Practical Session #4: Protein sequence analysis

https://www.uniprot.org/help/advanced_search
https://www.uniprot.org/help/customize
https://www.uniprot.org/help/query-fields
https://www.uniprot.org/help/sequence_annotation
https://www.uniprot.org/help/api

I. Find out the human protein entries at UniProt/SwissProt that fulfill the following criteria:

a) It contains a protein that is the ortholog to the mouse protein with refseq accession number NP_001343453.1.

b) It contains a protein with a binding site for androstan.

c) It contains a protein with a Zinc finger domain and exactly 7 transmembrane regions. (Tip: I'm cheating A LOT here, take a look here https://www.uniprot.org/help/query-fields , here: https://web.archive.org/web/20190608232443/https://www.uniprot.org/help/query-fields, and here: https://legacy.uniprot.org/   )

d) It contains a protein with a "nuclear localization signal" of 30 amino acids annotated as a motif. **This one is worse than the previous one! (recheck here https://www.uniprot.org/help/sequence_annotation)**

e) It contains a protein with a 3D structure already solved using x-ray with a resolution better than 3A and that belongs to the superfamily 2.40.160.10 (Porin) in the CATH domain structural classification.

f) It contains a protein with a motif (pattern) with accession number PS00449 at PROSITE database.

g) It contains a protein with a motif with the consensus pattern <M-L-C-C-[LIVM]-R-R.

h) It contains a protein containing the peptide EQAVETEPEP in its primary structure.

i) It contains a protein containing three domains classified by Pfam database as family receptor L (PF01030), GF_recep_IV (PF14843) and furin-like domain (PF00757).

j) It contains a protein with the with the PFAM domain "Piwi" that is believed to interact with UniProt protein Q96C10.

k) It contains a protein with this domain organization at Pfam:

l) It contains a protein with a GPS domain and a motif with sequence logo shown here below repeated 6 times.
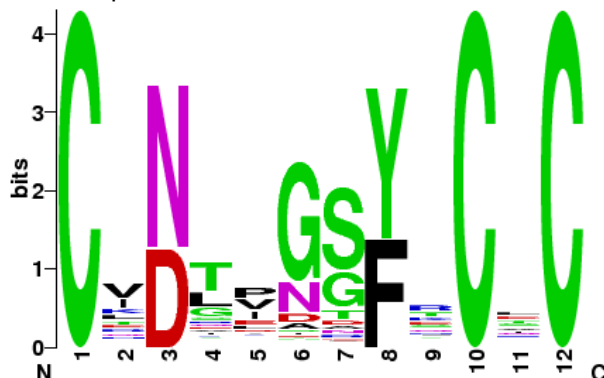
In the position specific scoring matrix (PSSM) at positions 2, 4, 5, 6, 7, 9 and 11 no amino acid has a score of zero.

Tips at http://prosite.expasy.org/prosuser.html#conv_pa