

DÚ1: Literární rešerše – vyhledat informace k zadanému genu/proteinu/enzymu

1) Zjistěte, co znamená zkratka

Zadaný protein: **TNF α** = tumor necrosis factor alpha (někdy označován jako kachektin; homotrimerický protein o 157 AMK; rodina cytokinů)

2) Vyhledejte 5 relevantních zdrojů o tomto enzymu

- alespoň jedno review
- alespoň jeden článek z roku 2017
- 1 článek starší než rok 2000

Používaná databáze k vyhledávání = PubMed. Využití filtrů na levé straně webu.

Journal Articles

- Huang J, Ning N, Zhang W. [Effects of paraquat on IL-6 and TNF- \$\alpha\$ in macrophages](#). *Exp Ther Med*. 2019 Mar;17(3):1783-1789. doi: 10.3892/etm.2018.7099. Epub 2018 Dec 14. PubMed PMID: 30783450; PubMed Central PMCID: PMC6364147.
[Related citations](#)
- Ming H, Tian A, Liu B, Hu Y, Liu C, Chen R, Cheng L. [Inflammatory cytokines tumor necrosis factor- \$\alpha\$, interleukin-8 and sleep monitoring in patients with obstructive sleep apnea syndrome](#). *Exp Ther Med*. 2019 Mar;17(3):1766-1770. doi: 10.3892/etm.2018.7110. Epub 2018 Dec 18. PubMed PMID: 30783447; PubMed Central PMCID: PMC6364239.
[Related citations](#)
- Luo Y, He H, Zhang J, Ou Y, Fan N. [Changes in serum TNF- \$\alpha\$, IL-18, and IL-6 concentrations in patients with chronic schizophrenia at admission and at discharge](#). *Compr Psychiatry*. 2019 Feb 8;90:82-87. doi: 10.1016/j.comppsy.2019.01.003. [Epub ahead of print] PubMed PMID: 30782515.
[Related citations](#)
- Ciebiera M, Włodarczyk M, Zgliczyńska M, Łukaszyk K, Męczałski B, Kobierzycki C, Łoziński T, Jakiel G. [The Role of Tumor Necrosis Factor \$\alpha\$ in the Biology of Uterine Fibroids and the Related Symptoms](#). *Int J Mol Sci*. 2018 Dec 4;19(12). pii: E3869. doi: 10.3390/ijms19123869. Review. PubMed PMID: 30518097; PubMed Central PMCID: PMC6321234.
[Free full text](#) [Related citations](#)
- Fryer AD, Jacoby DB, Wicher SA. [Protective Role of Eosinophils and TNFa after Ozone Inhalation](#). *Res Rep Health Eff Inst*. 2017 Mar;(191):1-41. PubMed PMID: 29659241.
[Related citations](#)
- Maugeri D, Russo MS, Franzé C, Motta V, Motta M, Destro G, Speciale S, Santangelo A, Panebianco P, Malaguarnera M. [Correlations between C-reactive protein, interleukin-6, tumor necrosis factor-alpha and body mass index during senile osteoporosis](#). *Arch Gerontol Geriatr*. 1998 Sep-Oct;27(2):159-63. PubMed PMID: 18653160.
[Cited in PMC](#) [Related citations](#)

review

2017

starší
než 2000

3) založte si WEB Endnote účet a importujte tyto zdroje do seznamu „My references“

My References

All My References (6)

[Untitled] (6)

Quick List (0)

Trash (1) Empty

My Groups

New Group (0)

New Group (0)

TRACK YOUR CITATIONS

Claim your researcher profile on Publons

Sort by: Year -- oldest to newest

Author	Year	Title
<input type="checkbox"/> Maugeri, D.	1998	Correlations between C-reactive protein, interleukin-6, tumor necrosis factor-alpha and body mass index during senile osteoporosis <i>Arch Gerontol Geriatr</i> Added to Library: 22 Feb 2019 Last Updated: 22 Feb 2019
<input type="checkbox"/> Fryer, A. D.	2017	Protective Role of Eosinophils and TNFa after Ozone Inhalation <i>Res Rep Health Eff Inst</i> Added to Library: 22 Feb 2019 Last Updated: 22 Feb 2019
<input type="checkbox"/> Ciebiera, M.	2018	The Role of Tumor Necrosis Factor alpha in the Biology of Uterine Fibroids and the Related Symptoms <i>Int J Mol Sci</i> Added to Library: 22 Feb 2019 Last Updated: 22 Feb 2019
<input type="checkbox"/> Huang, J.	2019	Effects of paraquat on IL-6 and TNF-alpha in macrophages <i>Exp Ther Med</i> Added to Library: 22 Feb 2019 Last Updated: 22 Feb 2019
<input type="checkbox"/> Luo, Y.	2019	Changes in serum TNF-alpha, IL-18, and IL-6 concentrations in patients with chronic schizophrenia at admission and at discharge <i>Compr Psychiatry</i> Added to Library: 22 Feb 2019 Last Updated: 22 Feb 2019
<input type="checkbox"/> Ming, H.	2019	Inflammatory cytokines tumor necrosis factor-alpha, interleukin-8 and sleep monitoring in patients with obstructive sleep apnea syndrome <i>Exp Ther Med</i> Added to Library: 22 Feb 2019 Last Updated: 22 Feb 2019

Show 10 per page

Page 1 of 1 Go

4) Zjistěte H-index vašeho oblíbeného profesora z FaF

Hledání Hirschova indexu skrze databázi Scopus. Lze najít další možné informace o publikační činnosti a citovanosti autora.

Scopus Preview

Author search

Sources

Help ▾

Register >

Login ▾

Author details

Print Email

Nováková, Lucie

Charles University, Prague Praha, Czech Republic
Author ID: 6603839921

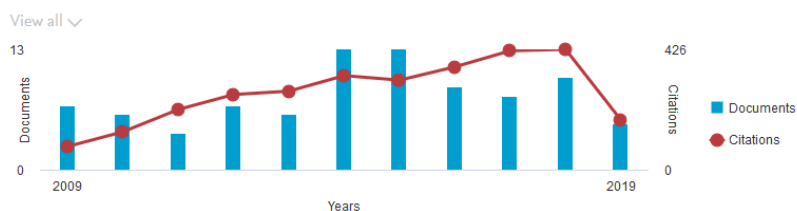
<http://orcid.org/0000-0003-3570-5871>

Other name formats: [Novakova, Lucie](#) [Nováková, L.](#)

Subject area:

[Chemistry](#) [Biochemistry, Genetics and Molecular Biology](#) [Pharmacology, Toxicology and Pharmaceutics](#)
[Chemical Engineering](#) [Environmental Science](#) [Health Professions](#) [Medicine](#) [Agricultural and Biological Sciences](#)

Document and citation trends:



Follow this Author

View potential author matches

h-index: 30

[View h-graph](#)

Documents by author

104

[Analyze author output](#)

Total citations

3135 by 2562 documents

DÚ2: Vyhledávání proteinů - pracujte s „vaším“ genem/proteinem/enzymem

1) Vyhledejte NCBI a UniProt přístupové kódy vašeho proteinu

NCBI: NP_000585.2

UniProt: P01375

UniProtKB - **P01375** (TNFA_HUMAN)

Display [BLAST](#) [Align](#) [Format](#) [Added to basket](#) [History](#)

Entry **Tumor necrosis factor** [Homo sapiens]

Gene **TNF**

Organism **Homo sapiens (Human)**

Status **Reviewed** - Annotation score: ●●●●● - Experimental evidence at protein level!

NCBI Reference Sequence: **NP_000585.2**

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

2) Najděte a uložte si „FASTA“ formát sekvence (jako text)

Na NCBI → [kliknout na FASTA](#)

tumor necrosis factor [Homo sapiens]

NCBI Reference Sequence: NP_000585.2

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

```
>NP_000585.2 tumor necrosis factor [Homo sapiens]
MSTESMIRDVELAEALPKKTGGPQGSRRCLFLSLFSFLIVAGATTLFCLLHFGVIGPQREEFPRDLSLI
SPLAQAVRSSSRTPSDKPVAVHVVANPQAEQGQLQWLNRRANALLANGVELRDNLVVPSEGLYLIYSQVLF
KGQGCPSSTHVLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPYIYLGGVFQLEKGDRLSA
EINRPDYLDFAESGQVYFGI IAL
```

3) Vyberte část proteinu od 100. do 150. aminokyseliny

Skrze program SMS/Range Extractor Protein; výběr oblasti 100..150 AMK

Sequence Manipulation Suite:

Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or ranges. The residues corresponding to the position, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor Protein to obtain subsequences using

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

```
>NP_000585.2 tumor necrosis factor [Homo sapiens]
MSTESMIRDVELAEALPKKTGGPQGSRRCLFLSLFSFLIVAGATTLFCLLHFGVIGPQREEF
PRDLSLISPLAQAVRSSSRTPSDKPVAVHVVANPQAEQGQLQWLNRRANALLANGVELRDNLV
PSEGLYLIYSQVLFKGQGCPSSTHVLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAK
PWYEPYIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGI IAL
```

Enter the residue positions or ranges to be extracted. Use "." to represent a range, and use a comma to separate entries. The word "center", to represent the beginning, end, middle, and length of the sequence. Arithmetic expressions can be included in the range of a sequence, the range '(end - 2)..' can be used to represent the last two residues. The range '(center + 1)..' can be used to represent the center, '(center + 1)..' (center + 30)' can be used to represent the center plus 30 residues.

100..150

Submit Clear Reset

• Sequence segments should be

*This page requires JavaScript. See [browser compatibility](#)

*You can [mirror this page](#) or use it off-line.

Sequence Manipulation Suite - Mozilla Firefox

www.bioinformatics.org/sms2/range_extract_protein.html

Range Extractor Protein results

Results for 233 residue sequence "NP_000585.2 tumor necrosis factor [Homo sapiens]-100..150 AMK" starting "GQLQWLNRRRA"

5.67 kDa

oblast 100..150 AMK

4) Spočítejte molekulovou hmotnost tohoto úseku proteinu

Výpočet Mr úseku proteinu/peptidu – pomocí programu SMS/Protein Molecular Weight

Sequence Manipulation Suite:

Protein Molecular Weight

Protein Molecular Weight accepts one or more protein sequences and calculates molecular weight. You can append copy number to the list. Use Protein Molecular Weight when you wish to predict the location of a protein of interest on a gel in relative

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

```
>NP_000585.2 tumor necrosis factor [Homo sapiens]-100..150 AMK
GQLQWLNRRANALLANGVELRDNLVVPSEGLYLIYSQVLFKGQGCPSSTHV
```

Submit Clear Reset

• Add 1 copies of Nothing to

*This page requires JavaScript. See [browser compatibility](#)

*You can [mirror this page](#) or use it off-line.

Sequence Manipulation Suite - Mozilla Firefox

www.bioinformatics.org/sms2/protein_molecular_weight.html

Protein Molecular Weight results

Results for 51 residue sequence "NP_000585.2 tumor necrosis factor [Homo sapiens]-100..150 AMK" starting "GQLQWLNRRRA"

5.67 kDa

zvolený úsek proteinu má molekulovou hmotnost 5,67

5) Zjistěte, kolikrát bude vaše celá sekvence štěpena trypsinem

Štěpení trypsinem simulováno pomocí programu PeptideCutter

Trypsin 19 8 19 20 28 29 60 65 78 82 107 108 120 141 158 166 174 179 204 207

Celý protein/sekvence TNFA bude trypsinem štěpen 19x.

DÚ 3: Analýza proteinů, podobnost

1) Obsahuje váš protein nějaké typické motivy?

Ano, obsahuje.

