[Bioinformatics: Computer methods in molecular biology](#)
ICGEB Trieste
Elisabeth Gasteiger, June 30, 2004
Protein databases, Sequence analysis, Similarity search and multiple alignment

Starting point: http://www.expasy.org **and** http://www.expasy.org/tools/ **in particular.**
Please use the ExPASy mirror sites.

General remark: Swiss-Prot is highly cross-referenced, so make use of these cross-references to get from Swiss-Prot to corresponding entries in other databases!

1) Look for the amino acid sequence of **human carbonic anhydrase 2** in Swiss-Prot.
   Retrieve the corresponding nucleic acid sequence entries from the EMBL and GenBank databases. (Why are there more than one?)
   Look at a PDB file. (Why are there more than one?)
   Find the chromosomal location of human carbonic anhydrase 2
   - o via MIM (Mendelian Inheritance in Man)
   - o or via GDB (Genome Data Base) – linked from GenBank
   Is carbonic anhydrase known for maize and drosophila? (possible appproaches: Swiss-Prot/TrEMBL or EMBL and/or the genomic databases MaizeDb/Flybase (as the gene might be known but not sequenced))

2) Perform a similarity search (blastp) with the sequence P23470 (**PTPG_HUMAN**). (Use the Blast form at the bottom of the Swiss-Prot entry on ExPASy with the default parameters, but limiting your search to human sequences.)
   (In particular: Look at the scores for sequence O60420 and the alignments)

   Compare the results obtained with or without the « low-complexity filter ».

   Re-run the similarity search, but this time select « identity blast ». Compare the results to your previous search, looking again at O60420.

   Run a blastp only with the domain "carbonic-anhydrase like" of PTPG_HUMAN (use the tools available after clicking on the domain positions in the feature table of the Swiss-Prot entry on ExPASy).

   Compare the blastp results of PTPG_HUMAN against UniRef100, UniRef90 and UniRef50. Try to find O60420 again.

3) Use SRS to extract from Swiss-Prot all protein entries for which GPI anchor attachment sites have been determined experimentally. Documents to consult: Swiss-Prot user manual, http://www.expasy.org/sprot/userman.html, and "Swiss-Prot annotation: how is biochemical information assigned to sequence entries": http://www.expasy.org/cgi-bin/lists?annbioch.txt)

4) Perform a multiple alignment of all alternative splice isoforms of the human gene GYG2 (Retrieve the corresponding sequences in fasta format from the « Alternative Products » comment in the Swiss-Prot entry). Compare the alignments obtained by using ClustalW and T-Coffee.

5) Determine the potential transmembrane regions of the protein P11244 (http://www.expasy.org/tools/#topology). Compare your results with the ones you get when using hydrobicity scales (ProtScale).
   Look at the 3D structure of this protein (PDB entry 1LDA) and visualize the transmembrane regions.

   Use ScanProsite to detect PROSITE motifs in P11244. Do not "exclude motifs with a high probability of occurrence". What is the biological significance of the matches found?

   Look at the Swiss-Prot annotation of P11244.