# 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

Vers	sion 1
Find hu • • •	man protein sequence <b>FGF3</b> 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?
-	boes the sequence contain any transmembrane regions:
Downlo	ad segence NM_001757.3
•	What does this sequence encode ?
•	Is there a complementary sequence to primer R1?
	>R1 GCTCTGACGCTCATGATGC
•	In which exon is the complemetary sequence for this primer?
	Is the primer suitable for PCR with Ta=60°C?

# Previous Exams

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?

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?

### How long is the longest fragment after trypsin digestion?

		Position of cleave	age site Name of cleaving	enzyme(s) Resulting peptide sequence (see explana	tions) Peptide length [aa	] Peptide mass [D
		24	Trypsin	MGLIWLLLLSLLEPGWPAAGPGAR	24	2532.085
	Destal	26	Trypsin	LR	2	287.362
SID Sale a Biomormatics Hesource	Ponal	27	Trypsin	R	1	174.203
		32	Trypsin	DAGGR	5	474.474
		44	Trypsin	GGVYEHLGGAPR	12	1212.330
		46	Trypsin	RR	2	330.390
PeptideCutter		47	Trypsin	K	1	146.189
		53	Trypsin	LYCATK	6	697.847
PeptideCutter [references /	documentation] predicts potential cleavag	63	Trypsin	YHLQLHPSGR	10	1207.357
		89	Trypsin	VNGSLENSAYSILEITAVEVGIVAIR	26	2718.100
Estas - UsiDasti/D (Osiss D	ent en T-EMPL ) mentein identifiere ID (n. e. A	95	Trypsin	GLFSGR	6	635.721
Enter a UniProtKB (Swiss-Pi	rot or Trembl) protein identifier, ID (e.g. A	101	Trypsin	YLAMNK	6	738.900
MGLIWLLLLSLLEPGWPAAGPGARI	RRDAGGRGGVYEHLGGAPRRRKLYC	102	Trypsin	R	1	174.203
ATKYHLOLHP	~	104	Trypsin	GR	2	231.255
COUNCELENCAVELLETTAVEVEL	VATECI ESCEVI AMMUDECTI VASEU	120	Trypsin	LYASEHYSAECEFVER	16	1933.079
SGRVNGSLENSAISILEIIAVEVGI	VAIRGLISGRILAMWRRGRLIASEN	132	Trypsin	IHELGYNTYASR	12	1423.548
1 SAECEFVER		135	Trypsin	LYR	3	450.538
IHELGYNTYASRLYRTVSSTPGARF	QPSAERLWYVSVNGKGRPRRGFKTR	144	Trypsin	TVSSTPGAR	9	874.949
RTQKSSLFLP		145	Trypsin	R	1	174.203
RVLDHRDHEMVROLOSGLPRPPGKG	VOPRRROKOSPDNLEPSHVOASRL	151	Trypsin	QPSAER	6	686.723
GSOLEASAH		160	Trypsin	LWYVSVNGK	9	1065.237
C C Q D D D D D D D D D D D D D D D D D		164	Trypsin	GRPR	4	484.559
		165	Trypsin	R	1	174.203
	Planas indicate the way you would like the class	168	Trypsin	GFK	3	350.418
	F lease indicate the way you would like the clea	vage site 170	Trypsin	TR	2	275.308
		171	Trypsin	R	1	174.203
	☑ Map of cleavage sites. Please select the number	174	Trypsin	IQK	3	375.425
	Initial of cleavage sites. Flease select the number	FOLAITIII 181	Trypsin	SSLFLPR	7	818.971
Perform   the cleavage of th	Table of sites sorted alphabetically by enzyme	and chen	l rypsin	VLDHR	5	638.724
	Table of eiter, control appropriately by enzyme	192	Trypsin	DHEMVR	6	785.873
	✓ Lable of sites, sorted sequentially by amino acid	a number 204	l rypsin	QLQSGLPRPPGK	12	1277.489
		209	l rypsin	GVQPR	5	555.635
Disease select		212	Trypsin	RRR	3	486.578
r lease, select		214	l rypsin		2	2/4.320
		//9	LEVDSID	USPUNLEPSHVUASK	15	1004 (5/

239

end of sequence

LGSQLEASAH

10

1012.087

O all available enzymes and chemicals

only the following selection of enzymes and chemicals

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?

### How many human similar reference proteins are in the databases?



S NCBI R	esources 🕑 How To 🕑		Sign in to NCBI
Protein	Protein v Advanced	Search	Help
GenPept 🚽		Send to: 🗸	Change region shown
fibrobla: NCBI Referer	st growth factor 3 precursor [Homo sapiens] ace Sequence: NP_005238.1		Customize view
Identical Protei	ns FASTA Graphics		Analyze this sequence
LOCUS	NP_005238 239 aa linear PRI 09-MAY-2018		Identify Conserved Domains
ACCESSION	NP_005238		Highlight Sequence Features
VERSION DBSOURCE	NP_005238.1 REFSEC: accession NM 005247.2		Find in this Sequence
KEYWORDS SOURCE ORGANISM	RefSeq. Homo sepiens (human) Homo sepiens		Show in Genome Data Viewer
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia: Futbaria: Eusechontoglines: Primates: Hanlorthini;		Articles about the EGE3 gene
REFERENCE	Catarrhini; Hominidae; Homo. 1 (residues 1 to 239)		Genetically confirmed thanatophoric dysplasia with fibroblast growth facto [Exp Mol Pathol. 2017]
TITLE	Jung W and Park SH. Genetically confirmed thanatophoric dysplasia with fibroblast growth factor recentor & mutation		Allelic loss at chromosome 11q13 alters FGF3 gene expression in a human br [Oncol Rep. 2014]

#### BLAST<sup>®</sup>» blastp sulte

Enter accession	number(s), gi(s	a) or FASTA	sequence(s)		lear	Query subrance 🕥
MGLIWLLLLSLI SGRVNGSLENSA IHELGYNTYASP RVLDHRDHEMVP	EPGWPAAGPGARL YSILEITAVEVGI LYRTVSSTPGARR QLQSGLPRPPGKG	RRDAGGRGG VAIRGLFSG QPSAERLWY SVQPRRRRQB	SVYEHLGGAPRRKLYC SVYEHLGGAPRRKLYC STYLAMNKRGRLYASEH ZVSVNGKGRPRRGFKTR KQSPDNLEPSHVQASRL	ATKYHLQLHP YSAECEFVER RTQKSSLFLP GSQLEASAH		From To
Or, upload file			Procházet	0		
Job Title						
Alian two or	Enter a desc	riptive title for	r your BLAST search 😡			
Align two or Choose Sea	Enter a desc more sequences urch Set	riptive title for	r your BLAST search 🥹			
Align two or Choose Sea Database	Enter a desc more sequences arch Set	proteins (re	r your BLAST search 😡	× 9		
Align two or Choose Sea Database Organism Optional	Enter a desc more sequences irch Set Reference Homo sapi	proteins (reviews)	r your BLAST search 🥹 fseq_protein) 606)	♥ 0	Exclude	•
☐ Align two or Choose Sea Database Organism Dptional	Enter a desc more sequences irch Set Reference Homo sapi Enter organis	proteins (re iens (taxid:90 sm common r	ryour BLAST search @ fseq_protein) 606) name, binomial, or tax id.	V 😥	Exclude	•
Align two or Choose Sea Database Organism Optional Exclude Optional	Enter a desc more sequences irch Set Reference Homo sapi Enter organis Models (	proteins (re proteins (re sm common r (XM/XP) [] (	ryour BLAST search @ fseq_protein) 606) name, binomial, or tax id. Jncultured/environmer	Only 20 top taxa v tal sample sequ	Exclude will be shown.	•

How many human similar reference proteins are in the databases?