NCBI/CD

Protein Classification
TNF domain-containing protein (domain architecture ID 10446394)
 TNF domain-containing protein

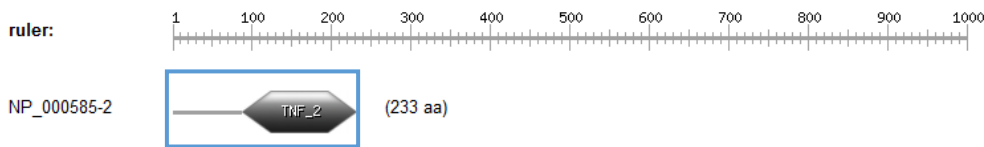
Graphical summary Zoom to residue level [show extra options >](#)

List of domain hits

Name	Accession	Description	Interval	E-value
TNF	pfam00229	TNF(tumor Necrosis Factor) family;	103-233	2.99e-53

ScanProsite

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



PS50049 TNF_2 TNF family profile :

89 - 233: score = 38.605
 VAHVVANPQAEQG--LQWLNRRANALLAnGVELRDNLVVPSEGLYLIYSQVLFKGGQGCP
 SHVLLHTIISRIVSYQIKVNLLSAIKSPcqrptpeGAEAKPWYEPIYLGGVFQLEKGD
 RLSAEINRPDYLDFAEaGQVYFGIIAL

SMART

Domains within *Homo sapiens* protein TNFA_HUMAN (P01375)
 Tumor necrosis factor

Information | Architecture | Interactions | Pathways | PTMs | Orthology

Length: 233 aa
 Source database: UniProt

Identifiers: TNFA_HUMAN, P01375, ENSP00000389492.2, ENSP00000389492, ENSP00000389698.2, ENSP00000389698.4, ENSP00000372988.4, ENSP00000372988, ENSP00000365280.3, ENSP00000365280, ENSP00000392A0A0U5J3S4, G3QT76_GORGO, G3QT76, A0A140T922_HUMAN, A0A140T922, B5BU06_HUMAN, B5BL

Source gene: ENSG00000230108

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than piece of sequence; the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start	End	E-value
transmembrane region	35	57	N/A
TNF	88	233	9.43e-56

Pfam

Summary

This is the summary of UniProt entry [TNFA_HUMAN](#) (P01375).

Description:	Tumor necrosis factor
Source organism:	Homo sapiens (Human) (NCBI taxonomy ID 9606) View Pfam proteome data.
Length:	233 amino acids
Reference Proteome:	<input checked="" type="checkbox"/>

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. It removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Click [Download](#) the data used to generate the domain graphic in JSON format.

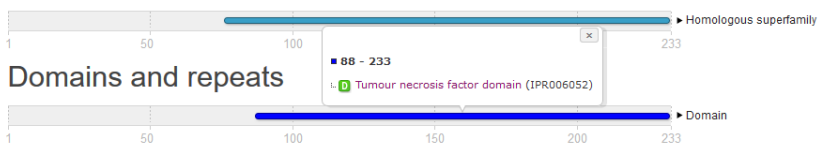
Source	Domain	Start	End
disorder	n/a	13	19
transmembrane	n/a	30	55
low_complexity	n/a	30	38
disorder	n/a	85	92
Pfam	TNF	102	233

InterPro

Protein family membership

- Tumour necrosis factor (IPR006053)
- Tumour necrosis factor alpha (IPR002959)

Homologous superfamilies



4) Vyberte pět podobných sekvencí (vyhledaných BLASTem) z jiných organismů – uložte si sekvence

>[Homo sapiens]

MSTESMIRDVELAEALPKKTGGPQGSRRCLFSLFSLVIVAGATTFLCLLHFGVIGPQREEFPRLDLSISPLAQAVRSSRTPSDKPAHVVANPQAEGLQWLNRRAN
ALLANGVELRDNLVVPSEGLYLIYSQVLFKQGQCPSTHVLLTHTISRIASVYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINRPDYLDFAE
SGQVYFGIIAL

>[Pan troglodytes]

MSTESMIRDVELAEALPKKTGGPQGSRRCLFSLFSLVIVAGATTFLCLLHFGVIGPQREEFPRLDLSISPLAQAGSSRTPSDKPAHVVANPQAEGLQWLNRRANA
LLANGVELRDNLVVPSEGLYLIYSQVLFKQGQCPSTHVLLTHTISRIASVYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINRPDYLDFAES
GQVYFGIIAL

>[Pongo abelii]

MSTESMIRDVELAEALPKKTGGPHGSRCLFSLFSLVIVAGATTFLCLLHFGVIGPQREEFPKDLISPLAQAVRSS
SRTPSDKPAHVVANPQAEGLQWLNRRANALLANGVELRDNLVVPSEGLYLIYSQVLFKQGQCPSTHVLLTHTISRIA
VSYQTKVNLLSAIKSPCQRETTEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINLPNYLDFAESGQVYFGIIAL

>[Ptilocolobus tephrosceles]

MSTESMIRDVELAEALPRKTAGPQGSRRWFLSFLVAGATTFLCLLHFGVIGPQREEFPKDLISPLAQAVRSSRTPSDKPAHVVANPQAEGLQWLNRRAN
NALLANGVELTDNLVVPSEGLYLIYSQVLFKQGQCPSNHVLLTHTISRIASVYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINRPDYLD
AESGQVYFGIIAL

>[Macaca mulatta]

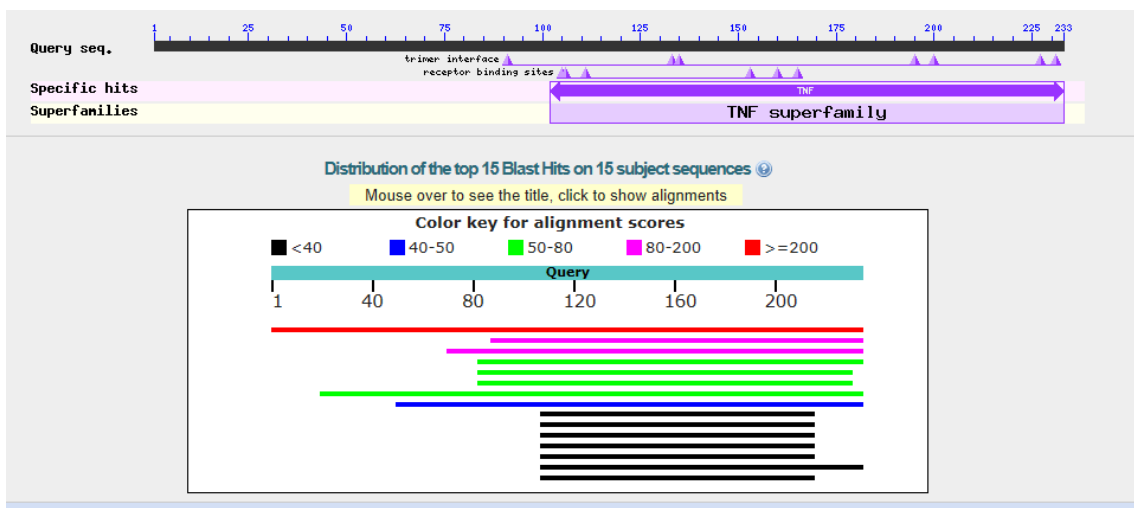
MSTESMIRDVELAEALPRKTAGPQGSRRWFLSFLVAGATTFLCLLHFGVIGPQREEFPKDPISPLAQAVRSSRTPSDKPAHVVANPQAEGLQWLNRRAN
NALLANGVELTDNLVVPSEGLYLIYSQVLFKQGQCPSNHVLLTHTISRIASVYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINLPDYLD
AESGQVYFGIIAL

>[Chlorocebus sabaeus]

MSTESMIRDVELAEALPRKTAGPQGSRRWFLSFLVAGATTFLCLLHFGVIGPQREEFPKDPISPLAQAVRSSRTPSDKPAHVVANPQAEGLQWLNRRAN
ANALLANGVELTDNLVVPSEGLYLIYSQVLFKQGQCPSNHVLLTHTISRIASVYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINLPDYLD
FAESGQVYFGIIAL

5) Vyhleďte podobné referenční sekvence jen u Homo sapiens – kolik sekvencí bylo nalezeno?

(je třeba zaškrtnout prohledávání „Refseq“ a omezit na Homo sapiens/human/)



Sequences producing significant alignments:

Select: All None Selected: 15

Alignments Download GapsPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/> tumor necrosis factor (Homo sapiens)	478	478	100%	9e-174	100.00%	NP_000585.2
<input checked="" type="checkbox"/> lymphotxin-alpha precursor (Homo sapiens)	88.2	88.2	63%	1e-20	33.77%	NP_000586.2
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 15 isoform VEGI-251 precursor (Homo sapiens)	82.8	82.8	70%	2e-18	28.57%	NP_005109.2
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 15 isoform VEGI-192 (Homo sapiens)	79.7	79.7	65%	9e-18	28.82%	NP_001191273.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 14 isoform 2 (Homo sapiens)	68.9	68.9	63%	1e-13	30.72%	NP_742011.2
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 14 isoform 1 (Homo sapiens)	68.9	68.9	63%	2e-13	30.72%	NP_003798.2
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 6 isoform 1 (Homo sapiens)	68.6	68.6	91%	4e-13	27.15%	NP_000630.1
<input checked="" type="checkbox"/> CD40 ligand (Homo sapiens)	48.1	48.1	78%	5e-06	26.40%	NP_000065.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 13 isoform zeta (Homo sapiens)	38.9	38.9	46%	0.005	31.48%	NP_001185552.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 13 isoform delta (Homo sapiens)	38.9	38.9	46%	0.005	31.48%	NP_001185551.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 13 isoform eta (Homo sapiens)	38.9	38.9	46%	0.005	31.48%	NP_001185553.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 13 isoform beta precursor (Homo sapiens)	38.5	38.5	46%	0.007	31.48%	NP_742084.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 13 isoform gamma (Homo sapiens)	38.5	38.5	46%	0.008	31.48%	NP_742085.1
<input checked="" type="checkbox"/> tumor necrosis factor ligand superfamily member 13 isoform alpha precursor (Homo sapiens)	38.5	38.5	54%	0.008	29.69%	NP_003799.1
<input checked="" type="checkbox"/> TNFSF12-TNFSF13 protein (Homo sapiens)	37.7	37.7	46%	0.018	31.48%	NP_742086.1

Možno nalézt 15 podobných enzymů nebo sekvencí.

DÚ 4: Proteinová bioinformatika III + shrnutí I

1) Porovnejte „vás“ protein se „stejným“ proteinem z myši – párové porovnání. Jaká je identita těchto dvou sekvencí? Proběhlo porovnání celé délky sekvence?

