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.

Possible Answers: gene mutations, gene duplication, gene deletion, change in expression levels, gene flow

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)

RNAs are made via transcription from their DNA templates. While there are multiple forms of RNA (e.g., tRNA, rRNA, siRNA, snRNA, tracrRNA, miRNA, etc.), some of which are functional in that form (e.g., ribozymes), students will be most familiar with mRNA, which is used as a template to create polypeptides that can fold into functional proteins/enzymes.

💡 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.)

Answer: 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)

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.

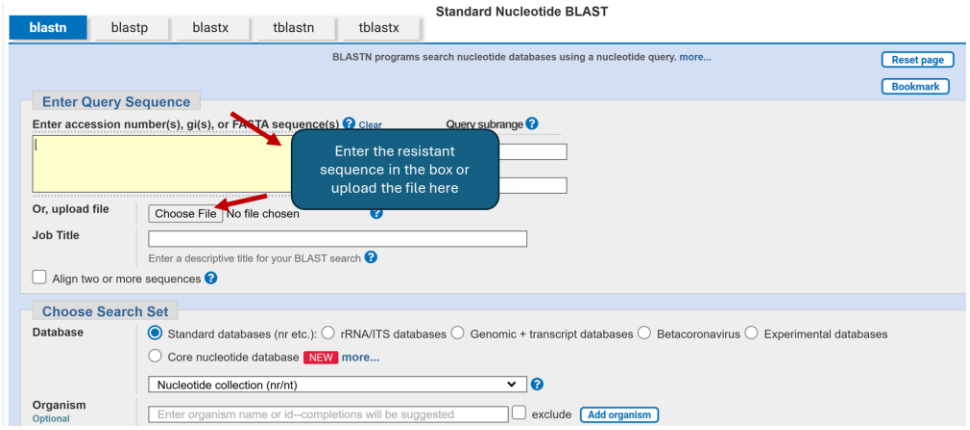
💡 Now that you are familiar with the physical and chemical properties of our compound of interest, **provide a brief description** of the effects on human health and **list 3** different disorders suspected to be associated with this chemical?

See teacher notes.

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).



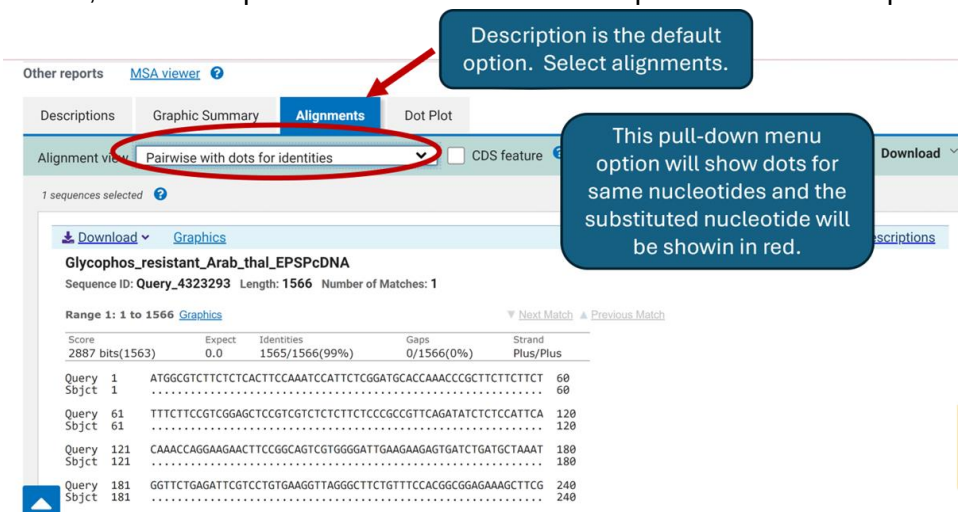
5. Click BLAST and analyze your results.

