

S1-15-rešitev

CRISPR-Cas je obrambni sistem, s katerim se lahko bakterije in arheje obranijo pred okužbami z bakteriofagi ali drugimi genskimi elementi. Sekvenciral si del genoma neznanе bakterije, za katerega sumiš, da vsebuje gene povezane s CRISPR-Cas obrambnim sistemom (za zaporedje glej nalogo).

1. Zanima te, kateri protein kodira sekvencirano zaporedje, zato uporabiš orodje [ORFfinder](#), ki se uporablja za identifikacijo odprtih bralnih okvirov - delov nukleotidnega zaporedja, ki v določenem bralnem okviru ne vsebujejo STOP kodonov. Na vrhu spletne strani je tudi krajši opis tega orodja in njegove uporabe. V polje poimenovano "enter query sequence" vstaviš zgornje zaporedje, ostalih nastavitev pa ne spreminjaš in na dnu strani pritisneš "submit" (slika 1). Rezultate ORFfinder-ja prikazuje slika 2.

Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

```
TTCTAAGCATTCAAAGCAATCAGTCTGTAATCGTAGCAACCAATTTGTCAATCCTGA  
TTTCACTCGGGTTTCTCAATAATGAGATTCAAACTGCGATGATCAATTTAAAT  
GGTTGAGGAGTTAAACATCAATCTAAATCTGTTACAAGCATGATTTCTGATCTCTT  
GATTTGTTTTCATCTGTGATCAAAAATTTAAGTTCAGAACTTCTCTCATATTTGTAACA  
ATATTGCACTATTTTGAATAACACAGACATCTTGGAGCACAGAAATTTGACCCCTCT  
TAATGGAATCGTTCATCTAATAAGGAAATTAAGATTCATCAAAAGCTCCCAAGAA  
ATACAAGTCTCTCATCGGAGTTTGAATACAATATTTCTTCCACATGATAAATCA
```

From: To:

Choose Search Parameters

Minimal ORF length (nt): 75

Genetic code: 1 Standard

ORF start codon to use:

"ATG" only

"ATG" and alternative initiation codons

Any sense codon

Ignore nested ORFs:

Start Search / Clear

Submit Clear

v to polje vnesesh podano zaporedje

Slika 1: ORFfinder

Open Reading Frame Viewer

Sequence

ORFs found: 24 Genetic code: 1 Start codon: 'ATG' only

ORF13 (1368 aa) Display ORF as... Mark

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF13	-	1	<5646	1540	4107 1368
ORF14	-	1	1539	670	870 289
ORF15	-	1	408	>1	408 135
ORF24	-	3	673	332	342 113
ORF21	-	2	2450	2229	222 73
ORF7	+	2	2060	2227	168 55
ORF8	+	2	2381	2533	153 50
ORF4	+	1	5236	5388	153 50
ORF23	-	2	1127	984	144 47
ORF9	+	2	2831	2968	138 45
ORF12	+	3	4267	4385	120 40

BLAST Database: Non-redundant protein sequences (nr)

BLAST

Slika 2: rezultat ORFfinder

- Koliko odprtih bralnih okvirjev najdeš? **24**
- Kateri izmed njih je pravilen in zakaj? Pravilen je odprti bralni okvir **13**, saj je **najdaljši ORF** tisti, ki po navadi kodira protein.

- Kateri protein kodira to zaporedje (ime ter uniprot accession code) in iz katerega organizma je? Za identifikacijo proteina, ki ga kodira to zaporedje uporabimo BLAST ter zbirko »non-redundant protein sequences«. Najlažje je, če do BLASTa dostopamo kar preko ORFfinderja (slika 2). Iz rezultatov v BLASTu (slika 3) razberemo, da gre za protein **CRISPR-associated endonuclease Cas9/Csn1** iz bakterije **Streptococcus pyogenes M1**. Da najdemo accession code proteina, ga z vpisom imena v iskalno vrstico poiščemo še v UniProtu: **Q99ZW2**

Slika 3: rezultati BLAST

- Kako velik je celoten genom tega organizma? V UniProtu gremo pri našem proteinu pod razdelek sequence in poiščemo povezavo do nukleotidnega zaporedja v GenBank (slika 4), tam razberemo, da ima celoten genom **1852433 bp**

Slika 4: Dostop do GenBanka iz UniProta

- Kateri ion je potreben za aktivnost tega proteina? **Mg²⁺**, saj je v UniProtu pod razdelkom function navedeno, da je magnezijev ion kofaktor proteina Cas9.
- Kakšna je funkcija tega proteina v bakterijskih celicah? **Protein Cas9 ima v kompleksu s crRNA in tracrRNA endonukelazno funkcijo, saj cepi linearno ali krožno tarčno dsDNA.**

2. CRISPER-Cas sistemi so zaradi nizke cene, ponovljivosti, učinkovitosti in preprostosti med najbolj uporabljenimi pristopi urejanja genomov.

