

UniProt database

1

- a. What is the AC of the UniProt entry for the human insulin? **P01308**
- b. How many isoforms for this protein are described in that entry? **2 isoforms**
- c. How many times has this entry been modified? **261 times; currently in version 262**
... and the protein sequence? **None; currently in version 1**
- d. With how many proteins does the human insulin interact? **487 interactors (BioGrid), 18 interactors (IntAct); databases do not always agree**

Feb 2022

Protein-protein interaction databases

BioGRID ⁱ	109842, 487 interactors
DIP ⁱ	DIP-6024N
IntAct ⁱ	P01308, 18 interactors
MINT ⁱ	P01308
STRING ⁱ	9606.ENSP00000380432

Feb 2021

Protein-protein Interaction databases

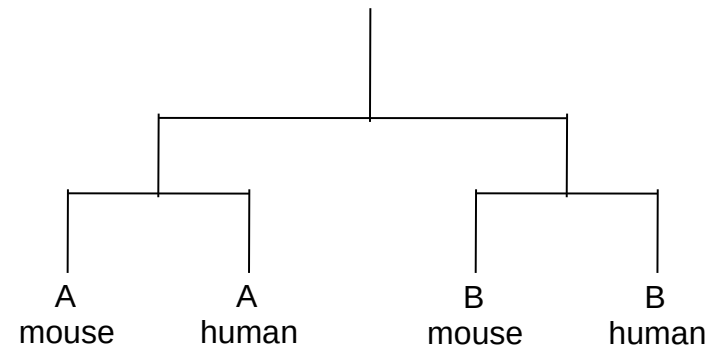
BioGRID ⁱ	109842, 24 interactors
DIP ⁱ	DIP-6024N
IntAct ⁱ	P01308, 18 interactors
MINT ⁱ	P01308
STRING ⁱ	9606.ENSP00000380432

Homology

2

Classify the following protein pairs based on their evolutionary relationship.
Note: proteins A and B have a common ancestor.

- a. Protein A mouse / Protein A human → Orthologs
- b. Protein A mouse / Protein B mouse → Paralogs
- c. Protein A mouse / Protein B human → Homologs
- d. Protein A human / Protein B mouse → Homologs
- e. Protein A human / Protein B human → Paralogs
- f. Protein B mouse / Protein B human → Orthologs



3

a. Using the human protein “P21741”, find its orthologous proteins in frog (*Xenopus laevis*) and get their UniProt AC. P48530, P48531

b. Check the identity between the orthologs (human – frog proteins).

P21741-P48530 = 61.1%, P21741-P48531 = 60.4%

c. Check the identity between the paralogs (frog – frog proteins).

P48530-P48531 = 97.9%



Human
(*Homo sapiens*)



Frog
(*Xenopus laevis*)

4

a. Based on the sequence of the “ATP synthase subunit a” protein from the extinct mammoth (*Mammuthus primigenius*) [Q38PR7], was the mammoth closer to the asian elephant (*Elephas maximus*) or to the african elephant (*Loxodonta africana*)? Use only SwissProt proteins.

M. primigenius (Q38PR7) – *E. maximus* (Q2I3G9) = 95.5%

M. primigenius (Q38PR7) – *L. africana* (Q9TA24) = 93.2%

b. Is there evidence enough to conclude if they are / are not closer? No.

c. Could you check with the “cytochrome b” protein too? [P92658] Use only SwissProt proteins.

M. primigenius (P92658) – *E. maximus* (O47885) = 96.3%

M. primigenius (P92658) – *L. africana* (P24958) = 97.9%



Woolly mammoth
(*Mammuthus primigenius*)



Asian elephant
(*Elephas maximus*)



African elephant
(*Loxodonta africana*)

5

a. Based solely on the sequence of the “Cytochrome b” protein (Q8SG72) from the extinct dodo (*Raphus cucullatus*), was the dodo closer to the Nicobar pigeon “*Caloenas nicobarica*” or to the chicken (*Gallus gallus*)? Use NCBI Blast.

R. cucullatus – *C. nicobarica* = 98.50%

R. cucullatus – *G. gallus* = 92.51%

b. There are more than 300 species of pigeons. Do the results differ if you consider the street pigeon (*Columba livia*)?