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

Arabidopsis thaliana RNA 3'-terminal phosphate cyclase/enolpyruvate transferase, alpha/beta (AT1G48860), mRNA Sequence ID: NM_103780.5
When determining with BLAST result to proceed with, consider the following:
NM_#: verified, supported by much evidence (refseq)

XM_#: predictive, experiments with less evidence at the moment
Other prefixes w/o _: individual uploads from individual researchers (least supported)

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?

One mutation at nucleotide 266

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

This is a one nucleotide change to the DNA. This mutation will be transcribed into mRNA as a U and then translated into an amino acid. The amino acid will be different in this case and change the primary polypeptide chain.

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

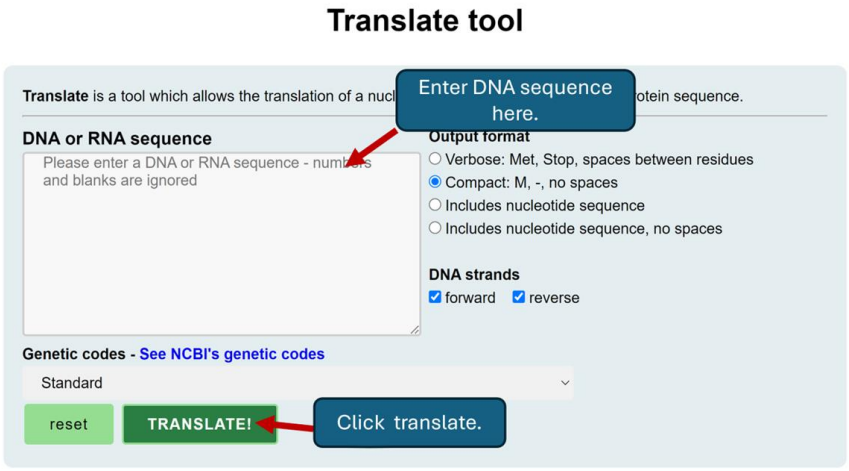
This will cause a mutation which will lead to a likely active site mutation.

There are 1-2 positively-charged AA residues in the vicinity (lys^{519} and arg^{386}) and will likely force a conformational change at the ligand site because there is a loss of a large and negatively-charged R-group from glutamic acid and replaced with alanine which has a smaller, uncharged methyl R-group

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?

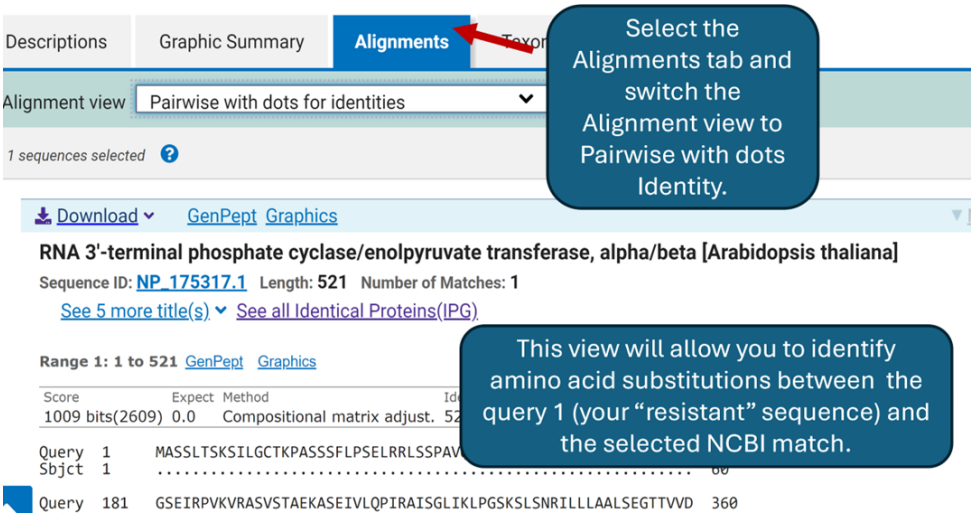
5'3' Frame 1 because it is entirely red which means that starting with the first methionine there is no stop codon until the end of the sequence. A dash in the reading frame represents a stop codon.