	BLAST » blast	p suite					
			19 sequences selected ( ?	Putative conserved de	omains have been detected	d, click on the image below for	detailed resu
	blastn blastp blast	<u>tblastn</u> <u>tblastx</u>	Queru	1 <sup>25</sup>	50 75 100	125 156 175	<sup>2</sup> 1 <sup>0</sup>
Protein BLAST	Enter Query Set	quence	Speci	fic hits	FGE THE	heperin binding site (glycine box) <u>A A A</u>	
FIOTEIII DLAST	Enter accession nu	mber(s) gi(s) or FASTA sequence(s) 🙆	aupor	Taniiles	ror supe	r+amily	
protein > protein	MGLIWLLLSLLEPG	WPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCA'		Distribut	ion of the top 19 Blast Hi	ts on 19 subject sequences	
	IHELGYNTYASRLYR RVLDHRDHEMVRQLQ	LEIIAVEVGIVAIRGLFSGRILAMNRKGRLIASEHI: IVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRR: SGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLG:		1 4	Query I I I 40 80 120	1 1 160 200	
	Or, upload file	Procházet 🥹					
	Job Hue	Enter a descriptive title for your BLAST search					
	Align two or more	e sequences 😡		_			
	Choose Search	Set					
l ≳ NCN Resource © How To ⊙ Protein Protein v Advanced	Sen in to XCBI Hep	Reference proteins (refseq_protein)	<b>∀</b> ⊌				
ComPeet- Compee fibroblast growth factor 3 precursor [Homo sapiens]	e region shown	Homo sapiens (taxid:9606)	Exclude	+			
NCBI Reference Sequence: NP_06528.1 Method Poteins FASTA Graebics Analyze O to D to	e this sequence	Enter organism common name, binomial, or tax id. On	nly 20 top taxa will be shown. (	Θ			
12000         WP_000218         239 es         110000         100000           120000         WP_000218         239 es         110000         100000           120000         WP_000228         Theory of the state of th	ISequence Features de al	Models (XM/XP) Uncultured/environmental	I sample sequences				
Automatical Messary Checkses (Textiles) Verdéhest Restences;         Autobies           Memains Notersis Notestain Verdéhest Restences;         Autobies;           Cestrilis Notersis Notestain Notestain Notestain Restences;         Genetical           RETERE :         Cestrilis Notestain Notestain Notestain Notestain Restences;           RETERE :         Genetical           Titude:         Genetical           Titude:         Genetical           Titude:         Genetical           Genetical :         Science :           Titude:         Genetical :           Genetical :         Genetical :           Titude:         Genetical :           Genetical :         Genetical :	about the FGF3 gene Z Query aly continued thanatophotic dysplasis batat growth tact (Exp Mol Pathot 2001) al deformation in a human te (Oncol Rep. 2014)	Enter an Entrez query to limit search 😡	You Tube Create co	ustom database			

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.

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) / <u>NP\_005238.1</u>

#### Fibroblast growth factor 3

-role in the regulation of embryonic development...



### Exam test:

		Version 1
		Find human protein sequence, FGF3
		What is the accession number and function of this protein?
Maximum 10	) noints	How long is the longest fragment after trypsin digestion?
	points	How many human similar reference proteins are in the databases?
8.5-10p:	1	Does the sequence contain any transmembrane regions?
7-8p:	2	
5-6.5p:	3	Download segence NM_001757.3
-		What does this sequence encode ?
		Is there a complementary sequence to primer R1?
		GCTCTGACGCTCATGATGC
		• In which evon is the complemetary sequence for this primer?
		• Is the primer suitable for PCR with Ta=60°C?
		Design <u>suitable</u> F <u>primer for this</u> experiment.

What does this sequence encode ?

Is there a complementary sequence to primer R1?

In which exon is the complemetary sequence for this primer?

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

Design suitable F primer for this experiment.

>R1 GCTCTGACGCTCATGATGC

What does this sequence encode ?

Is there a complementary sequence to primer R1?

In which exon is the complemetary sequence for this primer?

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

Design suitable F primer for this experiment.

>R1 GCTCTGACGCTCATGATGC

Sequence Analysis

#### Is there a complementary sequence to primer R1? >R1 GCTCTGACGCTCATGATGC Multalin: 280 300 310 330 350370 390 261 270 290320340360 380GCTCT-GACGC---TCATGATGC .....aCaTGAcGC... Consensus Reverse Complement results Sequence Manipulation Suite: Reverse Complement >Untitled reverse complement Format Conversion Reverse Complement converts a DNA sequence into its reverse, complement, or r -Combine FASTA it contains an ORF on the reverse strand. -EMBL to FASTA -EMBL Feature Extracto GCATCATGAGCGTCAGAGC -EMBL Trans Extractor -Filter DNA Paste the raw sequence or one or more FASTA sequences into the text area below -Filter Protein -GenBank to FASTA -GenBank Feature Extractor -GenBank Trans Extractor GCTCTGACGCTCATGATGC Consensus 521 530 540 570 590 610 620 630 640 -Genbank Trans Extracto -One to Three -Range Extractor DNA -Range Extractor Protein -Reverse Complement NH\_001757.3 CGATGAAAACAAATTTCTTTGGTACCCGAGATGTGTGCACAGAAT actccctctaataaaaccccaagggagagtggtgaacgtatcta<mark>gcatcatgagcgtcagagcccttaaaagcccc</mark>tgagagcccaga GCATCATGAGCGTCAGAGC -Split Codons Submit Clear Reset -Split FASTA -Three to One Consensus -Window Extractor DNA reverse-complement V 670 710 720 730 740 750 651 660 700 760 770 -Window Extractor Protein

1---

\_\_\_\_ 

\_\_\_\_

---\*-

---\*-

650

780

----

#### Is there a complementary sequence to primer R1?

>R1 GCTCTGACGCTCATGATGC

 $\rightarrow$ yes.

### NCBI: Graphic

Graphics -

Send to: -

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

NCBI Reference Sequence: NM 001757.3 GenBank FASTA

					୍ କ୍	II:		Search Results				*	b 🗶 bls • 🛬 🗭 T	racks 🧬
	1	180 150	280 250	380 35	8 400	450 500	550 60	Features Component	5 🖋 Sequence 🛛 T	racks			1,200	1,250
	Genes							Label ^		From	To	Strand		
	STS Markers	NP_0017481 carb_red_PTCR-like NBDP binding acetylatic phosphorylatic	phosphorylation		substrate bindin Glutathione b	g s active	exon CBR1,1984	GETCTGAGGCTGATGATGC		511	629	Negative	> 	>
													>	
0  100	150 200	250 300	350  400	450 50	00  550 4	600 <b>650</b>	700 750	800 850 900	).  950  1 K	1,050  1,100  1,150	1,200  1,2	50 <u>1,321</u> - 1 o	→ →	-
0  100 001757.3 •   Fin	150 200	250 <u>300</u>		1450  50	0 550 4	600 650	700.750	800 850 900	)  950  1 K	- lr'oeo - lr'ioo - lr'ieo	1,200  1,2	50 <u>1,321</u> - 1 o cks @ ? • <sup>se</sup>	of 1	-
0  100 0 <b>01757.3 ▼</b>   Fin	1:	250  300 ~ (\$\$\$) Q [688	350  400	450  5( 2, 100 111  619	0 1550	600 <b>650</b>	700	800  850  900  630	)	1,050  1,100  1,150 ×	1,200  1,2 Tools ▼ 😤 🕸 Trad	50 <u>1,321</u> - 1 o dks @ ? • se 656	of 1	2
01757.3 +   Find	G T G G C A C C		950	450	20. 550 4 A T C A T A G T	600 <b>650</b> T G A G A C T C	C G T C G C A G	900  950  90  530 A G A G C C C T C T C G G G	)  950  1K T T A A A A A A T T T T	1,050  1,100  1,150  640 G C T G C A C G A C G T	[1,200 [1,2 Tools • 🛣 🖨 Trad G C C C C G G G	50 1,32 cks @ ? • se 650 A G A T C T @ 1,11 X	polyA_signal_se 50  1,200	equence 1,250 cs shown:

What does this sequence encode ?

Is there a complementary sequence to primer R1?

In which exon is the complemetary sequence for this primer?

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

Design suitable F primer for this experiment.