EMBOSS Needle

```
Protein alignment Nucleotide alignment Web services Help & Documentation Bioinformatics Tools FAQ

# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#-----
#
# Aligned_sequences: 2
# 1: TNFA_HUMAN
# 2: TNFA_MOUSE
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 236
# Identity: 186/236 (78.8%)
# Similarity: 211/236 (89.4%)
# Gaps: 4/236 ( 1.7%)
# Score: 958.5
#-----

TNFA_HUMAN 1 MSTESHIRVDLAEAEALPKKTTGGPQGSRRCLFLSFLVAVGATTLFCL 50
TNFA_MOUSE 1 MSTESHIRVDLAEAEALPKKTTGGPQGSRRCLFLSFLVAVGATTLFCL 50
TNFA_HUMAN 51 LHFVIGPQREE--FPRDLSLISPLAQ--AVRSSRTPSDKPAHVAVANPQ 97
TNFA_MOUSE 51 LNFVIGPQREKDFPGLPLISSMAQTLTLRSSSQNSDKPAHVAVANHQ 100
TNFA_HUMAN 98 AEGQLQLNRRANALLANGVELRDNQLVVPSEGLYLIYSQVLFKGGQCP 147
TNFA_MOUSE 101 VEEQLNLSQRANALLANGHLDKDNQLVVPADGLYLYVYQVLFKGGQCP 158
TNFA_HUMAN 148 THVLLTHTISRIVAVSQTKNVLSAIKSPQRETPEGAEKPHYPIYLG 197
TNFA_MOUSE 151 -YVLLTHTVSRFAISVQEKVLLSAVKSQPKDTPEGAELKPHYPIYLG 199
TNFA_HUMAN 198 GVFLQLEKGRDLSAEINRPDYLDAESQVYFGIIAL 233
TNFA_MOUSE 200 GVFLQLEKGDQLSAEINLPKYLDFAESQVYFGVIAL 235
```

- Sekvence porovnány globálně
- Porovnány v celé délce, tj. 236 aminokyselin
- Identita činí 78,8 %, similarita 89,4 %

LALIGN

```
Protein alignment Nucleotide alignment Web services Help & Documentation Bioinformatics Tools FAQ

Threshold: E() < 10 score: 28
Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)
Parameters: BLS0 matrix (15/-5), open/ext: -12/-2
Scan time: 0.000

The best non-identical alignments are: 1s-w bits E(1) %_id %_sim aLen
sp|P06804|TNFA_MOUSE Tumor necrosis ( 235) 1214 360.3 1.9e-104 0.788 0.919 236
+-
+- 35 14.4 0.91 0.364 0.727 11
+- 31 13.3 1 0.294 0.706 17
+- 30 13.0 1 0.438 0.625 16
+- 29 12.7 1 0.294 0.529 34
+- 29 12.7 1 0.500 0.600 10
+- 29 12.7 1 0.667 1.000 6

>>>sp|P01375|TNFA_HUMAN, 233 aa vs lalign-I20190312-080347-0648-85385021-p2m.bsequence library

>>sp|P06804|TNFA_MOUSE Tumor necrosis factor OS=Mus musc (235 aa)
Waterman-Eggert score: 1214; 360.3 bits; E(1) < 1.9e-104
78.8% identity (91.9% similar) in 236 aa overlap (1-233:1-235)

10 20 30 40 50 60
sp|P01 MSTESHIRVDLAEAEALPKKTTGGPQGSRRCLFLSFLVAVGATTLFCLLHFVIGPQR
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
sp|P06 MSTESHIRVDLAEAEALPKKTTGGPQGSRRCLFLSFLVAVGATTLFCLLHFVIGPQR
10 20 30 40 50 60

70 80 90 100 110
sp|P01 EE-FPRDLSLISPLAQAV--RSSRTPSDKPAHVAVANPQAEGLQLNRRANALLANGV
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
sp|P06 DEKFPNGLPLISSMAQTLTLRSSSQNSDKPAHVAVANHQVEEQLNLSQRANALLANGH
70 80 90 100 110 120

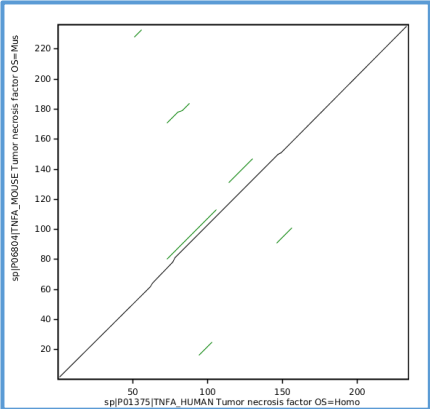
120 130 140 150 160 170
sp|P01 ELRDNQLVVPSEGLYLIYSQVLFKGGQCPSTHLLTHTISRIVAVSQTKNVLSAIKSPC
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
sp|P06 DLKDNQLVVPADGLYLYVYQVLFKGGQCP--YVLLTHTVSRFAISVQEKVLLSAVKSQ
130 140 150 160 170

180 190 200 210 220 230
sp|P01 QRETPEGAEKPHYPIYLGQVFLQLEKGRDLSAEINRPDYLDAESQVYFGIIAL
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
sp|P06 PKDTPEGAELKPHYPIYLGQVFLQLEKGRDLSAEINLPKYLDFAESQVYFGVIAL
180 190 200 210 220 230

>>>
Waterman-Eggert score: 35; 14.4 bits; E(1) < 0.91
36.4% identity (72.7% similar) in 11 aa overlap (146-156:91-101)

150
sp|P01 PSTHLLTHTI
: : : : :
sp|P06 PVAHVAVANHQV
```

- Sekvence porovnány lokálně
- Celkové skóre je 1214
- Identita činí 78,8 %, similarita činí 91,9 %



Grafický výstup lokálního porovnání aminokyselin Nejdelsí sekvence

2) Vytvořte mnohonásobné porovnání všech sekvencí z minulého úkolu (DÚ3) + vystřihněte „evoluční strom“

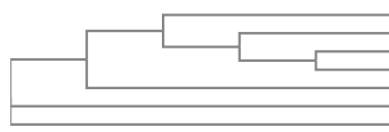
MultAlin

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
[Homo [Pan	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGVELDNQLVVPSEG													
[Pongo	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGVELDNQLVVPSEG													
[Piliocolobus	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGVELDNQLVVPSEG													
[Macaca	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGVELDNQLVVPSEG													
[Chlorocebus	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGVELDNQLVVPSEG													
Consensus	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGVELDNQLVVPSEG													

Clustal Omega

CLUSTAL O(1.2.4) multiple sequence alignment

[Mus	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60
[Piliocolobus	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60
[Macaca	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60
[Chlorocebus	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60
[Pongo	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60
[Homo	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60
[Pan	MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCLLHFGVIGPQR	60



[Mus 0.18412
[Piliocolobus 0.00536
[Macaca 0
[Chlorocebus 0.0043
[Pongo 0.01869
[Homo 0.00038
[Pan 0.00393

[Mus	DEKFPNGLPLISSMAQTLTRSSSQSSDKPVAVHVNANPQREGQLNRRANALLANGH	120
[Piliocolobus	E-EFPKDLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGV	117
[Macaca	E-EFPKDLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGV	117
[Chlorocebus	E-EFPKDLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGV	117
[Pongo	E-EFPKDLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGV	117
[Homo	E-EFPKDLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGV	117
[Pan	E-EFPKDLISPLAQAVRSSSRTPSDKPVAVHVNANPQREGQLNRRANALLANGV	116

[Mus	DLKDQNLVVPADGLVLYSIVLQFKGGCPDYVLLTHTISRIVSYQTKVNLSSAISKSP	179
[Piliocolobus	ELTDQNLVVPSEGLYLYSIVLQFKGGCPDMHLLTHTISRIVSYQTKVNLSSAISKSP	177
[Macaca	ELTDQNLVVPSEGLYLYSIVLQFKGGCPDMHLLTHTISRIVSYQTKVNLSSAISKSP	177
[Chlorocebus	ELTDQNLVVPSEGLYLYSIVLQFKGGCPDMHLLTHTISRIVSYQTKVNLSSAISKSP	177
[Pongo	ELRDQNLVVPSEGLYLYSIVLQFKGGCPDMHLLTHTISRIVSYQTKVNLSSAISKSP	177
[Homo	ELRDQNLVVPSEGLYLYSIVLQFKGGCPDMHLLTHTISRIVSYQTKVNLSSAISKSP	177
[Pan	ELRDQNLVVPSEGLYLYSIVLQFKGGCPDMHLLTHTISRIVSYQTKVNLSSAISKSP	176

[Mus	PKDTPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	235
[Piliocolobus	QRETPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	233
[Macaca	QRETPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	233
[Chlorocebus	QRETPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	233
[Pongo	QRETPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	233
[Homo	QRETPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	233
[Pan	QRETPGAEAKPWVEPIYLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIATL	232

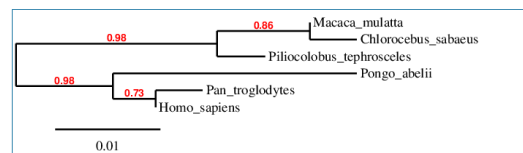


Figure 1: Phylogenetic tree

3) Má váš protein nějaké izofomy? Porovnejte je...

- protein TNFA má jednu potenciální izofomu
- jejich srovnání:

```

# Aligned_sequences: 2
# 1: TNFA_HUMAN
# 2: A0A140T922_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 233
# Identity: 171/233 (73.4%)
# Similarity: 171/233 (73.4%)
# Gaps: 62/233 (26.6%)
# Score: 878.0
#
#-----#
TNFA_HUMAN      1 MSTESHIRDELVEAEALPKKTGGPQGSRRCLFSLFSLFVAGATTFLCL  50
A0A140T922_HU  1 ----- 0
TNFA_HUMAN     51 LHFVIGPQREEFKDLISLISPLAQAVRSSSRTPSDKPVAVHVNANPQAE  100
A0A140T922_HU  1 -----FPRDLSLISPLAQAVRSSSRTPSDKPVAVHVNANPQAE  38
TNFA_HUMAN     101 QLQWLNRRANALLANGVELRDNLVVPSEGLYLYSIVLQFKGGCPSTHV  150
A0A140T922_HU  39 QLQWLNRRANALLANGVELRDNLVVPSEGLYLYSIVLQFKGGCPSTHV  88
TNFA_HUMAN     151 LLTHTISRIVSYQTKVNLSSAISKSPQRETPGAEAKPWVEPIYLGGVF  200
A0A140T922_HU  89 LLTHTISRIVSYQTKVNLSSAISKSPQRETPGAEAKPWVEPIYLGGVF  138
TNFA_HUMAN     201 QLEKGDRLSAEINRPDYLDFAESGQVYFGIATL  233
A0A140T922_HU  139 QLEKGDRLSAEINRPDYLDFAESGQVYFGIATL  171
#-----#

```

4) Byla určena 3D struktura vašeho proteinu? Vystřihněte jednu na ukázku.

Pomocí Protein Data Bank (PDB)

Structure Summary | **3D View** | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

1TNF Display Files Download Files

THE STRUCTURE OF TUMOR NECROSIS FACTOR-ALPHA AT 2.6 ANGSTROMS RESOLUTION. IMPLICATIONS FOR RECEPTOR BINDING

Note: Use your mouse to drag, rotate, and zoom in and out of the structure. Mouse-over to identify atoms and bonds. Mouse controls documentation.

Structure View | Electron Density Maps | Ligand View

Structure View Documentation

Assembly

Model

Symmetry

Style

Color

Ligand

Quality

Water Ions

Hydrogens Clashes

Default Structure View

Spin Center Fullscreen Screenshot Perspective Camera White background Focus

5) Je váš protein enzym? Jaké má enzymové číslo (EC)?

Zjištění charakteristiky proteinů/enzymů – pomocí webu Brenda.

Nejedná se o enzym.

DÚ5: Vyhledávání nukleotidových sekvencí

1) Vyhledejte si vaši nukleotidovou sekvenci. Zapište si přístupový kód.

- *NM_000594.4*

2) Vyberte pouze kódující sekvenci (CDS), uložte formát FASTA.

