

Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2024/2025

9/9

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 sequence **NM_001757.3**

- What does this sequence encode ?
- Is there a complementary sequence to primer R1?

```
>R1  
GCTCTGACGCTCATGATGC
```

- In which exon is the complementary 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:

MNWELLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVVWVTGASS
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA
ATKAVLQEFGRIDILVNNGGMSQRS LCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER
KQGKIVTVNSILGII SVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN
IVENSLAGEVT KTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY
MPTWAWWITNKMGGKKRIENFKSGVDADSSYFKIFKTKHD

- 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:

MNWELLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVWVVTGASS
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA
ATKAVLQEFGRIDILVNNGGMSQRSLCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER
KQGKIVTVNSILGII SVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN
IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY
MPTWAWWITNKMGGKKRIENFKSGVDADSSYFKIFKTKHD

- 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 alignment**
- 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

BLAST® » blastp suite » results for RID-0Y787PHZ016 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[← Edit Search](#) [Save Search](#) [Search Summary ▾](#) [? How to read this report?](#) [▶ BLAST Help Videos](#) [↶ Back to Traditional Results Page](#)

Job Title Protein Sequence

RID [0Y787PHZ016](#) Search expires on 04-29 14:15 pm [Download All ▾](#)

Program BLASTP [?](#) [Citation ▾](#)

Database nr [See details ▾](#)

Query ID lcl|Query_6710324

Description unnamed protein product

Molecule type amino acid

Query Length 339

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](#)

Compare these results against the new Clustered nr database [?](#) [BLAST](#)

Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments [Download ▾](#) [Select columns ▾](#) Show [?](#)

☒ select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

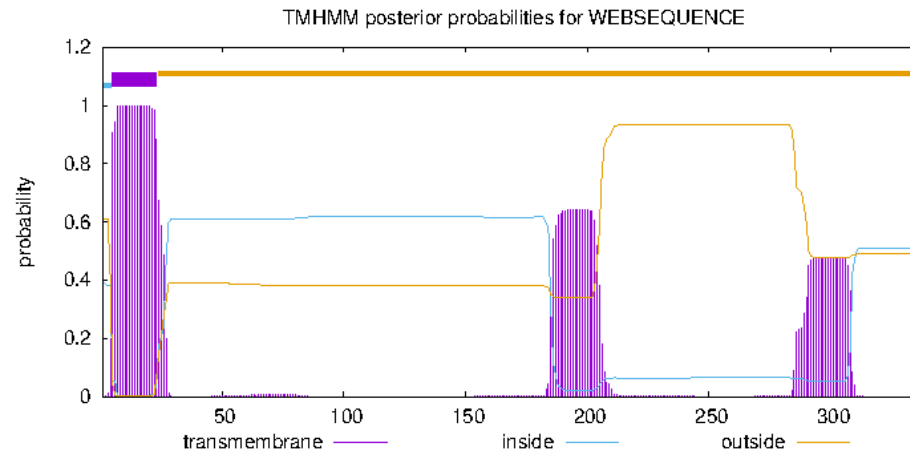
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	dehydrogenase/reductase SDR family member 7 isoform 1 precursor (Homo sapiens)	Homo sapiens	693	693	100%	0.0	100.00%	339	NP_057113.1
<input checked="" type="checkbox"/>	dehydrogenase/reductase SDR family member 7 isoform X1 (Pan paniscus)	Pan paniscus	691	691	100%	0.0	99.71%	339	XP_008955244.1

