Bioinformatic workshop - iGEM 2026 - AnTigone

- You have been given 2 mysterious **unknown amino acid sequences**
- Groups: 2-3 people
- The mission: Using bioinformatic tools, find which protein corresponds to each sequences
- What is the **function** of the proteins?
- Are there any similarities or differences between the proteins?

A:

>seq_A|

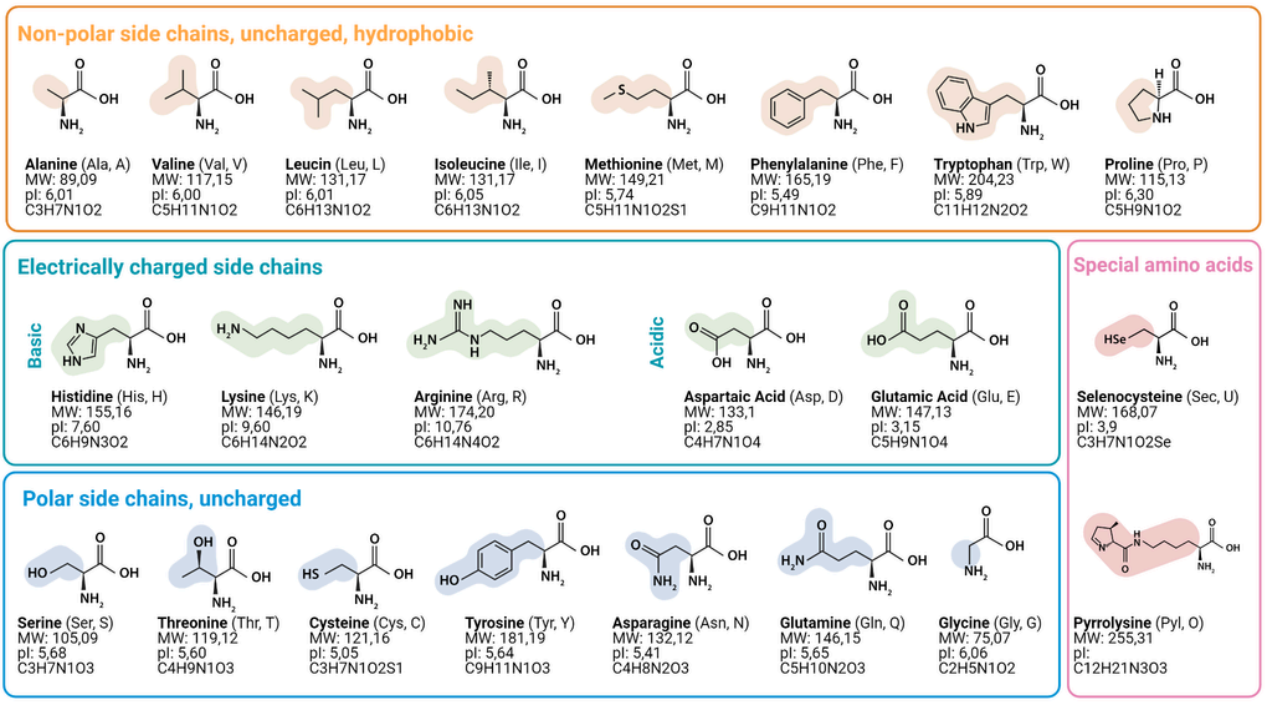
IQLIPPLINLLMSIEPDVIYAGHDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKSLPGFR
NLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSGQMLYFAPDLILNEQRMKESSFYSLCLT
MWQIPQEFVKLQVSQEEFLCMKVLLLLNTIPLEGLRSQTQFEEMRSSYIRELIKAIGLRQ
KGVVSSSQRFYQLTKLLDNLHDLVKQLHLYCLNTFIQSRALSVEFPEMMS

B:

>seq_B|

TVSHIEGYECQPIFLNVLEAIEPGVVCAGHDNNQPDSFAALLSSLNELGERQLVHVVKWA
KALPGFRNLHVDDQMAVIQYSWMGLMVFAMGWRSFTNVNSRMLYFAPDLVFNE
YRMHKSRMYSQCVRMRHLSQEFGWLQITPQEFLCMKALLLFSIIPVDGLKNQKFF
DELRMNYIKELD