- >NM_000594.4:178-879 Homo sapiens tumor necrosis factor (TNF), mRNA
ATGAGCACTGAAAGCATGATCCGGGACGTGGAGCTGGCCGAGGAGGCGCTCCCAAGAAGACAGGGGGG
CCCAGGGCTCCAGGCGGTGCTTGTCTCTCAGCCTCTTCTCTTCTGATCGTGGCAGGCGCCACCACGCT
CTTCTGCCTGCTGCACCTTTGGAGTGATCGGCCCCAGAGGGAAGATTCCCCAGGGACCTCTCTAATC
AGCCCTCTGCCCCAGGCAGTCAGATCATCTTCTCGAACCCCGAGTGACAAGCCTGTAGCCCATGTTGTAG
CAAACCTCAAGCTGAGGGGAGCTCCAGTGGCTGAACCGCCGGCCAATGCCCTCTGGCCAATGGCGT
GGAGCTGAGAGATAACCAGCTGGTGGTGCCATCAGAGGGCCTGTACCTCATCTACTCCCAGGTCCTCTTC
AAGGGCCAAGGCTGCCCTCCACCCATGTGCTCCTCACCCACACCATCAGCCGCATCGCCGTCTCTACC
AGACCAAGGTCAACCTCTCTCTGCCATCAAGAGCCCTGCCAGAGGGAGACCCAGAGGGGGGCTGAGGC
CAAGCCCTGGTATGAGCCCATCTATCTGGGAGGGTCTTCCAGCTGGAGAAGGGTGACCGACTCAGCGCT
GAGATCAATCGCCCGACTATCTCGACTTTGCCGAGTCTGGCAGGTCTACTTTGGGATCATTGCCCTGT
GA

ORIGIN

```
1 agcagagcgt cctcagcaa gacagcaga ggaacagcta agagggagag aagcaactac
61 agacccccc tgaaaaacaac cctcagcagc cacatccccc gacaagctgc caggcaggtt
121 cttctccctt cacataatga cccacagctc caccctctct cccctggaaa gacaccctc
181 agcaatgaaa gcatgacccg ggaagggag agggccagag aggcctcacc caagaagaca
241 agggggcccc agggctccag ggggctctg ttcccccagc tcttctctt cctgatctg
301 caagagcaca ccacgctctt ctgctctctg aactctgag tgatcgccc ccagaagaaa
361 gaattccaca gggacctctc tcaatcagc cctcggccc aggcagcag atcatctct
421 gaaaccocga gtagcaagcc tgragccat gttagcaaa acctcaagc tgaagggcag
481 ctccagctgc tgaacggccg ggcacatgpc ctcccgcca atgggtgga gctgagagat
541 aaccagctgg tggctccacc agagggctg taccctatct atcccaggc cctcttcaa
601 ggcacaagct gccctccacc caatggctc ctacccaca caatcagcc catcgccctc
661 tctaccaga caaaggtcaa cctctctctt gccatcaaga gccctgcca gaggagacc
721 ccagagggg ctgagggcca gccctggat gagccatct atctggagg gctcttcaa
781 ctgagaaag gtagccatct cagcctgag atcaatcgc ccaatctct ccaatctgca
841 gactctggc agcctactt tggacactt gccctggag gaggacgac atccaacctt
901 cccaaagcc tccctgccc caatccctt attaccccc cctcagaca ccccaacctt
961 ctctggctc aaaaagaaa ttggggctt aggtctgaa cccaagetta gaactttaag
1021 caacaagacc accacttoga aacctggat tcaagatgt gtggctgca cagtgaagt
1081 ctggcaacca ctagaatcc aacctgggc ctccagaact cactggggc tacagctttg
1141 atccttgaca ttggaatct ggagaccag gacccctgg ttctggccc aatgctgag
1201 gacttgaaa gacctcaact agaaattgac acaattgac cttaggctt cctcttcca
1261 gatgttcca gacttcctg agacacggag cccagccctc ccaatggagc cagctccctc
1321 tatttatgt tgaacttgg attatttat attatttat tatttatta ttacagatg
1381 aatgtatta ttggggagc cgggtatcc ttggggacc aatgtaggag ctgctctgc
1441 tcaacatgt ttcccgtaa aacggagctg aacaatagg tgttccatg tagccccctg
1501 gccctctgc cttctttga ttatgtttt taaatattt atctgattaa gttgtctaaa
1561 caatgctat ttggtagca actgcaact attgctgag cctctctcc cagggagtt
1621 gttctgtaa tcgcccact atccagctg gagaataaa gtttcttag aaaaaga
```

178..879

```
/gene="TNF"
/gene_synonym="DIF; TNF-alpha; TNFA; TNFSF2; TNLG1F"
/note="cachectin; TNF, monocyte-derived; TNF, macrophage-derived; APCI protein; tumor necrosis factor-alpha; TNF-a; tumor necrosis factor ligand superfamily member 2; tumor necrosis factor ligand 1F"
/codon_start=1
/product="tumor necrosis factor"
/protein_id=" NP_000585.2 "
/db_xref="CCDS: CCDS4702.1 "
/db_xref="GeneID: 7124 "
/db_xref="HGNC: HGNC:11892 "
/db_xref="MIM: 191160 "
/translation="MSTESMIRDVELAEELPKKTTGGPQGSRRCLFLSLFSLIVAGA
TTLFCLLHFGVIGPQREEFPRDLISLISPLAQAVRSSRIPSDKPVAVHVNPNQAEGL
QWLNRANALLANGVELRDNLVVPSEGLYLIYSQVLFKGGCPCSTHLLTHIISRIA
VSYQTQVNLLSAIKSPCQRETPEGAEAKPWEPYIYLVGGVQLEKDRLSAEINRPDYL
DFAESQGVYFGIIAL"
```

Details

Display:

FASTA

GenBank

Help

3) Na kterém chromozomu leží? Kolik má exonů? Jak je dlouhý 2. exon? Stáhněte jeho sekvenci.

- **Chromozom: 6**
- **Počet exonů: 4**

FEATURES

```
source
1..1678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.33"
gene
1..1678
/gene="TNF"
/gene_synonym="DIF; TNF-alpha; TNFA; TNFSF2; TNLG1F"
/note="tumor necrosis factor"
/db_xref="GeneID:7124"
/db_xref="HGNC:HGNC:11892"
/db_xref="MIM:191160"
exon
1 363
```

exon

Feature

1 of 4

NM_000594 : 1 segment

- **2. exon:**
>NM_000594.4:364-409 Homo sapiens tumor necrosis factor (TNF), mRNA
TTCCCAGGGACCTCTCTAATCAGCCCTCTGGCCCAGGCAGTCA
- **počet bp: 46** (409-364+1)

4) Obsahuje 2. exon nějaké „významné“ SNP (GMAF > 0,01)?

- **pro daný protein není tato možnost zobrazení k dispozici**

5) Kolik obsahuje 2. exon cytosinů? Kolik dinukleotidů CG?

Využití programu SMS/DNA Stats – vkládám FASTA formát sekvence 2. exonu

DNA Stats results

Results for 73 residue sequence "Untitled" starting "NMHmsanstu"

Pattern:	Times found:	Percentage:
g	9	12.33
a	11	15.07
t	13	17.81
c	21	28.77
n	5	6.85
u	1	1.37
r	4	5.48
y	0	0.00
s	4	5.48
w	0	0.00
k	0	0.00
m	4	5.48
b	0	0.00
d	0	0.00

tg	1	1.39
ta	1	1.39
tt	1	1.39
tc	7	9.72
tn	1	1.39
cg	0	0.00
ca	5	6.94
ct	7	9.72
cc	8	11.11
cn	0	0.00
ng	0	0.00
na	1	1.39
nt	0	0.00
nc	1	1.39
nn	0	0.00

2. exon obsahuje celkem 21 cytosinů (C) a žádný CG dinukleotid.

DÚ6: Nukleotidová bioinformatika II

1) Přeložte „vaši“ nukleotidovou sekvenci; ve kterém čtecím rámci se otevírá?

Překlad nukleotidové sekvence

Jedná se o překlad nukleotidové sekvence mRNA v aminokyselinovou podle genetického kódu. Volím celou sekvenci mRNA, nikoliv pouze CDS oblast. Použit program SMS/Translate.

Sequence Manipulation Suite:

Translate

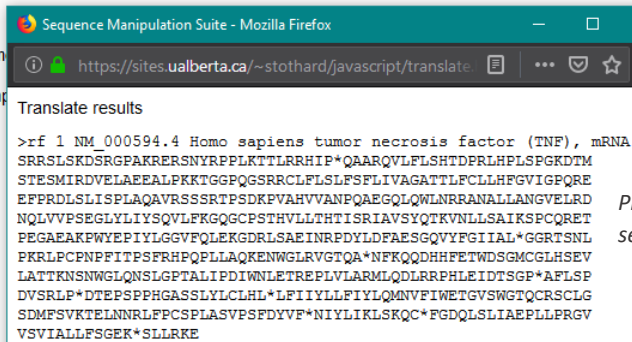
Translate accepts a DNA sequence and converts it into a protein in the reading frame

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

```
>NM_000594.4 Homo sapiens tumor necrosis factor (TNF), mRNA
AGCAGACGCTCCCTCAGCAAGGACAGCAGAGGACACGCTAAGAGGGAGAGAAGCAACTACA
GACCCCTCC
TGAAAACAACCCCTCAGACGCCACATCCCTGACAAGCTGCCAGGCAGGTTCTCTCCCTCTC
ACATACTGA
CCCACGGCTCCACCCCTCTCTCCCTGAAAAGGACACCATGAGCACTGAAAGCATGATCCG
GACGTGGAG
```

Submit Clear Reset

- Translate in on the strand.
- Use the genetic code.



Přeložená sekvence

Zjištění čtecího rámce

Pro ověření/zjištění čtecího rámce použít program NCBI/ORFfinder.

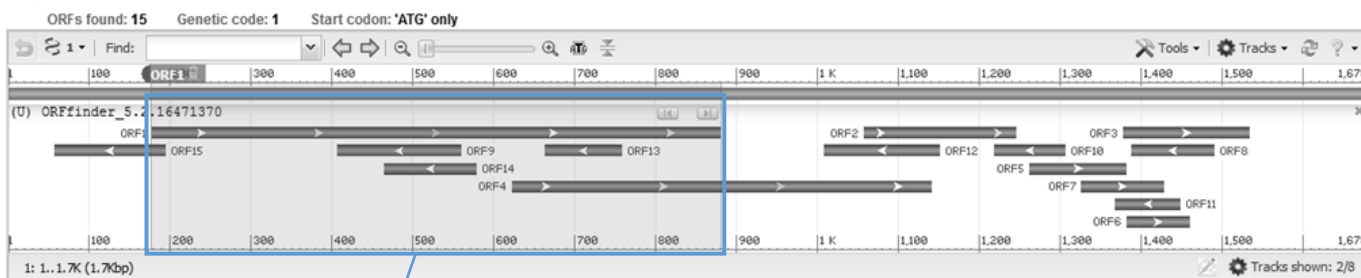
ORF = open reading frame = čtecí rámeček

Vkládán FASTA formát nukleotidové sekvence.

Open Reading Frame Viewer

Help

Sequence



```
ORF1 (233 aa) Display ORF as... Mark
>1c1|ORF1
MSTESMIRDVELAEALPKKTKGGPQGSRRCLFSLFSLFVAGATILFCL
LHFGVIGPQREFFRDLISPLAQAVSSSRTPSDKPVHVVANPQAEQ
QLQWLNRRANALLANGVELRDNQLVVPSEGLYLYSQTVEKQGCPSHVL
LLHTISRIAVSYQTKVNLNLSAIKSPCQRETPEGAEAKWYEPYLGCVF
QLEKGDRLSAEINRPDYLDFAESGQVYFGIIAL
```

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF1	+	1	178	879	702 233
ORF4	+	2	623	1141	519 172
ORF2	+	1	1057	1245	189 62
ORF3	+	1	1378	1533	156 51
ORF9	-	1	559	407	153 50
ORF12	-	3	1151	1008	144 47
ORF15	-	3	194	57	138 45
ORF5	+	2	1262	1381	120 39
ORF14	-	3	578	465	114 37

Nejdelší úsek (oblast 178 – 879; 233 aa) se otevírá ve čtecím rámci +1 (ORF1+), tzn., že se jedná kódující vláknou dvoušroubovice, resp. danou mRNA

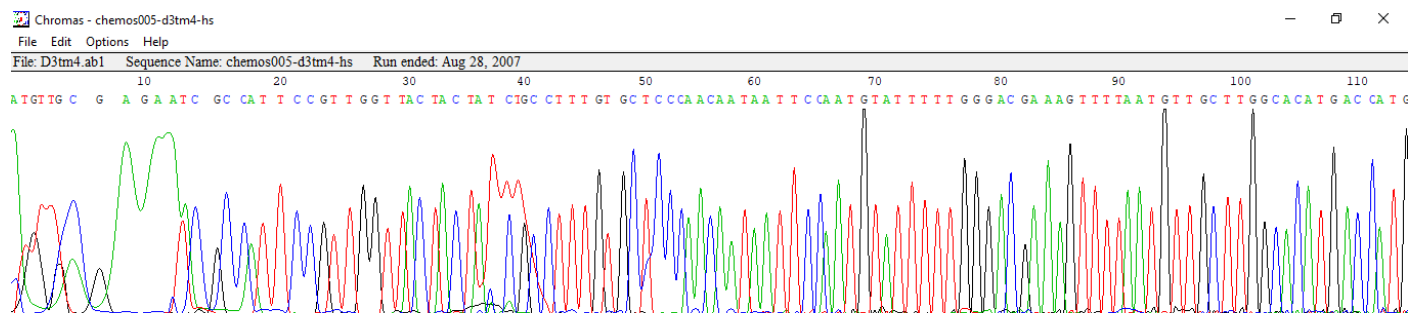
2) Porovnejte celou a CDS „vaši“ nukleotidové sekvence.

