Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2023/2024

Exam

Thursday 2.5.2024 16:30

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 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.

Previous Exams

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?

Find human protein sequence FGF3

What is the accession number and function of this protein?

How many valines does the sequence contain?

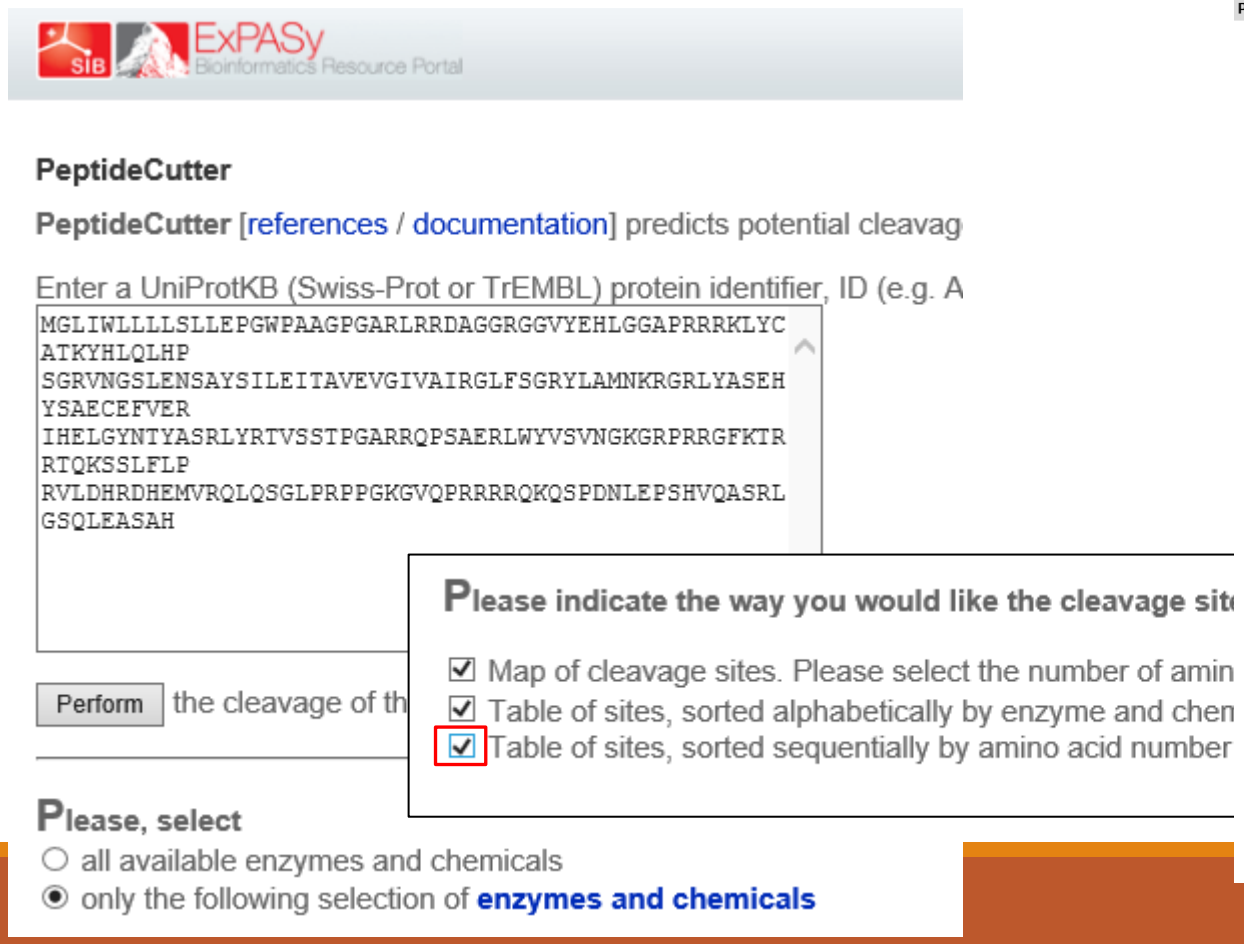
How long is the longest fragment after trypsin digestion?

How many human similar reference proteins are in the databases?

Does the sequence contain any transmembrane regions?

Find human protein sequence FGF3

How long is the longest fragment after trypsin digestion?



Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence (see explanations)	Peptide length [aa]	Peptide mass [Da]
24	Trypsin	MGLIWLLLLSLLEPGWPAAGPGAR	24	2532.085
26	Trypsin	LR	2	287.362
27	Trypsin	R	1	174.203
32	Trypsin	DAGGR	5	474.474
44	Trypsin	GGVYEHLGGAPR	12	1212.330
46	Trypsin	RR	2	330.390
47	Trypsin	K	1	146.189
53	Trypsin	LYCATK	6	697.847
63	Trypsin	YHLQLHPSGR	10	1207.357
89	Trypsin	VNGSLENSAYSILEITAVEVGIVAIR	26	2718.100
95	Trypsin	GLFSGR	6	635.721
101	Trypsin	YLAMNK	6	738.900
102	Trypsin	R	1	174.203
104	Trypsin	GR	2	231.255
120	Trypsin	LYASEHYSAECEFVER	16	1933.079
132	Trypsin	IHELGYNTYASR	12	1423.548
135	Trypsin	LYR	3	450.538
144	Trypsin	TVSSTPGAR	9	874.949
145	Trypsin	R	1	174.203
151	Trypsin	QPSAER	6	686.723
160	Trypsin	LWYVSVNGK	9	1065.237
164	Trypsin	GRPR	4	484.559
165	Trypsin	R	1	174.203
168	Trypsin	GFK	3	350.418
170	Trypsin	TR	2	275.308
171	Trypsin	R	1	174.203
174	Trypsin	TQK	3	375.425
181	Trypsin	SSLFLPR	7	818.971
186	Trypsin	VLDHR	5	638.724
192	Trypsin	DHEMVR	6	785.873
204	Trypsin	QLQSGLPRPPGK	12	1277.489
209	Trypsin	GVQPR	5	555.635
212	Trypsin	RRR	3	486.578
214	Trypsin	QK	2	274.320
229	Trypsin	QSPDNLEPSHVQASR	15	1664.752
239	end of sequence	LGSQLEASAH	10	1012.087

Find human protein sequence FGF3

What is the accession number and function of this protein?

How many valines does the sequence contain?

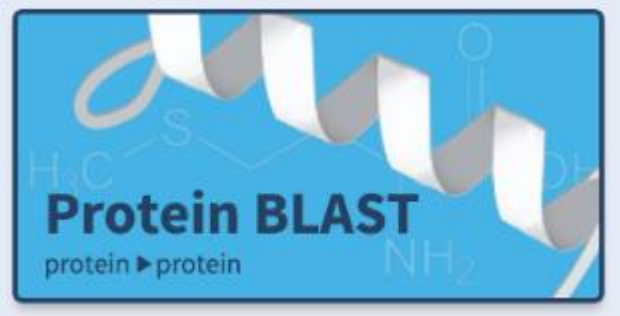
How long is the longest fragment after trypsin digestion?

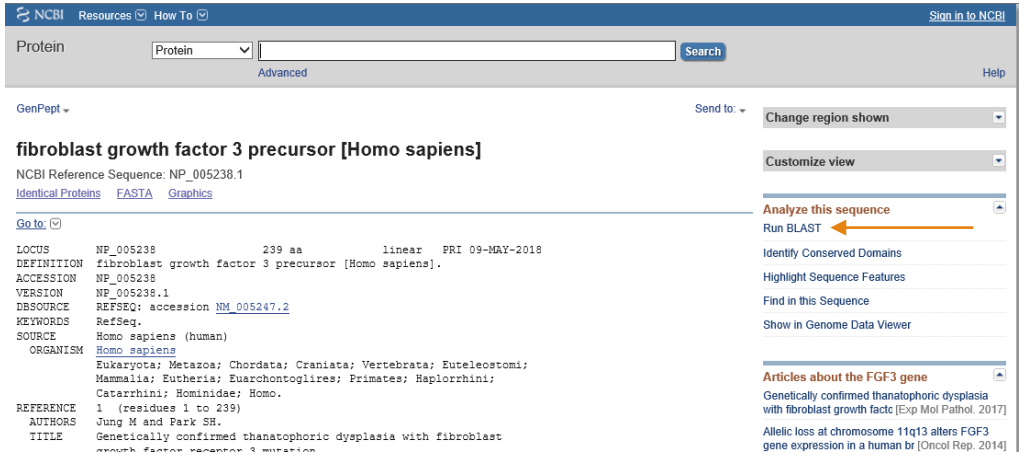
How many human similar reference proteins are in the databases?

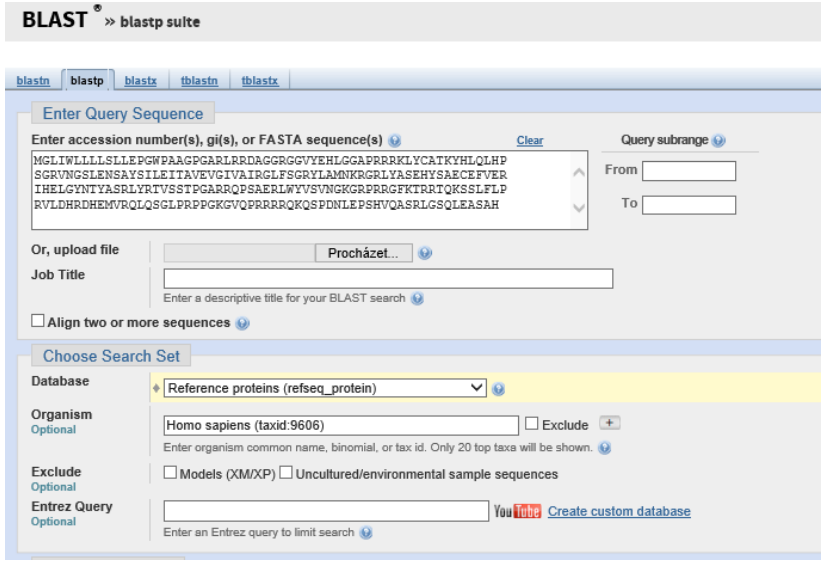
Does the sequence contain any transmembrane regions?

Find human protein sequence FGF3

How many human similar reference proteins are in the databases?

