Rešitev naloge

Kratki odgovori na vprašanja:

Z uporabo znanih podatkov s pomočjo Viterbijevega algoritma določi najverjetnejše skrito zaporedje.

GLLPP

Kateri je to protein?

prestin

Katera STOP kodona sta prisotna v mRNA zaporedju za izoforma X1 in X2 tega proteina?

tag in tga

Kakšna je funkcija tega proteina v celici?

pretvarja transmembranski električni potencial v mehanske premike, kar je pomembno pri sluhu

Kateri je akronim najbolj pogoste bolezni povezane s tem proteinom pri človeku?

DFNB61

Zakaj je ta protein precej pomemben pri kitu?

zaradi eholokacije

Koliko transmembranskih vijačnic vsebuje?

14 pri človeku, 9 pri belem kitu

Kaj pomeni anotacija Compositional bias v oddelku Family and Domains?

pomeni specializiran ali intrinzično neurejen del proteina, v katerem so določene aminokisline prekomerno zastopane

Zapiši enočrkovno kodo vezavnega mesta in njegovo vrstno številko v zaporedju belega kita.

S 398

V kakšnem ekspresijskem sistemu in s kakšno metodo je bila določena struktura?

homo sapiens, z elektronsko mikroskopijo

Poišči v izhodiščnem članku iz strukturne biologije kako aspirin vpliva na ta protein.

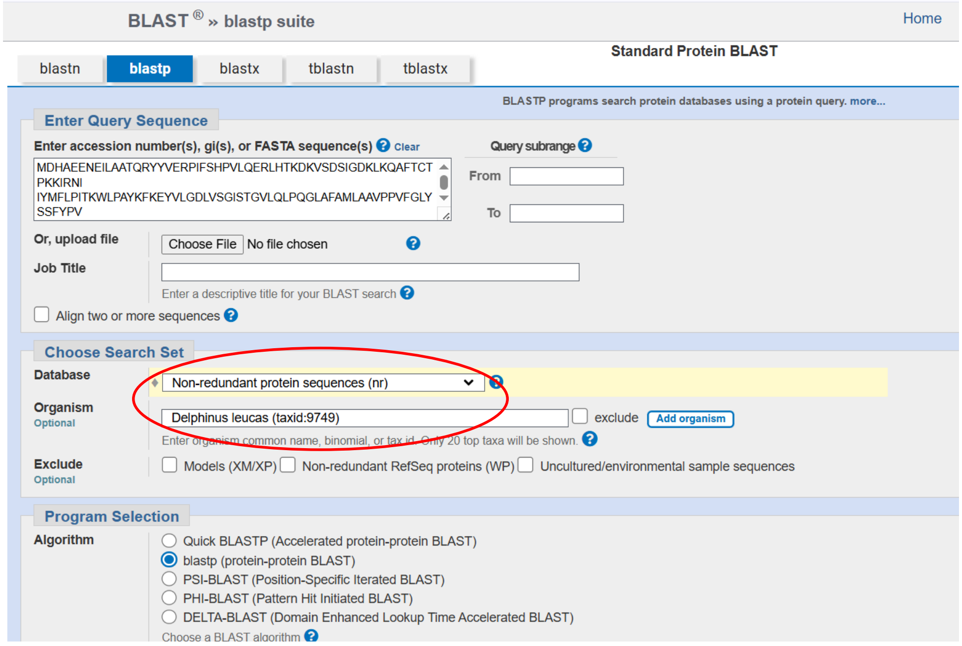
inhibira motorno aktivnost, zaradi česar po zaužitju velikih količin lahko pride do izgube sluha

Določitev skritega zaporedja:

- najprej določimo skrito zaporedje s pomočjo Viterbijevega algoritma, ki smo ga obravnavali na predavanjih
- dodatno je priložena Excel datoteka s podrobnostmi tega izračuna
- nepoznani del zaporedja pa ustreza aminokislinskimi ostanki GLLPP

Iskanje homologa:

- uporabimo blastp s spletne strani https://blast.ncbi.nlm.nih.gov/
- upišemo naše zaporedje in nastavimo iskanje po neredundantni bazi (nr) za organizem Delphinus leucas (taxid:9749) ter zaženemo blast



BLAST ® » blastp suite

Home

Standard Protein BLAST

blastn | **blastp** | blastx | tblastn | tblastx

BLASTP programs search protein databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

```
MDHAEENEILAATQRYYVERPIFSHPVLQERLHTKDKVSDSIGDKLKQAFTCT
PKKIRNI
IYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLPQGLAFAMLAAVPPVFGLY
SSFYPV
```

Query subrange

From

To

Or, upload file Choose File No file chosen

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: Non-redundant protein sequences (nr)

Organism (Optional): Delphinus leucas (taxid:9749) ☐ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude (Optional): ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

○ Quick BLASTP (Accelerated protein-protein BLAST)

● blastp (protein-protein BLAST)

○ PSI-BLAST (Position-Specific Iterated BLAST)

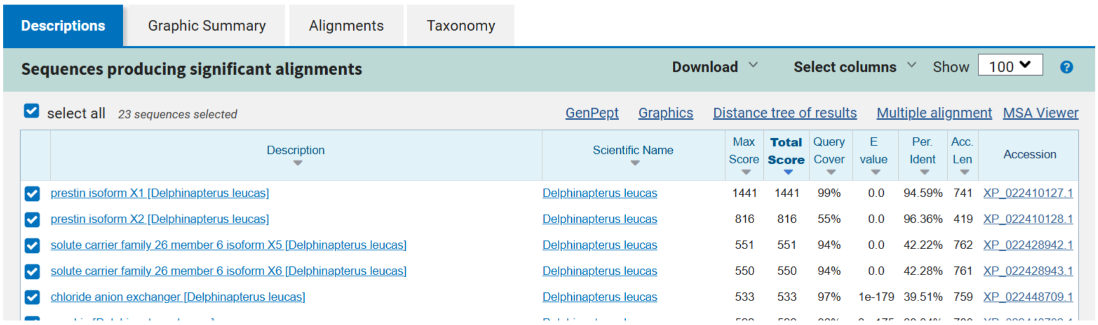
○ PHI-BLAST (Pattern Hit Initiated BLAST)

○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Slika 1. Prikaz nastavitvi v blastp.

- rezultat z največjim total score je izoformna oblika X1 proteina <u>prestina</u>:



Descriptions | Graphic Summary | Alignments | Taxonomy

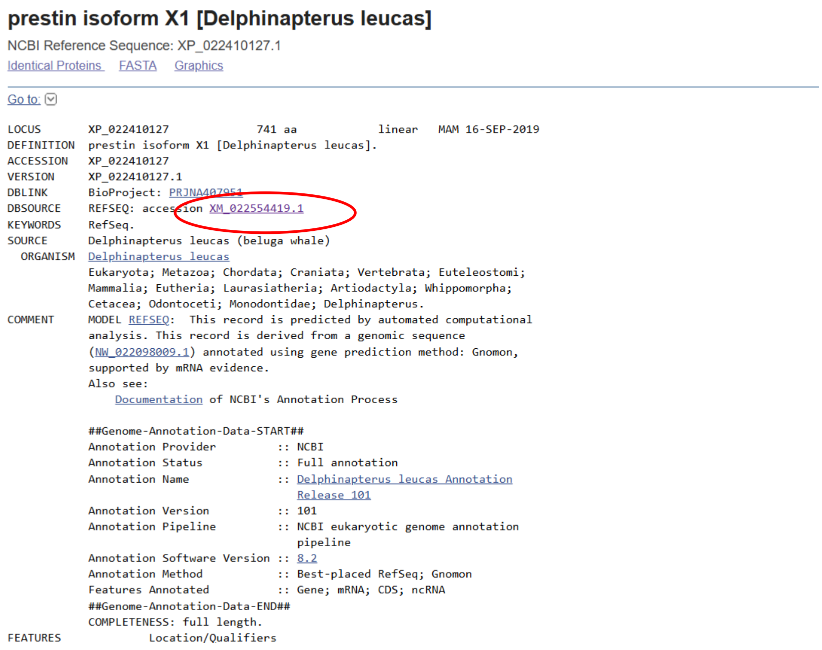
Sequences producing significant alignments — Download — Select columns — Show 100

☑ select all *23 sequences selected* — GenPept — Graphics — Distance tree of results — Multiple alignment — MSA Viewer

	Description	Scientific Name	Max Score	**Total Score**	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	prestin isoform X1 [Delphinapterus leucas]	Delphinapterus leucas	1441	1441	99%	0.0	94.59%	741	XP_022410127.1
☑	prestin isoform X2 [Delphinapterus leucas]	Delphinapterus leucas	816	816	55%	0.0	96.36%	419	XP_022410128.1
☑	solute carrier family 26 member 6 isoform X5 [Delphinapterus leucas]	Delphinapterus leucas	551	551	94%	0.0	42.22%	762	XP_022428942.1
☑	solute carrier family 26 member 6 isoform X6 [Delphinapterus leucas]	Delphinapterus leucas	550	550	94%	0.0	42.28%	761	XP_022428943.1
☑	chloride anion exchanger [Delphinapterus leucas]	Delphinapterus leucas	533	533	97%	1e-179	39.51%	759	XP_022448709.1

Slika 2. Rezultati, ki jih dobimo z blastom, sortirani po total score.

