V laboratoriju smo med čiščenjem zamrzovalnika naleteli na epruveto z označbo “hrček – vranica”, ki jo je očitno pustil nekdanji sodelavec. Izkazalo se je, da vsebuje fragment proteina zlatega hrčka, ki ga je preučeval v okviru starega projekta. Ker dokumentacije ni bilo veliko, smo se odločili, da pogledamo, za kaj pravzaprav gre – in kmalu ugotovili, da je fragment dovolj zanimiv, da si zasluži nadaljnjo analizo. To zaporedje je bilo:

SEFSGDDSFTIVNENTGKCIQPLSDWIVAQDCSETRSMLWKWVSQHRLFHLESQKCLGLDMTKAADNL
RMFRCDSSVLLWWKCEHHSLYSAAHYRLDLKDGYATASTNSSAVWKKGGSKENLCDQPYREIYTRDGN
SYGRPCEFPFLVGETWHHDCIRDENHSGPWCATTLNYEYDQKWGICLKPESGCEGNWEKNEQIGSCY
QFNNQEVLSWKEAYVSCQNQGADLLSIHSAAELAYITGKEDIARIVWIGLNQLYSARGWEWSDFKPLKFL
NWDPGTPSAPMIGGSSCARMDTETGLWRSVSCEAQQPYVCKKPLNNTVELPDVWTYSDTHCDVGWLP
QNGFCYLLANESGPWDAAHLKCKAFGGDLISIHSLADVEVVVTKLHSGDVKEEIWTGLRNVNSPTLFQW
SDGTEVTLTYWNENEPSVPYNKTPNCVSYLGKLGQWKVQSCEKKLRYVCKKKGEITNDTRSDKLCPPDE
GWKRHGETCYKIYENEVPFGTNCNLTITSRFEQEFLNDMMKKYDKSFQKYFWTGLRDADARGEYSWAAT
GGLKQAMTFSNWNFLQPASPGGCVAMSTGKTLGRWEVKSCRSFRALSICKKMSGPQEPEEATPKPDEP
CPEGWHTFPSNLSCYKVFHIERTVRRRTWEEAERFCQALGAHLPSFSHMNEVKEFLHLLQDQFSVQRW
LWIGLNKRSPDLQGSWQWSDRTPVSTVIMHREFQQDYDVRDCAAIKVLDNAWLRTWYYYDERKFGYLK
PFSCDAKLDWVCQIPKGSTLQVPDWYNPERTGIHGPPVIIDGSEYWFVEEPRLNYEEAVLYCASNHSFLA
TITTFTKLKAIRGKMENLSGEEQKWWVKANANPIDHYFLRTRPLWHRFSMLLDEECLQMSAKMWHLDL
NKRADCNDKLPFVCEKYNVSSLEKYSPDSAAKVQCTGKWIPFQNKCFLKVKSEPVTFSQASSTCHTYGG
TLPSVLSKSEQDFIISLLPEMETSLWIGLRWTAYDRISKWTDGRNLTYSNFHPLLVGRRLSIAAYFIDEESHY
HCALMLNLRKSPLTGTWNFTSCSERHSLSLCQKYSENEDGRPWETNSETVKYLNNLYKIISKPLTWHGAL
KECLNENMRLVSITDPYQQAFLSVQATLRNTSFWIGLSSQDDELNFGWSDGTYLHFSNWAVDNEKLDD
CVILDTDGFWKTADCDENQPGAICYYPGNETSKEVRPLNSAKCPSPAQSTPWVPFQNSCYNFMITKNRH
RTITQKEVHSLCQKLHSKAQILSIRNEEENNFVVEQLLYFNYIASWVMLGVTYENNSLMWFDKTALSYTH
WRAGRPAVKNHKFLAGLSTDGFWDIQSFNVIDETLHFYQHSILACKIEMVDYKEERNSTLPEFIPYEDGVY
NVIQKRVTWYQALSMCSQSGRHLASVHNPKEQLFLEDIVNRDGFPLWVGLSSHDGSESSFEWSDGSA
FDYIPWKSQGSPGNCVILDPKGTWKHENCLSVKDGAICYKPTKFKELASHAHSSKCPLVKRNGSQWVQ
YGDHCYSAEQALHTFAEAKKLCQELDHSATVVTIADENENKFVSRLMRENYNITMRVWLGLSQHSLDQS
WSWLDGLDVTFVKWENKSKNGDGKCSILIASNETWKKVECSRGYARVVCKVPLSPD

1) Za kateri protein najverjetneje gre. Navedi ime proteina.

