

**EE550**  
**Computational Biology**  
**Term Project**  
**Due 26.05.2022**

As your term project, you are asked to carry out research on a human protein or protein complex using online gene and protein databases and the applications provided by them. The project will consist of the following three specific aims.

**Specific aim 1: Identify a well-connected human protein of interest in the KEGG Pathway database**

For drug research, proteins of interest tend to be those proteins or protein complexes that interact with numerous other proteins or protein complexes. The KEGG Pathway database contains information on the signal transduction pathways known to date. In this phase of the project, you will browse through the molecular pathways in humans and identify a well-connected protein or protein complex in a signaling pathway, and briefly provide the background information on the pathway and the identified protein. Then, you will list all the proteins and protein complexes that your selected protein interacts with in all other pathways it figures.

**Specific aim 2: Identify common sequence elements among the set of interacting proteins**

Common sequence features of the proteins that interact with your selected protein are to be identified either using motif finding tools or via multiple sequence alignment procedure using Clustal Omega tool provided by the EMBL-EBI and grouped into protein clusters that bear significant sequence similarity. Next, prominent and common sequence features will be determined for each cluster. These common sequence elements will be indicative of their interactions with the original protein; and thus, they are to be evaluated in terms of specific sequence motifs or structural elements they may be associated with.

**Specific aim 3: Query sequence databases for additional interacting proteins**

Since the sequence elements are indicators for possible interaction with the original protein, they are to be queried in sequence databases to see if there are other proteins to which it may be interacting. Note that this search needs to include not only the human proteins but also viral and bacterial proteins in case they may figure in host-pathogen type of interactions. For each additional protein identified to possess the sequence features of interest, a literature search is to be carried out to see if there is any evidence indicating possible interactions between the newly identified protein and the original one.

Your work should be returned in the form of a 6-10page technical report written in 10pt Times New Roman font by e-mail as a single PDF document, complete with the following sections:

**Title** – The subject matter of your study with your name and school identification number

**Background** – The background information on your protein and its significance as well as the objective of this project

**Material and Method** – The specific methodology pursued to achieve the stated objective

**Results** – The results obtained by carrying out the methodology described above

**Discussion** – The significance of your results

**References**