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)?

yes

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

Nonsynonymous, amino acid change to alanine from glutamic acid

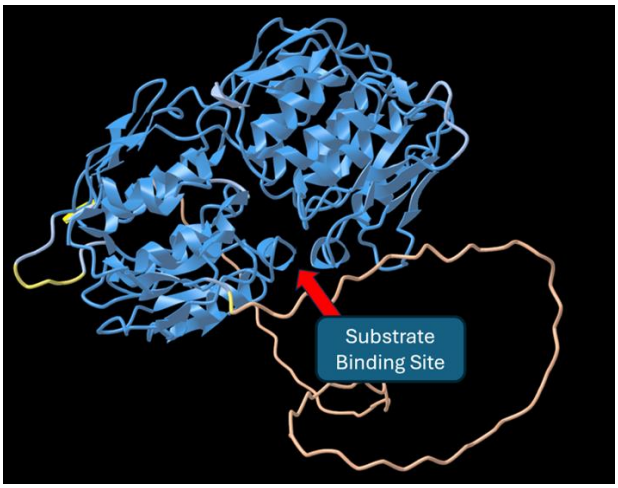
💡 What effect would this change have in the protein?

It will most likely cause an error in folding

V. Modeling of amino acid substitutions using iCn3D

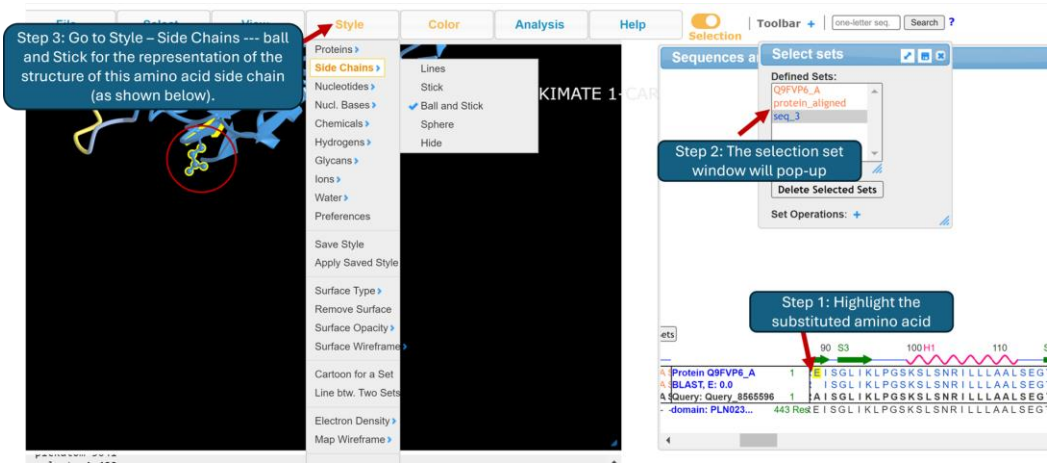
iCn3D is a web-based structure viewer tool that allows the interactive analysis of the structure. In this module, it will be used to model the structure of the protein of interest.

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)

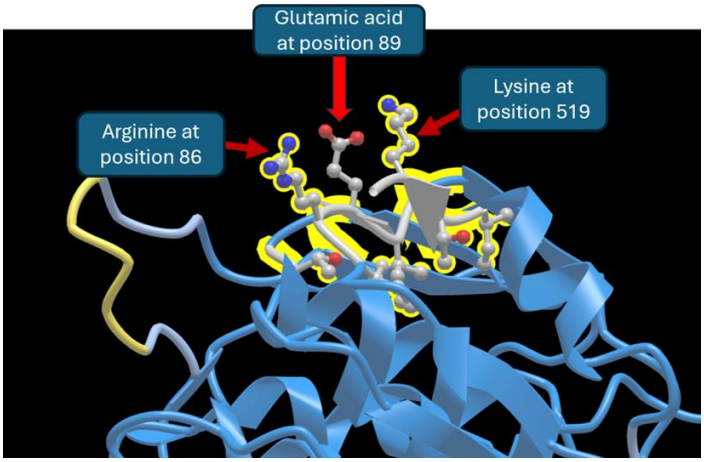


6. Use the bar under the sequence to look at the alignment of the two sequences. As previously shown in the protein alignment using BLASTp, there was a substitution at position 89 (glutamic acid to alanine, E → A).
7. Highlight the “E” and the location of this amino acid will then be highlighted in yellow 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 change of each of the labeled amino acids. What properties do they have in common?