Limfocit antigen 75

Job Title	Protein Sequence
RID	1VXN636Y016 *Search expires on 05-10 20:35 pm* Download All
Program	BLASTP Citation
Database	nr See details
Query ID	lcl|Query_12009323
Description	unnamed protein product
Molecule type	amino acid
Query Length	1639
Other reports	Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to | E value to | Query Coverage to

Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all *100 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
lymphocyte antigen 75 precursor [Mesocricetus auratus]	Mesocricetus auratus	3426	3426	100%	0.0	100.00%	1722	NP_001268813.1
lymphocyte antigen 75 isoform X2 [Mesocricetus auratus]	Mesocricetus auratus	3424	3424	100%	0.0	99.94%	1870	XP_040600776.1
lymphocyte antigen 75 isoform X1 [Mesocricetus auratus]	Mesocricetus auratus	3419	3419	100%	0.0	100.00%	1664	XP_040600775.1
lymphocyte antigen 75-like isoform 1 [Cricetulus griseus]	Cricetulus griseus	3261	3261	100%	0.0	94.45%	1722	ERE70064.1

2) Poišči, če obstaja tudi ta protein pri človeku. Išči po bazi eksperimentalno določenih struktur. Izmed proteinov z najvišjo vrednostjo Query Cover, izberi protein, ki je bil določen pri večji ločljivosti.

Gre za protein s PDB kodo 7JPT.

Your search is limited to records that include: human (taxid:9606)

Job Title	Protein Sequence
RID	1VXP5VB6013 *Search expires on 05-10 20:36 pm* Download All
Program	BLASTP Citation
Database	pdb See details
Query ID	lcl|Query_8548562
Description	unnamed protein product
Molecule type	amino acid
Query Length	1639
Other reports	Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to | E value to | Query Coverage to

Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all *81 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
Chain A, Lymphocyte antigen 75 [Homo sapiens]	Homo sapiens	2709	2709	100%	0.0	76.89%	1722	7JPU_A
Chain A, Lymphocyte antigen 75 [Homo sapiens]	Homo sapiens	2707	2707	100%	0.0	76.89%	1693	7JPT_A
Chain A, Lymphocyte antigen 75 [Homo sapiens]	Homo sapiens	1034	1275	99%	0.0	80.23%	603	8HBC_A
Chain A, Lymphocyte antigen 75 [Homo sapiens]	Homo sapiens	1033	1275	99%	0.0	80.37%	596	8K8H_A

Gledamo samo prva dva proteina in preko PDB baze glede na s pomočjo PDB kode določimo, kateri izmed dveh je bila določen pri višji ločljivosti.