R. cucullatus – *C. livia* = 95.88%

R. cucullatus – *G. gallus* = 92.51%



Dodo
(*Raphus cucullatus*)



Nicobar pigeon
(*Caloenas nicobarica*)



Chicken (rooster)
(*Gallus gallus*)



Pigeon
(*Columba livia*)

MSA

6

Given the following alignments,

classify them in:

- Pairwise / multiple
- Local / global

calculate their:

- % similarity
- % identity

```
>Protein_A
KKKYYWKKT
>Protein_B
AKKYYW
>Protein_C
RKRWWWRT
```

a) Protein_A YYWW
Protein_B YYWW

Pairwise
Local
100% similarity
100% identity

b) Protein_A KKKYYWKKT
Protein_B AKKYYW---

Pairwise
Global
60% similarity
60% identity

c) Protein_A KKKYYWKKT
Protein_B AKKYYW---
Protein_C AKRWWWRT
*:::**

Multiple
Global
60% similarity
30% identity

7

a. The proteins “P11582” and “P02226” are paralogs, but they differ in length (152 vs 161 amino acids). Is there an extra region in P02226, or the extra amino acids are dispersed along the protein? Use UniProt.

Extra region in P02226: “IGNESN”.

b. How could that have happened in evolution?

Deletion in P11582, or insertion in P02226.

c. Protein “P02226” is a globin. If you blast P02226 and use only SwissProt proteins as database, how many results do you get? Is this protein present in any bacterial proteome?

159 results; 15 in Bacteria and 144 in Eukaryota.

// FYI: Protein entry P02226 is one of the oldest in SwissProt; check the date! **November 1, 1988.**

8

a. Both “P17861” (XBP1_HUMAN) and “Q3SZZ2” (XBP1_BOVIN) are “X-box binding protein 1” proteins. Can you detect which region/s of these proteins is/are important for their function? Why? Use Clustal Omega.

What should you do to detect them? No. They are too similar. We would need a protein from a more distant organism.

b. Add the proteins “G5EE07” (G5EE07_CAEEL) and “Q8UVQ5” (Q8UVQ5_DANRE) to the study. Are you able to identify that region/s now? Why? Use Clustal Omega.

Yes. They are not as similar.

c. Check the positional annotations in the entry of the human protein. Was the region you identified annotated as a domain?

bZIP (basic-leucine zipper) domain in positions:

70-133 (human)

70-133 (cattle)

61-117 (worm)

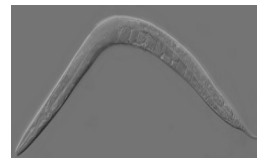
69-132 (zebrafish)



Human
(*Homo sapiens*)



Cattle
(*Bos taurus*)



Worm
(*Caenorhabditis elegans*)



Zebra fish
(*Danio rerio*)

9

a. All of the sequences in “*file1.fasta*” (<https://cbdm.uni-mainz.de/mb22/>) are homologs. How many groups of orthologs would you say there are in this file? Use SeaView.

Two groups of orthologs: Protein A & protein B.

b. What could you say about the history of this protein family?

E. coli has only one protein, and then it duplicated to form A and B. It is possible that *X.laevis_B* duplicated later to form B and C.

c. Would you say there is any wrongly annotated sequence?

X.tropicalis_B is wrongly annotated. It should be *X.tropicalis_A*, because they are in the same branch. The actual *X.tropicalis_B* is either not in the dataset or was lost during evolution.

Phylogeny

10

a. Using “*file2.fasta*” (<https://cbdm.uni-mainz.de/mb22/>), can you approximate to which taxonomic division belongs “proteinX”? *Primates*.

b. From which organism could it be? After guessing, check it.

Homo sapiens (human) or *Pan troglodytes* (chimpanzee); they are 100% identical.

Phylogeny

11

The sequences in “*file4.fasta*” (<https://cbdm.uni-mainz.de/mb22/>) are orthologs from dolphin, human, cattle, zebra fish, chicken, platypus and frog.

a. Build a phylogenetic tree using these sequences.

b. Check in UniProt > Taxonomy (<https://www.uniprot.org/taxonomy/>) if the phylogenetic tree built only from that one protein family is coherent with evolution.

Dolphin = *Delphinus delphis*

Human = *Homo sapiens*

Cattle = *Bos taurus*

Zebra fish = *Danio rerio*

Chicken = *Gallus gallus*

Platypus = *Ornithorhynchus anatinus*

Frog = *Xenopus laevis*

No, the tree fails to relate all amniotes in one branch.