

Practical 3

22 Feb 2024

Questions

1. Find the amino acid sequence of human mitochondrial β barrel membrane protein VDAC1 and its function? How many transmembrane segments are present in the protein?
2. Obtain the sequences of “transcription factors” with 50% sequence identity in FASTA format. List the count of sequences and count of clusters.
<http://www.uniprot.org/uniprot/>
3. How many protein sequences from Homo sapiens are obtained at identity cutoff of 100%, 90% and 50% sequence identity?
4. In UniProt, how many mouse (*Mus musculus*) protein sequences are manually annotated? And how many of these manually annotated protein sequences are associated with PDB (3D structures)?
Hint: Use “Advanced” search
5. Map first 10 UniProt IDs of above manually curated mouse protein sequences with 3D structures to STRING database. How many STRING IDs are mapped?
Hint: Use “Retrieve/ID mapping”, upload Uniprot IDs as text file
6. Using UniProt Statistics data, answer the following
 - a) What do you infer from the distribution of sequence length in UniProt?
 - b) The shortest and longest sequence in UniProtKB
 - c) Amino acid composition in percent for the complete database

Deadline: 3 Mar 2024