

Question 1

Procedure: Visit Uniprot website and search for “human mitochondrial β barrel membrane protein VDAC” in the search box. Then open the search result. This is the link :

<https://www.uniprot.org/uniprot/P21796>

Protein sequence from UniProt website is :

sp|P21796|VDAC1_HUMAN Voltage-dependent anion-selective channel protein 1 OS=Homo sapiens OX=9606 GN=VDAC1 PE=1 SV=2
MAVPPTYADLGKSARDVFTKGYGFGLIKLDLKTSENGLEFTSSGSANTETTKVTGSLET
KYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFSPTGKKNAKIKTGYKR
EHINLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMNFFETAKSRVTQSNFAVGKYKTDEFQL
HTNVNDGTEFGGSYQKVNKKLETAVNLAWTAGNSNTRFGIAAKYQIDPDACFSAKVNNS
SLIGLGYTQTLKPGIKLTLSALLDGKNVNAGGHKLGLGLEFQA

Functions

- It is involved in the formation of a channel through the outer mitochondrial membrane as well as the plasma membrane. The channel at the outer mitochondrial membrane allows small hydrophilic molecules to diffuse; in the plasma membrane, it regulates cell volume and is involved in apoptosis. At low or zero membrane potential, it adopts an open conformation, and at potentials greater than 30-40 mV, it adopts a closed conformation. The open state is anion-selective, whereas the closed state is cation-selective.
- Binds a variety of signalling molecules, including sphingolipid ceramide, phospholipid phosphatidylcholine, and sterol cholesterol.
- In depolarized mitochondria, acts downstream of PRKN and PINK1 to promote mitophagy or prevent apoptosis; polyubiquitination by PRKN promotes mitophagy, whereas monoubiquitination by PRKN reduces mitochondrial calcium influx, which inhibits apoptosis.
- It might be involved in the formation of the permeability transition pore complex (PTPC), which is responsible for the release of mitochondrial products that cause apoptosis.

Number of Transmembrane segments present in this protein: 19

Question 2

Procedure

Go to the website provided with the question and write transcription factors in the search box.

UniProtKB - "transcription factors"

UniProtKB 2021_04 results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Filter by:

- Reviewed (3,887) Swiss-Prot
- Unreviewed (112,777) TrEMBL
- Popular organisms
 - Human (787)
 - A. thaliana (735)

1 to 25 of 116,664 Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
Q08471	MSA1_YEAST	G1-specific transcription factors a...	MSA1 YOR066W, YOR29-17	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	629
Q8VWG0	Q8VWG0_ARATH	AtbZIP transcription factor	AtbZIP52 basic leucine-zipper 52, bZIP52, At1g06850, F4H5.7, F4H5.7	Arabidopsis thaliana (Mouse-ear cress)	337
A0A024R5Z0	A0A024R5Z0_HUMAN	Transcription factor 12 (HTF4, hell...	TCF12 hCG_40686	Homo sapiens (Human)	706
Q63934	PO4F2_MOUSE	POU domain, class 4, transcription ...	Pou4f2 Brn-3.2, Brn3b	Mus musculus (Mouse)	411

We'd like to inform you that we have updated our [Privacy Notice](#) to comply with Europe's new General Data Protection Regulation (GDPR) that applies since 25 May 2018. [Do not show this banner again](#)

It gives 116,664 results in result

The select 50% in identity clusters

UniProtKB - "transcription factors"

UniProtKB 2021_04 results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Filter by:

- Reviewed (3,887) Swiss-Prot
- Unreviewed (112,777) TrEMBL
- Popular organisms
 - Human (787)
 - A. thaliana (735)

1 to 25 of 116,664 Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
P08047	SP1_HUMAN	Transcription factor Sp1	SP1 TSFP1	Homo sapiens (Human)	785
Q99967	CITE2_HUMAN	Cbp/p300-interacting transactivator...	CITED2 MRG1	Homo sapiens (Human)	270
P47811	MK14_MOUSE	Mitogen-activated protein kinase 14	Mapk14 Crk1, Csbp1, Csbp2	Mus musculus (Mouse)	360
Q92481	AP2B_HUMAN	Transcription factor AP-2-beta	TFAP2B	Homo sapiens (Human)	460
D3ZS23	NLK_RAT	Serine/threonine-protein kinase NLK	Nlk	Rattus norvegicus (Rat)	527
P17208	PO4F1_MOUSE	POU domain, class 4, transcription ...	Pou4f1 Brn-3, Brn3, Brn3a	Mus musculus (Mouse)	421
Q9UBE8	NLK_HUMAN	Serine/threonine-protein kinase NLK	NLK LAK1	Homo sapiens (Human)	527
Q15759	MK11_HUMAN	Mitogen-activated protein kinase 11	MAPK11 PRKM11, SAPK2, SAPK2B	Homo sapiens (Human)	364
O54949	NLK_MOUSE	Serine/threonine-protein kinase NLK	Nlk	Mus musculus (Mouse)	527
Q39204	MYC2_ARATH	Transcription factor MYC2	MYC2 BHLH6, EN38, JAI1, JIN1, RAP1	Arabidopsis thaliana (Mouse-ear cress)	623
Q9C9L2	TCP15_ARATH	Transcription factor TCP15	TCP15 At1g69690, T6C23.11	Arabidopsis thaliana (Mouse-ear cress)	325
E2QWQ2	NLK_CANLF	Serine/threonine-protein kinase NLK	NLK	Canis lupus familiaris (Dog) (Canis familiaris)	527