>R1 GCTCTGACGCTCATGATGC

### In which exon is the complemetary sequence for this primer?

### NCBI: Graphic



>R1 GCTCTGACGCTCATGATGC

 $\rightarrow$  third exon

What does this sequence encode ?

Is there a complementary sequence to primer R1?

In which exon is the complemetary sequence for this primer?

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

(Design suitable F primer for this experiment.) 2025: We did not cover this one >R1 GCTCTGACGCTCATGATGC

#### Solution:

This sequence encodes carbonyl reductase 1 (CBR1).

Yes, there is complementary sequence to primer R1.

It is within 3rd exon.

At Ta=60°C the primer should work.

An example of suitable F primer: F:CAAGGTTGCTGATCCCACAC



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

NCBI Reference Sequence: NM\_001757.3

FASTA Graphics

														650
NH_001757.3 R Consensus	ĊGATG	AAAACAAAT	ттстттббте	ICCCGAGATGT	GTGCACAGA	ATTACTCCCTC	TAATAAAAC	CCCAAGGGAGF	IGTGGTGAACO	TATCTAGCAT GCAT GCAT	CATGAGCGTO CATGAGCGTO CATGAGCGTO	CAGAGCCCTTI CAGAGC CAGAGC	NAAAGCTGCAG	1 CCCAGA
NH_001757.3	651   6CT6C	660 AGCAGAAGT	670 TCCGCAGTGA	680 GACCATCACT	690 GAGGAGGAG	700 CTGGTGGGGGC1	710 CATGAACAA	720 GTTTGTGGAGG	730 ATACARAGAR	740 IGGGAGTGCAC	750 CAGAAGGAGG	760 GCTGGCCCA	770 SCRGCGCRTAC	780 1 6666T6

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 REGIO	N SIZE: 1	321						

..........

### Exam test II:



What does this sequence encode?

(Are there any significant single nucleotide polymorphisms (GMAF), in which exons?) **2025:** We did not cover this one

Would be the restriction endonucleases Ndel and Kpnl 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?

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:	Kpnl ggtac c	289
	Ndel ca tatg	none
	Nuel callary	none

ightarrow KpnI is not suitable

What does this sequence encode?

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

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

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

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

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

>protein-FGF3

MGLIWLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVER IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRQKQSPDNLEPSHVQASRLGSQLEASAH\*

### Download sequence NM 005247.2

#### Solution:

This gene encodes Fibroblast Growth Factor 3

GMAF are in the 1st and 3rd exon.

Restriction endonuclease KpnI is not suitable for cloning.

The FASTA of the protein is:

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

#### >protein-FGF3 MGLIWLLLSLLEPGWPAAGPGARLRRDAGGRGGVYEHLGGAPRRRKLYCATKYHLQLHP SGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAMNKRGRLYASEHYSAECEFVEF IHELGYNTYASRLYRTVSSTPGARRQPSAERLWYVSVNGKGRPRRGFKTRRTQKSSLFLP RVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH\*

FGF3 MOUSE

#### Homo sapiens fibroblast growth factor 3 (FGF3), mRNA # Aligned\_sequences: 2 # 1: EMBOSS\_001 # 2: FGF3\_MOUSE # Matrix: EBLOSUM6 # Gap\_penalty: 10.0 NCBI Reference Sequence: NM 005247.2 # Extend\_penalty: 0.5 # Length: 249 GMAF >= 0.01 197/249 (79.1%) 14 # Identity: 210/249 (84.3%) # Similarity: G/T | rs143969364 6/T | rs41538178 R/G rs35420992 14/249 ( 5.6%) ∉ Gaps: Score: 1021.5 Genes - Exon 14 EMBOSS\_001 1 MGLIWLLLSLEPGWPAAGPGARLERDAGGRGGVYEHLGGAPEREKLY FGF3\_MOUSE 1 MGLIWLLLSLLEPSWFTTGFGTRLRRDAGGRGGVYEHLGGAPRRRKLY Ndel caltatg none EMBOSS 001 51 ATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGIVAIRGLFSGRYLAM 51 ATKYHLQLHPSGRVNGSLENSAYSILEITAVEVGVVAIKGLFSGRYLAMN FGF3 MOUSE 101 KRGRLYASEHYSAECEFVERIHELGYNTYASRLYRTVSSTFGARRQPSAE EMBOSS 001 1.001 9190900 ..... 101 KRGRLYASDHYNAECEFVERIHELGYNTYASRLYRTGSSGFGAQRQPGAQ FGF3\_MOUSE Kpnl ggtac|c 289 EMBOSS\_001 151 RLWYVSVNGKGRPRRGFKTRRTQKSSLFLPRVLDHRDHEMVRQLQSGLPR 151 REWYVSVNGKGRPRRGFKTRRTQKSSLFLPRVLGHKDHEMVRLLQSSQPR FGF3 MOUSE EMBOSS 001 201 PPGKGVOPRRRQ-KOSPDNLEPSHVOASRLGSQLEASAH-

### Exam test:

2 parts, each five points

#### Maximum 10 points

9-10p: 1

7-8p: 2

5-6p: 3



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

Identify the protein.

Does it contain any typical domains?

How many cysteins contain the peptide?

What is the molecular weight of this peptide?

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

>S2

CGPSTSGTTSGPGPARPARARPRRPREETLTPEEEEKRRVRRERNKLAAAKCRNRRRELT DRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIPYEEGPGPGPLAEVRD LPGSAPAKEDGFSWLLPPPPPPPPPPPPPTQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSC

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

Identify the protein.

Does it contain any typical domains?

How many cysteins contain the peptide?

What is the molecular weight of this peptide?

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

Solution:

Peptide is probably human FosB homologue.Yes, it contains leucine zipper bZIP.Contains four cysteinsMw=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 Stats results Results for 180 residue sequence "Untitled" starting "CGPSTSGTTS"		
Pattern:	Times found:	Percentage:
A	15	8.33
В	0	0.00
С	4	2.22
R	6	2.22

#### Protein Molecular Weight results Results for 180 residue sequence "S2" starting "CGPSTSGTTS"

19.91 kDa



### protein fosB isoform 1 [Homo sapiens]

NCBI Reference Sequence:NP\_006723.2Identical ProteinsFASTAGraphics



cted feature details							
RLZ domair	RLZ domain						
s 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

### Version from 2023\_new to Moodle

#### Version I 19.5.2023

#### A: Download the sequence NM\_000946.3

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

#### B: Work with following peptide sequence:

LALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACLRGIGRWT NPQYRQFITILEATHRNQSSENKRQLAWYNFDFRSWPVDFHWEEPSSRKESRGGPSRRGV ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQG QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP LEACYSVLGWNHPGFAGSIGVPFPQNEANAMDVVVQFALHRLGFQPQDIIIYAWSIGGFT

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

NIH National Library of Medicine National Center for Biotechnology Information	<b></b> jostovap	
Nucleotide  Nucleotide Advanced	Search	٩le
GenBank 🗸	Send to: - Change region shown	•
Homo sapiens DNA primase subunit 1 (PRIM1), mRNA NCBI Reference Sequence: NM_000946.3	Customize view	•
<u>Go to:</u> ♥	Analyze this sequence	
LOCUS NM_000946 1423 bp mRNA linear PRI 03-APR-2024	Pick Primers	
DEFINITION Homo sapiens <mark>UNA primase subunit 1</mark> (PRIM1), mRNA. ACCESSION NM_000946	Highlight Sequence Features	
VERSION NM_000946.3	Find in this Sequence	
SOURCE Homo sapiens (human) ORGANISM <u>Homo sapiens</u> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Show in Genome Data Viewer	
Mammalia: Eutheria: Euarchontoglires: Primates: Haplorrhini:		

