

S1-02

DOKUMENT Z REŠITVAMI NALOG

1. NALOGA – UGANKA

- a) Za rešitev preprosto uporabimo orodje Blastp, v katerega prilepimo dano aminokislinsko zaporedje in iščemo po *Non-redundant protein sequences (nr)*. Kot rezultat iskanja dobimo tako tabelo:

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
cytochrome b [Okapia johnstoni]	Okapia johnstoni	762	762	100%	0.0	100.00%	379	YP_007626458.1
cytochrome b [Okapia johnstoni]	Okapia johnstoni	758	758	100%	0.0	99.47%	379	ABV29382.1
cytochrome b [Muntiacus vuquangensis]	Muntiacus vuqu...	728	728	100%	0.0	95.25%	380	YP_005296078.1
cytochrome b [Moschus berezovskii]	Moschus berez...	705	705	100%	0.0	93.67%	379	AUL75724.1
cytochrome b [Moschus berezovskii bjiangensis]	Moschus berez...	704	704	100%	0.0	93.40%	379	AAB94606.1
cytochrome b [Muntiacus sp. LOM2]	Muntiacus sp. L...	703	703	100%	0.0	94.20%	379	AQT00095.1
RecName: Full=Cytochrome b; AltName: Full=Complex III subunit 3; AltName: Full=Complex III subunit...	Moschus leuco...	703	703	100%	0.0	94.20%	379	O47584.1
cytochrome b [Muntiacus atherodes]	Muntiacus athe...	703	703	100%	0.0	94.72%	379	ARH02767.1
RecName: Full=Cytochrome b; AltName: Full=Complex III subunit 3; AltName: Full=Complex III subunit...	Moschus fuscus	702	702	100%	0.0	93.93%	379	O48309.1
cytochrome b [Moschus berezovskii]	Moschus berez...	702	702	100%	0.0	93.14%	379	XXL74818.1
cytochrome b [Muntiacus sp. JAV10]	Muntiacus sp. J...	702	702	100%	0.0	94.20%	379	AQT00435.1
cytochrome b [Dama dama]	Dama dama	701	701	100%	0.0	93.93%	379	YP_007626068.1
cytochrome b [Muntiacus sp. THA1]	Muntiacus sp. T...	701	701	100%	0.0	93.93%	379	AQS98857.1
cytochrome b [Moschus leucogaster]	Moschus leuco...	701	701	100%	0.0	93.67%	379	YP_009847206.1

Iz nje je jasno vidno 100 % ujemanje in prav tako identičnost, E-vrednost je neskončno majhna, kar pomeni, da lahko z zagotovostjo trdimo, da gre za protein citokrom b iz vrste *Okapia johnstoni*.

- b) Rešitev za to bomo poiskali v Uniprot-u (<https://www.uniprot.org/uniprotkb/Q6YLN1/entry>). V razdelku Names & Taxonomy najdemo podatek, da je protein kodiran v mitohondrijskem genomu (kar seveda glede na njegovo nalogo in mesto nahajanja v celici niti ni tako presenetljivo – oboje lahko najdemo v Uniprot-u pod razdelkoma *Function* in *Subcellular location*).

Odgovor na vprašanje torej ni pritrilen.

The screenshot shows the UniProt interface for protein Q6YLN1. The 'Names & Taxonomy' section is expanded, showing 'Protein names' (Recommended name: Cytochrome b, Alternative names: Complex III subunit 3, etc.) and 'Gene names' (Name: MT-CYB, Synonyms: COB, CYTB, MTCYB). The 'Encoded in' field is highlighted with a red box and contains the text 'Mitochondrion'.

- c) 42,8 kDa je navidezna molekulska masa tega proteina. To lahko najdemo pod razdelkom *Sequence*.

- d) Glede na podatke iz spleta, okapiji živijo v zelo omejenem območju v Demokratični republiki Kongo. So primer endemita za to območje, kar pomeni, da jih najdemo samo tam in nikjer drugje na svetu (v Sloveniji je primer take endemične vrste za ozko območje Bele krajine npr. *Proteus anguinus parkelj*).

Povezava do spletne enciklopedije Britannica:
<https://www.britannica.com/animal/okapi>

- e) Navkljub podobnemu imenu imata proteina drugačno funkcijo in si med seboj nista tako sorodna (homologna), kot bi lahko na prvo žogo zmotno mislili iz podobnosti v imenu. O funkciji si preberemo v razdelku *Function* v zbirki UniProt. Povezavi do obeh: <https://www.uniprot.org/uniprotkb/P00156/entry#sequences> (B) in <https://www.uniprot.org/uniprotkb/P99999/entry#sequences> (C).

