

Practical 3

Pradhyumn Jain, BS19B021

Question 1

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

MAVPPTYADLGKSARDVFTKGYGFGLIKLDLKTSENGLEFTSSGSANTETTKVTGSLET
KYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFSPTGKKNAKIKTGYKR
EHINLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMNFEAKSRVTQSNFAVGKYTDEFQL
HTNVNDGTEFGGSYQKVNKKLETAVNLAWTAGNSNTRFGIAAKYQIDPDACFSAKVNN
SLIGLGYTQTLKPGIKLTLALLDGKNVNAGGHKLGLGLEFQA.

Functions

- 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. 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
- Binds various signaling molecules, including the sphingolipid ceramide, the phospholipid phosphatidylcholine, and the sterol cholesterol.
- 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
- May participate in the formation of the permeability transition pore complex (PTPC) responsible for the release of mitochondrial products that triggers apoptosis.
- May mediate ATP export from cells

Number of transmembrane segments : 19

Question 2

Number of sequences (50% identity) = 16,050

Number of clusters (50% identity) = 244

First 5 sequences:

>UniRef50_Q9LD44 NAC transcription factor 56 n=455 Tax=Magnoliopsida TaxID=3398 RepID=NAC56_ARATH
MESDSSGGPPPPQPNLPPGFRFHPTDEELVVHYLKRKAASAPLPVAIIAEVDLYKFDPPW
ELPAKASFGQEWFYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPVLASDGNQKVGKK
ALVYVSGKPKGKSDWIMHEYRLIENKPNRPPGCDFGNKKNSLRLLDDWVLCRIYKKNN
ASRHVDNDKDHDMIDYIFRKIPPSLSMAAASGLHQHHNVSRSMNFFPGKFSGGGYGIF
SDGGNTSIYDGGGMINNIGTDSVDHDNADVVGLNHASSSGPMMANLKRTPVPYWPVA
DEEQDASPSKRPHGVGGGGGDCSNMSSSMEETPPLMQQGGVLGDGLFRTTYSQLPGLN
WYSS
>UniRef50_G3MP37 BTP domain-containing protein n=25 Tax=Ixodoidea TaxID=297308 RepID=G3MP37_AMBMU
MASVQHKRWGELPPLITETESGIAAIEREQITKPRPVPIQGAALYQPGQHNDNDNCERQS
PDYGTIDPITAHTIALIQHAKKVQSCITNLQQQDRDIKPLEAEYLPPPIPEGPCIPGLTTA
SPSNNSYNEIKSERKSSWKPPQLGAHATRVVLRRAVTAMCAHAGFDTTWNSVLEILTIV
CSDFYRRLCWQLRTLVDREALTGQTGFVDPLDHALHEIGMEGLSSLVQFFDYRVIGYNK
HMMKTCLKLHQYHVQVRQQRSPDDVKPVKVEEPIHFPTLDDGDESMADVEETSM
DELQRAIQSLEGMLRDDASRWISVQQQQQLPLLIGHRKGSTASSVGHLDADDEEIVIVS
DSPPLSAEVIGSLVDTVESGGLDASKLGRPPPLKRRRSEAAATYDFGV
>UniRef50_V4TQ85 NAC domain-containing protein n=257 Tax=Mesangiospermae TaxID=1437183 RepID=V4TQ85_CITCL
MENVSQVKGEDDQMDLPPGFRFHPTDEELITHYLYKKVLDVCFSCRAIGDVLNKNPEWE
LPWKAKMGEKEWYFFCMRDRKYPTGLRTNRATVSGYWKATGDKDEIYRGKSLVGMKKTLV
FYRGRAPKGEKSSWVMHEYRLDGKFSVHSLPKTAKVCAFSAVFSPVFFPLFLSTLFLTL
FLCFLVMQNEWLCRVFQKSGGKKTHISGLAGLGSFGNELGPPGLPLMDSSPDNGSKT
IKSVADSAVYSCFSNSIDLQRNQKTTTTIENFFNNPPPISVSSNCPDVFPRIPLSSNSFY
SPLSVVPVPSHAQFPGSVFMQDHSILRALIENQGSNMSQSFKTEREMISVSQDTGLTADMN
PEISSVSNLEMVVRPFNDHDAPSTSAGPVDVDFCFWNY
>UniRef50_A0A0E0RV17 Zn(2)-C6 fungal-type domain-containing protein n=61 Tax=Fusarium TaxID=5506 RepID=A0A0E0RV17_GIBZE
MAAKEGENEDNEVPVIRQKAHKKSRGLGCKNCKLRSVKCDSEKPSCKRCKSSGFVCSPTET
SPSSFFQLAHFNAGPVFVSVIGRSLGPIINPLRVVVKPVKEGMEIIVLDDFALKALERFRK
RTVFSVGTKKSKRVYSEGAFMLGLKHPFLMHVFIALAYLHDEHLNPNISASHRTPLAFHW
YQATALFHRRLATASSVQDPSTLPSSERDSLWTAGALLGAASFALLDVQSVDNVWPLKES
DPLDLWLKMSDGKKVWNLADPTRQESIFHELLAEKTSMPDGTKIIPPDVLPISIFYSVF
NLDASSASTNPYHTAASLLAQLLPHETDNTVIOFLTQLDPRFRKLLLEGKDPGRAM
LLAWWYTKAAVHSSWMMQRRSLVEGQAICIYLERYCTDVEGICELVKFPKRVFIEICRSNG
GTASEERGASVLGG
>UniRef50_Q08471 G1-specific transcription factors activator MSA1 n=25 Tax=Saccharomyces TaxID=4930 RepID=MSA1_YEAST
MDKSMIKKRGRPPITKDYPNPLQSPMAHSSMQVQKQGPHSFAKPLMKVGQSSPSPNKRRL
SIDHHNLAATTRKGRYRGVLLSTPTKSSSTNGSTPISTPSSNDSYNNTVFSETRKTFLQ
SSPIMTSSPAFQKKNYMFPSQEQLSLTITESTGKAVIAGSLFPSPSSKSSHLNNNN
KKIMQNEKIHKGSKKNAPKFEKRRILSLLKQMKNEKYCDTDTLPEAPPKPSRSDIIDTE
LPTIETASPIGSARNNNILLSQPPQSPSSAQLKPPSTPKSSLQFRMGFTPNVALNSV
SLSDTISKSTNAV GASNNNNQNGNSISNIADANTLLTLTNSPGVFLSPRNKMLPKSTAS
NEQQQEFVFFKSSGDPDLLTDDADGNWPEMLFNVSNTPRRQKCFNTPPSWINFGSPGLFS
PPRSSNMVNGTTVATASDSGNVHRQLQAQLEAQVQVQSQSNSTQROQQQRQFIPPP
INMNSPPQINIASPPHQMSRVSSIYFNKEKTTGVANMLGNTKSENLOPPANLFTA
AHGPSTPRNQEFQLPTLIECTPLIQQTMMNSLGTKYIPGTSISNSATPNLHGFPVGTGKAPS
SFDDSLKQNPYSNKKQDDARTALKRLIDDDQ