Amino acids:

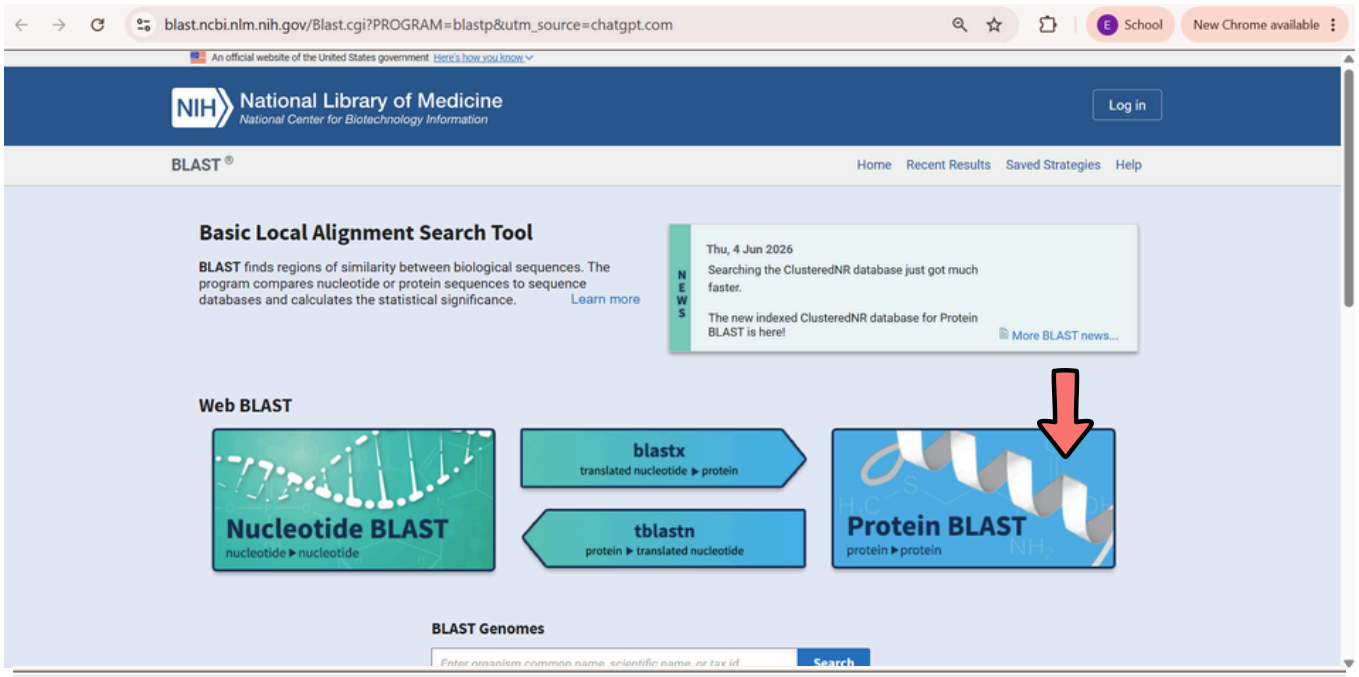


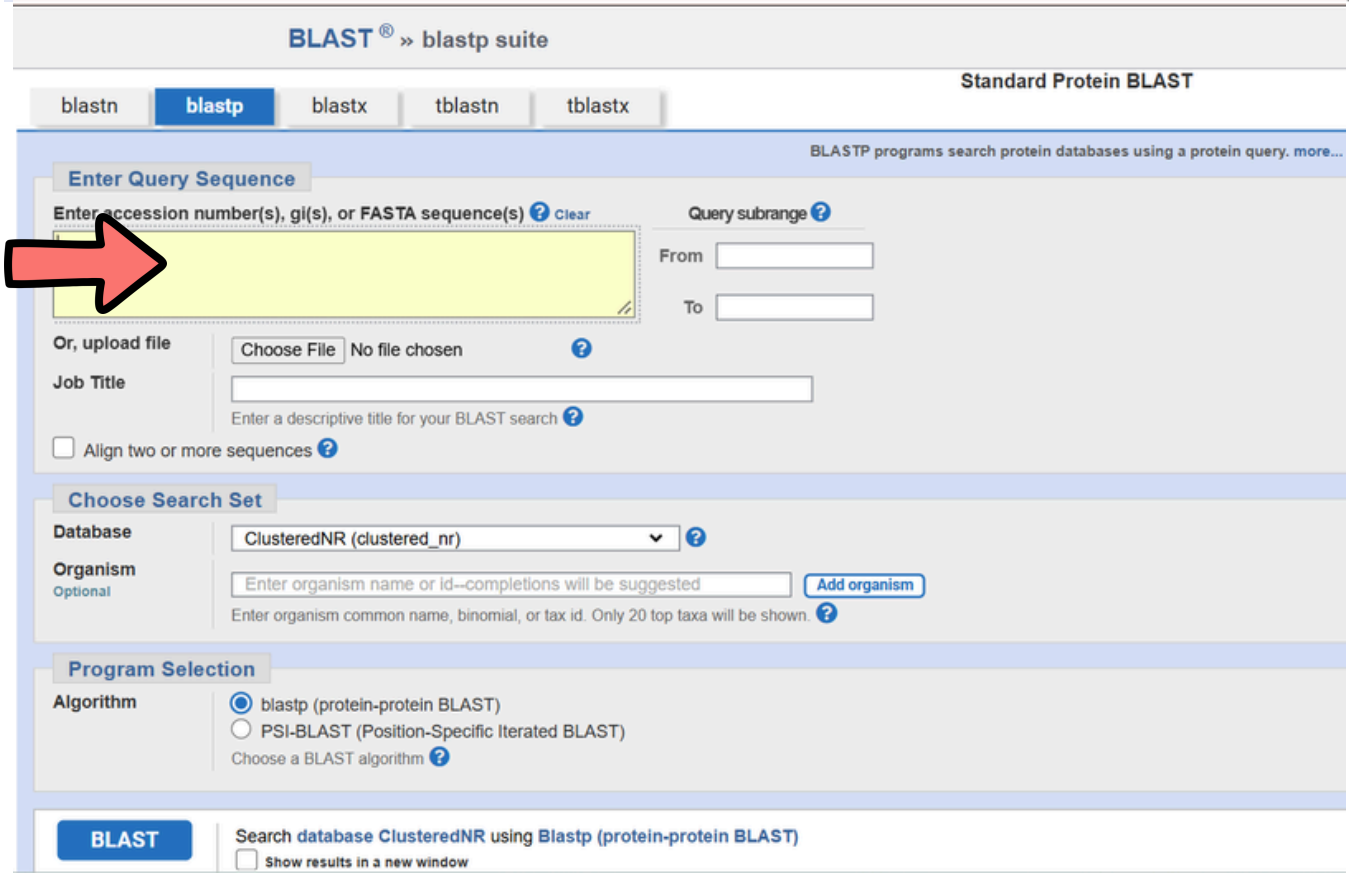
https://www.jpt.com/media/3e/92/82/1772095973/Amino%20Acid%20Chart.pdf

