Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2021/2022

10/10

Exam?

Thursday 14:00 in usual lecture time …

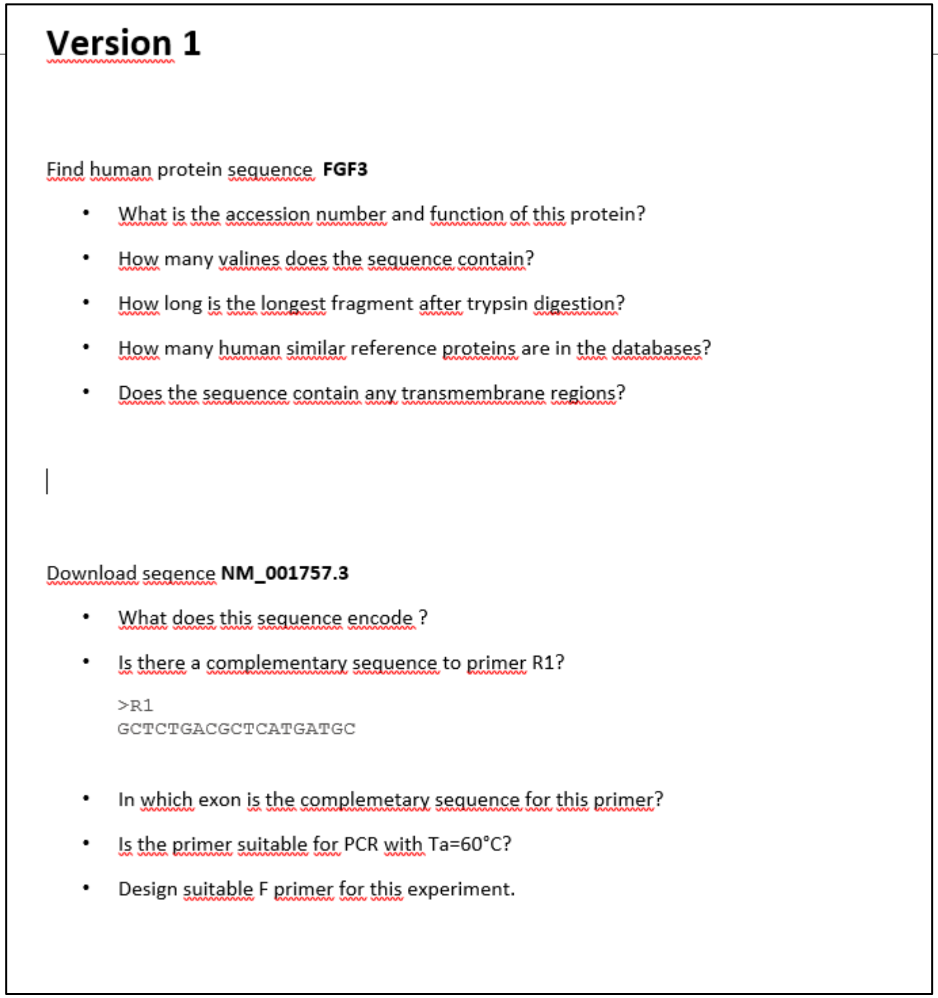
Moodle: files with versions
open „task“ for submisson (like for homeworks).

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.

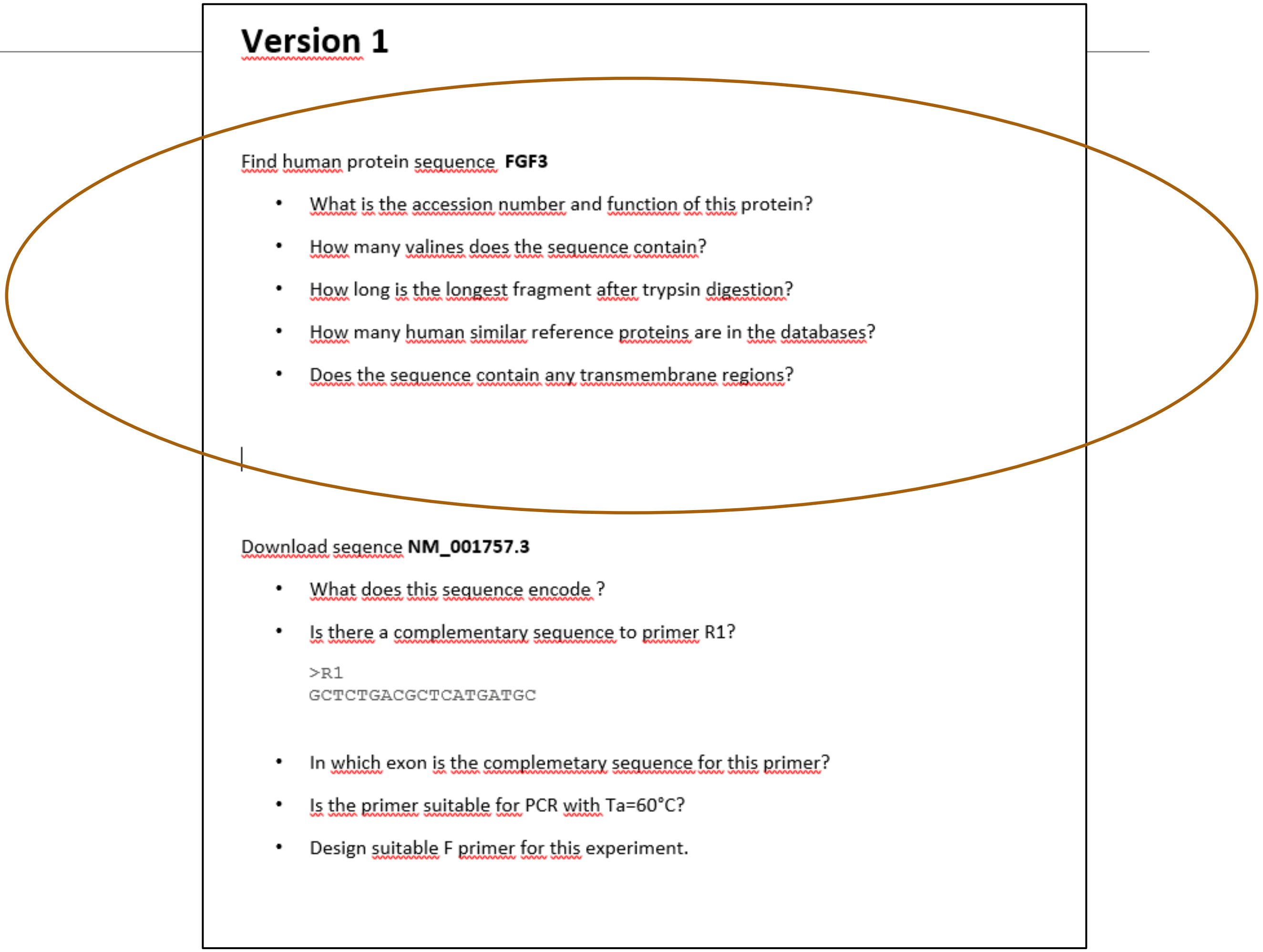
Previous Exams

Exam test:

2 parts, each five points

<u>Maximum 10 points</u>

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.

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**

What is the accession number and function of this protein?

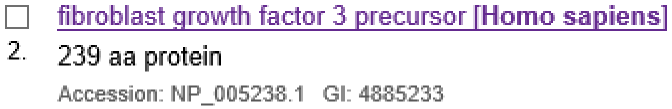
Uniprot:

☐	P11487	FGF3_HUMAN		**Fibroblast growth factor 3**	FGF3 INT2	Homo sapiens (Human)

→ role in the regulation of embryonic development, cell proliferation, and cell differentiation….

OR

NCBI/Protein:



☐ fibroblast growth factor 3 precursor [**Homo sapiens**]
2. 239 aa protein
Accession: NP_005238.1 GI: 4885233

Find human protein sequence **FGF3**

What is the accession number and function of this protein?

How many valines does the sequence contain?

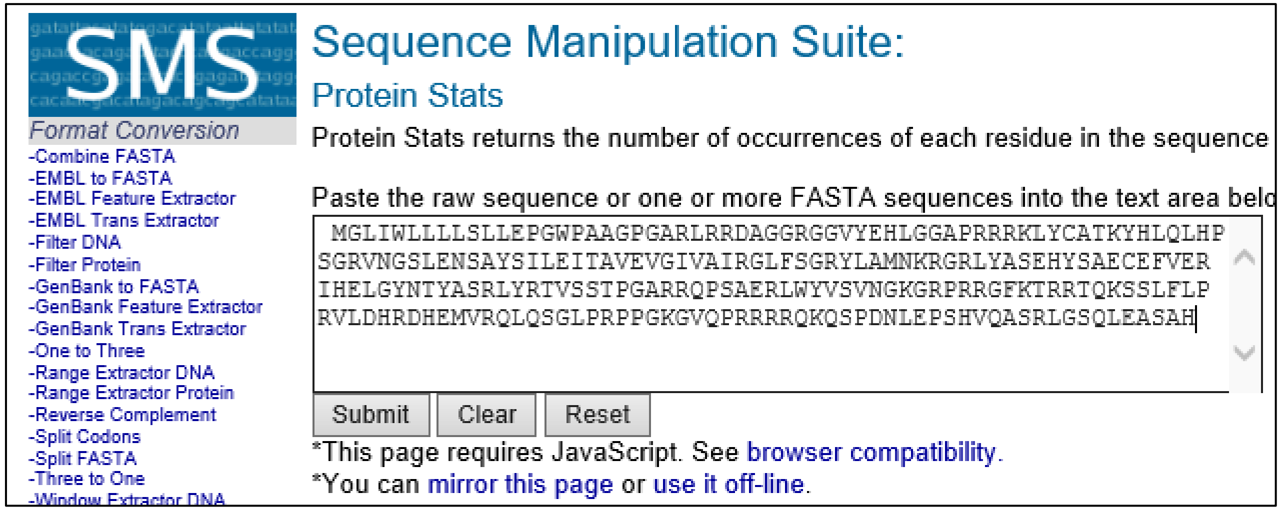
How long is the longest fragment after trypsin digestion?

How many human similar reference proteins are in the databases?

Does the sequence contain any transmembrane regions?

Find human protein sequence **FGF3**

How many valines does the sequence contain?



Sequence Manipulation Suite:

Protein Stats

Protein Stats returns the number of occurrences of each residue in the sequence

Paste the raw sequence or one or more FASTA sequences into the text area belo

```
 MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH
```

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

V	13	5.44

Find human protein sequence **FGF3**

What is the accession number and function of this protein?

How many valines does the sequence contain?

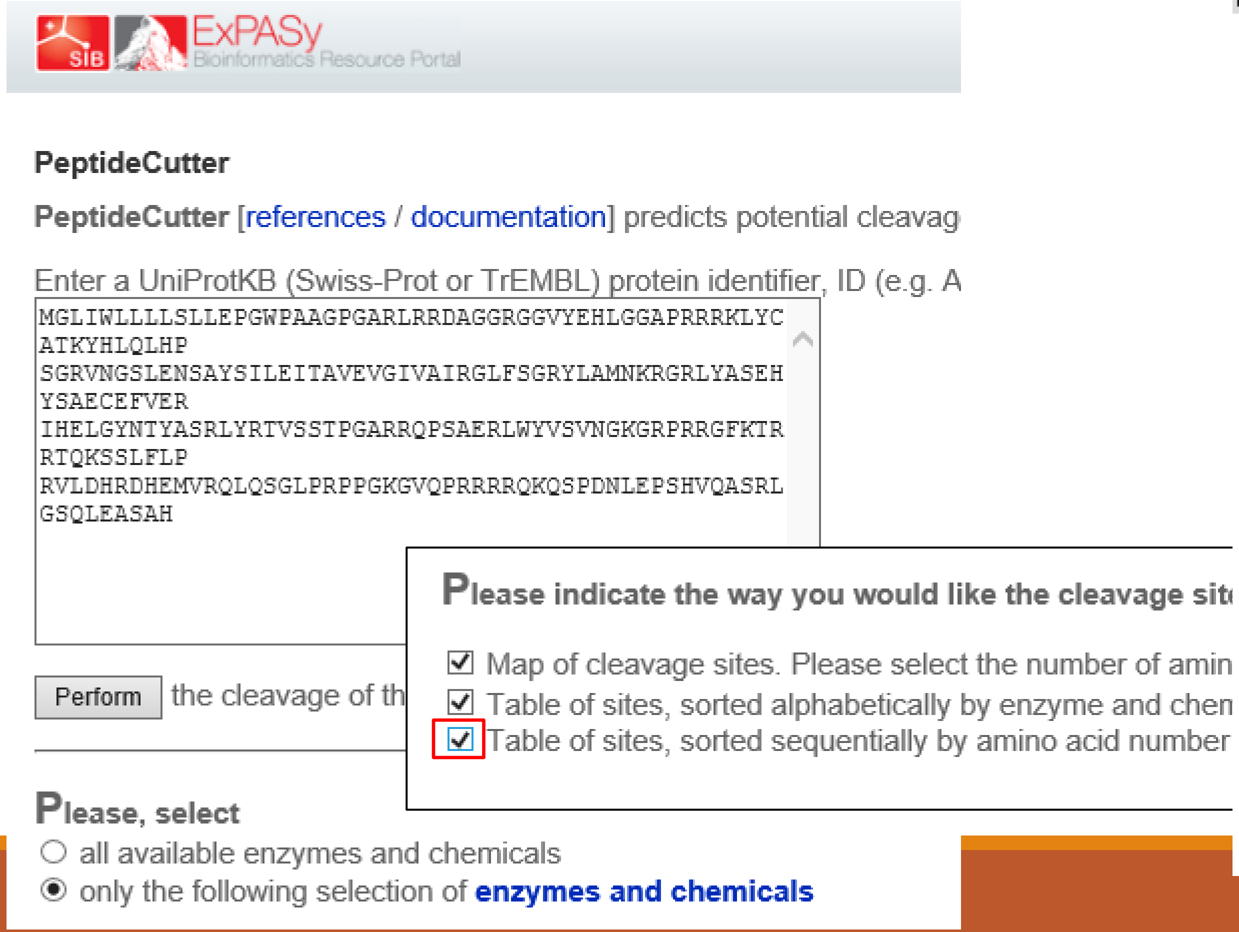
How long is the longest fragment after trypsin digestion?

