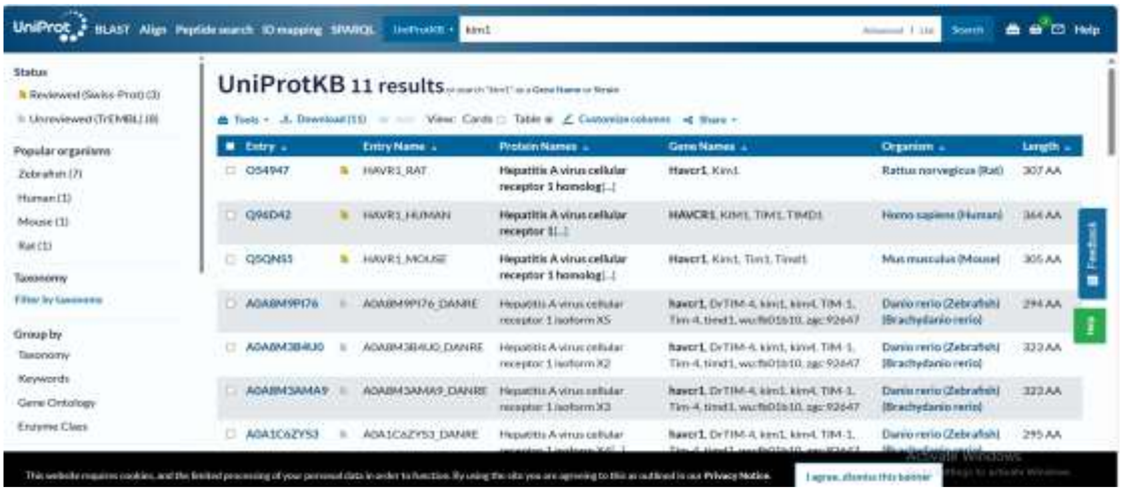
Rešitev naloge

1. DEL

Najprej v iskalno vrstico na UniProt vpišemo ime KIM-1. Gremo na prvi zadetek, ki je pri človeku, torej Homo sapiens. Njegov UniProt ID je Q96D42, priporočen naziv proteina pa je Hepatitis A virus cellular receptor 1. To najdemo pod zavihkom Names & Taxonomy, v razdelku Recommended name.
Pod Function preverimo interakcije z virusi - med njimi so:
Hepatitis A virus, Ebolavirus and Marburg virus, Dengue virus, Zika virus ter Chikungunya virus.

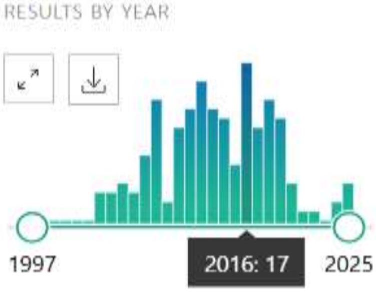


UniProtKB 11 results

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
O54947	HAVR1_RAT	**Hepatitis A virus cellular receptor 1 homolog**[...]	**Havcr1**, Kim1	Rattus norvegicus (Rat)	307 AA
Q96D42	HAVR1_HUMAN	**Hepatitis A virus cellular receptor 1**[...]	**HAVCR1**, KIM1, TIM1, TIMD1	Homo sapiens (Human)	364 AA
Q5QNS5	HAVR1_MOUSE	**Hepatitis A virus cellular receptor 1 homolog**[...]	**Havcr1**, Kim1, Tim1, Timd1	Mus musculus (Mouse)	305 AA
A0A8M9PI76	A0A8M9PI76_DANRE	Hepatitis A virus cellular receptor 1 isoform X5	**havcr1**, DrTIM-4, kim1, kim4, TIM-1, Tim-4, timd1, wu:fb01b10, zgc:92647	Danio rerio (Zebrafish) (Brachydanio rerio)	294 AA
A0A8M3B4U0	A0A8M3B4U0_DANRE	Hepatitis A virus cellular receptor 1 isoform X2	**havcr1**, DrTIM-4, kim1, kim4, TIM-1, Tim-4, timd1, wu:fb01b10, zgc:92647	Danio rerio (Zebrafish) (Brachydanio rerio)	323 AA
A0A8M3AMA9	A0A8M3AMA9_DANRE	Hepatitis A virus cellular receptor 1 isoform X3	**havcr1**, DrTIM-4, kim1, kim4, TIM-1, Tim-4, timd1, wu:fb01b10, zgc:92647	Danio rerio (Zebrafish) (Brachydanio rerio)	323 AA
A0A1C6ZY53	A0A1C6ZY53_DANRE	Hepatitis A virus cellular	**havcr1**, DrTIM-4, kim1, kim4, TIM-1,	Danio rerio (Zebrafish)	295 AA