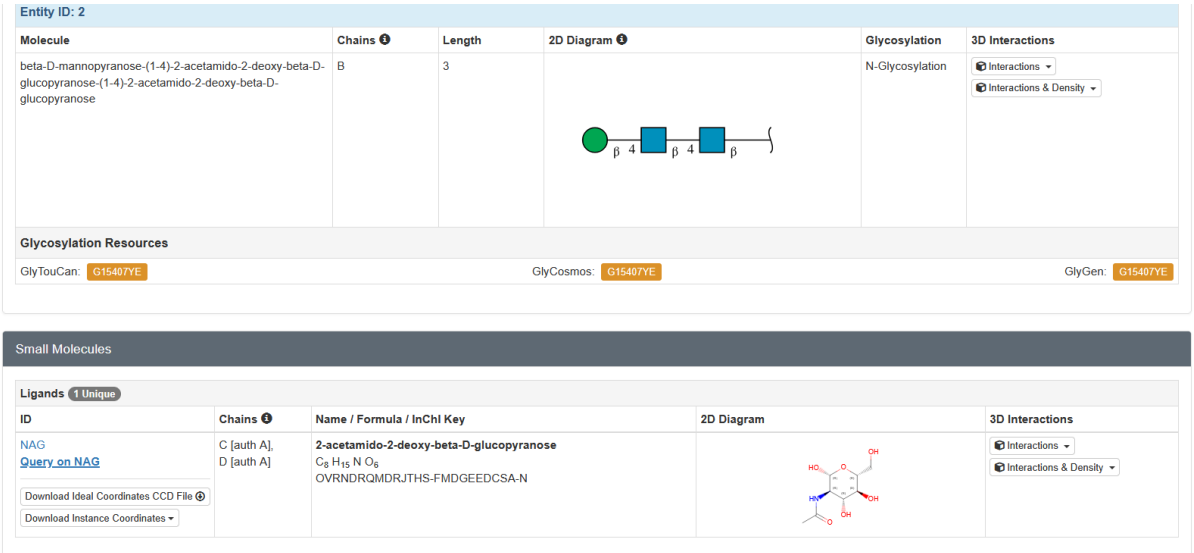
3) Kje je največja regija, ki je znanstveniki niso mogli zmodelirati (N-konec, v sredini, C-konec) Napiši, od katerega do katerega aminokislinskega ostanka ta regija sega.

Na C-koncu, gre za ak ostanke 1355-1693.

Vrstica unmodeled.

4) Kateri ligandi so kovalentno vezani v naši PDB strukturi? Preko katerih aminokislinskih ostankov so vezani (glede na številčenje vUSCF Chimera/Molstar)? Ali so vezani ligandi pričakovani glede na domene v proteinu?

NAG (2-acetamido-2-deoxy-beta-D-glucopyranose) in BMA preko dveh NAG (beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose). Na protein so vezani preko asparaginov 1076 in 1253. Da, ker protein sestavljajo lektinske domene, za katere je znano, da vežejo sladkorje.



Entity ID: 2

Molecule	Chains	Length	2D Diagram	Glycosylation	3D Interactions
beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose	B	3	β 4 β 4 β	N-Glycosylation	Interactions Interactions & Density