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?



PeptideCutter

PeptideCutter [references / documentation] predicts potential cleavag

Enter a UniProtKB (Swiss-Prot or TrEMBL) protein identifier, ID (e.g. A

```
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYC
ATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEH
YSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTR
RTQKSSLFLP
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRL
GSQLEASAH
```

the cleavage of th

Please indicate the way you would like the cleavage sit

☑ Map of cleavage sites. Please select the number of amin

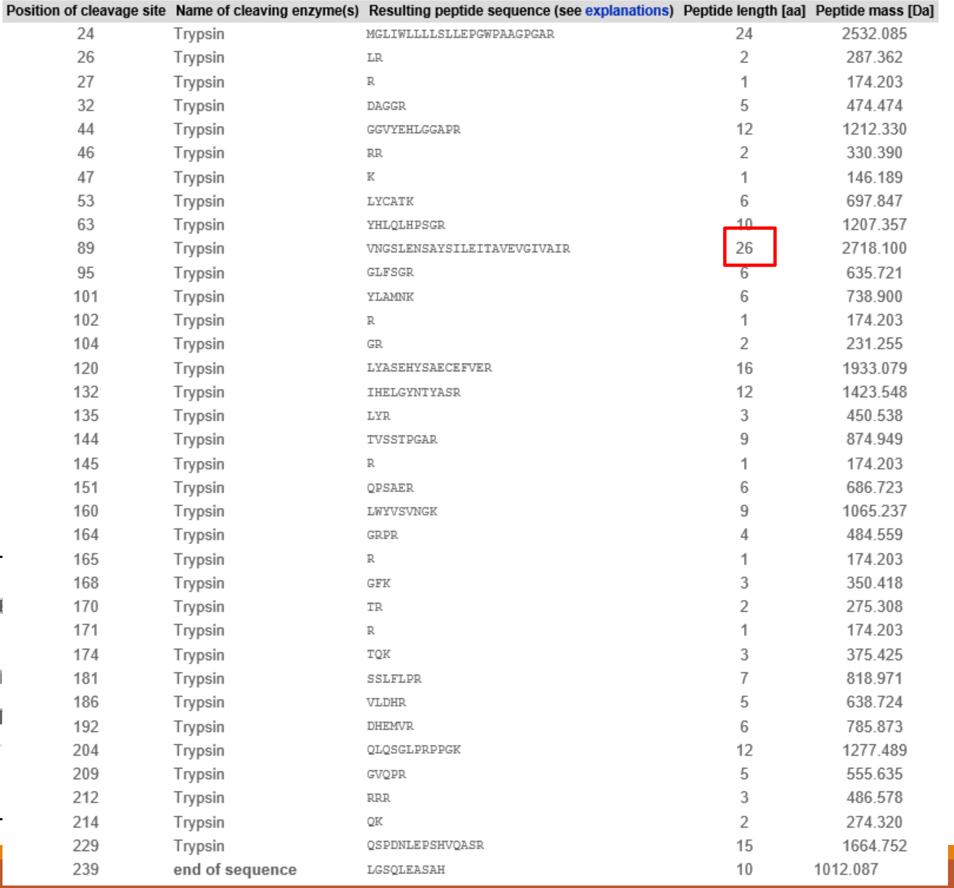
☑ Table of sites, sorted alphabetically by enzyme and chen

☑ Table of sites, sorted sequentially by amino acid number

Please, select

○ all available enzymes and chemicals

● only the following selection of **enzymes and chemicals**



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?

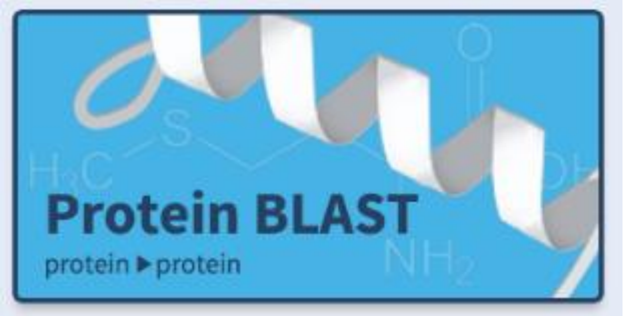
How long is the longest fragment after trypsin digestion?

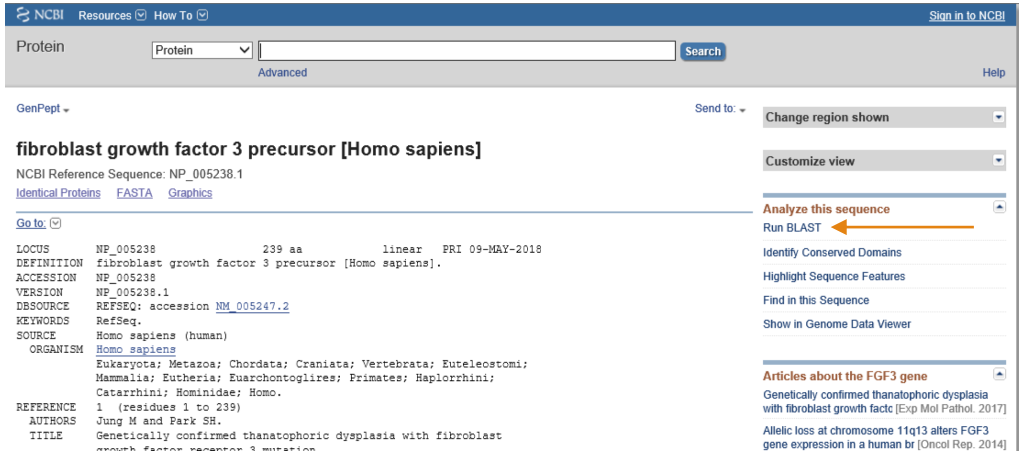
How many human similar reference proteins are in the databases?

Does the sequence contain any transmembrane regions?

Find human protein sequence **FGF3**

How many human similar reference proteins are in the databases?





fibroblast growth factor 3 precursor [Homo sapiens]

NCBI Reference Sequence: NP_005238.1

Identical Proteins FASTA Graphics

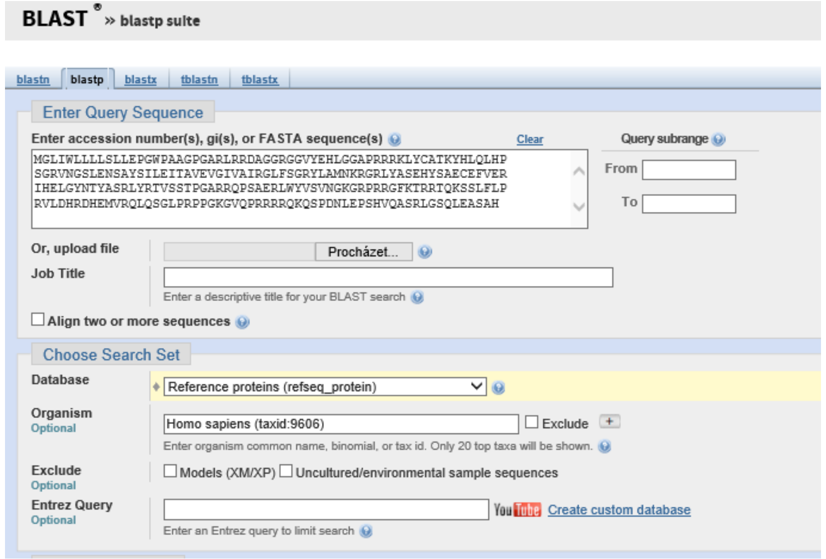
```
LOCUS       NP_005238                239 aa            linear   PRI 09-MAY-2018
DEFINITION  fibroblast growth factor 3 precursor [Homo sapiens].
ACCESSION   NP_005238
VERSION     NP_005238.1
DBSOURCE    REFSEQ: accession NM_005247.2
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 239)
  AUTHORS   Jung M and Park SH.
  TITLE     Genetically confirmed thanatophoric dysplasia with fibroblast
```

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence
- Show in Genome Data Viewer

Articles about the FGF3 gene

- Genetically confirmed thanatophoric dysplasia with fibroblast growth factc [Exp Mol Pathol. 2017]
- Allelic loss at chromosome 11q13 alters FGF3 gene expression in a human br [Oncol Rep. 2014]



Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH
```

Or, upload file

Job Title: Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: Reference proteins (refseq_protein)

Organism (Optional): Homo sapiens (taxid:9606) ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

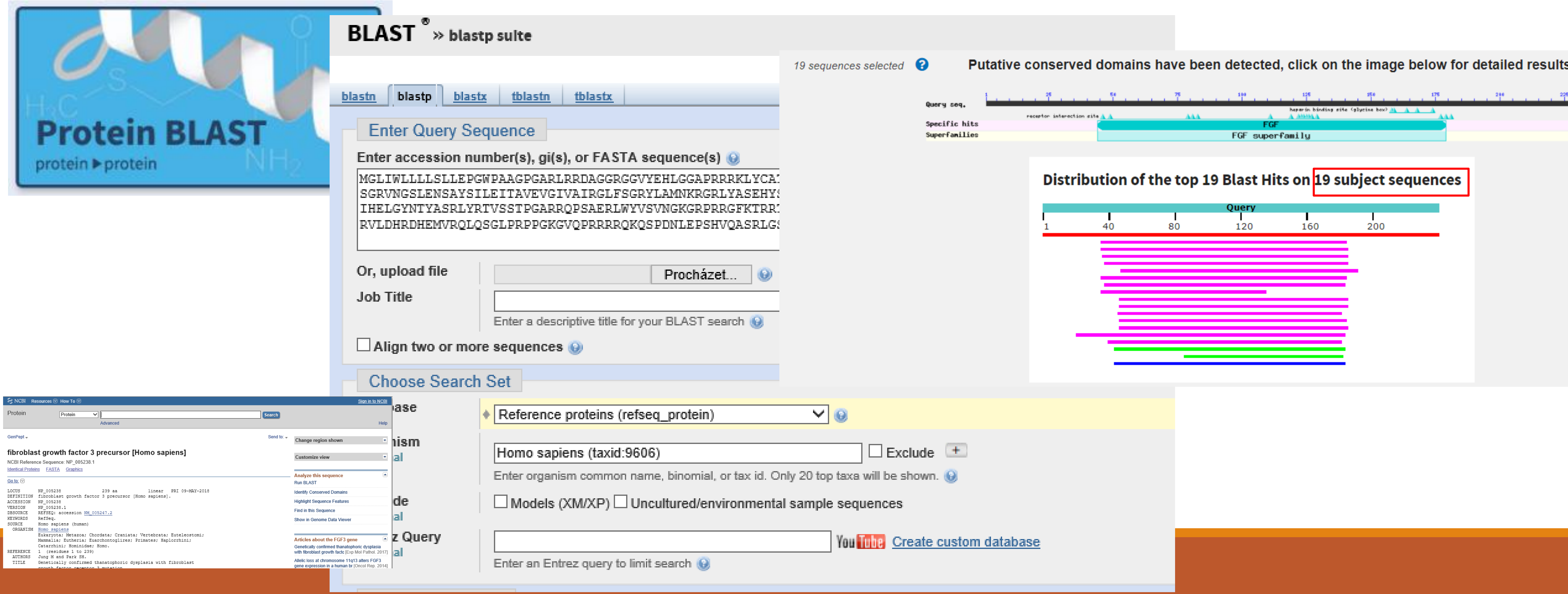
Exclude (Optional): ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query (Optional): Create custom database

Enter an Entrez query to limit search

Find human protein sequence **FGF3**

How many human similar reference proteins are in the databases?



Enter accession number(s), gi(s), or FASTA sequence(s)

```
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCA
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHY
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRR
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLG
```

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

ase: Reference proteins (refseq_protein)

ism
al: Homo sapiens (taxid:9606) ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

de
al: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

z Query
al:

Enter an Entrez query to limit search

Putative conserved domains have been detected, click on the image below for detailed results

Distribution of the top 19 Blast Hits on 19 subject sequences

fibroblast growth factor 3 precursor [Homo sapiens]

NCBI Reference Sequence: NP_005238.1

Identical Proteins FASTA Graphics

