In the Weeds: Genomic Analysis of Glyphosate Resistance in Crop Fields

Learning Objectives

After completing this activity, students will be able to:

- Summarize the relationships between transcription, translation, and the universal genetic code
- Utilize successfully NCBI and EXPASY databases and software, including: PubMed, PubChem, nucleotide BLAST, protein BLAST, and iCn3D
    - Analyze DNA and protein sequence information
    - Detect a mutation in an unknown sequence by comparing it to wild type sequence
- Predict how a specific DNA mutation can cause a phenotype change
- Understand the general mechanisms of action of herbicides in plants and their potential effects on humans
- Explore the role of selective pressures on evolution of organisms

Background

The Central Dogma of Molecular Biology explains the flow of information in living organisms. Genetic information of a cell is stored as DNA. Segments of DNA, known as genes, are transcribed by an RNA polymerase to produce messenger RNA (mRNA). In many species, the mRNA is processed to remove introns to produce a "mature" mRNA. These mRNAs are used in the process of translation, where ribosomes synthesize a polypeptide (protein) based on the sequence information present in the mRNA.

Scenario

Roundup® is a commonly-used herbicide sprayed early in the growing season to keep weeds from outcompeting crop seedlings. The active ingredient in Roundup® is glyphosate, which inhibits a protein needed for weeds and grasses to grow. "Roundup®-ready" crops are genetically modified so that they are not killed by the treatment.

You are a molecular biologist working with the state extension office in support of local agricultural operations. Over the last few years local farmers have noticed an increasing number of Roundup®-resistant weeds in their crop fields (i.e., the weeds are still present after the herbicide is used on the fields).

💡 What are some reasons for why this might be happening? Describe your hypotheses below.

Let's hypothesize that a ***mutation*** has arisen in a gene product from the Roundup®-resistant weeds.

💡 Describe in your own words what is meant by "gene product"?
(Hint: consider the Central Dogma of Molecular Biology)

💡 Find a good candidate gene where this mutation may have taken place, considering the specific plant protein on which glyphosate acts. Name the gene below.
(Hint: Use your favorite web browser to help with this question.)

Procedure

In order to help determine what is specifically causing the resistance in the weeds, you will be provided a file with a sequence from a Roundup®-sensitive plant and a second sequence from a Roundup®-resistant plant. As you consider how to analyze this data and test your hypothesis (above), look over the list of databases available to you from NCBI that you will be using in the following activities.

PubChem: offers freely accessible chemical information.

PubMed: searchable collection of citations and abstracts of bioscience-related literature

Nucleotide BLAST: computational tool for searching and comparing nucleotide sequence databases with a DNA or RNA query sequence (i.e., unknown)

Protein BLAST: computational tool for searching and comparing protein (polypeptide) sequence databases with a query sequence of amino acids

Expasy: tool used to translate DNA sequences of interest into amino acid sequences.

iCn3D ("I see in 3D"): 3-dimensional protein structure visualization tool that can also be used to align AI-predicted (AlphaFold) or experimentally-determined structures and determine significance of amino acid changes (resulting from gene mutations) on protein structure

I. Understanding the chemical structure of glyphosate

PubChem offers freely accessible chemical information. This tool can be used to explore the chemical and physical properties of glyphosate.

