analyses conducted so far, rationalize the next step to take, find and learn how to use your own unique sequence analysis free, on-line resource. Conduct the analysis, writing your own protocol and documenting the results along the way.

## TASK 1.6 (Project Exploration 2)

Complete the project documentation report that includes a series of questions similar to the ones provided below. Official report sheet provided separately as a word document.

1. Write the first six amino acids of the unnamed protein assigned to you.
2. Write one detail about the protein determined by BLAST
3. Paste a picture of the generated hydropathy plot for the unknown protein here.
4. *Is the unnamed protein an integral membrane protein?* Explain how the generated hydropathy plot supports/contradicts the idea that the unnamed protein is an integral membrane protein.
5. Write the highest scoring predicted monopartite NLS amino acid sequence, including score. If no result was listed, write ‘none’.
6. *Is the unnamed protein located in the nucleus?* Explain how the NLS mapper results supports/contradicts the idea that the unnamed protein is located in the nucleus.
7. Write the position of the ER export signal. If no result was listed, write ‘none’.
8. *Is the unnamed protein able to be exported from the ER?* Explain how the 3of5 website results supports/contradicts the idea that the unnamed protein is located in the ER or elsewhere in the endomembrane system.
9. Based on the data so far, what unique sequence analysis have you decided to perform? Rationalize your choice, in particular discussing what clues you have gotten from BLAST, hydropathy plot, NLS and ER export analyses.
10. Include your unique sequence analysis protocol here, particularly website used and if applicable, code used.
11. Write/describe the results from your unique sequence analysis.
12. Explain how these results support/contradict that your unnamed protein was housed in a particular location.
13. Based on your sequence analysis, hypothesize where the unnamed protein is located within the cell.

## TASK 1.7 (Project Documentation Draft)

Great leaders have mastered their listening skills. So, when you find yourself in the presence of a great leader, why stand in awkward silence? If you talk about it, they will listen. If what they hear is impressive, this small interaction may find you as the person who gets the next scholarship or grant or computer or promotion. Even if it the interaction doesn’t lead you to an advantage, at least you have had a satisfying conversation.

Let’s practice the 30 second blurb (“short description”). The subject is going to be our unnamed protein and what we have found out about it so far.

1. Develop one “talking points” index card with your name at the top and outline a 30 second ‘take home blurb’ presentation that includes:

* a logical progression through background information that led to the putative cellular location of the unnamed protein
  + BLAST detail
  + Hydropathy Plot
  + NLS
  + ER Export Signal
  + Your own analysis
* statement of the putative cellular location of the unnamed protein
* what are you thinking of naming your protein
* if time allows, how might it function in the cell

1. Each student is assigned to a break-out room. Once in the room, assign an order for each student to present their blurb (#1 to #4). Assign a student who will be a recorder (it is hard to talk and record). Assign one student to record the recorders presentation. Each student will present their ‘blurb’ to their own group and then another group. Practice your blurbs.
2. When your group is done preparing all ‘blurbs’, set a timer set to 30 seconds. The first student to present per group and their recorder go first. The remaining students are the audience.
   1. Student recorder, open voice recorder app and start recording.
   2. When instructed, student presenter starts their ‘blurb’ (can use index card)
   3. When instructed, student presenter stops their ‘blurb’
   4. Student recorder, stops recording and saves recording as CB-Fa20-pre-*1st name presenter*. At some point, share recording with instructor via Email ([alagier@grandview.edu](mailto:alagier@grandview.edu)) and if desired, cc: student presenter.
   5. Switch presenter/listener/recording roles and repeat 1-4 until all group members have presented their ‘blurbs’ for the first time.
   6. Assess presentation performance and modify index card (if needed). Use 3-5 minutes to practice again (if desired).
3. When all break-out room groups are done with practice round of presentations, instructor will switch student break-out rooms for a presentation session. Student presenter will join another break-out room, whose students are the audience. Audience choice is randomized by instructor.
   * + 1. Student recorder in that room, open voice recorder app and start recording.
       2. When instructed, student presenter starts their ‘blurb’ (no index card)
       3. When instructed, student presenter stops their ‘blurb’
       4. Student recorder, stops recording and saves recording as CB-Fa20-post-*1st name presenter*. At some point, share recording with instructor via Email (alagier@grandview.edu) and if desired, cc: student presenter.
4. Presenter/listener/recording roles will be switched and repeat Step 4.1-4.4 will be repeated until all group members have presented their ‘blurbs’ for to another group.

## TASK 1.8 (Project Documentation Final)

**Record a final recording of your project ‘blurb’ and submit recording before deadline for final project documentation.**

**When submit final recording in ‘submit text’ box make a comment about what you improved between the practice presentation sessions and the final recording.**