Arginine and lysine have basic side groups (+ charge, amino group). Glutamic acid has an acidic side group (- charge, carboxylic group).

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

Alanine is a neutral amino acid. The basic arginine and lysine would not have a strong affinity for alanine.

Assessment Answers

1. How can PubMed help you answer a research question? *It is a searchable database of nearly all biomedical/biological/chemical literature. You can determine what is known about a particular research problem.*

2. What do the various forms of BLAST do? *BLAST stands for Basic Local Alignment Search Tool. This algorithm is able to rapidly and accurately search for sequence similarity between a query sequence and billions of target/subject sequences in large databases of known sequence information and display the level of statistical significance of a match.*

3.What is a cDNA? A cDNA is a DNA copy of a mRNA. *By definition it will contain the sequence of the condoning strand of the gene.*

4. Do all DNA mutations change the protein? *No, only mutations that change an amino acid identity can change the protein structure.*

In which codon positions are mutations more likely to cause protein change? *By analyzing a codon table, it can be seen that in general mutations to first and second codon positions are more likely to change amino acid identity than third codon position mutations.*

What codon position has been altered in the Glyphosate resistant gene? *The second codon position for E (Glutamic acid) has been changed to code for A (Alanine).*

5. Had there been a mutation in the third codon position of E, would that have changed the amino acid to an A? *It depends on the base substitution at the third codon position, if the change is to an A, the codon would still code for E, if the change is to a C or T the change would code for an Aspartic acid (D).*

6. In what type of protein primary structure does the non synonymous amino acid substitution from E to A take place?

A beta pleated sheet type of protein secondary structure.

7. Could a change at the protein level be detectable at the mRNA level?

Yes! We just learned about The Central Dogma.

8. What amino acids is EPSPs involved in synthesizing? EPSPs is a central part of the anabolic shikimate pathway

The shikimate pathway is found in plants and some bacteria, it is not found in animals. The shikimate pathway is involved in the synthesis of aromatic amino acids Tyrosine, Phenylalanine and Tryptophan.

9. Could a mutation in this gene cause glyphosate resistance?

Yes, but it may not be the only factor contributing to glyphosate resistance.

Additional Teacher Information

This activity is divided into a number of modules; you are welcome to expand or delete these, as necessary, to work best for your class.

More detailed information

The gene that Roundup® acts on is part of 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase (Yuan et al., 2022). While point mutations may occur in this gene, they are not common because EPSP synthase is a critical enzyme in the shikimate pathway and likely to be under purifying selection.

As a single mutation, it is found in the DNA at position 266 and is an A→C change. The codon is a GAA→GCA change (Ngo et al., 2018).

A. The mutation, in codon 89, is a switch from glutamic acid to alanine; there is a loss of a large and negatively-charged R-group (replaced with a smaller, uncharged methyl R-group).
B. There are likely 1-2 positively-charged AA residues in the vicinity (lys^{519} and arg^{386})
    1. Less than 1.5Å is usually a covalent bond and 2.6-3.1Å is usually a H-bond
    2. Forces conformational change at the ligand site (Schönbrunn et al., 2001).
C. This is likely in an active site mutation (Schönbrunn et al., 2001).
D. This can manifest as overexpression (*Zea mays*) and double mutation (*Arabidopsis thaliana*) (Sammons et al., 2018).
E. In procedure image 2, note that it is labeled as cDNA. Remember, mRNA is the same sequence as cDNA.
F. After students run their alignment, the first two results are the same ascension number (sequence).
G. EPSP synthase gene: AT1G48860