Myšleno porovnat celou mRNA sekvence TNF a CDS oblast mRNA sekvence TNF; porovnání v programu MultAlin

NM_000594.4 – celá mRNA sekvence, NM_000594.4:170-879 – CDS sekvence

NM_000594.4	1	10	20	30	40	50	60	70	80	90	100	110	120	130
NH_000594.4:178-879	AGCAGACGCTCCCTCAGCAGGACAGCAGAGGACCCAGCTAGAGGGAGAGGACACTACAGACCCCTCGAARACACCTCAGACGCCACATCCCTGACAGCTGCCAGGACGGTCTCTCTCTCT													
Consensus													
NM_000594.4	131	140	150	160	170	180	190	200	210	220	230	240	250	260
NH_000594.4:178-879	CNCATCTACTGACCCAGGCTCACCCTCTCTCCCTGGARAGGACACCATGAGCAC TGARAGCATGATCCGGACGTGAGCTGCCCCAGAGGACGCTCCCCAGAGACAGGGGGCCCCAGGCTCCAG													
ConsensusATGAGCATGAGAGCATGATCCGGACGTGAGCTGCCCCAGAGGACGCTCCCCAGAGACAGGGGGCCCCAGGCTCCAG													
NM_000594.4	261	270	280	290	300	310	320	330	340	350	360	370	380	390
NH_000594.4:178-879	GCGGTGCTGTGTTCTGAGCTCTTCTCTCTGATCTGAGGACGCTGACAGCTCTCTGCTGCTGACCTTTGAGGATGATCGGCCCCAGAGGAGAGATTCGCCAGGACCTCTCTTAATGAGC													
ConsensusGCGGTGCTGTGTTCTGAGCTCTTCTCTCTGATCTGAGGACGCTGACAGCTCTCTGCTGCTGACCTTTGAGGATGATCGGCCCCAGAGGAGAGATTCGCCAGGACCTCTCTTAATGAGC													
NM_000594.4	391	400	410	420	430	440	450	460	470	480	490	500	510	520
NH_000594.4:178-879	CTCTGCCCCAGGACGTCAGATCATCTCTCGAACCCGAGTGACAGCTGATGCCCATGTTGTAGCAACCTCAGCTGAGGGGACGCTCCAGTGGCTGAAACCCGGGCCAATGCCCCGGCCCA													
ConsensusCTCTGCCCCAGGACGTCAGATCATCTCTCGAACCCGAGTGACAGCTGATGCCCATGTTGTAGCAACCTCAGCTGAGGGGACGCTCCAGTGGCTGAAACCCGGGCCAATGCCCCGGCCCA													
NM_000594.4	521	530	540	550	560	570	580	590	600	610	620	630	640	650
NH_000594.4:178-879	ATGCGGTGAGGCTGAGAGATACCCAGCTGGTGGTCCATCAGAGGGCTGACCTCATCTACCTCCAGCTCTCTCAGGGCCAGGCTGCCCTCCACCATGCTCTCCACACACCATCAGCCG													
ConsensusATGCGGTGAGGCTGAGAGATACCCAGCTGGTGGTCCATCAGAGGGCTGACCTCATCTACCTCCAGCTCTCTCAGGGCCAGGCTGCCCTCCACCATGCTCTCCACACACCATCAGCCG													
NM_000594.4	651	660	670	680	690	700	710	720	730	740	750	760	770	780
NH_000594.4:178-879	CATGCCCTCCCTACCGACCAAGGTCACCTCTCTGCCATCAGAGCCCTGCCAGAGGAGACCCAGAGGGGGCTGAGGCCAGGCTGGATGAGCCCATATCTGGAGGGGCTCTCCAG													
ConsensusCATGCCCTCCCTACCGACCAAGGTCACCTCTCTGCCATCAGAGCCCTGCCAGAGGAGACCCAGAGGGGGCTGAGGCCAGGCTGGATGAGCCCATATCTGGAGGGGCTCTCCAG													
NM_000594.4	781	790	800	810	820	830	840	850	860	870	880	890	900	910
NH_000594.4:178-879	CTGGAGAGGGGTGACCGACTCAGCGCTGAGATCARTCGCCCGACTATCTCGACTTTGCCAGTCTGGGAGGCTACTTTGGGATCATTGCCCTGAGGAGGACACATCCACCTCCCAACGCC													
ConsensusCTGGAGAGGGGTGACCGACTCAGCGCTGAGATCARTCGCCCGACTATCTCGACTTTGCCAGTCTGGGAGGCTACTTTGGGATCATTGCCCTGAGGAGGACACATCCACCTCCCAACGCC													
NM_000594.4	911	920	930	940	950	960	970	980	990	1000	1010	1020	1030	1040
NH_000594.4:178-879	TCCCTGCCCCAATCCCTTTATACCCCTCTCTCAGACCCCTCAGCTCTCTGCTCAGAGAGAGATGGGGCTTAGGGTCGAGACCCAGCTTAGAGCTTAAAGCAACAGACCCACCTTCGA													
ConsensusTCCCTGCCCCAATCCCTTTATACCCCTCTCTCAGACCCCTCAGCTCTCTGCTCAGAGAGAGATGGGGCTTAGGGTCGAGACCCAGCTTAGAGCTTAAAGCAACAGACCCACCTTCGA													
NM_000594.4	1041	1050	1060	1070	1080	1090	1100	1110	1120	1130	1140	1150	1160	1170
NH_000594.4:178-879	AACCTGGGATTCAGGATGTGTGGCTGCACAGTGAAGTGTGCGACCACTAAGATTCACACTGGGGCTCCAGACTCACTGGGGCTCAGCTTTGATCCCTGACATCTGGAATCTGAGACCAAG													
ConsensusAACCTGGGATTCAGGATGTGTGGCTGCACAGTGAAGTGTGCGACCACTAAGATTCACACTGGGGCTCCAGACTCACTGGGGCTCAGCTTTGATCCCTGACATCTGGAATCTGAGACCAAG													
NM_000594.4	1171	1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
NH_000594.4:178-879	GAGCCTTTGGTCTGGCCAGATGCTGCGAGACTGAGAGAGACCTCAGCTAGAAATGACACAGTGGACTTAGGCCCTCTCTCCAGATGTTCCAGACTCTCTGAGACACAGGACCCAGCCCTC													
ConsensusGAGCCTTTGGTCTGGCCAGATGCTGCGAGACTGAGAGAGACCTCAGCTAGAAATGACACAGTGGACTTAGGCCCTCTCTCCAGATGTTCCAGACTCTCTGAGACACAGGACCCAGCCCTC													
NM_000594.4	1301	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430
NH_000594.4:178-879	CCCATGAGCCAGCTCCCTCTATTATGTTGACCTGTGATATTATTATTATTATTATTATTATTATTATTACAGATGATGATTTATTGGGAGCCGGGTACTCTGGGGACCAATGTAGGAG													
ConsensusCCCATGAGCCAGCTCCCTCTATTATGTTGACCTGTGATATTATTATTATTATTATTATTATTATTATTATTACAGATGATGATTTATTGGGAGCCGGGTACTCTGGGGACCAATGTAGGAG													
NM_000594.4	1431	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560
NH_000594.4:178-879	CTGCTTGGCTCAGCATGTTTCCGTAAGACGGAGCTGACCAATAGGCTGTTCCCATGAGCCCTGGCCCTCTGCTCTTTTGTATGTTTTTAAATATTATCTGATTAGTGTCTAAA													
ConsensusCTGCTTGGCTCAGCATGTTTCCGTAAGACGGAGCTGACCAATAGGCTGTTCCCATGAGCCCTGGCCCTCTGCTCTTTTGTATGTTTTTAAATATTATCTGATTAGTGTCTAAA													
NM_000594.4	1561	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1679	
NH_000594.4:178-879	CAATGCTGATTTGGTACCACTGCTCACTATTGCTGAGCCTGCTCCCCAGGGAGTGTGCTGTGATCGCCCTACTATCAGTGGCAGAAATAGATTTGCTAGAAAGAAA													
ConsensusCAATGCTGATTTGGTACCACTGCTCACTATTGCTGAGCCTGCTCCCCAGGGAGTGTGCTGTGATCGCCCTACTATCAGTGGCAGAAATAGATTTGCTAGAAAGAAA													

3) Stáhněte si neznámou sekvenci v úkolu a otevřete ji programem chromas.



FASTA formát neznámé sekvence:

>D3tm4 sequence exported from chromatogram file

```

ATGTTGCGGAGAATCGCCATTCGGTGGTTACTACTATCTGCCTTTGTGCTCCCAACAATAATTCCAATGTATTTTTGGGACGAAAGTTTTAATGTTGCTTGGCACA
TGACCATGCTGAAATATATCTTCGGTCTGAACGCTGCATTTCTAGTCAACAGCGTTGCTCATATGTGGGGGTATAAACCTTATGACAAGAACATAGCTCCAACAC
AGAATTATATAGCAACATTGCGCACATTAGGCGAAGGTTTCCACAACACTACCATCACGCATTTCTTGGGATTATCGTGCCTCGGAACCTGGAAACAATTATCTTAA
CTTAACGACCAAGTTCATAGATTTCTCGCTGGATTGGCTGGGCTTATGATTTGAAAACGGTTCTGAAGATCTGCTTCAGAAAAGGATGGAGAGAAGCTGGTG
ACGGTACCAATCTTTGGGGGAGGGGTGATAAGAACATGAAAAAGATTATGTGAAATCTACAGATGTCCATGAATGACTCGAGCATGCATCTAGAGGGCCGC
ATCATGTAATTAGTTATGTCAGCTTACATTACGCCCTCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTA
    
```

4) Zkontrolujte přítomnost „vektorů“ a „očistěnou“ sekvenci identifikujte

Z chromasů použijte FASTA formát dané sekvence (viz výše) – projedte programem VecScreen – zjistíte kontaminaci vektorem

BLAST » vector contamination » RID-9ZDMAPHZ014

BLAST Results

Vecscreen

Job title: D3tm4 sequence exported from chromatogram

RID: 9ZDMAPHZ014 (Expires on 04-01 02:33 am)

Query ID: Id|Query_72425
 Description: D3tm4 sequence exported from chromatogram file
 Molecule type: nucleic acid
 Query Length: 615

Database Name: screen/UniVec
 Description: UniVec (build 10.0)
 Program: BLASTN 2.9.0+ » Citation

Other reports: Search Summary | Taxonomy reports | Distance tree of results | MSA viewer

