

# Základy praktické Bioinformatiky

---

PETRA MATOUŠKOVÁ

2019/2020

10/10

# Základy praktické bioinformatiky

Téma 10/10

---

## Opakování

**Cíle:**

Student bude dobře připraven na zkoušku 😊

# Zkouškové otázky 2017

---

# Zkoušková písemka:

---

2 úlohy, každá 5 bodů

Hodnocení: max celkem 10 bodů

9-10b: 1

7-8b: 2

5-6b: 3

## Verze 1

Najděte lidskou sekvenci **FGF3**

- Zapište přístupový kód a funkci tohoto proteinu.
- Kolik valinů obsahuje proteinová sekvence?
- Jak dlouhý je nejdelší peptid po štěpení trypsinem?
- Kolik lidských podobných referenčních proteinů je v databázích k nalezení?
- Identifikujte 80 aminokyselinu a zapište její mutaci na prolin.

Stáhněte si sekvenci **NM\_001757.3**

- Co tento gen kóduje?
- Obsahuje příslušná sekvence cílové místo pro zadaný primer R1?
- V jakém exonu leží cílová sekvence tohoto primeru?
- Bylo by možné použít tento primer při PCR s Ta 60°C?
- Navrhnete k tomuto primeru vhodný F primer.

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Uniprot:

<input type="checkbox"/>	<a href="#">P11487</a>	FGF3_HUMAN		<b>Fibroblast growth factor 3</b>	FGF3 INT2	<a href="#">Homo sapiens (Human)</a>
--------------------------	------------------------	------------	-----------------------------------------------------------------------------------	-----------------------------------	-----------	--------------------------------------

→ regulace vývoje embrya, proliferace, diferenciace...

Nebo

NCBI/Protein:

- [fibroblast growth factor 3 precursor \[Homo sapiens\]](#)
- 2. 239 aa protein  
Accession: NP\_005238.1 GI: 4885233

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

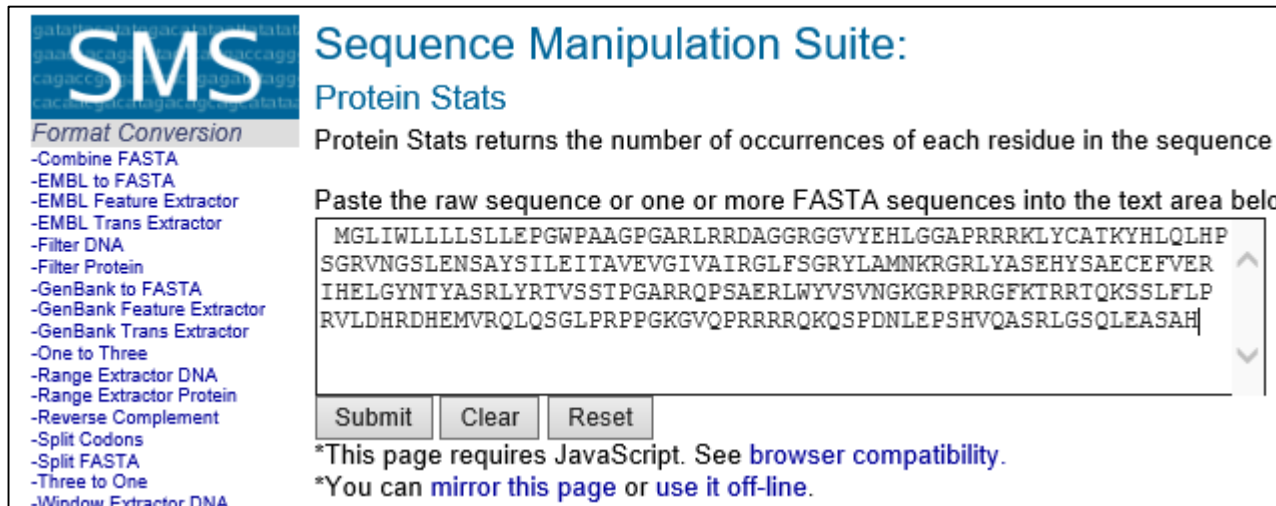
Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.



# Najděte lidskou sekvenci FGF3

Kolik valinů obsahuje proteinová sekvence?



**SMS** Sequence Manipulation Suite:  
Protein Stats

Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA

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 below

```
MGLIWLLLLSLLLEPGWPAAGPGARLRRDAGGRGGVYEHGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSACEFVER
IHELGYNTYASRLYRTVSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP
RVLDRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH
```

Submit Clear Reset

\*This page requires JavaScript. See [browser compatibility](#).  
\*You can [mirror this page](#) or [use it off-line](#).

V	13	5.44
---	----	------

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.

# Najděte lidskou sekvenci FGF3

Jak dlouhý je nejdelší peptid po štěpení trypsinem?



## PeptideCutter

PeptideCutter [[references](#) / [documentation](#)] predicts potential cleavag

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

```
MGLIWLLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLLGGAPRRRKLYC
ATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEH
YSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTR
RTQKSSLFLP
RVLDRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRL
GSQLEASAH
```

Perform

the cleavage of th

Please indicate the way you would like the cleavage sites

- Map of cleavage sites. Please select the number of amino acids
- Table of sites, sorted alphabetically by enzyme and chemical
- 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 <a href="#">explanations</a> )	Peptide length [aa]	Peptide mass [Da]
24	Trypsin	MGLIWLLLLLSLLEPGWPAAGPGAR	24	2532.085
26	Trypsin	LR	2	287.362
27	Trypsin	R	1	174.203
32	Trypsin	DAGGR	5	474.474
44	Trypsin	GGVYEHLLGGAPR	12	1212.330
46	Trypsin	RR	2	330.390
47	Trypsin	K	1	146.189
53	Trypsin	LYCAIK	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	IVSSTPGAR	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	QLQSGLRPPGK	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

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

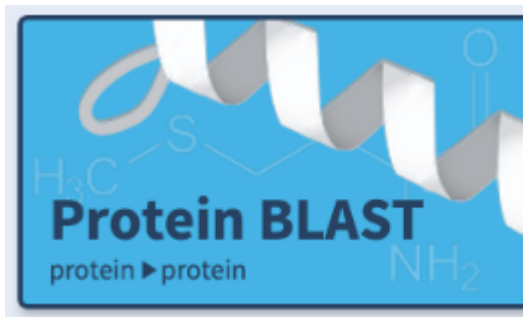
Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.

# Najděte lidskou sekvenci FGF3

Kolik lidských podobných proteinů je v databázích k nalezení?