Question 3

| Cutoff % | Count |
|----------|---------|
| 100 | 152,368 |
| 90 | 91,947 |
| 50 | 68,589 |

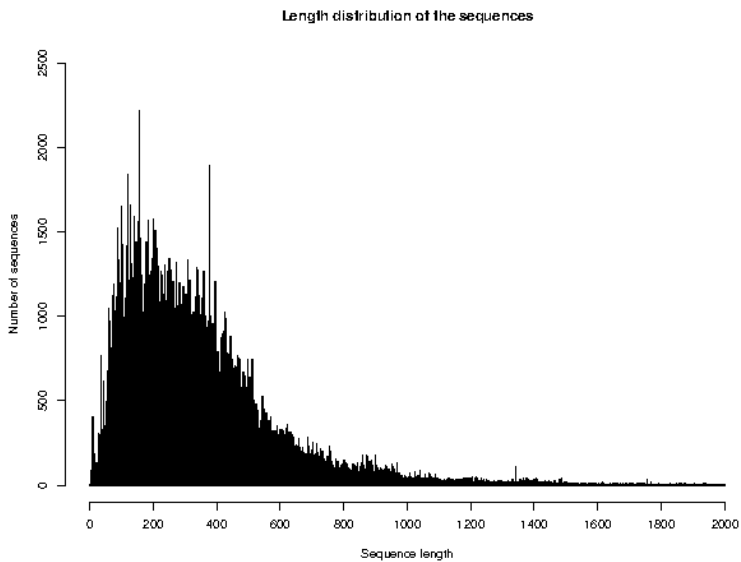
Question 4

Number of manually annotated sequences : 17,099
Number of Sequences associated with PDB : 2,062

Question 5

1,957 out of 2,062 identifiers from UniProtKB AC/ID were successfully mapped to 1,957 STRING IDs.

Question 6



a)

- The average sequence length in UniProtKB/Swiss-Prot is 360 amino acids.

b)

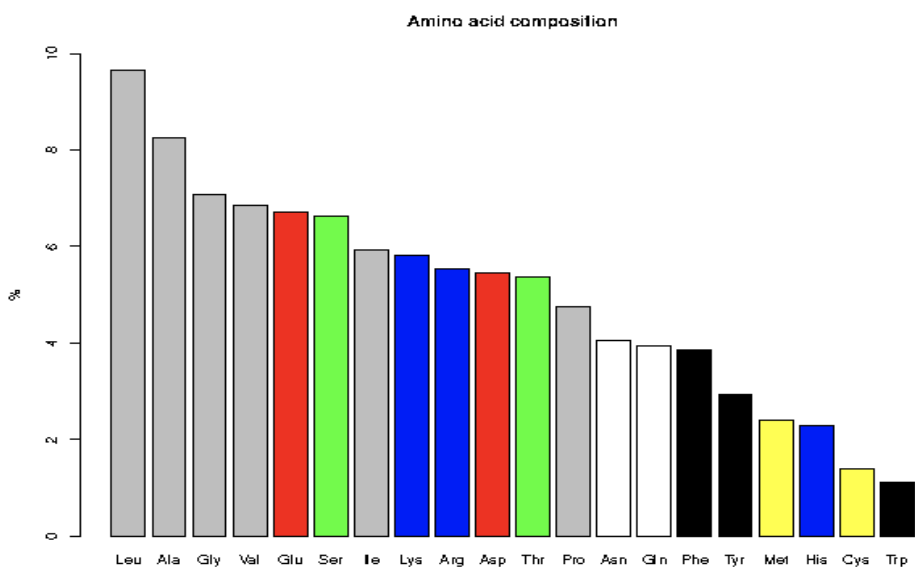
- The shortest sequence is GWA_SEPOF (P83570): 2 amino acids.
- The longest sequence is TITIN_MOUSE (A2ASS6): 35213 amino acids.

c)

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 |
|---------|-------|---------|-------|---------|------|



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