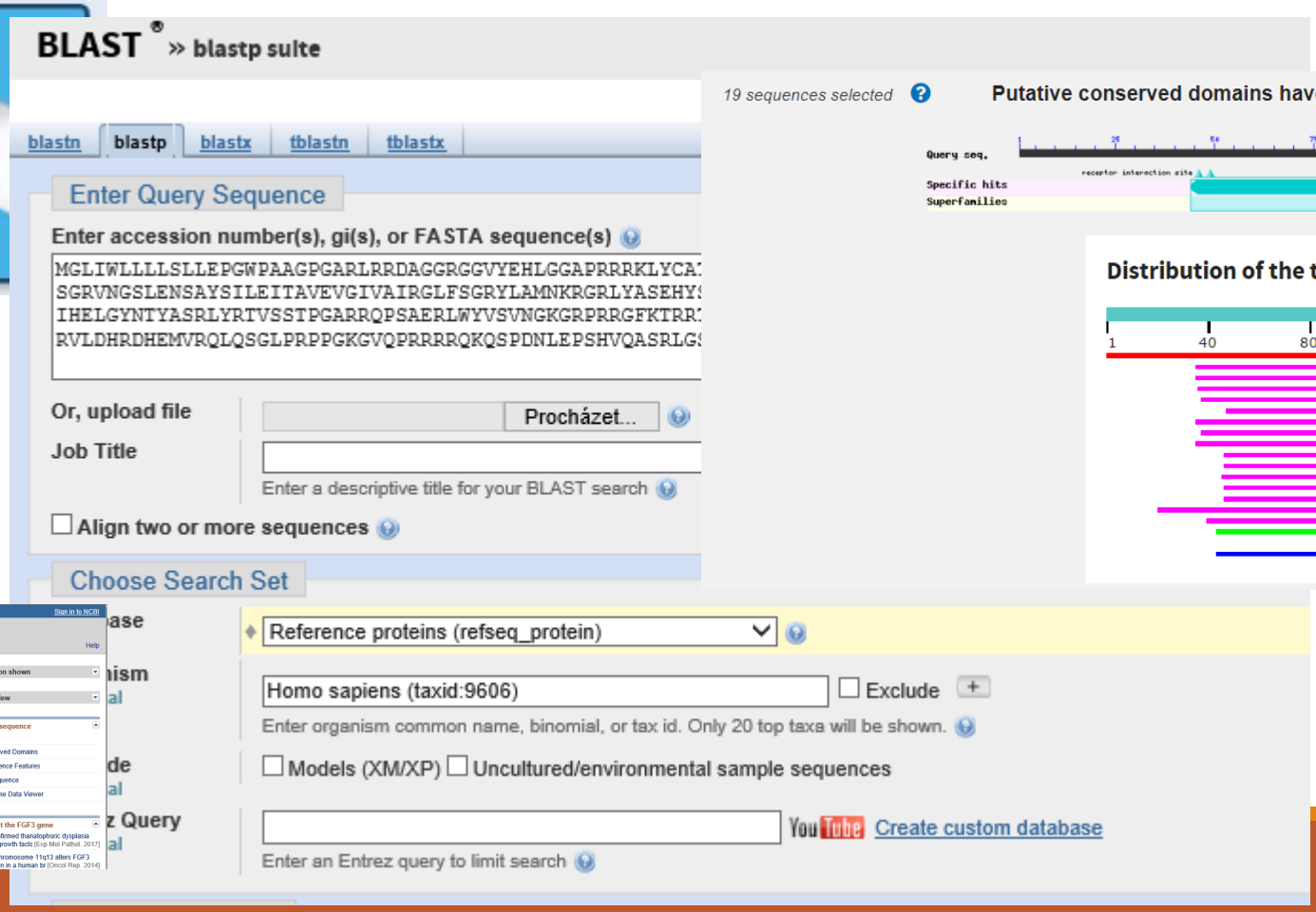


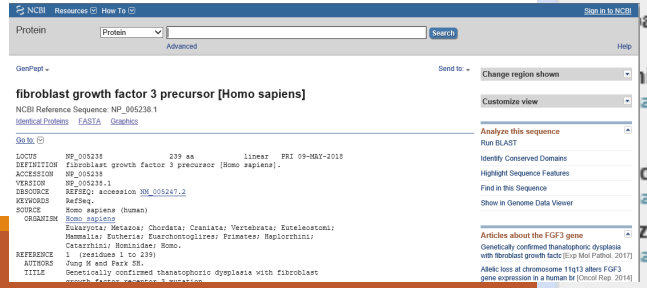




Find human protein sequence FGF3

How many human similar reference proteins are in the databases?





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?

Solution:

FGF3: P11487 (NP_005238.1), Fibroblast growth factor 3-role in the regulation of embryonic development…
Contains 13 valins.
Longest fragment after trypsin digestion has 26 aminoacids.
In the databes of reference sequences, there is 19 similar proteins.
There is no transmembrane helices.

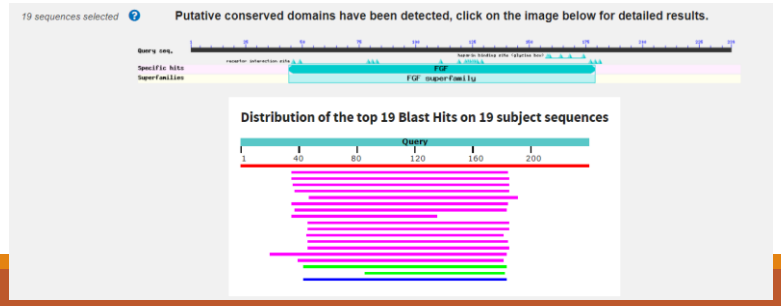
Find human protein sequence FGF3

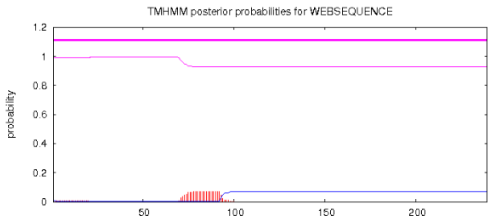
Solution:
FGF3: P11487 (NP_005238.1), Fibroblast growth factor 3-role in the regulation of embryonic development…
Contains 13 valins.
Longest fragment after trypsin digestion has 26 aminoacids.
In the databes of reference sequences, there is 19 similar proteins.
There is no transmembrane helices.

P11487 (FGF3_HUMAN) / NP_005238.1

Fibroblast growth factor 3
-role in the regulation of embryonic development…

V		13
Trypsin	VNGSLENSAYSILEITAVEVGIVAIR	26

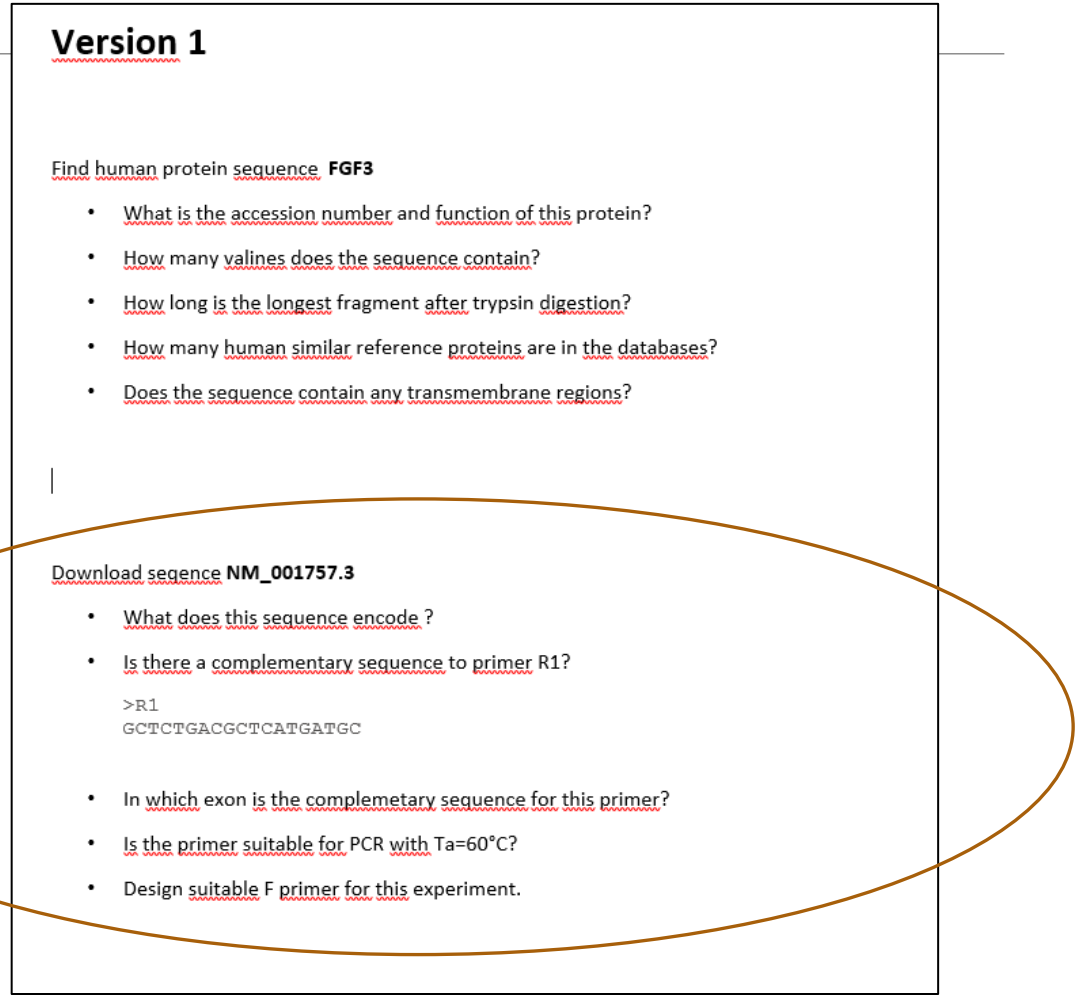




Exam test:

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 sequence **NM_001757.3**

What does this sequence encode ?

Is there a complementary sequence to primer R1?

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.

>R1
GCTCTGACGCTCATGATGC

Download seqence **NM_001757.3**

What does this sequence encode ?

Is there a complementary sequence to primer R1?

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.

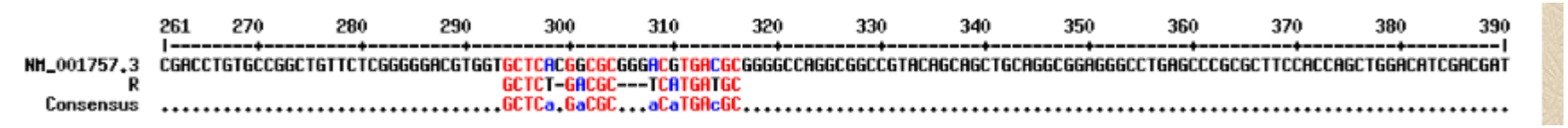
>R1
GCTCTGACGCTCATGATGC

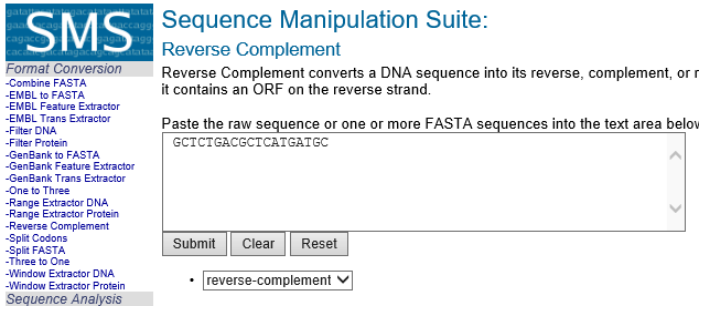
Download seqence **NM_001757.3**

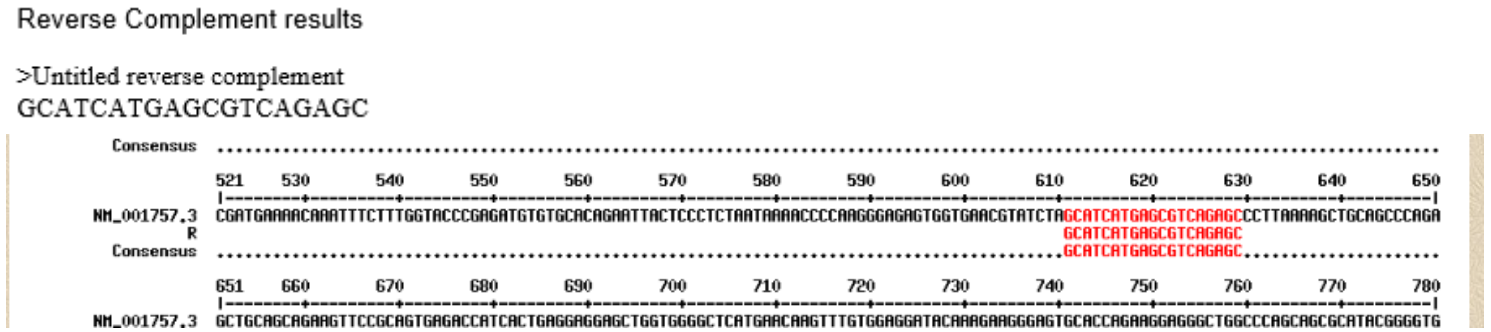
Is there a complementary sequence to primer R1?