**BLAST** » blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence

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

```
MGLIWLILLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLLGGAPRRRRLKLYCATKYHLQI
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSACEFV
IHELGYNTYASRLYRTVSSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRTRTQKSSLI
RVLDHRDHMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLASL
```

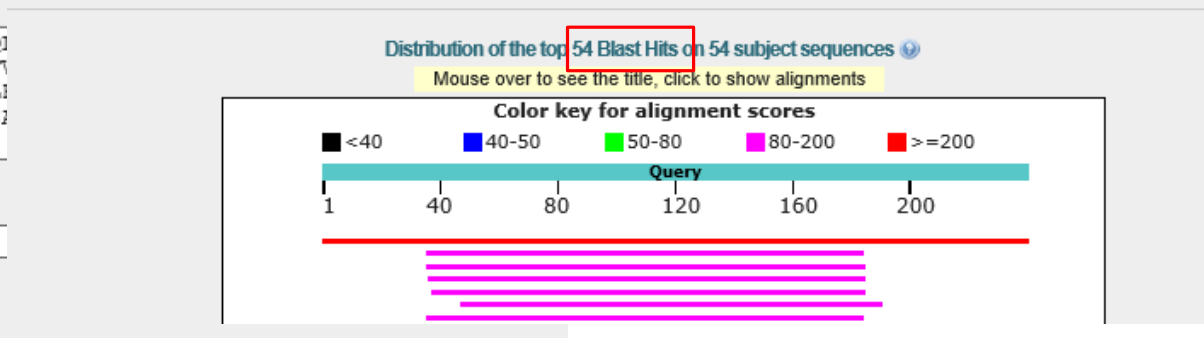
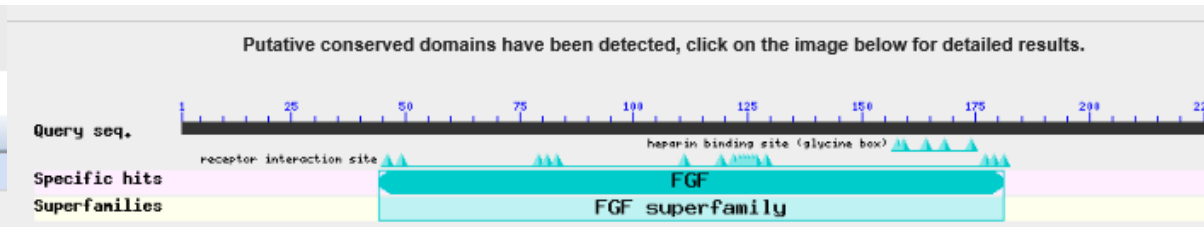
Or, upload file

Job Title

Procházet...

Enter a descriptive title for your BLAST search

Align two or more sequences



NCBI Resources | How To | Sign In to NCBI

Protein

fibroblast growth factor 3 precursor [Homo sapiens]

NCBI Reference Sequence: NP\_005238.1

Identical Protein FASTA Graphics

GenPept

Change region shown

Customize view

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in the Sequence

Show in Genome Data Viewer

Articles about the FGF3 gene

Genetically confirmed thymoplasma dysplasia with fibroblast growth factor 3 mutation [Chen et al. 2014]

Choose Search Set

Reference proteins (refseq\_protein)

Homo sapiens (taxid:9606)  Exclude

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

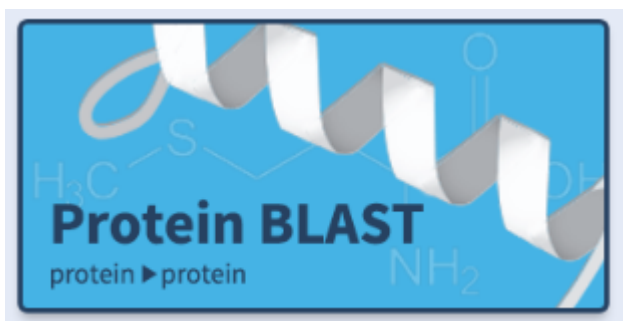
Models (XM/XP)  Uncultured/environmental sample sequences

Enter an Entrez query to limit search

YouTube Create custom database

# Najděte lidskou sekvenci FGF3

Kolik lidských podobných proteinů je v databázích k nalezení?



NCBI Resources How To Sign in to NCBI

Protein Protein Search

Advanced Help

GenPept Send to Change region shown Customize view

**fibroblast growth factor 3 precursor [Homo sapiens]**  
NCBI Reference Sequence: NP\_005238.1  
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to

LOCUS NP\_005238 239 aa linear FRI 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;  
Carnivora; Hominoidea; Homo.  
REFERENCE  
1 (residues 1 to 239)  
AUTHORS Jung M and Park SH.  
TITLE Genetically confirmed thanatophoric dysplasia with fibroblast growth factor 3 mutation

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 factor 3 mutation [Exp Mol Pathol. 2017]  
Allelic loss at chromosome 11q13 alters FGF3 gene expression in a human br [Oncol Rep. 2014]

BLAST » blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

MGLIWLILLSLEPGWFAAGPGARLRRDAGGRGGVYEHGGAPRRRKLKATKYHLQLHP  
SGRVNGLSENSAYSILEITAVEGVIVAIRGLFSGRYLAMNKRRLYASEHYSACEFVER  
IHELGYNTYASRLYRTVSTPGARRQPSAERLWVSVNGKRRPRRGFKTRRQKSSFLFP  
RVLDHRDHEMVRQLQSGLRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSLQLEASAH

From To

Or, upload file Procházet...  
Job Title  
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Reference proteins (refseq\_protein)

Organism Homo sapiens (taxid:9606) Exclude  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query  
Enter an Entrez query to limit search

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.



# Najděte lidskou sekvenci FGF3

Identifikujte 80 aminokyselinu a zapište mutaci na prolin.



## Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

## Sequence Analysis

- Codon Plot
- Codon Usage
- CoG Islands

## Sequence Manipulation Suite:

### Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or obtain subsequences using position information.

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

```
MGLIWLLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLLGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEEFVER
IHELGYNTYASRLYRTVSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSFLFP
RVLDRDHEMVRQLQSGLEPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH
```

Enter the residue positions or ranges to be extracted. Use ".." to represent a range, ; be included in the ranges. For example, to obtain the last three residues of a sequer be used.

### Range Extractor Protein results

>results for 239 residue sequence "Untitled" starting "MGLIWLLLLS"  
A

Alanin

→A80P

# Najděte lidskou sekvenci FGF3

---

Zapište přístupový kód (lidský) a funkci tohoto proteinu.

Kolik valinů obsahuje proteinová sekvence?

Jak dlouhý je nejdelší peptid po štěpení trypsinem?

Kolik lidských podobných proteinů je v databázích k nalezení?

Identifikujte 80. aminokyselinu a zapište mutaci na prolin.

## Výsledek:

FGF3: P11487 (NP\_005238.1), Fibroblast growth factor 3-regulace vývoje embrya

Obsahuje 13 valinů.

Nejdelší peptid po štěpení trypsinem bude mít 26 aminokyselin.

V databázi (referenčních sekvencí) je 54 podobných proteinů.

80. Aminokyselina je alanin, mutace na prolin bude zapsána A80P.

# Najděte lidskou sekvenci FGF3

## Výsledek:

FGF3: P11487 (NP\_005238.1), Fibroblast growth factor 3-regulace vývoje embrya

Obsahuje 13 valinů.

Nejdelší peptid po štěpení trypsinem bude mít 26 aminokyselin.

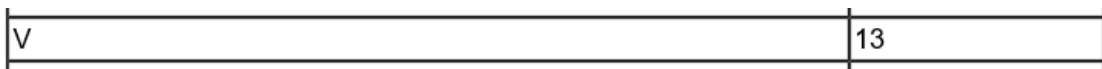
V databázi (referenčních sekvencí) je 54 podobných proteinů.

80. Aminokyselina je alanin, mutace na prolin bude zapsána A80P.

[P11487 \(FGF3\\_HUMAN\)](#) / [NP\\_005238.1](#)

### Fibroblast growth factor 3

-regulace vývoje embrya...



