

Practical 3

Kudari Pavani

BE19B023

1)

Procedure:

Uniprot website and search for “human mitochondrial β barrel membrane protein VDAC” in search box.

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

Amino Acid Sequence:

sp|P21796|VDAC1_HUMAN Voltage-dependent anion-selective channel protein 1
OS=Homo sapiens OX=9606 GN=VDAC1 PE=1 SV=2
MAVPPTYADLGKSARDVFTKGYGFLIKLDLKTSENGLEFTSSGSANTETTKVTGSLET
KYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFSPNTGKKNAKIKTGYKR
EHINLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMNFETAKSRVTQSNFAVGYKTDEFQL
HTNVNDGTEFGGSIYQKVNKKLETAVNLAWTAGNSNTRFGIAAKYQIDPDACFSAKVNNNS
SLIGLGYTQTLKPGIKLTLALLDGKNVNAGGHKLGLGLEFQA.

Function:

It forms a channel through outer membrane of mitochondria along with the plasma membrane whereby the outer mitochondrial membrane channel permits the diffusion of small hydrophilic molecules; in the plasma membrane which is involved in the apoptosis, cell and volume regulation. It accepts an open conformation at very low or zero membrane potential or adopts an closed conformation when potential is above 30-40mV. The closed has a cation selectivity whereas the open state is weak anion-selective. It may also participate in the production of permeability transition pore complex (PTPC) and it could be the triggering agent for release of mitochondrial products that initiates apoptosis.

Number of Transmembrane segments present in this protein= 19.

2) Method:

- Search for “transcription factors” in UniProt.
<https://www.uniprot.org/uniprot/?query=%22transcription+facto%22&sort=score>
- Select cluster identities for specific values.

The screenshot displays the UniProt website interface. At the top, the search bar contains the query "cluster:(uniprot:(transcription factors) identity:0.5)". Below the search bar, the results are titled "UniProtKB 2021_04 results". A summary box indicates that UniProtKB consists of two sections: "Reviewed (Swiss-Prot) - Manually annotated" and "Unreviewed (TrEMBL) - Computationally analyzed". Below this, a table of results is shown, filtered by "Popular organisms". The table has columns for Entry, Entry name, Protein names, Gene names, Organism, and Length. The results are sorted by "Score" in descending order.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q48LC2	Y1551_PSE14	UPF0260 protein PSPPH_1551	PSPPH_1551	Pseudomonas savastanoi pv. phaseolicola (strain 1448A / Race 6) (Pseudomonas syringae pv. phaseolicola (strain 1448A / Race 6))	149
B5Z9W0	Y173_HELPG	UPF0114 protein HPG27_173	HPG27_173	Helicobacter pylori (strain G27)	177
A6VEA3	XPT_PSEA7	Xanthine phosphoribosyltransferase	xpt PSA7_6072	Pseudomonas aeruginosa (strain PA7)	190
P69722	VIF_HV112	Virion infectivity factor	vif	Human immunodeficiency virus type 1 group M subtype B (isolate PCV12) (HIV-1)	192
Q3KGH1	Y1392_PSEPF	UPF0260 protein PFI01_1392	PFI01_1392	Pseudomonas fluorescens (strain Pf0-1)	149

Total number of sequences= 7,202,857,
50% Identity Clusters= **31,057**clusters

UniRef 2021_04 results

The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records. This hides redundant sequences and obtains complete coverage of the sequence space at three resolutions:

- UniRef100** combines identical sequences and sub-fragments with 11 or more residues from any organism into a single UniRef entry.
- UniRef90** is built by clustering UniRef100 sequences such that each cluster is composed of sequences that have at least 90% sequence identity to, and 80% overlap with, the longest sequence (a.k.a. seed sequence).
- UniRef50** is built by clustering UniRef90 seed sequences that have at least 50% sequence identity to, and 80% overlap with, the longest sequence in the cluster.

