Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2024/2025

Exam

Tuesday 6.5.2025 13:10

Moodle: files with the version
open „task" for submission (like for homeworks).

Usual Exam test:

2 parts, each five points

Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

Version 1

Find human protein sequence **FGF3**

- What is the accession number and function of this protein?

- How many valines does the sequence contain?

- How long is the longest fragment after trypsin digestion?

- How many human similar reference proteins are in the databases?

- Does the sequence contain any transmembrane regions?

Download segence **NM_001757.3**

- What does this sequence encode ?

- Is there a complementary sequence to primer R1?

  ```
  >R1
  GCTCTGACGCTCATGATGC
  ```

- In which exon is the complemetary sequence for this primer?

- Is the primer suitable for PCR with Ta=60°C?

- Design suitable F primer for this experiment.

Last year Exam
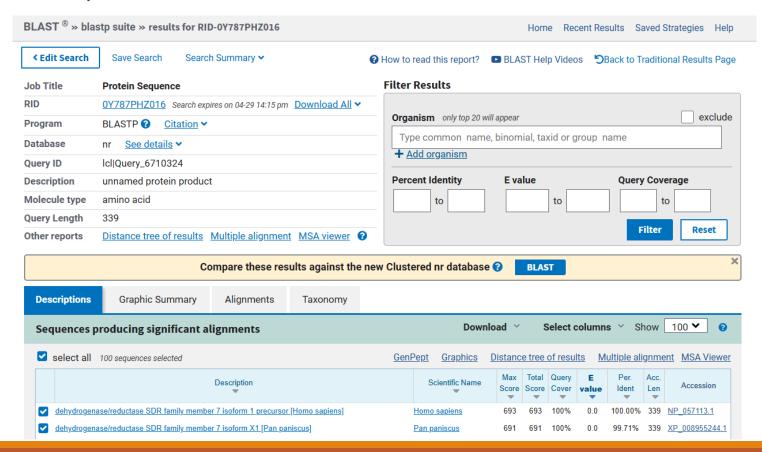
Work with the following sequence:

```
MNWELLLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVVWVTGASS
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA
ATKAVLQEFGRIDILVNNGGMSQRSLCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER
KQGKIVTVNSILGIISVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN
IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY
MPTWAWWITNKMGKKRIENFKSGVDADSSYFKIFKTKHD
```

- Identify the sequence.
- Does this protein have any transmembrane regions?
- How many isoforms this protein has? Align them.
- How many cysteins does the sequence have?
- Design primers to amplify the CDS of the respective gene.

Work with the following sequence:

```
MNWELLLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVVWVTGASS

GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA

ATKAVLQEFGRIDILVNNGGMSQRSLCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER

KQGKIVTVNSILGIISVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN

IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY

MPTWAWWITNKMGKKRIENFKSGVDADSSYFKIFKTKHD
```

- Identify the sequence. → BLASTp

- Does this protein have any transmembrane regions? → TMHMM, TopCons, Phobius…
  ….not Protter (just figure)

- How many isoforms this protein has? Align them. → Uniprot for isoforms, multalin for alingment

- How many cysteins does the sequence have? → SMS: Protein Stats

- Design primers to amplify the CDS of the respective gene. → Manual design (OligoCalc for check-similar Tm)
  SMS:Reverse complement for R primer

Work with the following sequence:

- Identify the sequence.
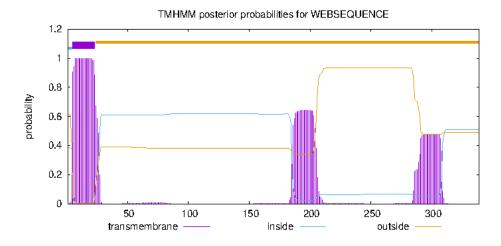
DHRS7

Work with the following sequence:

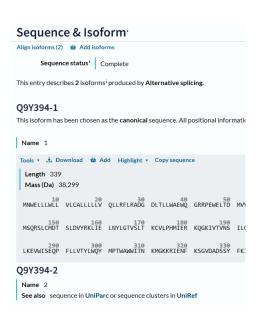- Does this protein have any transmembrane regions?