We'd like to inform you that we have updated our [Privacy Notice](#) to comply with Europe's new General Data Protection Regulation (GDPR) that applies since 25 May 2018. [Do not show this banner again](#)

It gives around 18,085 results

Question 3

There are a total of 1,926,090 sequences of “homo sapiens” in UniProt.

50% - 162,680

90% - 450,164

100% - 1,358,983

Question 4

17,090 sequences are manually annotated for “Mus Musculus”.

UniProtKB 2021_04 results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Filter by: [BLAST](#) [Align](#) [Download](#) [Add to basket](#) [Columns](#)

Expand search "Mus musculus (Mouse) [10090]" to include lower taxonomic ranks

Entry	Entry name	Protein names	Gene names	Organism	Length
Q88291	ZNF326_MOUSE	DBIRD complex subunit ZNF326	Znf326 Zan75, Zfp326, Zird	Mus musculus (Mouse)	580
Q8BG89	ZNF365_MOUSE	Protein ZNF365	Znf365 Dbz, Kiaa0844, Zfp365	Mus musculus (Mouse)	408
Q5YST1	ZDHHC20_MOUSE	Palmitoyltransferase ZDHHC20	Zdhhc20	Mus musculus (Mouse)	380

1 to 25 of 17,090 Show 25

After applying 3D database filter in pdb we get 2062 results

UniProtKB 2021_04 results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Filter by: [BLAST](#) [Align](#) [Download](#) [Add to basket](#) [Columns](#)

Expand search "Mus musculus (Mouse) [10090]" to include lower taxonomic ranks

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9Z0P5	TWF2_MOUSE	Twinfilin-2	Twf2 Ptk9l	Mus musculus (Mouse)	349
Q9Z2Y1	UBXN1_MOUSE	UBX domain-containing protein 1	Ubxn1 D19Etd721e, Saks1	Mus musculus (Mouse)	297
Q88878	ZFAN5_MOUSE	AN1-type zinc finger protein 5	Zfand5 Za20d2, Zfp216, Znf216	Mus musculus (Mouse)	213
Q70480	VAMP4_MOUSE	Vesicle-associated membrane protein...	Vamp4	Mus musculus (Mouse)	141

1 to 25 of 2,062 Show 25

Question 5

1957 of 2062 ids were matched

Results

1,957 out of 2,062 identifiers from UniProtKB AC/ID were successfully mapped to 1,957 STRING IDs.
[Click here to download unmapped identifier\(s\)](#)

[Download](#)

