📘 Bioinformatics assignment 2: Proteomics and PPIs

Deliverables: Jupyter notebook / Python code

Deadline: Wednesday, 3.12.2025 23:59

🧪 The Idea

You have to perform an end-to-end analysis of a real human protein involved in a disease. Select **one protein** from the list below (UniProt accessions provided):

Protein	UniProt
BRCA1	P38398
KRAS	P01116
EGFR	P00533
PTEN	P60484

📥 Detailed Assignment Steps

🔷 Step 1 — Retrieve core data from UniProt

Using the UniProt REST API obtain:

- Protein **sequence length**
- List of **Posttranslational Modifications** (phosphorylation, ubiquitination, glycosylation, etc.)
- Known **disease associations**
- Associated **PDB IDs**
- **STRING ID** (xref field)

Guiding questions

- What types of PTMs does your protein undergo?
- Are these related to its function or disease role?

🔷 Step 2 — Select one PDB structure and visualize it

Use the RCSB API to select **one PDB structure**, download it from RCSB database and visualize it with a tool of choice.

🔷 Step 3 — Build a STRING-based interaction network

Call the STRING REST API:

- Download **first-neighbor interactions** of the chosen protein
- Expand to **second neighbors (two hops)** using iterative API calls
- Build a **graph** using NetworkX where Nodes = proteins, edges = interactions (with scores)

Required outputs

- Graph visualization (NetworkX + spring layout)
- Number of nodes and edges

Guiding questions

- Does your protein act as a hub?
- How dense is the surrounding interactome?

🔷 Step 4 — Compute graph centralities

Calculate:

- Degree centrality
- Betweenness centrality
- Closeness centrality

Required outputs

- Table of top 10 nodes by each centrality
- Highlight your protein in the rankings

Guiding questions

- Is your protein a “bottleneck”, “hub”, or “peripheral node”?
- Which other proteins appear to be key regulators?

🔷 Step 5 — Simulate knockout of the protein

Remove the selected protein **and all its edges** from the graph.
Compute:

- Change in number of connected components
- Change in network diameter
- Centrality changes of other top nodes

Required outputs

- Before vs. after knockout metrics
- Visual comparison of the graph

Guiding questions

- Does the network fragment when the protein is removed?
- Which nodes take over regulatory roles?
- Could this protein be a potential therapeutic target?