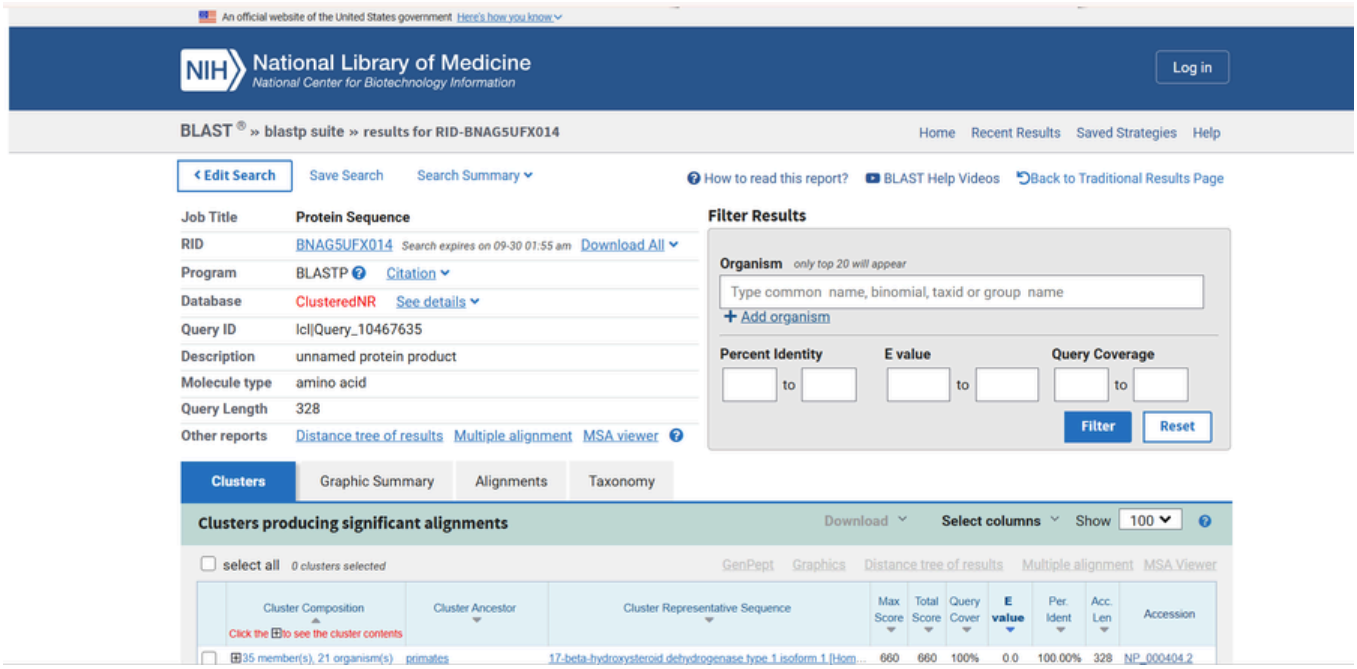
NCBI BLAST (Basic Local Alignement Search Tool)

BLAST is an algorithm and program for **comparing primary biological sequence** information, such as the **amino-acid** sequences of proteins, nucleotides of **DNA** and/or **RNA** sequences. For example, following the discovery of a previously unknown gene in the mouse, a scientist will typically perform a BLAST search of the human genome to see if humans carry a similar gene; BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of sequence (source: Wikipedia).
https://blast.ncbi.nlm.nih.gov/Blast.cgi .

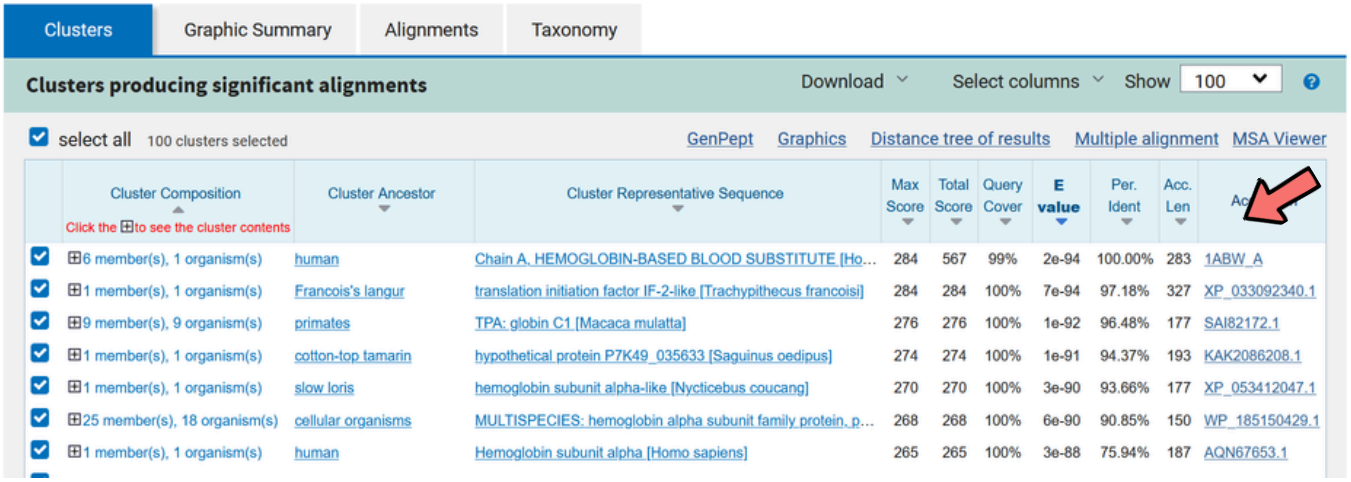
1. Copy the protein sequence into the protein BLAST tool, **blastp.**