Help | UniRef help video | Other tutorials and videos | Downloads

Filter by: 50% (31,057) | Map to: UniProtKB, UniParc | Demo | Help video

Cluster ID	Cluster name	Size	Cluster members	Organisms	Length	Identity
UniRef50_A0A010QZT8	Cluster: Uncharacterized protein	53	A0A010QZT8 A0A640QLN8 A0A379F9E8 A0A844P9I2 A0A110DXT3 A0A346YVP8 A0A7W2RXL3 A0A0Q8BV37 A0A4Z0E6V6 +43	Alteromonas sp. ALT199 Alteromonas sp. KUL150 Proteus vulgaris Alivibrio fischeri (Vibrio fischeri) Moritella viscosa Neorhizobium sp. SOG26 Colwellia sp. MB3u-64 Rhizobium sp. Root564 Methyloburum sp. Q1 Bordetella bronchiseptica (Alcaligenes bronchiseptica) And more	79	50%
UniRef50_A0A010REZ0	Cluster: Uncharacterized protein	963	A0A010REZ0 A0A423GRQ8 A0A0N9VUZ5 A0A2N8C416	Pseudomonas fluorescens HK44 Pseudomonas brassicaearum Pseudomonas fluorescens Pseudomonas sp. FW306-2-11AA	105	50%

3)

There are a total of sequences of “homo sapiens” in UniProt.

Sequence Identity	Number of clusters
50%	9258839
90%	2216520
100%	406431

4)

UniProtKB 2021_04 results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evoked 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.

Help | UniProtKB help video | Other tutorials and videos | Downloads

Filter by: Reviewed (17,090) | Popular organisms: Mouse (17,309) | Proteomes: G9900008589 (17,805) | View by: ...

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

Entry	Entry name	Protein name	Gene name	Organism	Length
Q8BQ91	ZNF126_MOUSE	DNB1D complex subunit ZNF126	Znf126 Znf126, Zfp126, Znf1	Mus musculus (Mouse)	388
Q8BQ98	ZNF365_MOUSE	Protein ZNF365	Znf365 [bc, Klaf8944, Zfp365]	Mus musculus (Mouse)	408
Q5Y1T1	ZNF126_MOUSE	Palindromic transferase ZNF126	Znf126	Mus musculus (Mouse)	368

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](#)

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 curation.

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 (2,862)
- Popular organisms
- Proteomes
- View by

Entry	Entry name	Protein names	Gene names	Organism	Length
Q07BP9	TNFR2_MOUSE	Tnfrsf1b-2	Tnfrsf1b	Mus musculus (Mouse)	348
Q03ZY1	UBOX1_MOUSE	UBOX domain-containing protein 1	Ubox1	Mus musculus (Mouse)	297
Q68518	ZFAND5_MOUSE	ARE-type zinc finger protein 5	Zfand5	Mus musculus (Mouse)	213
Q17460	VAMP4_MOUSE	Vesicle-associated membrane protein 4	Vamp4	Mus musculus (Mouse)	141
Q17460	ZFAND5_MOUSE	ARE-type zinc finger protein 5	Zfand5	Mus musculus (Mouse)	213

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.

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

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

5)

Procedure:

1. Search in UniProt with search query –
2. database:(type:pdb) AND reviewed:yes AND organism:"Mus musculus (Mouse) [10090]"
3. Select only the entry column and download the identifiers as a list.
4. In Retrieve/ID Mapping, paste these identifiers.
5. Under select options:
 - a. From – UniProtKB
 - b. To – STRING.... And submit.