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
```



Work with the following sequence:

- How many isoforms this protein has? Align them.

2 isoforms in uniprot

Sequence & Isoformⁱ

Align isoforms (2) [Add isoforms](#)

Sequence statusⁱ | Complete

This entry describes 2 isoformsⁱ produced by **Alternative splicing**.

Q9Y394-1

This isoform has been chosen as the **canonical** sequence. All positional informati

Name 1

Tools [Download](#) [Add](#) [Highlight](#) [Copy sequence](#)

Length 339
Mass (Da) 38,299

10 20 30 40 50 60 70 80 90 100 110 120 130

MINWELLWLVLVLCALLLLVQLRLFRADGDLTLLEHAGRRPEMELTDHVVYVWGSSGDEELAYQLSKLGYSVLSSRRPHLEKPKRRLLENGHKEKDIIVLPDLTDTGSHEHATKAVLQDFG

150 160 170 180 190

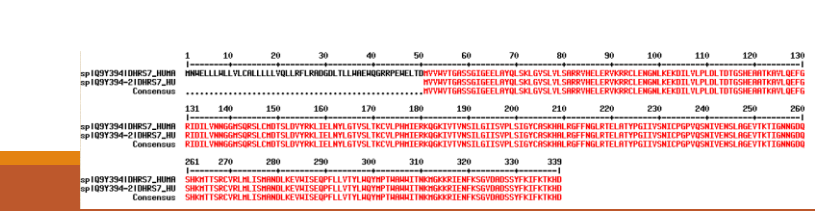
MSQRSLSMDTSLDVYRKLIENLYLGTVSLTKCVLPKHIERKQGIKIVTVNSIL

200 210 220 230 240 250 260

LKEVWISQEPFLLVTVLWQYMPMTAWMIITNKMGGKRIENFKSGVDADSSYFK

Q9Y394-2

Name 2
See also sequence in UniParc or sequence clusters in UniRef



4 isoforms in NCBI

DHRS7 – dehydrogenase/reductase 7

Homo sapiens (human)

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

Gene ID: 51635

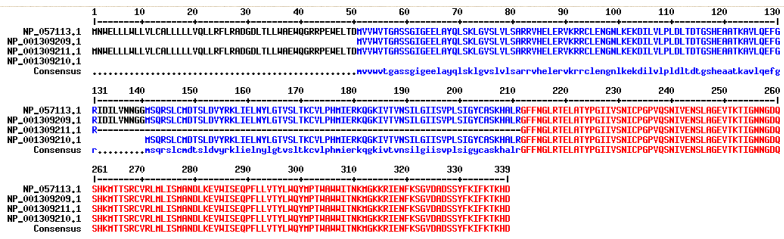
- RefSeq products
- Orthologs
- Genome Data Viewer

New - Visualize gene across multiple species

RefSeq Sequences

Showing 4 of 4 (by status, accession number)

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



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.

```
MNWELLWLLVLCALLLLLVQLRFLRADGDLTLLWAEWQGRPEWELTDMVVWVTGASS
GIGEELAYQLSKLGVSLVLSARRVHELKRRCLNGLKEKDILVPLDLTDTGSHEA
ATKAVLQEFGRIDLNVNNGGMSQRSCLMDTSLDVYRKLIENYLGTVSLTKCVLPHMIER
KQGIKIVTVNSILGIISVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN
IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY
MPTWAWWITNKMGGKRIENFKSGVDADSSYFKIFKTKHD
```

*This page requires JavaScript. See [browser compatibility](#).

*You c

Sequence Manipulation Suite – Škola – Microsoft Edge

about:blank

Protein Stats results

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

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

RefSeq products Orthologs Genome Data Viewer

New - Visualize gene across multiple species

RefSeq Sequences

Showing 4 of 4 (by status, accession number)

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

FASTA

Homo sapiens dehydrogenase/reductase 7 (DHRS7),

NCBI Reference Sequence: NM_016029.4

GenBank Graphics

>NM_016029.4:90-1109 Homo sapiens dehydrogenase/reductase 7 (DHRS7), transcribed mRNA

