** SIB course: Protein sequence databases**

**Exam – March 2020**

**http://education.expasy.org/cours/SIB\_UniProtKB\_2020\_exa/**

**First Name:**

**Last Name:**

**Question Protein sequence databases**

**From UniProtKB**

**- Look for the Candida Albicans reference proteome**

What is the Proteome ID of the Candida Albicans Reference Proteome ?

To which strain does it correspond ?

How many proteins ?

How many Swiss-Prot entries ?

How many TrEMBL entries ?

How many proteins have been proved to exist at the protein level ? (Protein existence)

How many proteins with a transmembrane domain (KW Transmembrane) and a phosphohistidine?

Where does the information about the phosphohistidine site come from?

**Question: Automated annotation**

You have isolated in a screen protein Zgc:171429 (A8KB11\_DANRE) from zebrafish (Danio rerio). Can you find out using HAMAP-Scan (<https://hamap.expasy.org/>) what kind of protein it is (Protein name?) and by which HAMAP annotation rule it will be annotated?

The following is an excerpt of the annotation rule:



If you submitted A8KB11\_DANRE to HAMAP Scan & Annotate, which subcellular location annotation would you find in the resulting protein annotation:

1. Cytoplasm
2. Mitochondrion
3. Nothing

**Question: How to mine enzyme data ?**

**Enzyme-catalysed reactions are annotated in UniProtKB in the Function / Catalytic activity section, using the Rhea annotation (at pH 7.3)**

*Biological context*: Trimethylamine (TMA) is produced by the gut microbiome from dietary choline. TMA is then absorbed and converted to the pro-atherogenic molecule trimethylamine N-oxide (TMAO).

Find the bacteria and the human enzymes involved in the reactions.

* Use the compounds names
* Use the InChIKey (ChEBI)

Beware the pH and the names / synonyms of the different compounds.