Graphic Summary

Distribution of Vector Matches on the Query Sequence

Match to Vector: ■ Strong ■ Moderate ■ Weak
 Segment of suspect origin: ■

Segments matching vector:
 Strong match: 511-615
 Weak match: 499-510

Alignments

gn|uv|U30497.1:3230-4814 Cloning vector pAS2-1
 Sequence ID: Length: 1585 Number of Matches: 1

Range	Score	Expect	Identities	Gaps	Strand
242 to 333	184 bits(92)	4e-44	92/92(100%)	0/92(0%)	Plus/Minus

gn|uv|AF102577.1:1-176 Cloning vector pMyrCAM
 Sequence ID: Length: 176 Number of Matches: 1

Range	Score	Expect	Identities	Gaps	Strand
115 to 176	124 bits(62)	5e-26	62/62(100%)	0/62(0%)	Plus/Plus

gn|uv|Y10545.1:345-446 Cloning vector pZeRO-2T
 Sequence ID: Length: 102 Number of Matches: 1

Range	Score	Expect	Identities	Gaps	Strand
14 to 36	46.5 bits(23)	0.017	23/23(100%)	0/23(0%)	Plus/Plus

Oblast silné shody – *suspektní kontaminace vektorem v oblasti 511 – 615*

3 možné shody (vektory)

Sequence Manipulation Suite:

Range Extractor DNA

Range Extractor DNA accepts a DNA sequence along with a set of positions or ranges. The bases corresponding to the positions or ranges are extracted from the sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using

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

```
GGCGAAGGTTTCACAACACTACATCACGCAATTCCTTGGGAAATTCGTGCTCGGAACCT
GSAACAATTAATCTTAACCTACAGCCAGGTCATAGATTCCTGCTCGGATTCGGCTGG
GCTTATGATTTGAAAGGCTCGTGAAGATCTGCTCAGAAAAGGATGGAGAACTGGT
GACGATCCACATCTTTGGGGAGGGGTGATAAGACATGAAAAGGATTAATGTAAATCT
ACAGATGTCATGAATGACTCGAGCATGCATCTAGAGGCGCGCATCATGTAATAGTTAT
GTCACGCTTACATTCACGCGCTCCCCCCACATCCGCTCTAACCGAAAAGGAGGTTAG
ACACCTGAAGTCTA
```

Enter the base positions or ranges to be extracted. Use "." to represent a range, and use a comma to separate entries. The words "start", "end", "center", and "length" can be used to represent the beginning, end, center, and length of a range, respectively. The range "(end - 2).end" can be used to represent the last two bases of a range. The range "(center + 1).(center + 30)" can be used to represent a 30-base window centered on a specific position.

1..615

- Obtain bases from the direct
- Sequence segments should be returned in FASTA format

*This page requires JavaScript. See below for instructions on how to use it off-line.

Mon Nov 6 02:58:29 2017
 Valid XHTML 1.0, Valid CSS

Dalším krokem je očištění sekvence pomocí SMS/Range extraktor DNA

Získám FASTA formát očištěný od kontaminace

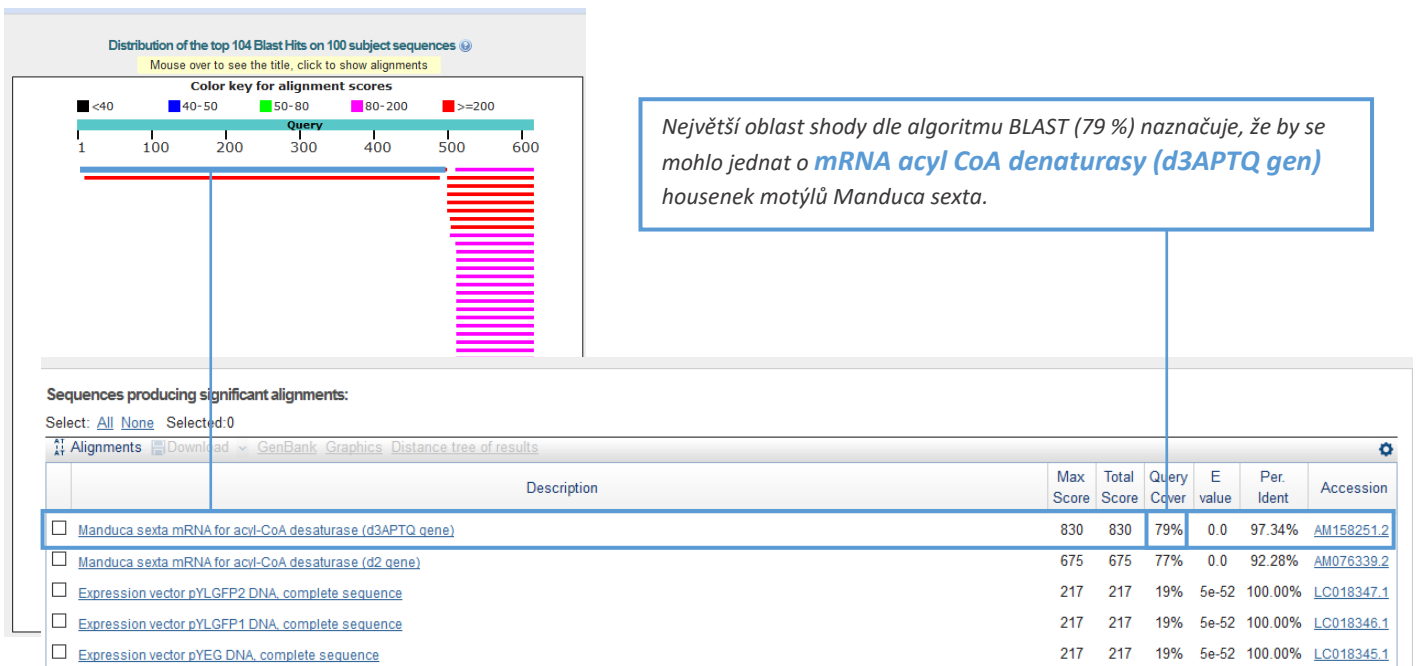
www.bioinformatics.org/sms2/range_extract_dna.html

Range Extractor DNA results

Results for 615-residue sequence "D3tm4 sequence exported from chromatogram file" starting "ATGT"

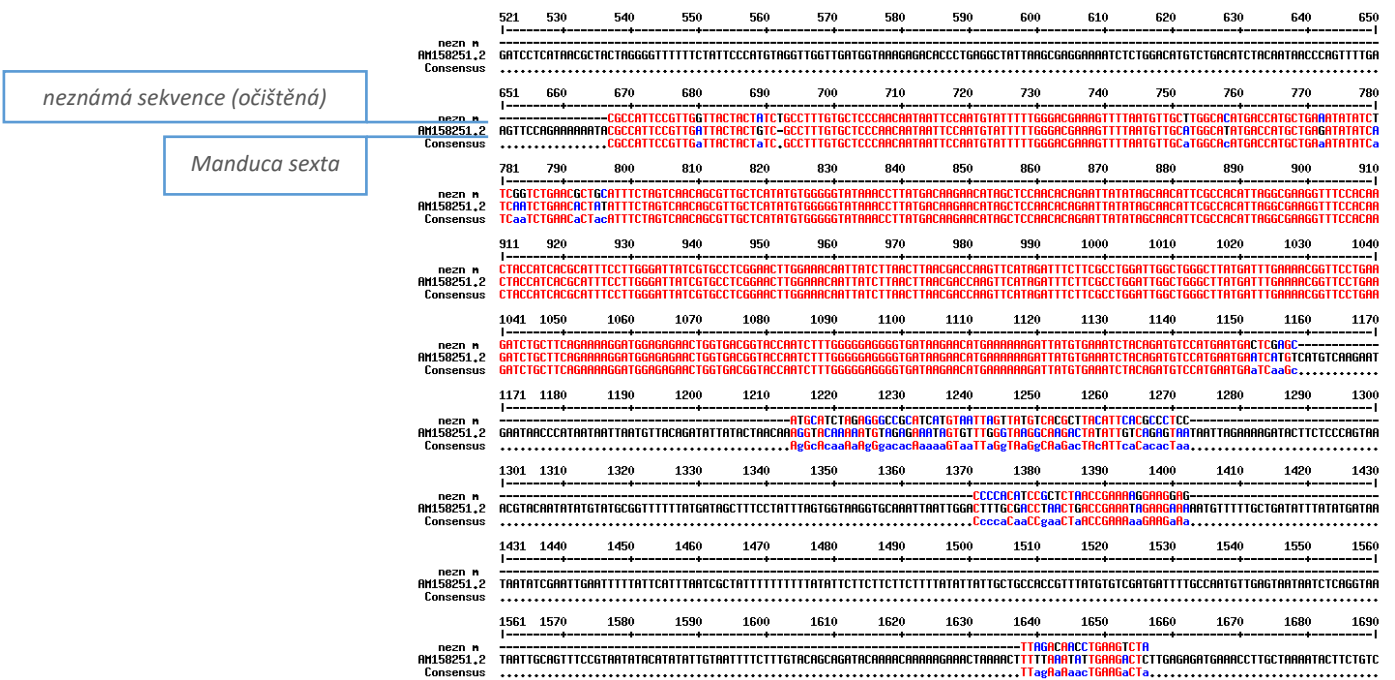
```
ATGTTGGGAAATCCCAATTCCTGGTACTACTATCGCTTTGTGCTCCCAACAATA
ATTCCAATGATTTTTGGGAGAAAGTTTAAATGTTGCTTGGCAGATGACCAATGCTGAAA
TATATCTGGGCTGAAAGCTGCAATTTCTAGTCAACAGGCTTGCATATGTTGGGGTAT
AAACCTTATGACAAGAAGATAGTCCCAACAGAAATATATAGCAACATTCGCCACATTA
GGCGAAGTTTCACAACACTACATCACGCAATTCCTTGGGAAATTCGTGCTCGGAACCT
GSAACAATTAATCTTAACCTACAGCCAGGTCATAGATTCCTGCTCGGATTCGGCTGG
GCTTATGATTTGAAAGGCTCGTGAAGATCTGCTCAGAAAAGGATGGAGAACTGGT
GACGATCCACATCTTTGGGGAGGGGTGATAAGACATGAAAAGGATTAATGTAAATCT
ACAGATGTCATGAATGACTCGAGCATGCATCTAGAGGCGCGCATCATGTAATAGTTAT
GTCACGCTTACATTCACGCGCTCCCCCCACATCCGCTCTAACCGAAAAGGAGGTTAG
ACACCTGAAGTCTA
```


To dál znamená vzít očištěný FASTA formát, projít ho BLASTem, najít oblast největší shody a získat její FASTA formát.



Největší oblast shody dle algoritmu BLAST (79 %) naznačuje, že by se mohlo jednat o **mRNA acyl CoA denaturasy (d3APTQ gen)** housenek motýlů *Manduca sexta*.

Dál provedu srovnání očištěného FASTA formátu neznámé sekvence získané z Chromasu a nejpravděpodobnější sekvenci z BLASTu (tedy mRNA acetyl-CoA denaturasy u housenek *Manduca sexta*).