```
LOCUS       NP_005238               239 aa            linear   PRI 09-MAY-2018
DEFINITION  fibroblast growth factor 3 precursor [Homo sapiens].
ACCESSION   NP_005238
VERSION     NP_005238.1
DBSOURCE    REFSEQ: accession NM_005247.2
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 239)
  AUTHORS   Jung M and Park SH.
  TITLE     Genetically confirmed thanatophoric dysplasia with fibroblast
```

Analyze this sequence
- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence
- Show in Genome Data Viewer

Articles about the FGF3 gene
- Genetically confirmed thanatophoric dysplasia with fibroblast growth factc [Exp Mol Pathol. 2017]
- Allelic loss at chromosome 11q13 alters FGF3 gene expression in a human br [Oncol Rep. 2014]

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?

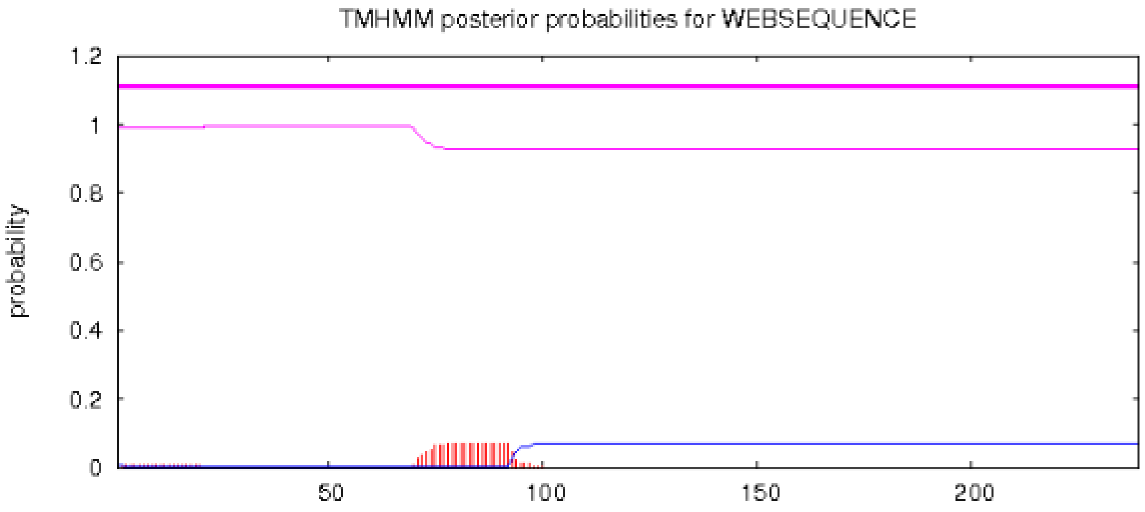
How many human similar reference proteins are in the databases?

Does the sequence contain any transmembrane regions?

Find human protein sequence **FGF3**

Does the sequence contain any transmembrane regions?

→Use any of the transmembrane helices predicting program.



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.

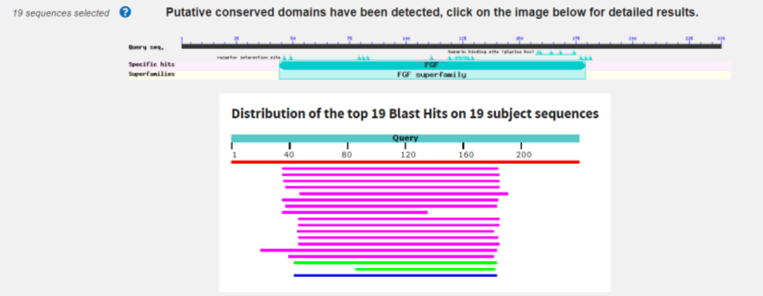
P11487 (FGF3_HUMAN) / NP_005238.1

Fibroblast growth factor 3

-role in the regulation of embryonic development…

V	13

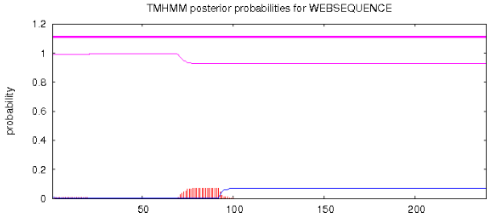
Trypsin	VNGSLENSAYSILEITAVEVGIVAIR	26



19 sequences selected

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 19 Blast Hits on 19 subject sequences

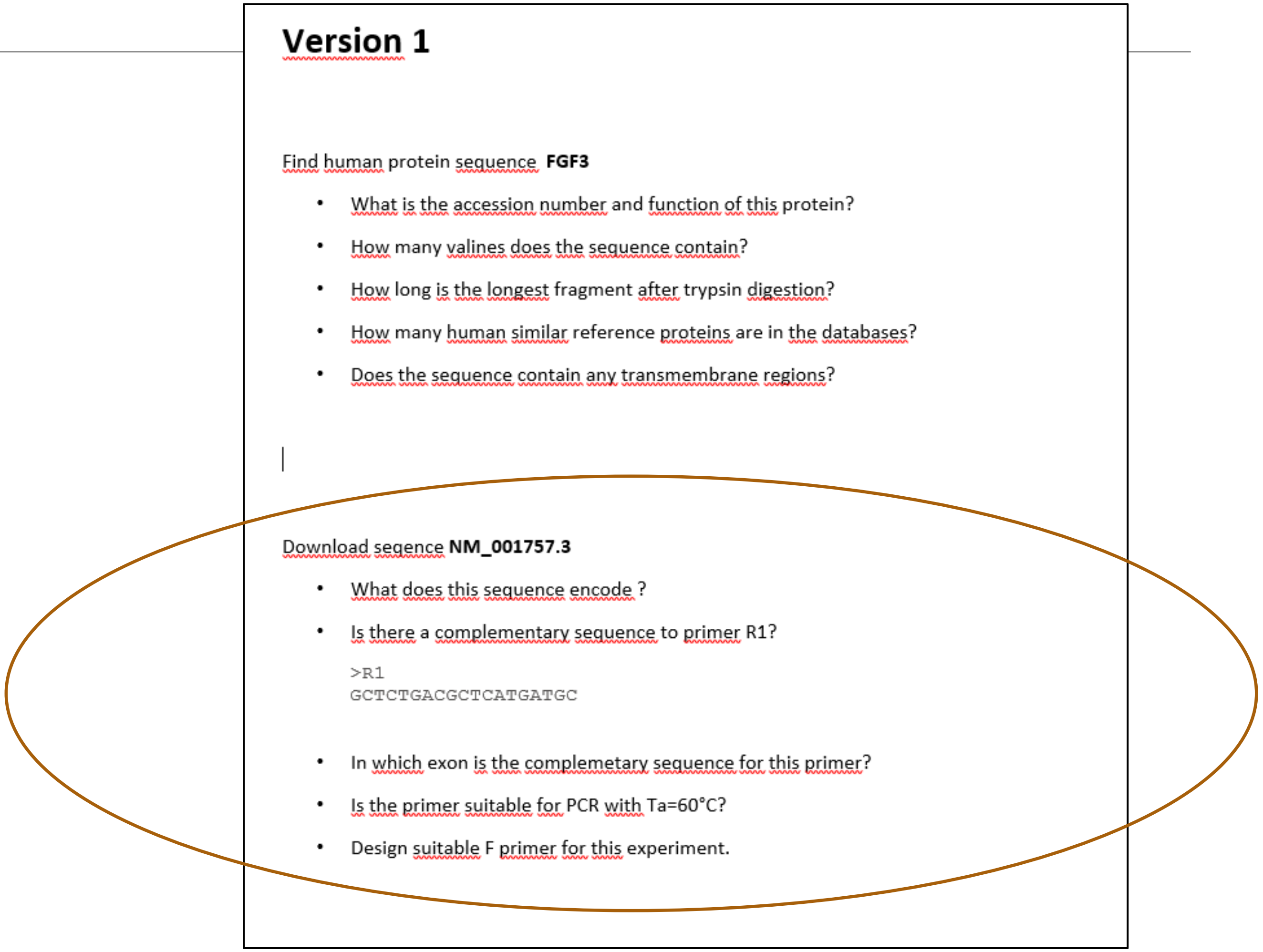


TMHMM posterior probabilities for WEBSEQUENCE

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 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?

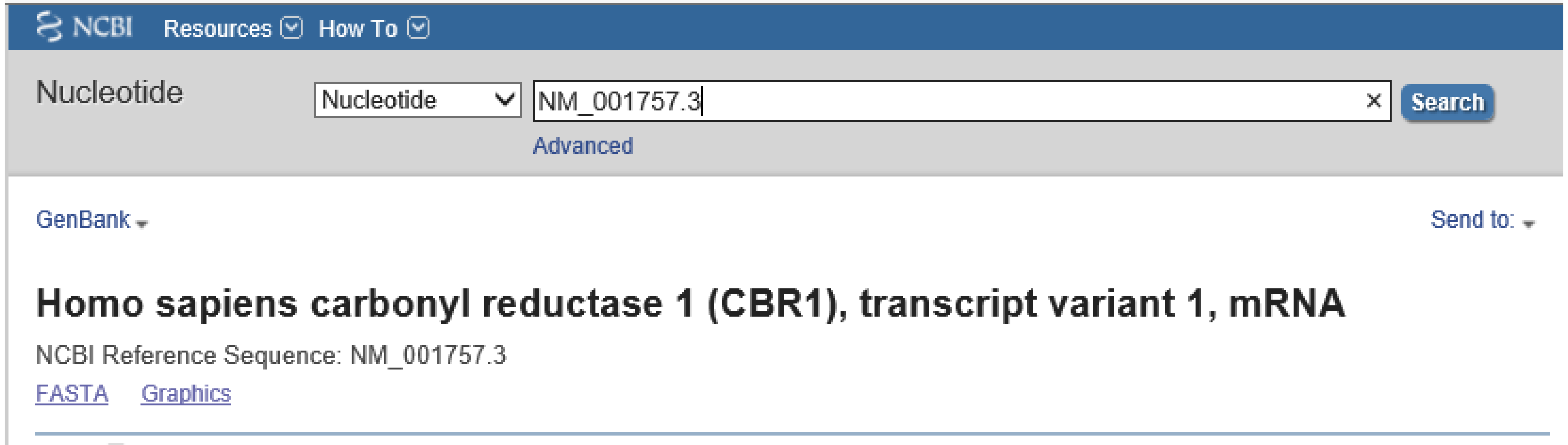
Is the primer suitable for PCR with Ta=60°C?

Design suitable F primer for this experiment (gene detection).

```
>R1
GCTCTGACGCTCATGATGC
```

Download seqence **NM_001757.3**

What does this sequence encode ?



Nucleotide: NM_001757.3

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

NCBI Reference Sequence: NM_001757.3

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.

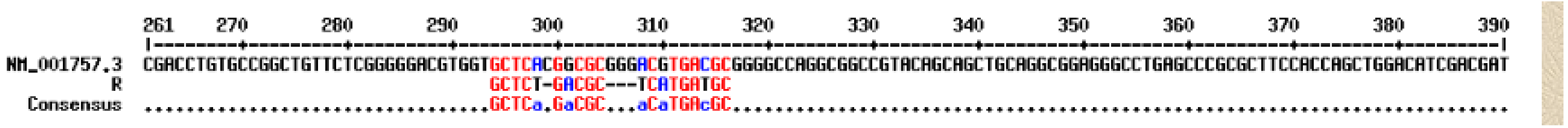
>R1
`GCTCTGACGCTCATGATGC`

Download seqence **NM_001757.3**

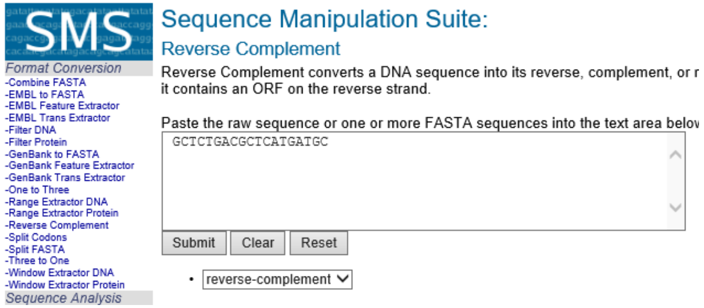
Is there a complementary sequence to primer R1?

>R1
GCTCTGACGCTCATGATGC

Multalin:



```
             261    270       280       290       300       310       320       330       340       350       360       370       380       390
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_001757.3  CGACCTGTGCCGGCTGTTCTCGGGGGACGTGGTGCTCACGGCGCGGGACGTGACGCGGGGCCAGGCGGCCGTACAGCAGCTGCAGGCGGAGGGCCTGAGCCCGCGCTTCCACCAGCTGGACATCGACGAT
          R                                   GCTCT-GACGC---TCATGATGC
  Consensus  .................................GCTCa.GaCGC...aCaTGAcGC..........................................................................
```



Sequence Manipulation Suite:

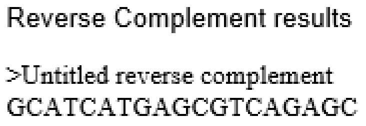
Reverse Complement

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

Paste the raw sequence or one or more FASTA sequences into the text area belov

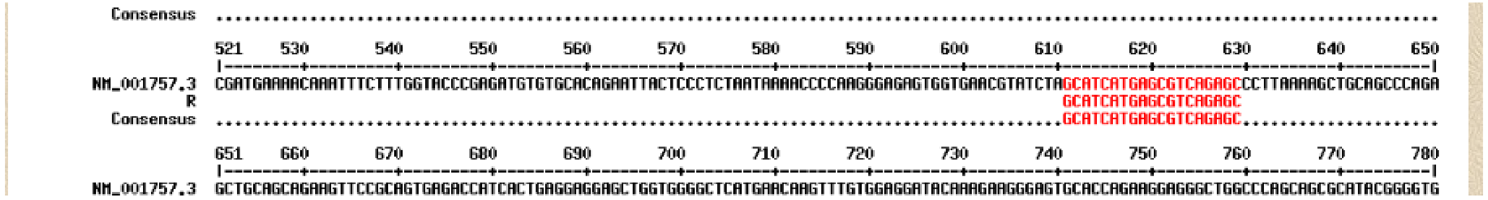
GCTCTGACGCTCATGATGC

- reverse-complement



Reverse Complement results

>Untitled reverse complement
GCATCATGAGCGTCAGAGC



```
  Consensus  ..................................................................................................................................