Animal metabolic processes do not include the shikimate pathway, which is found in plants, fungi, and bacteria. This implies that animals, including humans, do not suffer negative consequences from judicious use of glyphosate-containing products. Glyphosate residue is degraded by soil microorganisms with a full cycle of up to 280 days. A by-product of glyphosate metabolism is aminomethyl-phosphonic acid (AMPA) which has a full cycle degradation of up to 958 days (Costas-Ferreira, et al., 2022).

Multiple studies indicate negative impacts of glyphosate use, whether through direct contact with glyphosate or through exposure to AMPA (von Ehrenstein et al. (2019); Martinez et al. (2020); Martinez and Al-Ahmoud (2019)). These impacts include crossing of blood-brain barrier, increased incidences of autism, interference with cell death pathways, and reduced cell viability.

The purposes of this original version of activities include exploring the effects of single nucleotide alterations, determining the resulting amino acid substitutions, and predicting protein structure changes. Please note that if you modify the herbicide, target, or gene sequence you may not be able to include all of the section activities and/or use the provided answers to student questions.

Bibliography

Schönbrunn, E., Eschenburg, S., Shuttleworth, W. A., Schloss, J. V., Amrhein, N., Evans, J. N., & Kabsch, W. (2001). Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *PNAS*, *98*(4), 1376–1380. https://doi.org/10.1073/pnas.98.4.1376

Ngo, T. D., Malone, J. M., Boutsalis, P., Gill, G., & Preston, C. (2018). EPSPS gene amplification conferring resistance to glyphosate in windmill grass (*Chloris truncata*) in Australia. *Pest manag. sci.*, *74*(5), 1101–1108. https://doi.org/10.1002/ps.4573

Yuan, Y., Zhou, Z.. Zhan, Y., Ke, X., Yan, Y., Lin, M., Li, P.,Jiang, S., Wang, J., Lu, W. (2022) A Highly Glyphosate-Resistant EPSPS Mutant from Laboratory Evolution. *Appl. Sci.*,*12*, 5723. https://doi.org/10.3390/app12115723

Sammons, R. D., You, J., Qi, Y., Flasinski, S., Kavanaugh, C., Washam, J., Ostrander, E., Wang, D., & Heck, G. (2018). Evaluation of glyphosate resistance in *Arabidopsis thaliana* expressing an altered target site EPSPS. *Pest manag. sci.*, *74*(5), 1174–1183. https://doi.org/10.1002/ps.4654

Costas-Ferriera, C., Durán, R., Faro, L. (2022). Toxic effect of glyphosate on the nervous system: a systematic review. Int. J. Mol. Sci. 2022, 23, 4605. https://doi.org/10.3390/ijms23094605

von Ehrenstein, O. S., Ling, C., Cui, X., Cockburn, M., Park, A. S., Yu, F., Wu, J., & Ritz, B. (2019). Prenatal and infant exposure to ambient pesticides and autism spectrum disorder in children: population based case-control study. *BMJ (Clinical research ed.)*, *364*, l962. https://doi.org/10.1136/bmj.l962

Martínez, M. A., Rodríguez, J. L., Lopez-Torres, B., Martínez, M., Martínez-Larrañaga, M. R., Maximiliano, J. E., Anadón, A., & Ares, I. (2020). Use of human neuroblastoma SH-SY5Y cells to evaluate glyphosate-induced effects on oxidative stress, neuronal development and cell death signaling pathways. *Environ. int.*, *135*, 105414. https://doi.org/10.1016/j.envint.2019.105414

Martinez, A., & Al-Ahmad, A. J. (2019). Effects of glyphosate and aminomethylphosphonic acid on an isogenic model of the human blood-brain barrier. *Toxicol. Iett.*, *304*, 39–49. https://doi.org/10.1016/j.toxlet.2018.12.013