>R1
GCTCTGACGCTCATGATGC

Multalin:







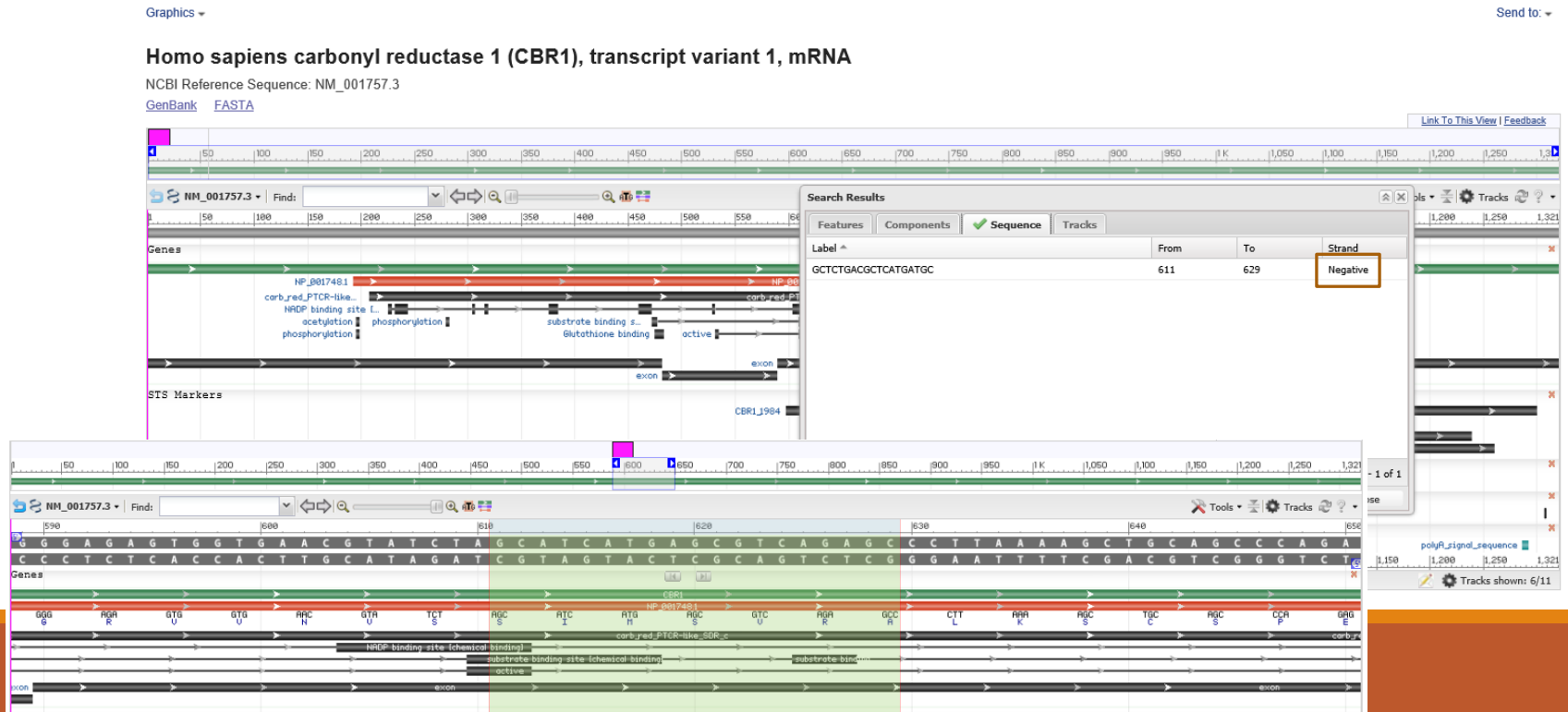
→yes.

Download seqence NM_001757.3

Is there a complementary sequence to primer R1?

>R1
GCTCTGACGCTCATGATGC

NCBI: Graphic



→yes.

Download seqence **NM_001757.3**

What does this sequence encode ?

Is there a complementary sequence to primer R1?

In which exon is the complemetary sequence for this primer?

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

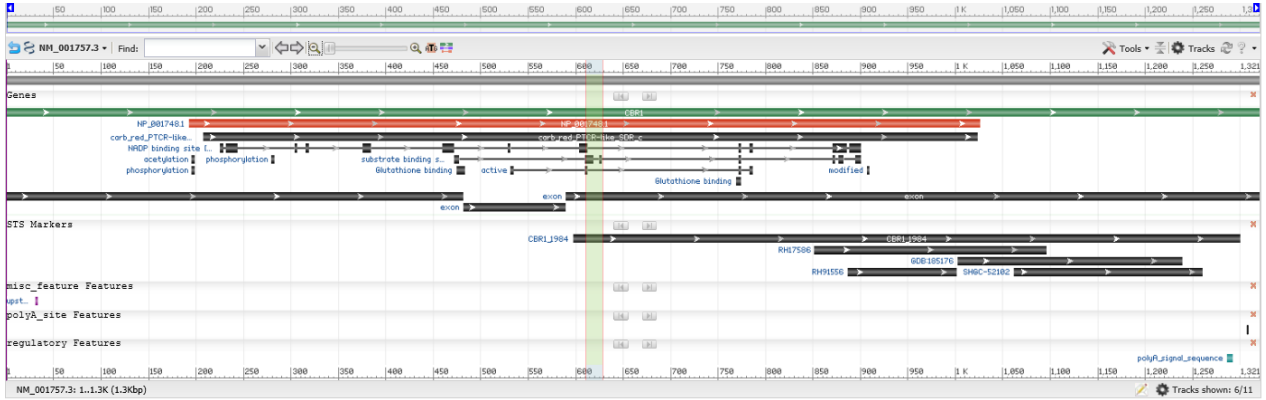
Design suitable F primer for this experiment.

>R1
GCTCTGACGCTCATGATGC

Download seqence **NM_001757.3**

In which exon is the complemetary sequence for this primer?

NCBI: Graphic



>R1
GCTCTGACGCTCATGATGC

→ third exon

Download seqence **NM_001757.3**

What does this sequence encode ?

Is there a complementary sequence to primer R1?

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.) We did not cover this part

>R1
GCTCTGACGCTCATGATGC

Download seqence **NM_001757.3**

Solution:

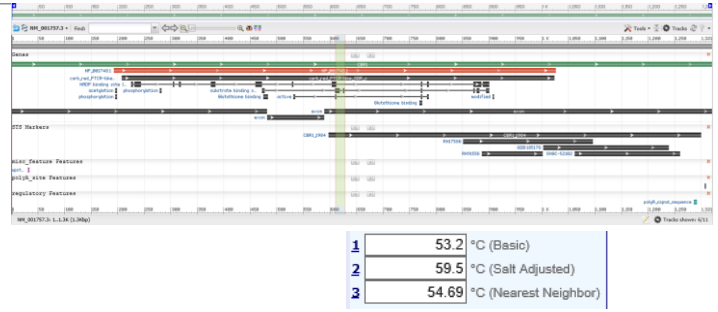
This sequence encodes carbonyl reductase 1 (CBR1).

Yes, there is complementary sequence to primer R1.

It is within 3rd exon.

At Ta=60°C the primer should work.

An example of suitable F primer: F:CAAGGTTGCTGATCCCACAC



1	53.2	°C (Basic)
2	59.5	°C (Salt Adjusted)
3	54.69	°C (Nearest Neighbor)

Homo sapiens carbonyl reductase 1 (CBR1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_001757.3

FASTA Graphics

```
                                                                                                                    650
NM_001757.3  CGATGAAAACAAATTTCTTTGGTACCCGAGATGTGTGCACAGAATTACTCCCTCTAATAAAACCCCAAGGGAGAGTGGTGAACGTATCTAGCATCATGAGCGTCAGAGCCCTTAAAAGCTGCAGCCCAGA
R                                                                                                                   GCATCATGAGCGTCAGAGC
Consensus    ..................................................................................................GCATCATGAGCGTCAGAGC..................

                 651       660       670       680       690       700       710       720       730       740       750       760       770       780
                 |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_001757.3  GCTGCAGCAGAAGTTCCGCAGTGAGACCATCACTGAGGAGGAGCTGGTGGGGCTCATGAACAAGTTTGTGGAGGATACAAAGAAGGGAGTGCACCAGAAGGAGGGCTGGCCCAGCAGCGCATACGGGGTG
```

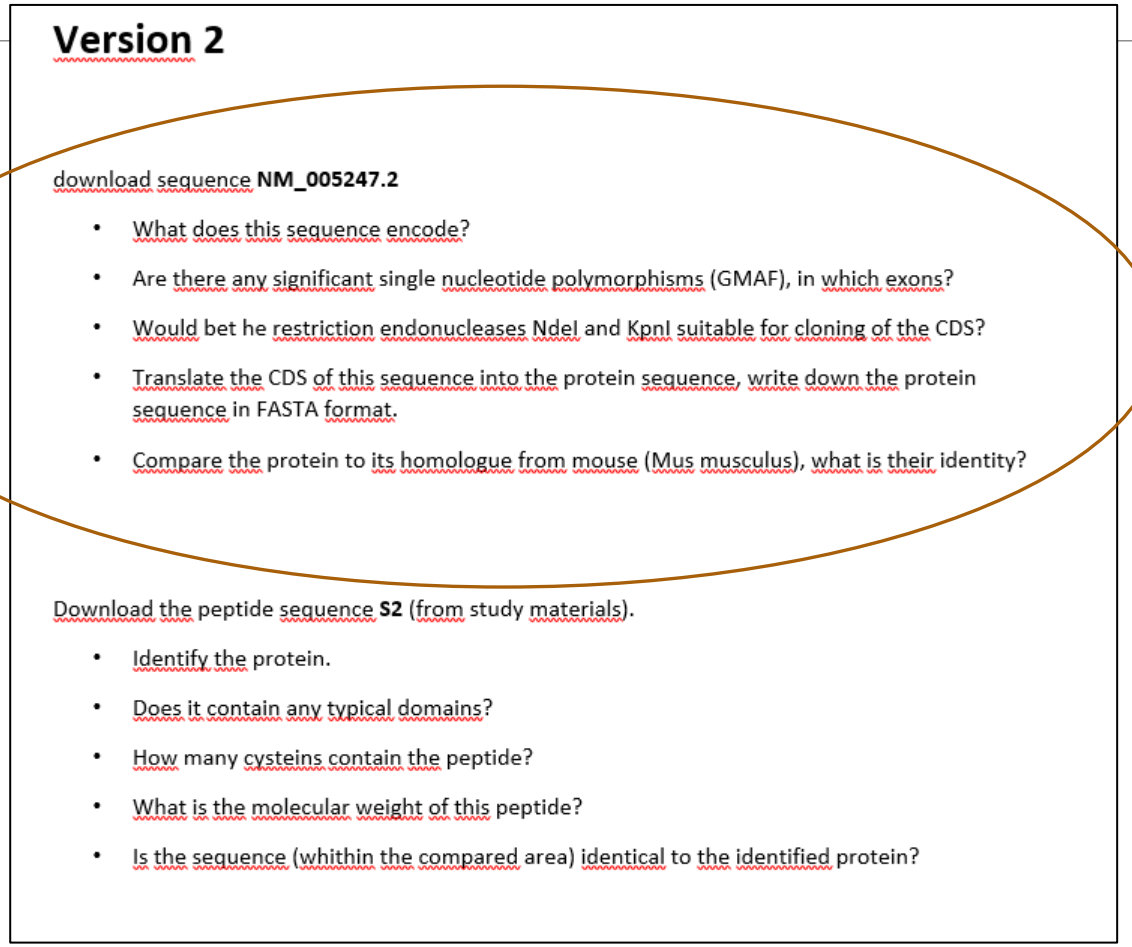
```
Template masking not selected
No mispriming library specified
Using 1-based sequence positions
OLIGO             start  len      tm      gc%   any th  3' th  hairpin  seq
LEFT PRIMER         477   20   59.12   55.00     0.00   0.00     0.00  CAAGGTTGCTGATCCCACAC
RIGHT PRIMER        629   19   59.07   57.89     0.00   1.25     0.00  GCTCTGACGCTCATGATGC
SEQUENCE SIZE: 1321
INCLUDED REGION SIZE: 1321

PRODUCT SIZE: 153, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00
```

Exam test II:

2 parts, each five points

Maximum 10 points

8.5-10p:	1
7-8p:	2
5-6.5p:	3



Version 2

download sequence **NM_005247.2**

- What does this sequence encode?
- Are there any significant single nucleotide polymorphisms (GMAF), in which exons?
- Would bet he restriction endonucleases NdeI and KpnI suitable for cloning of the CDS?
- Translate the CDS of this sequence into the protein sequence, write down the protein sequence in FASTA format.
- Compare the protein to its homologue from mouse (Mus musculus), what is their identity?