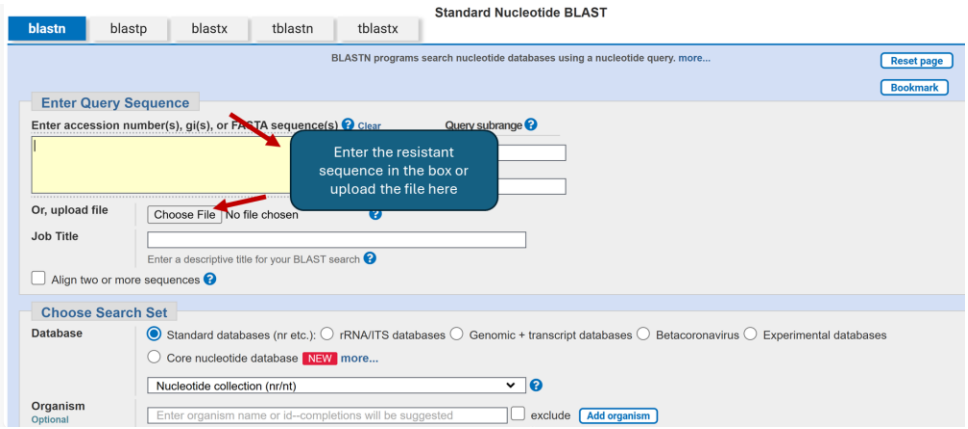
1. Access the website https://pubchem.ncbi.nlm.nih.gov/
2. On the search bar, type “glyphosate”
3. Tabs on the right side will allow you to explore many chemical and physical properties of the compound, including: toxicity, agrochemical information, associated disorders, and diseases, among others.

💡 Now that you are familiar with the physical and chemical properties of our compound of interest, **provide a brief description** of the effects on human health and **list 3** different disorders suspected to be associated with this chemical?

II. DNA sequence alignment of glyphosate sensitive and resistant weeds

The Basic Local Alignment Search Tool (BLAST) can be used to find the similarities between biological sequences. As a researcher, you are interested in learning more about the differences at the genetic level between the resistant and sensitive strains. You are provided with the nucleotide sequence of the resistant strain and you will use BLAST to identify the species it is similar to and how it compares to the sensitive strain. Follow the steps below to achieve your goal.

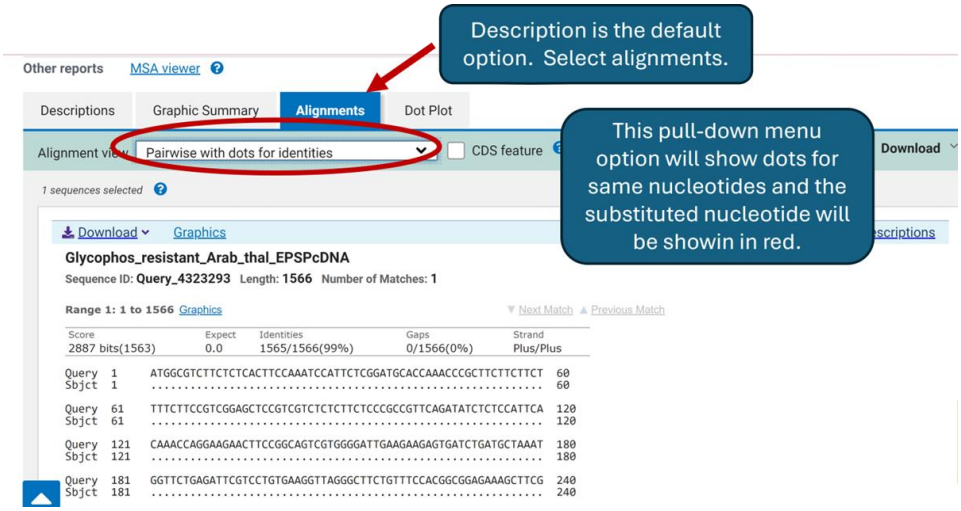
1. Access https://blast.ncbi.nlm.nih.gov/Blast.cgi.
2. Click on Nucleotide Blast.
3. Your instructor provided the resistant sequence (as text or a file).
4. You may paste the sequence into the search box or you may upload a file (as shown in the picture below).



5. Click BLAST and analyze your results.

💡 Based on your results, the best match is for which organism?