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice.

i.) Gremo na PubMed in v iskalno vrstico vpišemo priporočeno ime: Hepatitis A virus cellular receptor 1. Nato pri filtrih označimo možnost Review. Na levi strani, pod Results by year, opazimo, da je največ člankov izšlo v letu 2016.



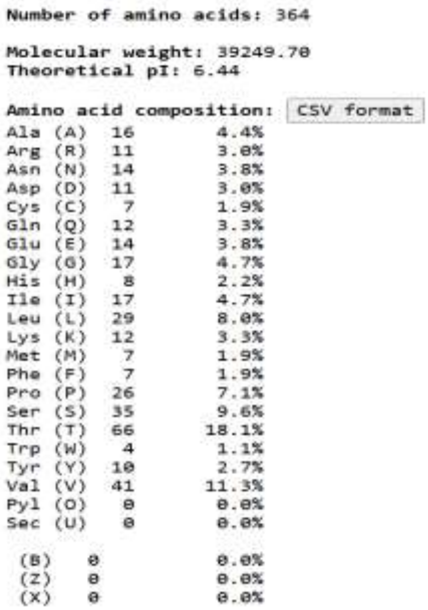
b.) Na UniProt, pod Function, piše da lahko zaznamo v urinu.

c.) Na UniProt, pod Disease & Variants, odgovor je na sliki prikazan.

+	Mutagenesis	338	About 50% loss of ubiquitination. 1 Publication
+	Mutagenesis	338	Complete loss of ubiquitination; when associated with R-346. 1 Publication
+	Mutagenesis	346	About 50% loss of ubiquitination. 1 Publication
+	Mutagenesis	346	Complete loss of ubiquitination; when associated with R-338. 1 Publication

d.) Protein je dolg 364 a.k. ostankov.

- i.) Na ProtParam vnesemo aminokislinsko zaporedje za ta protein. Opazimo, da se najpogosteje pojavljajo treonin (Thr, T), valin (Val, V) in serin (Ser, S).
- ii.) Izoelektricna tocka je 6.44, pri pH=8, je deprotoniran, to pomeni da uporabljamo anionski izmenjevalec.