Trypsin

VNGSLENSAYSILEITAVEVGIVAIR

26



Range Extractor Protein results

>results for 239 residue sequence "Untitled" starting "MGLIWLLLLLS"

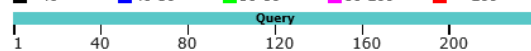
A

Distribution of the top 54 Blast Hits on 54 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >=200



A80P

# Zkoušková písemka

---

2 úlohy, každá 5 bodů

## Verze 1

Najděte lidskou sekvenci **FGF3**

- Zapište přístupový kód a funkci tohoto proteinu.
- Kolik valinů obsahuje proteinová sekvence?
- Jak dlouhý je nejdelší peptid po štěpení trypsinem?
- Kolik lidských podobných referenčních proteinů je v databázích k nalezení?
- Identifikujte 80 aminokyselinu a zapište její mutaci na prolin.

Stáhněte si sekvenci **NM\_001757.3**

- Co tento gen kóduje?
- Obsahuje příslušná sekvence cílové místo pro zadaný primer R1?
- V jakém exonu leží cílová sekvence tohoto primeru?
- Bylo by možné použít tento primer při PCR s Ta 60°C?
- Navrhněte k tomuto primeru vhodný F primer.

# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?

Obsahuje příslušná sekvence cílové místo pro zadaný **primer R1**?

V jakém exonu leží cílová sekvence tohoto primeru?

>R1

GCTCTGACGCTCATGATGC

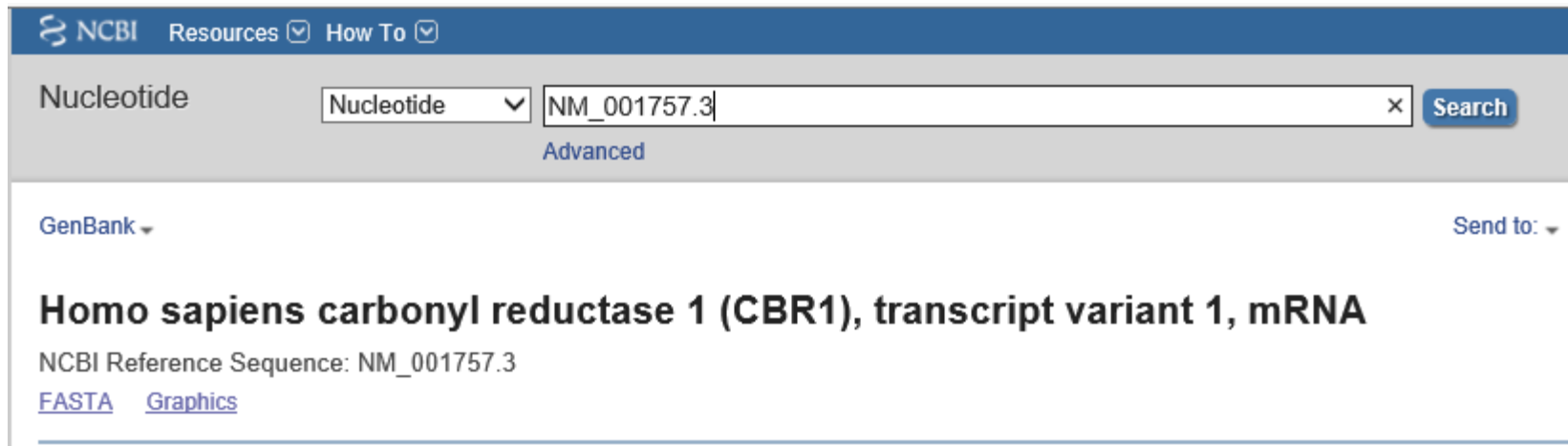
Bylo by možné použít tento primer při PCR s Ta 60°C?

Navrhněte k tomuto primeru vhodný F primer.

# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?



NCBI Resources How To

Nucleotide Nucleotide NM\_001757.3 Search

Advanced

GenBank Send to:

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

NCBI Reference Sequence: NM\_001757.3

[FASTA](#) [Graphics](#)

# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?

Obsahuje příslušná sekvence cílové místo pro zadaný **primer R1**?

>R1

GCTCTGACGCTCATGATGC

V jakém exonu leží cílová sekvence tohoto primeru?

Bylo by možné použít tento primer při PCR s Ta 60°C?

Navrhněte k tomuto primeru vhodný F primer.

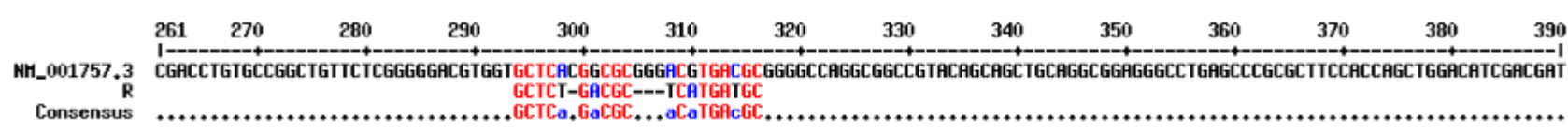
# Stáhněte si sekvenci NM\_001757.3

Obsahuje příslušná sekvence cílové místo pro zadaný primer R1?

>R1

GCTCTGACGCTCATGATGC

Multalin:



Sequence Manipulation Suite:

Reverse Complement

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

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

GCTCTGACGCTCATGATGC

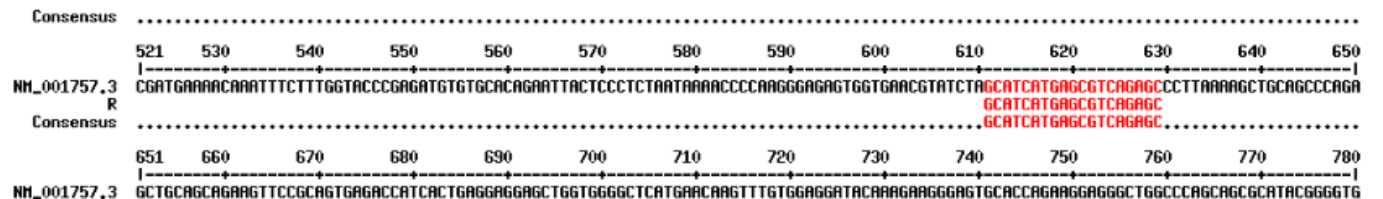
Submit Clear Reset

reverse-complement

Reverse Complement results

>Untitled reverse complement

GCATCATGAGCGTCAGAGC



→Ano.



# Stáhněte si sekvenci NM\_001757.3

Obsahuje příslušná sekvence cílové místo pro zadaný primer R1?

