

Q1

Please find the UniProt accession for human GPR75. Please find its function, publications, Pfam domain(s), 50% similar proteins in its UniProt webpage.

Use the mapping tool in UniProt website to map the UniProt accessions of 50% similar proteins to GI numbers. Attach the mapping table as your answer. Then please write a script using EFetch to download the full record text file for all GIs. Please attach your script and the text files you download as your answer. For the usage of EFetch, please read <https://www.ncbi.nlm.nih.gov/books/NBK25499/#chapter4.EFetch>

Q2

List the major databases of nucleic acid sequences, protein sequences, protein families and protein structures.

Q3

What are the three major structural biology methods? What are the limits of these methods?

Q4

Use NCBI Assembly and NCBI Taxonomy to find out how many chordata species has been whole genome sequenced? Note, not the number of assemblies. Use Taxonkit tool to find the class of these species, how many species in these classes have been sequenced?