```
ATGAAC TGGAGCTGCTGCTGGCTGCTGGTCTGTGCGCTGCTCCTGCTCTTGGTGCAGCTGCTGC
GCTTCTCTGAGGGCTGACGGCGACCTGACGCTACTATGGGCCGAGTGGCAGGGACGACGCCAGAATGGGA
GCTGACTGATATGGTGGTGGGTGACTGGAGCCTCGAGTGAATTGGTGAGGAGCTGGCTTACAGTTG
TCTAACTAGGAGTTTCTCTTGTGCTGCAGCCAGAAGTGCATGAGCTGGAAGGGTGAAAGGAAGAT
GCCTAGAGAATGGCAATTTAAAGAAAAAGATATACTTGTTTTGCCCTTGACCTGACCGACACTGGTTTC
CCATGAAGCGGCTACCAAGCTGTTCTCAGGAGTTTGGTAGAATCGACATTCTGGTCAACATGGTGGA
ATGTCACAGCGTTCTCTGTGCATGGATACCAGCTTGGATGTCACAGAAAGCTAATAGAGCTTAACTACT
TAGGGACGGTGTCTTGCAGAAATGTGTTCTGCTCATGATCGAGAGGAAGCAAGGAAAGATTGTTAC
TGTGAATAGCATCTGGGTATCATATCTGTACCTCTTCCATTGGATACTGTGCTAGCAAGCATGCTCTC
CGGGGTTTTTTAATGGCTTTCGAACAGAAGTTGCCACATACCCAGGTATAATAGTTCTTAACATTTGCC
CAGGACCTGTGCAATCAAAATATTGGGAGAATTCCTAGCTGGAGAAGTCACAAAGACTATAGGCAATAA
TGGAGACCAGTCCCAAGATGACAACAGCTGTTGTGTGCGGCTGATGTTAATCAGCATGGCCAATGAT
TTGAAAGAAGTTTGGATCTCAGAACACCTTTCTTGTAGTAGAACATATTTGTGGCAATACATGCCAACCT
GGGCTGGTGGATAACCAAGATGGGGAAGAAAGGATTGAGAACTTTAAGAGTGGTGGATGCAGA
CTCTTCTATTTTAAATCTTTAAGACAAAACATGACTGA
```

Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
GC and Molecular Weight calculations are for single-stranded DNA or RNA

Nucleotide base codes

ATG AAC TGG GAG CTG CTG C

Reverse Complement Strand(5' to 3') is:

GCA GCA GCT CCC AGT TCA T

5' modification (if any) 3' modification (if any) select ss/ds and DNA or RNA

5 (Minimum base pairs required for single primer self-dimerization) 4 (Minimum base pairs required for a hairpin)

50 mM Primer 50 mM Salt (Na⁺)

Calculate Swap Strands BLAST Check Self-Complementarity

Physical Constants

Length: 19 bases

GC content: 58 %

Molecular Weight: 5868.9

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 4.839 microMolar and contains 28.4 micrograms

Melting Temperature (T_m) Calculations

1 53 °C (Basic)

2 50 °C (Salt Adjusted)

3 55 °C (Nearest Neighbor)

Thermodynamic Constants

Sequence Manipulation Suite:

Reverse Complement

Reverse Complement converts a DNA sequence into its reverse, complement, or reverse-complement of a sequence if it contains an ORF on the reverse strand.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100,000 characters.

```
GCCAATGAT
TTGAAGAAGATTTGGATCTCAGAACCACTCTTCTGTAGTAACATATTTGGCAATACA
TGCCAACT
GGGCTGGTGGATACCAACAAGATGGGGAAGAAAGATTTGAGAACTTTAAGAGTGGTGT
GGATGCAGA