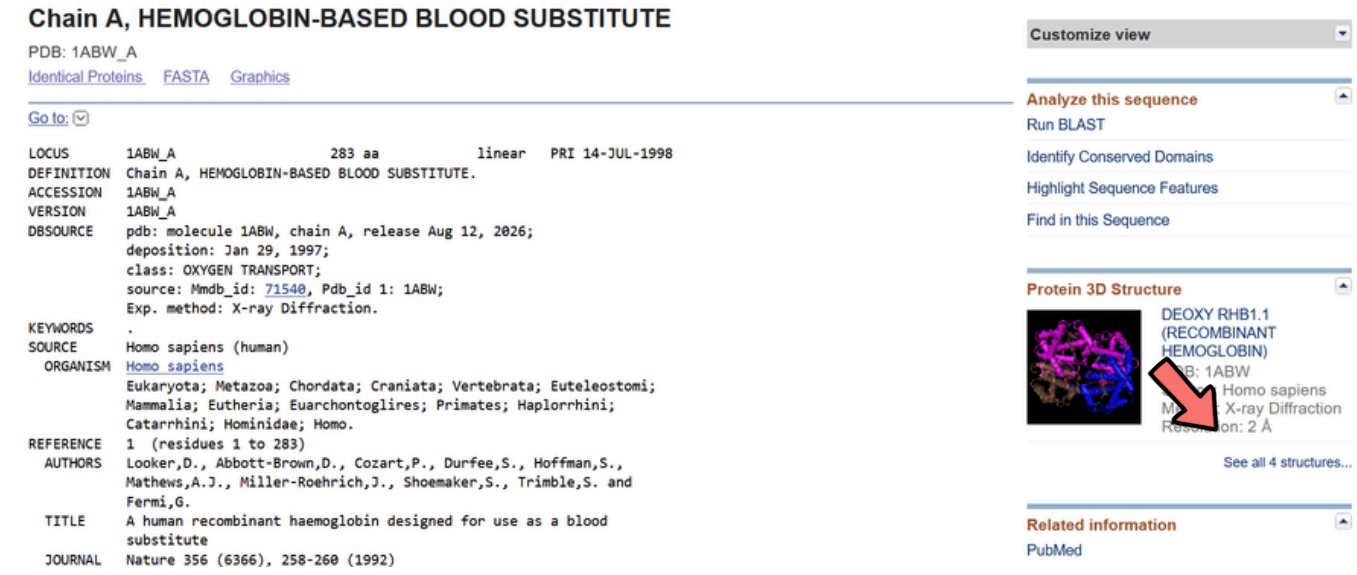


2. Choose the protein with the **highest identity.**
3. What is the protein **called**?
4. What **organism** does it come from?
5. Click on the **BLAST identifier** of your chosen protein (Hemoglobin sub-unit A in the example). A new page will open, with information about who originally characterized the sequence and its properties.



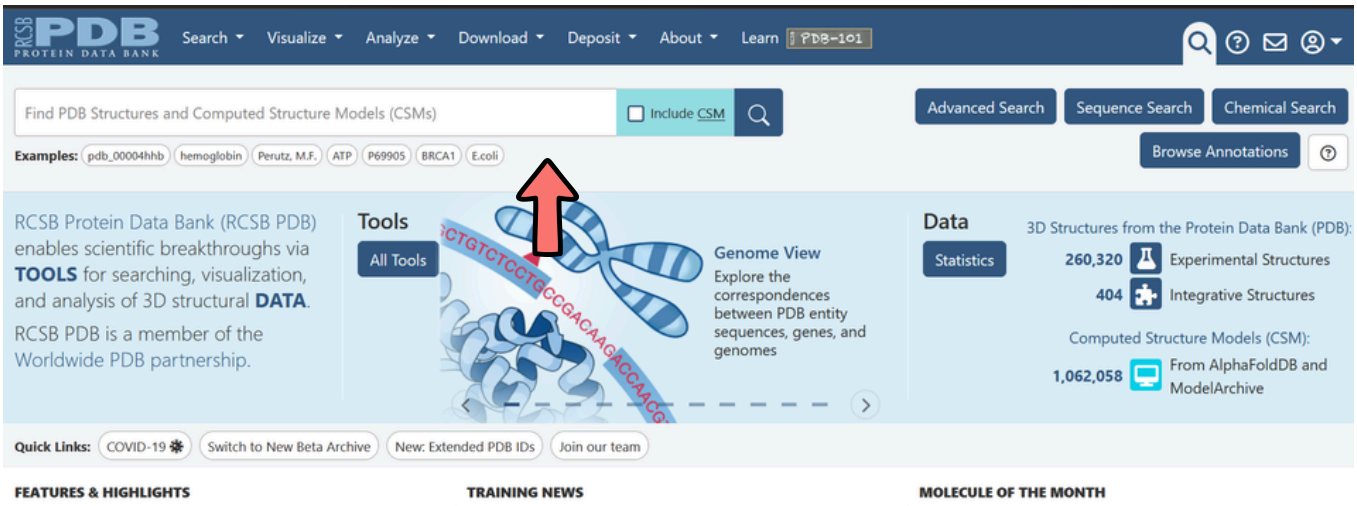
PDB (Protein Data Bank)