- odpremo ta zapis v bazi GenBank
- s klikom na RefSEQ lahko odpremo izhodiščno mRNA zaporedje
- izberemo CDS in vidimo, da se ta konča s STOP kodonom <u>tag</u>



prestin isoform X1 [Delphinapterus leucas]

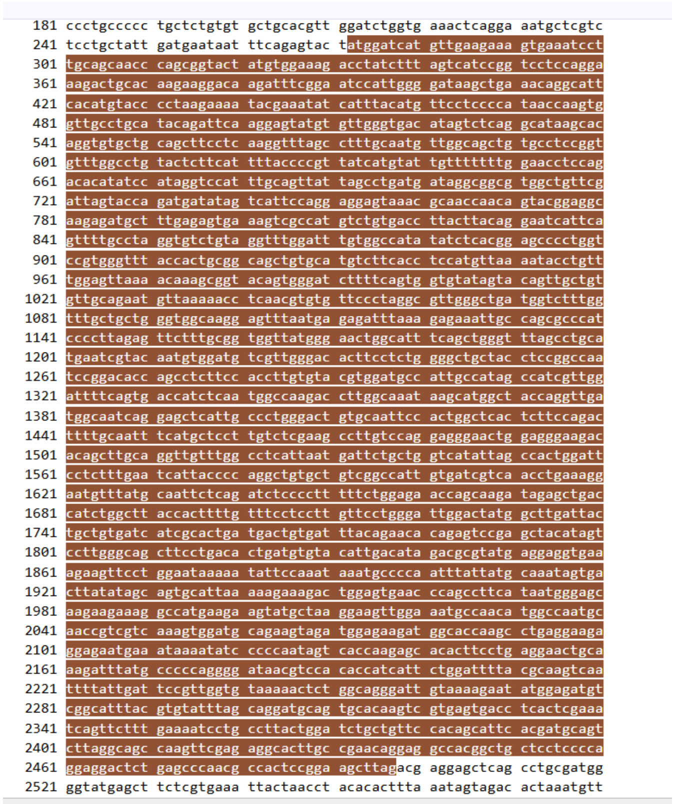
NCBI Reference Sequence: XP_022410127.1

Identical Proteins FASTA Graphics

Go to:

```
LOCUS       XP_022410127             741 aa            linear   MAM 16-SEP-2019
DEFINITION  prestin isoform X1 [Delphinapterus leucas].
ACCESSION   XP_022410127
VERSION     XP_022410127.1
DBLINK      BioProject: PRJNA407951
DBSOURCE    REFSEQ: accession XM_022554419.1
KEYWORDS    RefSeq.
SOURCE      Delphinapterus leucas (beluga whale)
  ORGANISM  Delphinapterus leucas
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Artiodactyla; Whippomorpha;
            Cetacea; Odontoceti; Monodontidae; Delphinapterus.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_022098009.1) annotated using gene prediction method: Gnomon,
            supported by mRNA evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Name             :: Delphinapterus leucas Annotation
                                           Release 101
            Annotation Version          :: 101
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 8.2
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
```

Slika 3. Prikaz zapisa tega proteina v GenBank, obkrožena je povezava do mRNA.



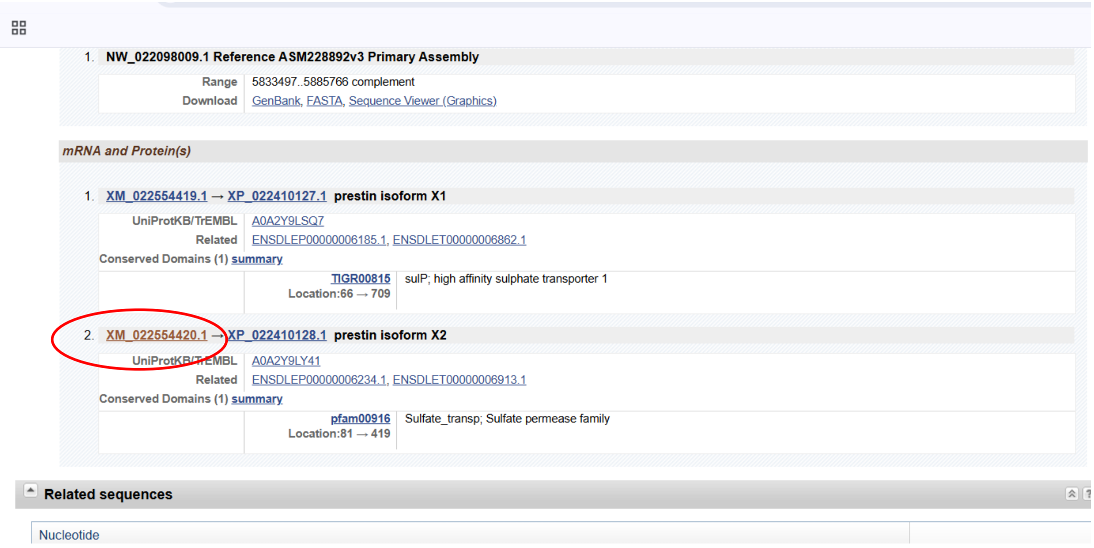
```
  181 ccctgccccc tgctctgtgt gctgcacgtt ggatctggtg aaactcagga aatgctcgtc
  241 tcctgctatt gatgaataat ttcagagtac tatggatcat gttgaagaaa gtgaaatcct
  301 tgcagcaacc cagcggtact atgtggaaag acctatcttt agtcatccgg tcctccagga
  361 aagactgcac aagaaggaca agatttcgga atccattggg gataagctga aacaggcatt
  421 cacatgtacc cctaagaaaa tacgaaatat catttacatg ttcctcccca taaccaagtg
  481 gttgcctgca tacagattca aggagtatgt gttgggtgac atagtctcag gcataagcac
  541 aggtgtgctg cagcttcctc aaggtttagc ctttgcaatg ttggcagctg tgcctccggt
  601 gtttggcctg tactcttcat tttaccccgt tatcatgtat tgttttttg gaacctccag
  661 acacatatcc ataggtccat ttgcagttat tagcctgatg ataggcggcg tggctgttcg
  721 attagtacca gatgatatag tcattccagg aggagtaaac gcaaccaaca gtacggaggc
  781 aagagatgct ttgagagtga aagtcgccat gtctgtgacc ttacttacag gaatcattca
  841 gttttgccta ggtgtctgta ggtttggatt tgtggccata tatctcacgg agcccctggt
  901 ccgtgggttt accactgcgg cagctgtgca tgtcttcacc tccatgttaa aatacctgtt
  961 tggagttaaa acaaagcggt acagtgggat cttttcagtg gtgtatagta cagttgctgt
 1021 gttgcagaat gttaaaaacc tcaacgtgtg ttccctaggc gttgggctga tggtctttgg
 1081 tttgctgctg ggtggcaagg agtttaatga gagatttaaa gagaaattgc cagcgcccat
 1141 cccccttagag ttctttgcgg tggttatggg aactggcatt tcagctgggt ttagcctgca
 1201 tgaatcgtac aatgtggatg tcgttgggac acttcctctg ggctgctac ctccggccaa
 1261 tccggacacc agcctcttcc accttgtgta cgtggatgcc attgccatag ccatcgttgg
 1321 attttcagtg accatctcaa tggccaagac cttggcaaat aagcatggct accaggttga
 1381 tggcaatcag gagctcattg ccctgggact gtgcaattcc actggctcac tcttccagac
 1441 ttttgcaatt tcatgctcct tgtctcgaag ccttgtccag gagggaactg gagggaagac
 1501 acagcttgca ggttgtttgg cctcattaat gattctgctg gtcatattag ccactggatt
 1561 cctctttgaa tcattacccc aggctgtgct gtcggccatt gtgatcgtca acctgaaagg
 1621 aatgtttatg caattctcag atctcccctt tttctggaga accagcaaga tagagctgac
 1681 catctggctt accacttttg tttcctcctt gttcctggga ttggactatg gcttgattac
 1741 tgctgtgatc atcgcactga tgactgtgat ttacagaaca cagagtccga gctacatagt
 1801 ccttgggcag cttcctgaca ctgatgtgta cattgacata gacgcgtatg aggaggtgaa
 1861 agaagttcct ggaataaaaa tattccaaat aaatgcccca atttattatg caaatagtga
 1921 cttatatagc agtgcattaa aaagaaagac tggagtgaac ccagccttca taatgggagc
 1981 aagaagaaag gccatgaaga agtatgctaa ggaagttgga aatgccaaca tggccaatgc
 2041 aaccgtcgtc aaagtggatg cagaagtaga tggagaagat ggcaccaagc ctgaggaaga
 2101 ggagaatgaa ataaaatatc ccccaatagt caccaagagc acacttcctg aggaactgca
 2161 aagatttatg cccccagggg ataacgtcca caccatcatt ctggatttta cgcaagtcaa
 2221 ttttattgat tccgttggtg taaaaactct ggcagggatt gtaaaagaat atggagatgt
 2281 cggcatttac gtgtatttag caggatgcag tgcacaagtc gtgagtgacc tcactcgaaa
 2341 tcagttcttt gaaaatcctg ccttactgga tctgctgttc cacagcattc acgatgcagt
 2401 cttaggcagc caagttcgag aggcacttgc cgaacaggag gccacggctg ctcctcccca
 2461 ggaggactct gagcccaacg ccactccgga agcttagacg aggagctcag cctgcgatgg
 2521 ggtatgagct tctcgtgaaa ttactaacct acacacttta aatagtagac actaaatgtt
```

Slika 4. Izbrana CDS regija mRNA zapisa, očitno je tag STOP kodon.

- do X2 izoformne variante tega proteina lahko pridemo na isti način iz okna z rezultati blasta, lahko pa direktno iz GeneBank s pomočjo GeneID

```
                     /chromosome="Unknown"
                     /sex="female"
                     /tissue_type="blood"
                     /dev_stage="adult"
     Protein         1..741
                     /product="prestin isoform X1"
                     /calculated_mol_wt=80774
     Region          66..709
                     /region_name="PRK11660"
                     /note="putative transporter; Provisional; cl46980"
                     /db_xref="CDD:481320"
     CDS             1..741
                     /gene="SLC26A5"
                     /coded_by="XM_022554419.1:272..2497"
                     /db_xref="GeneID:111164632"
ORIGIN
        1 mdhveeseil aatqryyver pifshpvlqe rlhkkdkise sigdklkqaf tctpkkirni
       61 iymflpitkw lpayrfkeyv lgdivsgist gvlqlpqgla famlaavppv fglyssfypv
      121 imycffgtsr hisigpfavi slmiggvavr lvpddivipg gvnatnstea rdalrvkvam
      181 svtlltgiiq fclgvcrfgf vaiylteplv rgfttaaavh vftsmlkylf gvktkrysgi
      241 fsvvystvav lqnvknlnvc slgvglmvfg lllggkefne rfkeklpapi pleffavvmg
      301 tgisagfslh esynvdvvgt lplgllppan pdtslfhlvy vdaiaiaivg fsvtismakt
      361 lankhgyqvd gnqelialgl cnstgslfqt faiscslsrs lvqegtggkt qlagclaslm
      421 illvilatgf lfeslpqavl saivivnlkg mfmqfsdlpf fwrtskielt iwlttfvssl
      481 flgldyglit aviialmtvi yrtqspsyiv lgqlpdtdvy ididayeevk evpgikifqi
      541 napiyyansd lyssalkrkt gvnpafimga rrkamkkyak evgnanmana tvvkvdaevd
      601 gedgtkpeee eneikyppiv tkstlpeelq rfmppgdnvh tiildftqvn fidsvgvktl
      661 agivkeygdv giyvylagcs aqvvsdltrn qffenpalld llfhsihdav lgsqvreala
      721 eqeataappq edsepnatpe a
//
```

Slika 5. V zapisu X1 izoforma najdemo GeneID.