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

CDS:	CDS	<pre>/inference="alignment:Spl 261288 /gene="PRIM1" /gene_synonym="p49; PDIL" /EC number="2.7.7.102"</pre>	ign:2.1.0"	1) 1288-26+1= 1263nt (bp)
Transl	ated protein			
		primase polypeptide 1, 49кµа; UNA primase 49 кµа subu 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: <u>5557</u> " /db_xref="HGNC:HGNC:9369"	nıt;	<pre>261288 //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 (DNA) subunit 1"</pre>
DN	A primase sr	nall subunit [Homo sa	piens]	/codon_start=1 /product="DNA primase small subunit" /product=1
NCBI	Reference Sequence	e: NP_000937.1		/db_xref="Const: <u>CCDS44926.1</u> " /db_xref="GeneID: 5557 "
Identio	cal Proteins FASTA	Graphics		/db_vref="HGNC: <u>HGNC:9369</u> " /db_vref="MIM: <u>176635</u> " /translation="METFDPTELPELLKLYYRRLFPYSQYYRWLNYGGVIKNYFQHRE
Go to:				FSFTLKDDIYIRYQSFNNQSDLEKEMQKMNPYKIDIGAVYSHRPNQHNTVKLGAFQAQ EKELVFDIDMTDYDDVRRCCSSADICPKCWTLMTMAIRIIDRALKEDFGFKHRLWVYS GRRGVHCWVCDESVRKLSSAVRSGIVEYLSLVKGGQDVKKKVHLSEKIHPFIRKSINI
LOCUS	NP_000937	<mark>_420_aa_</mark>	linear PRI 03-APR-202	24 IKKYFEEYALVNQDILENKESWDKILALVPETIHDELQQSFQKSHNSLQRWEHLKKVA SRYQNNIKNDKYGPWLEWEIMLQYCFPRLDINVSKGINHLLKSPFSVHPKTGRISVPI
DEFIN	IITION DNA primase	small subunit [Homo sapiens].		DLQKVDQFDFFTVFTSFTCKELDAISTNEEEKEENEAESDVKHKTKDYKKTSLAPYV KVFEHFLENLDKSRKGELLKKSDLQKDF"
	SION NP_000937			Details 🕑 Display: FASTA GenBank Hein 🗶
	IRCE REFSEO: acc	ession NM 000946.3		
KEYWO	RDS RefSea: MAN	E Select.		

CDS

#### 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 Other possibility: Homo sapiens DNA primase subunit 1 (PRIM1), mRNA. DEFINITION ACCESSION NM 000946 Graphics -Send to: -Homo sapiens DNA primase subunit 1 (PRIM1), mRNA NCBI Reference Sequence: NM 000946.3 GenBank FASTA Link To This View | Feedback |50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 11 |1,050 |1,100 |1,150 |1,200 |1,250 |1,300 |1,350 1,42 CDS 5 S NM 000946.3 - Find: ~ (\$\$) Q []= 🗨 🚛 📑 😤 📡 Tools 🗸 🄹 Tracks 🛛 📩 Download 🗸 🎘 🤋 150 200 250 300 350 400 450 500 550 600 650 700 750 850 900 950 |1 K 1,050 1,100 1,150 1,200 1,250 1,300 1,350 1,423 100 800 Genes 14



#### Other option for translated protein:

Format Conversion

-Combine FASTA -EMBL to FASTA -EMBL Feature Extractor

-Pairwise Align DNA -Pairwise Align Protein -PCR Primer Stats -PCR Products





#### Translate

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

-EMBL Trans Extractor -Filter DNA -Filter Protein -GenBank to FASTA -GenBank Trans Extractor -GenBank Trans Extractor -One to Three -Range Extractor DNA	NM_000946.3:26-1288 Homo sap (PRIM1), mRNA ATGGAGACGTTTGACCCCACCGAGCTGCCC TTCCCTACT CTCAGTACTATCGCTGGCTCAACTACGGTG	iens DNA primase subunit 1 GAGCTGCTTAAACTTTATTACCGGAGGCTCT	
-Range Extractor Protein -Reverse Complement -Split Codons -Split FASTA	Submit Clear Reset	Sequence Manipulation Suite – Pracovní – Microsoft Edge –	
-Three to One -Window Extractor DNA -Window Extractor Protein Sequence Analysis	<ul> <li>Translate in reading frame 1</li> <li>Use the standard (1)</li> </ul>	about:blank Translate results	
Sequence Analysis -Codon Piot -Codon Usage -CpG Islands -DNA Molecular Weight -DNA Stats -Fuzzy Search DNA -Fuzzy Search DNA -Fuzzy Search Protein -Ident and Sim -Mutil Rev Trans -Mutate for Digest -ORF Finder -Pairwise Align Codons -Pairwise Align Codons -Pairwise Align DNA -Pairwise Align Protein -PCR Primer Stats	*This page requires JavaScript. Se *You can mirror this page or use it Sun 14 Jun 00:37:01 2020 Valid XHTML 1.0; Valid CSS	<pre>&gt;rf 1 NM_000946.3:26-1288 Homo sapiens DNA primase subunit 1 (PRIM1), mRNA METFDPTELPELLKLYYRRLFPYSQYYRWLNYGGVIKNYFQHREFSFTLKDDIYIRYQSF NNQSDLEKEMQKMNPYKIDIGAVYSHRPNQHNTVKLGAFQAQEKELVFDIDMTDYDDVRR CCSSADICPKCWTLMTMAIRIIDRALKEDFGFKHRLWVYSGRRGVHCWVCDESVRKLSSA VRSGIVEYLSLVKGQQDVKKKVHLSEKIHPFIRKSINIIKKYFEEYALVNQDILENKESW DKILALVPETIHDELQQSFQKSHNSLQRWEHLKKVASRYQNNIKNDKYGPWLEWEIMLQY CFPRLDINVSKGINHLLKSPFSVHPKTGRISVPIDLQKVDQFDPFTVPTISFICRELDAI STNEEEKEENEAESDVKHRTRDYKKTSLAPYVKVFEHFLENLDKSRKGELLKKSDLQKDF *</pre>	

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

Range Extractor Protein results

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

Protein Molecular Weight results

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

Μ

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 ATGGAGACGTTTGACCCCACC GAGCTGCCCGAGCTGCCCGAGCTGCTTAAACTTTATTACCGGAGGCTCTTTCCCTACT CTCAGTACTATCGCTGGCTCAACTACGGTGGAGTGATAAAGAATTACTTTCAACACCGTGAATTTTCATT CACATTGAAAGATGATATTTACATTCGCTACCAATCCTTCAACAACCAGAGTGATCTGGAAAAGGAGATG CAGAAAATGAATCCATACAAGATTGATATAGGCGCAGTATATTCTCACAGACCCCAATCAACACAATACAG TGAAGCTGGGAGCTTTCCAGGCTCAGGAAAAAGAACTGGTATTTGACATTGACATGACAGACTATGACGA TGTGAGGAGATGTTGTAGTTCTGCAGACATATGTCCTAAGTGCTGGACCCTCATGACAATGGCCATACGC ATCATTGACAGAGCATTGAAGGAGGACTTTGGATTTAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAG GTGTTCATTGTTGGGTCTGTGATGAATCAGTTAGAAAACTGTCTTCTGCAGTACGTTCTGGGATAGTTGA TTTATCAGAAAATCTATAAACATAATAAAAAAAATACTTTGAAGAATATGCCTTGGTTAATCAAGATATTC TCGAAAATAAAGAAAGCTGGGATAAGATTTTAGCCCTTGTTCCTGAAACAATTCATGATGAACTTCAACA AAGCTTCCAAAAGTCTCACAATTCACTTCAGCGTTGGGAGCACTTGAAGAAAGTAGCCAGCAGATATCAG AATAACATCAAAAATGACAAATATGGACCCTGGCTGGAGTGGGAGATTATGCTCCAGTACTGTTTTCCAC GGCTGGATATCAATGTCAGCAAAGGAATCAATCATCTACTGAAGAGCCCTTTTAGTGTTCATCCTAAAAC AGGTCGCATATCTGTGCCTATTGATTTGCAGAAAGTGGACCAGTTTGATCCATTTACTGTTCCGACCATA AGCTTCATCTGCCGTGAATTGGATGCCATTTCCACTAATGAAGAGGAAAAAGAGGAGAATGAAGCTGAAT CTGATGTCAAACATAGAACCAGAGATTATAAGAAGACCAGTCTAGCACCTTATGTGAAAGTTTTTGAACA TTTTCTTGAAAATCTGGATAAATCCCGAAAAGGAGAACTTCTTAAGAAGAGTGATTTACAAAAAGATTTC

