

Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2023/2024

9/10

Exam

Thursday 2.5.2024 16:30

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

Usual Exam test:

2 parts, each five points

Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

Version 1

Find human protein sequence **FGF3**

- What is the accession number and function of this protein?
- How many valines does the sequence contain?
- How long is the longest fragment after trypsin digestion?
- How many human similar reference proteins are in the databases?
- Does the sequence contain any transmembrane regions?

Download sequence **NM_001757.3**

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

```
>R1  
GCTCTGACGCTCATGATGC
```

- In which exon is the complementary sequence for this primer?
- Is the primer suitable for PCR with Ta=60°C?
- Design suitable F primer for this experiment.

Previous Exams

Find human protein sequence FGF3

What is the accession number and function of this protein?

How many valines does the sequence contain?

How long is the longest fragment after trypsin digestion?

How many human similar reference proteins are in the databases?

Does the sequence contain any transmembrane regions?

Find human protein sequence FGF3

What is the accession number and function of this protein?

How many valines does the sequence contain?

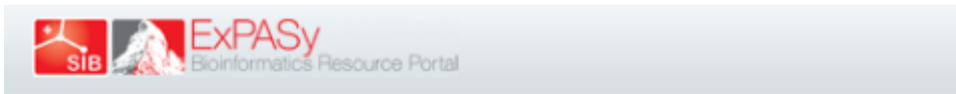
How long is the longest fragment after trypsin digestion?

How many human similar reference proteins are in the databases?

Does the sequence contain any transmembrane regions?

Find human protein sequence FGF3

How long is the longest fragment after trypsin digestion?



PeptideCutter

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

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

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

Perform

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	MGLIWLLLLLLEPGWPAAGPGAR	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	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	LYASEHYSACEFVER	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	VLDR	5	638.724
192	Trypsin	DHEMVR	6	785.873
204	Trypsin	QLQSLPRPPGK	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?



NCBI Resources | How To | Sign in to NCBI

Protein Advanced Help

GenPept Send to:

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; Catarrhini; Hominidae; 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) Query subrange

MGLIWL...
SGRVN...
IHELGY...
RVLDHR...

From
To

Or, upload file

Job Title

Align two or more sequences

Choose Search Set

Database +

Organism Exclude +

Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query [YouTube](#) [Create custom database](#)

Find human protein sequence FGF3

How many human similar reference proteins are in the databases?



BLAST » blastp suite

blastn | **blastp** | blastx | tblastn | tblastx

Enter Query Sequence

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

```
MGLIWLILLSLLEPGWPAAGPGARLRRDAGGRGGVYEHGGAPRRRKLKLYCA  
SGRVNGSLNSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHY  
IHELGYNTYASRLYRTVVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTR  
RVLDHRDHEMVRQLQSGLPFRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLG
```

Or, upload file

Job Title

Align two or more sequences

Choose Search Set

19 sequences selected

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

Query seq. receptor interaction site FGF heparin binding site (glycine box)

Specific hits FGF

Superfamilies FGF superfamily

Distribution of the top 19 Blast Hits on 19 subject sequences

The chart shows the alignment of 19 subject sequences (represented by horizontal bars) against a query sequence (represented by a teal bar at the top). The x-axis is labeled from 1 to 200. The bars are color-coded: purple, blue, green, and red. A red box highlights the text "19 subject sequences" in the chart's title.

NCBI Resources | How To | Sign in to NCBI

Protein | Search

fibroblast growth factor 3 precursor [Homo sapiens]

NCBI Reference Sequence: NP_005238.1

Identical Proteins | 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

LOCUS NP_005238.1 239 aa linear PRI 09-09-2010

DEFINITION fibroblast growth factor 3 precursor [Homo sapiens].

ACCESSION NP_005238

VERSION NP_005238.1

DBSOURCE REFSeq; accession: NP_005238.1

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhina; Catarrhini; Hominoidea; Hominidae; Homo.

REFERENCE 1. (residues 1 to 239)

ATTNONS Tony W and Park DE. Genetically confirmed thanatophoric dysplasia with fibroblast growth factor 3 mutations. *Am J Hum Genet* 2014

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

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.

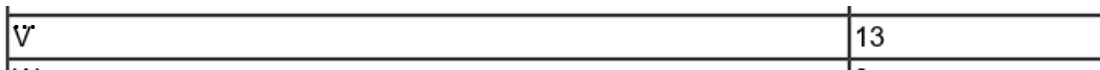
In the databes of reference sequences, there is 19 similar proteins.

There is no transmembrane helices.

[P11487 \(FGF3_HUMAN\)](#) / [NP_005238.1](#)

Fibroblast growth factor 3

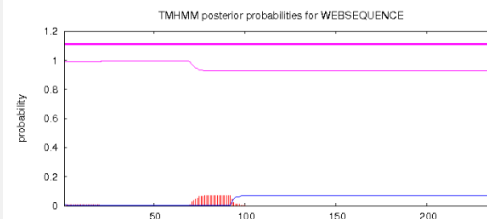
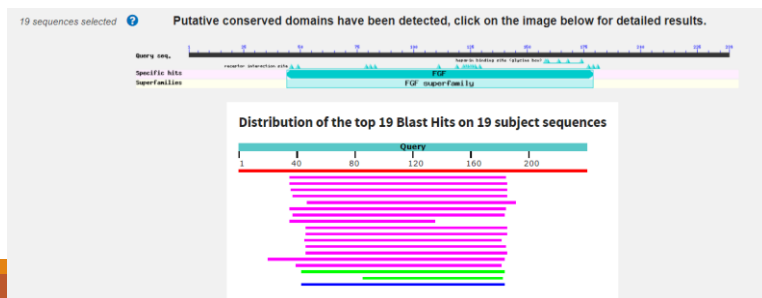
-role in the regulation of embryonic development...



Trypsin

VNGSLENSAYSILEITAVEVGIVAIR

26



Exam test:

Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

Version 1

Find human protein sequence **FGF3**

- What is the accession number and function of this protein?
- How many valines does the sequence contain?
- How long is the longest fragment after trypsin digestion?
- How many human similar reference proteins are in the databases?
- Does the sequence contain any transmembrane regions?