1. **NW_022098009.1 Reference ASM228892v3 Primary Assembly**

Range	5833497..5885766 complement
Download	GenBank, FASTA, Sequence Viewer (Graphics)

mRNA and Protein(s)

1. XM_022554419.1 → XP_022410127.1 **prestin isoform X1**

UniProtKB/TrEMBL	A0A2Y9LSQ7
Related	ENSDLEP00000006185.1, ENSDLET00000006862.1

Conserved Domains (1) summary

TIGR00815 Location:66 → 709	sulP; high affinity sulphate transporter 1

2. XM_022554420.1 → XP_022410128.1 **prestin isoform X2**

UniProtKB/TrEMBL	A0A2Y9LY41
Related	ENSDLEP00000006234.1, ENSDLET00000006913.1

Conserved Domains (1) summary

pfam00916 Location:81 → 419	Sulfate_transp; Sulfate permease family

Related sequences

Nucleotide

Slika 6. Med zapisi za ta gen najdemo mRNA zaporedje za X2 izoformno obliko

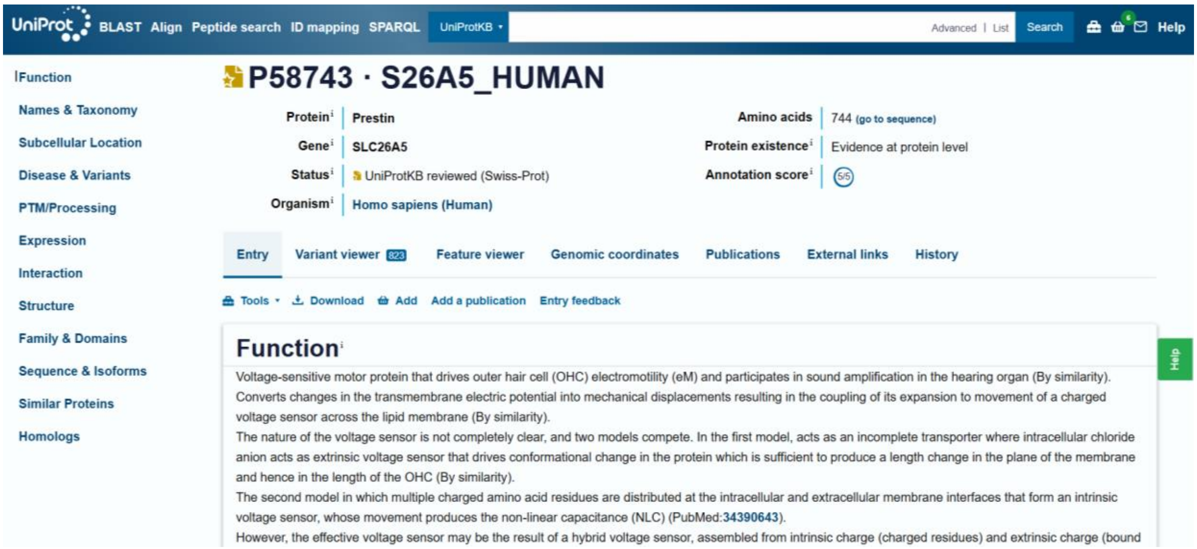
- spet izberemo CDS regijo tega zapisa in odčitamo STOP kodon izoforma X2: <u>tga</u>

```
                     EFNERFKEKLPAPIPLEFFAVVMGTGISAGFSLHESYNVDVVGTLPLGLLPPANPDTS
                     LFHLVYVDAIAIAIVGFSVTISMAKTLANKHGYQVDGNQELIALGLCNSTGSLFQTFA
                     ISCSLSRSLVQEGTGGKTQIHAPVSCV"
ORIGIN
        1 ccgcccaggg cagcgccggg cagcgccggc cgggccaagg agcaagctct cccatctcca
       61 gctctcagcc tcacagatca agatttcact gaactgcact agacctccca tctggggacc
      121 ctgccccctg ctctgtgtgc tgcacgttgg atctggtgaa actcaggaaa tgctcgtctc
      181 ctgctattga tgaataattt cagagtacta tggatcatgt tgaagaaagt gaaatccttg
      241 cagcaaccca gcggtactat gtggaaagac ctatctttag tcatccggtc ctccaggaaa
      301 gactgcacaa gaaggacaag atttcggaat ccattgggga taagctgaaa caggcattca
      361 catgtacccc taagaaaata cgaaatatca tttacatgtt cctccccata accaagtggt
      421 tgcctgcata cagattcaag gagtatgtgt tgggtgacat agtctcaggc ataagcacag
      481 gtgtgctgca gcttcctcaa ggtttagcct ttgcaatgtt ggcagctgtg cctccggtgt
      541 ttggcctgta ctcttcattt taccccgtta tcatgtattg tttttttgga acctccagac
      601 acatatccat aggtccattt gcagttatta gcctgatgat aggcggcgtg gctgttcgat
      661 tagtaccaga tgatatagtc attccaggag gagtaaacgc aaccaacagt acggaggcaa
      721 gagatgcttt gagagtgaaa gtcgccatgt ctgtgacctt acttacagga atcattcagt
      781 tttgcctagg tgtctgtagg tttggatttg tggccatata tctcacggag cccctggtcc
      841 gtgggtttac cactgcggca gctgtgcatg tcttcacctc catgttaaaa tacctgtttg
      901 gagttaaaac aaagcggtac agtgggatct tttcagtggt gtatagtaca gttgctgtgt
      961 tgcagaatgt taaaaacctc aacgtgtgtt ccctaggcgt tgggctgatg gtctttggtt
     1021 tgctgctggg tggcaaggag tttaatgaga gatttaaaga gaaattgcca gcgcccatcc
     1081 ccttagagtt ctttgcggtg gttatgggaa ctggcatttc agctgggttt agcctgcatg
     1141 aatcgtacaa tgtggatgtc gttgggacac ttcctctggg gctgctacct ccggccaatc
     1201 cggacaccag cctcttccac cttgtgtacg tggatgccat tgccatagcc atcgttggat
     1261 tttcagtgac catctcaatg gccaagacct tggcaaataa gcatggctac caggttgatg
     1321 gcaatcagga gctcattgcc ctgggactgt gcaattccac tggctcactc ttccagactt
     1381 ttgcaatttc atgctccttg tctcgaagcc ttgtccagga gggaactgga gggaagacac
     1441 agatccatgc tccagtgagc tgtgtatgaa tcagagaaag gcatgattgg ttgaagcaaa
     1501 tcacaaggaa ctgaatctac agaaaagggg tatattcgcc agtcaaatcg aaaaattaaa
     1561 caacactccc tgctaccagt aaacctcaaa gttcagacat aggatgttgt catgaaagcc
     1621 tgag
//
```

Slika 7. mRNA zaporedje za izoform X2, očitno je tga STOP kodon.

Iskanje lastnosti proteina:

- v neredundantni bazi UniProt poiščemo zapise za človeški prestin (P58743) ter izoformno obliko X1 prestina belega kita (A0A2Y9LSQ7)
- v dobro anotiranem zapisu za človeški protein lahko odčitamo njegove funkcije
- pomembnost tega proteina pri belem kitu lahko sami sklepamo, lahko pa poiščemo relevantne znanstvene članke (kot sta *Hear That? Bats and Whales Share Sonar Protein. (2021). [Dataset]. AAAS Articles DO Group. American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/article.30946* ali *Li, Y., Liu, Z., Shi, P., & Zhang, J. (2010). The hearing gene Prestin unites echolocating bats and whales. Current Biology, 20(2), R55–R56. https://doi.org/10.1016/j.cub.2009.11.042*)



P58743 · S26A5_HUMAN

Protein: Prestin
Gene: SLC26A5
Status: UniProtKB reviewed (Swiss-Prot)
Organism: Homo sapiens (Human)
Amino acids: 744 (go to sequence)
Protein existence: Evidence at protein level
Annotation score: 5/5

Function

Voltage-sensitive motor protein that drives outer hair cell (OHC) electromotility (eM) and participates in sound amplification in the hearing organ (By similarity). Converts changes in the transmembrane electric potential into mechanical displacements resulting in the coupling of its expansion to movement of a charged voltage sensor across the lipid membrane (By similarity).

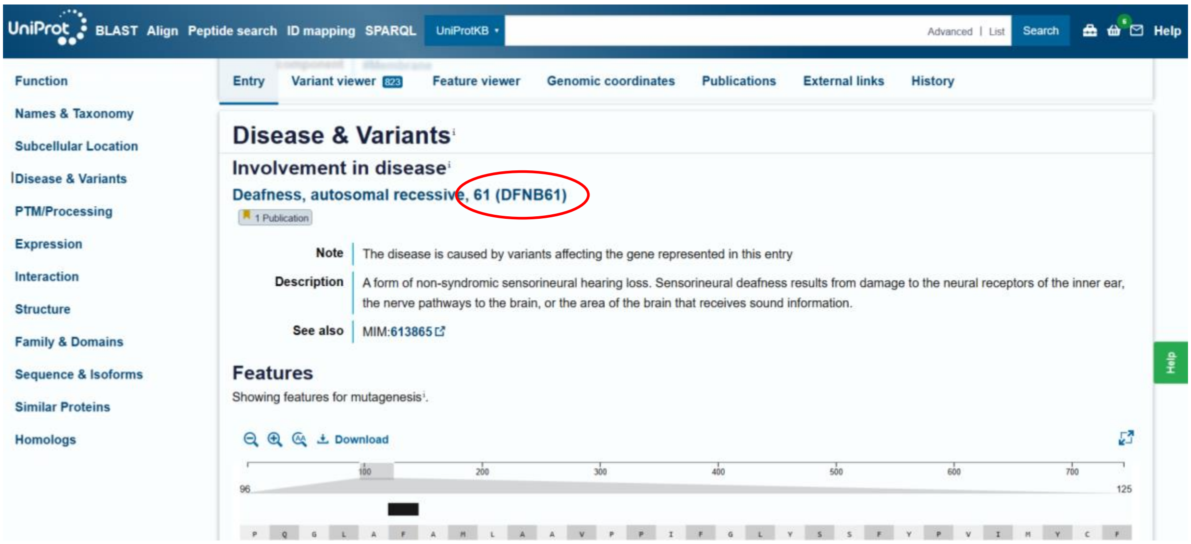
The nature of the voltage sensor is not completely clear, and two models compete. In the first model, acts as an incomplete transporter where intracellular chloride anion acts as extrinsic voltage sensor that drives conformational change in the protein which is sufficient to produce a length change in the plane of the membrane and hence in the length of the OHC (By similarity).

The second model in which multiple charged amino acid residues are distributed at the intracellular and extracellular membrane interfaces that form an intrinsic voltage sensor, whose movement produces the non-linear capacitance (NLC) (PubMed:34390643).