CTCTTCTATTTTAAATCTTTAAGACAAAACATGACTGA
```

Submit Clear Reset

Sequence Manipulation Suite - Skola - Microsoft Edge

about:blank

Reverse Complement results

>NM_016029.4:90-1109 Homo sapiens dehydrogenase/reductase 7 (DHRS7), transcribed mRNA

TCAGTCACTGTTTTGTCTTAAAGATTTTAAATAAGAGAAGTCTGCGATCCACACCACTCTT

AAAGTCTCTCAATCCTTTTCTCCCACTCTGTTGGTTATCCACAGGCCAGTGTGGCAT

GTATTGGCCACAAATATGTTACTAACAGAAAGTGTGTTCTGAGATCCAACTTCTTCAA

ATCATTGGCCATGCTGATTAAACATCAGCCGCACACACAGACTGGTTGTCTCTTGGGA

CTGGTCTCCATTATTGCTATAGTCTTTGTGACCTTCTCAGCTAGGGAAATCTCCACAA

ATTGATTGACAGAGTCTGGGCAATGTTAGAACTATTATACCTGGTATGTGGCAAG

TTCTGTTGGAAGGCCATTAAGAAACCCGGAGAGATGCTCTGAGCAGCATATCCAA

GGAAAGAGTACAGATATGATACCCAGGATGCTATTACAGTAACAACTTTGCTGTGTT

CCTCTGATCATGTGAGGCAGAACACATTTTGTCAAGGACACCGTCCCTAAGTAGTTAAG

CTCTATTAGCTTCTGTAGACATCCCACTGGTATCCATGCACAGAGAACGCTGGGACAT

TCACCAATTGTTGACAGAAATGCGATTCTACCAAACTCCTGGAGAACAGCTTTGGTAGC

CGCTTCATGAGAACAGTGTGCTGAGTCAAGTCAAGGGGCAAAAGATATATCTTTTCTTT

TAAATGCGATCTCTGAGGACATCTTTTCAACCTTTTCAGCTCAGCAGCTCTCTGGC

TGACAGCAGAGAGAACTCTAGTTTAGACAACTGGTAAGCCAGCTCTCCCAATGCC

ACTCGAGGCTCCAGTCAACACACACATATCAGTCAAGCTCCCATCTCTGGGCTCGTGC

CTGCCACTCGGCCATAGTAGCGTCAAGTCTGCCCTCAGCCCTCAGGAAGCGCAGCAGCTG

CACCAAGAGCAGGACAGCGGCACAGCAGCAGCCACAGCAGCAGCTCCAGTTCTAT

Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
GC and Molecular Weight calculations are for single-stranded DNA or RNA

Nucleotide base codes

TCA GTC ATG TTT TGT CTT AAA GAT TTT AA

Reverse Complement Strand(5' to 3') is:

TTA AAA TCT TTA AGA CAA AAC ATG ACT GA

5' modification (if any) 3' modification (if any) select ss/ds and DNA or RNA

5 (Minimum base pairs required for single primer self-dimerization) 4 (Minimum base pairs required for a hairpin)

50 mM Primer 50 mM Salt (Na⁺)

Calculate Swap Strands BLAST Check Self-Complementarity

Physical Constants

Length: 29 bases

GC content: 24 %

Molecular Weight: 8886.9

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 4.839 microMolar and contains 28.4 micrograms

Melting Temperature (T_m) Calculations

1 52 °C (Basic)

2 60 °C (Salt Adjusted)

3 54 °C (Nearest Neighbor)

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 sequence **NM_001757.3**

- What does this sequence encode ?
- Is there a complementary sequence to primer R1?

