# Lab sheet#7
## Protein sequence alignment and ExPASy Tools

**Objective:**

- To know how to use BLSTP, and find similarities between protein sequences.
- To know how to use Clustal W/Omega, and find similarities between multible sequences.
- Translation of a nucleotide sequence to a protein sequence using ExPASy.
- Analysis of protein primary structure and computing various physical and chemical parameters

## Part 1:

### A. Protein sequence search using NCBI database and BLASTP web page:

- Navigate to the **NCBI BLASTP** web form and write the accession number (**CAG33009.1**) of homo sapiens X-ray repair cross complementing 1 (**XRCC1**) into the query window.
  - Choose the "non-redundant protein sequences (nr)" as the database to be searched.
  - To save lots of time, restrict your search to the organism under study in the "Organism field". In this example we are looking for sequence similarity with **mouse** (taxid:10088).
  - Pick "blastp (protein-protein BLAST)" as the program to be used in the search.
  - Launch the search by clicking on the "BLAST" button.

### B. Multiple protein sequence alignment using Clustal Omega:

**Multiple Sequence Alignment (MSA)** is a way of arranging three or more biological sequences (protein or nucleic acid) to identify regions of similarity that may be a consequence of functional or structural relationships between the sequences.

Use Clustal Omega to perform multiple sequence aliment as follows:

1. Navigate to the **NCBI BLASTP** web form and write the accession number (**NP_065826.2**) of homo sapiens Estrogen-induced gene 121 protein into the query window.
   - Choose the "non-redundant protein sequences (nr)" as the database to be searched.
   - Pick "blastp (protein-protein BLAST)" as the program to be used in the search.
   - Launch the search by clicking on the "BLAST" button.
2. In BLAST result page, click **Taxonomy Reports**. Showing the similar proteins in other organisms.
3. In organism report, copy the accession number of the first hit under interested organisms.
4. Paste these accession numbers with (,) between them in NCBI protein database search box. Change the Display setting to **FASTA text** and copy all sequences.

5. Open **Clustal Omega** and paste the sequences in the input box. Change the first part of each sequence to the organism name.
6. Pick "**protein**" as the sequences to be aligned are proteins. Choose " **ClustalW with character count**" as the output format and click **Submit**.
7. Click "**show colors**" in the result page.
8. All the sequences now are aligned under each other, with different signs represented under each amino acid position **(*, . , : , -)** showing the extent of similarity.
9. Download the alignment file.

## Part 2:

### A. Translation of a nucleotide sequence to a protein sequence using ExPASy web page:

1. Retrieve the FASTA format of homo sapiens X-ray repair cross complementing 1 (XRCC1) mRNA sequence (**NM_006297.2**) from the NCBI GenBank database.
2. Use the translate tool on ExPASy website:
   - Paste the retrieved XRCC1 mRNA sequence into the box.
   - Click the "**Translate sequence**" button.
   - Choose the most reliable predicted protein sequence.

### B. Primary Structure Analysis of a Protein Using ProtParam on the ExPasy server:

1. Search for ProtParm, click the first link.
2. Copy the protein sequence of XRCC1 protein (**CAG33009.1**), then paste in the ProtParm webpage.
3. Click compute parameters.
4. Find out the Molecular weight of the protein.
5. How many Cysteine amino acids located in the protein sequence.
6. What is the total number of negatively charged residues.
7. Find out the Estimated half-life of the protein in mammals and yeast.