However, the effective voltage sensor may be the result of a hybrid voltage sensor, assembled from intrinsic charge (charged residues) and extrinsic charge (bound

Slika 8. Oddelek Function opisuje funkcije prestina kot motoričnega proteina precej pomembnega pri sluhu.

- v oddelku *Disease & Variants* pa poiščemo najbolj pogosto bolezen povezano s tem proteinom



Disease & Variants

Involvement in disease

Deafness, autosomal recessive, 61 (DFNB61)
1 Publication

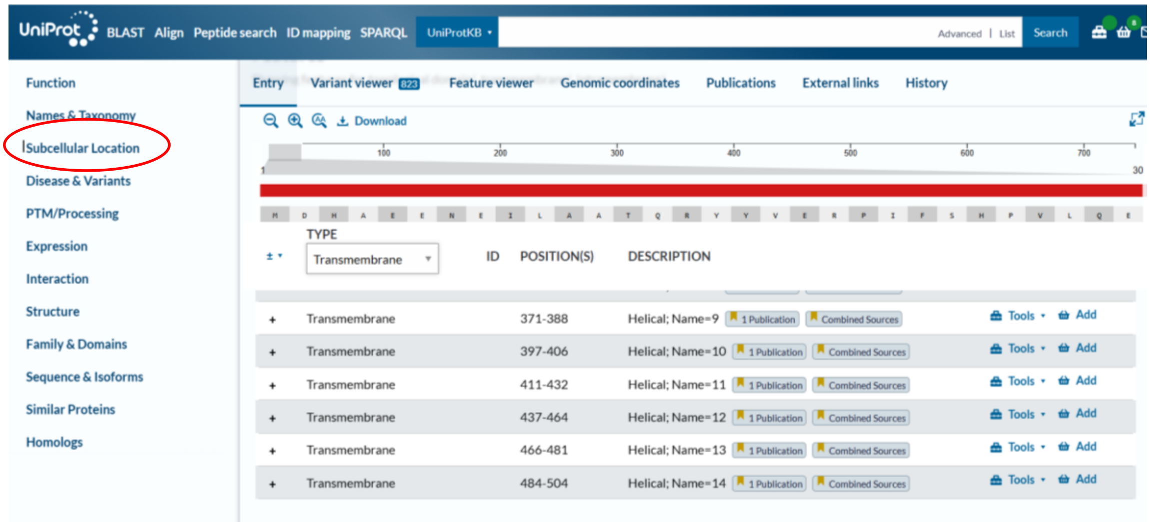
Note: The disease is caused by variants affecting the gene represented in this entry
Description: A form of non-syndromic sensorineural hearing loss. Sensorineural deafness results from damage to the neural receptors of the inner ear, the nerve pathways to the brain, or the area of the brain that receives sound information.
See also: MIM:613865

Features

Showing features for mutagenesis.

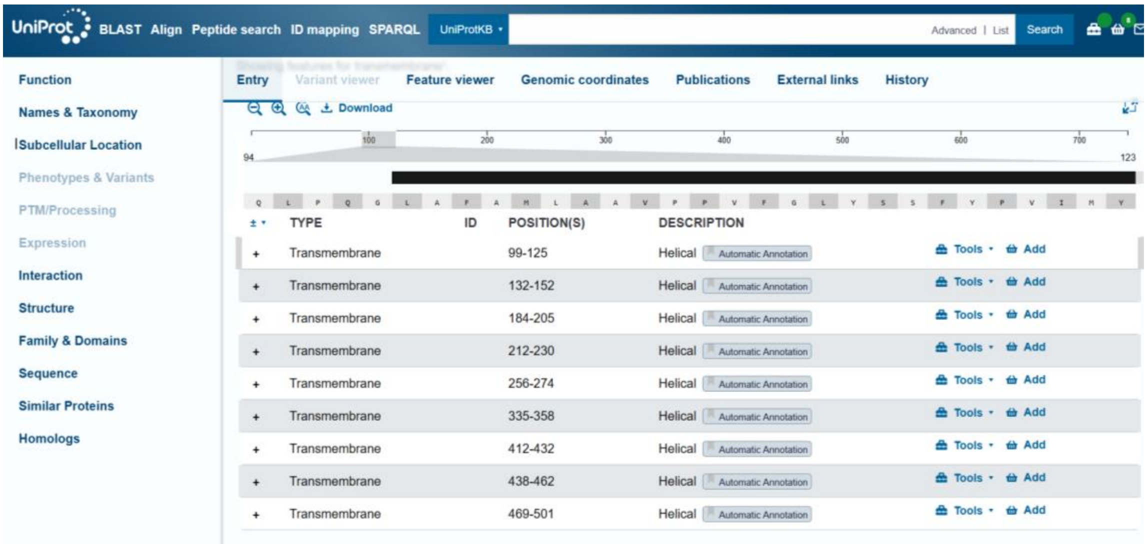
Slika 9. Bolezen je 'Deafness, autosomal recessive 61', s kratico DFNB61

- število transmembranskih vijačnic poiščemo v oddelku *Subcellular Location*



TYPE	ID	POSITION(S)	DESCRIPTION
Transmembrane		371-388	Helical; Name=9 1 Publication Combined Sources
Transmembrane		397-406	Helical; Name=10 1 Publication Combined Sources
Transmembrane		411-432	Helical; Name=11 1 Publication Combined Sources
Transmembrane		437-464	Helical; Name=12 1 Publication Combined Sources
Transmembrane		466-481	Helical; Name=13 1 Publication Combined Sources
Transmembrane		484-504	Helical; Name=14 1 Publication Combined Sources

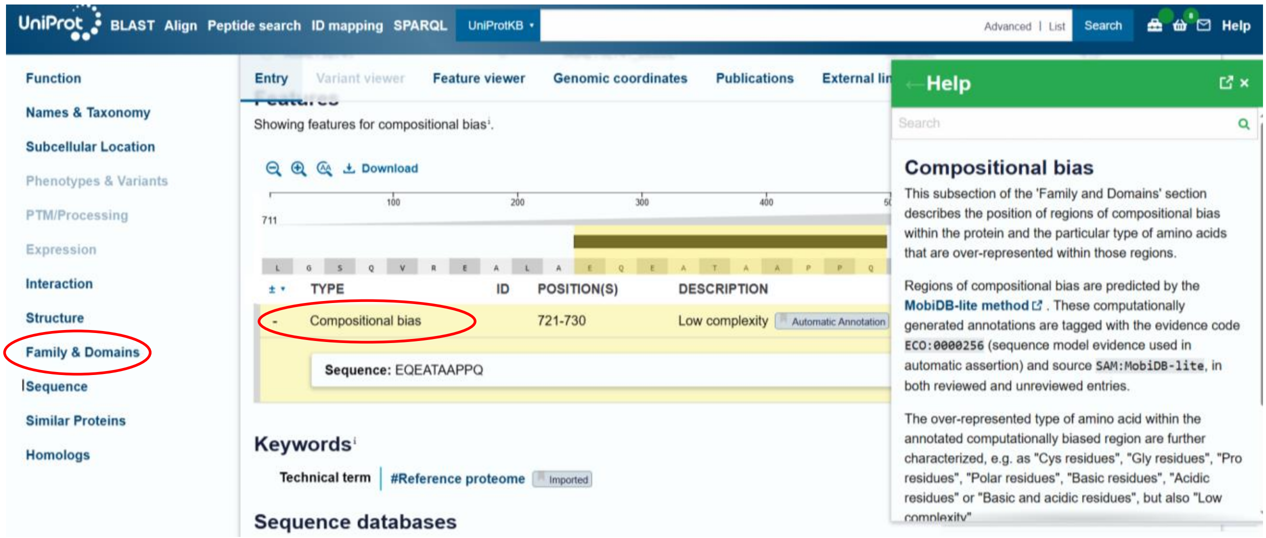
Slika 10. Pri človeškem proteinu je prisotnih 14 transmembranskih vijačnic.



TYPE	ID	POSITION(S)	DESCRIPTION
Transmembrane		99-125	Helical Automatic Annotation
Transmembrane		132-152	Helical Automatic Annotation
Transmembrane		184-205	Helical Automatic Annotation
Transmembrane		212-230	Helical Automatic Annotation
Transmembrane		256-274	Helical Automatic Annotation
Transmembrane		335-358	Helical Automatic Annotation
Transmembrane		412-432	Helical Automatic Annotation
Transmembrane		438-462	Helical Automatic Annotation
Transmembrane		469-501	Helical Automatic Annotation

Slika 11. Pri prestinu belega kita je pa 9 transmembranskih vijačnic.

- pri prestinu belega kita pa poiščemo še pomen anotacije *Compositional bias*, lahko pa preverimo z uporabo dodatnih spletnih vir



Showing features for compositional bias.

TYPE	ID	POSITION(S)	DESCRIPTION
Compositional bias		721-730	Low complexity Automatic Annotation

Sequence: EQEATAAPPQ

Keywords

Technical term #Reference proteome Imported

Sequence databases

Help

Compositional bias

This subsection of the 'Family and Domains' section describes the position of regions of compositional bias within the protein and the particular type of amino acids that are over-represented within those regions.

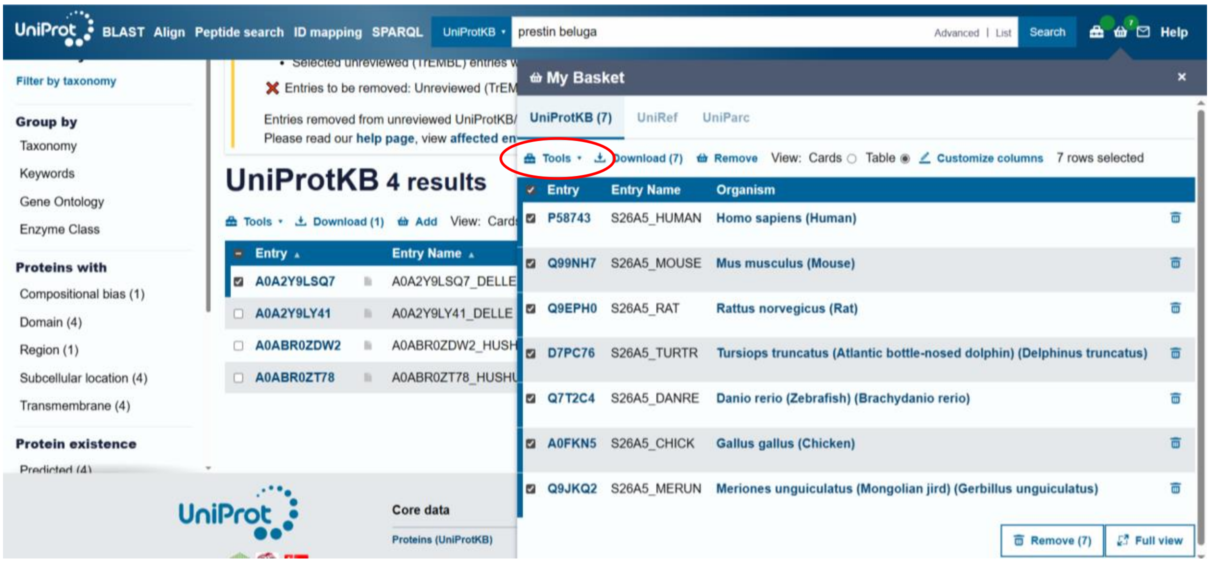
Regions of compositional bias are predicted by the **MobiDB-lite method**. These computationally generated annotations are tagged with the evidence code `ECO:0000256` (sequence model evidence used in automatic assertion) and source `SAM:MobiDB-lite`, in both reviewed and unreviewed entries.