```
>R1  
GCTCTGACGCTCATGATGC
```

- In which exon is the complementary 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 Scal?
- 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 sequence in NCBI
- How long is the second exon? → NCBI: highlight sequence features or Graphic
- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or Scal? → Get CDS only and SMS: Restriction summary
- What fragments are formed after cleavage by all these enzymes at once? → SMS: Restriction digest
- Compare how similar is the corresponding protein to the respective mouse homologue? → Get protein and mouse protein and Lalign or Needle pairwise comparison

Download from database the sequence NM_001282692.1

- What does this mRNA encode?

FMO1

GenBank ▾

Homo sapiens flavin containing dimethylaniline monooxygenase 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 monooxygenase 1 (FMO1), transcript variant 1, mRNA.				
ACCESSION	NM_001282692				

Download from database the sequence NM_001282692.1

- How long is the second exon?

$$298..486 : 486-298+1 = 189\text{nt}$$

GenBank

Send to:

Change region shown

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

NCBI Reference Sequence: NM_001282692.1

FASTA Graphics

Go to:

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

NM_001282692

2198 bp

mRNA

linear

PRI 09-FEB-2005

Homo sapiens

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

NM_001282692

NM_001282692.1

RefSeq.

Homo sapiens (human)

61

atgacctaat

cttcaaaat

ctttttctg

gcttgtaga

gcagccaagg

gtggggtaga

121

gcttgtaga

aaaagcctg

cttcattct

ctcatcgag

agaacatg

caagcaggt

181

gccattgtg

gagctgggt

cagcgctct

gcctccatc

atgctgtct

ggaaagaaga

241

ctggagccca

ctgcttga

gagagcag

gacctggg

ggctgtggag

attaccgaa

301

catgtgaag

agggcagcg

cagctctac

angctgtgg

tttccaaacg

ctgcaggag

361

atgtctgtt

actcagact

tccattccca

gaagattatc

caaaactatg

gccaaattc

421

caattctgg

aatattcaa

aagtatgca

aaccatttg

accttctga

acacattcaa

481

ttcagccca

aagctgtcg

tgtaacaaa

tgctcagtt

ctgctgtct

tgccaatgg

541

gaggttgta

ctatgcata

agaaagcaac

gagtcagca

ctttgtatc

tgctatgtg

601

tgactctgt

ttcttctaa

tctctattg

ccactggat

ctttccagg

tattatgac

661

ttaaagcgc

agtaattca

taggcgcaac

tataagcac

cagatattat

taaggacaa

721

agagctctg

tgattgaat

gggaattct

ggcacagaca

ttgctgtga

ggccagccc

781

ctggcgaaa

aggtgttct

cagcacacc

ggagggagat

gggtatcg

ccgaatttt

841

gactgggct

accatggga

catgtgttc

atgacagct

ttcagaacat

gttgagaatt

901

tccctccca

cccaattgt

gactgtgtg

atggagcaa

agataaacaa

ctggctcaat

961

catgcaaat

acggcttaat

accagaagac

aggactcag

tgaagagtt

tgctctaat

1021

gatgagctc

caggagcat

catactggg

aaagtgtta

tcaggccaag

cataaaagag

1081

gtaaaagaaa

actctgtcat

attatacaat

acttcaaa

aagagccat

tgacatcatt

1141

gtctttgca

ctggatacac

atttgcttc

cccttctgt

atgagctgt

agtgaagtt

1201

gaagatggc

agggctcat

gtacaaagt

atcttccctg

cacatctga

aaagccaacc

1261

ctggccatta

ttggctcat

caaaccttg

ggctcatga

tactacagg

agaacacaa

1321

ctcagtg

gggtgag

ctctgaag

ctgaataagt

taccacacc

aagtgctac

1381

atagagaaa

taagtacag

gaagaagaac

aagccaggt

gcttggtct

gtgtctatg

1441

aaagtttgc

atcagattta

tatcacatc

atagatgac

tcttgacct

tataatgaca

1501

aaaccacac

ttgtctcat

gtctcaaac

gctccatc

tggtctgac

cgctctctt

1561

ggccactgc

cacatacca

gttctcggt

actggccag

gaaaatgga

aggagccaga

1621

aatgccatc

tgaccacgt

ggagccaaca

ttcagggtca

tcaagctcg

agttgtaca

1681

gagctccat

ctctcttga

aagtttctt

aaagtctta

gctttctgc

ttgtctgtg

1741

gctattttt

tgattttct

ataagtaaa

gactctctaa

atggagatg

cacagagtag

1801

atttacaatg

ctccaattcc

tctctacag

caatattgc

ttcacagtta

taaaactgat

1861

tcaaatagta

aaggcaccc

tctcgcttc

ctggctgac

cagggtcac

cactgtgatt

1921

ctcgagctc

tccagctcc

acttctaag

ctagagaagt

ataactaaga

ctttgtgtca

1981

tttgaaggt

gttgaaga

tacaggttca

tttgaaga

aaagctgttc

gtgacagcat

2041

cttgagcat

catactctt

tccatataa

actattttca

cagatctcaa

ctaaacacc

2101

ttaattcaca

aatgatgtg

ttgtgtgaa

atggtgtctt

tatagtactg

gcttattaaa

exon

Feature

2 of 8

1 segment

298..486

/gene="FMO1"

/inference="alignment:SpLinn2"

Details

Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), trar

NCBI Reference Sequence: NM_001282692.1

[GenBank](#) [FASTA](#)

Genes

NP_001269621.1

FMO1

FMO-like

exon

misc_feature Features

upstream in...

polyA_site Features

regulatory Features

Cited Variations, dbSNP b157 v2

rs10912675 | T/C

rs1126692 | A/C/G

rs606396

BLAST nr: [NM_001282692.1 \(298..486\)](#)

BLAST to Genome: [NM_001282692.1 \(298..486\)](#)

FASTA record: [NM_001282692.1 \(298..486\)](#)

Download from database the sequence NM_001282692.1

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

BamHI, Scal- cuts 1x

EcoRI –does not cut

BamI tgg cca	17, 381, 1056, 11
BamHI g gatcc	1378
EcoRI ggaattc	none
EcoRI glaattc	none
Scal agt act	521

FASTA

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

NCBI Reference Sequence: NM_001282692.1

[GenBank](#) [Graphics](#)