Download sequence **NM_001757.3**

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

```
>R1  
GCTCTGACGCTCATGATGC
```

- In which exon is the complementary sequence for this primer?
- Is the primer suitable for PCR with Ta=60°C?
- Design suitable F primer for this experiment.

Download sequence NM_001757.3

What does this sequence encode ?

Is there a complementary sequence to primer R1?

>R1

GCTCTGACGCTCATGATGC

In which exon is the complementary sequence for this primer?

Is the primer suitable for PCR with $T_a=60^\circ\text{C}$?

Design suitable F primer for this experiment.

Download sequence NM_001757.3

What does this sequence encode ?

Is there a complementary sequence to primer R1?

>R1

GCTCTGACGCTCATGATGC

In which exon is the complementary sequence for this primer?

Is the primer suitable for PCR with $T_a=60^\circ\text{C}$?

Design suitable F primer for this experiment.

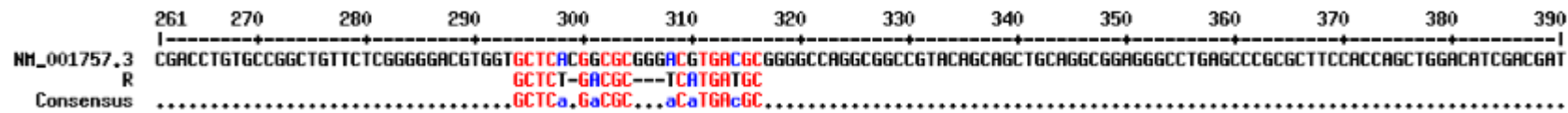
Download sequence NM_001757.3

Is there a complementary sequence to 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
```

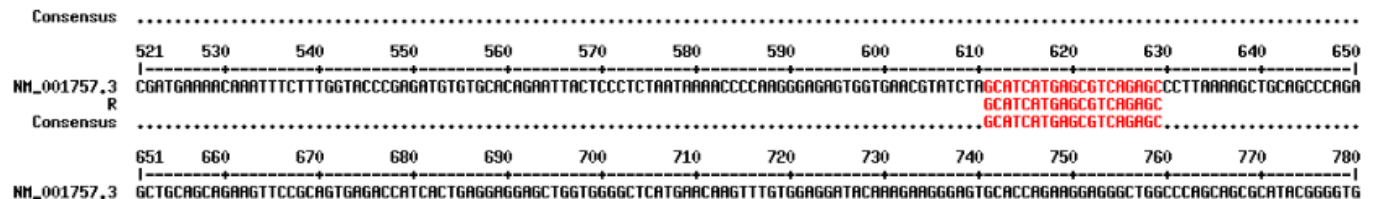
Submit Clear Reset

reverse-complement

Reverse Complement results

>Untitled reverse complement

GCATCATGAGCGTCAGAGC



→yes.

Download sequence NM_001757.3

Is there a complementary sequence to primer R1?

>R1

GCTCTGACGCTCATGATGC

NCBI: Graphic

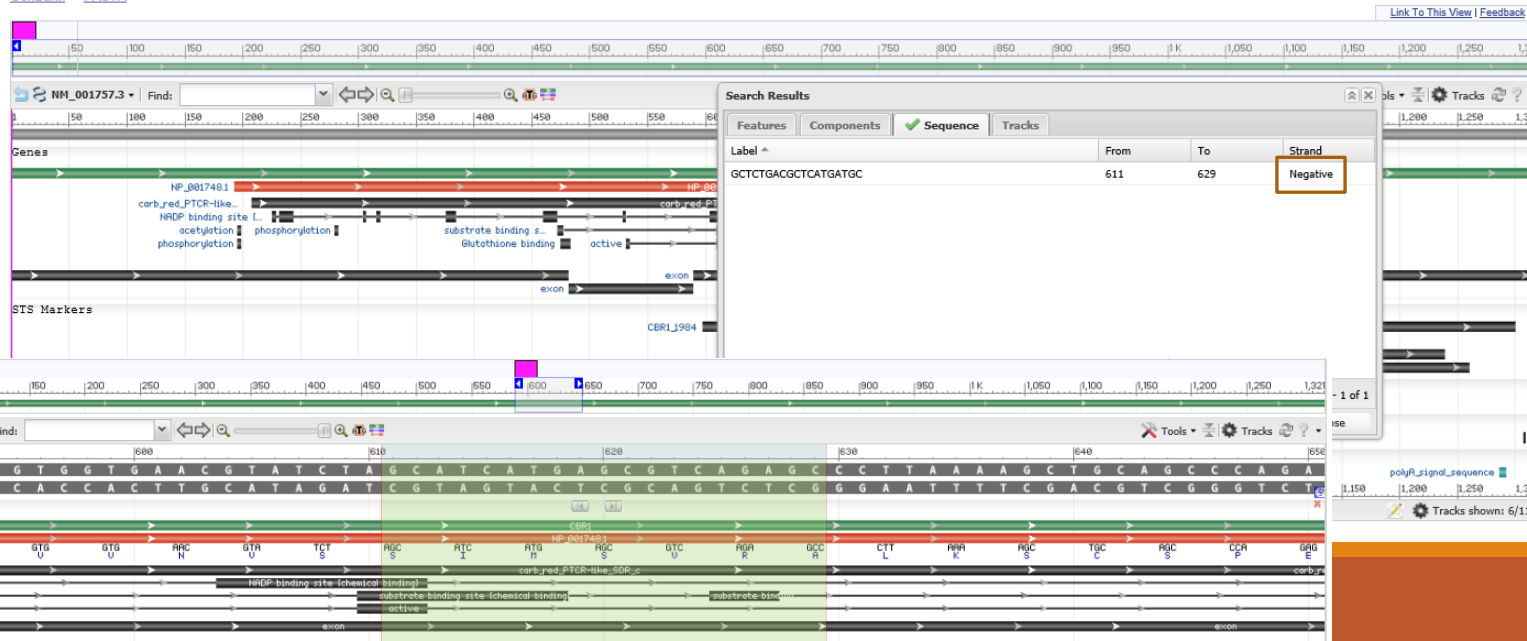
Graphics ▾

Send to: ▾

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

NCBI Reference Sequence: NM_001757.3

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



→yes.

Download sequence NM_001757.3

What does this sequence encode ?

Is there a complementary sequence to primer R1?

>R1

GCTCTGACGCTCATGATGC

In which exon is the complementary sequence for this primer?

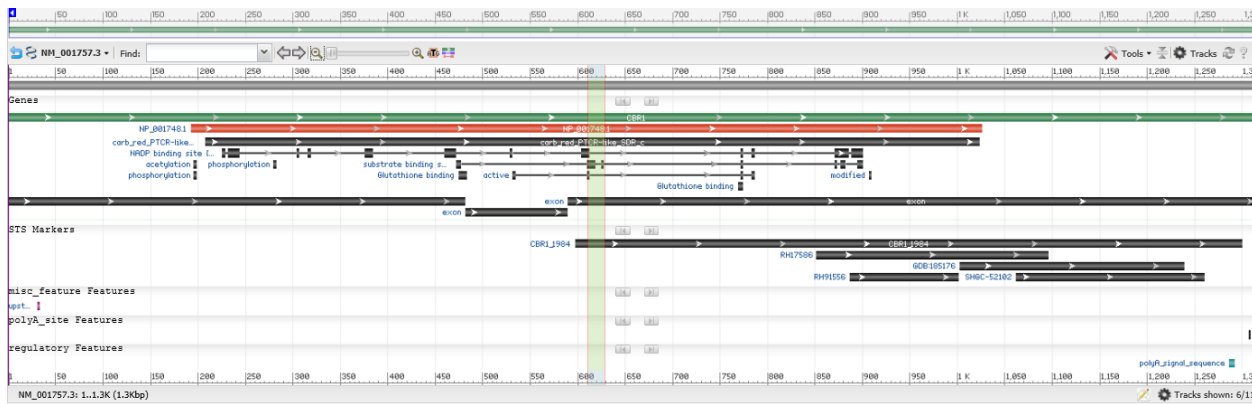
Is the primer suitable for PCR with $T_a=60^\circ\text{C}$?

Design suitable F primer for this experiment.

Download sequence NM_001757.3

In which exon is the complementary sequence for this primer?

NCBI: Graphic



>R1

GCTCTGACGCTCATGATGC

→ third exon

Download sequence NM_001757.3

What does this sequence encode ?

Is there a complementary sequence to primer R1?

>R1

GCTCTGACGCTCATGATGC

In which exon is the complementary sequence for this primer?

Is the primer suitable for PCR with $T_a=60^\circ\text{C}$?

(Design suitable F primer for this experiment.)

We did not cover this part

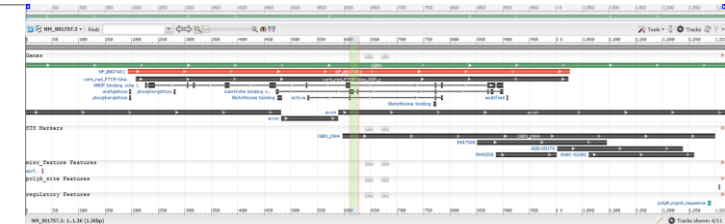
Download sequence NM_001757.3

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](#)