Oligo Calc: Oligonucleotide Properties	Calculator
Enter Oligonucleotide Sequence Below OD calculations are for single-stranded DNA or I	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)     Sel       ✓     ✓     Ss	lect molecule
50         nM         Primer           50         mM         Salt (Na <sup>+</sup> )	0 nanometers
Calculate Swap Strands BLAST mfol	d
Physical Constants	Melting Temperature (T <sub>M</sub> ) 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 <sup>5</sup> / <sub>2</sub> and contains 28.8 micrograms.	3 52.18 °C' (Nearest Neighbor)
Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.	
RInK 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), TCAGAAATCTTTTTGTAAATCACTCTTCTTAAGAAGTTCTCCCTTTTCGGGATTTATCCAG ATTTTCAAGAAAATGTTCAAAAACTTTCACATAAGGTGCTAGACTGGTCTTCTTATAATC TCTGGTTCTATGTTTGACATCAGATTCAGCTTCATTCTCCTCTTTTTCCTCTTCATTAGT GGAAATGGCATCCAATTCACGGCAGATGAAGCTTATGGTCGGAACAGTAAATGGATCAAA CTGGTCCACTTTCTGCAAATCAATAGGCACAGATATGCGACCTGTTTTAGGATGAACACT AAAAGGGCTCTTCAGTAGATGATTGATTCCTTTGCTGACATTGATATCCAGCCGTGGAAA ACAGTACTGGAGCATAATCTCCCCACTCCAGCCAGGGTCCATATTTGTCATTTTGATGTT ATTCTGATATCTGCTGGCTACTTTCTTCAAGTGCTCCCAACGCTGAAGTGAATTGTGAGA CTTTTGGAAGCTTTGTTGAAGTTCATCATGAATTGTTTCAGGAACAAGGGCTAAAATCTT ATCCCAGCTTTCTTTATTTCCGAGAATATCTTGATTAACCAAGGCATATTCTTCAAAGTA TTTTTTTTATTATGTTTATAGATTTTCTGATAAAAGGGTGAATTTTTTCACTTAGGTGAAC TTTTTTTTAACGTCTTGACCACCCTTTACAAGGCTCAAATACTCAACTATCCCCAGAACG TACTGCAGAAGACAGTTTTCTAACTGATTCATCACAGACCCCAACAATGAACACCTCT TCCAGAATATACCCAGAGACGATGCTTAAATCCAAAGTCCTCCTTCAATGCTCTGTCAAT GATGCGTATGGCCATTGTCATGAGGGTCCAGCACTTAGGACATATGTCTGCAGAACTACA ACATCTCCTCACATCGTCATAGTCTGTCATGTCAAATACCAGTTCTTTTTCCTG AGCCTGGAAAGCTCCCAGCTTCACTGTATTGTGTTGATTGGGTCTGTGAGAATATACTGC GCCTATATCAATCTTGTATGGATTCATTTTCTGCATCTCCTTTTCCAGATCACTCTGGTT

#### Oligo Calc: Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below OD calculations are for single-stranded DNA or F	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)       Self         50 nM Primer           50 mM Salt (Na <sup>+</sup> )       1 Measured Absorbance at 260         Calculate         Swap Strands         BLAST	ect molecule DNA V 0 nanometers
Physical Constants	Melting Temperature (T <sub>M</sub> ) Calculations
Length:       27       Molecular Weight:       8198.4       4       GC content:       26       %         1 ml of a sol'n with an Absorbance of       1       at 260 nm       is       3.55       microMolar       5       and contains       29.1       micrograms.	1         50.6         °C (Basic)           2         59.2         °C (Salt Adjusted)           3         52.83         °C (Nearest Neighbor)
Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.	
RInK 33.404 cal/(°K*mol)	deltaH 203.6 Kcal/mol
deltaG 30.1 Kcal/mol	deltaS 543 cal/(°K*mol)
Deprecated Hairpin/Self dimerization calculations         5 • (Minimum base pairs required for single primer self-dimerization)         4 • (Minimum base pairs required for a hairpin)	Check Self-Complementarity

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

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

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

		10000
	>F	
110	ATG GAG ACG TTT GAC CCC AC	
	>R nc	
No.	AAG AGT GAT TTA CAA AAA GAT TTC TGA	
1		
2	VMM 000046 2026 1988 Home conjens DMA primose suburit 1 (DDTM1) mDNA	
	ATCCACE ACCESSED AND SAPIENS DWA PRIMASE SUBURIL 1 (PRIMI), MRWA	
ŝ		-
8		
L		

	,	10	20	20		51	E41	20	201	50	100	110	120	130
F	ī			ятселено	GTT TERCOCO	IC					+		·····	
R_rc NH_010946.3126-1288 NH_010946.3 Consensus	сттяс	CETEGOGRE	TTECREBET	ATGGAGAC RATGGAGAC Lategagao	GTT NERCECCO GTT NERCECCO gEX.Egacecco	CONNECTED CONNECTED	CEGNECTICT	I BARAC TT TA T I BARACTT TA T	TRCCGGABGC TRCCGGABGC	ICTITICCTN ICTITICCTN	CTCTCRGTRC	TRTCGCTGGC TRTCGCTGGC	TCARCTREGG TCARCTREGG	TGGAGTG TGGAGTG
	131	140	150	150	179	190	190	200	210	220	230	240	250	260
F R_rc NH_010346.3126-1288 NH_010346.3 Consensus	ATAAA	GRATTRCTT GARTTRCTT	TCRACACCES	GARETTER GARETTER	TTCHCHTTGH TTCHCHTTGH	ANCAT GATA TA T Ancat gata ta t	T THEM IT LIGE T THEM IT LIGE	INCORPLECT	TERREARCEA	GAGTGATCTG	GNANASGAGA	TGCRGRARATI TGCRGRARATI	GARTCORTRO	ANGATTG
	261 	270	290	290	300	319	329	330	30)	350	350	370	380	390
F R_rc NH_010945.3125-1288 NH_010945.3 Concensus	ATATA Atata	GGCGCRGTR GGCGCRGTR	TATICICAC	GACCOARTO	ARICRICIA TAIC Aricrical Taic	IST SHIGCT S	GEASCITICO GEASCITICO	REGET CREGA	ARAGARCTS	GTATT TGACA	TIGACATGACI	RGACTRTGAC	GATETEASGA	GRIGIIG
	391	400	419	429	439	48	450	450	470	480	490	500	510	520
F R_rc NH_010945.3125-1298 NH_010946.3	18611 18611	CTECHERCH	TREFETCE	ET EC TGERC	CETER TRACK	A TIGGICCH THE	GERTCHTTGR	CHIGHIGLATTE	ANGEAGERCT Angeagerct	TIGGATTTAR	GERICETETE	IGGGTRTRTT IGGGTRTRTT	TGGARGGAG TGGARGGAG	AGGTGTT AGGTGTT
Consensus	521	520	549	559	550	529	589	539	6M	F10	620	630	640	650
F R_rc NH_010345.3125-1288 NH_010345.3	CATTS CATTS	1 1666 1C 16	TGATGAATC	GT THGANA		CING THE GATTE	I SCEATRET D	CHICTRITTICS	GCCTTETARA	GEGIGETCRE	GRCGT TARAPA	NGARAGTICA REARAGTICA	CTRASTGAR	ARARTTC
Consensus		650	670	680	670	704	716	2780	720	7.00	250	750	770	280
F R_rc NH_000345,3125-1288 NH_000345,3 Consensus	HCCCT HCCCT	TTINICAGA	AAATCTATA	ACREATING	ARABAR TRICT T	T GARAGAN THE D T GARAGAN THE D	GDC TT BG TT N GDC TT BG TT N	RTCHREATH T	TCTCGRARAT	ANAGANAGC 1 ANAGANAGC 1	gegrennen ti Gegrennen ti	TTRECCT	GTTCCTGRAR STTCCTGRAR	CARTTER
	781	790	809	819	829	839	849	850	850	870	880	890	900	910
F R_rc NH_010945.3:25-1288 NH_010946.3 Consensus	TGATG	ARCTICARC	REALTICS	RARGICICS	CRATTCRCT TO CRATTCRCT TO	CRECETTEES	ASCACTTERN ASCACTTERN	SING TRECC	RECEGRETATION	AGARTARCAT AGARTARCAT	CRARARTGAC	ARATATGGACI Aratatggaci	CTEGETEGR	STG5GAG STG5GAG
	911 	320	930	340	350	350	\$79	390	390	1000	1010	1020	1030	1040
F R_rc NH_010946.3125-1298 NH_010946.3 Concensus	ATTAT ATTAT	GETECRGTR Getecratr		CSGCTSGAT CSGCTSGAT	A TICKA DE DOK A TICKA DE DOK	GCHINAISGAN T GCHINAISGAN T	CHRITCHT CT IN Chritcht Ct In	CTGRABABOC CTGRABABOC	CTTTTRGTGT CTTTTRGTGT	TCATCCT NAM	ACASSICSCA ACASSICSCA	TRICTOTOCC	TATIGATITS	CAGAAAG
	1941	1050	1960	1979	1989	1090	1100	1119	1120	1170	1140	1150	1160	1170
F R_rc NH_000345.3125-1288 NH_000345.3 Consensus	tggac Tggac	CRGTTTGAT	CONTINUE	TTECHNER	TRANSCITICATO Fransiciticato	CINGLOCIEST SAIR CINGLOCIEST SAIR	T TGEN TGECH	TTCCHCIRR	terrerer Terrerer	ANAGAGGAGA Anagaggaga	ATGRASCTGA Atgrasctga	ATC TGATGTC ATC TGATGTC	ABACATASAN ABACATASAN	ocrease Ocrease
	1171	1180	1190	1200	1219	1229	1220	1249	1250	1250	1270	1280	1290	1300
F R_rc NH_010345.3125-1288 NH_010346.3 Concessor	TTATA	AGAAGACCA	GTCTRGCACO	TTATETEAA TTATETEAA	AGT TT TTGAR AGT TT TTGAR		ANART ET GGR Anart et ggr	THERT DODGE	ANASGAGNAC		GRETGRETTER GRETGRETTER GRETGRETTER		TCTGA TCTGA TCTGARGACA	RECTCC
	1301	1310	1329	1330	1340	1359	1350	1379	1380	1390	1400	1410	1429423	
F R_rc N1_010345,3125-1283 N1_010345,3125-2833 ST_constraints	тсяня	ссяттятья	ATATCTIC	ECTTERREE	RCREATCRAR	INCITICANGE	GECHI TTANT	NH TH TODOS	GRECTRETR	TETETCTINE	ACCTCRARGT	ABATTITECT	TGAGARATAR	