>R1

GCTCTGACGCTCATGATGC

## NCBI: Graphic

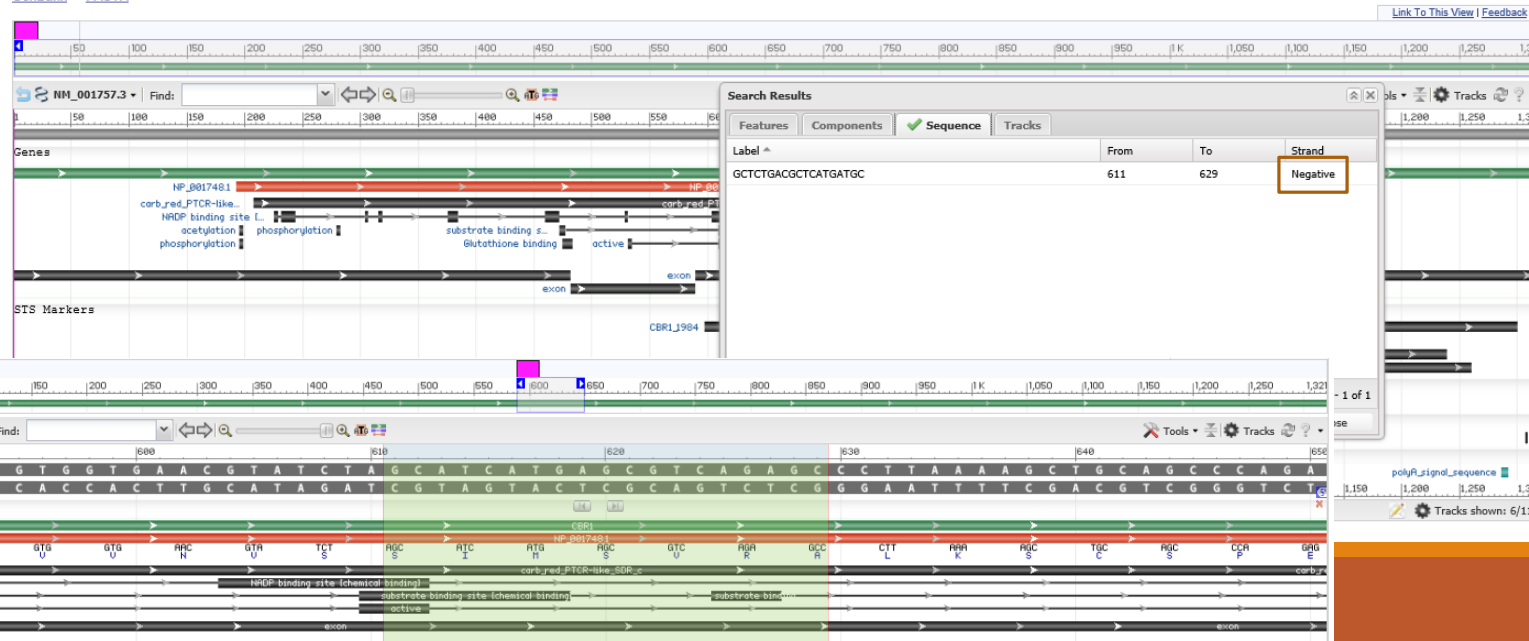
Graphics ▾

Send to: ▾

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

NCBI Reference Sequence: NM\_001757.3

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



→Ano.

# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?

Obsahuje příslušná sekvence cílové místo pro zadaný **primer R1**?

>R1

GCTCTGACGCTCATGATGC

V jakém exonu leží cílová sekvence tohoto primeru?

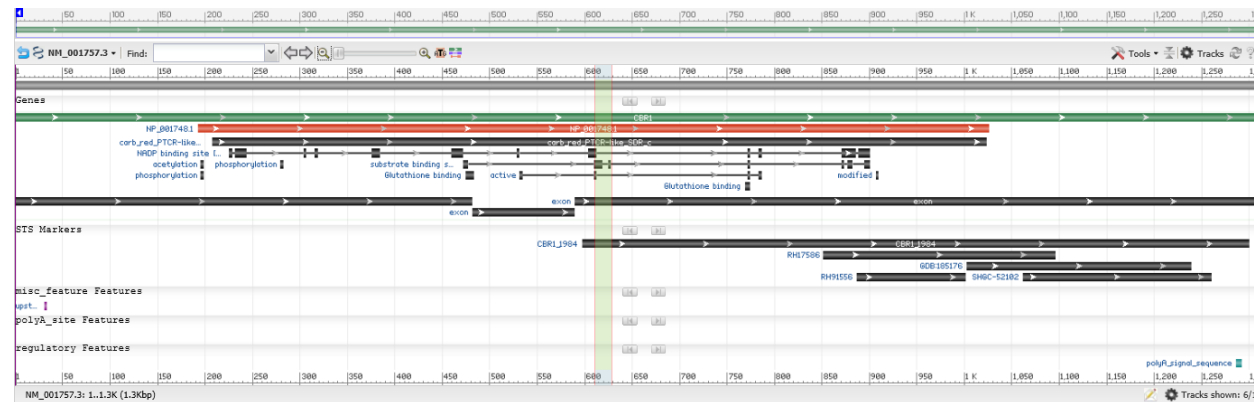
Bylo by možné použít tento primer při PCR s Ta 60°C?

Navrhněte k tomuto primeru vhodný F primer.

# Stáhněte si sekvenci NM\_001757.3

V jakém exonu leží cílová sekvence tohoto primeru?

NCBI: Graphic



>R1

GCTCTGACGCTCATGATGC

→ ve třetím exonu

# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?

Obsahuje příslušná sekvence cílové místo pro zadaný **primer R1**?

>R1

GCTCTGACGCTCATGATGC

V jakém exonu leží cílová sekvence tohoto primeru?

Bylo by možné použít tento primer při PCR s  $T_a$  60°C?

Navrhněte k tomuto primeru vhodný F primer.

# Stáhněte si sekvenci NM\_001757.3

Bylo by možné použít tento primer při PCR s Ta 60°C?

OligoCalc:

>R1

GCTCTGACGCTCATGATGC

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 AIG 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

50 nM Primer 1 Measured Absorbance at 260 nanometers

50 mM Salt (Na<sup>+</sup>)

Calculate Swap Strands BLAST mfold

Physical Constants Melting Temperature (T<sub>m</sub>) Calculations

Length: 19 Molecular Weight: 5779.8<sup>g</sup> GC content: 58%

1 ml of a sol'n with an Absorbance of 1 at 260 nm

is 5.265 microMolar<sup>g</sup> and contains 30.4 micrograms.

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)

Check Self-Complementarity

→Ano.

# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?

Obsahuje příslušná sekvence cílové místo pro zadaný primer R1?

>R1

GCTCTGACGCTCATGATGC

V jakém exonu leží cílová sekvence tohoto primeru?

Bylo by možné použít tento primer při PCR s Ta 60°C?

Navrhněte k tomuto primeru vhodný F primer (pro detekci genu).



# Stáhněte si sekvenci NM\_001757.3

---

Co tento gen kóduje?

Obsahuje příslušná sekvence cílové místo pro zadaný **primer R1**?

>R1

GCTCTGACGCTCATGATGC

V jakém exonu leží cílová sekvence tohoto primeru?

Bylo by možné použít tento primer při PCR s Ta 60°C?

Navrhněte k tomuto primeru vhodný F primer.

## Výsledek:

Tento gen je karbonyl reductasa 1 (CBR1).

Cílové místo sekvence obsahuje.

Leží ve třetím exonu.

Při 60°C by šel použít.

Vhodný F primer by mohl být například : F:CAAGGTTGCTGATCCCACAC



# Stáhněte si sekvenci NM\_001757.3

## Výsledek:

Tento gen je karbonyl reduktasa 1 (CBR1).

Cílové místo sekvence obsahuje.

Leží ve třetím exonu.

Při 60°C by šel použít.