The over-represented type of amino acid within the annotated computationally biased region are further characterized, e.g. as "Cys residues", "Gly residues", "Pro residues", "Polar residues", "Basic residues", "Acidic residues" or "Basic and acidic residues", but also "Low complexity"

Slika 12. Razlaga anotacije Compositional bias iz UniProta.

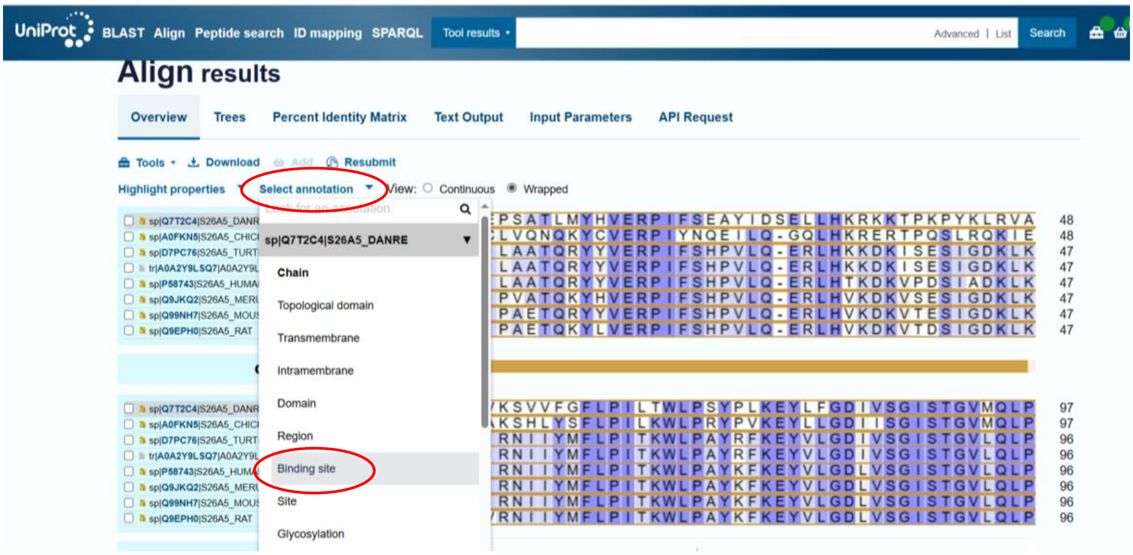
Poravnava več zaporedij in iskanje vezavnega mesta:

- izberemo nekaj dobro anotiranih zaporedij za prestin in zapis za belega kita, ter jih damo v *Basket* v UniProtu
- naredimo poravnavo teh zaporedij



☑	Entry	Entry Name	Organism
☑	P58743	S26A5_HUMAN	Homo sapiens (Human)
☑	Q99NH7	S26A5_MOUSE	Mus musculus (Mouse)
☑	Q9EPH0	S26A5_RAT	Rattus norvegicus (Rat)
☑	D7PC76	S26A5_TURTR	Tursiops truncatus (Atlantic bottle-nosed dolphin) (Delphinus truncatus)
☑	Q7T2C4	S26A5_DANRE	Danio rerio (Zebrafish) (Brachydanio rerio)
☑	A0FKN5	S26A5_CHICK	Gallus gallus (Chicken)
☑	Q9JKQ2	S26A5_MERUN	Meriones unguiculatus (Mongolian jird) (Gerbillus unguiculatus)

Slika 12. Dodamo vsa zaporedja v Basket, izberemo jih in pod Tools izberemo Align



```
sp|Q7T2C4|S26A5_DANR   PSATLMYHVERPIFSEAYIDSELLHKRKKTPKPYKLRVA   48
sp|A0FKN5|S26A5_CHICI  LVQNQKYCVERPIYNQEILQ-GQLHKRERTPQSLRQKIE  48
sp|D7PC76|S26A5_TURT   LAATQRYYVERPIFSHPVLQ-ERLHKKDKISESIGDKLK  47
tr|A0A2Y9LSQ7|A0A2Y9L  LAATQRYYVERPIFSHPVLQ-ERLHKKDKISESIGDKLK  47
sp|P58743|S26A5_HUMA   LAATQRYYVERPIFSHPVLQ-ERLHTKDKVPDSIADKLK  47
sp|Q9JKQ2|S26A5_MERU   PVATQKYHVERPIFSHPVLQ-ERLHVKDKVSESIGDKLK  47
sp|Q99NH7|S26A5_MOUS   PAETQRYYVERPIFSHPVLQ-ERLHVKDKVTESIGDKLK  47
sp|Q9EPH0|S26A5_RAT    PAETQKYLVERPIFSHPVLQ-ERLHVKDKVTDSIGDKLK  47

sp|Q7T2C4|S26A5_DANR   KSVVFGFLPILTWLPSYPLKEYLFGDIVSGISTGVMQLP  97
sp|A0FKN5|S26A5_CHICI  KSHLYSFLPILKWLPRYPVKEYLLGDIISGISTGVMQLP  97
sp|D7PC76|S26A5_TURT   RNIIYMFLPITKWLPAYRFKEYVLGDIVSGISTGVLQLP  96
tr|A0A2Y9LSQ7|A0A2Y9L  RNIIYMFLPITKWLPAYRFKEYVLGDIVSGISTGVLQLP  96
sp|P58743|S26A5_HUMA   RNIIYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLP  96
sp|Q9JKQ2|S26A5_MERU   RNIIYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLP  96
sp|Q99NH7|S26A5_MOUS   RNIIYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLP  96
sp|Q9EPH0|S26A5_RAT    RNIIYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLP  96
```

Slika 13. Rezultati poravnave, obkroženo je kako izberemo anotacijo.

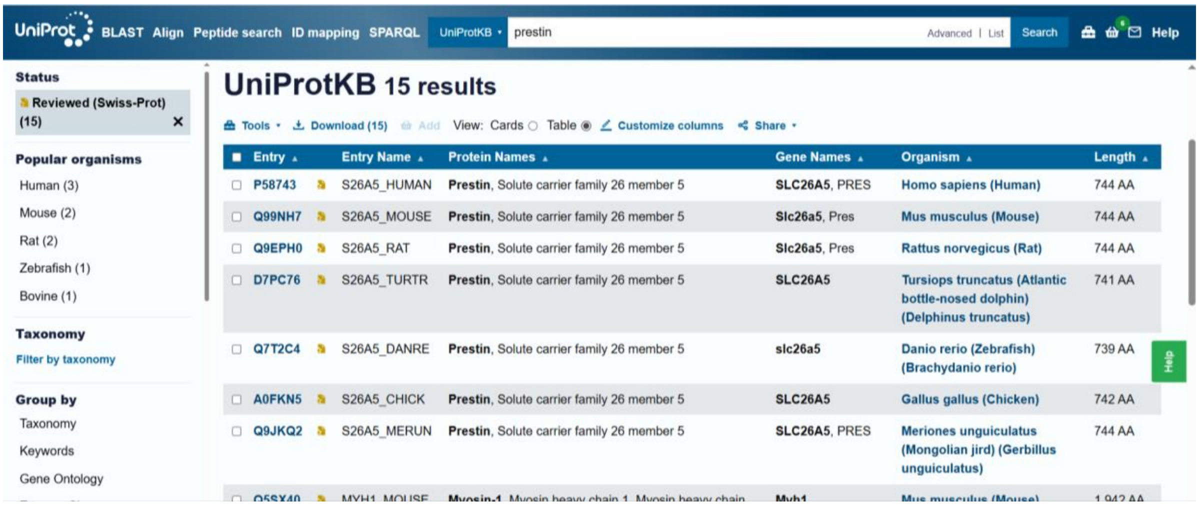
- v oknu z rezultati poravnave, izberemo eno od dobro anotiranih zaporedij in vklopimo prikaz anotacije za vezavno mesto
- poiščemo označeni aminokislinski ostanek, očitno je to <u>serin na mestu 398</u> v našem proteinu

```
☐ sp|Q9JKQ2|S26A5_MERUN              FHLVYVDAIAIAIVGFSVTISMAKTLANKHGYQVDGNQELIALGICNSI   384
☐ sp|Q99NH7|S26A5_MOUSE              FHLVYVDAIAIAIVGFSVTISMAKTLANKHGYQVDGNQELIALGICNSI   384
☐ sp|Q9EPH0|S26A5_RAT                FHLVYVDAIAIAIVGFSVTISMAKTLANKHGYQVDGNQELIALGICNSI   384

                 Q7T2C4:Binding site

☐ sp|Q7T2C4|S26A5_DANRE              SSFFHTFVVTASMSRSLVQESTGGHTEIAGLLASLLVLLVVVAIGFVFQ   436
☐ sp|A0FKN5|S26A5_CHICK              GSFFQSFPITCSMSRSLVQESTGGKTQIAGALSSIMVLLVIVAIGYLFE   439
☐ sp|D7PC76|S26A5_TURTR              GSLFQTFAISCSLSRSLVQEGTGGKTQLAGCLASLMILLVILATGFLFE   433
☐ tr|A0A2Y9LSQ7|A0A2Y9LSQ7_DELLE     GSLFQTFAISCSLSRSLVQEGTGGKTQLAGCLASLMILLVILATGFLFE   433
☐ sp|P58743|S26A5_HUMAN              GSLFQTFSISCSLSRSLVQEGTGGKTQLAGCLASLMILLVILATGFLFE   433
☐ sp|Q9JKQ2|S26A5_MERUN              GSLFQTFSISCSLSRSLVQEGTGGKTQLAGCLASLMILLVILATGFLFE   433
☐ sp|Q99NH7|S26A5_MOUSE              GSLFQTFSISCSLSRSLVQEGTGGKTQLAGCLASLMILLVILATGFLFE   433
☐ sp|Q9EPH0|S26A5_RAT                GSLFQTFSISCSLSRSLVQEGTGGKTQLAGCLASLMILLVILATGFLFE   433

                 Q7T2C4:Binding site

☐ sp|Q7T2C4|S26A5_DANRE              PLPTTVLAAIIMVNLLGMFKQTRDIPVLWRKSKIELAIWLVSFFASVLL   485
☐ sp|A0FKN5|S26A5_CHICK              PLPQTVLAAIVMVNLKGMFKQFADVAHFWRTSKIELAIWVVAFVASLFL   488
☐ sp|D7PC76|S26A5_TURTR              SLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSLFL   482
☐ tr|A0A2Y9LSQ7|A0A2Y9LSQ7_DELLE     SLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSLFL   482
☐ sp|P58743|S26A5_HUMAN              SLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSLFL   482
☐ sp|Q9JKQ2|S26A5_MERUN              SLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSLFL   482
☐ sp|Q99NH7|S26A5_MOUSE              SLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSLFL   482
☐ sp|Q9EPH0|S26A5_RAT                SLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSLFL   482
```

Slika 14. V zaporedju poiščemo aminokislinski ostanek, ki je anotiran kot vezavno mesto.