- Koliko člankov objavljenih med letoma 2022 in 2024 vsebuje v naslovu besedo CRISPR (vključno z navedenima letoma)? Uporabimo napredno iskanje na PubMedu: pod razdelek *title* napišemo CRISPR, pod *date - publication* pa ustrezna datuma. **7698 člankov**
 - Koliko člankov, katerih besedilo je prosto dostopno, pa vsebuje besedo cas9? V iskalno vrstico napišemo cas9 in med filtri izberemo *free full text*. **29563 člankov**
3. Iskani protein je večdomenski. Za aktivacijo in opravljanje svoje funkcije mora tvoriti kompleks z gRNA (vodilno RNA). Ta je sestavljena iz dveh RNA molekul: tracrRNA (»trans-activating crRNA«) in crRNA, ki omogočata prepoznavo tarčnega zaporedja. Poleg tega pa vsebuje še domene za vezavo tarčne DNA ter nukleazne domene.
- Katere izmed domen zgornjega proteina so nukleazne? Zapiši njihova imena ter katere aminokislinske ostanke zajemajo. V UniProtu pod razdelkom *Family & Domains* preberemo, da ima Cas9 2 vrsti nukleaznih domen: **RuvC-podobna domena** med aminokislinskimi ostanki **1-62, 718-765 in 925-1102** in **HNH domena** med aminokislinskimi ostanki **770-921**.
 - Kateri aminokislinski ostanki so v aktivnem mestu nukleaznih domen? Odgovor najdemo v UniProtu pod razdelkom *Function*: pri **RuvC-podobni domeni je v aktivnem mestu D10, pri HNH domeni pa H840**.
 - Katera molekula inhibira nukleazno aktivnost? Odgovor najdemo v UniProtu pod razdelkom *Function*: **EDTA**
 - Katere sekundarne strukture so prisotne v strukturi HNH domene? V UniProtu pod razdelkom *Family & Domains* poiščemo direktno povezavo do orodja InterPro, kjer s klikom na HNH domeno izvemo več podrobnosti o njeni strukturi (slika 5). Ugotovimo, da **HNH domena vsebuje eno beta ploskev iz dveh antiparalelnih beta trakov, ki ju na vsaki strani obdaja po ena alfa vijačnica**. Alternativna rešitev: na PDB poiščemo strukturo Cas9 proteina ter HNH domeno preučimo v ChimeriX.