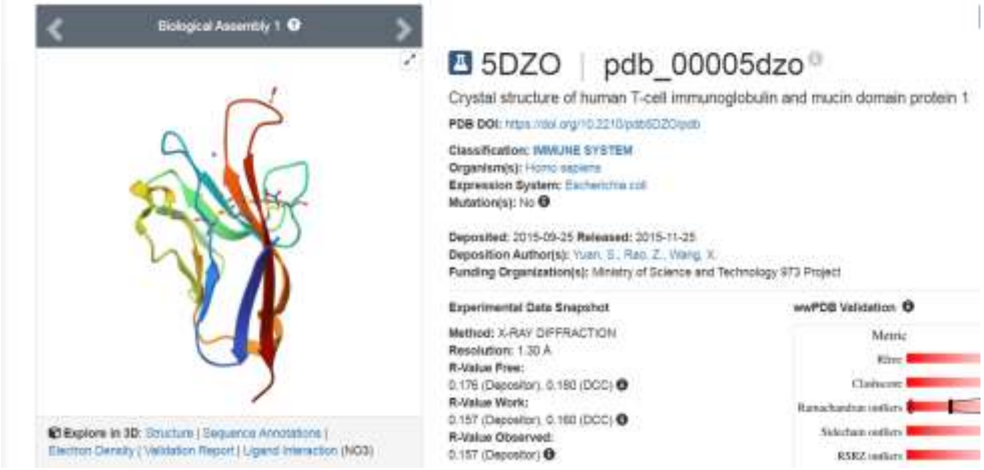


```
Number of amino acids: 364

Molecular weight: 39249.70
Theoretical pI: 6.44

Amino acid composition:  CSV format
Ala (A)  16       4.4%
Arg (R)  11       3.0%
Asn (N)  14       3.8%
Asp (D)  11       3.0%
Cys (C)   7       1.9%
Gln (Q)  12       3.3%
Glu (E)  14       3.8%
Gly (G)  17       4.7%
His (H)   8       2.2%
Ile (I)  17       4.7%
Leu (L)  29       8.0%
Lys (K)  12       3.3%
Met (M)   7       1.9%
Phe (F)   7       1.9%
Pro (P)  26       7.1%
Ser (S)  35       9.6%
Thr (T)  66      18.1%
Trp (W)   4       1.1%
Tyr (Y)  10       2.7%
Val (V)  41      11.3%
Pyl (O)   0       0.0%
Sec (U)   0       0.0%

 (B)   0          0.0%
 (Z)   0          0.0%
 (X)   0          0.0%
```

- iii.) Vstavimo aminokislinsko zaporedje v BLASTp in vidimo, da je najbolj podoben protein, razen pri človeku, najden pri vrsti Pan troglodytes, s kodo XP_016809598.2.

e.) Na UniProt pod Structure najdemo PDB kodo 5DZO.

- i.) Na PDB strani opazimo, da je bila za določitev strukture uporabljena metoda X-ray diffraction, in da je bil protein izražen v celicah Escherichia coli.
- ii.) Manjka propeptid.
- iii.) Opazimo ligande: nitratni in natrijev ion.



5DZO | pdb_00005dzo

Crystal structure of human T-cell immunoglobulin and mucin domain protein 1

PDB DOI: https://doi.org/10.2210/pdb5DZO/pdb

Classification: IMMUNE SYSTEM
Organism(s): Homo sapiens
Expression System: Escherichia coli
Mutation(s): No

Deposited: 2015-09-25 **Released:** 2015-11-25
Deposition Author(s): Yuan, S., Rao, Z., Wang, X.
Funding Organization(s): Ministry of Science and Technology 973 Project

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.30 Å
R-Value Free:
0.176 (Depositor), 0.180 (DCC)
R-Value Work:
0.157 (Depositor), 0.160 (DCC)
R-Value Observed:
0.157 (Depositor)

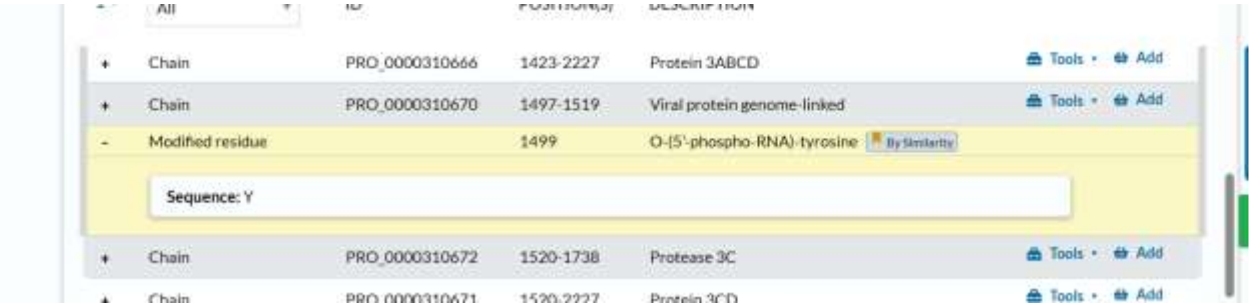
Explore in 3D: Structure | Sequence Annotations | Electron Density | Validation Report | Ligand Interaction (NO3)