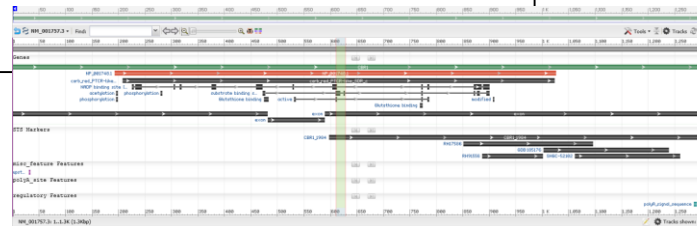
Vhodný F primer by mohl být například : F:CAAGGTTGCTGATCCCACAC

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

NCBI Reference Sequence: NM\_001757.3

[FASTA](#) [Graphics](#)

```
NM_001757.3 CGATGAAACAAATTTCTTTGGTACCCGAGATGTGTGCACAGATTACTCCCTTAATAAAACCCAGGGAGAGTGGTGAACGATCTAGCATCATGAGCGTCAGAGCCCTTAAAGCTGC  
R  
Consensus .....GCATCATGAGCGTCAGAGC  
GCATCATGAGCGTCAGAGC  
.....  
NM_001757.3 | 651 660 670 680 690 700 710 720 730 740 750 760 770  
NM_001757.3 GCTGCAGCAGAGTTCGCACTGAGCCATCCTGAGGAGGAGCTGGTGGGGCTCATGACACAGTTTGTGGAGGATACAAAGAGGGAGTGCACCAGAGGGAGGGCTGGCCACGACGCGCAT
```



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

Template masking not selected  
No mispriming library specified  
Using 1-based sequence positions

OLIGO	start	len	tm	gc%	any th	3' th	hairpin	seg
LEFT PRIMER	477	20	59.12	55.00	0.00	0.00	0.00	CAAGGTTGCTGATCCCACAG
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

# Zkoušková písemka

---

2 úlohy, každá 5 bodů

## Verze 2

Stáhněte si sekvenci **NM\_005247.2**

- Co tato sekvence kóduje?
- Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?
- Porovnejte příslušný protein s myším (Mus musculus) homologem, jaká je jejich shoda?
- Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy NdeI a KpnI
- Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

Stáhněte si k tomuto příkladu sekvenci peptidu **S2**.

- K jakému lidskému proteinu tento peptid pravděpodobně patří?
- Obsahuje peptid nějaké typické domény?
- Jaká je molekulová hmotnost tohoto peptidu?
- Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?
- Zapište identifikovan(ou)é mutac(i)e.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

NCBI/Nucleotide: **Homo sapiens fibroblast growth factor 3 (FGF3), mRNA**

NCBI Reference Sequence: NM\_005247.2

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

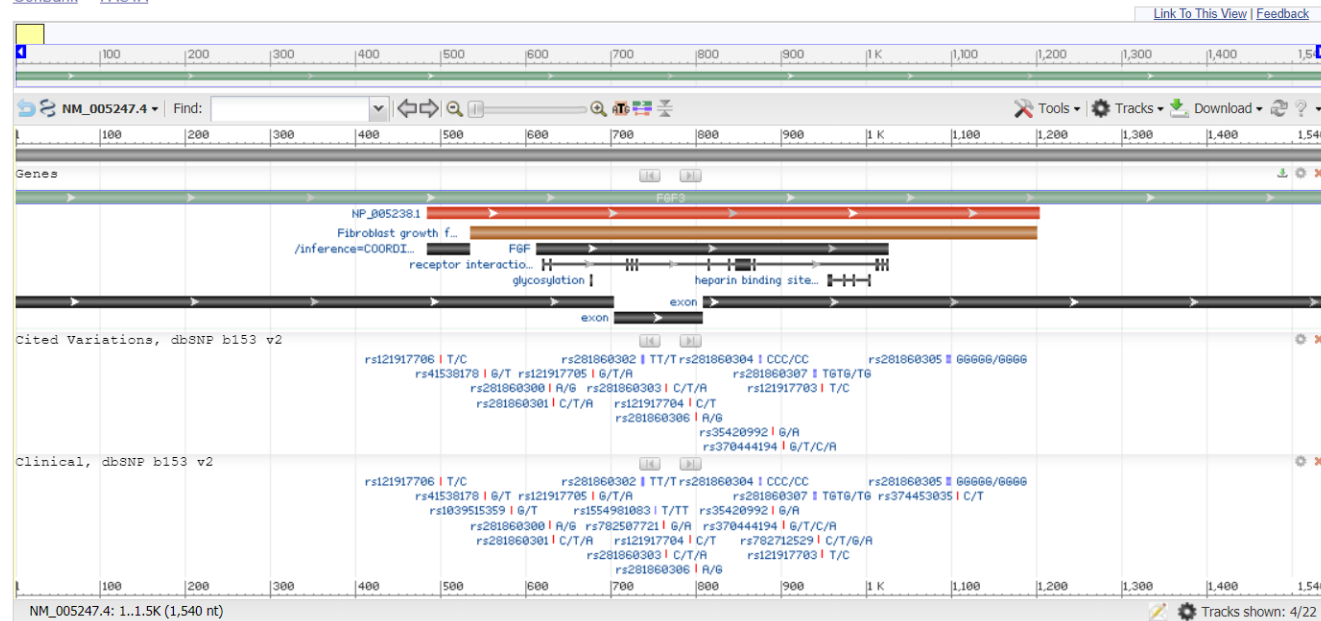
Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

- ve všech třech:

## Homo sapiens fibroblast growth factor 3 (FGF3), mRNA

NCBI Reference Sequence: NM\_005247.4

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



# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.



# Stáhněte si nukleotidovou sekvenci NM\_005247.2

Porovnejte příslušný protein s myším (Mus musculus) homologem, jaká je jejich shoda?

The image shows a GenBank record for NM\_005247.2. It includes a DNA sequence and a summary table with the following columns:

GETTING STARTED	RESOURCES	POPULAR	FEATURED	NCBI INFORMATION
NCBI Education	Chemicals & Bioassays	PubMed	Genetic Testing Registry	About NCBI
NCBI Help Manual	Data & Software	BioRxiv	PubMed Health	Research at NCBI
NCBI Handbook	DNA & RNA	PubMed Central	GenBank	NCBI News & Blog
Training & Tutorials	Domains & Structures	PubMed Health	Reference Sequences	NCBI FTP Site
Submit Data	Genes & Expression	BLAST	Gene Expression Omnibus	NCBI on Facebook
	Genetics & Medicine	Nucleotide	Genome Data Viewer	NCBI on Twitter
	Genomes & Maps	Genomes	Human Genome	NCBI on YouTube
	Homology	SNP	Mouse Genome	
	Literature	Gene	Influenza Virus	
	Proteins	Protein	Primer-BLAST	
	Sequence Analysis	PubChem	Sequence Read Archive	
	Taxonomy			
	Variation			

At the bottom, there is a navigation bar with buttons for 'Details', 'Display', 'FASTA', 'GenBank', and 'Help'.

## Emboss Needle:

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

EMBOSS_001      1  MGLIWLLLLLLEPGWPAAGPGARLRDRDAGGGRGVYVHLLGGAPRRKLYC  50
|||||
FGF3_MOUSE      1  MGLIWLLLLLLEPSPFTTGPGLRRDAGGGRGVYVHLLGGAPRRKLYC    50

EMBOSS_001      51  ATKYHLQLHPGSRVNGSLNSAYSILEITAVEVGVIVAIRGLFSGRYLMM  100
|||||
FGF3_MOUSE      51  ATKYHLQLHPGSRVNGSLNSAYSILEITAVEVGVIVAIRGLFSGRYLMM  100

EMBOSS_001      101 KRGRLYASEHYSAECEFVERIHLYNTYASRLYRVSTPGARRQPSPAE   150
|||||
FGF3_MOUSE      101 KRGRLYASDHYNAECEFVERIHLYNTYASRLYRTGSSGPGARRQPGAQ   150

EMBOSS_001      151 RLWYVSVNGKGRFRFGFKTRRTQKSSLFLPRLVLRDHRDHEMVRQLQGLPR  200
|.....|
FGF3_MOUSE      151 RFWYVSVNGKGRFRFGFKTRRTQKSSLFLPRLVLRDHRDHEMVRQLQSQFR  200