6. You will get multiple sequences (maybe over 100). You will pick the top matching sequence from the search (the most statistically significant E-value). Ideally, you will select a single "hit/organism" and you will work on the DNA sequence alignment between your resistant sequence and a match sequence (your instructor will discuss with you the proper parameters for your selection).
7. Click on Alignments as shown in the figure below and change to the "pairwise with dots for identities" view you will be able to identify any nucleotide change. If you review the results, the dots represent the same nucleotides present on both sequences.



💡 Based on your results, can you identify any differences between the 2 sequences?

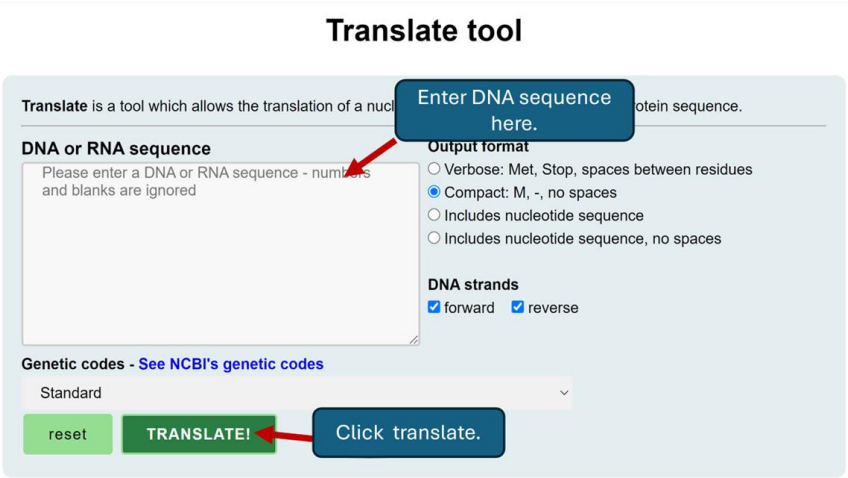
💡 If a change was identified, what does this change represent?

💡 Based on your understanding of the Central Dogma of Molecular Biology, which could be the implications of this change?

III. The Central Dogma: Using online resources to translate a nucleotide sequence

A **codon** is defined as a sequence of nucleotides which is translated into an amino acid. Currently, bioinformatics resources can transcribe and translate a sequence of imputed DNA. The Swiss Institute of Bioinformatics offers the Expasy tool allowing us to translate our DNA sequence of interest into an amino acid sequence.

1. Use https://web.expasy.org/translate/ to translate the “sensitive” DNA sequence.



2. Your results will include 6 different Open Reading Frames (ORFs). An ORF is identified as a fragment of DNA that is the correct frame identifying the start codon. Since a codon is a triplet, you will get three different options. Since genes can be arranged in the forward or reverse direction, the total number of possibilities will be 6 (as shown by your results). Make sure that compact; M, - no spaces is selected and that the forward and reverse options are selected.

💡 Based on your results, which one is the correct reading frame for this protein product? Why?

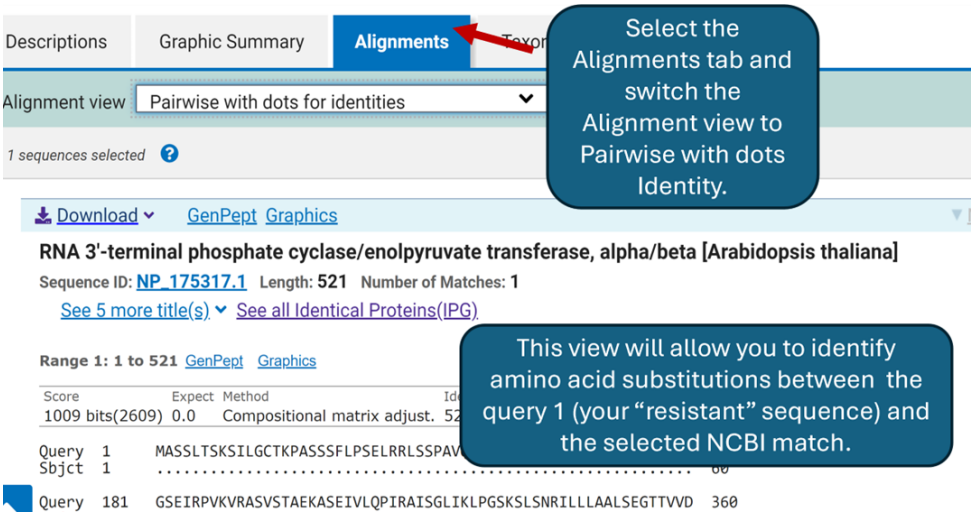
3. For purposes of this activity, we will be using the +1 translational reading frame. Your result will show the translated amino acid sequence. You may copy/paste this amino acid sequence and save it as a word pad file with extension .txt.
4. Repeat the steps above for the “resistant” plant.