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
NO3 Query on NO3	B [auth A], C [auth A]	**NITRATE ION** $N O_3$ NHNBFGGVMKEFGY-UHFFFAOYSA-N		
NA Query on NA	D [auth A]	**SODIUM ION** Na FKNQFGJONOIPTF-UHFFFAOYSA-N		

2. DEL

Na strani UniProt po taksonomiji poiščemo virus Hepatitis A in odpremo prvi zadetek s kodo A3FMB2.

a.) V razdelku PTM/Processing najdemo odgovor prikazan na sliki: Modificirana aminokislina je tirozin na poziciji 1499, označen kot: O-(5'-phospho-RNA)-tyrosine.



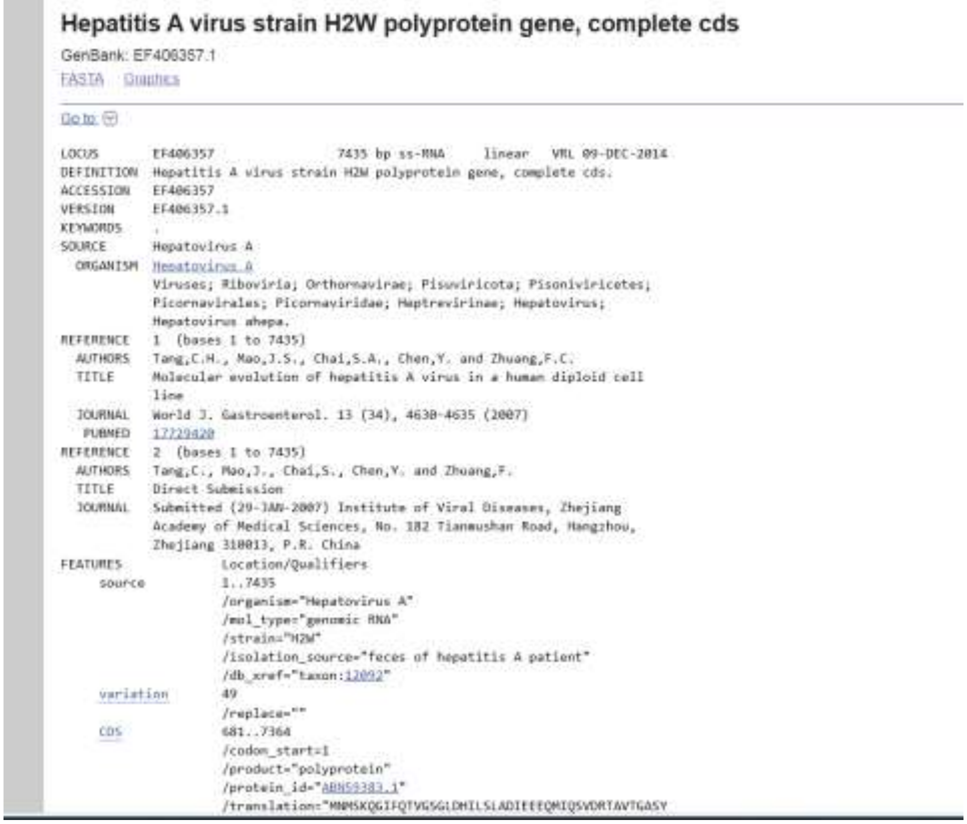
+	Chain	PRO_0000310666	1423-2227	Protein 3ABCD
+	Chain	PRO_0000310670	1497-1519	Viral protein genome-linked
-	Modified residue		1499	O-(5'-phospho-RNA)-tyrosine By Similarity
	Sequence: Y			
+	Chain	PRO_0000310672	1520-1738	Protease 3C
+	Chain	PRO_0000310671	1520-2227	Protein 3CD

b.) Ne bi uporabili bakterijske celice za izražanje, ker ne tvorijo disulfidne vezi, nač protein pa jih tvori, zato bi bilo izražanje napačno.