EMBOSS_001      201 PFGKGVQPRRRRQ-KQSPDNLEPSHWQSRLQGLASAH-----  239
|.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
FGF3_MOUSE      201 APGEKQPRRRQKQKQSPGD----HGKMETLSTRATPSTQLHTGGLAVA  245
```

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

KpnI ggtac c	289
NdeI ca tatg	none

→ KpnI vhodné nebude

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

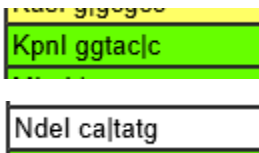
Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy NdeI a KpnI

```
>NM_005247.2:492-1211 Homo sapiens fibroblast growth factor 3 (FGF3), mRNA
ATGGGCCTAATCTGGCTGCTACTGCTCAGCCTGCTGGAGCCCGGCTGGCCCCGAGCGGGCCCTGGGGCGC
GGTTGCGGGCGCGATGCGGGCGGGCCGTGGCGGGCTCTACGAGCACCTTGGCGGGGGCGCCCCGGCGCCGCAA
GCTCTACTGCGCCACGAAGTACCACCTCCAGCTGCACCCGAGCGGCCGCGTCAACGGCAGCCTGGAGAAC
AGCGCCTACAGTATTTTGGAGATAACGGCAGTGGAGGTGGGCATTGTGGCCATCAGGGGTCTCTTCTCCG
GGCGGTACCTGGCCATGAACAAGAGGGGACGACTCTATGCTTCGGAGCACTACAGCGCCGAGTGGCGAGTT
TGTGGAGCGGATCCACGAGCTGGGCTATAATACGTATGCCCTCCCGGCTGTACCGGACGGTGTCTAGTACG
CCTGGGGCCCCGCCGAGCCAGCCGAGAGACTGTGGTACGTGTCTGTGAACGGCAAGGGCCGGCCCC
GCAGGGGCTTCAAGACCCGCCGACACAGAAGTCCCTCCCTGTTCCCTGCCCCGCGTGTGGACCACAGGGA
CCACGAGATGGTGCGGCAGCTACAGAGTGGGCTGCCAGACCCCTGGTAAGGGGGTCCAGCCCCGACGG
CGGCGGCAGAAGCAGAGCCCGGATAACCTGGAGCCCTCTCACGTTTCAGGCTTCGAGACTGGGCTCCCAG
TGGAGGCCAGTGCGCACTAG
```



Například:

F:ttt catatg **ATGGGCCTAATCTGGC**

R:ttt ggtacc **CTAGTGCGCACTG**

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

---

Co tato sekvence kóduje?

X Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?

Porovnejte příslušný protein s myším (*Mus musculus*) homologem, jaká je jejich shoda?

Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy NdeI a KpnI

Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

## Výsledek:

Tento gen je Fibroblast Growth Factor 3

Významné polymorfismy jsou v prvním a třetím exonu. X

Shoda v rámci celé sekvence je 79.1 %.

Primery mohou být například takovéto: F:TTT CAT ATG ATG GGC CTA ATC TGG C , R: TTT GGT ACC CTA GTG CGC ACT G

Restriční endonukleasa KpnI vhodná pro klonování není (štěpí sekvenci v pozici 289).

# Stáhněte si nukleotidovou sekvenci NM\_005247.2

## Výsledek:

Tento gen je Fibroblast Growth Factor 3

Významné polymorfismy jsou v prvním a třetím exonu. X

Shoda v rámci celé sekvence je 79.1 %.

Primery mohou být například takovéto: F:TTT CAT ATG ATG GGC CTA ATC TGG C, R: TTT GGT ACC CTA GTG CGC ACT G

Restrikční endonukleasa KpnI vhodná pro klonování není (štěpí sekvenci v pozici 289).

## Homo sapiens fibroblast growth factor 3 (FGF3), mRNA

NCBI Reference Sequence: NM\_005247.2



```
#-----  
#  
# Aligned_sequences: 2  
# 1: EMBOSSE_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 MGLIWLILLSLLEPGWPAAGFARLRDAGSGGYYEHLGGAFRRKLYC 50  
FGF3_MOUSE 1 MGLIWLILLSLLEPGWPTTGPGRLRDAGSGGYYEHLGGAFRRKLYC 50  
EMBOSS_001 51 ATKYHLQLHPSGRVNSLENSAYSILEITAVEVGIVAIRGLFSGRYLAM 100  
FGF3_MOUSE 51 ATKYHLQLHPSGRVNSLENSAYSILEITAVEVGIVAIRGLFSGRYLAM 100  
EMBOSS_001 101 KRGRLYASEHYSAECEPVERIHELGYNTYASRLYRIVSSTFGARRQPSAE 150  
FGF3_MOUSE 101 KRGRLYASDHNAECEPVERIHELGYNTYASRLYRIVSSTFGARRQPSAE 150  
EMBOSS_001 151 RLWYVSVNGWRFRGFRRTQKSSLLFPLRVLDHRDHMVRQLQSGSLFR 200  
FGF3_MOUSE 151 RFWYVSVNGWRFRGFRRTQKSSLLFPLRVLDHRDHMVRQLQSGSLFR 200  
EMBOSS_001 201 PFGKGVQPRRRQ-KQSPFNLEPSHVQASRLGSLASAH----- 239  
FGF3_MOUSE 201 APFGSGVQPRRRQKQSPFGD----HGKMETLSTRAIFSTQLHTGGLAVA 245
```

Například:

F:ttt catatg **ATGGGCCTAATCTGGC**

R:ttt ggtacc **CTAGTGCGCACTG**

NdeI ca tatg	none
KpnI ggtac c	289



# Zkoušková písemka

---

2 úlohy, každá 5 bodů

## Verze 2

Stáhněte si sekvenci **NM\_005247.2**

- Co tato sekvence kóduje?
- Ve kterých exonech se vyskytují významné polymorfismy (GMAF)?
- Porovnejte příslušný protein s myším (Mus musculus) homologem, jaká je jejich shoda?
- Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy NdeI a KpnI
- Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.

Stáhněte si k tomuto příkladu sekvenci peptidu **S2**.

- K jakému lidskému proteinu tento peptid pravděpodobně patří?
- Obsahuje peptid nějaké typické domény?
- Jaká je molekulová hmotnost tohoto peptidu?
- Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?
- Zapište identifikovan(ou)é mutac(i)e.

Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

>S2

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEKRRVRRERNKLAAAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVPSC
```

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

**>S2**

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEEKRRVRRER NKLA AAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKI PYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

K jakému lidskému proteinu tento peptid pravděpodobně patří?

The image shows a BLAST search interface with the following components:

- BLAST >> blastp suite** header.
- Navigation tabs: `blastn`, `blastp`, `blastx`, `tblastn`, `tblastx`.
- Enter Query Sequence** section:
  - Input field: `>S2`  
`CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEKRRVRRERNKLA AAKCRNRRRELT`  
`DRLQAE TDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKI PYEEGPGGPLAEVRD`  
`LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLIASLFTHSEVQVLGDPFPVWNPSC`
  - Buttons: `Clear`, `Query subrange` (From: , To: )
  - Options: `Or, upload file` (Procházet...), `Job Title` (S2), `Align two or more sequences` (checkbox).
- Choose Search Set** section:
  - Database: `Reference proteins (refseq_protein)`
  - Organism: `Homo sapiens (taxid:9606)` (Optional)
  - Exclude: `Models (XM/XP)`, `Uncultured/environmental sample sequences` (checkboxes).
  - Entrez Query:  (Optional)
- Results** section:
  - Putative conserved domains: `coiled coil`, `dimer interface`.
  - Specific hits: `bZIP_Fos`.
  - Superfamilies: `bZIP superfamily`.
  - Distribution of the top 44 Blast Hits on 44 subject sequences (Color key for alignment scores: `<40`, `40-50`, `50-80`, `80-200`, `>=200`).