1957 of 2062 ids were matched

Results

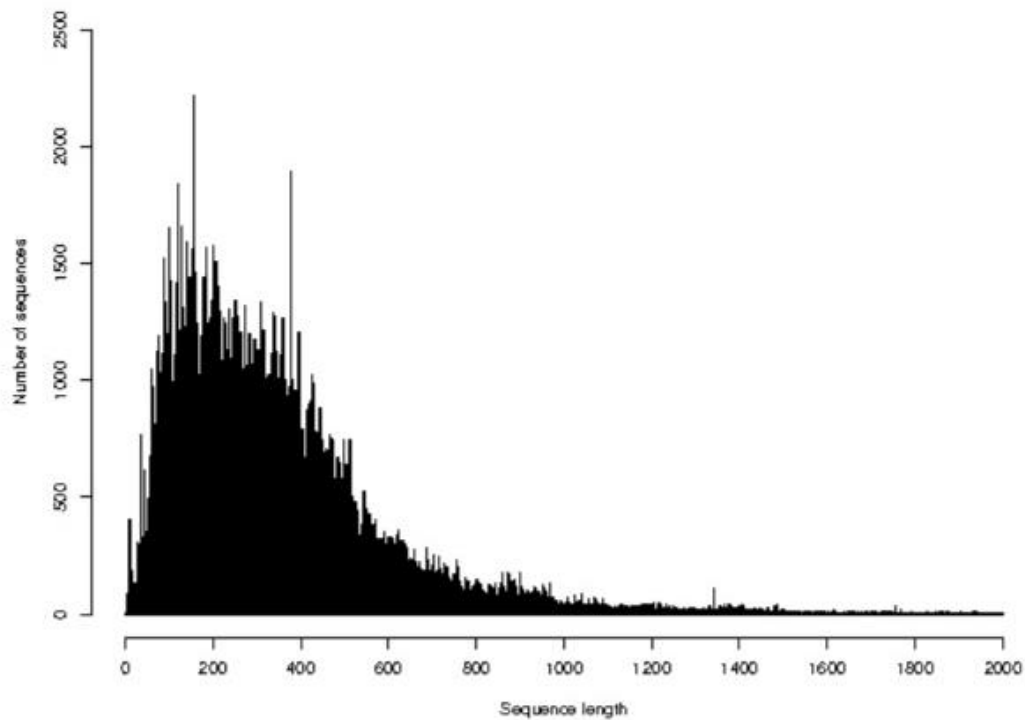
1,957 out of 2,062 identifiers from UniProtKB AC/ID were successfully mapped to 1,957 STRING IDs.
Click here to download unmapped identifier(s)

Download

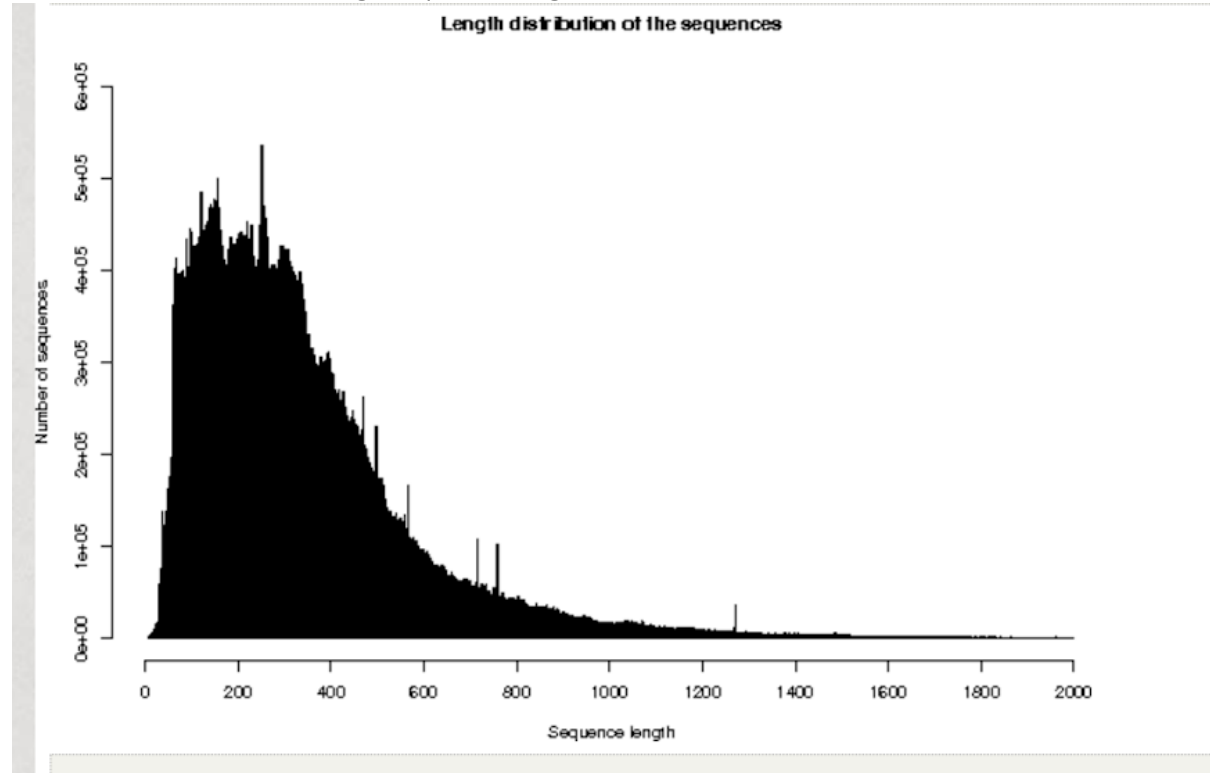
From	To
S4R2P9	10090.ENSMUSP000000138735
Q922X8	10090.ENSMUSP000000131029
Q922X2	10090.ENSMUSP000000033805
Q922X1	10090.ENSMUSP000000130023
Q922W9	10090.ENSMUSP000000075687
Q922U1	10090.ENSMUSP000000088057
Q922U0	10090.ENSMUSP000000029082
Q922M6	10090.ENSMUSP000000078303
Q922L7	10090.ENSMUSP000000060028
Q922H7	10090.ENSMUSP000000037328
Q922G9	10090.ENSMUSP000000082374
Q922G6	10090.ENSMUSP000000021347
Q922D7	10090.ENSMUSP000000032469
Q922D6	10090.ENSMUSP000000033770
Q922B5	10090.ENSMUSP000000040993
Q921XA	10090.ENSMUSP000000065720