IV. Using protein BLAST to determine whether the single base substitution causes a change in polypeptide primary structure

BLASTP is a tool that will allow us to input the translated amino acid sequence and compare it to a protein database. As a result, you will get multiple matches, but it is important to select the one with the highest percent of identity. This approach will allow the identification of any amino acid change (s) at a specific position.

1. Access https://blast.ncbi.nlm.nih.gov/Blast.cgi.
2. Select the “blastp” option from the tabs.
3. Paste your amino acid sequence from Expasy in the Enter Query Sequence section. *Note that when you click on the sequence, it may take you to a different window and it will be easier to copy the sequence from the fasta format file.*
4. Under Choose Search Set: Standards: Databases click the arrow and select non-redundant protein sequences (nr),
5. Under Program Selection: Algorithm select BLASTP (protein-protein BLAST).
6. Once the sequence is pasted (or the file uploaded) click the BLAST button.
7. Once your query is completed, a new window will open.

8. The results window will offer multiple sequences/organisms. Select the one with the highest max score/ percent identity.
9. Make sure the Alignments tab is selected and that you select “Pairwise with dots for identities” from the pull down menu.



💡 From this output view, is there any amino acid substitution when you compare your submitted sequence (Query 1) and your match (Subject 1)?

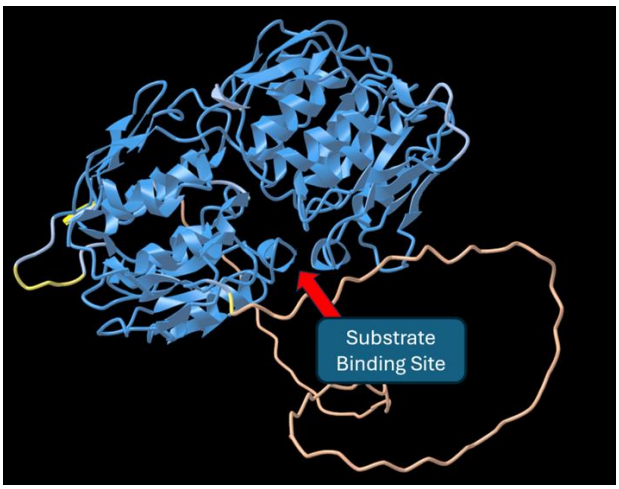
💡 What type of amino acid substitution is this nonsynonymous or synonymous? Explain.

💡 What effect would this change have in the protein?