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

#### Get the 3rd exon:

	Connect with NLM	National Library of Medicine	Web Policies	Help	/inference="alignment:Splign:2.1.0"
		8600 Rockville Pike	FOIA	Accessi	t
exon V Feature 3 of 1	3 NM_000946 : 1 segment				Details 🕑 Display: <u>FASTA</u> <u>GenBank</u> <u>Help</u> 💌

#### **Restriction summary:**



#### 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 proceeding 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 sap (PRIM1), mRNA CCCAATCAACACAATAAGTGAAGCTGGAA TGACATTG ACATGACAGACTATGACGATGTGAGGAGAT	iens DNA primase <u>subunit</u> 1 GCTTTCCAGGCTCAGGAAAAAGAACTGGT GTTGTAG	ATT	
Submit Clear Baset	Sequence Manipulation Suite – Pracovn	ıí – Microsoft Edge —	×
Submit Clear Reset	i) about:blank		Ð
Treat sequences as linear	Nael gcc ggc	none	
*This page requires JavaScript_Se	Narl gg cgcc	none	
*You can mirror this page or use it	Ncol c catgg	none	
	Ndel ca tatg	none	
Sun 14 Jun 00:37:00 2020	Ndell  gatc	none	
Valid XHTML 1.0; Valid CSS	NgoMIV g ccggc	none	
	Nhel g ctagc	none	
	Nlalli catg	76	
	Notl gc ggccgc	none	
	Nrul tealcas	nono	

#### **B: Work with following peptide sequence:**

LALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACLRGIGRWT NPQYRQFITILEATHRNQSSENKRQLANYNFDFRSWPVDFHWEEPSSRKESRGGPSRRGV ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQG QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP LEACYSVLGWNHPGFAGSTGVPFPQNEANAMDVVVQFAIHRLGFQPQDIIIYAWSIGGFT

- 1) To which human protein this peptide probably belongs?
- 2) Does this peptide contain any transmembrane helices?
- 3) What is molecular weigth of this peptide?
- 4) How does a given peptide sequence differ from an identified (human) sequence (within the same segment)?
   2025: We did not cover this one
- 5) Write down the identified mutation. **2025:** We did not cover this one

#### **B: Work with following peptide sequence:**

LALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACLRGIGRWT NPQYRQFITILEATHRNQSSENKRQLANYNFDFRSWPVDFHWEEPSSRKESRGGPSRRGV ALLRPEPLHRGTADTLLNRVKKLPCOITSYLVAHTLGRRMLYPGSVYLLOKALMPVLLOG QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP LEACYSVLGWNHPGFAGSTGVPFPONEANAMDVVVOFAIHRLGFOPODIIIYAWSIGGFT

- To which human protein this peptide probably belongs?
- Does this peptide contain any transmembrane helices? 2)
- What is molecular weigth of this peptide? 3)
- How does a given peptide sequence differ from an identified (human) sequence (within the same 4) segment)? **2025:** We did not cover this one
- Write down the identified mutation. 5)

**2025:** We did not cover this one

→BLASTp

→TMHMM, TopCons...