6)

A. a) In **UniProtKB/Swiss-Prot**: average sequence length: 360 amino acids.



In **UniProtKB/TrEMBL**: average sequence length: 348 amino acids.



b)

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

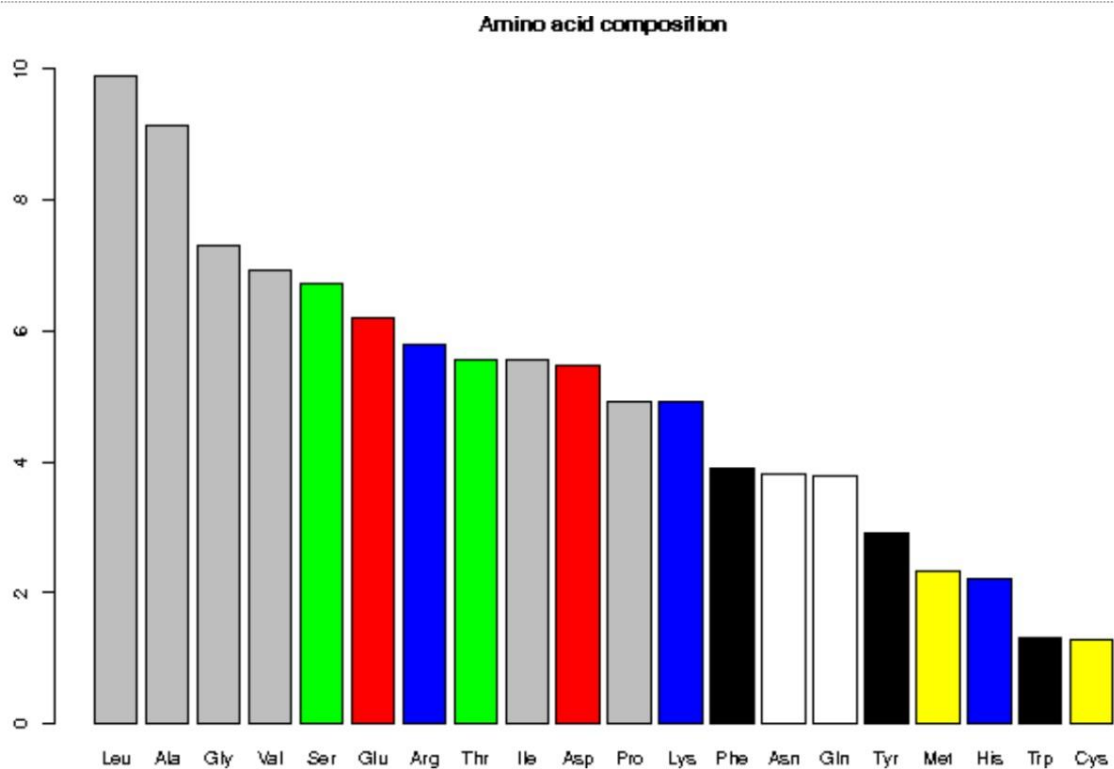
The longest sequence in UniProtKB = 35,213 AAs with sequence ID=A2ASS6