Že tukaj dvomimo o smiselnosti globalne poravnave, dodaten dvom pa se pojavi pri razliki v dolžini aminokislinskega zaporedja (B: 380 AKO; C: 105 AKO). Zato je smiselna kvečjemu priprava lokalne poravnave. Dobimo sledeče podatke:

```
# Aligned_sequences: 2
# 1: c_b
# 2: c_c
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 10
# Identity:      6/10 (60.0%)
# Similarity:   7/10 (70.0%)
# Gaps:         0/10 ( 0.0%)
# Score: 26.0
#
#
#=====
c_b          195 LHLFLHETG    204
              |||...||
c_c          33 LHGLFGRKTG   42
```

Po izvedbi vidimo, da je poravnanih zgolj 10 aminokislinskih ostankov (res kratka regija), z identičnostjo 6 AKO, še eden pa je podoben. Vrednost poravnave je nizka, kar skupaj nakazuje na to, da sta proteina evolucijsko po sorodnosti precej oddaljena. Sam rezultat na neko podobno domeno v proteinih ne nakazuje, do ujemanja bi lahko prišlo tudi naključno. To vse skupaj potrjuje, da proteina med sabo nista evolucijsko bližnje sorodna.

2. NALOGA

- a) Priprava poravnave v N-ESPript pokaže, da smo pridobili vzorce dveh različnih oseb, in sicer sta vzorca 1 in 3 od mrtvega poglavarja, vzorec 2 pa tako pripada domnevemu storilcu. To sklepamo iz primerjave aminokislinskih ostankov, ki vidimo da se razlikuje samo pri vzorcu 2, IN SICER Z ZAGOTOVOSTJO NA 7. KODONU. Iz tega sklepamo, da je z raziskavo smiselno nadaljevati.

Seq1	ATGGTGCATCTGACACCCGAGGAGAAGTCCTCCTTACTGGCCTGTGGGC	AAGGTAA
Seq2	ATGGTNCAYYNACCCNGTGGAGAAGTCYCYGTNACNGCNCTNTGGGC	AAGGTAA
Seq3	ATGGTGCATCTGACACCCGAGGAGAAGTCCTCCTTACTGGCCTGTGGGC	AAGGTAA

	70	80	90	100	110	120
Seq1	CTGGATGAAGT	TGGTGGT	GAGGCCCTGGGCAG	CTGCT	GTTGGTCTACCCCT	TGGACCCAG
Seq2	CTGGATGAAG	Y Y GG Y GG Y	GAGGCCCTGGGCAG	NCTGCT	NTGGTCTACCCCT	TGGACCCAG
Seq3	CTGGATGAAG	TGGTGGT	GAGGCCCTGGGCAG	CTGCT	GTTGGTCTACCCCT	TGGACCCAG

	130	140	150	160	170	180
Seq1	AGGTCTTTG	AGTCTT	GGGGATCTG	CCAC	CTTGA	TGCT
Seq2	AGGTCTTT	NGAN	TTCN	TTGGGGATCTG	NNNAC	NCNNGAN
Seq3	AGGTCTTTG	AGTCTT	GGGGATCTG	CCAC	CTTGA	TGCT

	190	200	210	220	230	240
Seq1	CTGAAGGCTCAT	TGGCAAGAAAGT	GCTCGGTCGCTTTA	GTGATGGCCTGGCT	CACTTGGAC	
Seq2	CTGAAGGCTCAT	TGGCAAGAAAGT	GCTCGGTCGCTTTA	GTGATGGCCTGGCT	CACTTGGAC	
Seq3	CTGAAGGCTCAT	TGGCAAGAAAGT	GCTCGGTCGCTTTA	GTGATGGCCTGGCT	CACTTGGAC	

	250	260	270	280	290	300
Seq1	AACCTCAAGGGCACCT	TTGGCCCACTGAGT	GAGCTGCAAGCT	CA	GTGGAT	
Seq2	AACCTCAAGGGCACCT	TTGGCCCACTGAGT	GAGCTGCAAGCT	CA	GTGGAT	
Seq3	AACCTCAAGGGCACCT	TTGGCCCACTGAGT	GAGCTGCAAGCT	CA	GTGGAT	