             521    530       540       550       560       570       580       590       600       610       620       630       640       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
```

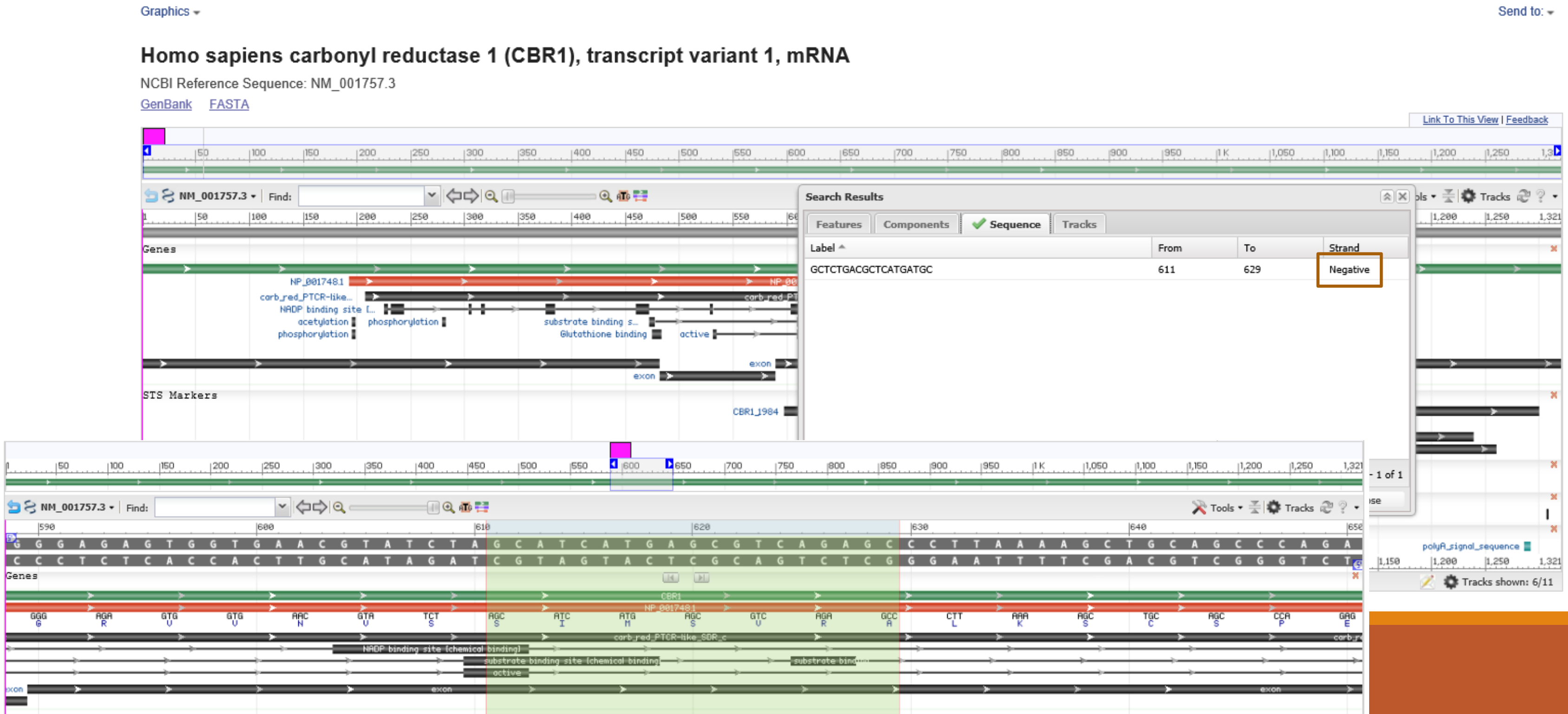
→yes.

Download seqence **NM_001757.3**

Is there a complementary sequence to primer R1?

>R1
GCTCTGACGCTCATGATGC

NCBI: Graphic



Graphics

Send to:

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

NCBI Reference Sequence: NM_001757.3

GenBank FASTA

Search Results

Label	From	To	Strand
GCTCTGACGCTCATGATGC	611	629	Negative

→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?

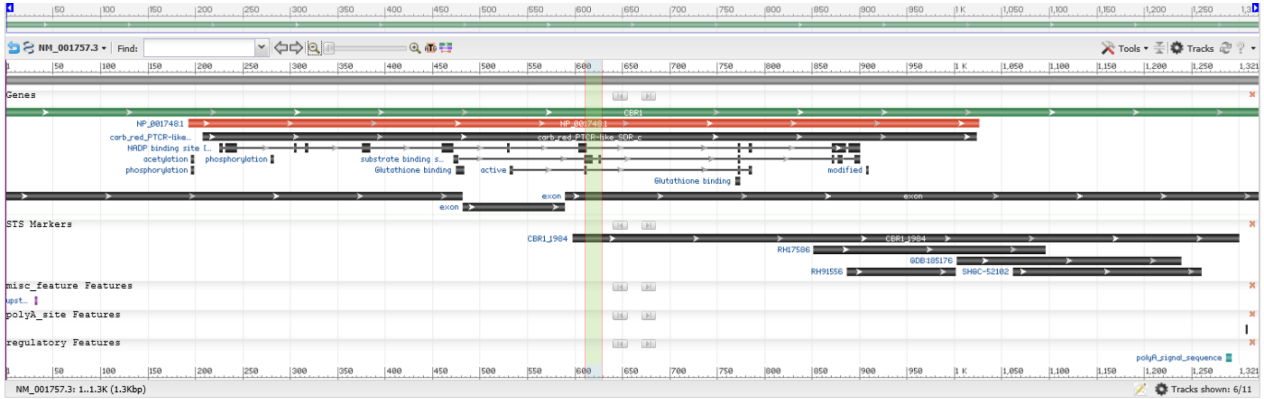
Design suitable F primer for this experiment.

>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?

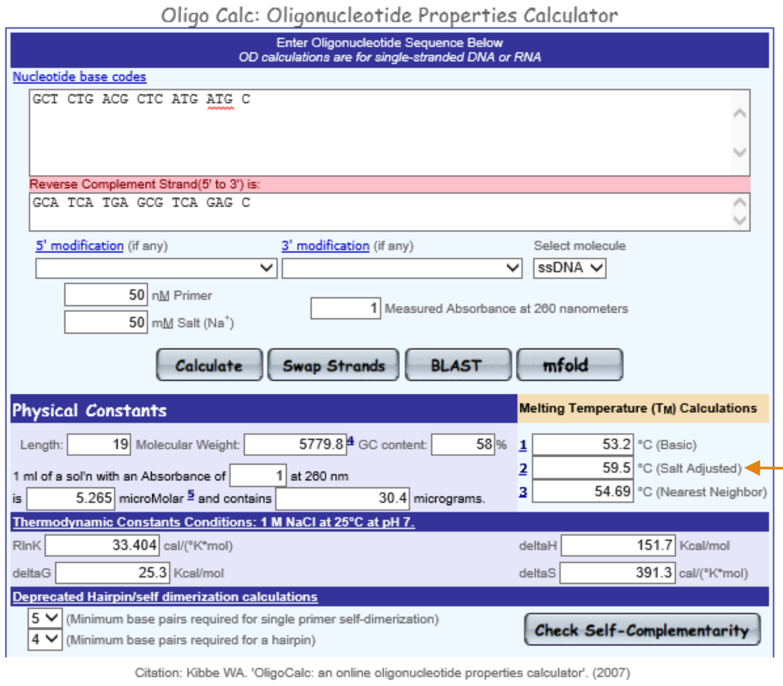
Design suitable F primer for this experiment.

```
>R1
GCTCTGACGCTCATGATGC
```

Download seqence **NM_001757.3**

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

OligoCalc:



Oligo Calc: Oligonucleotide Properties Calculator

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

Nucleotide base codes

```
GCT CTG ACG CTC ATG ATG C
```

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

```
GCA TCA TGA GCG TCA GAG C
```

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

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

Physical Constants

Length: 19 Molecular Weight: 5779.8[4] GC content: 58 %

1 ml of a sol'n with an Absorbance of 1 at 260 nm
is 5.265 microMolar [5] and contains 30.4 micrograms.

Melting Temperature (TM) Calculations

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

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

RlnK 33.404 cal/(°K*mol)
deltaH 151.7 Kcal/mol
deltaG 25.3 Kcal/mol
deltaS 391.3 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)

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

>R1
GCTCTGACGCTCATGATGC

→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.

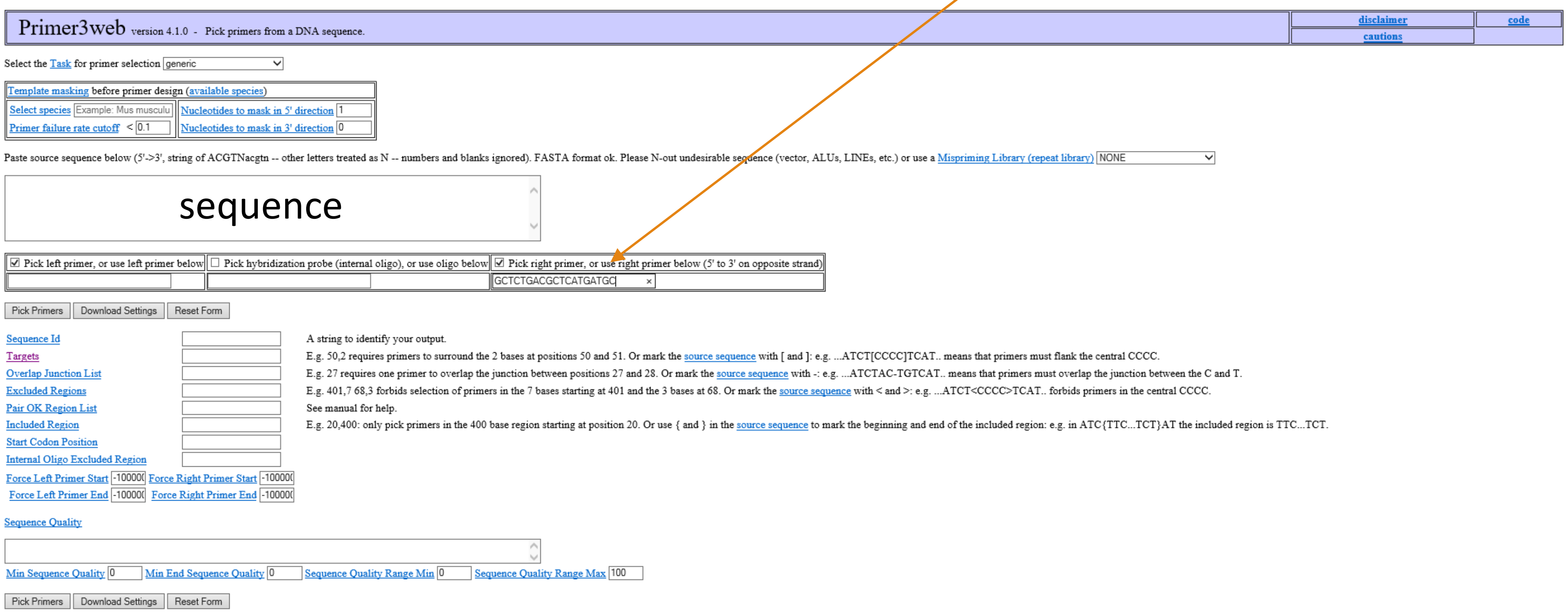
```
>R1
GCTCTGACGCTCATGATGC
```

Download seqence **NM_001757.3**

Design suitable F primer for this experiment.

>R1
GCTCTGACGCTCATGATGC

Primer3:



Primer3web version 4.1.0 - Pick primers from a DNA sequence.

Select the Task for primer selection: generic

Template masking before primer design (available species)

Select species: Example: Mus musculu
Primer failure rate cutoff < 0.1
Nucleotides to mask in 5' direction: 1
Nucleotides to mask in 3' direction: 0

Paste source sequence below (5'->3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a Mispriming Library (repeat library): NONE

sequence

☑ Pick left primer, or use left primer below
☐ Pick hybridization probe (internal oligo), or use oligo below
☑ Pick right primer, or use right primer below (5' to 3' on opposite strand): GCTCTGACGCTCATGATGC

Sequence Id: A string to identify your output.
Targets: E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the source sequence with [and]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.
Overlap Junction List: E.g. 27 requires one primer to overlap the junction between positions 27 and 28. Or mark the source sequence with -: e.g. ...ATCTAC-TGTCAT.. means that primers must overlap the junction between the C and T.
Excluded Regions: E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the source sequence with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.
Pair OK Region List: See manual for help.
Included Region: E.g. 20,400: only pick primers in the 400 base region starting at position 20. Or use { and } in the source sequence to mark the beginning and end of the included region: e.g. in ATC{TTC...TCT}AT the included region is TTC...TCT.
Start Codon Position
Internal Oligo Excluded Region
Force Left Primer Start: -100000 Force Right Primer Start: -100000
Force Left Primer End: -100000 Force Right Primer End: -100000

Sequence Quality

Min Sequence Quality: 0 Min End Sequence Quality: 0 Sequence Quality Range Min: 0 Sequence Quality Range Max: 100

Download seqence **NM_001757.3**

Design suitable F primer for this experiment.

Primer3:

Primer3 Output