```

.....................................................................
NM_001757.3      CGATGAARACAAATTTCTTTGGTACCCGAGATGTGTGCACAGATTACTCCCTCTAATAAACCACAGGGAGAGTGGTGACGATCTAGCATCATGAGCGTCAGAGCCCTTAAAGCTGCAGCCCAGA
R               GCATCATGAGCGTCAGAGC
Consensus       .....GCATCATGAGCGTCAGAGC.....
651  660  670  680  690  700  710  720  730  740  750  760  770  780
NM_001757.3      GCTGCAGCAGAGTCCGCGTGGAGCCATCACTGAGGAGGAGCTGGTGGGGCTCATGACCAAGTTTGTGGAGGATACAAAGAGGGAGTGCACCAAGAGGAGGGCTGGCCAGCAGCGCATACGGGGTG
    
```

```

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

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

Exam test II:

2 parts, each five points

Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

Version 2

download sequence NM_005247.2

- What does this sequence encode?
- Are there any significant single nucleotide polymorphisms (GMAF), in which exons?
- Would 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 the peptide sequence S2 (from study materials).

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

Download sequence NM_005247.2

What does this sequence encode?

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

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

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

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

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

Download sequence NM_005247.2

What does this sequence encode?

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

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

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

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

Restriction summary:

KpnI ggtac c	289
NdeI caltatg	none

→ KpnI is not suitable

Download sequence NM_005247.2

What does this sequence encode?

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

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

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

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

Solution:

This gene encodes Fibroblast Growth Factor 3

GMAF are in the 1st and 3rd exon.

Restriction endonuclease KpnI is not suitable for cloning.

The FASTA of the protein is:

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

```
>protein-FGF3
MGLIWLLLLSLEPGWPAAGPGARLRRDAGGRGGVYEHGGAPRRRKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSFLP
RVLDHRDHEMVRQLQSGLEPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH*
```

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 identity of human and mouse FGF3 proteins is 79.1 %.

```
>protein-FGF3
```

```
MGLIWLLLLLLEPGWPAAGPGARLRRDAGGRGGVYEHGGAPRRRKLKLYCATKYHLQLHP
SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEPVER
IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSFLFP
RVLDHRDHMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSLQLEASAH*
```