Download the peptide sequence **S2** (from study materials).

- Identify the protein.
- Does it contain any typical domains?
- How many cysteins contain the peptide?
- What is the molecular weight of this peptide?
- Is the sequence (whithin the compared area) identical to the identified protein?

Download sequence **NM_005247.2**

What does this sequence encode?

(Are there any significant single nucleotide polymorphisms (GMAF), in which exons?)

We did not cover this part (or it does not work anymore)

Would be the restriction endonucleases NdeI and KpnI suitable for cloning of the CDS?

Translate the CDS of this sequence into the protein sequence, write down the protein sequence in FASTA format.

Compare the protein to its homologue from mouse (Mus musculus), what is their identity?

Download sequence **NM_005247.2**

What does this sequence encode?

Are there any significant single nucleotide polymorphisms (GMAF), in which exons?

Would be the restriction endonucleases NdeI and KpnI suitable for cloning of the CDS?

Translate the CDS of this sequence into the protein sequence, write down the protein sequence in FASTA format.

Compare the protein to its homologue from mouse (Mus musculus), what is their identity?

Restriction summary:

KpnI ggtac|c	289
NdeI ca|tatg	none

→ KpnI is not suitable

Download sequence **NM_005247.2**

What does this sequence encode?

Are there any significant single nucleotide polymorphisms (GMAF), in which exons?

Would be the restriction endonucleases NdeI and KpnI suitable for cloning of the CDS?

Translate the CDS of this sequence into the protein sequence, write down the protein sequence in FASTA format.

Compare the protein to its homologue from mouse (Mus musculus), what is their identity?

Solution:
This gene encodes Fibroblast Growth Factor 3
GMAF are in the 1st and 3rd exon.
Restriction endonuclease KpnI is not suitable for cloning.
The FASTA of the protein is:
The indentity of human and mouse FGF3 proteins is 79.1 %.

```
>protein-FGF3
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH*
```

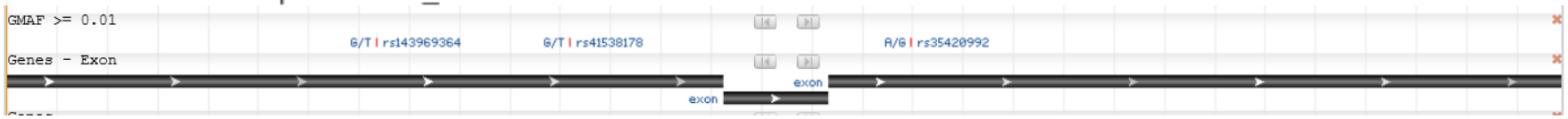
Download sequence **NM_005247.2**

<u>Solution:</u>
This gene encodes Fibroblast Growth Factor 3
GMAF are in the 1st and 3rd exon.
Restriction endonuclease KpnI is not suitable for cloning.
The FASTA of the protein is:
The indentity of human and mouse FGF3 proteins is 79.1 %.

```
>protein-FGF3
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH*
```

Homo sapiens fibroblast growth factor 3 (FGF3), mRNA

NCBI Reference Sequence: NM_005247.2



Ndel ca|tatg	none
KpnI ggtac|c	289

```
#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: FGF3_MOUSE
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 249
# Identity:     197/249 (79.1%)
# Similarity:   210/249 (84.3%)
# Gaps:          14/249 ( 5.6%)
# Score: 1021.5
#
#
#=======================================

EMBOSS_001         1 MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYC     50
                     ||||||||||||||.||..|||.||||||||||||||||||||||||||
FGF3_MOUSE         1 MGLIWLLLLSLLEPSWPTTGPGTRLRRDAGGRGGVYEHLGGAPRRRKLYC     50

EMBOSS_001        51 ATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMN    100
                     ||||||||||||||||||||||||||||||||||:|||:||||||||||||
FGF3_MOUSE        51 ATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGVVAIKGLFSGRYLAMN    100

EMBOSS_001       101 KRGRLYASEHYSAECEFVERIHELGYNTYASRLYRTVSSTPGARRQPSAE    150
                     ||||||||:||:||||||||||||||||||||||||.||.|||:|||.|:
FGF3_MOUSE       101 KRGRLYASDHYNAECEFVERIHELGYNTYASRLYRTGSSGPGAQRQPGAQ    150

EMBOSS_001       151 RLWYVSVNGKGRPRRGFKTRRTQKSSLFLPRVLDHRDHEMVRQLQSGLPR    200
                     |.||||||||||||||||||||||||||||||.|:||||||.|||..||
FGF3_MOUSE       151 RPWYVSVNGKGRPRRGFKTRRTQKSSLFLPRVLGHKDHEMVRLLQSSQPR    200

EMBOSS_001       201 PPGKGVQPRRRRQ-KQSPDNLEPSHVQASRLGSQLEASAH---------    239
                     .||:|.|||:||| ||||.:    |.:...|.::...|..
FGF3_MOUSE       201 APGEGSQPRQRRQKKQSPGD----HGKMETLSTRATPSTQLHTGGLAVA    245
```

Exam test:

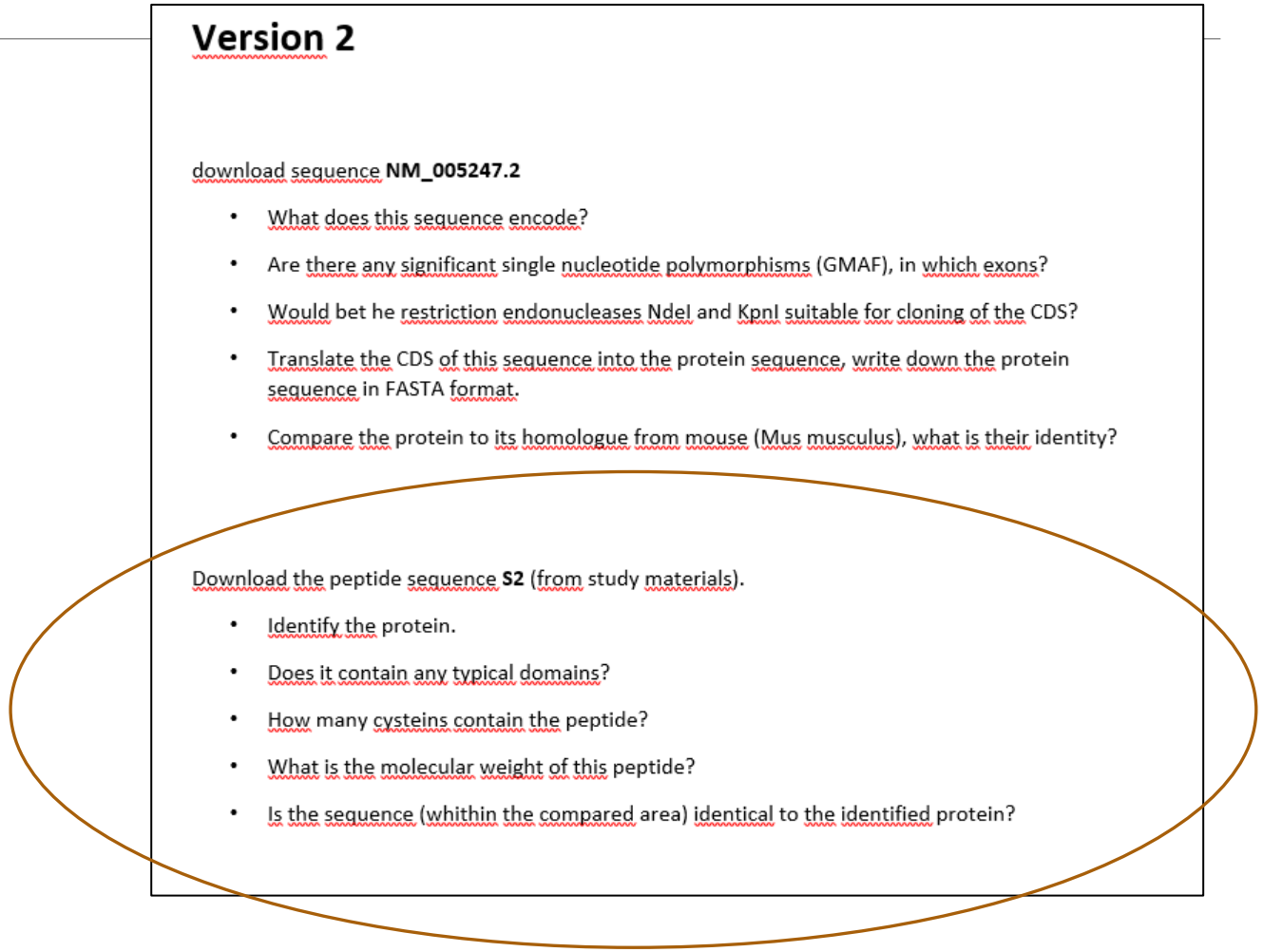
2 parts, each five points

Maximum 10 points

9-10p: 1

7-8p: 2

5-6p: 3



Version 2

download sequence **NM_005247.2**

- What does this sequence encode?
- Are there any significant single nucleotide polymorphisms (GMAF), in which exons?
- Would bet he restriction endonucleases NdeI and KpnI suitable for cloning of the CDS?
- Translate the CDS of this sequence into the protein sequence, write down the protein sequence in FASTA format.
- Compare the protein to its homologue from mouse (Mus musculus), what is their identity?

Download the peptide sequence **S2** (from study materials).

- Identify the protein.
- Does it contain any typical domains?
- How many cysteins contain the peptide?
- What is the molecular weight of this peptide?
- Is the sequence (whithin the compared area) identical to the identified protein?

Download the peptide sequence **S2** (from study materials).

Identify the protein.

Does it contain any typical domains?

How many cysteins contain the peptide?

What is the molecular weight of this peptide?

Is the sequence (whithin the compared area) identical to the identified protein?

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

Download the peptide sequence **S2** (from study materials).

Identify the protein.

Does it contain any typical domains?

How many cysteins contain the peptide?

What is the molecular weight of this peptide?

Is the sequence (whithin the compared area) identical to the identified protein?

Solution:
Peptide is probably human FosB homologue.
Yes, it contains leucine zipper bZIP.
Contains four cysteins
Mw=19.1 kDa.
It differs from the human homologue in the first and the last aminoacid.

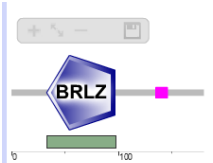
Download the peptide sequence **S2** (from study materials).

<u>Solution:</u>
Peptide is probably human FosB homologue.
Yes, it contains leucine zipper bZIP.
Contains four cysteins
Mw=19.1 kDa.
It differs from the human homologue in the first and the last aminoacid.

protein fosB isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_006723.2

Identical Proteins FASTA Graphics



Selected feature details

BRLZ domain

This is a SMART **BRLZ** domain (full annotation).