```
PRIMER PICKING RESULTS FOR NM_001757.3 Homo sapiens carbonyl reductase 1 (CBR1), transcript variant 1, mRNA

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

    1 AGGGGTAGGGATGGTTCAGCCCCGCCCCGCTAGGGCGGGGCCTGCGCCTGCGCGCTCAGC


   61 GGCCGGGCGTGTAACCCACGGGTGCGCGCCCACGACCGCCAGACTCGAGCAGTCTCTGGA


  121 ACACGCTGCGGGGCTCCCGGGCCTGAGCCAGGTCTGTTCTCCACGCAGGTGTTCCGCGCG


  181 CCCCGTTCAGCCATGTCGTCCGGCATCCATGTAGCGCTGGTGACTGGAGGCAACAAGGGC


  241 ATCGGCTTGGCCATCGTGCGCGACCTGTGCCGGCTGTTCTCGGGGGACGTGGTGCTCACG


  301 GCGCGGGACGTGACGCGGGGCCAGGCGGCCGTACAGCAGCTGCAGGCGGAGGGCCTGAGC


  361 CCGCGCTTCCACCAGCTGGACATCGACGATCTGCAGAGCATCCGCGCCCTGCGCGACTTC


  421 CTGCGCAAGGAGTACGGGGGCCTGGACGTGCTGGTCAACAACGCGGGCATCGCCTTCAAG
                                                              >>>>

  481 GTTGCTGATCCCACACCCTTTCATATTCAAGCTGAAGTGACGATGAAAACAAATTTCTTT
      >>>>>>>>>>>>>>>>

  541 GGTACCCGAGATGTGTGCACAGAATTACTCCCTCTAATAAAACCCCAAGGGAGAGTGGTG


  601 AACGTATCTAGCATCATGAGCGTCAGAGCCCTTAAAAGCTGCAGCCCAGAGCTGCAGCAG
                <<<<<<<<<<<<<<<<<<<
```

>R1
GCTCTGACGCTCATGATGC

example:
>F
CAAGGTTGCTGATCCCACAC

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`

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`

Download seqence **NM_001757.3**

Solution:

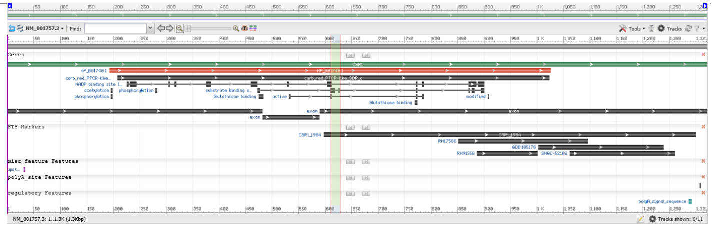
This sequence encodes carbonyl reductase 1 (CBR1).

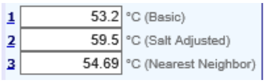
Yes, there is complementary sequence to primer R1.

It is within 3rd exon.

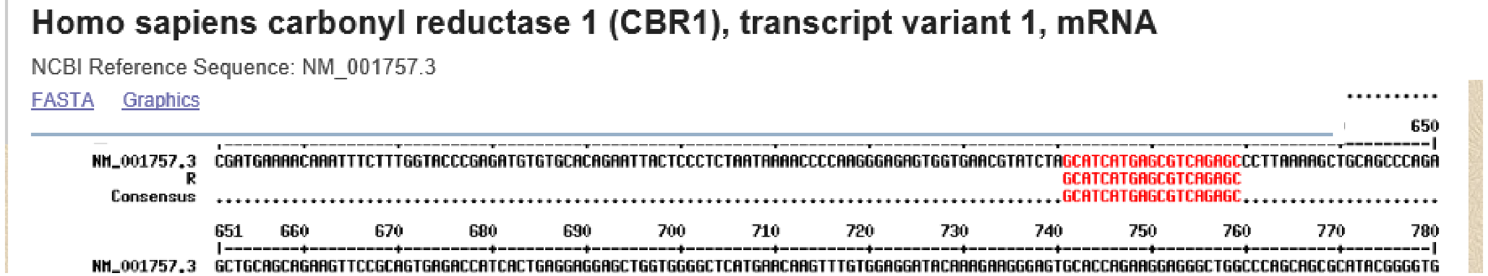
At Ta=60°C the primer should work.

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)

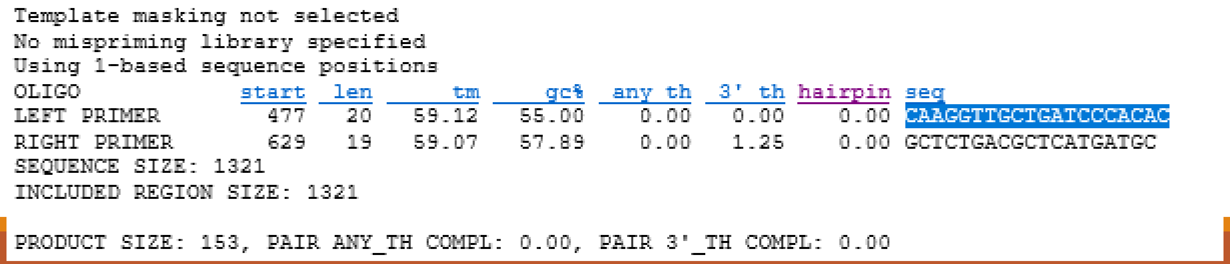


Homo sapiens carbonyl reductase 1 (CBR1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_001757.3

FASTA Graphics

```
                                                                                                                      ...........
                                                                                                                              650
              |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
NM_001757.3   CGATGAAAACAAATTTCTTTGGTACCCGAGATGTGTGCACAGAATTACTCCCTCTAATAAACCCCAAGGGAGAGTGGTGAACGTATCTAGCATCATGAGCGTCAGAGCCCTTAAAAGCTGCAGCCCAGA
          R                                                                                             GCATCATGAGCGTCAGAGC
  Consensus   ..........................................................................................GCATCATGAGCGTCAGAGC...................

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



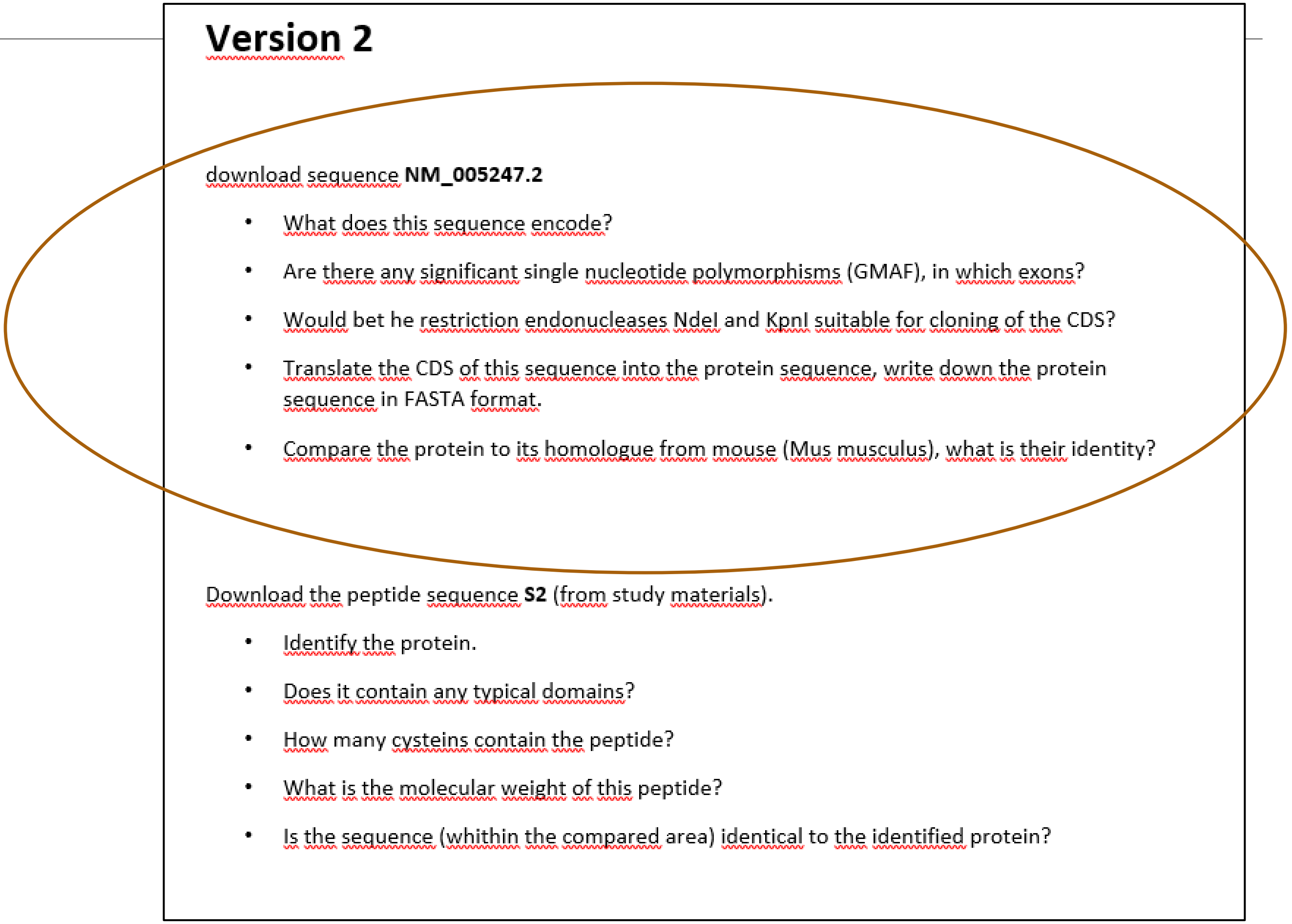
```
Template masking not selected
No mispriming library specified
Using 1-based sequence positions
OLIGO            start  len     tm     gc%   any th  3' th hairpin seq
LEFT PRIMER        477   20  59.12   55.00    0.00   0.00    0.00 CAAGGTTGCTGATCCCACAC
RIGHT PRIMER       629   19  59.07   57.89 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

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?

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?

Download sequence **NM_005247.2**

What does this sequence encode?

NCBI/Nucleotide:

Homo sapiens fibroblast growth factor 3 (FGF3), mRNA

NCBI Reference Sequence: NM_005247.2

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?

Download sequence **NM_005247.2**

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

NCBI / Graphic: (Track: Sequence variation)

→yes, in the first and third exons

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?

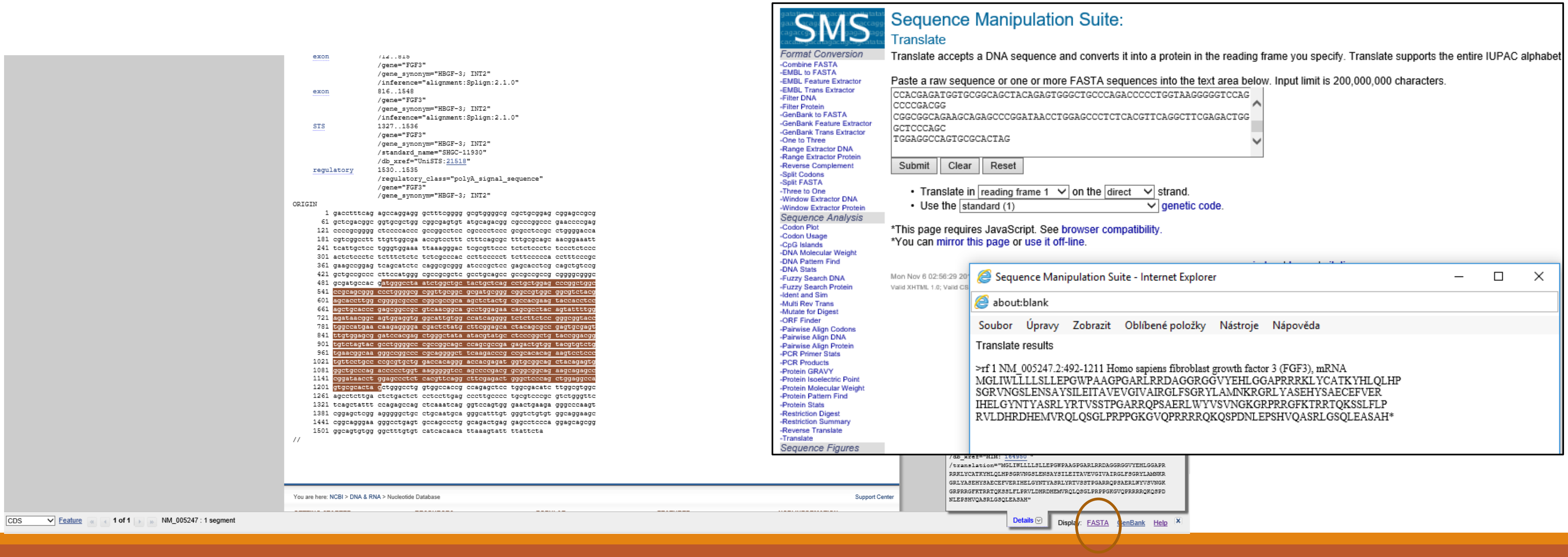
Translate the CDS of this sequence into the protein sequence, write down the protein sequence in FASTA format.

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

Download sequence **NM_005247.2**

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

NCBI:flat file- highlight sequence features „CDS“ – Translate