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: EMBOS_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 MGLIWLLLLLLEPGWPAAGPGARLRRDAGGRGGVYEHGGAPRRRKLKLYCATKYHLQLHP 50
FGF3_MOUSE     1 MGLIWLLLLLLEPGWPTTGPGRRLRRDAGGRGGVYEHGGAPRRRKLKLYCATKYHLQLHP 50
EMBOSS_001     51 ATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASE 100
FGF3_MOUSE     51 ATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASE 100
EMBOSS_001     101 KRGRLYASEHYSAECEPVERIHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSV 150
FGF3_MOUSE     101 KRGRLYASDHNAECEPVERIHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSV 150
EMBOSS_001     151 RLVYVSVNGKGRPRRGFKTRRTQKSSFLFRLVLDHRDHMVRQLQSGLPRPPGKGVQ 200
FGF3_MOUSE     151 RLVYVSVNGKGRPRRGFKTRRTQKSSFLFRLVLDHRDHMVRQLQSGLPRPPGKGVQ 200
EMBOSS_001     201 PFGKGVQPRRRRQ-KQSPDNLEPSHVQASRLGSLQLEASAH----- 239
FGF3_MOUSE     201 APFGKGVQPRRRRQ-KQSPDNLEPSHVQASRLGSLQLEASAH----- 239
EMBOSS_001     245 ----- 245
FGF3_MOUSE     245 ----- 245

```

Exam test:

2 parts, each five points

Maximum 10 points

9-10p: 1

7-8p: 2

5-6p: 3

Version 2

download sequence NM_005247.2

- What does this sequence encode?
- Are there any significant single nucleotide polymorphisms (GMAF), in which exons?
- Would 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 the peptide sequence S2 (from study materials).

- Identify the protein.
- Does it contain any typical domains?
- How many cysteines contain the peptide?
- What is the molecular weight of this peptide?
- Is the sequence (within 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 cysteines contain the peptide?

What is the molecular weight of this peptide?

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

>S2

```
CGPSTSGTTS GPGPARPARARPRRPREETLTPEEEEKRRVRRER NKLA AAKCRNRRRELT  
DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKI PYEEGPGPGPLAEVRD  
LPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNL TASFTHSEVQVLGDPFPV VNPSC
```

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 (within 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

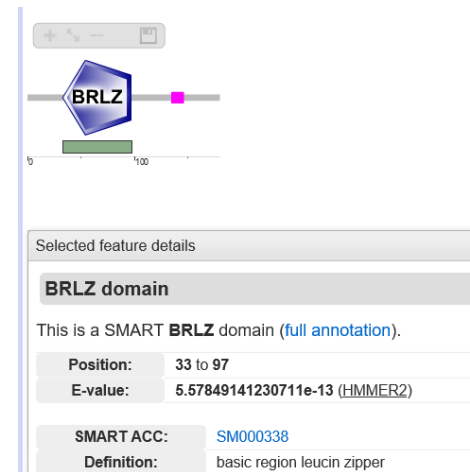
Mw=19.1 kDa.

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](#)



Protein Stats results

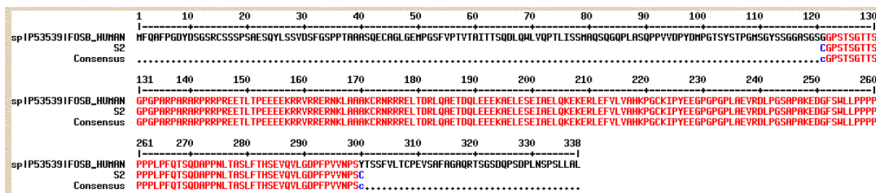
Results for 180 residue sequence "Untitled" starting "CGPSTSGTTS"

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



Version from 2023_new to Moodle

Version I 19.5.2023

A: Download the sequence NM_000946.3

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

B: Work with following peptide sequence:

```
LALASVFWSISYYSSPFAFFLYRKGYSLSKVVVFFSHYAGTLLLLLAGVACLGRIGRWT  
NPQYRQFITILEATHRNQSSSENKRQLANYNDFRSPVDFHWEEPPSSRKE SRGGPSRRGV  
ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVAHTLGRMLYPGSVYLLQKALMPVLLQG  
QARLVEECNGRRAKLLACDGENEIDTMEVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP  
LEACYSVLGWNHHPGFAGSTGVFPQNEANAMDVVVQFAIHRLLGFQPDII IYAWSIGGFT
```

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


A: Download the sequence NM_000946.3

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

A: Download the sequence NM_000946.3

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

What does this sequence encode?

 **National Library of Medicine**
National Center for Biotechnology Information

jostovap

Nucleotide [Help](#)

Advanced

GenBank

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

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

LOCUS	NM_000946	1423 bp	mRNA	linear	PRI 03-APR-2024
DEFINITION	Homo sapiens DNA primase subunit 1 (PRIM1), mRNA.				
ACCESSION	NM_000946				
VERSION	NM_000946.3				
KEYWORDS	RefSeq; MANE Select.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;				

Analyze this sequence

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

[Show in Genome Data Viewer](#)

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

CDS:

[CDS](#)

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

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

Translated protein

```
primase polypeptide 1, 49kDa; DNA primase 49 kDa subunit;  
primase (DNA) subunit 1"  
/codon_start=1  
/product="DNA primase small subunit"  
/protein_id="NP_000937.1"  
/db_xref="CCDS:CCDS44926.1"  
/db_xref="GeneID:5557"  
/db_xref="HGNC:HGNC:9369"
```

DNA primase small subunit [Homo sapiens]

NCBI Reference Sequence: NP_000937.1

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

Go to:

LOCUS	NP_000937	420 aa	linear	PRI 03-APR-2024
DEFINITION	DNA primase small subunit [Homo sapiens].			
ACCESSION	NP_000937			
VERSION	NP_000937.1			
DBSOURCE	REFSEQ: accession NM_000946.3			
KEYWORDS	RefSeq; MANE Select.			