	310	320	330	340	350	360
Seq1	CCTGAGA	ACTTCAGGCT	CTCGGCAACGTG	CTG	GCTCTG	TGTGCTGGCCCATCACTTTGGC
Seq2	CCTGAGA	ACTTCAGGCT	CTCGGCAACGTG	CTG	GCTCTG	TGTGCTGGCCCATCACTTTGGC
Seq3	CCTGAGA	ACTTCAGGCT	CTCGGCAACGTG	CTG	GCTCTG	TGTGCTGGCCCATCACTTTGGC

	370	380	390	400	410	420
Seq1	AAAGAA	ATCACCCACCAGTGCA	CGC	CGCCTATCAGAAAGTGGT	GCTGGTGT	GCCTAA
Seq2	AAAGAA	ATCACCCACCAGTGCA	CGC	CGCCTATCAGAAAGTGGT	GCTGGTGT	GCCTAA
Seq3	AAAGAA	ATCACCCACCAGTGCA	CGC	CGCCTATCAGAAAGTGGT	GCTGGTGT	GCCTAA

	430	440
Seq1	GCCCTGGCCCA	CAAGTATCACTAA
Seq2	GCCCTGGCCCA	CAAGTATCACTAA
Seq3	GCCCTGGCCCA	CAAGTATCACTAA

- b) Za prvo zaporedje dobimo:

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens hemoglobin subunit beta (HBB), mRNA	Homo sapiens	821	821	100%	0.0	100.00%	628	NM_000518.5
<input checked="" type="checkbox"/> TPA_inf. Homo sapiens GLNA1 gene for globin A1	Homo sapiens	821	821	100%	0.0	100.00%	444	LT548086.1
<input checked="" type="checkbox"/> Homo sapiens cDNA FLJ92086. Homo sapiens hemoglobin_beta (HBB), mRNA	Homo sapiens	821	821	100%	0.0	100.00%	494	AK311825.1
<input checked="" type="checkbox"/> PREDICTED: Pan. paniscus hemoglobin subunit beta (LOC100976465), mRNA	Pan. paniscus	815	815	100%	0.0	99.77%	643	XM_003819029.5
<input checked="" type="checkbox"/> PREDICTED: Pan. troglodytes hemoglobin subunit beta (HBB), mRNA	Pan. troglodytes	815	815	100%	0.0	99.77%	639	XM_508242.5
<input checked="" type="checkbox"/> Homo sapiens full open reading frame cDNA clone RZPD0834D0222D for gene HBB, hemoglobin, bet...	Homo sapiens	815	815	100%	0.0	99.77%	444	CR538530.1
<input checked="" type="checkbox"/> Homo sapiens hemoglobin beta subunit variant (HBB) mRNA, complete cds	Homo sapiens	815	815	100%	0.0	99.77%	579	AF181989.1
<input checked="" type="checkbox"/> Homo sapiens hemoglobin_beta, mRNA (cDNA clone MGC:14540 IMAGE:4292125), complete cds	Homo sapiens	809	809	100%	0.0	99.55%	658	BC007075.1
<input checked="" type="checkbox"/> Homo sapiens full open reading frame cDNA clone RZPD0834E0633D for gene HBB, hemoglobin, bet...	Homo sapiens	809	809	99%	0.0	99.77%	441	CR541913.1
<input checked="" type="checkbox"/> Synthetic construct Homo sapiens clone FLH028849.01X hemoglobin beta (HBB) mRNA, complete cds	synthetic constr...	809	809	100%	0.0	99.55%	444	AY890157.1

Za drugo zaporedje dobimo:

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Human sickle cell beta-globin mRNA, complete cds	Homo sapiens	665	665	100%	0.0	100.00%	468	M25079.1
<input checked="" type="checkbox"/> Human sickle beta-hemoglobin mRNA	Homo sapiens	662	662	100%	0.0	90.99%	576	M25113.1
<input checked="" type="checkbox"/> Human messenger RNA for beta-globin	Homo sapiens	662	662	100%	0.0	91.22%	549	V00500.1
<input checked="" type="checkbox"/> Homo sapiens hemoglobin beta chain variant Hb S-Wake (HBB) mRNA, complete cds	Homo sapiens	654	654	100%	0.0	90.09%	534	AY136510.1
<input checked="" type="checkbox"/> Human messenger RNA for beta-globin	Homo sapiens	649	649	95%	0.0	91.51%	626	V00497.1
<input checked="" type="checkbox"/> Homo sapiens hemoglobin subunit beta (HBB), mRNA	Homo sapiens	645	645	95%	7e-180	91.27%	628	NM_000518.5
<input checked="" type="checkbox"/> PREDICTED: Pan. paniscus hemoglobin subunit beta (LOC100976465), mRNA	Pan. paniscus	645	645	95%	7e-180	91.27%	643	XM_003819029.5
<input checked="" type="checkbox"/> TPA_inf. Homo sapiens GLNA1 gene for globin A1	Homo sapiens	645	645	95%	7e-180	91.27%	444	LT548096.1
<input checked="" type="checkbox"/> PREDICTED: Pan. troglodytes hemoglobin subunit beta (HBB), mRNA	Pan. troglodytes	645	645	95%	7e-180	91.27%	639	XM_508242.5