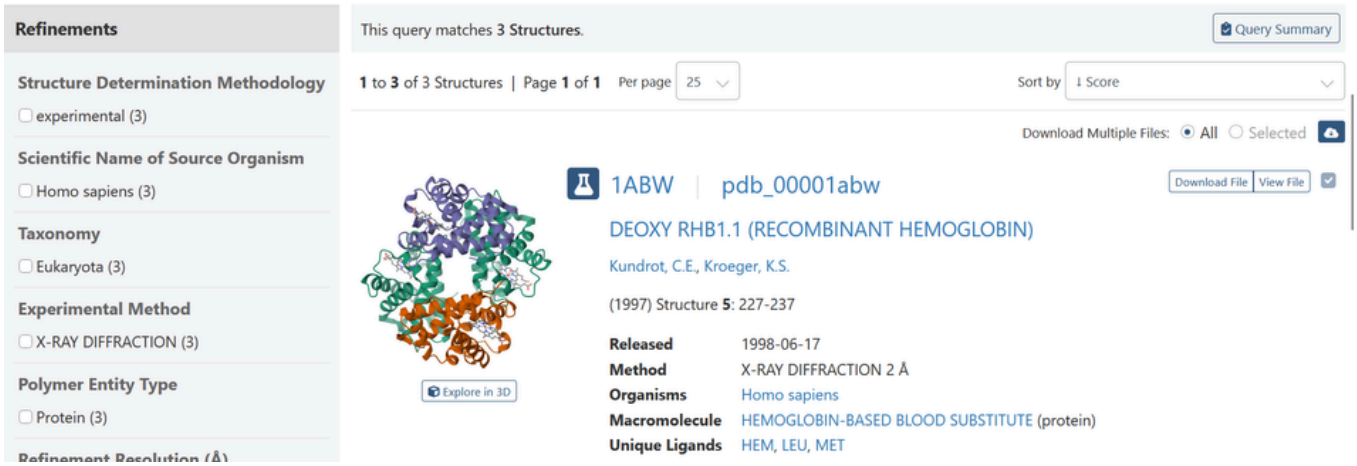
Look for the **PDB 3D structure** in the NCBI record. On the right side of the page, it is possible to visualize the protein 3D structure. Among the information provided, the **PDB identifier** can be found. This is required to search for the same entry on the **PDB database**, which contains the three-dimensional structural data of large biological molecules such as proteins and nucleic acids, along with their physical and chemical properties, experimental data, ligands (molecules they bind and interact with) and more.



1. Copy the PDB/Uniprot **Accession number** -> The database identifier.
2. Paste the Accession number into the UniProt/PDB webpage.