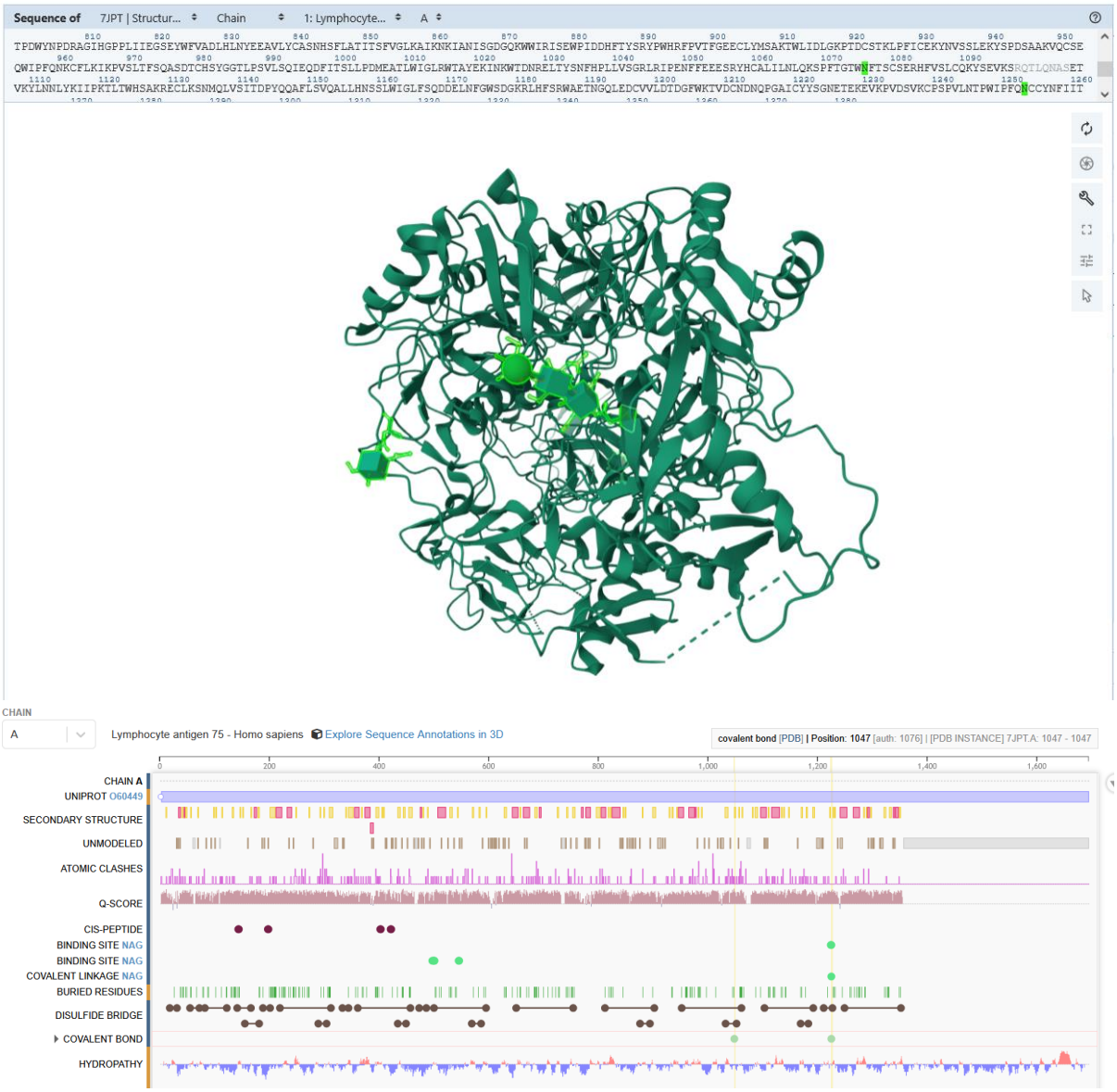
Glycosylation Resources

GlyTouCan: G15407YE GlyCosmos: G15407YE GlyGen: G15407YE

Small Molecules

Ligands 1 Unique

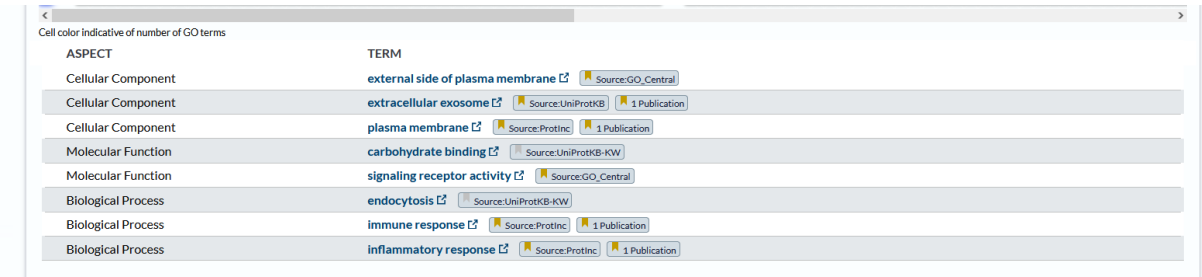
ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
NAG Query on NAG Download Ideal Coordinates CCD File Download Instance Coordinates	C [auth A], D [auth A]	**2-acetamido-2-deoxy-beta-D-glucopyranose** $C_8H_{15}NO_6$ OVRNDRQMDRJTHS-FMDGEEDCSA-N		Interactions Interactions & Density



Ligande lahko določimo iz Molstar ali pod kategorijo Structure summaries v PDB. Za določitev mest vezave lahko to neposredno razberemo iz zaporedja nad strukturo v Molstar ali pa pod kategorijo Sequence>Covalent bonds.