```
>NM_001282692.1:154-1764 Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA
ATGCAGGAGAACATGGCCAAGCGAGTTGCCATTGTGGGAGCTGGGGTCAGCGGCCTGGCCTCCATCAAGT
GCTGTCTGGAAGAAGGACTGGAGCCACCTGCTTTGAGAGGAGCGATGACCTTGGGGGCTGTGGAGATT
CACCGAACATGTTGAAGAAGGCAGAGCCAGTCTCTACAAGTCTGTGGTTTCCAACAGCTGCAAGGAGATG
TCTTGTTACTCAGACTTTCCATTCCAGAAGATTATCCAAACTATGTGCCAAATTCCTCAATTCCTGGAAT
ATCTCAAAATGTATGCAAAACCACTTTGACCTTCTGAAACACATTCAATTCAAGACCAAAAGTCTGCAGTGT
AACAAAATGCTCAGATTCTGCTGTCTCTGGCCAATGGGAGGTGGTCACTATGCATGAAGAGAAGCAAGAG
TCAGCCATCTTTGATGCTGTCATGGTCTGCACTGGCTTTCTTACTAATCCTTATTTGCACTGGATTCTCT
TTCCAGGTATTAATGCTTTTAAAGGCGAGTACTTTTCATAGCGGCGAATATAGCATTCAGATATATTTAA
```

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

Sequence Manipulation Suite:

Restriction Digest

Restriction Digest cleaves a DNA sequence in a virtual restriction digest, with one, two, or three restriction enzymes that produced them. You can digest linear or circular molecules, and even a mixture of molecules (by entering

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100,000,00

GCCAGAAAT
GCCATCATGACCCAGTGGGACCGAACATTCAAGGTCATCAAGCTCGAGTTGTACAAGAGT
CTCCATCTC
CCTTTGAAAGTTTCTTAAAGCTTTAGCTTCTGGCTTGTGTGGCTATTTTTCTGAT
TTTTCTATA

Submit Clear Reset

- Treat sequences as molecules.
- Digest with and and

*This is a screenshot of a web browser showing a page titled "Sequence Manipulation Suite - Škola - Microsoft Edge". The page content is mostly obscured by a large, semi-transparent watermark that reads "Škola".

about:blank

Restriction Digest results

>857 bp linear fragment from linear panel NM_01282692.1:154-1764 Homo sa
 AATTCTCATAGCGCGCAATATAAGCATGCTCAGATATATTTAAGACACAGAGATGCTTGGTGA
 TTGGAATGGGAATTTCTGGCAGACAGGATCTGCTGTGGAGCGCAGC CACTTGGCGGAAAAGG
 TGCTCTCAGCACACCGGAGGAGGATGGGTGATGCACGCGAATATCTCTCTCTGGCGCTACC
 CATGGGACATGGTGTGTGATGACACCGTGTTCAGAACATGTTGAGAAATTCCTCCCTCCACCC
 CAATTGCTGACTTGCTGATGGACGCAAGATATAACACTGGTTCAGATCATCAATATACG
 TGATATACACATCATGCTGCTGTAAGAGAGTGTGATGAGAGAGTGTGATGAGAGAGTGTGAG
 GACGCATCATCATCGGGAAGTGTCTTCATGACGCAAGCATATAAAGAGAGTAAGAAGAACT
 CTGTCAATTATTAACATATCTCAAAGGAGAGGCGTATGACATCATGTTCTTGCTGCACCT
 GATACACATTTGCTTTCCCTCTTGCTGATGAGTCTGTAGTGAAGTGTGAAGATGGCCGAC
 CTGCTCATGTACAGATATATCTCTCGCATCATCTGCAAAAGCACCACTCGTGCCATTATGTG
 GCTCTACCAACCGCTCGCATCATGATCATCAGAGAGAAACACAGCTCGTGGGCTGT
 TGTGAGTCTGGAAGTGTGTAATCAAGTGTATACACCACAGAGTATGATAGAGAGAAATTA
 ATCTGAGAGAGAGAGAGTGTGTTGGCTGTGTGCTCATCGCAAGGCTTACCAT
 CAGATATCATCATCATATGATGAACCTGTGAGCTATATCATCTCAAGGCTCCAACTCTGT
 TCTCTGCTGCTCAAGC

>520 bp linear fragment from linear parent NM_001282692.1:154-1764 Homo sa
 AGTCAGGAGACAGTCGCCAAGCAGTGTGCCATTGTGGGAGCTGGGGTCTCAGGCGCTGGCC
 TCATCAAGTGTCTGTGAGAAAGCACTGTGGAGCCAGCTCCTCTGAGGAGGAGAGCTGATC
 CTGGGGGGCTGTGAGATTCAGGCAGATGTTGTGAAGAGAGCGAGAGCGCTCTCAGAC
 TCTGTGTTGTTGTCAGGAGATCTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
 GATTATCCAAACTGATGTGCGAAATCTCAATTTCTGGAAATCTCTCAAAATGTATCTGCAAC
 CACTTTGACCTCTGAAACAACTCAATTCAGACCAAGATCTCGAGTGTACCAAAATGTC
 TCAGATTCTCTGTGCTGTGGCCATCTGGAGGTGGTGTGCTCATCTGAGGAGAGAGCAAGAG
 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
 CTGAGTCCCTTCTCAGGATGATATGCTCTTAAAGGCACT

>234 bp linear fragment from linear parent NM_001282692.1:154-1764 Homo sa
GATCCACATCTGGCTCTGACCGTCTCTTTTGCCCCATGCTCACCATTACCAAGTCCCGGTG
ACTGAGCCCAGGAAAAATGGGAAGGAGCCGAAATGCCATCATGACCCAGTGGGACCGAACA
TCAGAGGCACGACGAGGTGGGACGACGACGACGACGACGACGACGACGACGACGACGACG

Download from database the sequence NM_001282692.1

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

Input sequence ⓘ

Sequence type

☒ Protein ☐ DNA

Paste your first sequence here - or use the example sequence