```
26..1288  
/gene="PRIM1"  
/gene_synonym="p49; PDIL"  
/EC_number="2.7.7.102 "  
/note="primase p49 subunit; DNA primase subunit 48;  
primase, DNA, polypeptide 1 (49kDa); DNA primase 1;  
primase polypeptide 1, 49kDa; DNA primase 49 kDa subunit;  
primase (DNA) subunit 1"  
/codon_start=1  
/product="DNA primase small subunit"  
/protein_id="NP_000937.1 "  
/db_xref="CCDS:CCDS44926.1 "  
/db_xref="GeneID:5557 "  
/db_xref="HGNC:HGNC:9369 "  
/db_xref="MIM:176635 "  
/translation="METFDPTLPELLKLYRRLFPYQYRRLNYGGVIKNYFQHRE  
FSFTLKDDIYIRYQSFNNQSDLEKEMQKINPYKIDIGAVYSHRPNQHTVKLGAFQAQ  
EKELVFDIDMTDYDDVRRCCSSADICPKCWTLMTMAIRIIDRALKEDFGFKHRLIIVYS  
GRRGVHCWVCDSEVRKLSAVRSIGIVEYLSLVKGGQDVKKVHLSKIHPPFKRSINI  
IKKYFEEYALVNQDILENKESWDKILALVPETIHDELQQSFKSHNSLQRWEHLKKVA  
SRYNQNIKNDKYGPWLEWIMLQYCFPRLDINVSKGINHLKSPFSVHPKTRGISVPI  
DLQKVDQDFPPTVPTISFICRELDAISTNEEKEEAEASDVKXRTDRYKTS LAPYV  
KVFHFLENLDSRKGELKKSLQKDF"
```

[Details](#)

Display: [FASTA](#) [GenBank](#) [Help](#)

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

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

Go to:

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

Other possibility:

Graphics

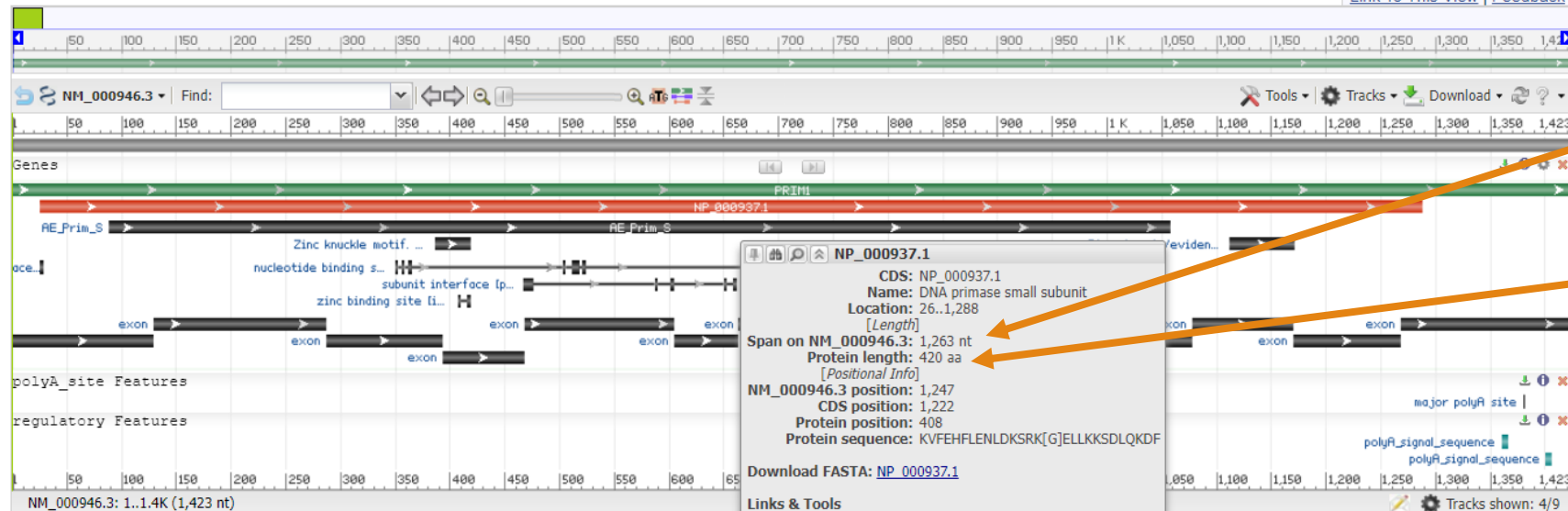
Send to:

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

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

[Link To This View](#) | [Feedback](#)



CDS

protein

Other option for translated protein:

Translate CDS:



Sequence Manipulation Suite:

Translate

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
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim
- Multi Rev Trans
- Mutate for Digest
- ORF Finder
- Pairwise Align Codons
- Pairwise Align DNA
- Pairwise Align Protein
- PCR Primer Stats
- PCR Products

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

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

```
>NM_000946.3:26-1288 Homo sapiens DNA primase subunit 1  
(PRIM1), mRNA  
ATGGAGACGTTTGACCCACCGAGCTGCCCGAGCTGCTTAAACTTTATTACCGAGGCTCT  
TTCCCTACT  
CTCAGTACTATCGCTGGCTCAACTACGGTGGAGTGATAAAGAATTACTTTCAACACCGTGA  
ATTTTCATT
```

Submit Clear Reset

- Translate in
- Use the

*This page requires JavaScript. See
*You can mirror this page or use it

Sun 14 Jun 00:37:01 2020
Valid XHTML 1.0; Valid CSS

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

Range Extractor Protein results