```
     exon            712..815
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
                     /inference="alignment:Splign:2.1.0"
     exon            816..1548
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
                     /inference="alignment:Splign:2.1.0"
     STS             1327..1536
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
                     /standard_name="SHGC-11930"
                     /db_xref="UniSTS:21518"
     regulatory      1530..1535
                     /regulatory_class="polyA_signal_sequence"
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
ORIGIN
        1 gacctttcag agccaggagg gctttcgggg gcgtggggcg cgctgcggag cggagccgcg
       61 gctcgacggc ggtgcgctgg cggcgagtgt atgcagacgg cgcccggccc gaaccccgag
      121 ccccgcgggg ctccccaccc gccggcctcc cgcccctccc gcgcctccgc ctggggacca
      181 cgtcggcctt ttgttggcga accgtccttt ctttcagcgc tttgcgcagc aacggaaatt
      241 tcattgctcc tgggtggaaa ttaaagggac tcgcgttccc tctctccctc tccctctccc
      301 actctccctc tctttctctc tctcgcccac ccttcccccct tcttccccca cctttcccgc
      361 gaagccggag tcagcatctc caggcgcggg atcccgctcc gagcacctcg cagctgtccg
      421 gctgccgccc cttccatggg cgccgcgctc gcctgcagcc gccgccgccg cggggcgggc
      481 gcgatgccac gatgggccta atctggctgc tactgctcag cctgctggag cccggctggc
      541 ccgcagcggg ccctggggcg cggttgcggc gcgatgcggg cggccgtggc ggcgtctacg
      601 agcaccttgg cggggcgccc cggcgccgca agctctactg cgccacgaag taccacctcc
      661 agctgcaccc gagcggccgc gtcaacggca gcctggagaa cagcgcctac agtattttgg
      721 agataacggc agtggaggtg ggcattgtgg ccatcagggg tctcttctcc gggcggtacc
      781 tggccatgaa caagagggga cgactctatg cttcggagca ctacagcgcc gagtgcgagt
      841 ttgtggagcg gatccacgag ctgggctata atacgtatgc ctcccggctg taccggacgg
      901 tgtctagtac gcctggggcc cgccggcagc ccagcgccga gagactgtgg tacgtgtctg
      961 tgaacggcaa gggccggccc cgcaggggct tcaagacccg ccgcacacag aagtcctccc
     1021 tgttcctgcc ccgcgtgctg gaccacaggg accacgagat ggtgcggcag ctacagagtg
     1081 ggctgcccag acccccctggt aaggggtcc agccccgacg gcggcggcag aagcagagcc
     1141 cggataacct ggagccctct cacgttcagg cttcgagact gggctcccag ctggaggcca
     1201 gtgcgcacta gctgggcctg gtggccaccg ccagagctcc tggcgacatc ttggcgtggc
     1261 agcctcttga ctctgactct cctccttgag cccttgcccc tgcgtcccgc gtctgggttc
     1321 tcagctattt ccagagccag ctcaaatcag ggtccagtgg gaactgaaga gggcccaagt
     1381 cggagctcgg agggggctgc ctgcaatgca gggcatttgt gggtctgtgt ggcaggaagc
     1441 cggcagggaa gggcctgagt gccagccctg gcagactgag gagcctccca ggagcagcgg
     1501 ggcagtgtgg ggctttgtgt catcacaaca ttaaagtatt ttattcta
//
```

You are here: NCBI > DNA & RNA > Nucleotide Database

Support Center

CDS Feature 1 of 1 NM_005247 : 1 segment

```
/db_xref="MIM: 164950"
/translation="MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPR
RRKLYCATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKR
GRLYASEHYSAECEFVERIHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGK
GRPRRGFKTRRTQKSSLFLPRVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPD
NLEPSHVQASRLGSQLEASAH"
```

Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200,000,000 characters.

```
CCACGAGATGGTGCGGCAGCTACAGAGTGGGCTGCCCAGACCCCCTGGTAAGGGGGTCCAG
CCCCGACGG
CGGCGGCAGAAGCAGAGCCCGGATAACCTGGAGCCCTCTCACGTTCAGGCTTCGAGACTGG
GCTCCCAGC
TGGAGGCCAGTGCGCACTAG
```

Submit | Clear | Reset

- Translate in reading frame 1 on the direct strand.
- Use the standard (1) genetic code.

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

Mon Nov 6 02:56:29 20
Valid XHTML 1.0; Valid CS

Translate results

```
>rf 1 NM_005247.2:492-1211 Homo sapiens fibroblast growth factor 3 (FGF3), mRNA
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP
RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH*
```

Download sequence **NM_005247.2**

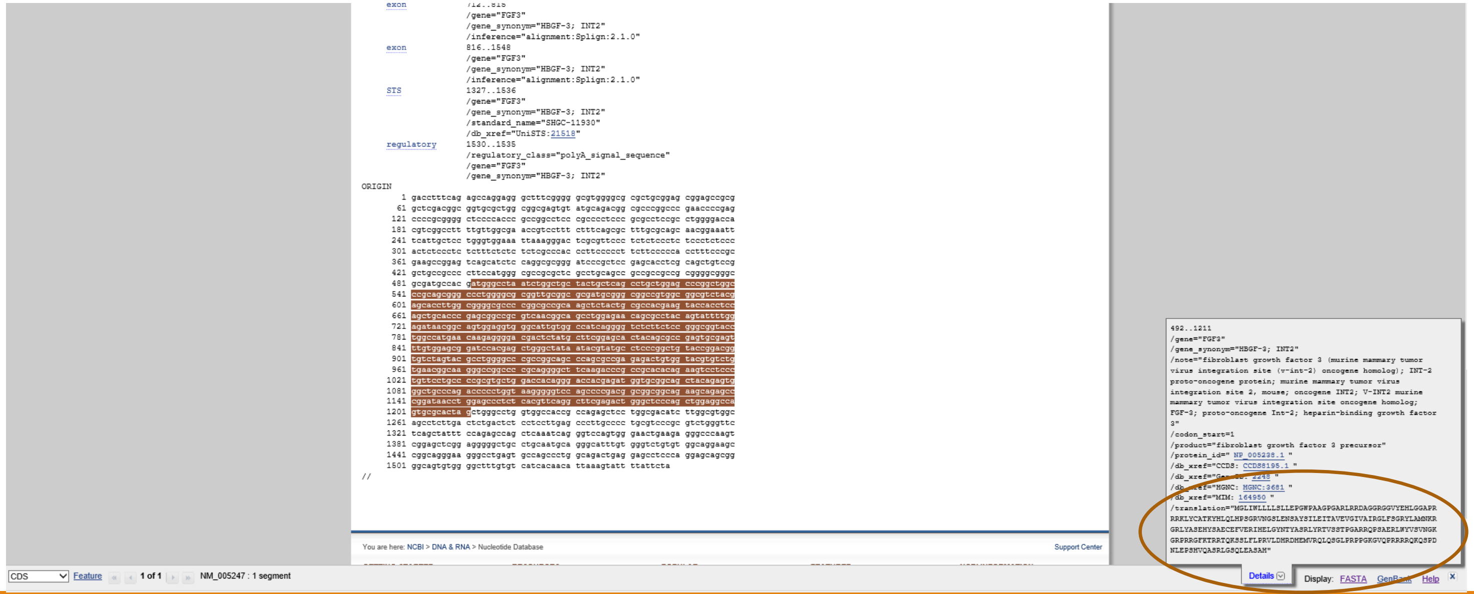
Translate the CDS of this sequence into the protein sequence, write down the protein sequence in **FASTA format**.

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

Download sequence **NM_005247.2**

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

NCBI:flat file- highlight sequence features „CDS“ – translated sequence „in the box“



```
     exon            712..815
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
                     /inference="alignment:Splign:2.1.0"
     exon            816..1548
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
                     /inference="alignment:Splign:2.1.0"
     STS             1327..1536
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
                     /standard_name="SHGC-11930"
                     /db_xref="UniSTS:21518"
     regulatory      1530..1535
                     /regulatory_class="polyA_signal_sequence"
                     /gene="FGF3"
                     /gene_synonym="HBGF-3; INT2"
ORIGIN
        1 gacctttcag agccaggagg gctttcgggg gcgtggggcg cgctgcggag cggagccgcg
       61 gctcgacggc ggtgcgctgg cggcgagtgt atgcagacgg cgcccggccc gaaccccgag
      121 ccccgcgggg ctccccaccc gccggcctcc cgcccctccc gcgcctccgc ctggggacca
      181 cgtcggcctt ttgttggcga accgtccttt ctttcagcgc tttgcgcagc aacggaaatt
      241 tcattgctcc tgggtggaaa ttaaagggac tcgcgttccc tctctccctc tccctctccc
      301 actctccctc tctttctctc tctcgcccac ccttccccct tcttccccca cctttcccgc
      361 gaagccggag tcagcatctc caggcgcggg atcccgctcc gagcacctcg cagctgtccg
      421 gctgccgccc cttccatggg cgccgcgctc gcctgcagcc gccgccgccg cggggcgggc
      481 gcgatgccac gatgggccta atctggctgc tactgctcag cctgctggag cccggctggc
      541 ccgcagcggg ccctggggcg cggttgcggc gcgatgcggg cggccgtggc ggcgtctacg
      601 agcaccttgg cggggcgccc cggcgccgca agctctactg cgccacgaag taccacctcc
      661 agctgcaccc gagcggccgc gtcaacggca gcctggagaa cagcgcctac agtattttgg
      721 agataacggc agtggaggtg ggcattgtgg ccatcagggg tctcttctcc gggcggtacc
      781 tggccatgaa caagagggga cgactctatg cttcggagca ctacagcgcc gagtgcgagt
      841 ttgtggagcg gatccacgag ctgggctata atacgtatgc ctcccggctg taccggacgg
      901 tgtctagtac gcctggggcc cgccggcagc ccagcgccga gagactgtgg tacgtgtctg
      961 tgaacggcaa gggccggccc cgcagggggct tcaagacccg ccgcacacag aagtcctccc
     1021 tgttcctgcc ccgcgtgctg gaccacaggg accacgagat ggtgcggcag ctacagagtg
     1081 ggctgcccag acccccctggt aagggggtcc agccccgacg gcggcggcag aagcagagcc
     1141 cggataacct ggagccctct cacgttcagg cttcgagact gggctcccag ctggaggcca
     1201 gtgcgcacta gctgggcctg gtggccaccg ccagagctcc tggcgacatc ttggcgtggc
     1261 agcctcttga ctctgactct cctccttgag cccttgcccc tgcgtcccgc gtctgggttc
     1321 tcagctattt ccagagccag ctcaaatcag ggtccagtgg gaactgaaga gggcccaagt
     1381 cggagctcgg aggggctgc ctgcaatgca gggcatttgt gggtctgtgt ggcaggaagc
     1441 cggcagggaa gggcctgagt gccagccctg gcagactgag gagcctccca ggagcagcgg
     1501 ggcagtgtgg ggctttgtgt catcacaaca ttaaagtatt ttattcta
//
```

You are here: NCBI > DNA & RNA > Nucleotide Database

Support Center