5) Pri katerih procesih v celici sodeluje iskani protein?

Endocytosis, immune response, inflammatory response.



Cell color indicative of number of GO terms

ASPECT	TERM
Cellular Component	external side of plasma membrane [Source:GO_Central]
Cellular Component	extracellular exosome [Source:UniProtKB] [1 Publication]
Cellular Component	plasma membrane [Source:ProtInc] [1 Publication]
Molecular Function	carbohydrate binding [Source:UniProtKB-KW]
Molecular Function	signaling receptor activity [Source:GO_Central]
Biological Process	endocytosis [Source:UniProtKB-KW]
Biological Process	immune response [Source:ProtInc] [1 Publication]
Biological Process	inflammatory response [Source:ProtInc] [1 Publication]

UniProt pod kategorijo Function oz. na GO annotations.

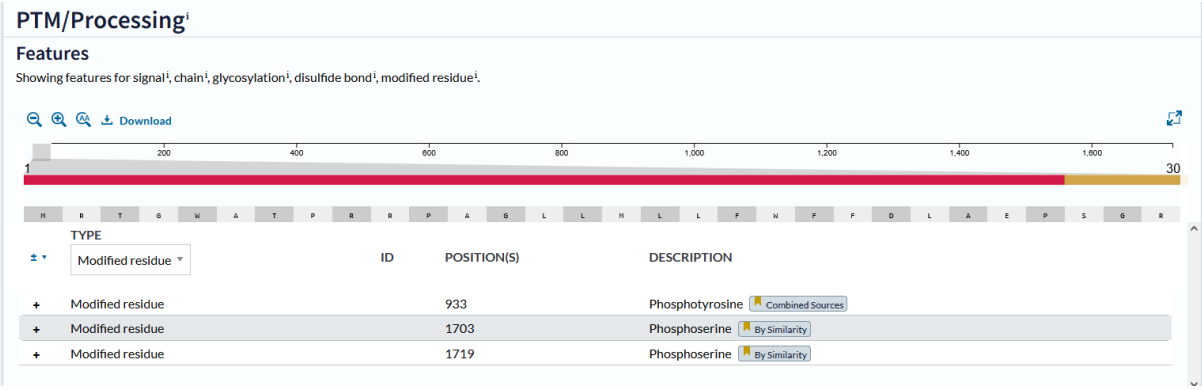
6) Kje v proteomu je zapisan gen za protein?

Na kromosomu 2.

UniProt kategorija Names and Taxonomy ali v PDB v kategoriji Genome.

7) Protein ima modificirane ostanke. Kateri so modificirani ostanki in kje se nahajajo?

933 fosfotirozin, 1703 fosfoserin, 1719 fosfoserin.