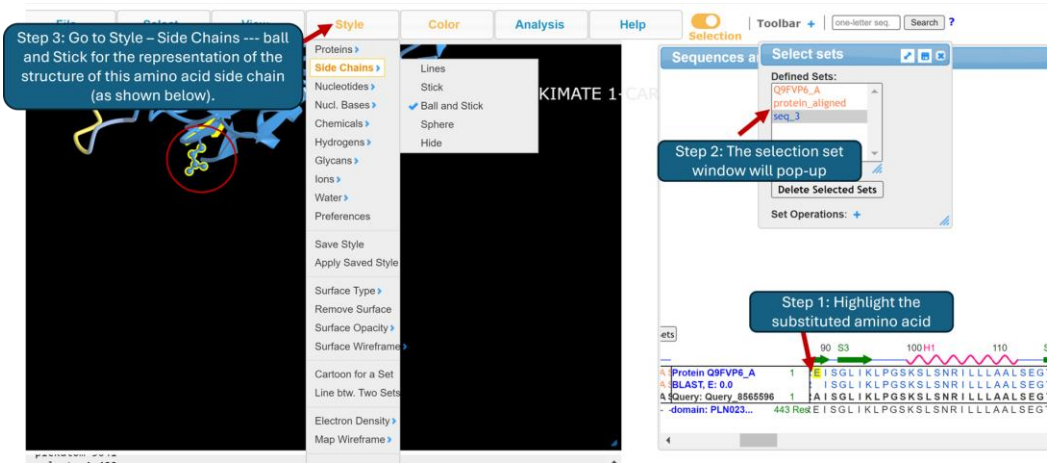
V. Modeling of amino acid substitutions using iCn3D

iCn3D is a web-based structure viewer tool that allows the interactive analysis of the structure. In this module, it will be used to model the structure of the protein of interest.

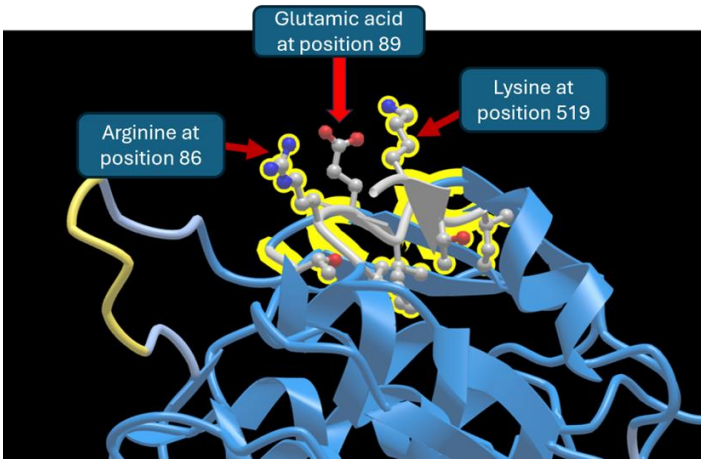
1. From the BLASTP page, click “Alpha-fold Structure” (right side of the page).
2. iCn3D will open and load your protein.
3. On the right side of the page, the sequences and annotations window shows a pairwise alignment of your query sequence (“resistant gene”) and the target sequence “wild type” gene.
4. Once you get your protein model, feel free to play around with it. Use your mouse to move the protein structure in the window on the left.
5. If you move the structure around, can you predict, based on the structure of the enzyme, the substrate binding site. (as shown in the figure below)



6. Use the bar under the sequence to look at the alignment of the two sequences. As previously shown in the protein alignment using BLASTp, there was a substitution at position 89 (glutamic acid to alanine, E → A).
7. Highlight the "E" and the location of this amino acid will then be highlighted in yellow on the structural diagram shown in the left panel (as shown on the right in the figure below). After selection, go to Style > Side Chains > Ball and Stick for observation of the amino acid residue as part of this model.
8. Using the Style and Color drop down menus (see arrow on the left hand side of the figure below), change style and colors in your structure.



9. You may also select/highlight specific regions close to the amino acid of interest to visualize the interactions with other side chains. For this you will have to highlight the region you are interested in as you did before and change the style to ball and stick for those specific amino acids.
10. The picture below shows glutamic acid and two other amino acids: arginine at position 86 and lysine at position 519.



💡 Look at the side chains of each of the labeled amino acids. What properties do they have in common?

💡 Based on your observations, can you predict what effect the amino acid substitution could have (glutamic acid by alanine)?