

Example of a Tutorial Worksheet

Tutorial on Databases and Servers

1. **UniProt** database– <https://www.uniprot.org> - information about proteins, partly manually annotated and part from computer algorithms
search for ACE2 human, UniprotID [Q9BYF1](#)

additional names

functions

amino acid sequence in FASTA format

GOterms – Gene Ontology, standardized terminology

E.C. number

Cellular location

2. **BLAST** server– <https://blast.ncbi.nlm.nih.gov/Blast.cgi> - searching databases for similar DNA or protein sequences (and structures) and aligning them
Search with FASTA sequence from ACE2
What are top hits with “nonredundant” selected? What % identity? What % coverage?

What are top hits with model organisms? What % identity? What % coverage?

What are top nonhuman hits? What % identity? What % coverage?

Make multisequence alignments with different parameters

What are top PDB hits? What are the PDBIDs? What are % coverage? What are % identity? – fill in worksheet

3. MoonProt database—www.moonlightingproteins.org – our lab's database of proteins with 2 or more functions
what are other functions of ACE2?

4. TMHMM server - <http://www.cbs.dtu.dk/services/TMHMM/> - to look for TM helices, although this doesn't find beta-barrel TM proteins

How many TM helices does ACE2 have?

Where are they located in the amino acid sequence?

How many TM helices does PGI have?

```
>sp|P06744|G6PI_HUMAN Glucose-6-phosphate isomerase OS=Homo sapiens OX=9606
GN=GPI PE=1 SV=4
MAALTRDPQFQKLQQWYREHRSELNLRRLFDANKDRFNHFSLTLNTNHGHILVDYSKNLV
TEDVMRMLVLDLAKSRGVEAARERMFNGEKINYTEGRAVLHVALRNRNTPILVDGKDVMP
EVNKVLDKMKSFQQRVRSRSGDWKGYTGKTITDVINIGIGGSDLGPLMVTEALKPYSGGPR
VWYVSNIDGTHIAKTLAQLNPESLFI IASKTFTTQETITNAETAKEWFLQAAKDPSAVA
KHFVALSTNTTKVKEFGIDPQNMFEFWDWVGGRYSLWSAIGLSIALHVGFDNFEQLLSGA
HWMDQHFRTPLEKNAPVLLALLGIWYINCFGCETHAMPLPYDQYLHRFAAYFQQGDMESEN
GKYITKSGTRVDHQTGPIVWGEPTNGQHAFYQLIHQGTKMIPCDFLIPVQTQHPIRKGL
HHKILLANFLAQTEALMRGKSTEEARKELQAAGKSPEDLERLLPHKVFEGNRPTNSIVFT
KLTFFMLGALVAMYEHKIFVQGI IWDINSFDQWGVELGKQLAKKIEPELDGSAQVTSHDA
STNGLINFIKQQREARVQ
```

How many TM helices does CXCR4 have?

>sp|P61073|CXCR4_HUMAN C-X-C chemokine receptor type 4 OS=Homo sapiens
OX=9606 GN=CXCR4 PE=1 SV=1

MEGISIYTS DNYTEEMGSGDYDSMKEPCFREANANFNKIFLPTIYSIIFLTGIVGNGLVI
LVMGYQKKLRSM TDKYRLHLSVADLLFVITL PFWAVDAVANWYFGNFLCKAVHVIYTVNL
YSSVLILAFISLDRYLAI V HATNSQRPRKLLAEKV VYVGVWIPALLLTIPDFIFANVSEA
DDRYICDRFY PNDLWVVVFQFQHIMVGLILPGIVILSCYCI I I SKLSHSGKHQKRKALKT
TVILILAFFACWLPYYIGISIDSFILLEI IKQGCEFENTVHKWISITEALAFFHCCLNPI
LYAFLGAKFKTSAQH ALTSVSRGSSLKILSKGKRGGHSSVSTESSESSFHSS

How many TM helices does Tar have?

>sp|P07017|MCP2_ECOLI Methyl-accepting chemotaxis protein II OS=Escherichia
coli (strain K12) OX=83333 GN=tar PE=1 SV=2

MINRIRVV TLLVMVLGV FALLQLISGSLFFSSLHHSQKSFVVS NQLREQQ GELTSTWDLM
LQTRINLSRS AVRMMMDSSNQSN AKVELLDSARKTLAQAATHYK KFKSMAPLPEMVATS
RNIDEKYKNYYTAL TELIDYLDYGNTGAYFAQPTQGMQ NAMGEAFAQYALSSEKLYRDIV
TDNADDYRFAQWQLAVIALV VVLLILLVAWYGIRRM LLLTPLAKIIAHIREIAGGNLANTLT
IDGRSEMGDLAQSVSHMQ RSLTDTVTHVREGSDAIYAGTREIAAGNTDLSSRTEQQASAL
EETAASMEQLTATVKQNADNARQASQLAQSASDTAQHG GKVV DGVVKT MHEIADSSKKIA
DIISVIDGIAFQTNILALNA AVEAARAGEQGRGF AVVAGEVRNLASRSAQA AKEIKALIE
DSVSRVDTGSVLVESAGETMNNIVNAVTRVTDIMGEIASASDEQSRGIDQVALAVSEMDR
VTQQNASLVQESAAAAA ALEEQASRLTQAVSAFRLAASPLTNKPQT PSRPASEQPPAQPR
LRIAEQDPNWETF