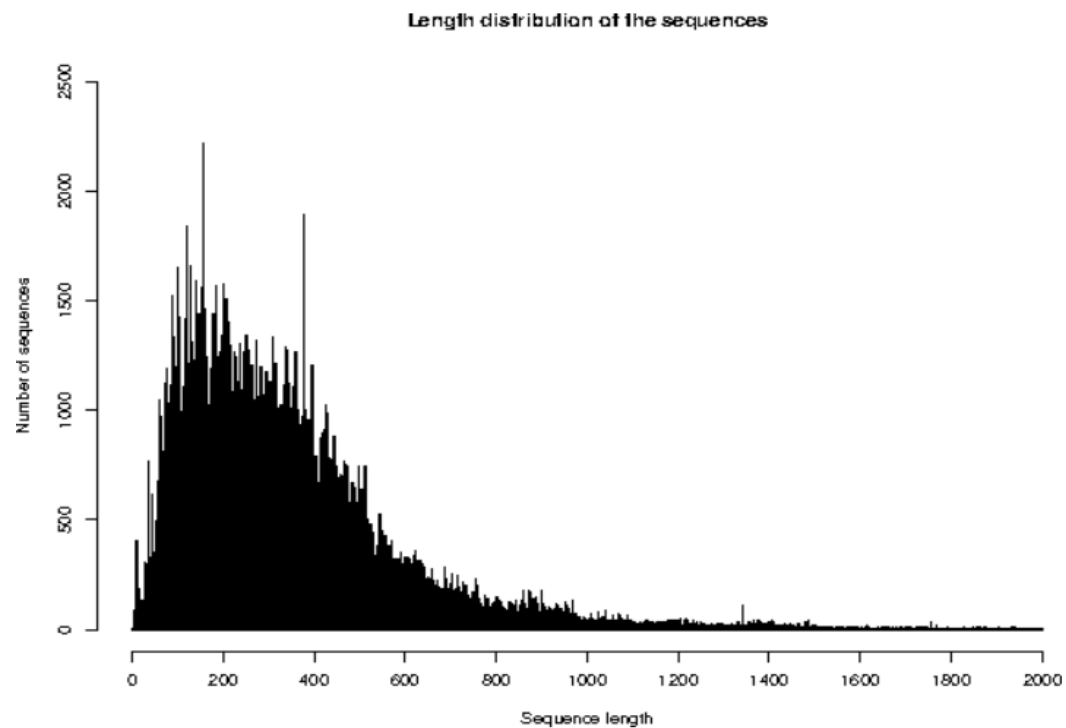
◀ 1 to 25 of 1,957 ▶

From	To
S4R2P9	10090.ENSMUSP00000138735
Q9ZZX8	10090.ENSMUSP00000131029
Q9ZZX2	10090.ENSMUSP00000033805
Q9ZZX1	10090.ENSMUSP00000130023
Q9ZZW9	10090.ENSMUSP00000075687
Q9ZZU1	10090.ENSMUSP00000088057
Q9ZZU0	10090.ENSMUSP00000029082
Q9ZZM6	10090.ENSMUSP00000078303
Q9ZZL7	10090.ENSMUSP00000060028
Q9ZZH7	10090.ENSMUSP00000037328
Q9ZZG9	10090.ENSMUSP000000082374
Q9ZZG6	10090.ENSMUSP00000021347
Q9ZZD7	10090.ENSMUSP00000032469

Question 6

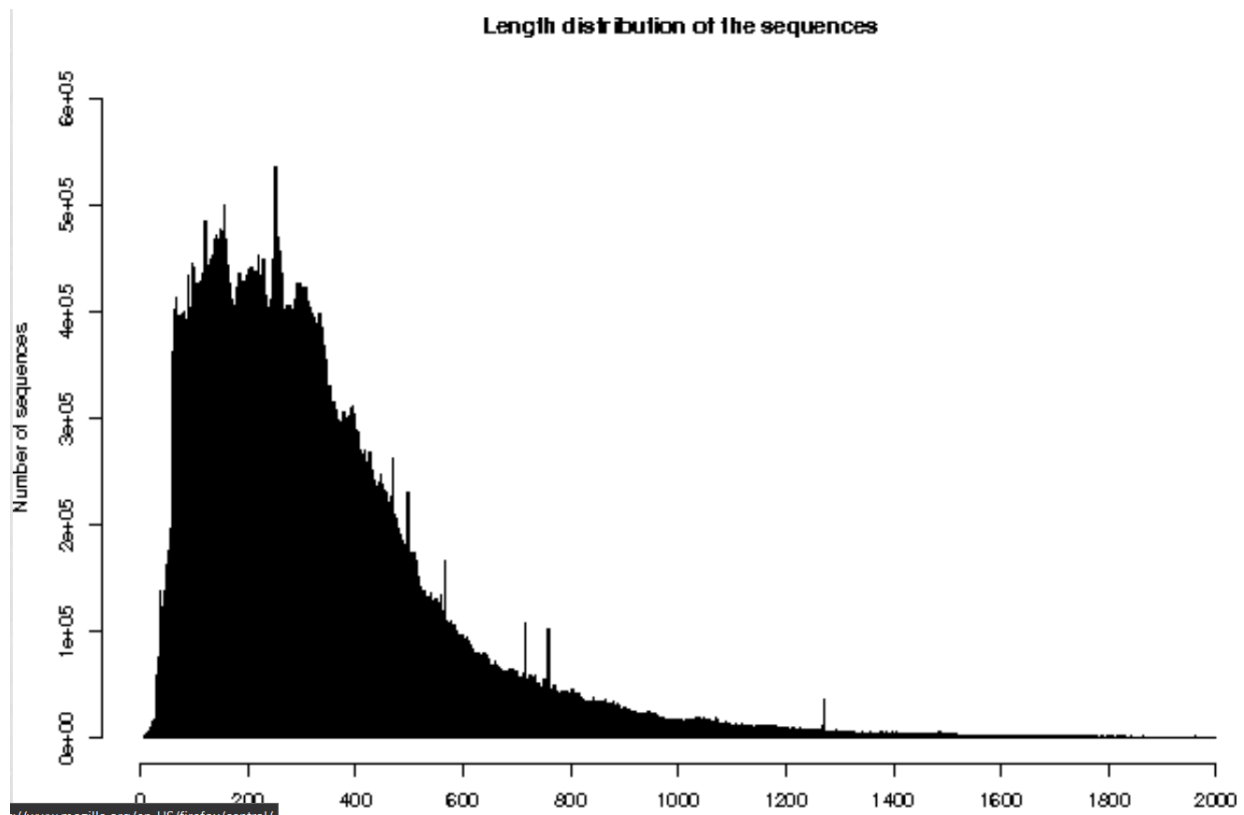
a)