```
>results for 420 residue sequence "Untitled" starting "METFDPTLPE"  
M
```

Protein Molecular Weight results

Results for 420 residue sequence "Untitled" starting "METFDPTLPE"
49.91 kDa

Other option for translated protein: **math**

CDS=1263nt

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

Manually design primers for PCR amplification of the CDS.

F: ATG GAG ACG TTT GAC CCC AC

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

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

>NM_000946.3:26-1288 Homo sapiens DNA primase subunit 1 (PRIM1), mRNA
ATGGAGACGTTTGACCCCACCGAGCTGCCGAGCTGCTTAACTTTATTACCGGAGGCTCTTTCCTACT
CTCAGTACTATCGCTGGCTCACTACGGTGGAGTGATAAAGAATTACTTTCAACACCGTGAATTTTCATT
CACATTGAAAGATGATATTTACATTTCGCTACCAATCCTTCAACAACCAGAGTGATCTGGAAAAGGAGATG
CAGAAAATGAATCCATACAAGATTGATATAGGCGCAGTATATTCTCACAGACCCAATCAACACAATACAG
TGAAGCTGGGAGCTTTCCAGGCTCAGGAAAAAGAAGTGGTATTGACATTGACATGACAGACTATGACGA
TGTGAGGAGATGTTGTAGTTCGACAGACATATGTCCTAAGTGC TGGACCCATGACAATGGCCATACGC
ATCATTGACAGAGCATTGAAGGAGGACTTTGGATTTAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAG
GTGTTTCATTGTTGGGTCTGTGATGAATCAGTTAGAAAACGTCTTCTGCAGTACGTTCTGGGATAGTTGA
GTATTTGAGCCTTGTAAGGGTGGTCAAGACGTTAAAAAGAAAGTTCACCTAAGTGAAAAAATTCACCCCT
TTTATCAGAAAATCTATAAACATAATAAAAAAATACTTTGAAGAAATATGCCTTGGTTAATCAAGATATTC
TCGAAAATAAAGAAAGCTGGGATAAGATTTAGCCCTTGTTCGAAACAATTCATGATGAACTTCAACA
AAGCTTCCAAAAGTCTCACAATTCACCTCAGCGTTGGGAGCAC TTGAAGAAAGTAGCCAGCAGATATCAG
AATAACATCAAAAATGACAAATATGGACCC TGGCTGGAGTGGGAGATTATGCTCCAGTACTGTTTTCCAC
GGCTGGATATCAATGTGACAAAGGAATCAATCATCTACTGAAGAGCCCTTTTAGTGTTTCATCCTAAAAC
AGGTCGCATATCTGTGCCTATTGATTTGCAGAAAGTGGACCAGTTGATCCATTTACTGTTCCGACCATA
AGCTTCATCTGCCGTGAATTGGATGCCATTTCCACTAATGAAGAGGAAAAAGAGGAGAATGAAGCTGAAT
CTGATGTCAAACATAGAACCAGAGATTATAAGAAGACCAGTCTAGCACCTTATGTGAAAAGTTTTTGAACA
TTTTCTTGAAAATCTGGATAAATCCCAGAAAAGGAGAAGTCTTAAGAAGAGTGATTTACAAAAAGATTTCC
TGA

Oligo Calc: Oligonucleotide Properties Calculator

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

Nucleotide base codes

ATG GAG ACG TTT GAC CCC AC

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

GTG GGG TCA AAC GTC TCC AT

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

50 nM Primer 1 Measured Absorbance at 260 nanometers

50 mM Salt (Na⁺)

Calculate Swap Strands BLAST mfold

Physical Constants		Melting Temperature (T _M) Calculations	
Length:	20	Molecular Weight:	6102.4
GC content:	55%	1	53.8 °C (Basic)
1 ml of a sol'n with an Absorbance of 1 at 260 nm		2	60.5 °C (Salt Adjusted)
is 4.718 microMolar and contains 28.8 micrograms.		3	52.18 °C (Nearest Neighbor)

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

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

Deprecated Hairpin/self dimerization calculations

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

4 (Minimum base pairs required for a hairpin)

Check Self-Complementarity

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

R:TCA GAA ATC TTT TTG TAA ATC ACT CTT

Reverse complement the whole CDS:

Reverse Complement results

```
>NM_000946.3:26-1288 Homo sapiens DNA primase subunit 1 (PRIM1),  
TCAGAAATCTTTTGTAAATCACTCTTCTTAAGAAGTTCTCCTTTTCGGGATTTATCCAG  
ATTTTCAAGAAAATGTTCAAAAACCTTTCACATAAGGTGCTAGACTGGTCTTCTTATAATC  
TCTGGTTCTATGTTTGACATCAGATTCAGCTTCATTCTCCTCTTTTTCCTCTTCATTAGT  
GGAAATGGCATCCAATTCACGGCAGATGAAGCTTATGGTCGGAACAGTAAATGGATCAA  
CTGGTCCACTTTCTGCAAATCAATAGGCACAGATATGCGACCTGTTTTAGGATGAACACT  
AAAAGGGCTCTTCAGTAGATGATTGATTCCTTTGCTGACATTGATATCCAGCCGTGGAAA  
ACAGTACTGGAGCATAATCTCCCACTCCAGCCAGGGTCCATATTTGTCATTTTTGATGTT  
ATTCTGATATCTGCTGGCTACTTTCTTCAAGTGCTCCCAACGCTGAAGTGAATTGTGAGA  
CTTTTGAAGCTTTGTTGAAGTTCATCATGAATTGTTTCAGGAACAAGGGCTAAAATCTT  
ATCCAGCTTTCTTTATTTTCGAGAATATCTTGATTAACCAAGGCATATTCTTCAAAGTA  
TTTTTTTATTATGTTTATAGATTTCTGATAAAAGGGTGAATTTTTTCACTTAGGTGAAC  
TTTCTTTTAAACGTCTTGACCACCCTTTACAAGGCTCAAATACTCAACTATCCCAGAACG  
TACTGCAGAAGACAGTTTCTAACTGATTCATCACAGACCCAACAATGAACACCTCTCCT  
TCCAGAATATACCCAGAGACGATGCTTAAATCCAAAGTCTCCTTCAATGCTCTGTCAAT  
GATGCGTATGGCCATTGTCATGAGGGTCCAGCACTTAGGACATATGTCTGCAGAACTACA  
ACATCTCCTCACATCGTCATAGTCTGTCAATGTCAAATACCAGTTCTTTTTCCTG  
AGCCTGGAAAGCTCCAGCTTCACTGTATTGTGTTGATTGGGTCTGTGAGAATACTGC  
GCCTATATCAATCTTGTATGGATTCATTTTCTGCATCTCCTTTTCCAGATCACTCTGGTT
```

Oligo Calc: Oligonucleotide Properties Calculator

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

Nucleotide base codes

TCA GAA ATC TTT TTG TAA ATC ACT CTT

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

AAG AGT GAT TTA CAA AAA GAT TTC TGA

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

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

Calculate Swap Strands BLAST mfold

Physical Constants Melting Temperature (T_M) Calculations

Length: 27 Molecular Weight: 8198.4 GC content: 26%
1 ml of a sol'n with an Absorbance of 1 at 260 nm
is 3.55 microMolar and contains 29.1 micrograms.

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.
RlnK 33.404 cal/(°K*mol) deltaH 203.6 Kcal/mol
deltaG 30.1 Kcal/mol deltaS 543 cal/(°K*mol)

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

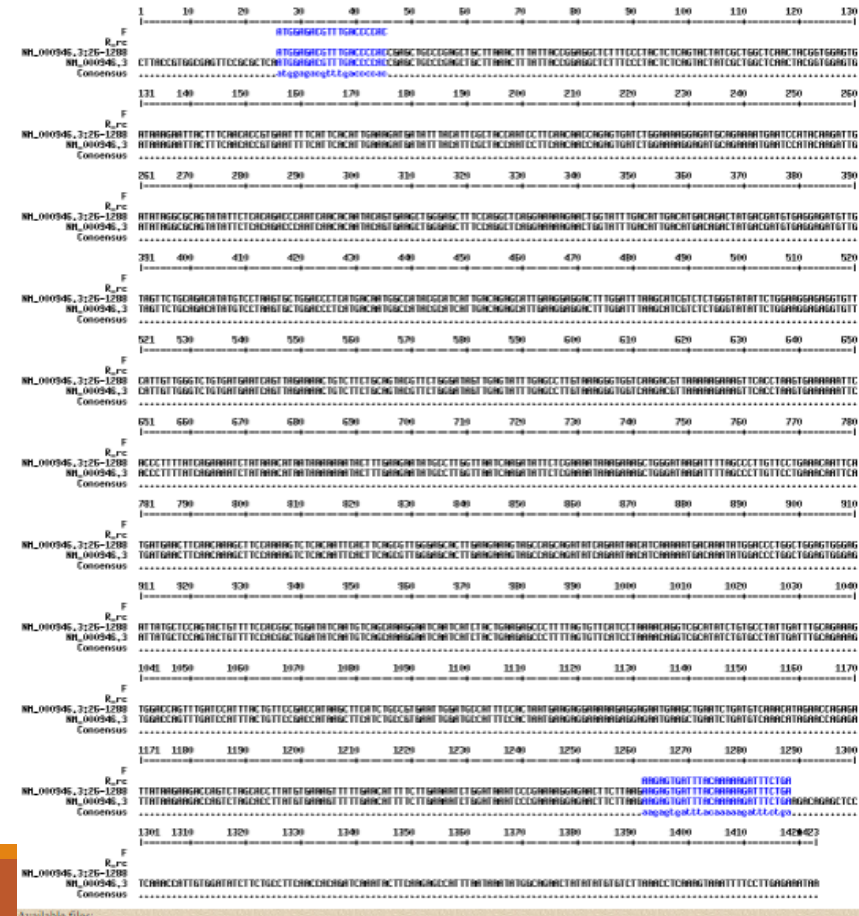
Check Self-Complementarity

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

Multalin (don't forget to reverse complement the primer for comparison)

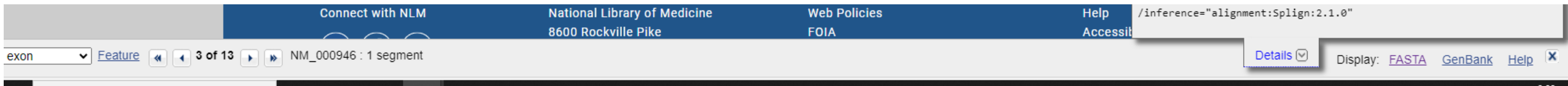
```
>F
ATG GAG ACG TTT GAC CCC Aq
>R rc
AAG AGT GAT TTA CAA AAA GAT TTC TGA