UniProt/PDB
https://www.uniprot.org/
https://www.rcsb.org/





2. What is the function of each protein?
3. View the 3D protein structure
4. Repeat this for both unknown sequences to find the corresponding proteins

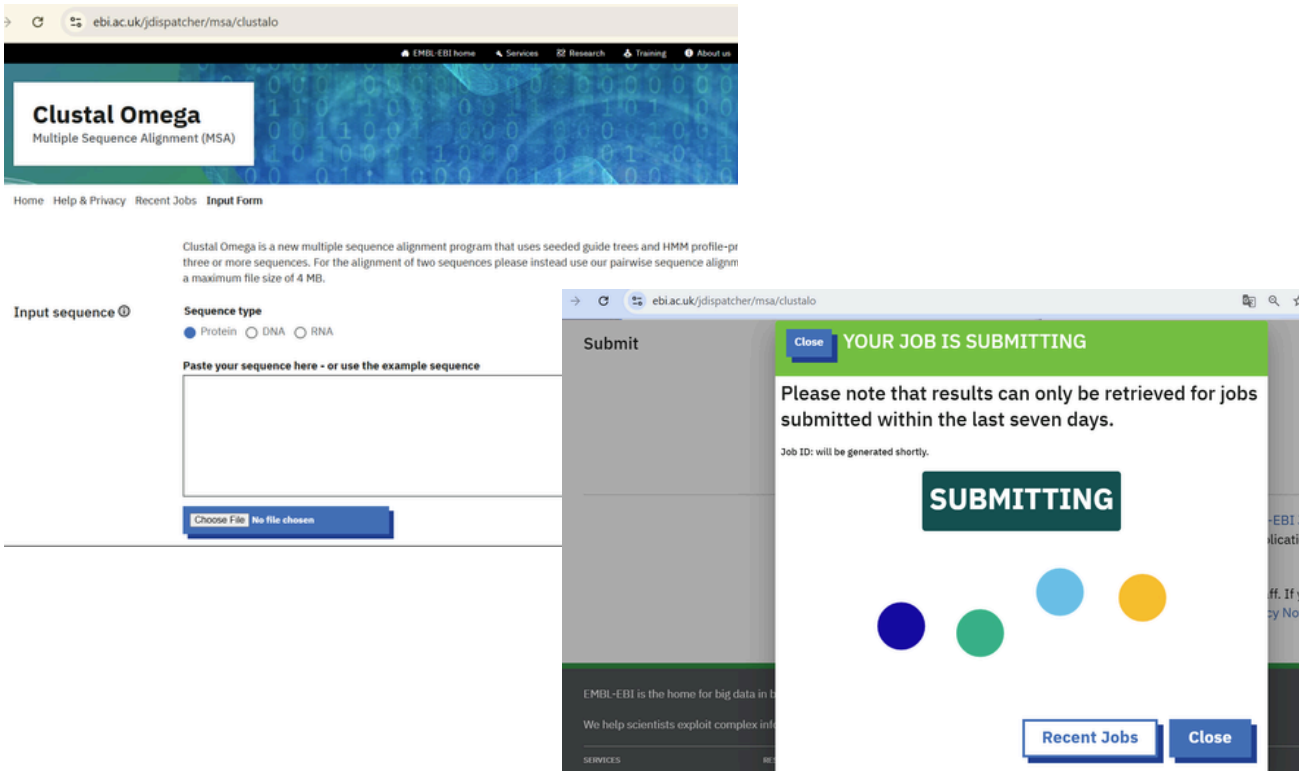
Clustal Omega - Multiple Sequence Analysis

Clustal Omega is a program used for multiple sequence alignement. While BLAST allows to compare the given sequence to a library of already known sequences, to find the closest match, Omega is used to compare multiple sequences and look for similarities and differences, and identify conserved regions. Imagine to have two nucleotide sequences:

AATGCATCCGGCCA
AA_G**T**ATCCG**T**CCA

In this case it is easy to point out where the two sequences differ. Gene and protein sequences however are very long and complex to compare. This tool does the job and provides a score for the level of **homology** between sequences. This is very useful to see how much a molecule that is common between different organism has changed and how much it has been **conserved.**
https://www.ebi.ac.uk/jdispatcher/msa/clustalo

1. Paste your sequences into Clustal Omega
2. Which regions are **conserved**?



Present your final results:

- Which proteins did you find?
- How do they differ from each other in function?
- What were the conserved regions?



Resources to learn more:

- **The Mysterious Protein: A Bioinformatics Expedition**: The Mysterious Protein: A Bioinformatics Expedition
- EMBL recorded webinar: **A guide to UniProt for students**: https://www.ebi.ac.uk/training/events/guide-uniprot-students-2025/