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

NCBI Reference Sequence: NP\_006723.2

[Identical Proteins](#) [FASTA](#) [Graphics](#)

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

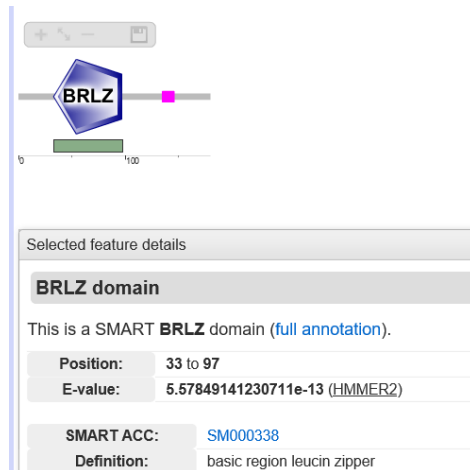
>S2

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVPSC
```

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

Obsahuje peptid nějaké typické domény?

PFAM, CDD, SMART..



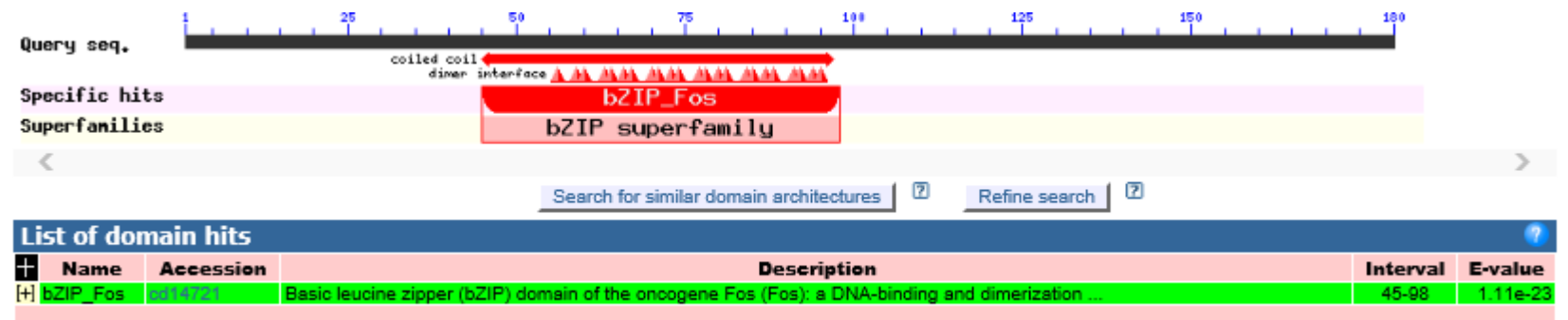
Selected feature details

**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



Query seq. 25 50 75 100 125 150 180

coiled coil  
dimer interface

Specific hits  
Superfamilies

bZIP\_Fos  
bZIP superfamily

Search for similar domain architectures ? Refine search ?

**List of domain hits**

Name	Accession	Description	Interval	E-value
bZIP_Fos	SM014721	Basic leucine zipper (bZIP) domain of the oncogene Fos (Fos): a DNA-binding and dimerization ...	45-98	1.11e-23

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

>S2

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEEEKRRVRRER NKLA AAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKI PYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPV VNPSC
```



# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

Jaká je molekulová hmotnost tohoto peptidu?



## Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

## Sequence Analysis

- Codon Plot
- Codon Usage
- CpG Islands

## Sequence Manipulation Suite:

### Protein Molecular Weight

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

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

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEKRRVRRERNKLAAAKCRNRRRELT
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGGGLAEVRD
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVPSC
```

Submit

Clear

Reset

• Add  copies of  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

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

>S2

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEEKRRVRRER NKLA AAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKI PYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPV VNPSC
```

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

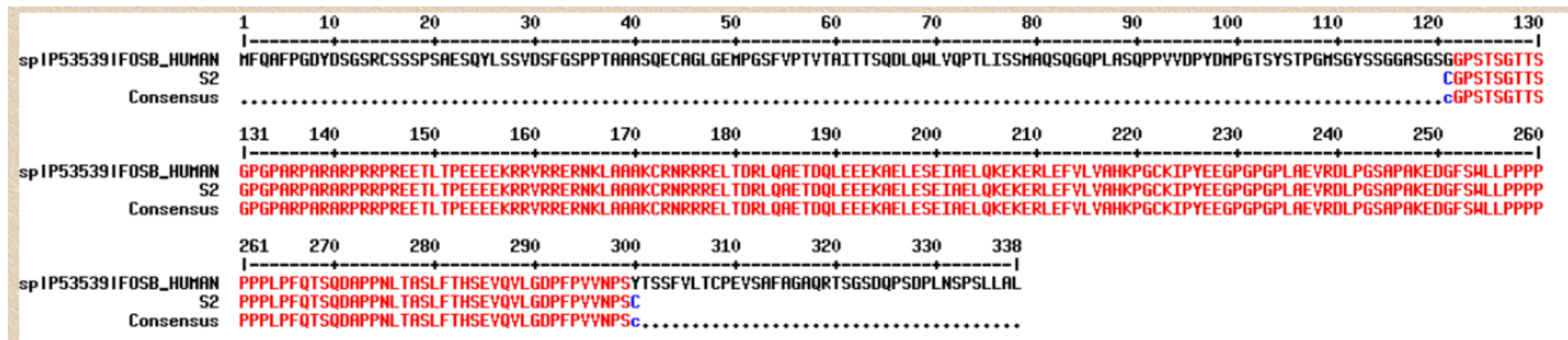
Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Porovnání-Multalin:

## protein fosB isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP\_006723.2

[Identical Proteins](#) [FASTA](#) [Graphics](#)



→Liší se v první a poslední aminokyselině

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

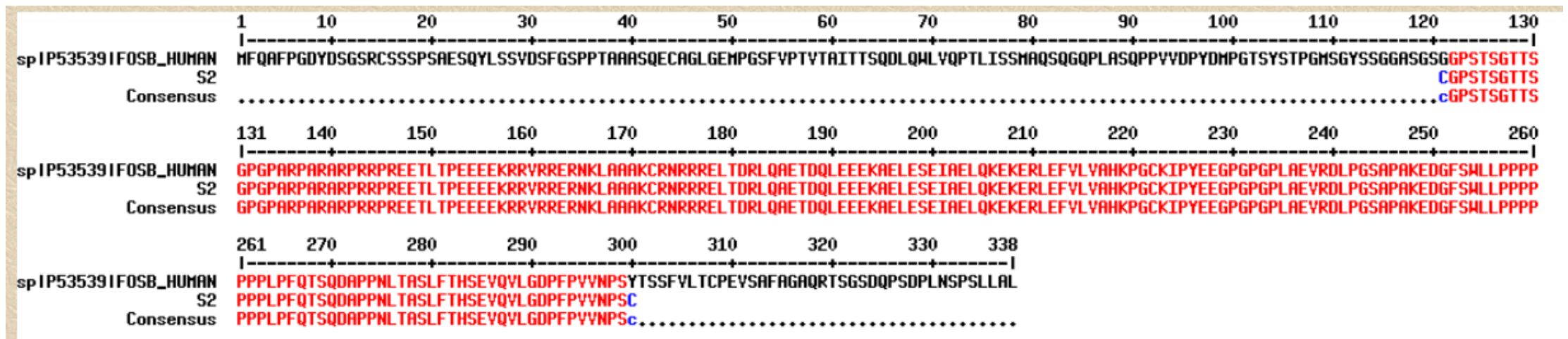
Zapište identifikovan(ou)é mutac(i)e.