Sequence Length Distribution in UniProtKB/Swiss-Prot(<https://web.expasy.org/docs/relnotes/relnstat.html>):



Sequence Length Distribution in

UniProtKB/TrEMBLstats(<https://www.ebi.ac.uk/uniprot/TrEMBLstats>):



In TrEMBL:

The shortest sequence in UniProtKB = 7 AAs, with sequence ID = A0A1Y7VI41

The longest sequence in UniProtKB = 45,354 AAs with sequence ID = A0A5A9P0L4

In Swiss Plot :

The shortest sequence in UniProtKB = 2 AAs, which sequence ID = P0DPR3

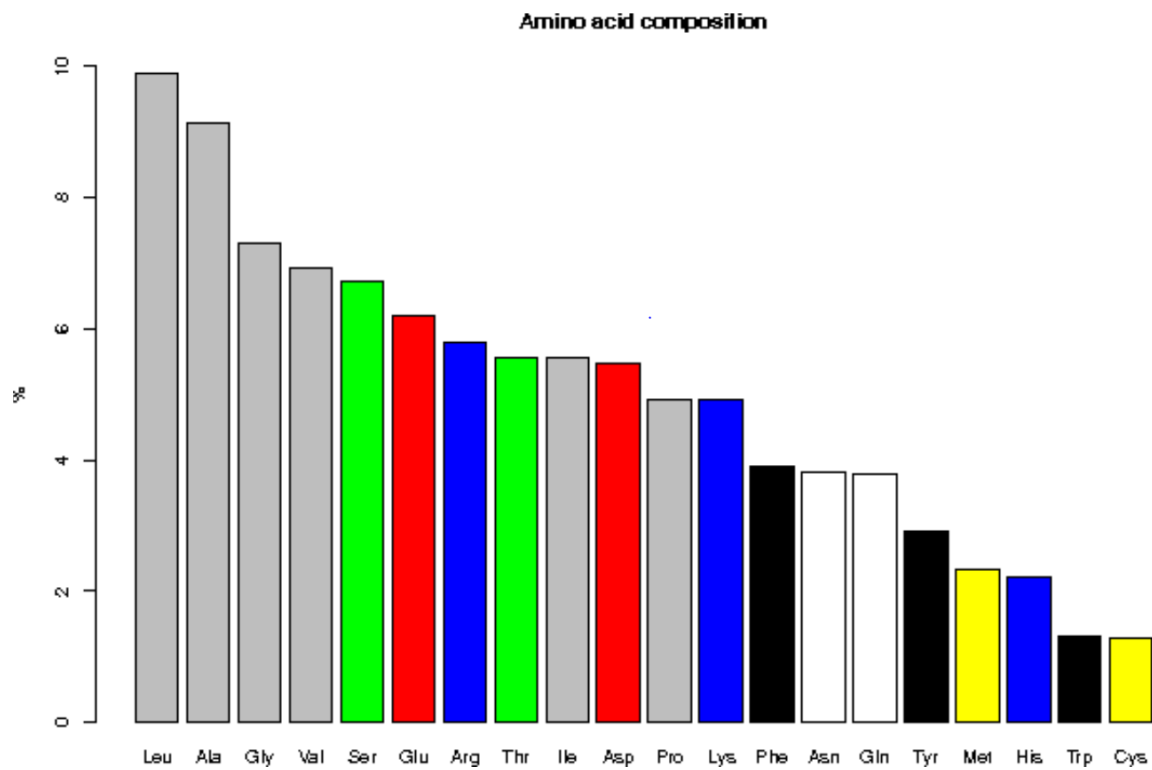
The longest sequence in UniProtKB = 35,213 AAs with sequence ID = A2ASS6

c) As per TrEMBL:

5. AMINO ACID COMPOSITION

5.1 Composition in percent for the complete database

Ala (A)	9.13	Gln (Q)	3.77	Leu (L)	9.88	Ser (S)	6.71
Arg (R)	5.80	Glu (E)	6.18	Lys (K)	4.91	Thr (T)	5.57
Asn (N)	3.80	Gly (G)	7.30	Met (M)	2.34	Trp (W)	1.30
Asp (D)	5.47	His (H)	2.20	Phe (F)	3.90	Tyr (Y)	2.90
Cys (C)	1.27	Ile (I)	5.56	Pro (P)	4.92	Val (V)	6.93
Asx (B)	0	Glx (Z)	0	Xaa (X)	0.08		



As per swiss plot