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



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

Get the 3rd exon:



Restriction summary:

Answer: **yes NlaIII**

Sequence Manipulation Suite:

Restriction Summary

Restriction Summary accepts a DNA sequence and returns the number and positions of commonly used restriction endonuclease cut sites. Use this program on a segment of DNA.

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

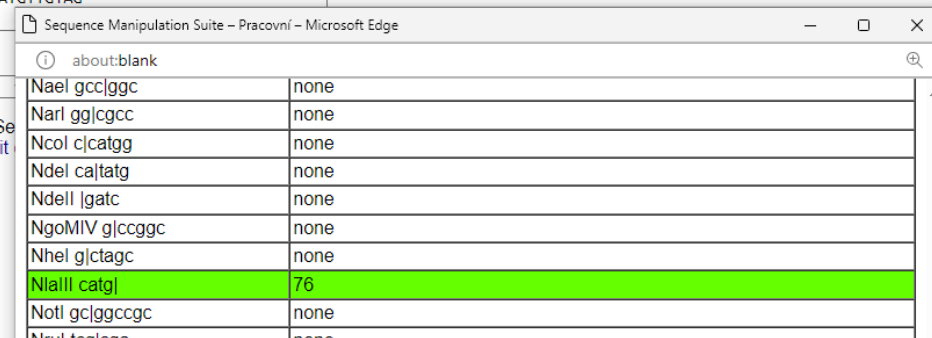
```
>NM_000946.3:287-393 Homo sapiens DNA primase subunit 1  
(PRIM1), mRNA  
CCCAATCAACACAATACAGTGAAGCTGGGAGCTTCCAGGCTCAGGAAAAGAACTGGTATT  
TGACATTG  
ACATGACAGACTATGACGATGTGAGGAGATGTTGTAG
```

Submit Clear Reset

• Treat sequences as linear

*This page requires JavaScript. See
*You can mirror this page or use it

Sun 14 Jun 00:37:00 2020
Valid XHTML 1.0; Valid CSS



Enzyme	Cut Site	Position
NaeI	gcc ggc	none
NarI	gg cgcc	none
NcoI	c catgg	none
NdeI	ca tatg	none
NdeII	gatc	none
NgoMIV	g ccggc	none
NheI	g ctagc	none
NlaIII	catG	76
NotI	gc ggccgc	none
NruI	tc agca	none

B: Work with following peptide sequence:

LALASVFWSISYYSSPFAFFYLRYRKGYSLSKVVVPSHYAGTLLLLLAGVACLRGIGRWT
NPQYRQFITILEATHRNQSSSENKRQLANYNFDFRSWPVDFHWEEPSRKESTRGGPSRRGV
ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVVAHTLGRRMLYPGSVYLLQKALMPVLLQG
QARLVEECNGRRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP
LEACYSVLGWNHPGFAGSTGVFPFPQNEANAMDVVVQFAIHRLGFQPQDIIIIYAWSIGGFT

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

To which human protein this peptide probably belongs?