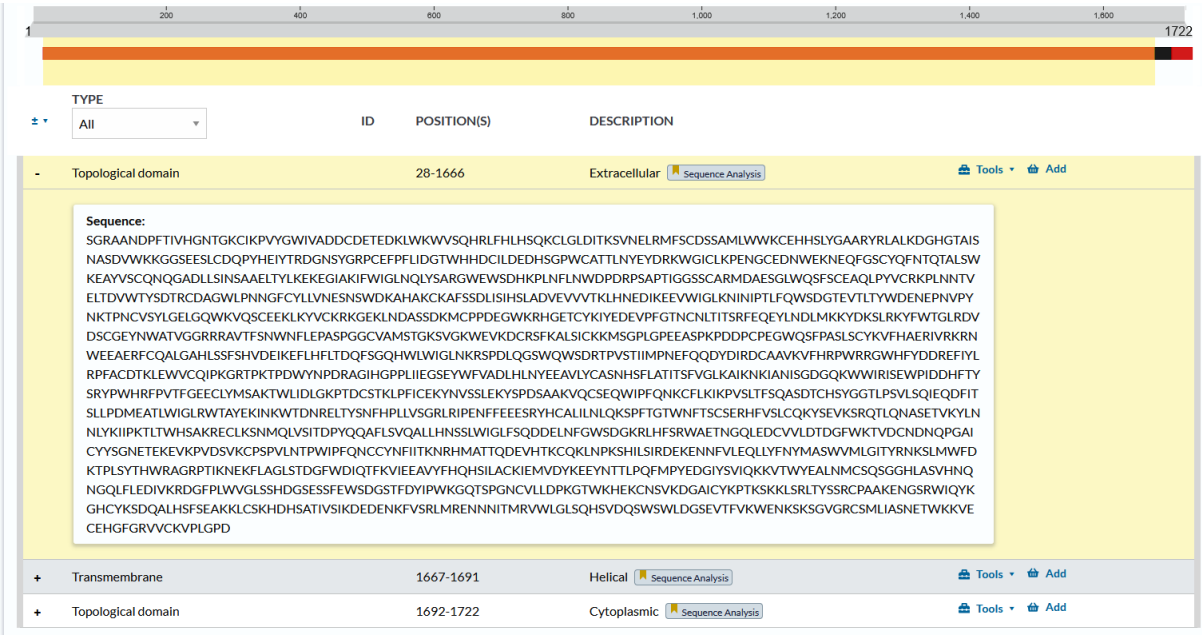
V UniProt v kategoriji PTM/Processing.

8) Koliko izoformnih produktov lahko dobimo z alternativnim izrezovanjem.

5

V UniProt pod kategorijo Sequence and Isoforms.

9) Zanima nas v vodi topen del proteina, ki predstavlja večji delež celotnega proteina. Kakšna je razlika v ekstinkcijskem koeficientu med tem delom proteina, če so vsi cisteini oksidirani ali vsi cisteini reducirani.




```
Extinction coefficients:
Extinction coefficients are in units of  M-1 cm-1, at 280 nm measured in water.
Ext. coefficient   484360
Abs 0.1% (=1 g/l)   2.565, assuming all pairs of Cys residues form cystines


Ext. coefficient   480360
Abs 0.1% (=1 g/l)   2.544, assuming all Cys residues are reduced
```

Izberemo ustrezen del zaporedja v UniProt, ki ga izberemo pod Subcellular location. To zaporedje na to prenesemo v ProtParam.

484360- 480360 = 4000; razlika je 4000.

10) Ker imamo nečist vzorec smo se ga odločili izolirati. Kakšen ionski izmenjevalec bomo uporabili, če želimo izolirati celoten protein pri pH 6.8?

Potrebovali bomo anionski izmenjevalec.

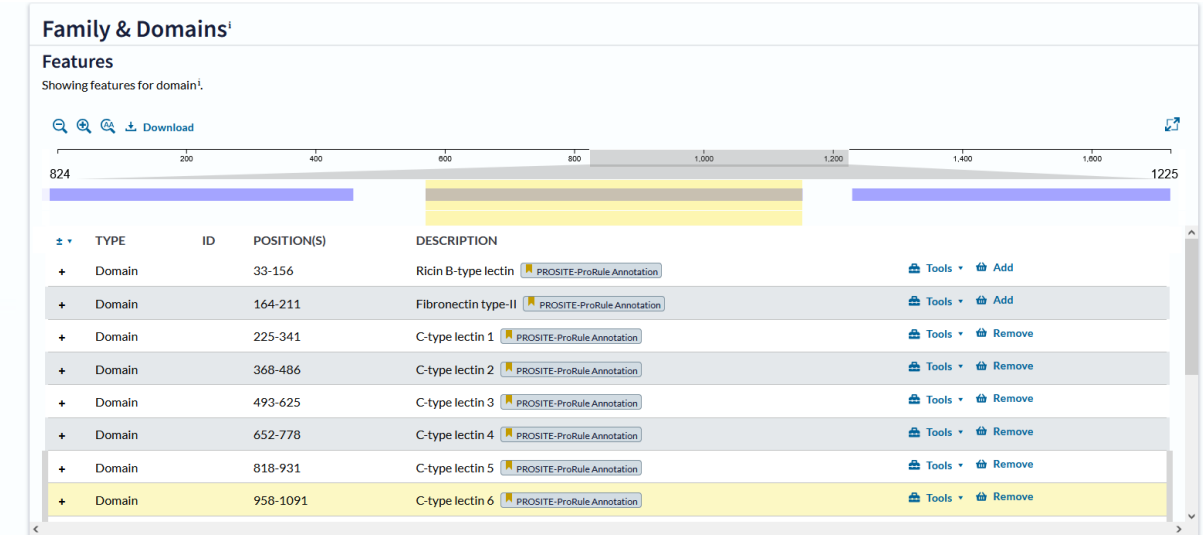
```
Number of amino acids: 1639

Molecular weight: 188821.09
Theoretical pI: 6.18
```