Position:	**33** to **97**
E-value:	**5.57849141230711e-13** (HMMER2)
SMART ACC:	SM000338
Definition:	basic region leucin zipper

Protein Stats results

Results for 180 residue sequence "Untitled" starting "CGPSTSGTTS"

Pattern:	Times found:	Percentage:
A	15	8.33
B	0	0.00
C	4	2.22
D	6	3.33

Protein Molecular Weight results

Results for 180 residue sequence "S2" starting "CGPSTSGTTS"

19.91 kDa

```
                    1       10        20        30        40        50        60        70        80        90       100       110       120       130
                    |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
sp|P53539|FOSB_HUMAN MFQAFPGDYDSGSRCSSSPSAESQYLSSVDSFGSPPTAAASQECAGLGEMPGSFVPTVTAITTSQDLQWLVQPTLISSMAQSQGQPLASQPPVVDPYDMPGTSYSTPGMSGYSSGGASGSGGPSTSGTTS
                S2                                                                                                                                CGPSTSGTTS
          Consensus ..................................................................................................................................cGPSTSGTTS

                    131     140       150       160       170       180       190       200       210       220       230       240       250       260
                    |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
sp|P53539|FOSB_HUMAN GPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPP
                S2 GPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPP
          Consensus GPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPP

                    261     270       280       290       300       310       320       330      338
                    |--------+---------+---------+---------+---------+---------+---------+---------|
sp|P53539|FOSB_HUMAN PPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSYTSSFVLTCPEVSAFAGAQRTSGSDQPSDPLNSPSLLAL
                S2 PPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
          Consensus PPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSc.................................
```

Version from 2023_new to Moodle

Version I 19.5.2023

A: Download the sequence NM_000946.3

1. What does this sequence encode?
2. How long is the coding sequence and how long will the translated protein be?
3. Manually design primers for PCR amplification of the CDS.
4. Compare the designed primers with the sequence of the mRNA and the CDS.
5. Is there a restriction enzyme that would cut the sequence of the third exon exactly once?

B: Work with following peptide sequence:

```
LALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACLRGIGRWT
NPQYRQFITILEATHRNQSSENKRQLANYNFDFRSWPVDFHWEEPSSRKESRGGPSRRGV
ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQG
QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP
LEACYSVLGWNHPGFAGSTGVPFPQNEANAMDVVVQFAIHRLGFQPQDIIIYAWSIGGFT
```

1) To which human protein this peptide probably belongs?
2) Does this peptide contain any transmembrane helices?
3) What is molecular weigth of this peptide?
4) How does a given peptide sequence differ from an identified (human) sequence (within the same segment)?
5) Write down the identified mutation.

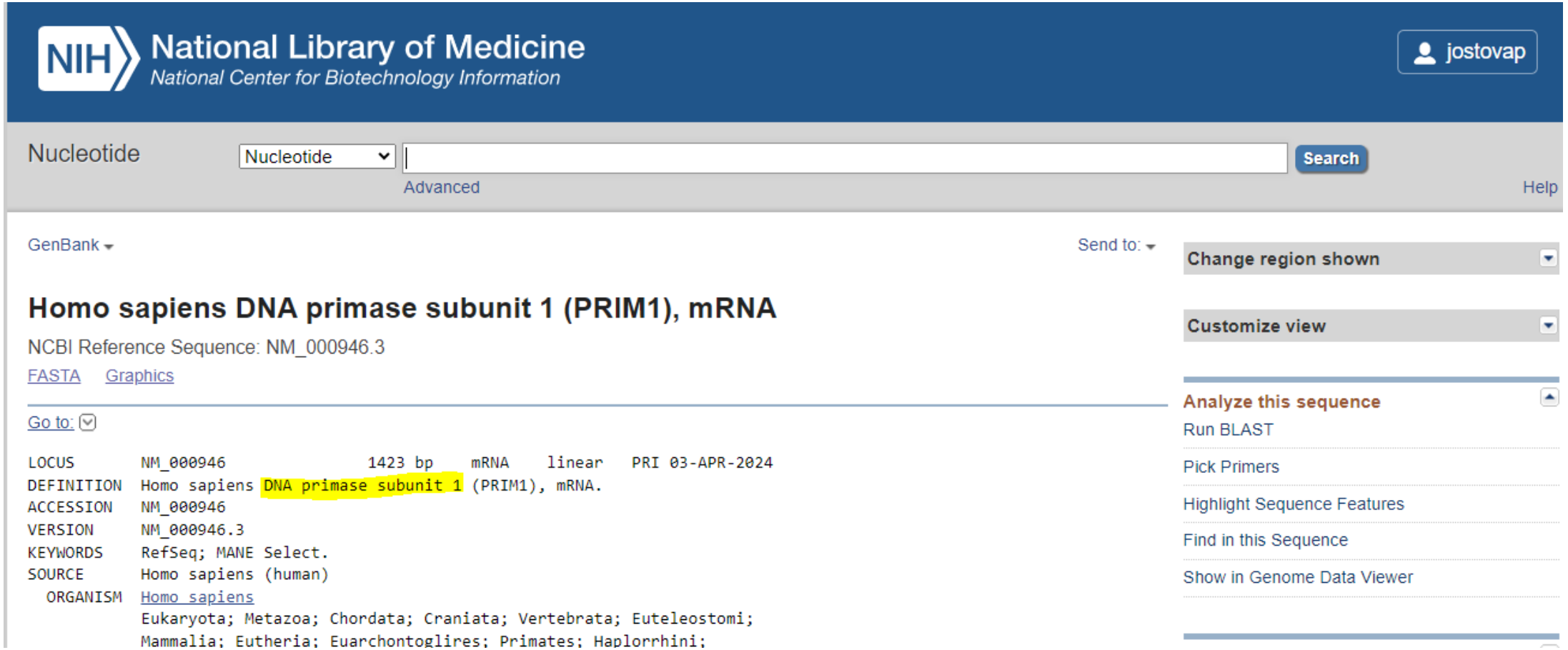
A: Download the sequence NM_000946.3

1. What does this sequence encode?
2. How long is the coding sequence and how long will the translated protein be?
3. Manually design primers for PCR amplification of the CDS.
4. Compare the designed primers with the sequence of the mRNA and the CDS.
5. Is there a restriction enzyme that would cut the sequence of the third exon exactly once?

A: Download the sequence NM_000946.3

1. What does this sequence encode?
2. How long is the coding sequence and how long will the translated protein be?
3. Manually design primers for PCR amplification of the CDS.
4. Compare the designed primers with the sequence of the mRNA and the CDS.
5. Is there a restriction enzyme that would cut the sequence of the third exon exactly once?

What does this sequence encode?



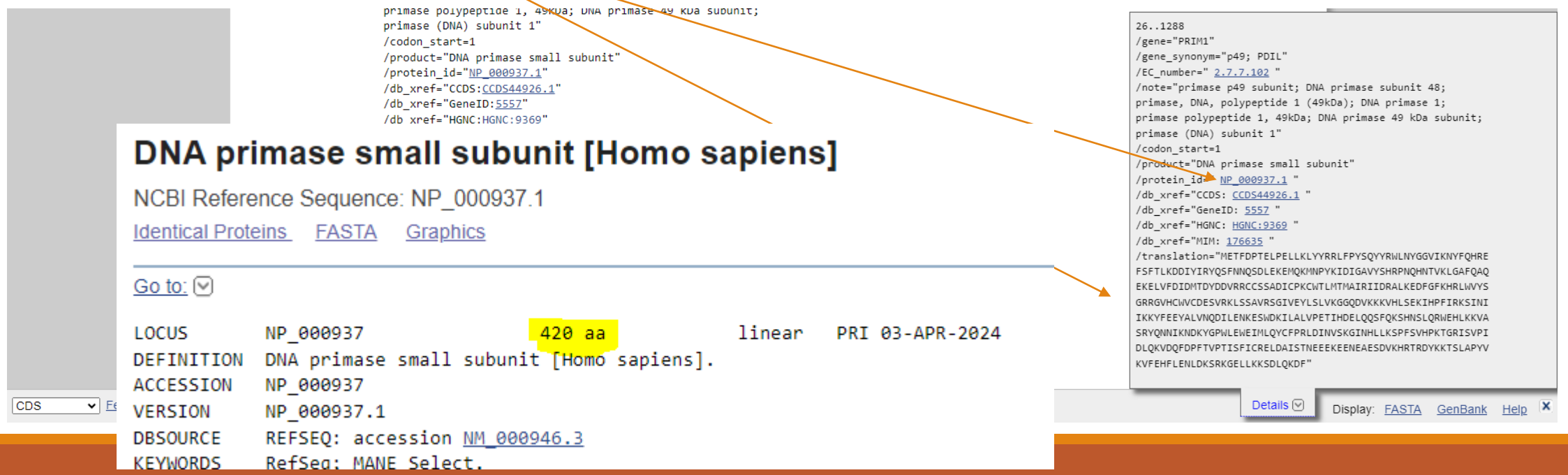
How long is the coding sequence and how long will the translated protein be?

CDS: CDS

```
/inference="alignment:Splign:2.1.0"
26..1288
/gene="PRIM1"
/gene_synonym="p49; PDIL"
/EC_number="2.7.7.102"
```

1) 1288-26+1= 1263nt (bp)

Translated protein



Other possibility:

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

FASTA Graphics

Go to:

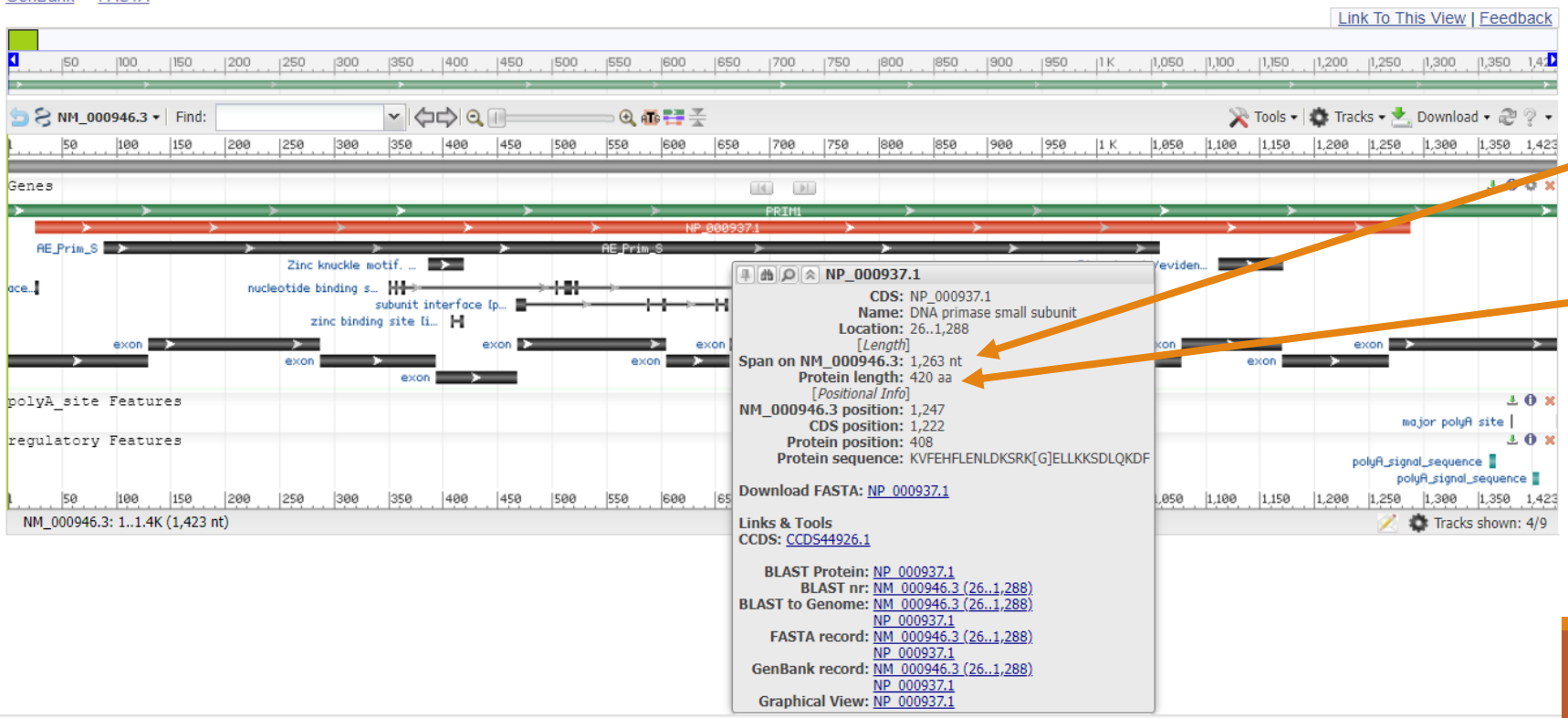
```
LOCUS       NM_000946              1423 bp    mRNA    linear   PRI 03-APR-2024
DEFINITION  Homo sapiens DNA primase subunit 1 (PRIM1), mRNA.
ACCESSION   NM_000946
```