Yes 1:

TMHMM result

```
# WEBSEQUENCE Length: 339
# WEBSEQUENCE Number of predicted TMHs:  1
# WEBSEQUENCE Exp number of AAs in TMHs: 44.121
# WEBSEQUENCE Exp number, first 60 AAs:  20.90194
# WEBSEQUENCE Total prob of N-in:        0.39052
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE     TMHMM2.0        inside      1     4
WEBSEQUENCE     TMHMM2.0        TMhelix     5    23
WEBSEQUENCE     TMHMM2.0        outside    24   339
```


TMHMM posterior probabilities for WEBSEQUENCE

Work with the following sequence:

- How many isoforms this protein has? Align them.

2 isoforms in uniprot

4 isoforms in NCBI

Work with the following sequence:

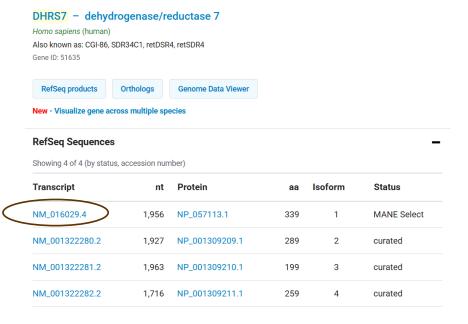- How many cysteins does the sequence have?

7 cysteins:

Sequence Manipulation Suite:

Protein Stats

Protein Stats returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each res

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 500,000,000 characters.

```
MNWELLLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVVWVTGASS
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA
ATKAVLQEFGRIDILVNNGGMSQRSLCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER
KQGKIVTVNSILGIISVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN
IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY
MPTWAWWITNKMGKKRIENFKSGVDADSSYFKIFKTKHD
```

Submit Clear Reset

*This page requires JavaScript. See browser compatibility.
*You c

Sun 1 O
Valid XHT

Sequence Manipulation Suite – Škola – Microsoft Edge — ☐ ✕

ⓘ about:blank

Protein Stats results

Results for 339 residue sequence "Untitled" starting "MNWELLLWLL"

Pattern:	Times found:	Percentage:
A	16	4.72
B	0	0.00
C	7	2.06
D	14	4.13

Work with the following sequence:

- Design primers to amplify the CDS of the respective gene.

DHRS7 – dehydrogenase/reductase 7

Homo sapiens (human)

Also known as: CGI-86, SDR34C1, retDSR4, retSDR4

Gene ID: 51635

Transcript	nt	Protein	aa	Isoform	Status
NM_016029.4	1,956	NP_057113.1	339	1	MANE Select
NM_001322280.2	1,927	NP_001309209.1	289	2	curated
NM_001322281.2	1,963	NP_001309210.1	199	3	curated
NM_001322282.2	1,716	NP_001309211.1	259	4	curated

F: ATG AAC TGG GAG CTG CTG C
R: TCA GTC ATG TTT TGT CTT AAA GAT TTT A

Exam test:

Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

Version 1

Find human protein sequence **FGF3**

- What is the accession number and function of this protein?

- How many valines does the sequence contain?

- How long is the longest fragment after trypsin digestion?

- How many human similar reference proteins are in the databases?

- Does the sequence contain any transmembrane regions?

Download seqence **NM_001757.3**

- What does this sequence encode ?

- Is there a complementary sequence to primer R1?

  >R1
  GCTCTGACGCTCATGATGC

- In which exon is the complemetary sequence for this primer?

- Is the primer suitable for PCR with Ta=60°C?

- Design suitable F primer for this experiment.

Download from database the sequence NM_001282692.1

- What does this mRNA encode?

- How long is the second exon?

- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or ScaI?

- What fragments are formed after cleavage by all these enzymes at once?