```
>sp|Q01740|FMO1_HUMAN Flavin-containing monooxygenase 1 OS=Homo sapiens C
MAKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEGRASLYKSVVS
NSCKEMSCYSDFPFPEYDYPNYVPNSQFLEYLKMYANHFDLLKHIQFKTKVCSVTKCSDSA
VSGQWEVVTMHEEKQESAIFDAVMVCTGFLTNPYLPLDSFPGINAFKQYFHSRQYKHPD
IFKDKRVLVIGMNSGTDIAVEASHLAEKVFLSTTGGGWVISRIFDSGYPWDMVMFTRFQ
NMLRNSLPTPIVTWLMERKINNWLHNHANYGLIPEDRTQLKEFVLNDELPGRIITGKVIFR
PSIKEVKENSIVFNNTSKEEPIIIVFATGYTFAFPFLDES VVKVEDGQASLYKYIFPAH
LQSTLALQTLKPLQCMPTQCTQPMWVWVLYKQVWVLRDQVAFETWLRKVKQSMF
```

Zvolit soubor Nevybrán žádný soubor

Paste your second sequence here - or use the example sequence

```
>sp|P50285|FMO1_MOUSE Flavin-containing monooxygenase 1 OS=Mus musculus O
MVKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEGRASLYKSVVS
NSSREMSCYPDFPFPEYDYPNFVPSLFLLEYLKLYSTQFNLQRCIYFNTKVCISITKRPDFA
VSGQWEVVTVTNGKQNSAIFDAVMVCTGFLTNPHLPLDSFPGILTFKGEYFHSRQYKHPD
IFKDKRVLVVG MNSGTDIAVEASHLAKKVFLSTTGGAWVISRVFDSGYPWDMIFMTRFQ
NMLRNLLPTPIVSWLISKMNWFWNHVNYGVAPEDRTQLREPVLNDELPGRIITGKVFIK
PSIKEVKENSIVFNNTPKKEEPIIIVFATGYTFAFPFLDES VVKVEDGQASLYKYIFPAH
```

```
-----
#
# Aligned_sequences: 2
# 1: FMO1_HUMAN
# 2: FMO1_MOUSE
# Matrix: EBL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 532
# Identity:      443/532 (83.3%)
# Similarity:    485/532 (91.2%)
# Gaps:          0/532 ( 0.0%)
# Score: 2396.0
#
#
#=====
```

FMO1_HUMAN	1	MAKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEG	50
		.	
FMO1_MOUSE	1	MVKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEG	50
FMO1_HUMAN	51	RASLYKSVVSNCKEMSCYSDFPFPEYDYPNYVPNSQFLEYLKMYANHFDL	100
FMO1_MOUSE	51	RASLYKSVVSNSSREMSCYPDFPFPEYDYPNFVPSLFLLEYLKLYSTQFNL	100
FMO1_HUMAN	101	LKHIQFKTKVCSVTKCSDSAVSGQWEVVTMHEEKQESAIFDAVMVCTGFL	150
		
FMO1_MOUSE	101	QRCIYFNTKVCISITKRPDFAVSGQWEVVTVTNGKQNSAIFDAVMVCTGFL	150

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)