Odkrijemo, da ima poglavar normalen človeški hemoglobin, medtem ko je iz vzorca 2 po izvedbi blastn jasno, da imamo opravka z mutirano obliko HBB, kar povzroči anemijo srpastih celic.

- c) Y pomeni C ali T, N pomeni katerokoli bazo. Vzrok za ta pojav je star zapis za mutiran hemoglobin, ki sem ga vzel iz GenBank (iz 1976, popravljen 1993).
- d) Pod Accession razdelkom po opravljenem Blast-n na strani GenBank poiščemo transkript za aminokislinsko zaporedje za kodirajočo sekvenco normalnega HBB.

```

CDS
51..494
/gene="HBB"
/gene_synonym="beta-globin; CD113t-C; ECYT6"
/note="beta globin chain; hemoglobin, beta; hemoglobin
beta subunit; Hb Monza protein"
/codon_start=1
/product="hemoglobin subunit beta"
/protein_id="NP_000509.1"
/db_xref="CCDS:CCDS7753.1"
/db_xref="GeneID:3043"
/db_xref="HGNC:HGNC:4827"
/db_xref="MIM:141900"
/translation="MVHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVPWQRFFE
SFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLNLRKGTFTLSELHCDKLHVDPE
NFRLLGNVLVCLAHFGKKEFTPPVQAAAYOKVVAGVANALAHKYH"

```

Kot osnovo je smiselno ta transkript uporabiti za obe rešitvi, saj bi pri starejšem zadetku morali ročno vnašati aminokislino na osnovi poravnave iz točke a) in tabele genskega koda. Slednjo pa potrebujemo, da ugotovimo, kaj za našo primarno strukturo HBB pomeni mutacija 7. (6.) kodona iz GAG → GTG. Zaradi te mutacije pride do spremembe aminokislino iz glutamata v valin. Mutacijo torej zapišemo E7V oz. E6V (v literaturi običajno start kodon ni vštet).

- e) Našli smo še en fragment, ki pripada plazmodiju malarije, in sicer kodira zapis za 18 S rRNA. To pomeni, da je bil domnevni napadalec okužen z malarijo.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Plasmodium malariae Sierra Leone P1 gene for 18S rRNA, partial sequence	Plasmodium m...	586	586	100%	3e-162	100.00%	317	LC859598.1
<input checked="" type="checkbox"/> Plasmodium cf. malariae type 1 small subunit ribosomal RNA gene, complete sequence	Plasmodium cf...	586	586	100%	3e-162	100.00%	2152	AF487999.1
<input checked="" type="checkbox"/> Plasmodium malariae rRNA (PmUG01_03031900), rRNA	Plasmodium m...	588	588	100%	1e-156	98.75%	2157	XR_003751927.1
<input checked="" type="checkbox"/> Plasmodium malariae clone B13 small subunit ribosomal RNA gene, partial sequence	Plasmodium m...	568	568	100%	1e-156	98.75%	1070	OR139045.1
<input checked="" type="checkbox"/> Plasmodium malariae clone A8 small subunit ribosomal RNA gene, partial sequence	Plasmodium m...	568	568	100%	1e-156	98.75%	1069	OR139043.1

- f) V tropskih predelih sveta imajo ljudje s heterozigotno mutacijo za HBB v resnici evolucijsko prednost pri okužbi s plazmodijem malarije. Plazmodij vstopa v eritrocite, kjer porablja kisik, se razmnožuje in povečuje oksidativni stres. Če so eritrociti deformirani, ne vežejo kisika tako učinkovito in jih telo prej odstrani iz krvnega obtoka in razgradi. Tako je tudi napredovanje okužbe upočasnjeno. PMID: npr. 34503910 (možnosti je ogromno – namen te točke je prebrati in izvedeti kaj novega).