- Compare how similar is the corresponding protein to the respective mouse homologue?

Download from database the sequence NM_001282692.1

- What does this mRNA encode? → Finding the sequnce in NCBI

- How long is the second exon? → NCBI: highlit sequence features or Graphic

- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or ScaI? → Get CDS only and SMS: Restriction summary

- What fragments are formed after cleavage by all these enzymes at once? → SMS: Restiction digest

- Compare how similar is the corresponding protein to the respective mouse homologue? → Get protein and mouse protein and Laling or Needle pairwise comparison

Download from database the sequence NM_001282692.1

- What does this mRNA encode?

FMO1

GenBank ▾

Homo sapiens flavin containing dimethylaniline monoxygenase 1 (FMO1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_001282692.1

FASTA   Graphics

Go to: ☑

```
LOCUS       NM_001282692            2198 bp    mRNA    linear   PRI 09-FEB-2025
DEFINITION  Homo sapiens flavin containing dimethylaniline monoxygenase 1
            (FMO1), transcript variant 1, mRNA.
ACCESSION   NM_001282692
```

Download from database the sequence NM_001282692.1

- How long is the second exon?

```
298..486 : 486-298+1 = 189nt
```

Download from database the sequence NM_001282692.1

- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or ScaI?

BamHI, ScaI- cuts 1x

EcoRI –does not cut

Ball tgg|cca	17, 381, 1056, 11
BamHI g|gatcc	1378
EcoRI g|aattc	none
ScaI agt|act	521

FASTA ▾

Homo sapiens flavin containing dimethylaniline monoxygenase 1 (F transcript variant 1, mRNA

NCBI Reference Sequence: NM_001282692.1

GenBank   Graphics

>NM_001282692.1:154-1764 Homo sapiens flavin containing dimethylaniline monoxygenase 1 (FMO1), transcript variant 1, mRNA
ATGCAGGAGAACATGGCCAAGCGAGTTGCCATTGTGGGAGCTGGGGTCAGCGGCCTGGCCTCCATCAAGT
GCTGTCTGGAAGAAGGACTGGAGCCCACCTGCTTTGAGAGGAGCGATGACCTTGGGGGGCTGTGGAGATT
CACCGAACATGTTGAAGAAGGCAGAGCCAGTCTCTACAAGTCTGTGGTTTCCAACAGCTGCAAGGAGATG
TCTTGTTACTCAGACTTTCCATTCCCAGAAGATTATCCAAACTATGTGCCAAATTCTCAATTCCTGGAAT
ATCTCAAAATGTATGCAAACCACTTTGACCTTCTGAAACACATTCAATTCAAGACCAAAGTCTGCAGTGT
AACAAAATGCTCAGATTCTGCTGTCTCTGGCCAATGGGAGGTGGTCACTATGCATGAAGAGAAGCAAGAG
TCAGCCATCTTTGATGCTGTCATGGTCTGCACTGGCTTTCTTACTAATCCTTATTTGCCACTGGATTCCT
TTCCAGGTATTAATGCCTTTAAAGGGCCAGTACTTTCATAGCCGGCAATATAAGCATCCAGATATATTTAA

Download from database the sequence NM_001282692.1

- What fragments are formed after cleavage by all these enzymes at once?

Three fragments: 857 + 520 +234 bp

Download from database the sequence NM_001282692.1

- Compare how similar is the corresponding protein to the respective mouse homologue?

Previous tasks-hints (solutions in other pdf)

Find human protein sequence FGF3

What is the accession number and function of this protein? ⟶ Uniprot or NCBI

How many valines does the sequence contain? ⟶ SMS: Protein Stats

How long is the longest fragment after trypsin digestion? ⟶ Peptide cutter-include the third table

How many human similar reference proteins are in the databases? ⟶ BLASTp-looking for RefSeq
in homo sapiens

Does the sequence contain any transmembrane regions?

TMHMM, TopCons, Phobius…
….not Protter (just figure)