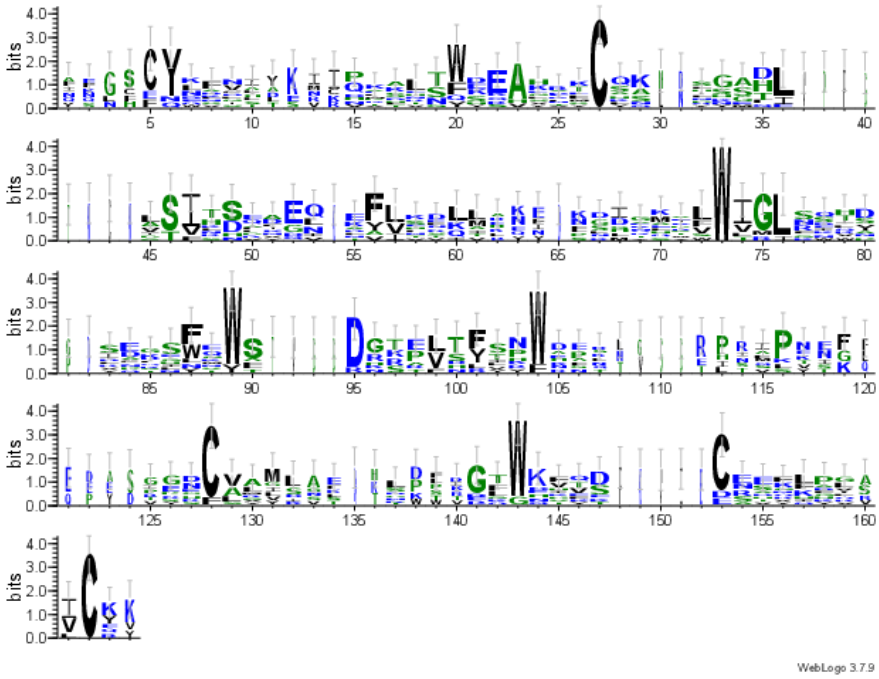
V ProtParam prenesemo zaporedje celotnega proteina in dobimo izračunan teoretičen pI, na podlagi katerega določimo naboj proteina pri pH 6.8 in tako tudi ustrezen izmenjevalec.

11) V našem proteinu imamo več ponavljajočih domen. Zanima nas ohranjenost domen, zato poglejte, kateri aminokislinski ostanki so v teh domenah najbolj ohranjeni. Odgovor

podajte kot ime in tričrkovna oznaka aminokislin in ne mesto v zaporedju. Kot odgovor dodaj še WebLogo sliko poravnav in iz nje poišči odgovor.

Ohranjeni so cisteini (CYS) in triptofani (TRP).