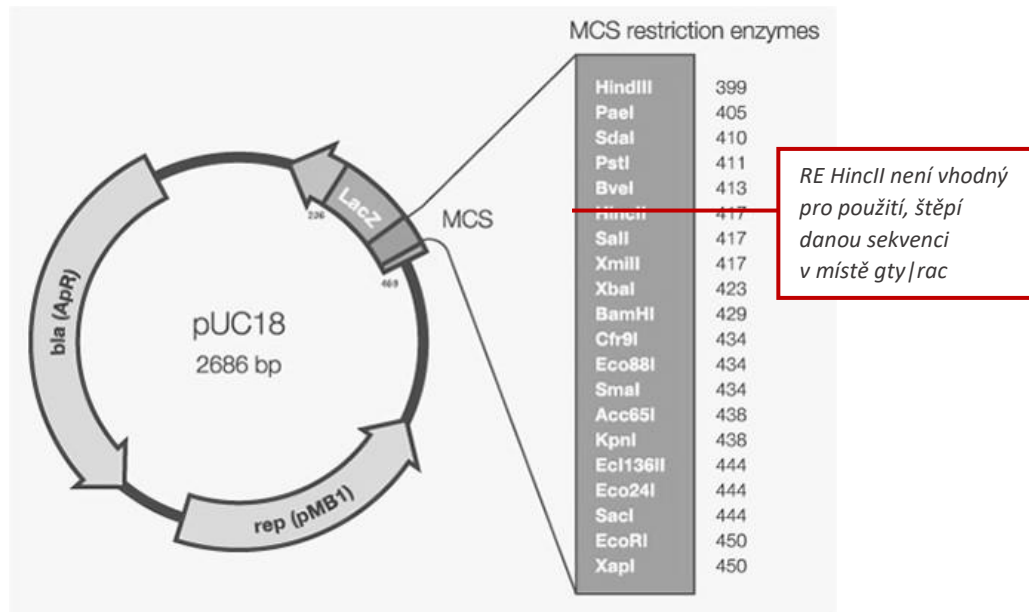
DÚ7: Primery a RE analýza

1) Zjistěte, zda a kolikrát štěpí celou sekvenci vašeho genu enzymy: *XhoI*, *BamHI*, *SacI*

(vezmu mRNA – její FASTA formát vložím do SMS/RestrictionSummary a vyberu potřebné RE)

- *XhoI* (c|tcgag) – neštěpí
- *BamHI* (g|gatcc) – neštěpí
- *SacI* (ccgc|gg) – neštěpí

2) Navrhněte primery tak, aby se vaše kódující sekvence (CDS) dala vložit do plasmidu pUC18



Zkontrolujte, které RE budou vhodné z nabídky MCS (nebude-li možné vaši sekvenci vložit do tohoto plasmidu, navrhněte primery pro kódující oblast NQO1)

Restrikční enzymy MCS vhodné k použití: *HindIII* (a|agctt), *PstI* (ctgca|g), *Sall* (g|tcgac), *XbaI* (t|ctaga), *BamHI* (g|gatcc), *SmaI* (ccc|ggg), *KpnI* (ggtac|c), *SacI* (gagct|c), *EcoRI* (g|aatc)

Navrhněte F a R primer tak aby T_m nebyla větší než 65°C

(primery můžou navrhnout a jejich vlastnosti porovnat přes OligoCalc nebo Primer Blast)

- **Forward primer:** ATG AGC ACT GAA AGC ATG ATC ($T_m = 57,5^\circ\text{C}$)

Nucleotide base codes	Melting Temperature (T_m) Calculations	
ATG AGC ACT GAA AGC ATG ATC	1	50.5 °C (Basic)
Reverse Complement Strand(5' to 3') is:	2	57.5 °C (Salt Adjusted)
GAT CAT GCT TTC AGT GCT CAT	3	52.26 °C (Nearest Neighbor)

- **Reverse primer:** TCA CAG GGC AAT GAT CCC AA ($T_m = 58,4^\circ\text{C}$)

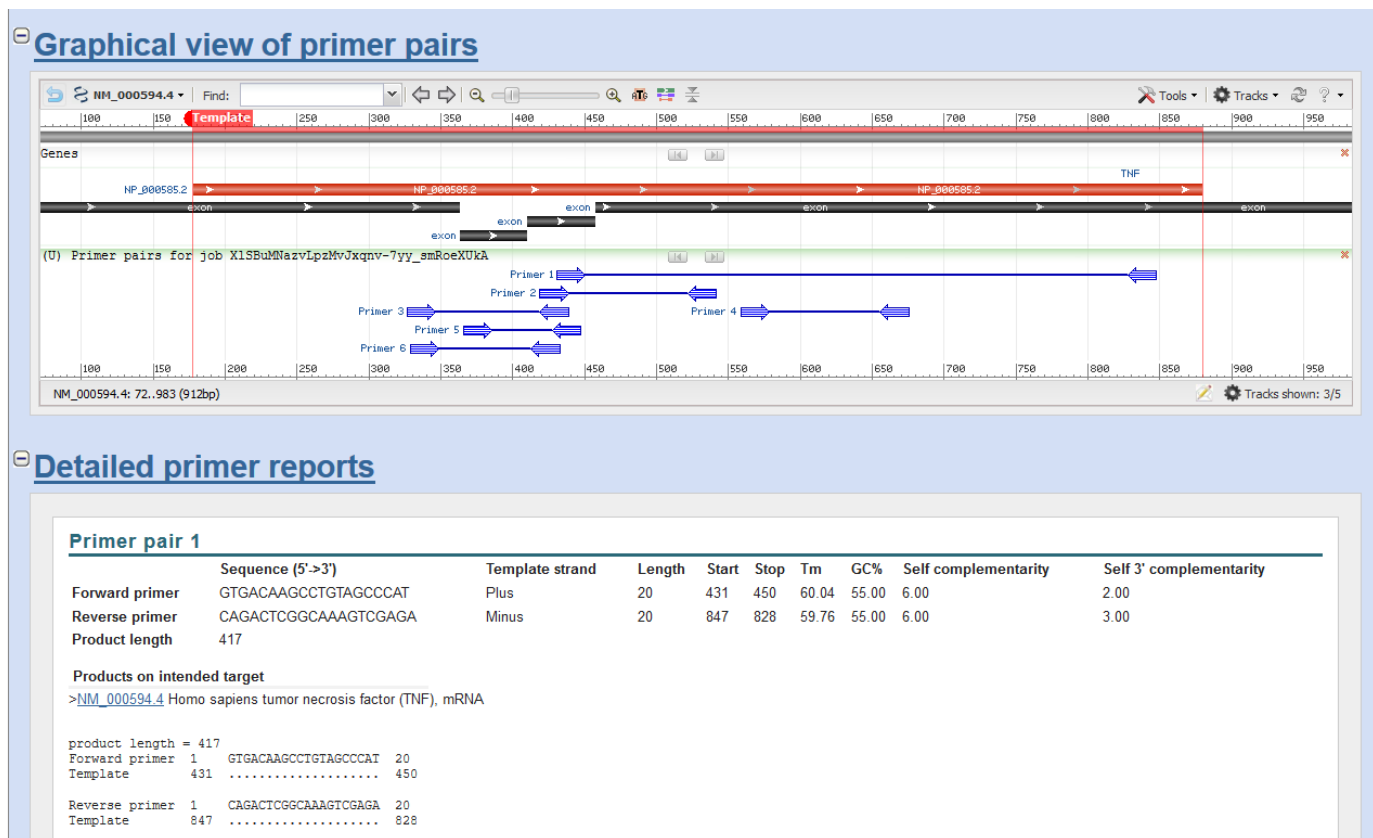
Nucleotide base codes	Melting Temperature (T_m) Calculations	
TTG GGA TCA TTG CCC TGT GA	1	51.8 °C (Basic)
Reverse Complement Strand(5' to 3') is:	2	58.4 °C (Salt Adjusted)
TCA CAG GGC AAT GAT CCC AA	3	52.38 °C (Nearest Neighbor)

Pokud bychom chtěli proces dokončit, je třeba k danému primeru vybrat restriční enzym z nabízených MCS. Pro forward primer vybírám ze spodní části nabídky, pro reverse primer z části horní:

Primery – finální část:

- **Forward primer:** ATG AGC ACT GAA AGC ATG ATC
 - zvolený RE: EcoRI (štěpí v místě g|aattc)
 - variabilní nukleotidy: př. AAA
 - celá sekvence: AAA GAA TTC ATG AGC ACT GAA AGC ATG ATC ($T_m = 66,7^\circ\text{C}$)
 - pokud bychom se chtěli dostat pod teplotu $T_m = 65^\circ\text{C}$, pak je potřeba ubrat množství nukleotidů
 - daná sekvence primeru určená pro vklad do plasmidu: AAA GAA TTC ATG AGC ACT GAA AGC ATG AT ($T_m = 64,6^\circ\text{C}$)
- **Reverse primer:** TCA CAG GGC AAT GAT CCC AA
 - zvolený RE: HindIII (štěpí v místě a|aacctt)
 - variabilní nukleotidy: AAA
 - celá sekvence: AAA AAG CTT TCA CAG GGC AAT GAT CCC AA ($T_m = 67,4^\circ\text{C}$)
 - daná sekvence primeru určená pro vklad do plasmidu: AAA AAG CTT TCA CAG GGC AAT GAT CC ($T_m = 64,6^\circ\text{C}$)

Pokud bychom tvořili primery skrze Primer Blast, nabízeloby se několik možností primerů. V tomto případě možnosti TNF se nabízí 6 párů primerů. U zvolených primerů lze vidět jednotlivé charakteristiky viz obrázek.



c) Primer-BLAST – kontrola specifity vybraných primerů

Vybrané primery **nejsoú specifické** – mohly by zahrnovat amplifikaci i dalších variant, jako například: *Homo sapiens DnaJ (Hsp40 member)*, *transmembrane protein 63C (TMEM63C)*, *sprouty related EVH1 domain containing 1 (SPRED1)* a spoustu dalších.



Detailed primer reports

Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GCTGCCCTCCACCC	Plus	15	608	622	58.78	80.00	3.00	0.00
Reverse primer	TCTGGTAGGAGACGGCG	Minus	17	670	654	57.61	64.71	4.00	3.00
Product length	63								

Products on intended target

>NM_000594.4 Homo sapiens tumor necrosis factor (TNF), mRNA

```

product length = 63
Forward primer 1 GCTGCCCTCCACCC 15
Template       608 ..... 622
Reverse primer 1 TCTGGTAGGAGACGGCG 17
Template       670 ..... 654
    
```

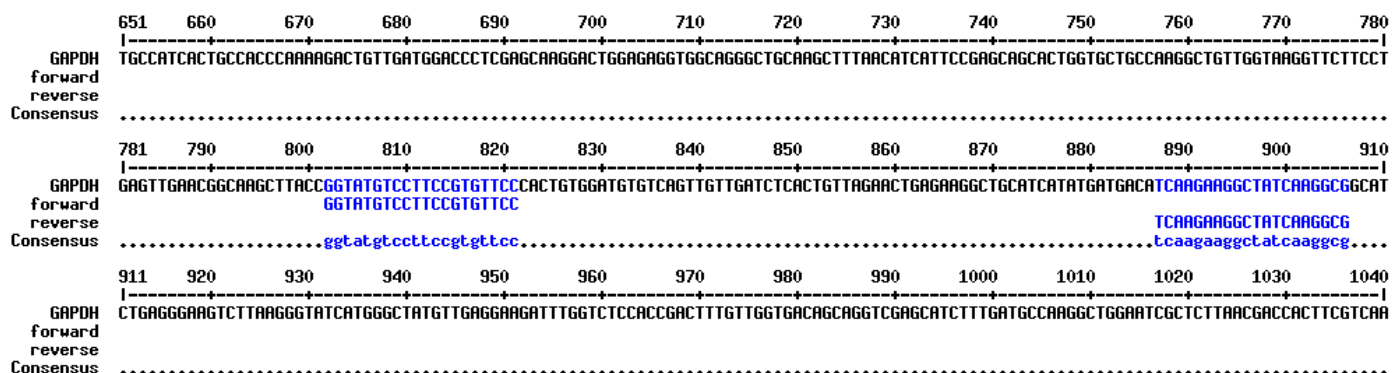
2) Zkontrolujte primery pro GAPDH z publikace (sekvence k nalezení v Supporting information 1 - u článku): MultAlin, Primer3, mFOLD, specifita?

