

Assignment 3

1. When we search for human mitochondrial β barrel membrane, we get three VDAC proteins, the amino acid sequence and its function is given below:

- i. **VDAC1:**

- a. **Amino acid sequence:**

MAVPPTYADLGKSARDVFTKGYGFGLIKLDLKTSENGLEFTSSGSANTETTKVTGS
LETKYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFSFNTGKKNAKI
KTGYKREHINLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMFETAKSRVTQSNFAV
GYKTDEFQLHTNVNDGTEFGGSIYQKVNKKLETAVNLAWTAGNSNTRFGIAAKYQID
PDACFSAKVNNSSLIGLGYTQTLKPGIKLTLALLDGKNVNAGGHKLGLGLEFQA

- b. **Function:**

Forms a channel through the mitochondrial outer membrane and also the plasma membrane. The channel at the outer mitochondrial membrane allows diffusion of small hydrophilic molecules; in the plasma membrane it is involved in cell volume regulation and apoptosis. It adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV. The open state has a weak anion selectivity whereas the closed state is cation-selective (PubMed:11845315, PubMed:18755977, PubMed:20230784, PubMed:8420959).

Binds various signaling molecules, including the sphingolipid ceramide, the phospholipid phosphatidylcholine, and the sterol cholesterol (PubMed:31015432).

In depolarized mitochondria, acts downstream of PRKN and PINK1 to promote mitophagy or prevent apoptosis; polyubiquitination by PRKN promotes mitophagy, while monoubiquitination by PRKN decreases mitochondrial calcium influx which ultimately inhibits apoptosis (PubMed:32047033).

May participate in the formation of the permeability transition pore complex (PTPC) responsible for the release of mitochondrial products that triggers apoptosis (PubMed:15033708, PubMed:25296756).

May mediate ATP export from cells (PubMed:30061676).

- c. **Transmembrane segments:**

19

- ii. **VDAC2:**

- a. **Amino acid sequence:**

MATHGQTCARPMCIPPSYADLGKAARDIFNKGFGFGLVKLDVKTKSCSGVEFST
SGSSNTDTGKVTGTLETYKWCEYGLTFTEKWNTDNTLGTEIAIEDQICQGLKL

TFDITTFSPNTGKKSGKIKSSYKRECINLGCDVDFDFAGPAIHGSAVFGYEGWLA
GYQMTFDSAKSKLTRNNFAVGYRTGDFQLHTNVNDGTEFGGSIYQKVCEDLDTS
VNLAWTSGTNCTRFGIAAKYQLDPTASISAKVNNSSLIGVGYTQTLRPGVKLTL
SALVDGKSINAGGHKVGLALELEA

b. Function:

Forms a channel through the mitochondrial outer membrane that allows diffusion of small hydrophilic molecules (By similarity).

The channel adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV (By similarity).

The open state has a weak anion selectivity whereas the closed state is cation-selective (By similarity).

Binds various lipids, including the sphingolipid ceramide, the phospholipid phosphatidylcholine, and the sterol cholesterol (PubMed:31015432).

Binding of ceramide promotes the mitochondrial outer membrane permeabilization (MOMP) apoptotic pathway (PubMed:31015432).

iii. VDAC3:

a. Amino acid sequence:

MCNTPTYCDLGKAAKDVFENKGYGFGMVKIDLKTKSCSGVEFSTSGHAYTDTGKA
SGNLETKYKVCNYGLTFTQKWNTDNTLGTEISWENKLAEGLKLTLDTIFVPNTG
KKSGKLKASYKRDCFSVGSNVDIDFSGPTIYGWAVLAFEGWLAGYQMSFDTAKS
KLSQNNFALGYKAADFQLHTHVNDGTEFGGSIYQKVNEKIETSINLAWTAGSNN
TRFGIAAKYMLDCRTSLSAKVNNASLIGLGYTQTLRPGVKLTLALIDGKNFSA
GGHKVGLGFELEA

b. Function:

Forms a channel through the mitochondrial outer membrane that allows diffusion of small hydrophilic molecules.