 $\rightarrow$ SMS/Protein Molecular weight

# To which human protein this peptide probably belongs?

		BERGI # DIG	Stp Suite // Tesuits for RID-STOXNRMD015		Home Recent	tesuits saved strategies freip
	BLASTp	< Edit Search	Save Search Search Summary 🗸	How to read this report?	BLAST Help Videos	Back to Traditional Results Page
		Job Title	Protein Sequence	Filter Results		
	BLAST <sup>®</sup> » blastp suite	RID	316XNRMD013 Search expires on 05-01 15:20 pm Download All 🗙			
blastn bla	stop blastx tblastn tblastx Standard Protein BL	Program	BLASTP 😧 <u>Citation</u> 🗸	Organism only top 20 v	will appear	exclude
	BLASTP programs search protein databases us	Database	refseq_protein <u>See details</u> ✓	Type common nam	ne, binomial, taxid or group	name
Enter Query S	equence	Query ID	lcl Query_7536389			
Enter accession nu	Imber(s), gi(s), or FASTA sequence(s) 😮 Clear Query subrange 🖓	Description	unnamed protein product	Percent Identity	E value	Query Coverage
LALASVFWSISYYSS RGIGRWT	PFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACL	Molecule type	amino acid	to	to	to
NPQYRQFITILEATHF RGGPSRRGV		Query Length	300			
Or, upload file	Zvolit soubor Nevybrán žádný soubor 🕜	Other reports	Distance tree of results Multiple alignment MSA viewer @			Filter Reset
Job Title						
_	Enter a descriptive title for your BLAST search 💡	Descriptions	Graphic Summary Alignments Taxonomy			
Align two or mor	e sequences 😧	Sequences	producing significant alignments	Dowr	nload 🎽 Select colur	nns × Show 100 × Q
Choose Searc	h Set	Sequences	producing significant augminents	50		
Databasos	Try experimental clustered nr database	🗹 select all	100 sequences selected	GenPept Graphics	Distance tree of results	Multiple alignment MSA Viewer
Databases	Standard databases (nr etc.): Experimental databases For more info see What is clustered nr?		Description	Scientific Name	Max Total Query	Per. Acc.
Compare	Select to compare standard and experimental database 2		Decempton		Score Score Cover val	ue Ident Len Accession
			Ŧ		<b>* * *</b>	· · ·
Standard		phosphatid	ylserine lipase ABHD16A isoform X1 (Gorilla gorilla gorilla)	Gorilla gorilla gorilla	625 625 100% 0	.0 99.67% 601 <u>XP 004043725.4</u>
Standard Database	Reference proteins (refseq_protein)	<ul> <li>phosphatid</li> <li>phosphatid</li> </ul>	ylserine lipase ABHD16A isoform X1 [Gorilla gorilla gorilla] ylserine lipase ABHD16A isoform X3 [Sapajus apella]	<u>Gorilla gorilla gorilla</u> <u>Sapajus apella</u>	625 625 100% 0 624 624 100% 0	•         •         •           .0         99.67%         601         XP 004043725.4           .0         99.67%         457         XP 032124883.1
Standard Database Organism		<ul> <li>phosphatidt</li> <li>phosphatidt</li> <li>phosphatidt</li> <li>phosphatidt</li> </ul>	ylserine lipase ABHD16A isoform X1 [Gorilla gorilla] ylserine lipase ABHD16A isoform X3 [Sapajus apella] ylserine lipase ABHD16A isoform X3 [Callithrix jacchus]	<u>Gorilla gorilla gorilla</u> <u>Sapajus apella</u> <u>Callithrix jacchus</u>	625         625         100%         0           624         624         100%         0           624         624         100%         0	VP         VP         004043725.4           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1
Standard Database Organism Optional		<ul> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> </ul>	ylserine lipase ABHD16A isoform X1 [Gorilla gorilla] ylserine lipase ABHD16A isoform X3 [Sapajus apella] ylserine lipase ABHD16A isoform X3 [Callithrix jacchus] ylserine lipase ABHD16A isoform X1 [Callithrix jacchus]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0	99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1
Standard Database Organism Optional Exclude		<ul> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> <li>phosphatid</li> </ul>	ylserine lipase ABHD16A isoform X1 [Gorilla gorilla gorilla] ylserine lipase ABHD16A isoform X3 [Sapajus apella] ylserine lipase ABHD16A isoform X3 [Callithrix jacchus] ylserine lipase ABHD16A isoform X1 [Callithrix jacchus] ylserine lipase ABHD16A isoform X1 [Aotus nancymaae]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus Aotus nancymaae	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1
Standard Database Organism Optional Exclude Optional		<ul> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>prosphalid</li> <li>protein ABI</li> </ul>	ylserine lipase ABHD16A isoform X1 [Gorilla gorilla] ylserine lipase ABHD16A isoform X3 [Sapajus apella] ylserine lipase ABHD16A isoform X3 [Callithrix jacchus] ylserine lipase ABHD16A isoform X1 [Callithrix jacchus] ylserine lipase ABHD16A isoform X1 [Aolus nancymaae] HD16A isoform X1 [Theropithecus gelada]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus Aolus nancymaae Theropithecus gelada	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1           0         99.00%         601         XP         025238181.1
Standard Database Organism Optional Exclude Optional		<ul> <li>chosphalid</li> </ul>	viserine lipase ABHD16A isoform X1 [Gorilla, gorilla, gorilla) viserine lipase ABHD16A isoform X3 [Sapajus apella) viserine lipase ABHD16A isoform X3 [Callithrix jacchus] viserine lipase ABHD16A isoform X1.[Callithrix jacchus] viserine lipase ABHD16A isoform X1.[Aolus nancymaae] HD16A isoform X1.[Theropithecus gelada] viserine lipase ABHD16A isoform a [Homo sapiens]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus Aotus nancymaae Theropithecus gelada Homo sapiens	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1           0         99.00%         601         XP         025238181.1           0         99.67%         558         NP         066983.1
Standard Database Organism Optional Exclude Optional Program Select		<ul> <li>phosphalid</li> </ul>	ylserine lipase ABHD16A isoform X1 [Gorilla gorilla] ylserine lipase ABHD16A isoform X3 [Sapajus apella] ylserine lipase ABHD16A isoform X3 [Callithrix jacchus] ylserine lipase ABHD16A isoform X1.[Callithrix jacchus] ylserine lipase ABHD16A isoform X1.[Aotus nancymaae] HD16A isoform X1.[Theropithecus gelada] ylserine lipase ABHD16A isoform a [Homo sapiens] ylserine lipase ABHD16A isoform X2 [Aotus nancymaae]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus Aotus nancymaae Theropithecus gelada Homo sapiens Aotus nancymaae	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1           0         99.07%         601         XP         025238181.1           0         99.67%         558         NP         066983.1           0         99.67%         522         XP         012292100.1
Standard Database Organism Optional Exclude Optional Program Select Algorithm		phosphalid     protein ABH     protein ABH	viserine lipase ABHD16A isoform X1 [Gorilla gorilla] viserine lipase ABHD16A isoform X3 [Sapajus apelia] viserine lipase ABHD16A isoform X3 [Callithrix jacchus] viserine lipase ABHD16A isoform X1 [Callithrix jacchus] viserine lipase ABHD16A isoform X1 [Aotus nancymaae] HD16A isoform X1 [Theropithecus gelada] viserine lipase ABHD16A isoform a [Homo sapiens] viserine lipase ABHD16A isoform X2 [Aotus nancymaae] HD16A isoform X2 [Macaca mulatta]	Gorilla gorilla gorilla         Sapajus apella         Callithrix jacchus         Callithrix jacchus         Aotus nancymaae         Theropithecus gelada         Homo sapiens         Aotus nancymaae         Macaca mulatta	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1           0         99.00%         601         XP         025238181.1           0         99.67%         558         NP         066983.1           0         99.67%         522         XP         012292100.1           0         99.00%         587         XP         028703550.1
Standard Database Organism Optional Exclude Optional Program Select Algorithm		<ul> <li>phosphalid</li> <li>protein ABH</li> <li>phosphalid</li> <li>phosphalid</li> </ul>	viserine lipase ABHD16A isoform X1 [Gorilla gorilla] viserine lipase ABHD16A isoform X3 [Sapajus apella] viserine lipase ABHD16A isoform X3 [Callithrix jacchus] viserine lipase ABHD16A isoform X1 [Callithrix jacchus] viserine lipase ABHD16A isoform X1 [Aolus nancymaae] HD16A isoform X1 [Theropithecus gelada] viserine lipase ABHD16A isoform a [Homo sapiens] viserine lipase ABHD16A isoform X2 [Aolus nancymaae] HD16A isoform X2 [Macaca mulatta] viserine lipase ABHD16A isoform X1 [Sapajus apella]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus Aolus nancymaae Theropithecus gelada Homo sapiens Aolus nancymaae Macaca mulatta Sapajus apella	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         02292099.1           0         99.00%         601         XP         025238181.1           0         99.67%         558         NP         066983.1           0         99.67%         522         XP         012292100.1           0         99.00%         587         XP         028703550.1           0         99.67%         558         XP         032124881.1
Standard Database Organism Optional Exclude Optional Program Select Algorithm		<ul> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>protein ABH</li> <li>phosphalid</li> <li>protein ABH</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> </ul>	viserine lipase ABHD16A isoform X1 [Gorilla gorilla] viserine lipase ABHD16A isoform X3 [Sapajus apella] viserine lipase ABHD16A isoform X3 [Callithrix jacchus] viserine lipase ABHD16A isoform X1 [Callithrix jacchus] viserine lipase ABHD16A isoform X1 [Aotus nancymaae] HD16A isoform X1 [Theropithecus gelada] viserine lipase ABHD16A isoform at [Homo sapiens] viserine lipase ABHD16A isoform X2 [Aotus nancymaae] HD16A isoform X2 [Macaca mulatta] viserine lipase ABHD16A isoform X1 [Sapajus apella] viserine lipase ABHD16A isoform X1 [Sapajus apella]	Gorilla gorilla gorilla Sapajus apella Callithrix jacchus Callithrix jacchus Aotus nancymaae Theropithecus gelada Homo sapiens Aotus nancymaae Macaca mulatta Sapajus apella Saimiri boliviensis boliviensis	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1           0         99.067%         558         NP         066983.1           0         99.67%         522         XP         012292100.1           0         99.67%         522         XP         028703550.1           0         99.067%         558         XP         032124881.1           0         99.67%         558         XP         03212481.1
Standard Database Organism Optional Exclude Optional Program Select Algorithm		<ul> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>phosphalid</li> <li>protein ABH</li> <li>phosphalid</li> <li>protein ABH</li> <li>phosphalid</li> <li>phosphalid</li> <li>protein ABH</li> </ul>	VIserine lipase ABHD16A isoform X1 [Gorilla gorilla] VIserine lipase ABHD16A isoform X3 [Sapajus apella] VIserine lipase ABHD16A isoform X3 [Callithrix jacchus] VIserine lipase ABHD16A isoform X1 [Callithrix jacchus] VIserine lipase ABHD16A isoform X1 [Aotus nancymaae] HD16A isoform X1 [Theropithecus gelada] VIserine lipase ABHD16A isoform x2 [Aotus nancymaae] HD16A isoform X2 [Macaca mulatta] VIserine lipase ABHD16A isoform X1 [Sapajus apella] VIserine lipase ABHD16A isoform X1 [Sapajus apella] VIserine lipase ABHD16A isoform X1 [Sapajus apella] VIserine lipase ABHD16A isoform X1 [Sapajus apella]	Gorilla gorilla gorilla Sapajus apella Callithrix Jacchus Callithrix Jacchus Callithrix Jacchus Aotus nancymaae Theropithecus gelada Homo sapiens Aotus nancymaae Macaca mulatta Sapajus apella Saimiri boliviensis boliviensis Macaca mulatta	625         625         100%         0           624         624         100%         0           624         624         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0           623         623         100%         0	V         V         V           0         99.67%         601         XP         004043725.4           0         99.67%         457         XP         032124883.1           0         99.67%         457         XP         035151139.1           0         99.67%         558         XP         035151136.1           0         99.67%         558         XP         012292099.1           0         99.67%         558         NP         069983.1           0         99.67%         558         NP         028703550.1           0         99.00%         587         XP         032124881.1           0         99.67%         558         XP         032124281.1           0         99.67%         558         XP         03212481.1           0         99.67%         558         XP         03212488.11           0         99.67%         558         XP         03212488.11           0         99.67%         558         XP         0324439.11           0         99.00%         601         XP         00394439.11