Graphics Send to:

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

GenBank FASTA

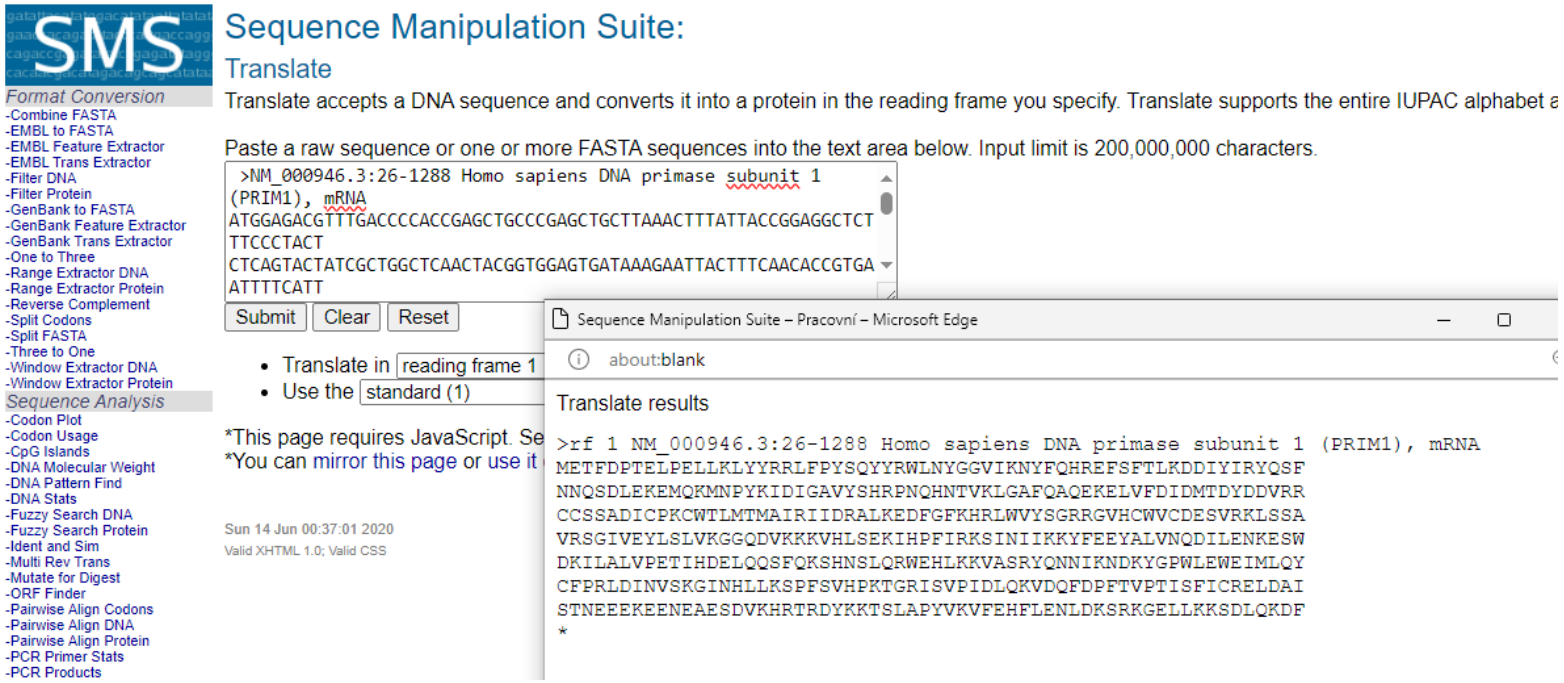


CDS

protein

Other option for translated protein:

Translate CDS:



And copy the protein to protein stats or range extractor protein Mw…all will tell you the length:

Range Extractor Protein results

>results for 420 residue sequence "Untitled" starting "METFDPTELP"
M

Protein Molecular Weight results

Results for 420 residue sequence "Untitled" starting "METFDPTELP"

49.91 kDa

Other option for translated protein: **math**

CDS=1263nt

Triplets for each AA: 1263/3=421 - 1 stop codon (no AA) = 420 AA

Manually design primers for PCR amplification of the CDS.

F: ATG GAG ACG TTT GAC CCC AC

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

GenBank Graphics

>NM_000946.3:26-1288 Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

ATGGAGACGTTTGACCCCACCGAGCTGCCCGAGCTGCTTAAACTTTATTACCGGAGGCTCTTTCCCTACT
CTCAGTACTATCGCTGGCTCAACTACGGTGGAGTGATAAAGAATTACTTTCAACACCGTGAATTTTCATT
CACATTGAAAGATGATATTTACATTCGCTACCAATCCTTCAACAACCAGAGTGATCTGGAAAAGGAGATG
CAGAAAATGAATCCATACAAGATTGATATAGGCGCAGTATATTCTCACAGACCCAATCAACACAATACAG
TGAAGCTGGGAGCTTTCCAGGCTCAGGAAAAAGAACTGGTATTTGACATTGACATGACAGACTATGACGA
TGTGAGGAGATGTTGTAGTTCTGCAGACATATGTCCTAAGTGCTGGACCCTCATGACAATGGCCATACGC
ATCATTGACAGAGCATTGAAGGAGGACTTTGGATTTAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAG
GTGTTCATTGTTGGGTCTGTGATGAATCAGTTAGAAAACTGTCTTCTGCAGTACGTTCTGGGATAGTTGA
GTATTTGAGCCTTGTAAAGGGTGGTCAAGACGTTAAAAAGAAAGTTCACCTAAGTGAAAAAATTCACCCT
TTTATCAGAAAATCTATAAACATAATAAAAAAATACTTTGAAGAATATGCCTTGGTTAATCAAGATATTC
TCGAAAATAAAGAAAGCTGGGATAAGATTTTAGCCCTTGTTCCTGAAACAATTCATGATGAACTTCAACA
AAGCTTCCAAAAGTCTCACAATTCACTTCAGCGTTGGGAGCACTTGAAGAAAGTAGCCAGCAGATATCAG
AATAACATCAAAAATGACAAATATGGACCCTGGCTGGAGTGGGAGATTATGCTCCAGTACTGTTTTCCAC
GGCTGGATATCAATGTCAGCAAAGGAATCAATCATCTACTGAAGAGCCCTTTTAGTGTTCATCCTAAAAC
AGGTCGCATATCTGTGCCTATTGATTTGCAGAAAGTGGACCAGTTTGATCCATTTACTGTTCCGACCATA
AGCTTCATCTGCCGTGAATTGGATGCCATTTCCACTAATGAAGAGGAAAAAGAGGAGAATGAAGCTGAAT
CTGATGTCAAACATAGAACCAGAGATTATAAGAAGACCAGTCTAGCACCTTATGTGAAAGTTTTTGAACA
TTTTCTTGAAAATCTGGATAAATCCCGAAAAGGAGAACTTCTTAAGAAGAGTGATTTACAAAAAGATTTC
TGA

Oligo Calc: Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD calculations are for single-stranded DNA or RNA

Nucleotide base codes

ATG GAG ACG TTT GAC CCC AC

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

GTG GGG TCA AAC GTC TCC AT

5' modification (if any) | 3' modification (if any) | Select molecule: ssDNA

50 nM Primer | 50 mM Salt (Na^+) | 1 Measured Absorbance at 260 nanometers

Calculate | Swap Strands | BLAST | mfold

Physical Constants

Length: 20 Molecular Weight: 6102.4 GC content: 55 %

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 4.718 microMolar.5 and contains 28.8 micrograms.

Melting Temperature (T_M) Calculations

1 53.8 °C (Basic)
2 60.5 °C (Salt Adjusted)
3 52.18 °C (Nearest Neighbor)

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.

RlnK 33.404 cal/(°K*mol) | deltaH 172.2 Kcal/mol
deltaG 26.6 Kcal/mol | deltaS 453.5 cal/(°K*mol)

Deprecated Hairpin/self dimerization calculations

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

Check Self-Complementarity

Citation: Kibbe WA. 'OligoCalc: an online oligonucleotide properties calculator'. (2007)

R:TCA GAA ATC TTT TTG TAA ATC ACT CTT

Reverse complement the whole CDS:

Reverse Complement results

>NM_000946.3:26-1288 Homo sapiens DNA primase subunit 1 (PRIM1),
TCAGAAATCTTTTTGTAAATCACTCTTCTTAAGAAGTTCTCCTTTTCGGGATTTATCCAG
ATTTTCAAGAAAATGTTCAAAAACTTTCACATAAGGTGCTAGACTGGTCTTCTTATAATC
TCTGGTTCTATGTTTGACATCAGATTCAGCTTCATTCTCCTCTTTTTCCTCTTCATTAGT
GGAAATGGCATCCAATTCACGGCAGATGAAGCTTATGGTCGGAACAGTAAATGGATCAAA
CTGGTCCACTTTCTGCAAATCAATAGGCACAGATATGCGACCTGTTTTAGGATGAACACT
AAAAGGGCTCTTCAGTAGATGATTGATTCCTTTGCTGACATTGATATCCAGCCGTGGAAA
ACAGTACTGGAGCATAATCTCCCACTCCAGCCAGGGTCCATATTTGTCATTTTTGATGTT
ATTCTGATATCTGCTGGCTACTTTCTTCAAGTGCTCCCAACGCTGAAGTGAATTGTGAGA
CTTTTGGAAGCTTTGTTGAAGTTCATCATGAATTGTTTCAGGAACAAGGGCTAAAATCTT
ATCCCAGCTTTCTTTATTTTCGAGAATATCTTGATTAACCAAGGCATATTCTTCAAAGTA
TTTTTTTATTATGTTTATAGATTTTCTGATAAAAGGGTGAATTTTTTCACTTAGGTGAAC
TTTCTTTTTAACGTCTTGACCACCCTTTACAAGGCTCAAATACTCAACTATCCCAGAACG
TACTGCAGAAGACAGTTTTCTAACTGATTCATCACAGACCCAACAATGAACACCTCTCCT
TCCAGAATATACCCAGAGACGATGCTTAAATCCAAAGTCCTCCTTCAATGCTCTGTCAAT
GATGCGTATGGCCATTGTCATGAGGGTCCAGCACTTAGGACATATGTCTGCAGAACTACA
ACATCTCCTCACATCGTCATAGTCTGTCATGTCAATGTCAAATACCAGTTCTTTTTCCTG
AGCCTGGAAAGCTCCCAGCTTCACTGTATTGTGTTGATTGGGTCTGTGAGAATATACTGC
GCCTATATCAATCTTGTATGGATTCATTTTCTGCATCTCCTTTTCCAGATCACTCTGGTT

Oligo Calc: Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD calculations are for single-stranded DNA or RNA

Nucleotide base codes

TCA GAA ATC TTT TTG TAA ATC ACT CTT

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

AAG AGT GAT TTA CAA AAA GAT TTC TGA

5' modification (if any) | 3' modification (if any) | Select molecule: ssDNA

50 nM Primer
50 mM Salt (Na^+)
1 Measured Absorbance at 260 nanometers

Calculate | Swap Strands | BLAST | mfold

Physical Constants

Length: 27 Molecular Weight: 8198.4 GC content: 26 %

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 3.55 microMolar and contains 29.1 micrograms.