```
492..1211
/gene="FGF3"
/gene_synonym="HBGF-3; INT2"
/note="fibroblast growth factor 3 (murine mammary tumor
virus integration site (v-int-2) oncogene homolog); INT-2
proto-oncogene protein; murine mammary tumor virus
integration site 2, mouse; oncogene INT2; V-INT2 murine
mammary tumor virus integration site oncogene homolog;
FGF-3; proto-oncogene Int-2; heparin-binding growth factor
3"
/codon_start=1
/product="fibroblast growth factor 3 precursor"
/protein_id=" NP_005238.1 "
/db_xref="CCDS: CCDS8195.1 "
/db_xref="GeneID: 2248 "
/db_xref="HGNC: HGNC:3681 "
/db_xref="MIM: 164950 "
/translation="MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPR
RRKLYCATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKR
GRLYASEHYSAECEFVERIHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGK
GRPRRGFKTRRTQKSSLFLPRVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPD
NLEPSHVQASRLGSQLEASAH"
```

Download sequence **NM_005247.2**

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

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

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.

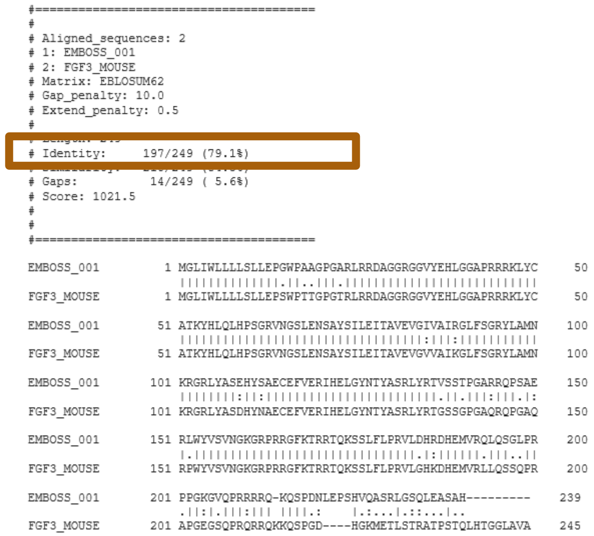
Compare the protein to its homologue from mouse (Mus musculus), what is their identity?

Download sequence **NM_005247.2**

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

→Find mouse homologue-Uniprot (FGF3)

Emboss Needle:



```
#=======================================
#
# 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:   [illegible]
# 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
```

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**

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:

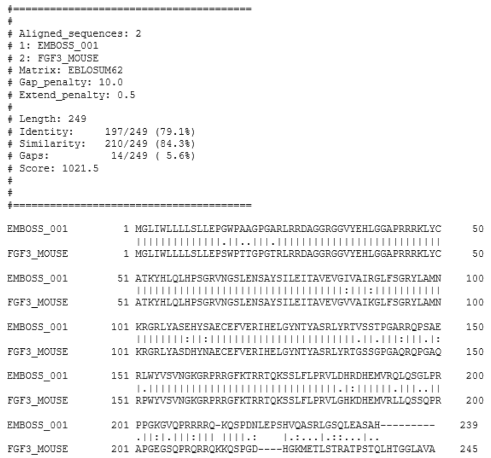
The indentity of human and mouse FGF3 proteins is 79.1 %.

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

Homo sapiens fibroblast growth factor 3 (FGF3), mRNA

NCBI Reference Sequence: NM_005247.2

NdeI 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:

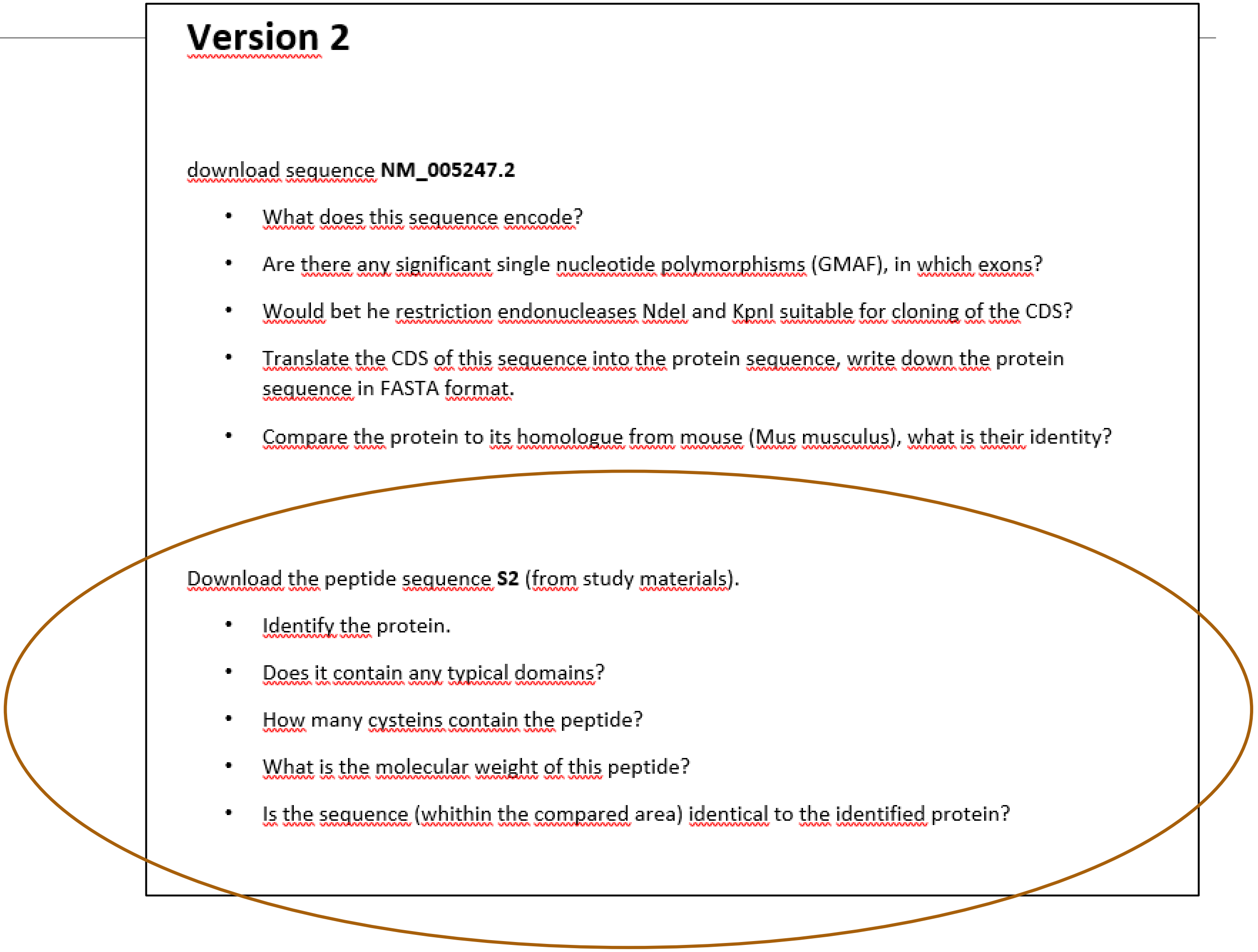
2 parts, each five points

<u>Maximum 10 points</u>

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?

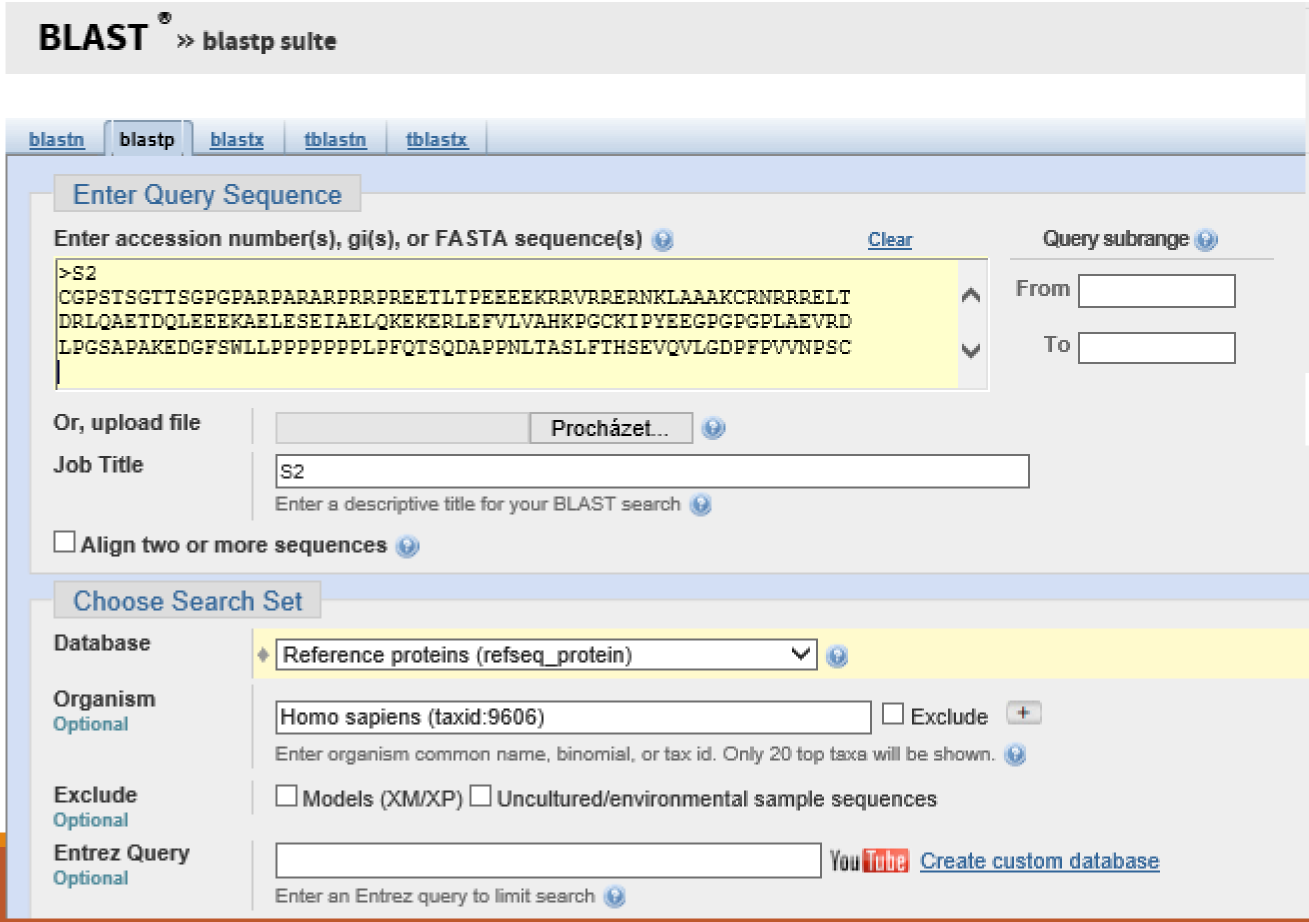
What is the molecular weight of this peptide?

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

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

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

Identify the protein.



Enter accession number(s), gi(s), or FASTA sequence(s)

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

Or, upload file

Job Title: S2

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Database: Reference proteins (refseq_protein)

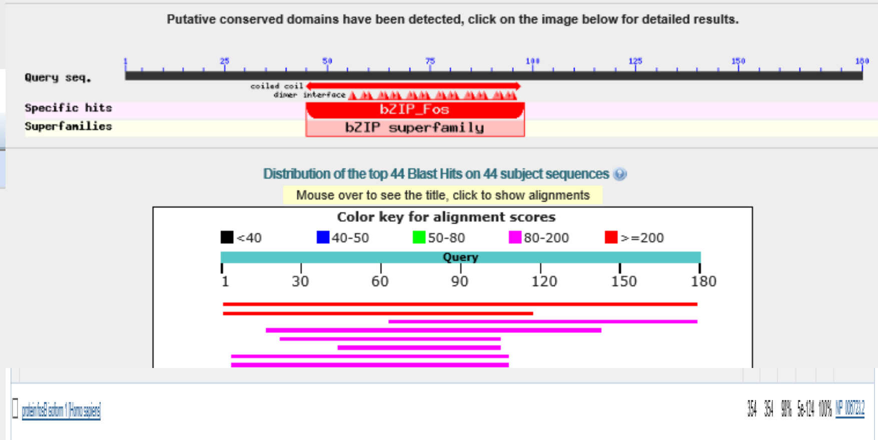
Organism (Optional): Homo sapiens (taxid:9606) ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude (Optional): ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query (Optional):

Enter an Entrez query to limit search



Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 44 Blast Hits on 44 subject sequences

Mouse over to see the title, click to show alignments

→ **protein fosB isoform 1 [Homo sapiens]**

NCBI Reference Sequence: NP_006723.2

Identical Proteins FASTA Graphics

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?

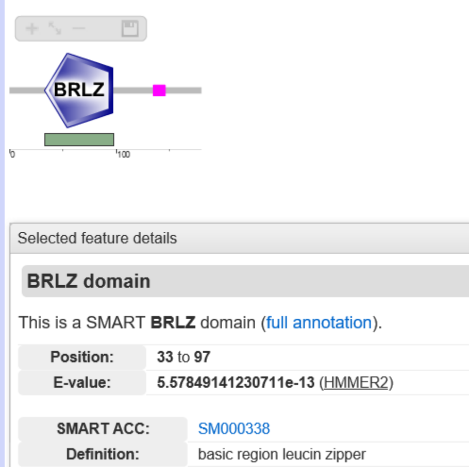
Is the sequence (whithin the compared area) identical to the identified protein?

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

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

Does it contain any typical domains?

PFAM, CDD, SMART..



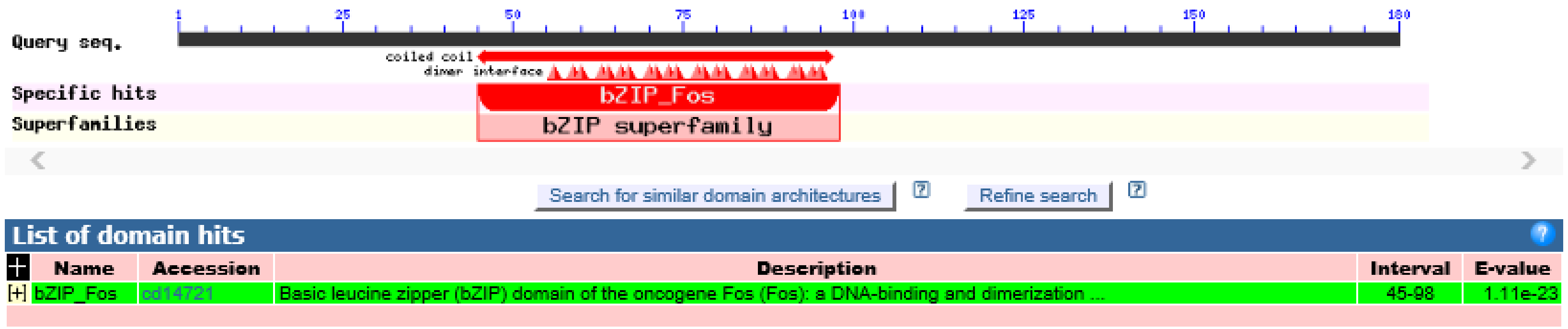
Selected feature details

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



List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	bZIP_Fos	cd14721	Basic leucine zipper (bZIP) domain of the oncogene Fos (Fos): a DNA-binding and dimerization ...	45-98	1.11e-23

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

Identify the protein.

Does it contain any typical domains?

How many cysteins contain the peptide?

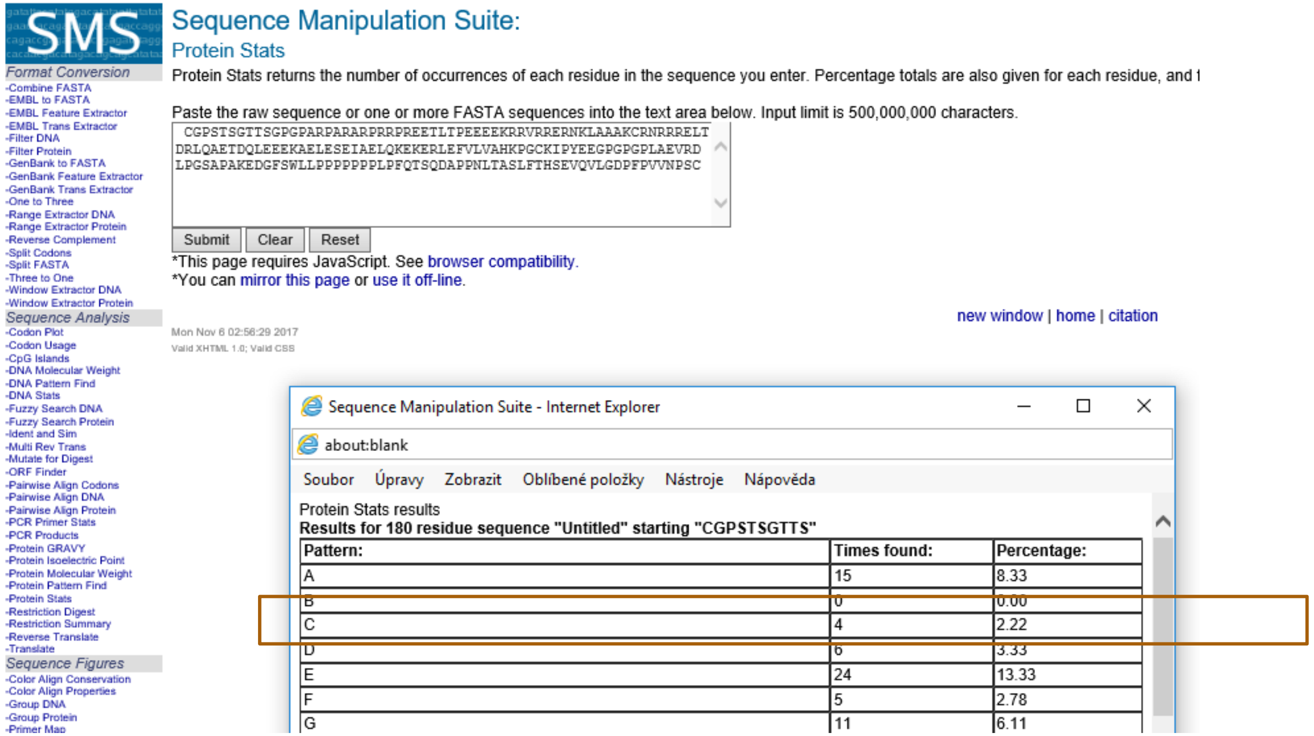
What is the molecular weight of this peptide?

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

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

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

How many cysteins contain the peptide?



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 residue, and t

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