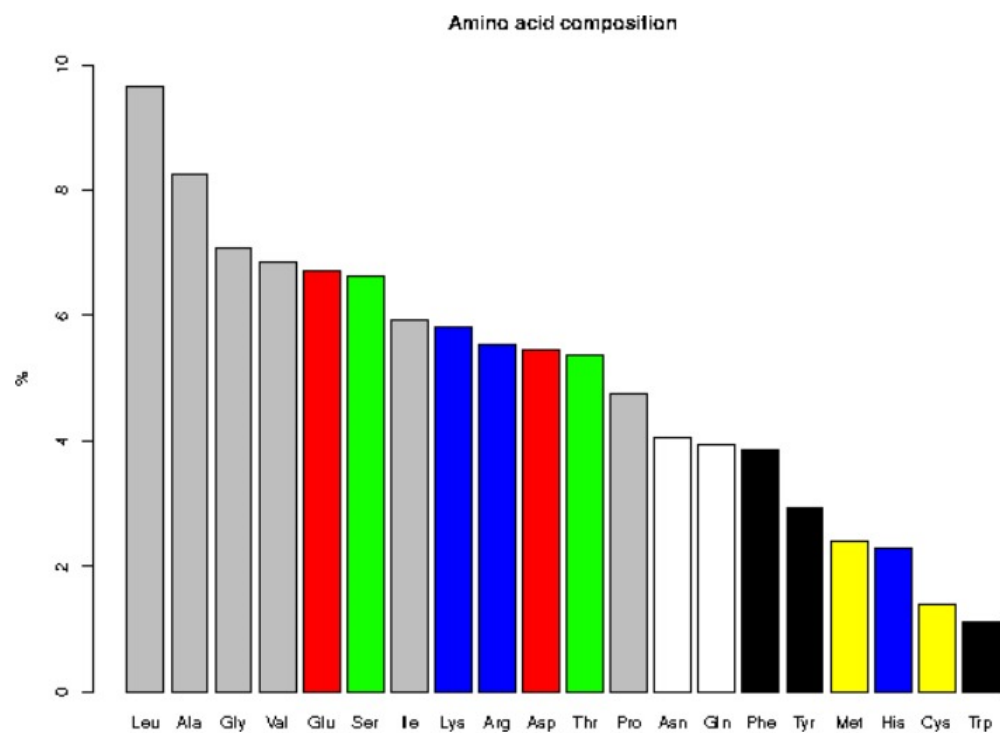
c)

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		





Legend: gray = aliphatic, red = acidic, green = small hydroxy,
 blue = basic, black = aromatic, white = amide, yellow = sulfur

6. AMINO ACID COMPOSITION

6.1 Composition in percent for the complete database

Ala (A)	8.25	Gln (Q)	3.93	Leu (L)	9.65	Ser (S)	6.64
Arg (R)	5.53	Glu (E)	6.72	Lys (K)	5.80	Thr (T)	5.35
Asn (N)	4.06	Gly (G)	7.07	Met (M)	2.41	Trp (W)	1.10
Asp (D)	5.46	His (H)	2.27	Phe (F)	3.86	Tyr (Y)	2.92
Cys (C)	1.38	Ile (I)	5.91	Pro (P)	4.74	Val (V)	6.86
Asx (B)	0.000	Glx (Z)	0.000	Xaa (X)	0.00		