c.) Gremo na prvi link pri UniProt ki vodi do GenBank, kod je EF406357.1, dolzina kodirajoča regija znaša 7364-681+1=6684 a.k. ostankov.

i.) genomska ssRNA, gre za RNA virus

ii.) motiv atttctccccg se nahaja na mestih 5131-5140 in to lahko ga hitro najdemo s funkcijo Ctrl + F tako, da zaporedje vpišemo v iskalno vrstico.

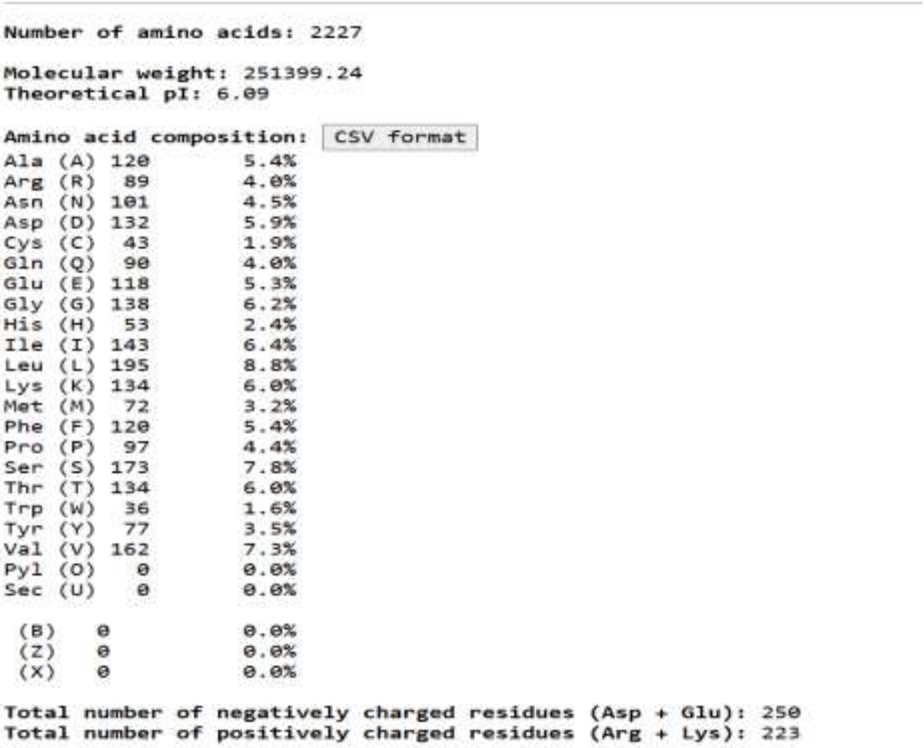


Hepatitis A virus strain H2W polyprotein gene, complete cds

GenBank: EF406357.1

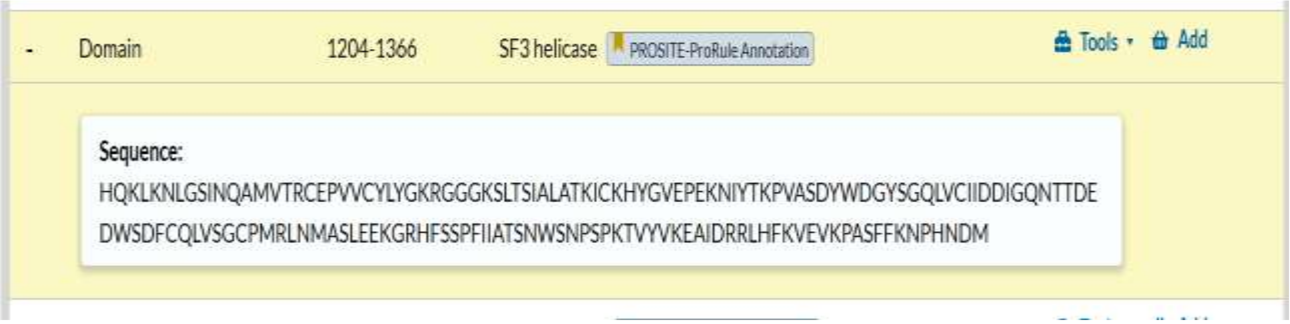
```
LOCUS       EF406357                7435 bp ss-RNA     linear   VRL 09-DEC-2014
DEFINITION  Hepatitis A virus strain H2W polyprotein gene, complete cds.
ACCESSION   EF406357
VERSION     EF406357.1
KEYWORDS    .
SOURCE      Hepatovirus A
  ORGANISM  Hepatovirus A
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Picornavirales; Picornaviridae; Heptrevirinae; Hepatovirus;
            Hepatovirus ahepa.
REFERENCE   1  (bases 1 to 7435)
  AUTHORS   Tang,C.H., Mao,J.S., Chai,S.A., Chen,Y. and Zhuang,F.C.
  TITLE     Molecular evolution of hepatitis A virus in a human diploid cell
            line
  JOURNAL   World J. Gastroenterol. 13 (34), 4630-4635 (2007)
   PUBMED   17729420
REFERENCE   2  (bases 1 to 7435)
  AUTHORS   Tang,C., Mao,J., Chai,S., Chen,Y. and Zhuang,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2007) Institute of Viral Diseases, Zhejiang
            Academy of Medical Sciences, No. 182 Tianmushan Road, Hangzhou,
            Zhejiang 310013, P.R. China
FEATURES             Location/Qualifiers
     source          1..7435
                     /organism="Hepatovirus A"
                     /mol_type="genomic RNA"
                     /strain="H2W"