Iskanje strukture s pomočjo filogenetskega drevesa:

- v UniProtu obstaja 7 Swiss-Prot recenziranih zapisov za prestin:

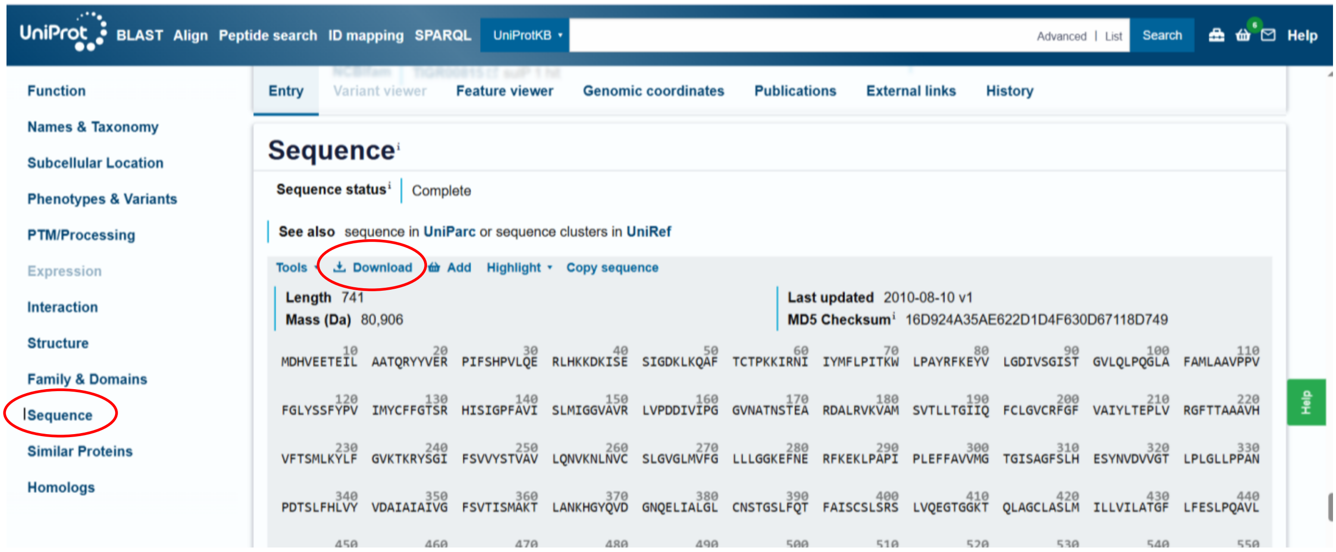


UniProtKB 15 results

	Entry	Entry Name	Protein Names	Gene Names	Organism	Length
☐	P58743	S26A5_HUMAN	**Prestin**, Solute carrier family 26 member 5	**SLC26A5**, PRES	Homo sapiens (Human)	744 AA
☐	Q99NH7	S26A5_MOUSE	**Prestin**, Solute carrier family 26 member 5	**Slc26a5**, Pres	Mus musculus (Mouse)	744 AA
☐	Q9EPH0	S26A5_RAT	**Prestin**, Solute carrier family 26 member 5	**Slc26a5**, Pres	Rattus norvegicus (Rat)	744 AA
☐	D7PC76	S26A5_TURTR	**Prestin**, Solute carrier family 26 member 5	**SLC26A5**	Tursiops truncatus (Atlantic bottle-nosed dolphin) (Delphinus truncatus)	741 AA
☐	Q7T2C4	S26A5_DANRE	**Prestin**, Solute carrier family 26 member 5	**slc26a5**	Danio rerio (Zebrafish) (Brachydanio rerio)	739 AA
☐	A0FKN5	S26A5_CHICK	**Prestin**, Solute carrier family 26 member 5	**SLC26A5**	Gallus gallus (Chicken)	742 AA
☐	Q9JKQ2	S26A5_MERUN	**Prestin**, Solute carrier family 26 member 5	**SLC26A5**, PRES	Meriones unguiculatus (Mongolian jird) (Gerbillus unguiculatus)	744 AA
☐	Q5SX40	MYH1_MOUSE	**Myosin-1**, Myosin heavy chain 1, Myosin heavy chain	**Myh1**	Mus musculus (Mouse)	1,942 AA

Slika 15. Prikaz vseh Swiss-Prot anotiranih zaporedij za prestin.

- za vsaki od teh zapisov odpremo in kopiramo zaporedje v FASTA formatu in jih vse damo v eno tekstualno datoteko skupaj z našim proteinom
- pri tem pa ustrezno imenujemo zapiske, da jih lahko kasneje razlikujemo po ogranizmu



Sequence

Sequence status Complete

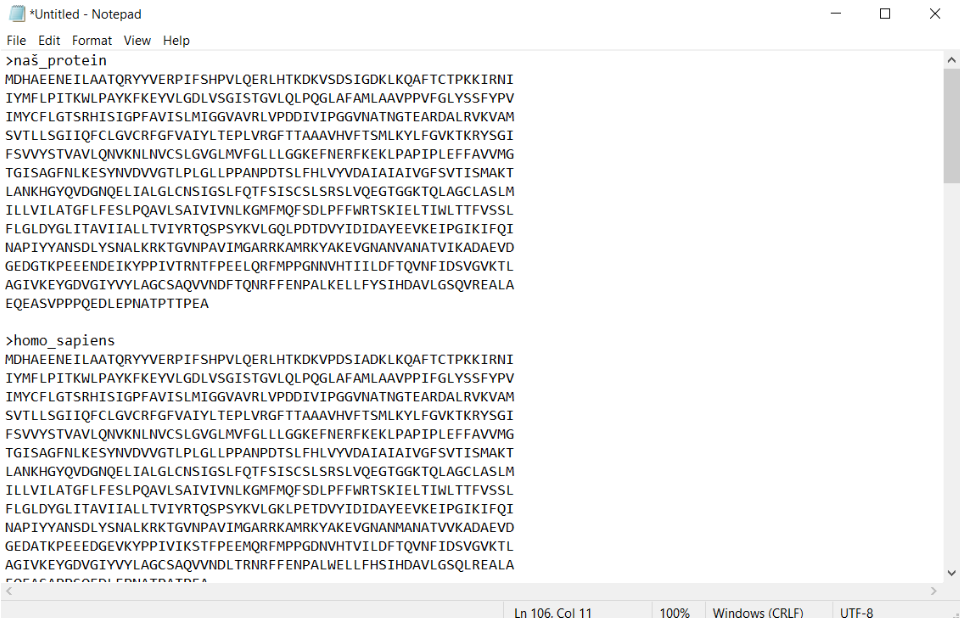
See also sequence in UniParc or sequence clusters in UniRef

Length 741
Mass (Da) 80,906

Last updated 2010-08-10 v1
MD5 Checksum 16D924A35AE622D1D4F630D67118D749

```
        10         20         30         40         50         60         70         80         90        100        110
MDHVEETEIL AATQRYYVER PIFSHPVLQE RLHKKDKISE SIGDKLKQAF TCTPKKIRNI IYMFLPITKW LPAYRFKEYV LGDIVSGIST GVLQLPQGLA FAMLAAVPPV

       120        130        140        150        160        170        180        190        200        210        220
FGLYSSFYPV IMYCFFGTSR HISIGPFAVI SLMIGGVAVR LVPDDIVIPG GVNATNSTEA RDALRVKVAM SVTLLTGIIQ FCLGVCRFGF VAIYLTEPLV RGFTTAAAVH

       230        240        250        260        270        280        290        300        310        320        330
VFTSMLKYLF GVKTKRYSGI FSVVYSTVAV LQNVKNLNVC SLGVGLMVFG LLLGGKEFNE RFKEKLPAPI PLEFFAVVMG TGISAGFSLH ESYNVDVVGT LPLGLLPPAN

       340        350        360        370        380        390        400        410        420        430        440
PDTSLFHLVY VDAIAIAIVG FSVTISMAKT LANKHGYQVD GNQELIALGL CNSTGSLFQT FAISCSLSRS LVQEGTGGKT QLAGCLASLM ILLVILATGF LFESLPQAVL

       450        460        470        480        490        500        510        520        530        540        550
```