BLASTp

BLAST® » blastp suite

blastn **blastp** blastx tblastn tblastx

Standard Protein BLAST

BLASTP programs search protein databases using

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From To

LALASVFWISYISYSSPFAFFYLRYKGYLSLSKVVPFSHYAGTLLLLAGVACL
RGIQRWT
NPQYRQFITILEATHRNQSENKRQLANYNFDFRSWPVDFHWEPPSSRKES
RGGPSRRGV

Or, upload file [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases Standard databases (nr etc.): **New** Experimental databases [Try experimental clustered nr database](#) [?](#)
For more info see [What is clustered nr?](#)

Compare Select to compare standard and experimental database [?](#)

Standard

Database [?](#)

Organism exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm [?](#)

BLAST® » blastp suite » results for RID-316XNRMD013 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

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

Job Title **Protein Sequence**

RID [316XNRMD013](#) Search expires on 05-01 15:20 pm [Download All](#) [v](#)

Program BLASTP [?](#) [Citation](#) [v](#)

Database refseq_protein [See details](#) [v](#)

Query ID lcl|Query_7536389

Description unnamed protein product

Molecule type amino acid

Query Length 300

Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

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

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

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

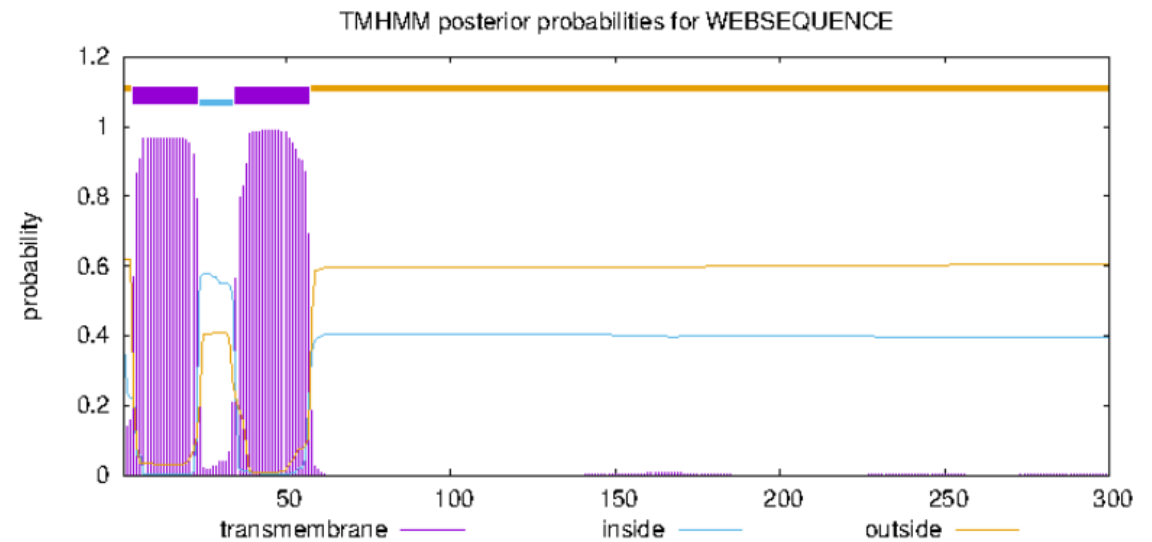
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X1 (Gorilla gorilla gorilla)	Gorilla gorilla gorilla	625	625	100%	0.0	99.67%	601	XP_004043725.4
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X3 (Sapajus apella)	Sapajus apella	624	624	100%	0.0	99.67%	457	XP_032124883.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X3 (Callithrix jacchus)	Callithrix jacchus	624	624	100%	0.0	99.67%	457	XP_035151139.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X1 (Callithrix jacchus)	Callithrix jacchus	623	623	100%	0.0	99.67%	558	XP_035151136.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X1 (Aotus nancymaae)	Aotus nancymaae	623	623	100%	0.0	99.67%	558	XP_012292099.1
<input checked="" type="checkbox"/>	protein ABHD16A isoform X1 (Theropithecus gelada)	Theropithecus gelada	623	623	100%	0.0	99.00%	601	XP_025238181.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform a (Homo sapiens)	Homo sapiens	623	623	100%	0.0	99.67%	558	NP_066983.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X2 (Aotus nancymaae)	Aotus nancymaae	623	623	100%	0.0	99.67%	522	XP_012292100.1
<input checked="" type="checkbox"/>	protein ABHD16A isoform X2 (Macaca mulatta)	Macaca mulatta	623	623	100%	0.0	99.00%	587	XP_028703550.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X1 (Sapajus apella)	Sapajus apella	623	623	100%	0.0	99.67%	558	XP_032124881.1
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X1 (Saimiri boliviensis boliviensis)	Saimiri boliviensis boliviensis	623	623	100%	0.0	99.67%	558	XP_003944391.1
<input checked="" type="checkbox"/>	protein ABHD16A isoform X1 (Macaca mulatta)	Macaca mulatta	623	623	100%	0.0	99.00%	601	XP_00112776.2
<input checked="" type="checkbox"/>	phosphatidylserine lipase ABHD16A isoform X2 (Callithrix jacchus)	Callithrix jacchus	622	622	100%	0.0	99.67%	523	XP_035151137.1

Does this peptide contain any transmembrane helices?

Yes 2.

TMHMM result

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



What is molecular weight of this peptide?

SMS

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
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim

Sequence Manipulation Suite: Protein Molecular Weight

Protein Molecular Weight accepts one or more protein sequences and calculates molecular weight. You can append copies of wish to predict the location of a protein of interest on a gel in relation to a set of protein standards.

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

```
LALASVFWSISYYSSPFAFFYLRYKGYLSLSKVVPFSHYAGTLLLLLAGVACLGRIGRWT
NPQYRQFITLEATHRNQSSSENKRQLANYNFDFRSWPVDFHWEESPSSRKE SRGGPSRRGV
ALLRPEPLHRGTADTLNLRVKKLPCQITSYLVAHTLGRMLYPGSVYLLQKALMPVLLQG
QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPPQGQKLVICCEGNAGFYEVGCVSTP
LEACYSVLGWNHPGFAGSTGVPPFQNEANAMDVVVVQFAIHLRGFQPQDIIYAWSIGGFT
```

Submit Clear Reset

• Add copies of

Protein Molecular Weight results
Results for 300 residue sequence "Untitled" starting "LALASVFWSI"
33.67 kDa

*This page requires JavaScript. See [this page](#) for more information.
*You can mirror this page or use it [here](#).

Sun 14 Jun 00:36:59 2020
Valid XHTML 1.0 - Valid CSS

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

Differs in 1 aminoacid