>S2

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEEEKRRVRRER NKLA AAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKI PYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC
```

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

Zapište identifikovan(ou)é mutac(i)e.



→G121C

→Y300C

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

---

K jakému lidskému proteinu tento peptid pravděpodobně patří?

Obsahuje peptid nějaké typické domény?

Jaká je molekulová hmotnost tohoto peptidu?

Jak se liší zadaná sekvence peptidu od identifikované (lidské) sekvence (v rámci shodného úseku)?

Zapište identifikovan(ou)é mutac(i)e.

## Výsledek:

Peptid je nejpravděpodobněji protein FosB.

Ano-obsahuje bZIP (leucinový zip).

Mw=19.1 kDa.

Liší se ve dvou aminokyselinách, první a poslední.

Zapsané mutace vůči původnímu proteinu by byly: G121C, Y300C.

# Stáhněte si k tomuto příkladu sekvenci peptidu S2.

## Výsledek:

Peptid je nejpravděpodobněji protein FosB.

Ano-obsahuje bZIP (leucinový zip).

Mw=19.1 kDa.

Liší se ve dvou aminokyselinách, první a poslední.

Zapsané mutace vůči původnímu proteinu by byly: G121C, Y300C.

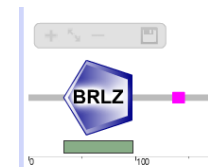
	1	10	20	30	40	50	60	70	80	90	100	110	120	130
sp P53539 FOSB_HUMAN	HFQRFPGDYDGSRCSSSPSRESQYLSSYDFFGSPPTAAASQECAGLGEHPGFSFYPTVTAITTSQDLQALVQPTLSSHAQSQGQPLASQPPVYDYPGTSYSTPGMSGYSSGGASGSG													CGPSTSGTTS
S2														CGPSTSGTTS
Consensus														CGPSTSGTTS
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
sp P53539 FOSB_HUMAN	GGPARPARARPRRPREETLTPEEEKRRVREANLAAKCRNRRELTDRLQRETDQLEEKAELESETAELQKEKERLEFVLYAHKPGCKIPYEEGGPGPLAEVRLPGSAPAKEDGFSALLPPP													GGPSTSGTTS
S2	GGPARPARARPRRPREETLTPEEEKRRVREANLAAKCRNRRELTDRLQRETDQLEEKAELESETAELQKEKERLEFVLYAHKPGCKIPYEEGGPGPLAEVRLPGSAPAKEDGFSALLPPP													GGPSTSGTTS
Consensus	GGPARPARARPRRPREETLTPEEEKRRVREANLAAKCRNRRELTDRLQRETDQLEEKAELESETAELQKEKERLEFVLYAHKPGCKIPYEEGGPGPLAEVRLPGSAPAKEDGFSALLPPP													GGPSTSGTTS
	261	270	280	290	300	310	320	330	338					
sp P53539 FOSB_HUMAN	PPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPEPVVNPSTYTSFVLTCPVSAFAGQRTSGSDQPSDPLNPSLLAL													
S2	PPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPEPVVNPSC													
Consensus	PPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPEPVVNPSC													

G121C  
Y300C

## protein fosB isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP\_006723.2

[Identical Proteins](#) [FASTA](#) [Graphics](#)



Selected feature details	
<b>BRLZ domain</b>	
This is a SMART <b>BRLZ</b> domain ( <a href="#">full annotation</a> ).	
Position:	33 to 97
E-value:	5.57849141230711e-13 (HMMER2)
SMART ACC:	SM000338
Definition:	basic region leucin zipper

## Protein Molecular Weight results

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

19.91 kDa

# Další „zkouškové verze“

---



# Stáhněte si sekvenci S1

---

Zjistěte, co tato sekvence pravděpodobně kóduje.

Zjistěte jaký je nejdelší čtecí rámeček.

Odovídá tato sekvence identifikovanému proteinu (z prvního bodu).

Obsahuje identifikovaný protein nějaké transmembránové úseky?

Jaké fragmenty vzniknou po štěpení enzymy BamHI, KpnI, NdeI?

# Stáhněte si sekvenci S1

---

Zjistěte, co tato sekvence pravděpodobně kóduje.

[BLASTn](#)

Zjistěte jaký je nejdelší čtecí rámeček.

[ORFfinder](#)

Odovídá tato sekvence identifikovanému proteinu (z prvního bodu).

[Multalin](#)

Obsahuje identifikovaný protein nějaké transmembránové úseky?

[Př. TMMOD](#)

Jaké fragmenty vzniknou po štěpení enzymy BamHI, KpnI, NdeI.

[Restriction Digest](#)



# Najděte proteinovou sekvenci lidské CBR1

---

Zapište přístupový kód a funkci tohoto proteinu.

Najděte také příslušnou nukleotidovou sekvenci.

Jak dlouhý je 2. exon?

Kolik CG nukleotidů obsahuje 3'UTR oblast tohoto genu?

Zjistěte jakou mutaci byste vnesli do tohoto genu při použití zadaných mutačních primerů.

```
F:TGC CGG CTG TTC GGC GGG GAC GTG GTG  
R:CAC CAC GTC CCC GCC GAA CAG CCG GCA
```

# Najděte proteinovou sekvenci lidské CBR1

---

Zapište přístupový kód a funkci tohoto proteinu.

[Uniprot / NCBI-protein](#)

Najděte také příslušnou nukleotidovou sekvenci.

[NCBI nucleotide / refseq odkaz skrz Uniprot](#)

Jak dlouhý je 2. exon?

[NCBI Graphic/ highlight sequence features](#)

Kolik CG nukleotidů obsahuje 3'UTR oblast tohoto genu?

[DNA stats](#)

Zjistěte jakou mutaci byste vnesli do tohoto genu při použití zadaných mutačních primerů.

[Multalin, Graphic, Genetic code](#)

# Najděte proteinovou sekvenci lidské CBR1

Zapište přístupový kód a funkci tohoto proteinu.

Najděte také příslušnou nukleotidovou sekvenci.

Jak dlouhý je 2. exon?

Kolik CG nukleotidů obsahuje 3' UTR oblast tohoto genu?

Zjistěte jakou mutaci byste vnesli do tohoto genu při použití zadaných mutačních primerů.

## P16152 (CBR1\_HUMAN)

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

NCBI Reference Sequence: NM\_001757.3

```
361 ccgcgcttcc accagctgga catgacgat ctgcagagca tccgcgcctc gcgcgacttc
421 ctgcgcaagg agtacggggg cctggacgtg ctggtcaaca acgcgggcat cgccttcaag
481 gttgctgata ccacacctt tcatattcaa gctgaagtga cgatgaaaac aaattctttt
541 ggtaccggag atgtgtgcac agaattactc cctctaataa aaccccaagg gagagtgtg
601 aacgtatcta gcatcatgag cgtcagagcc cttaaaagct gcagccocaga gctgcagcag
661 caattacaga atagagcat caatagaga caatagaga caatagaga caatagaga
```

```
482..589
/gene="CBR1"
/gene_synonym="CBR; hCBR1; SDR21C1"
/inference="alignment:Splice:1.39.8"
```

108 nt

3' UTR oblast (nepřekládaná)

927..1028

Obsahuje 2 dinukleotidy CG