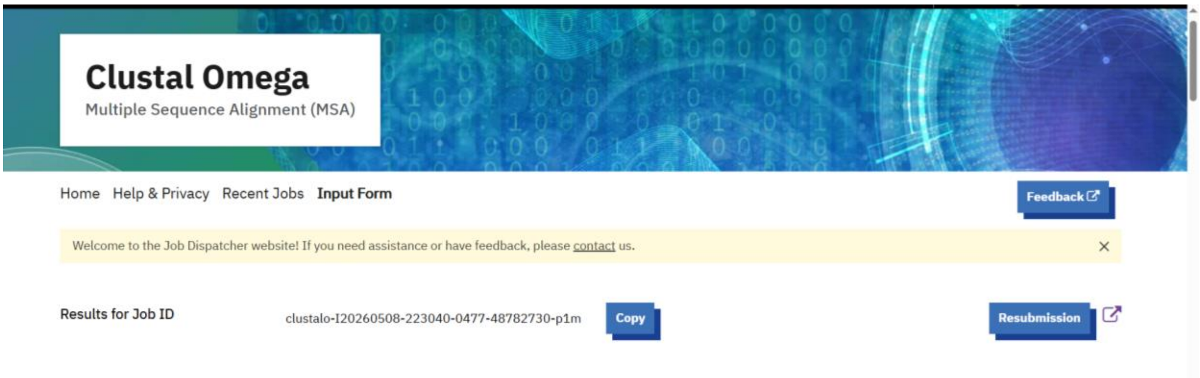
Slika 16. Za vsaki zapis izberemo Download pod oddelkom Sequence da dobimo zaporedje v FASTA formatu.



```
>naš_protein
MDHAEENEILAATQRYYVERPIFSHPVLQERLHTKDKVSDSIGDKLKQAFTCTPKKIRNI
IYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLPQGLAFAMLAAVPPVFGLYSSFYPV
IMYCFLGTSRHISIGPFAVISLMIGGVAVRLVPDDIVIPGGVNATNGTEARDALRVKVAM
SVTLLSGIIQFCLGVCRFGFVAIYLTEPLVRGFTTAAAVHVFTSMLKYLFGVKTKRYSGI
FSVVYSTVAVLQNVKNLNVCSLGVGLMVFGLLLGGKEFNERFKEKLPAPIPLEFFAVVMG
TGISAGFNLKESYNVDVVGTLPLGLLPPANPDTSLFHLVYVDAIAIAIVGFSVTISMAKT
LANKHGYQVDGNQELIALGLCNSIGSLFQTFSISCSLSRSLVQEGTGGKTQLAGCLASLM
ILLVILATGFLFESLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSL
FLGLDYGLITAVIIALLTVIYRTQSPSYKVLGQLPDTDVYIDIDAYEEVKEIPGIKIFQI
NAPIYYANSDLYSNALKRKTGVNPAVIMGARRKAMRKYAKEVGNANVANATVIKADAEVD
GEDGTKPEEENDEIKYPPIVTRNTFPEELQRFMPPGNNVHTIILDFTQVNFIDSVGVKTL
AGIVKEYGDVGIYVYLAGCSAQVVNDFTQNRFFENPALKELLFYSIHDAVLGSQVREALA
EQEASVPPPQEDLEPNATPTTPEA

>homo_sapiens
MDHAEENEILAATQRYYVERPIFSHPVLQERLHTKDKVPDSIADKLKQAFTCTPKKIRNI
IYMFLPITKWLPAYKFKEYVLGDLVSGISTGVLQLPQGLAFAMLAAVPPIFGLYSSFYPV
IMYCFLGTSRHISIGPFAVISLMIGGVAVRLVPDDIVIPGGVNATNGTEARDALRVKVAM
SVTLLSGIIQFCLGVCRFGFVAIYLTEPLVRGFTTAAAVHVFTSMLKYLFGVKTKRYSGI
FSVVYSTVAVLQNVKNLNVCSLGVGLMVFGLLLGGKEFNERFKEKLPAPIPLEFFAVVMG
TGISAGFNLKESYNVDVVGTLPLGLLPPANPDTSLFHLVYVDAIAIAIVGFSVTISMAKT
LANKHGYQVDGNQELIALGLCNSIGSLFQTFSISCSLSRSLVQEGTGGKTQLAGCLASLM
ILLVILATGFLFESLPQAVLSAIVIVNLKGMFMQFSDLPFFWRTSKIELTIWLTTFVSSL
FLGLDYGLITAVIIALLTVIYRTQSPSYKVLGKLPETDVYIDIDAYEEVKEIPGIKIFQI
NAPIYYANSDLYSNALKRKTGVNPAVIMGARRKAMRKYAKEVGNANMANATVVKADAEVD
GEDATKPEEEDGEVKYPPIVIKSTFPEEMQRFMPPGDNVHTVILDFTQVNFIDSVGVKTL
AGIVKEYGDVGIYVYLAGCSAQVVNDLTRNRFFENPALWELLFHSIHDAVLGSQLREALA
```

Slika 17. Zaporedja damo v tekstualno datoteko skupaj s našim proteinom da jih lahko hkrati kopiramo.

- odpremo program za poravnavo več zaporedij, kot je Clustal Omega, in v njega kopiramo vsa zaporedja:

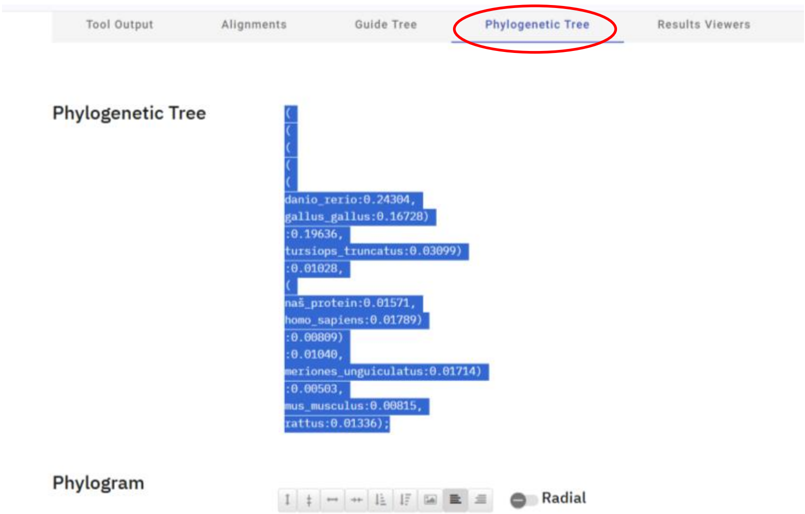


Welcome to the Job Dispatcher website! If you need assistance or have feedback, please contact us.

Results for Job ID clustalo-I20260508-223040-0477-48782730-p1m

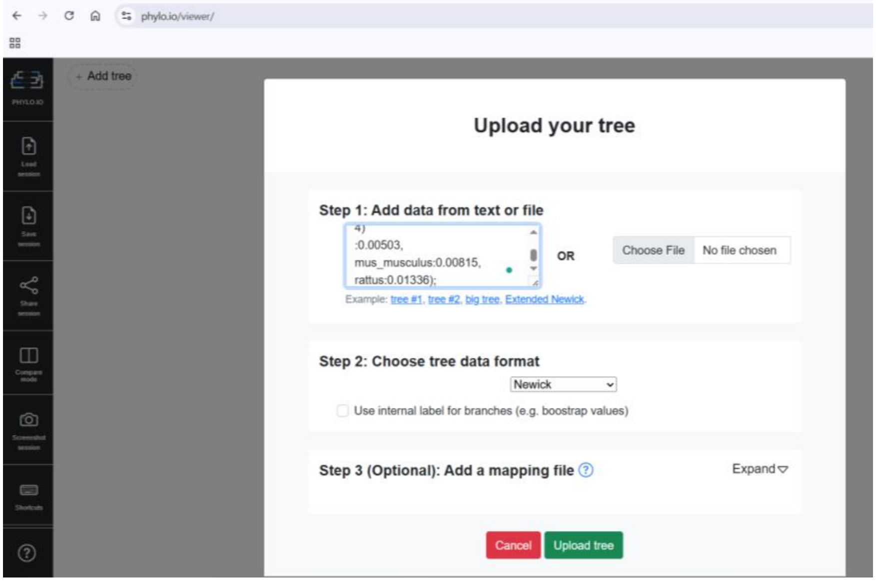
Slika 18. Prikaz sučelja Clustal Omega programa, ki ga lahko uporabimo za poravnavo zaporedij.

- v Clustal Omega dobimo zapis koordinat za filogenetsko drevo, ta lahko kopiramo in damo v program za vizualizacijo kot je phylo.io
- odčitamo kateri protein je evolucijski najbolj podoben prestinu belega kita



```
(
(
(
(
(
danio_rerio:0.24304,
gallus_gallus:0.16728)
:0.19636,
tursiops_truncatus:0.03099)
:0.01028,
(
naš_protein:0.01571,
homo_sapiens:0.01789)
:0.00809)
:0.01040,
meriones_unguiculatus:0.01714)
:0.00503,
mus_musculus:0.00815,
rattus:0.01336);
```

Slika 19. V Clustal Omega lahko kopiramo filogenetsko drevo.



Step 1: Add data from text or file

```
4)
:0.00503,
mus_musculus:0.00815,
rattus:0.01336);
```

No file chosen

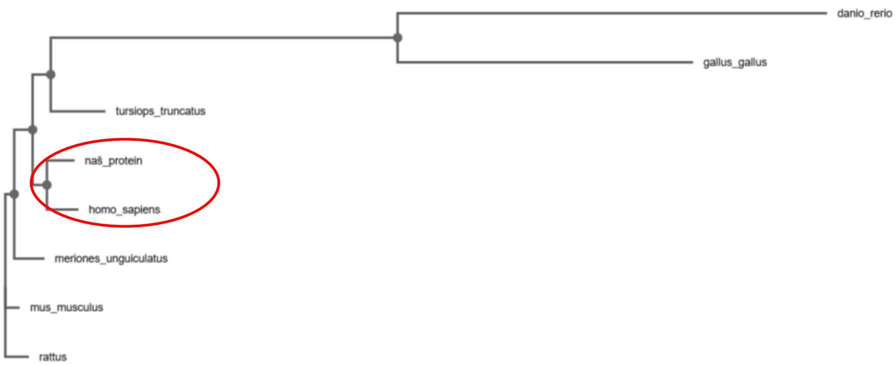
Step 2: Choose tree data format

Newick

☐ Use internal label for branches (e.g. boostrap values)

Step 3 (Optional): Add a mapping file

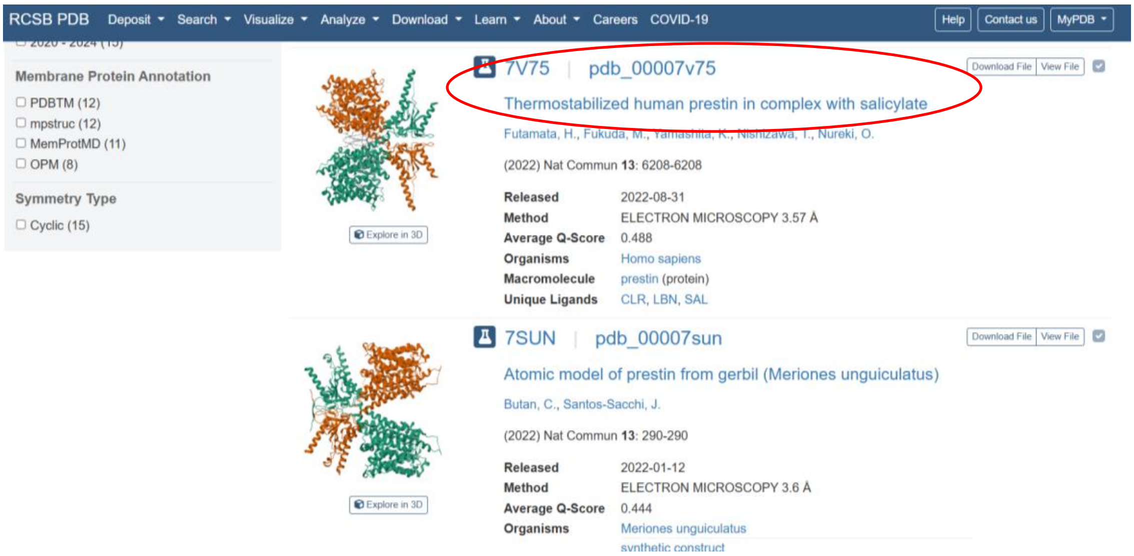
Slika 20. To drevo lahko prenesemo na program za vizualizacijo.