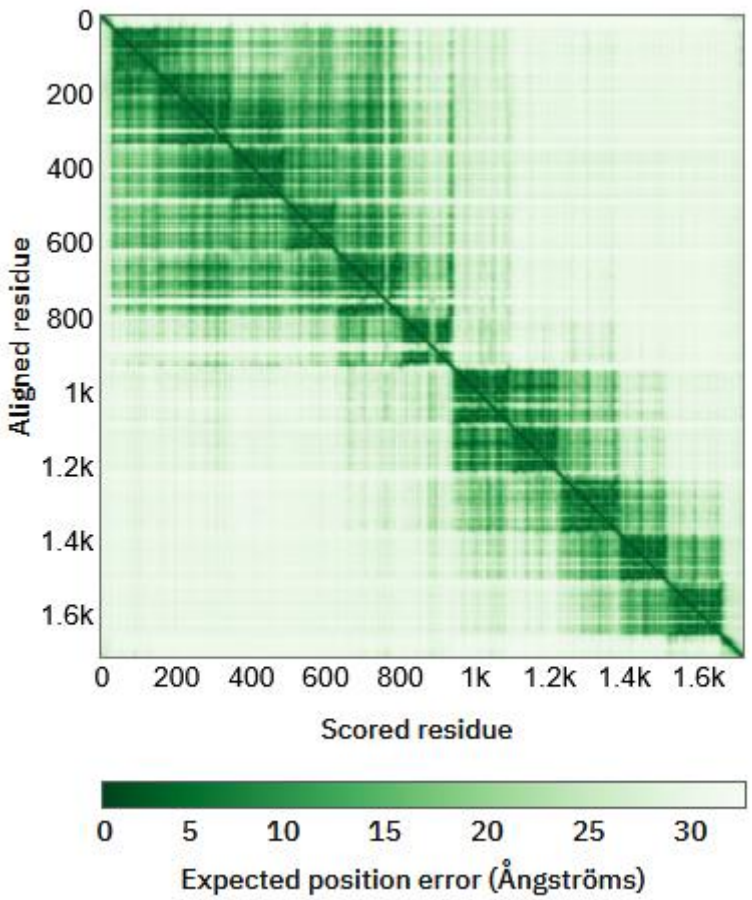




V UniProt pod kategorijo Family and Domains najdemo ustrezne domene, jih dodamo košarico in izvedemo poravnavo. Potem poravnave uvozimo v WebLogo in naredimo sliko.

12) Ali podatki iz baze UniProt (kot so funkcijske domene, transmembranske regije, neurejene regije) potrjujejo napovedane strukturne značilnosti, prikazane v PAE grafu iz AlphaFold?

Na grafu PAE lahko od diagonali vidimo 11 kvadratov, ki so temnejše zelene barve, kar se sklada s podatki v UniProt, saj protein sestavlja 11 domen, ki imajo visoko zanesljivost napovedi glede na ostale ostanke v domeni, domene pa povezujejo linkerjji, ki niso tako zelo urejeni.



13) Preko UniProt dostopaj do zaporedja za mRNA našega proteina, v primeru, da imaš na voljo več referenčnih sekvenc, izberi prvo. Katero signalno regijo lahko najdemo v zapisu, poleg signalnega peptida. Kakšno je njeno zaporedje?

PoliA signalna sekvenca. Zaporedje AATAAA.

```
                     /inference="alignment:Splign:2.1.0"
regulatory           8671..8676
                     /regulatory_class="polyA_signal_sequence"
                     /gene="LY75-CD302"
                     /gene_synonym="CD205; CLEC13B; DEC-205; gp200-MR6; Ly-75;
                     LY75"
                     /note="hexamer: AATAAA"
polyA_site           8696
                     /gene="LY75-CD302"
                     /gene_synonym="CD205; CLEC13B; DEC-205; gp200-MR6; Ly-75;
                     LY75"
                     /note="major polyA site"
regulatory           8899..8904
                     /regulatory_class="polyA_signal_sequence"
                     /gene="LY75-CD302"
                     /gene_synonym="CD205; CLEC13B; DEC-205; gp200-MR6; Ly-75;
                     LY75"
                     /note="hexamer: AATAAA"
polyA_site           8919
                     /gene="LY75-CD302"
                     /gene_synonym="CD205; CLEC13B; DEC-205; gp200-MR6; Ly-75;
                     LY75"
                     /note="major polyA site"
```

V UniProt pod kategorijo Sequence and Isogorms izbereš prvo referenčno sekvenco, nato pa odgovor iščeš v GenBank.