```
 CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

Mon Nov 6 02:56:29 2017
Valid XHTML 1.0; Valid CSS

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
E	24	13.33
F	5	2.78
G	11	6.11

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

Identify the protein.

Does it contain any typical domains?

How many cysteins contain the peptide?

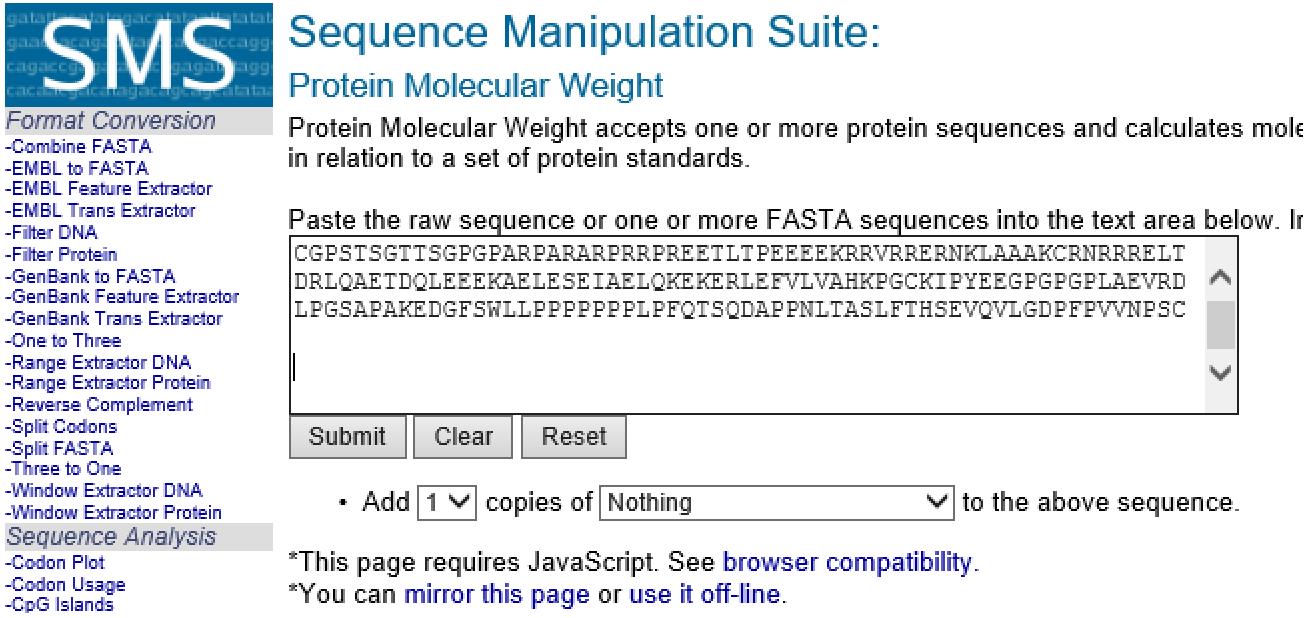
What is the molecular weight of this peptide?

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

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

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

What is the molecular weight of this peptide?



Sequence Manipulation Suite:

Protein Molecular Weight

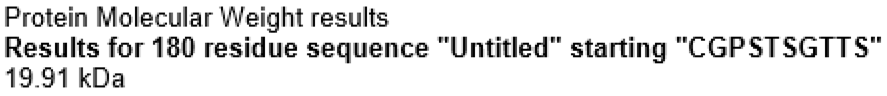
Protein Molecular Weight accepts one or more protein sequences and calculates mole
in relation to a set of protein standards.

Paste the raw sequence or one or more FASTA sequences into the text area below. Ir

```
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

- Add 1 copies of Nothing to the above sequence.

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.



Protein Molecular Weight results
Results for 180 residue sequence "Untitled" starting "CGPSTSGTTS"
19.91 kDa

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?

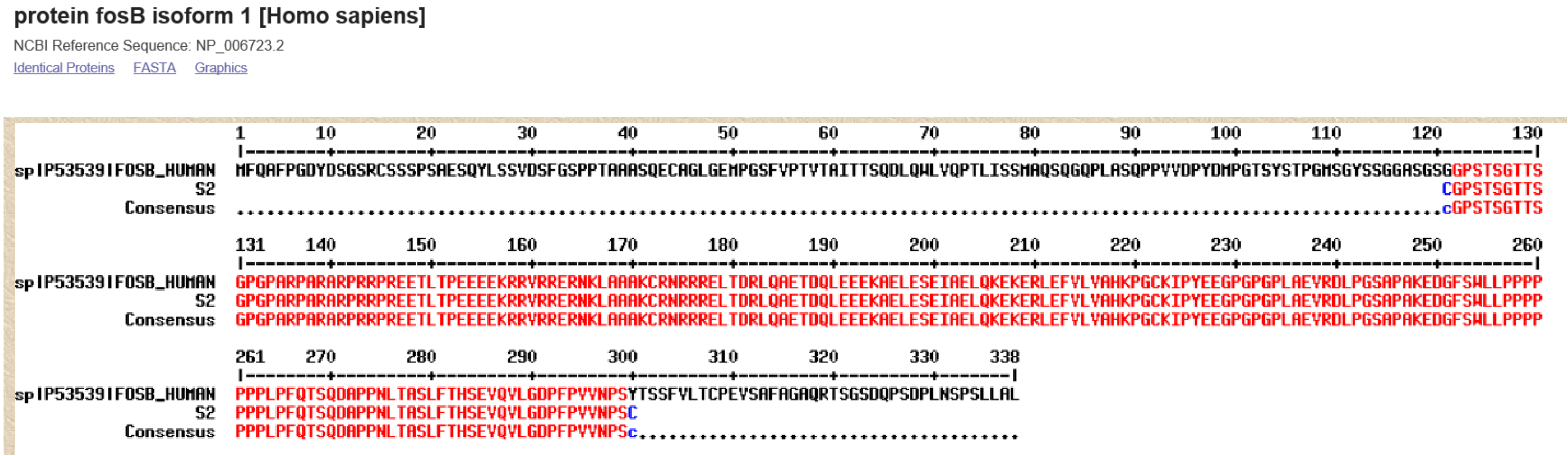
Is the sequence (whithin the compared area) identical to the identified protein?

```
>S2
CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

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

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

Comparison-Multalin:



protein fosB isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_006723.2

Identical Proteins FASTA Graphics

```
                      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  GPGPARPARARPRRPREETLTPEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPP
                  S2  GPGPARPARARPRRPREETLTPEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPP
           Consensus  GPGPARPARARPRRPREETLTPEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPP

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

→It differs in the first and the last aminoacid

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).

Solution:

Peptide is probably human FosB homologue.

Yes, it contains leucine zipper bZIP.

Contains four cysteins

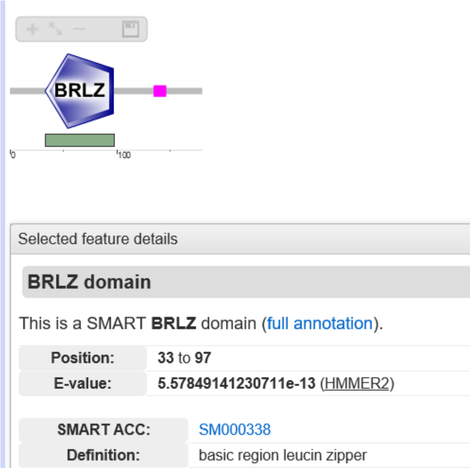
Mw=19.1 kDa.

It differs from the human homologue in the first and the last aminoacid.

protein fosB isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_006723.2

Identical Proteins FASTA Graphics

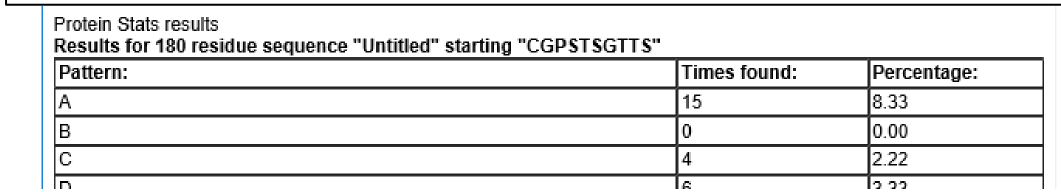


Selected feature details

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

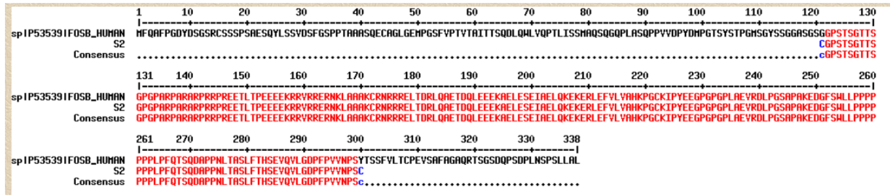
Results for 180 residue sequence "Untitled" starting "CGPSTSGTTS"

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......................................
```