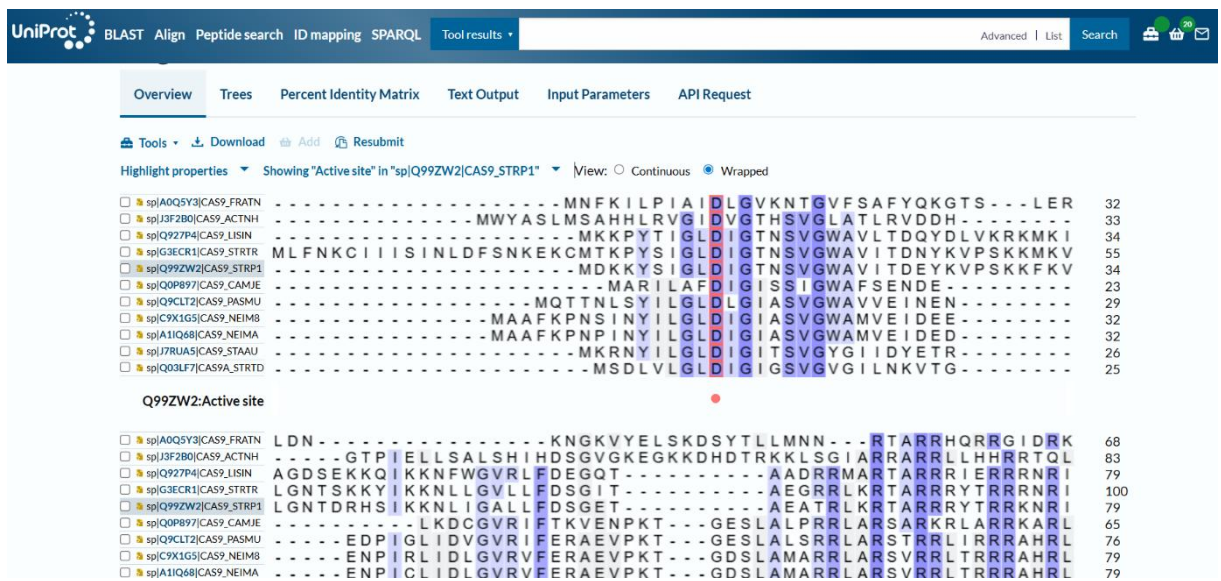


Slika 5: Dostop do podrobnosti o HNH domeni z uporabo orodja InterPro

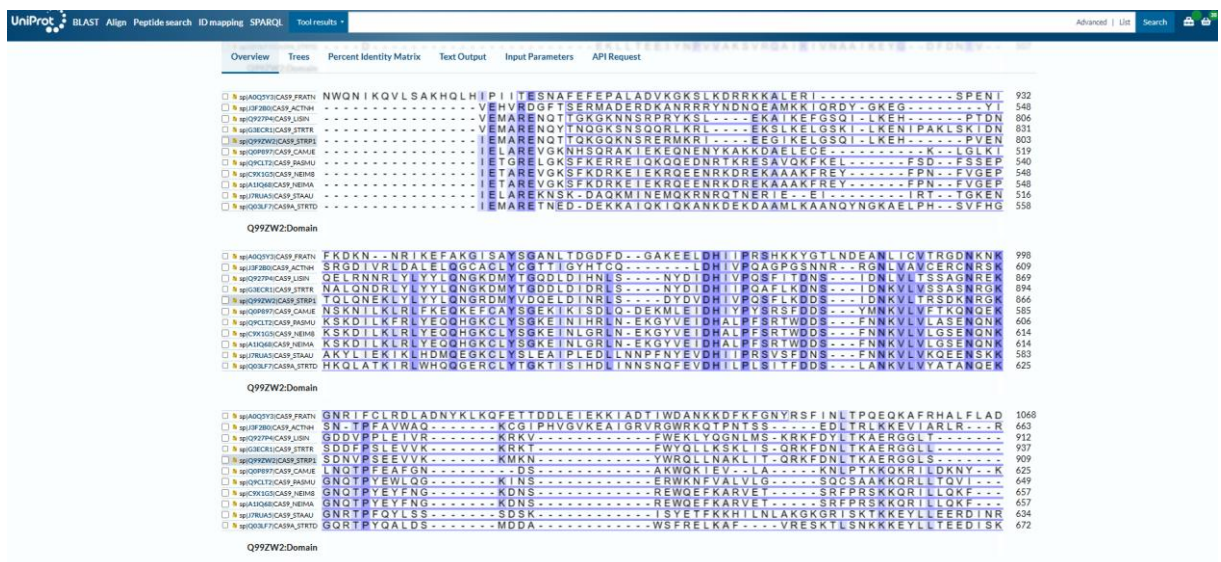
4. Poišči iskani protein pri vsaj še desetih drugih organizmih ter s pomočjo poravnave več zaporedij odgovori na spodnja vprašanja:

Za reševanje tega dela naloge lahko uporabimo orodje Clustal Omega, še lažje pa je, če poravnavo naredimo kar v UniProtu (slika 6).

- Zakaj je v poravnavi prisotnih toliko vrzeli in nepodobnih delov? Kljub temu da so proteini Cas9 iz različnih organizmov funkcijo enaki, se lahko precej razlikujejo v aminokislinskem zaporedju in njegovi dolžini, kar vpliva na njihovo specifičnost in učinkovitost, zato je v poravnavi prisotnih veliko vrzeli in nepodobnih delov.
- Katere domene oz. aminokislinski ostanki so najbolj ohranjeni in zakaj? Iz poravnave lahko razberemo, da so najbolj ohranjeni aminokislinski ostanki v aktivnih mestih (slika 6) ter nukleazne domene (slika 7), saj so le-te ključne za funkcijo Cas9 proteina.



Slika 6: Poravnava zaporedij Cas9 proteinov iz več organizmov ter ohranjenost aktivnega mesta



Slika 7: Poravnava zaporedij Cas9 proteinov iz več organizmov ter podobnost HNH domene

5. Za pravilno delovanje proteina je ključnega pomena njegova struktura. Ko proučevani protein tvori kompleks z gRNA, pride do konformacijskih sprememb, ki omogočajo njegovo aktivacijo. Do ustreznih PDB kod pridemo preko UniProt-a pod razdelkom *Structure* in nato v PDBju poiščemo odgovore na preostala vprašanja.

- Poišči strukturi proteina brez dodanih ligandov ter v kompleksu z vodilno RNA in tarčno DNA. Kakšni sta njuni PDB kodi? Cas9 brez ligandov: **4CMP**, Cas9 v kompleksu z vodilno RNA in tarčno DNA: **4O08**
- S katerima metodama so določili ti dve strukturi? **XRD**
- Kateri ekspresijski sistem so uporabili za določanje strukture proteina? Cas9 brez ligandov: **Escherichia coli BL21(DE3)**, Cas9 v kompleksu z vodilno RNA in tarčno DNA: **Escherichia coli**
- Kateri del strukture omogoča vezavo RNA molekul? Pod razdelkom *literature* v PDB preberemo, da je za vezavo sgRNA in DNA ključen **prepoznavni režanj** («recognition lobe«)