Melting Temperature (T_M) Calculations

1 50.6 °C (Basic)
2 59.2 °C (Salt Adjusted)
3 52.83 °C (Nearest Neighbor)

Thermodynamic Constants Conditions: 1 M NaCl at 25ºC at pH 7.

RlnK 33.404 cal/(°K*mol) deltaH 203.6 Kcal/mol
deltaG 30.1 Kcal/mol deltaS 543 cal/(°K*mol)

Deprecated Hairpin/self dimerization calculations

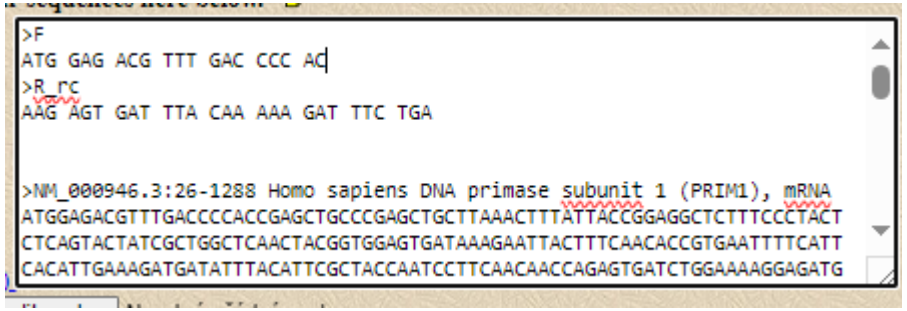
5 (Minimum base pairs required for single primer self-dimerization)
4 (Minimum base pairs required for a hairpin)

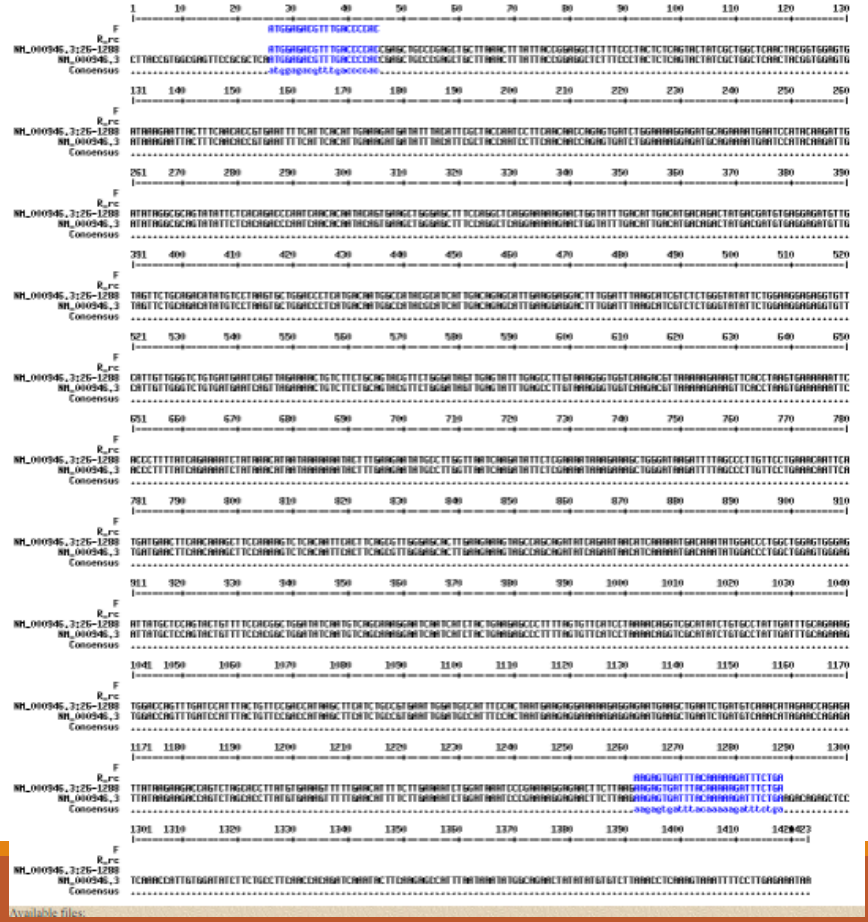
Check Self-Complementarity

Citation: Kibbe WA, 'OligoCalc: an online oligonucleotide properties calculator' (2007)

Compare the designed primers with the sequence of the mRNA and the CDS.

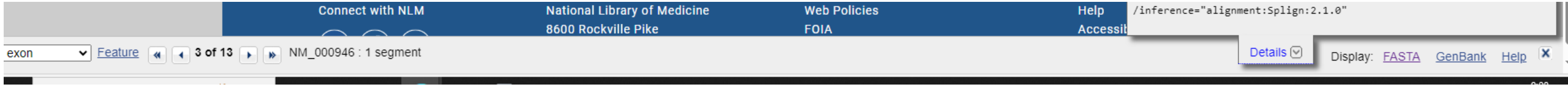
Multalin (don´t forget to reverse complement the primer for comparison





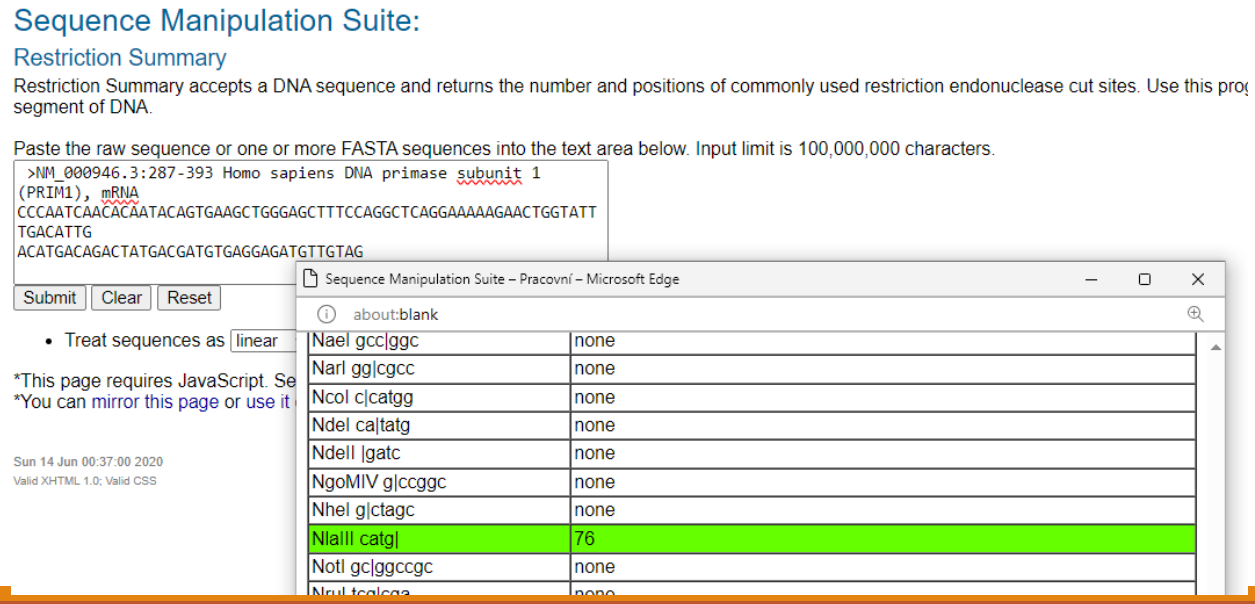
Is there a restriction enzyme that would cut the sequence of **the third exon** exactly once?

Get the 3rd exon:



Restriction summary:

Answer:yes NlaIII



B: Work with following peptide sequence:

LALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACLRGIGRWT

NPQYRQFITILEATHRNQSSENKRQLANYNFDFRSWPVDFHWEEPSSRKESRGGPSRRGV

ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQG

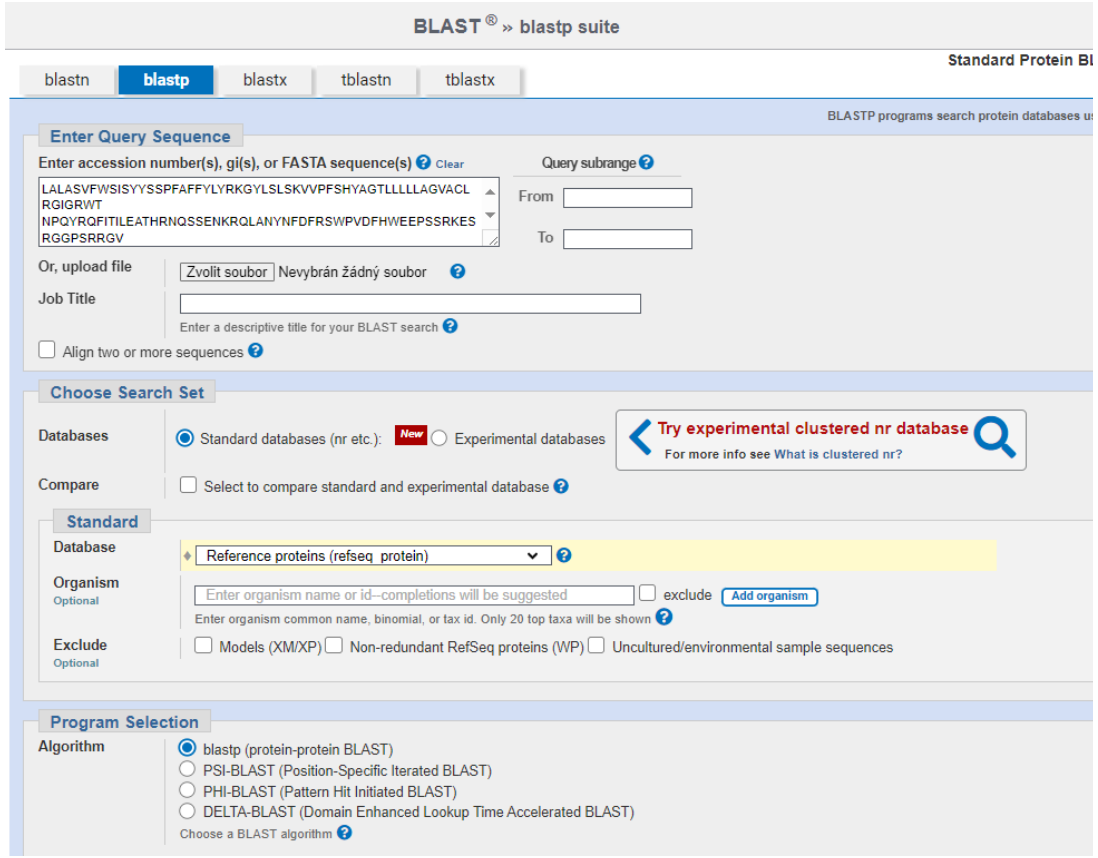
QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP

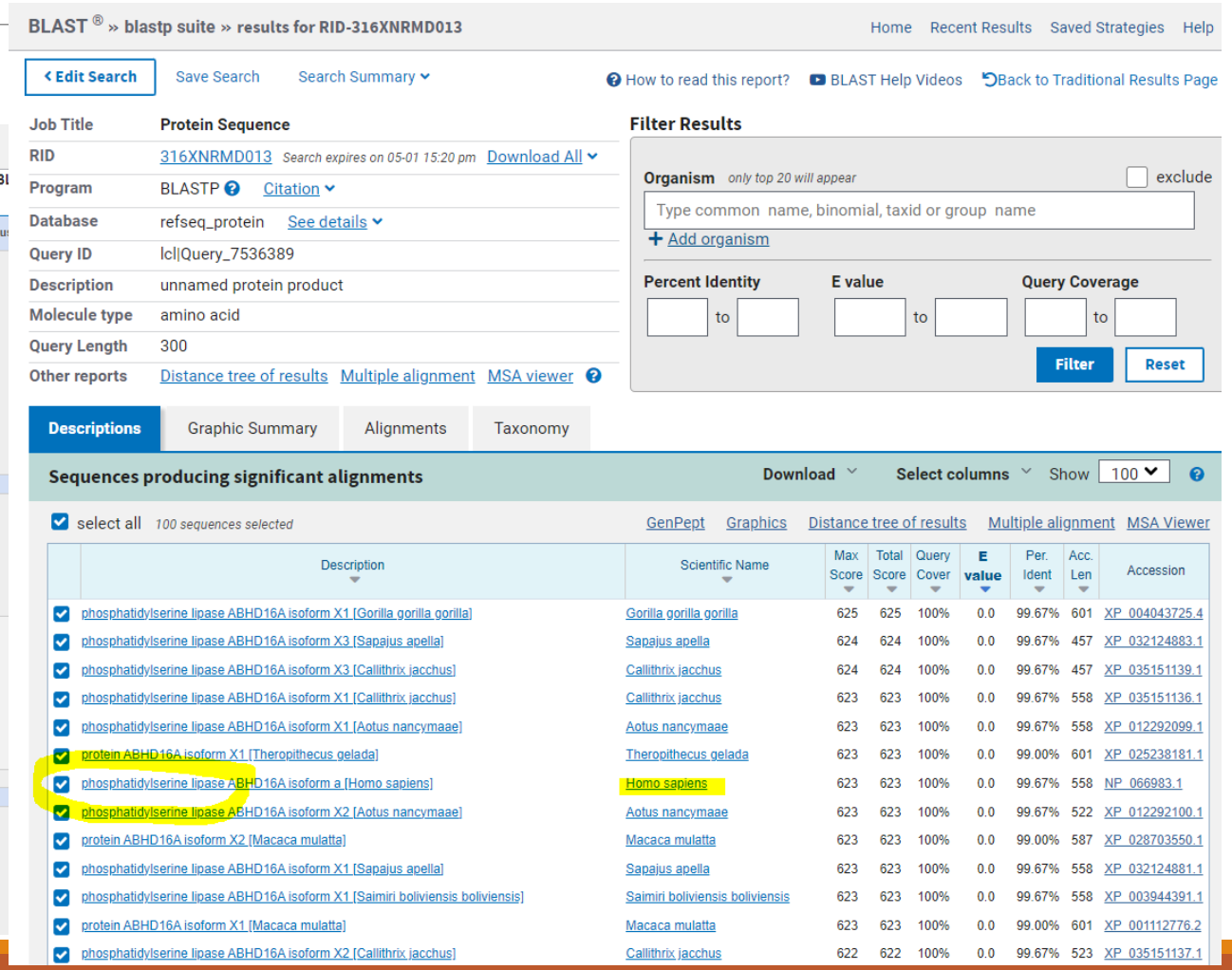
LEACYSVLGWNHPGFAGSTGVPFPQNEANAMDVVVQFAIHRLGFQPQDIIIYAWSIGGFT

1) To which human protein this peptide probably belongs?
2) Does this peptide contain any transmembrane helices?
3) What is molecular weigth of this peptide?
4) How does a given peptide sequence differ from an identified (human) sequence (within the same segment)?
5) Write down the identified mutation. -We did not cover this one.

To which human protein this peptide probably belongs?

BLASTp



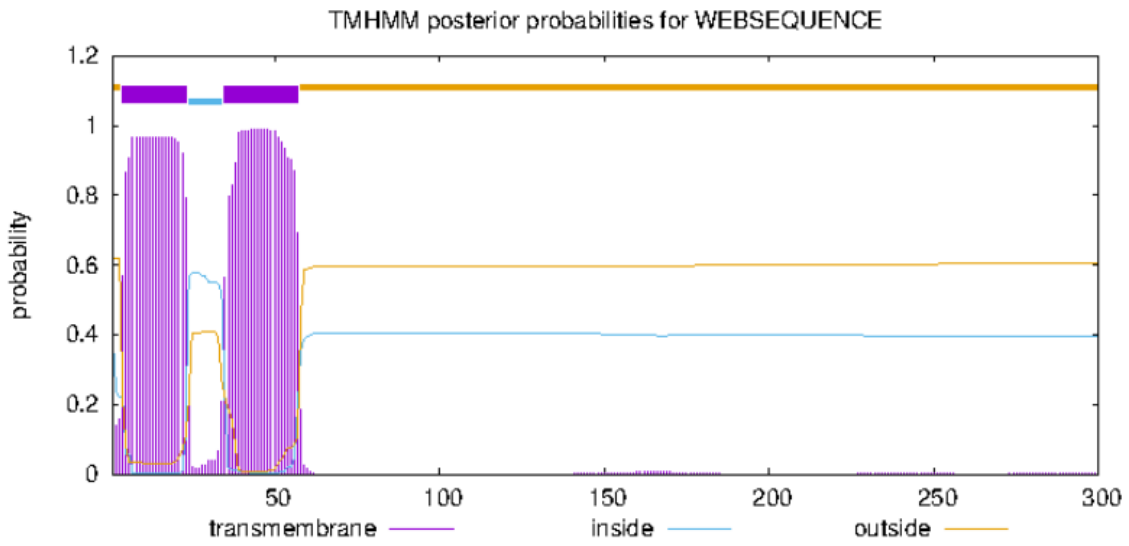


Does this peptide contain any transmembrane helices?

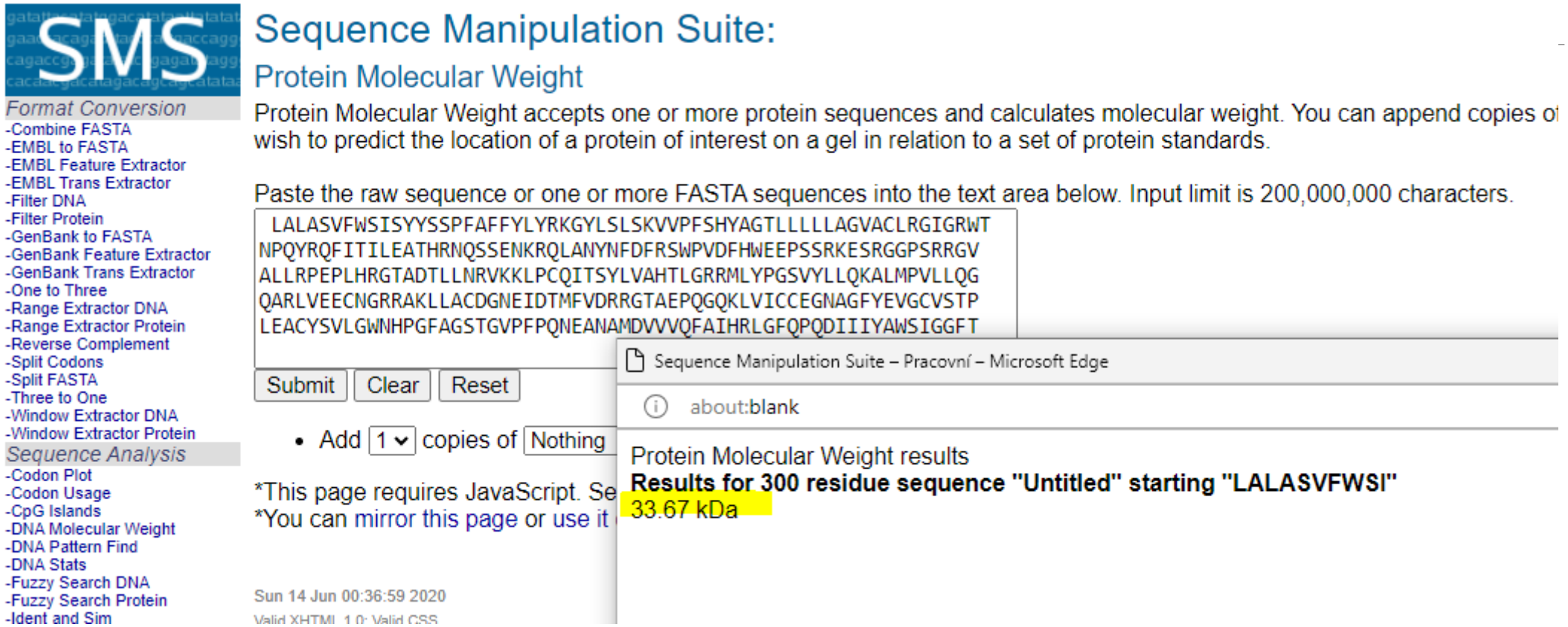
Yes 2.

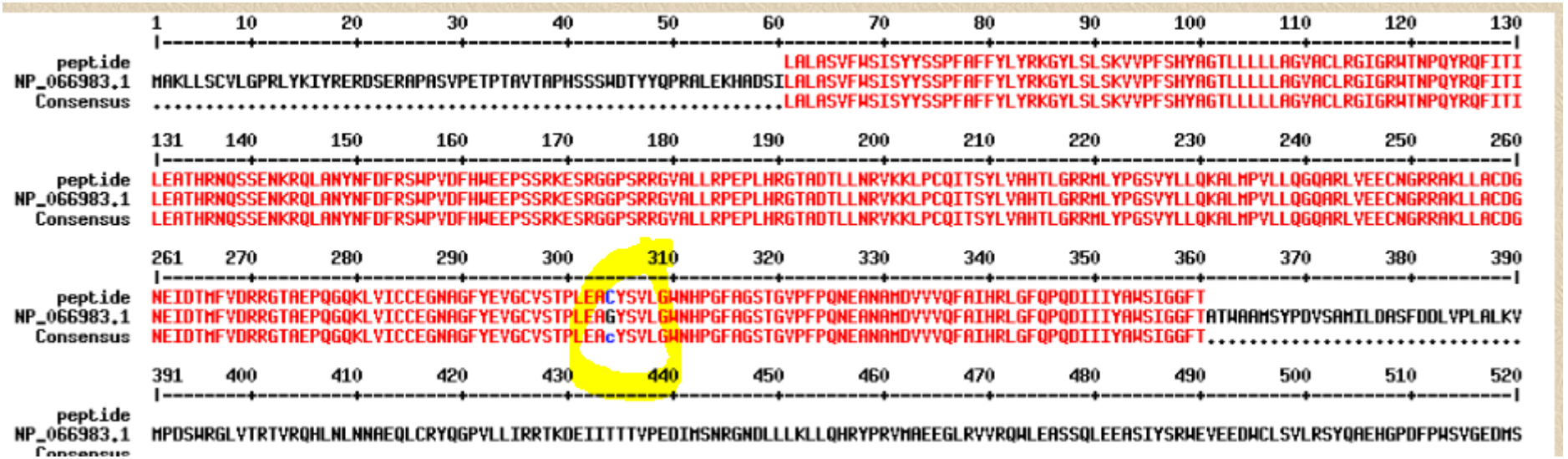
TMHMM result

```
# WEBSEQUENCE Length: 300
# WEBSEQUENCE Number of predicted TMHs:  2
# WEBSEQUENCE Exp number of AAs in TMHs: 41.2004100000001
# WEBSEQUENCE Exp number, first 60 AAs:  40.9446
# WEBSEQUENCE Total prob of N-in:        0.37925
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE     TMHMM2.0        outside      1     3
WEBSEQUENCE     TMHMM2.0        TMhelix      4    23
WEBSEQUENCE     TMHMM2.0        inside      24    34
WEBSEQUENCE     TMHMM2.0        TMhelix     35    57
WEBSEQUENCE     TMHMM2.0        outside     58   300
```



What is molecular weigth of this peptide?



How does a given peptide sequence differ from an identified (human) sequence (within the same segment)?

Differs in 1 aminoacid