2. The obtained sequences of "transcription factors" with 50% sequence identity in FASTA format:

[illegible]

The count of sequences is 11,019,551 and the count of clusters is 31,057

3. When searched in UniProt, the total number of Homo Sapiens protein sequences found are which 406,431 have 100%, 2,216,520 have 90%, 9,258,839 have 50% sequence identity.
4. In UniProt, 17,090 mouse (Mus Musculus) protein sequences are manually annotated. In this manually annotated sequences, 2,062 protein sequences are associated with PDB (3D structures).
5. When mapped the UniProt IDs with 3D structures to STRING database, we get 1,957 STRING IDs mapped.
6. Using UniProt Statistics data, we can obtain the following points:
 - a. From the distribution of sequence length. We infer that it indicates the number of amino acids in the canonical sequence displayed by default in the entry's Sequence section.
 - b. There are two protein sequences with the shortest sequence length of 2 units. They are:
 - i. Protein: T cell receptor delta diversity 1
Gene: TRDD1
Organism: Homo sapiens
More data: [TRDD1 - T cell receptor delta diversity 1 - Homo sapiens \(Human\) - TRDD1 gene & protein \(uniprot.org\)](#)
 - ii. Protein: Neuropeptide GWa
Gene: N/A
Organism: Sepia Officinalis
More data: [Neuropeptide GWa - Sepia officinalis \(Common cuttlefish\) \(uniprot.org\)](#)

And one protein sequence with the longest length of 45,354 units. Details of that protein are provided below:

- i. Protein: Peptidylprolyl isomerase
Gene: E1301 Tti010109

Organism: *Triplophysa tibetana*

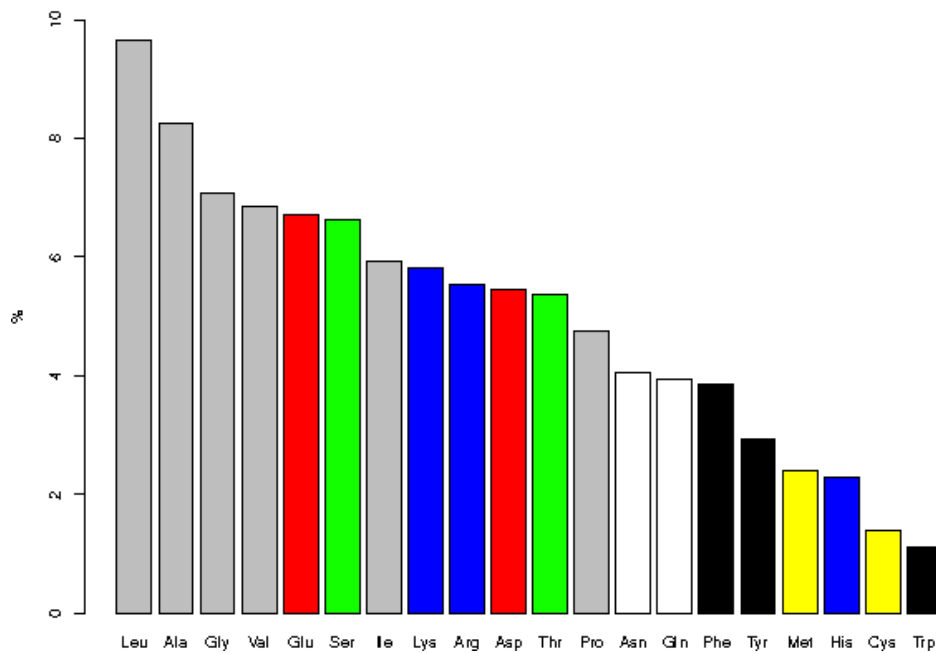
More data: [E1301 Tti010109 - Peptidylprolyl isomerase - Triplophysa tibetana - E1301 Tti010109 gene & protein \(uniprot.org\)](#)

c. Amino acid composition for the complete database in percentage:

For Swiss-Prot:

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

Amino acid composition



For TrEMBL:

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