Slika 21. Očitno je človeški protein evolucijski najbližji našem.

Ugotovitev lastnosti strukture:

- strukturo človeškega prestina poiščemo na bazi PDB
- v tem zapisu lahko najdemo ekspresijski sistem in metodo določitve strukture



7V75 | pdb_00007v75

Thermostabilized human prestin in complex with salicylate

Futamata, H., Fukuda, M., Yamashita, K., Nishizawa, T., Nureki, O.

(2022) Nat Commun **13**: 6208-6208

Released 2022-08-31
Method ELECTRON MICROSCOPY 3.57 Å
Average Q-Score 0.488
Organisms Homo sapiens
Macromolecule prestin (protein)
Unique Ligands CLR, LBN, SAL

7SUN | pdb_00007sun

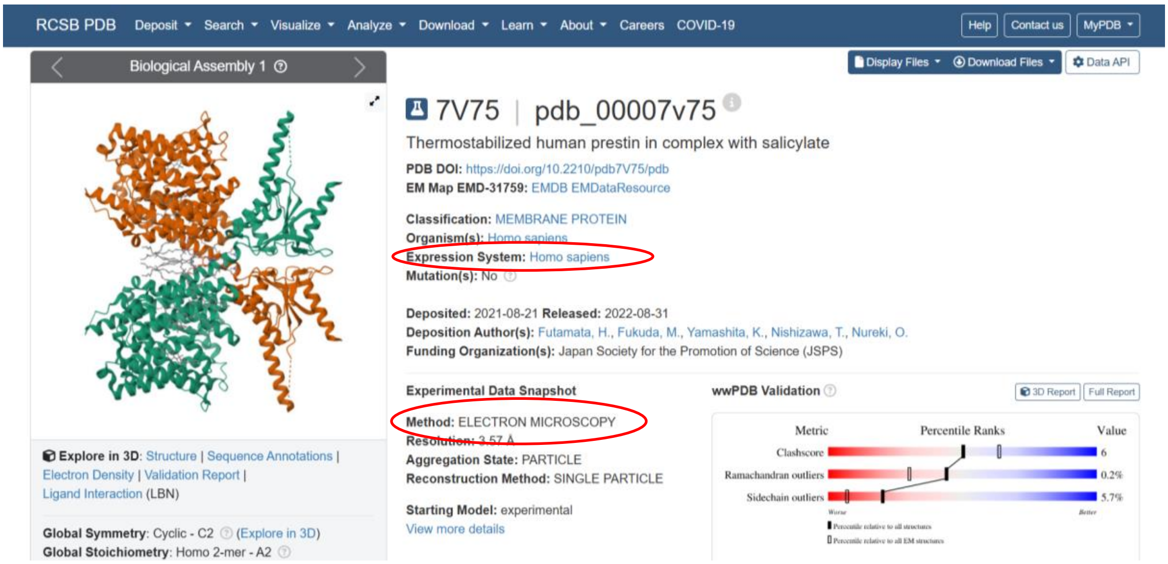
Atomic model of prestin from gerbil (Meriones unguiculatus)

Butan, C., Santos-Sacchi, J.

(2022) Nat Commun **13**: 290-290

Released 2022-01-12
Method ELECTRON MICROSCOPY 3.6 Å
Average Q-Score 0.444
Organisms Meriones unguiculatus
synthetic construct

Slika 22. Na PDB izberemo človeški prestin v kompleksu s salicilatom.



7V75 | pdb_00007v75

Thermostabilized human prestin in complex with salicylate

PDB DOI: https://doi.org/10.2210/pdb7V75/pdb
EM Map EMD-31759: EMDB EMDataResource

Classification: MEMBRANE PROTEIN
Organism(s): Homo sapiens
Expression System: Homo sapiens
Mutation(s): No

Deposited: 2021-08-21 **Released:** 2022-08-31
Deposition Author(s): Futamata, H., Fukuda, M., Yamashita, K., Nishizawa, T., Nureki, O.
Funding Organization(s): Japan Society for the Promotion of Science (JSPS)

Experimental Data Snapshot

Method: ELECTRON MICROSCOPY
Resolution: 3.57 Å
Aggregation State: PARTICLE
Reconstruction Method: SINGLE PARTICLE

Starting Model: experimental
View more details

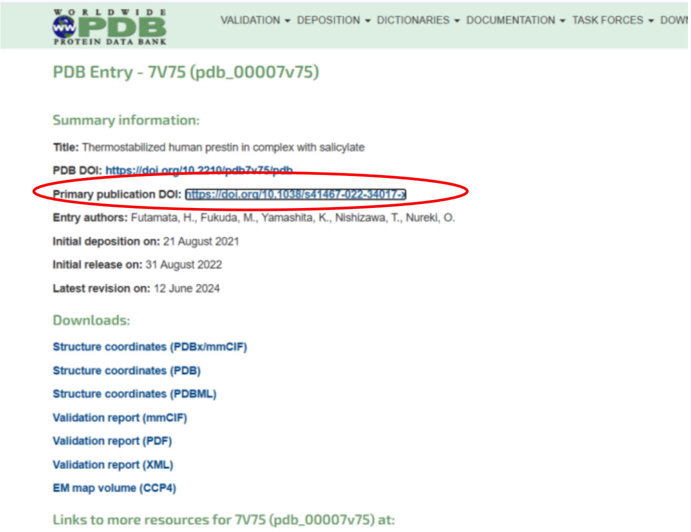
wwPDB Validation

Explore in 3D: Structure | Sequence Annotations | Electron Density | Validation Report | Ligand Interaction (LBN)

Global Symmetry: Cyclic - C2 (Explore in 3D)
Global Stoichiometry: Homo 2-mer - A2

Slika 23. Zapis iskanega proteina, obkrožene so iskane lastnosti.

- potem kliknemo na PDB DOI, da lahko dostopimo do izhodišnega članka iz strukturne biologije



PDB Entry - 7V75 (pdb_00007v75)

Summary information:

Title: Thermostabilized human prestin in complex with salicylate

PDB DOI: https://doi.org/10.2210/pdb7v75/pdb

Primary publication DOI: https://doi.org/10.1038/s41467-022-34017-x

Entry authors: Futamata, H., Fukuda, M., Yamashita, K., Nishizawa, T., Nureki, O.

Initial deposition on: 21 August 2021

Initial release on: 31 August 2022

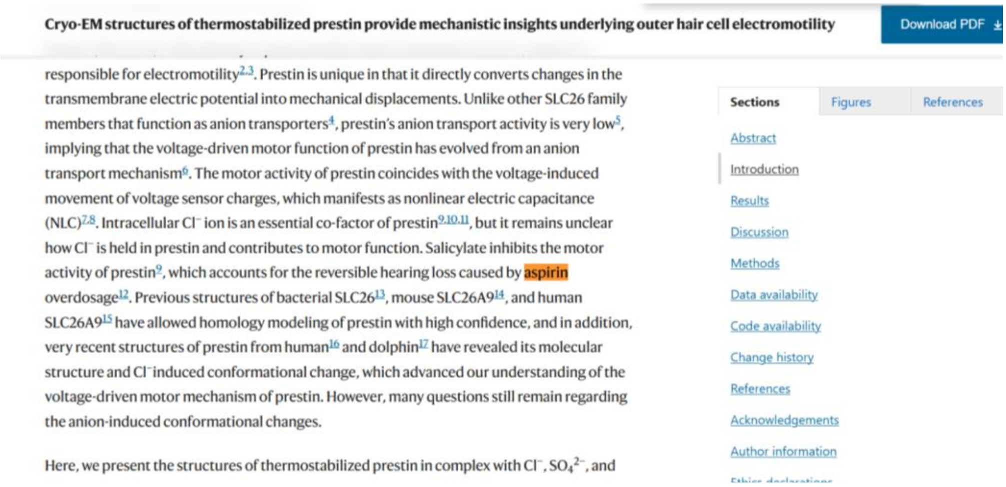
Latest revision on: 12 June 2024

Downloads:

Structure coordinates (PDBx/mmCIF)
Structure coordinates (PDB)
Structure coordinates (PDBML)
Validation report (mmCIF)
Validation report (PDF)
Validation report (XML)
EM map volume (CCP4)

Links to more resources for 7V75 (pdb_00007v75) at:

Slika 24. Najdemo link do izhodišnega članka.



Cryo-EM structures of thermostabilized prestin provide mechanistic insights underlying outer hair cell electromotility

responsible for electromotility[2,3]. Prestin is unique in that it directly converts changes in the transmembrane electric potential into mechanical displacements. Unlike other SLC26 family members that function as anion transporters[4], prestin's anion transport activity is very low[5], implying that the voltage-driven motor function of prestin has evolved from an anion transport mechanism[6]. The motor activity of prestin coincides with the voltage-induced movement of voltage sensor charges, which manifests as nonlinear electric capacitance (NLC)[7,8]. Intracellular Cl^- ion is an essential co-factor of prestin[9,10,11], but it remains unclear how Cl^- is held in prestin and contributes to motor function. Salicylate inhibits the motor activity of prestin[9], which accounts for the reversible hearing loss caused by aspirin overdosage[12]. Previous structures of bacterial SLC26[13], mouse SLC26A9[14], and human SLC26A9[15] have allowed homology modeling of prestin with high confidence, and in addition, very recent structures of prestin from human[16] and dolphin[17] have revealed its molecular structure and Cl^- induced conformational change, which advanced our understanding of the voltage-driven motor mechanism of prestin. However, many questions still remain regarding the anion-induced conformational changes.

Here, we present the structures of thermostabilized prestin in complex with Cl^-, SO_4^{2-}, and

Slika 25. V uvodu članka najdemo iskano informacijo o vplivu aspirina.

- ta članek nam pove o inhibitornem učinku aspirina na naš protein