# Does this peptide contain any transmembrane helices?

#### **TMHMM** result

# WEBSEQUENCE # WEBSEQUENCE # WEBSEQUENCE # WEBSEQUENCE # WEBSEQUENCE WEBSEQUENCE WEBSEQUENCE WEBSEQUENCE WEBSEQUENCE WEBSEQUENCE WEBSEQUENCE	Length: 300 Number of predic Exp number of A/ Exp number, firs Total prob of N- POSSIBLE N-term TMHMM2.0 TMHMM2.0 TMHMM2.0 TMHMM2.0 TMHMM2.0	ted TMHs: 2 s in TMHs: 4 in: 6 signal seque outside TMhelix inside TMhelix outside	2 41.20041 40.9446 3.37925 ence 1 4 24 35 58	3 23 34 57 300			
		TMHMM post	terior pr	obabilities for	WEBSEQUEN	ICE	
1.2	,						
-							
1	ana <b>ana</b> k						1
≥ 0.8							1
pilit of the second							
eqo 0.6							
ā 04							
0.4							
0.2							
o 💵							
	50	100		150 incide	200	250 Sutsida	300
	transmento	2018		inside —		outside	

Yes 2.

### What is molecular weigth of this peptide?



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.

LALASVFWSISYYSSPFAFFYLYRKGYLSLSKVVPFSHYAGTLLLLLAGVACLRGIGRWT NPQYRQFITILEATHRNQSSENKRQLANYNFDFRSWPVDFHWEEPSSRKESRGGPSRRGV ALLRPEPLHRGTADTLLNRVKKLPCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQG QARLVEECNGRRAKLLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTP LEACYSVLGWNHPGFAGSTGVPFPQNEANAMDVVVQFAIHRLGFQPQDIIIYAWSIGGFT

Submit Clear Deast	🕒 Sequence Manipulation Suite – Pracovní – Microsoft Edge
	i) about: <b>blank</b>
Add 1      copies of Nothing	Protein Molecular Weight results
*This page requires JavaScript. Se *You can mirror this page or use it	Results for 300 residue sequence "Untitled" starting "LALASVFWSI" 33.67 kDa
Sun 14 Jun 00:36:59 2020	

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

Differs in 1 aminoacid

**2025:** We did not cover this one

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
peptide NP_066983,1 Consensus	MAKLL	SCVLGPRLY	KIYRERDS	ERAPASVPETP	AVTAPHSS	SHDTYYQP	RALEKHADSI	LALASVFNSIS LALASVFNSIS LALASVFNSIS	YYSSPFAFF YYSSPFAFF YYSSPFAFF	YLYRKGYLSL YLYRKGYLSL YLYRKGYLSL	Skvvpfshyag Skvvpfshyag Skvvpfshyag	TLLLLAGVAC TLLLLAGVAC TLLLLAGVAC	LRGIGRATNPO LRGIGRATNPO LRGIGRATNPO	YRQFITI YRQFITI YRQFITI
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
peptide NP_066983.1 Consensus	LEATH LEATH LEATH	RNQSSENKR RNQSSENKR RNQSSENKR	QLANYNFDI Qlanynfdi Qlanynfdi	FRSHPYDFHHEE Frshpydfhhee Frshpydfhhee	PSSRKESR PSSRKESR PSSRKESR	GGPSRRGV GGPSRRGV GGPSRRGV	ALLRPEPLHR ALLRPEPLHR ALLRPEPLHR	GTADTLLNRYK GTADTLLNRYK GTADTLLNRYK	KLPCQITSY Klpcqitsy Klpcqitsy	. VAHTLGRRM LVAHTLGRRM LVAHTLGRRM	LYPGSVYLLQK Lypgsvyllqk Lypgsvyllqk	ALHPYLLQGQA Alhpyllqgqa Alhpyllqgqa	RLVEECNGRRA Rlveecngrra Rlveecngrra	KLLACDG KLLACDG KLLACDG
	261	270	280	290	300	310	320	330	340	350	360	370	380	390
peptide NP_066983.1 Consensus	NEIDTI NEIDTI NEIDTI	NFYDRRGTA NFYDRRGTA NFYDRRGTA	EPQGQKLY EPQGQKLY EPQGQKLY	LCCEGNAGFYEN LCCEGNAGFYEN LCCEGNAGFYEN	/GCVSTPLE /GCVSTPLE /GCVSTPLE	ACYSYLGH ACYSYLGH ACYSYLGH ACYSYLGH	NHPGFAGSTG NHPGFAGSTG NHPGFAGSTG	VPFPQNEANAM VPFPQNEANAM VPFPQNEANAM	DYYYQFAIH DYYYQFAIH DYYYQFAIH	RLGFQPQDII RLGFQPQDII RLGFQPQDII	LYANSIGGFT LYANSIGGFT LYANSIGGFT	THAAMSYPDVS	AHILDASFDDL	VPLALKY
	391 	400	410	420	430	440	450	460	470	480	490	500	510	520
peptide NP_066983.1	HPDSH	RGLVTRTVR	QHLNLNNAI	EQLCRYQGPYLL	IRRTKDEI	ITTTYPED	IMSNRGNDLL	LKLLQHRYPRY	HAEEGLRVVI	RQHLEASSQLI	EEASIYSRHE	EEDHCLSVLRS	YQAEHGPDFPH	SYGEDHS