                     /isolation_source="feces of hepatitis A patient"
                     /db_xref="taxon:12092"
     variation       49
                     /replace=""
     CDS             681..7364
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="ABN59383.1"
                     /translation="MNMSKQGIFQTVGSGLDHILSLADIEEEQMIQSVDRTAVTGASY
```

d.) Gremo na ProtParam, delež zrecunamo za negativno nabite a.k. ostankov kot 250/2227 in za pozitivne kot 223/2227.



```
Number of amino acids: 2227

Molecular weight: 251399.24
Theoretical pI: 6.09

Amino acid composition:  CSV format
Ala (A) 120        5.4%
Arg (R)  89        4.0%
Asn (N) 101        4.5%
Asp (D) 132        5.9%
Cys (C)  43        1.9%
Gln (Q)  90        4.0%
Glu (E) 118        5.3%
Gly (G) 138        6.2%
His (H)  53        2.4%
Ile (I) 143        6.4%
Leu (L) 195        8.8%
Lys (K) 134        6.0%
Met (M)  72        3.2%
Phe (F) 120        5.4%
Pro (P)  97        4.4%
Ser (S) 173        7.8%
Thr (T) 134        6.0%
Trp (W)  36        1.6%
Tyr (Y)  77        3.5%
Val (V) 162        7.3%
Pyl (O)   0        0.0%
Sec (U)   0        0.0%

 (B)   0           0.0%
 (Z)   0           0.0%
 (X)   0           0.0%

Total number of negatively charged residues (Asp + Glu): 250
Total number of positively charged residues (Arg + Lys): 223
```

e.) Od 1204 do 1366 sega SF3 helicase, in to najdemo na UniProt, pod Family & Domains.



-	Domain	1204-1366	SF3 helicase PROSITE-ProRule Annotation

Sequence:

HQKLKNLGSINQAMVTRCEPVVCYLYGKRGGGKSLTSIALATKICKHYGVEPEKNIYTKPVASDYWDGYSGQLVCIIDDIGQNTTDE
DWSDFCQLVSGCPMRLNMASLEEKGRHFSSPFIIATSNWSNPSPKTVYVKEAIDRRLHFKVEVKPASFFKNPHNDM