Protein Networks and Systems Biology

ELIXIR 2015, Bologna, Italy

**Web resources describing results of PPI experiments**

Luana Licata

Bioinformatics and Computational Biology Unit

Molecular Genetics Laboratory

University of Rome Tor Vergata, Italy

**Introduction**

## **UniProt, the protein sequence database**

UniProt (Universal Protein Resource) is the world's most complete collection of proteins information. UniProt is a central repository of protein sequence and function made by cooperation between the European Bioinformatics Institute, the Swiss Institute of Bioinformatics and the Protein Information Resource, Washington. The UniProt consortium goal is to provide high quality database functioning as a stable, complete, and accurately annotated protein sequence knowledgebase. The query interfaces is freely available to the scientific community.

The **UniProt Knowledgebase** (UniProtKB) is divided into two sections UniProtKB/SwissProt, which is manually curated, and UniProtKB/TrEMBL, which is automatically maintained.

In this practical part of the course, you will learn how to access the entries in the database and which information each entry contains.

## **Step1: Exploring UniProtKB**

Go to the web page: <http://www.uniprot.org> and use quick search bar to search the term ‘Ebola’ then click Search.

This allows you to search all the entries in UniProtKB where a protein has been identified as playing a role in the human disease, ebola.

**Q1:** How many proteins play a role in the human disease ebola? How many reviewed proteins and between then how many are human proteins?

From the results list, select Q05128, the **Matrix protein VP40** from Zaire Ebola Virus (strain Mayinda-76). Go to the display panel and select ‘Name & taxonomy’ field. Click on the hyperlink ‘Proteome’ to access all proteins within that reference proteome.

## **Q2:** How many proteins are described in the Proteomes - Zaire ebolavirus (strain Mayinga-76)?

**Step2: UniProt and PDB**

Go back to the Q05128 protein page, go to the display panel and select ‘Cross-reference’ field.

**Q3:** How many times has the protein been crystallized?

Click the RCSB PDBi hyperlink for crystal 1h2c to access the details of this crystal.

**Q4.** Does this protein crystallize as a monomer or a multimer? Which is the exact protein **Stoichiometry?**

**Step3: UniProt and PPI**

Go back to the main search page; type the protein ID ‘PTPN1 (P18031)’ and click search. Go to the display panel and select ‘Interaction’ field.

**Q5:** How many binary interactions are associated to this protein?

Among the several interactors, it is shown that PTPN1 interacts with BCAR1(P56945). Click on the hyperlink to look at this protein in IntAct.

**Q6**: What reaction does this enzyme catalyse?

#### Go back to the UniProt display panel and select ‘Cross-reference’ field. Then go to Protein-protein interaction databases.

**Q7:** How many databases contain interactions in which PTPN1 is involved?

**Step4: Protein isoforms**

Go back to the main search page; type the protein ID ‘**DLG1** (Q12959)’ and click search. Go to the display panel and select ‘Sequence’ field.

**Q8:** How many isoforms have been identified? All of these are mapped within UniProtKB, and given stable identifiers.

Press the button “Align” to see how all the isoforms differ.

Return to the Sequence section.

**Q9:** Which isoform has a shorten form?

Go to the display panel and select ‘Family & Domains’ field.

**Q10:** Which domains contain the protein?

**Q11:** To which family belong?

**Q13:** Which is the InterPro ID corresponding to the SH3 domain??