Potřebné informace:

- sekvence:
- primery (zadáni v 5'-3' směru):
 - forward: GGTATGTCCTTCCGTGTTCC
 - reverse: AGTTCCTCCGATAGTTCGCG (komplementárně původní řetězec: TCAAGAAGGCTATCAAGGCG)

a) MultAlin: srovnání organismu a primerů – jestli jsou opravdu vhodné pro použití v dané sekvenci, kde primery začínají

Primery **odpovídají dané sekvenci**; forward primer: 802-821, reverse primer: 887-906



DÚ9: Návrh primerů

1) Navrhnete mutační primery pro mutaci 100. aminokyseliny "vašeho" proteinu:

Identifikujte 100. AMK (X), запиšte plánovanou mutaci ve tvaru: X100G

Identifikujte příslušný kodon

Pro identifikaci 100. AMK (X) využívám program SMS Range Extractor Protein. Vkládám FASTA formát sekvence „svého“ proteinu a zadávám hodnotu 100 (jako 100. aminokyselina).

Po provedení identifikace bylo zjištěno, že 100. AMK je Glycin. Jelikož ale chceme simulovat mutaci právě jako záměnu za Glycin, je třeba zvolit jinou než 100. AMK.
Určena 10. AMK

Sequence Manipulation Suite:

Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or ranges. The residues corresponding to the positions or ranges are in a sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor Protein to obtain subsequences using position information.

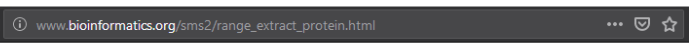
Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 500,000,000 characters.

```
>AQY77150.1 TNF [Homo sapiens]
MSTESMIRDVLAEEALPKKTKGGPQSSRRCLFSLFSLIVAGATTLFCLLHFGVIGPQRE
EFPRLSLI
SPLAQAVRSSSRTPSDKPVAVHVVANPQAEGLQWLNRANALLANGVELRDNLVVPSEGL
YLIYSQVLF
KGGQCPSTHLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPYILGGVE
LEKGDRLSA
```

Enter the residue positions or ranges to be extracted. Use "." to represent a range, and use a comma to separate entries. The words 'start', 'end', 'center', 'center + 1', '(center - 2)', 'center', '(center + 1)', '(center + 30)' can

100

Submit Clear Reset



Range Extractor Protein results

- Sequence segments should be extracted for 233 residue sequence "AQY77150.1 TNF [Homo sapiens]" starting "MSTESMIRDV"

Sequence Manipulation Suite:

Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or ranges. The residues corresponding to the positions or ranges are in a sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor Protein to obtain subsequences using position information.

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

```
>AQY77150.1 TNF [Homo sapiens]
MSTESMIRDVLAEEALPKKTKGGPQSSRRCLFSLFSLIVAGATTLFCLLHFGVIGPQRE
EFPRLSLI
SPLAQAVRSSSRTPSDKPVAVHVVANPQAEGLQWLNRANALLANGVELRDNLVVPSEGL
YLIYSQVLF
KGGQCPSTHLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPYILGGVE
LEKGDRLSA
```

Enter the residue positions or ranges to be extracted. Use "." to represent a range, and use a comma to separate entries. The words 'start', 'end', 'center', 'center + 1', '(center - 2)', 'center', '(center + 1)', '(center + 30)' can

10

Submit Clear Reset



Range Extractor Protein results

- Sequence segments should be extracted for 233 residue sequence "AQY77150.1 TNF [Homo sapiens]" starting "MSTESMIRDV"

10. AMK je Valin, lze tedy simulovat mutaci.

10. AMK = Valin

plánovaná mutace: V10G

V další fázi budeme identifikovat příslušný kodon. Využit opět SMS Range Extractor – ale DNA. Dosazují FASTA formát CDS sekvence mRNA. Pokud bych identifikovala 100. aminokyselinu, volila bych rozsah nukleotidové sekvence 298..300. Jelikož mě ale zajímá 10. aminokyselina, rozsah upraven na 28..30.

Musí se shodovat kodon s produktem, tedy výslednou aminokyselinou z předchozího kroku. Ověření správnosti například s pomocí této tabulky.

		SECOND BASE						
FIRST BASE	UUU	Phenylalanine	UUC	Serine	UAU	Tyrosine	UGU	Cysteine
	UUA	Leucine	UCA	Stop codon	UAA	Stop codon	UGA	Stop codon
	UUG	Leucine	UGG	Stop codon	UAG	Stop codon	UGG	Tryptophan
	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine
	CUC	Leucine	CCC	Proline	CAC	Glutamate	CGC	Arginine
	CUA	Leucine	CCA	Proline	CAA	Glutamate	CGA	Arginine
	CUG	Leucine	CCG	Proline	CAG	Glutamate	CGG	Arginine
	AUU	Isoleucine	ACU	Threonine	AUU	Asparagine	AGU	Serine
	AUC	Isoleucine	ACC	Threonine	AUA	Asparagine	AGC	Serine
	AUA	Methionine start codon	ACA	Threonine	AAG	Lysine	AGA	Arginine
	AUG	Methionine start codon	ACG	Threonine	AAG	Lysine	AGG	Arginine
	GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine
	GUC	Valine	GCC	Alanine	GAC	Aspartic acid	GGC	Glycine
	GUA	Valine	GCA	Alanine	GAA	Glutamic acid	GGA	Glycine
	GUG	Valine	GCG	Alanine	GAG	Glutamic acid	GGG	Glycine

Sequence Manipulation Suite:

Range Extractor DNA

Range Extractor DNA accepts a DNA sequence along with a set of positions or ranges. The bases corresponding to the positions or ranges are in a sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using position information.

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

```
>NM_000594.4:178-879 Homo sapiens tumor necrosis factor (TNF), mRNA
ATGAGCCTGAAAGCATGATCCGGGACGTGGAGCTGGCCGAGGAGCGCTCCCCAAGAAG
CAGGGGGGC
CCGAGGGCTCCAGGGCGGCTTGTTCCTCAGCCTCTTCTCCTCCTGATCGTGGCAGGCG
CACCACGCT
CTTCTGCTGCTGCACTTTGGAGTGTATCGGCCCCAGAGGGAAGATTCCCCAGGGACCT
```

Enter the base positions or ranges to be extracted. Use "." to represent a range, and use a comma to separate entries. The words 'start', 'end', 'center', 'center + 1', '(center - 2)', 'center', '(center + 1)', '(center + 30)' can

28..30

Submit Clear Reset



Range Extractor DNA results

- Obtain bases from the [] results for 702 residue sequence "NM_000594.4:178-879 Homo sapiens tumor necros
- Sequence segments should be extracted for 702 residue sequence "NM_000594.4:178-879 Homo sapiens tumor necros

Výsledný kodon GTG odpovídá Valinu.

Mutace by se dala jiným způsobem zapsat rovněž: V(GTG)10G(GGC)

- navrhnete mutační primery se záměnou v glycin (GGC)

Pro návrh mutačních primerů vycházím z programu SMS Range Extractor DNA; vložení FASTA formátu CDS sekvence. Pro návrh mutačních primerů je nutno vybrat 5 kodonů před mutací a 5 kodonů za mutací, pro 10. aminokyselinu tak volím rozmezí 13..45.

Sequence Manipulation Suite:

Range Extractor DNA

Range Extractor DNA accepts a DNA sequence along with a set of positions or ranges. The bases corresponding to the positions or ranges are returned sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor DNA to obtain subsequences using position informatior

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

```
>NM_000594.4:178-879 Homo sapiens tumor necrosis factor (TNF), mRNA
ATGAGCACTGAAAGCATGATCCGGGACGTGGAGCTGGCCGAGGAGCGCTCCCAAGAAGA
CAGGGGGCC
CCCAAGGCTCCAGGCGGTGCTTCTCCTCAGCCTCTTCTCCTCCTGATCGTGGCAGGGCGC
CACCACGCT
CTTCTGCCTGCTGCACTTTGGAGTATCGGCCCCAGAGGGAAGAGTTCCCAAGGACCT
```

Enter the base positions or ranges to be extracted. Use "-" to represent a range, and use a comma to separate entries. The words 'start', 'end', 'center', 'center + 1', '(center + 1)', '(center + 30)' can be used to represent positions relative to the start of the sequence. The range '(end - 2, center + 1), (center + 30)' can be used to represent a range centered on the start of the sequence.

Sequence Manipulation Suite - Mozilla Firefox

www.bioinformatics.org/sms2/range_extract_dna.html

13..45

Submit Clear Reset

Range Extractor DNA results

>results for 702 residue sequence "NM_000594.4:178-879 Homo sapiens tumor necrosis factor (TNF)

AGCATGATCCGGGACGTGGAGCTGGCCGAGGAG

• Obtain bases from the

• Sequence segments sh

sekvence 5 kodonů před, GTG, 5 kodonů za

V následujícím kroku se vrátím k FASTA formátu CDS – zvolím Graphics; v políčku Find vyplním sekvenci z předchozího kroku (AGCATGATCCGGGACGTGGAGCTGGCCGAGGAG). Dalo by se označit za určitý kontrolní krok (vizualizace sekvence).

Homo sapiens tumor necrosis factor (TNF), mRNA

NCBI Reference Sequence: NM_000594.4

GenBank FASTA

Label	From	To	Strand
AGCATGATCCGGGACGTGGAGCTGGCCGAGGAG	190	222	Positive

sekvence

daný kodon pro mutaci

Následujícím krokem bude manuální mutace Valinu (V) za Glycin (G), respektive GTG → GGC.

Fyziologická sekvence: AGCATGATCCGGGACGTGGAGCTGGCCGAGGAG

Mutovaná sekvence: AGCATGATCCGGGACGGCGAGCTGGCCGAGGAG

Oligo Calc: Oligonucleotide Properties Calculator

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

Nucleotide base codes
AGC ATG ATC CGG GAC GGC GAG CTG GCC GAG GAG

Forward mutační primer

Reverse Complement Strand(5' to 3') is:
CTC CTC GGC CAG CTC GCC GTC CCG GAT CAT GCT

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

Nucleotide base codes
CTC CTC GGC CAG CTC GCC GTC CCG GAT CAT GCT

Reverse mutační primer

Reverse Complement Strand(5' to 3') is:
AGC ATG ATC CGG GAC GCC GAG CTG GCC GAG GAG

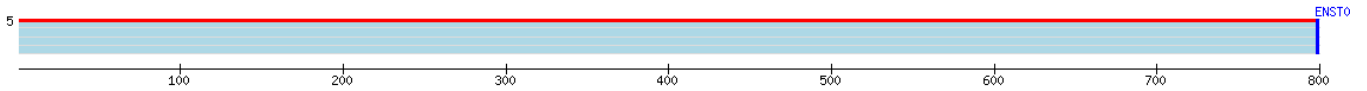
Mutační primery:

- **Forward (F):** 5'-AGC ATG ATC CGG GAC GGC GAG CTG GCC GAG GAG-3'
- **Reverse (R):** 5'-CTC CTC GGC CAG CTC GCC GTC CCG GAT CAT GCT-3'

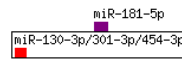
2) Podívejte se, zda jsou pro „vás“ gen predikované nějaké konzervované mikroRNA.

Pro TNF jsou známy **2 konzervované mikroRNA** (shodných 6 nukleotidů): miR-130-3p/301-3p/454-3p; miR-181-5p

Human TNF ENST00000449264.2 3' UTR length: 799



Conserved sites for miRNA families broadly conserved among vertebrates



- [Show poorly conserved sites for miRNA families conserved among vertebrates]
- [Show conserved sites for miRNA families conserved only among mammals]
- [Show poorly conserved sites for miRNA families conserved among mammals]
- [Show sites for poorly conserved but confidently annotated miRNA families]
- [Show sites for other miRBase annotations, most of which are miRNA* sequences or RNA fragments misannotated as miRNAs]

Key:

Sites with higher probability of preferential conservation

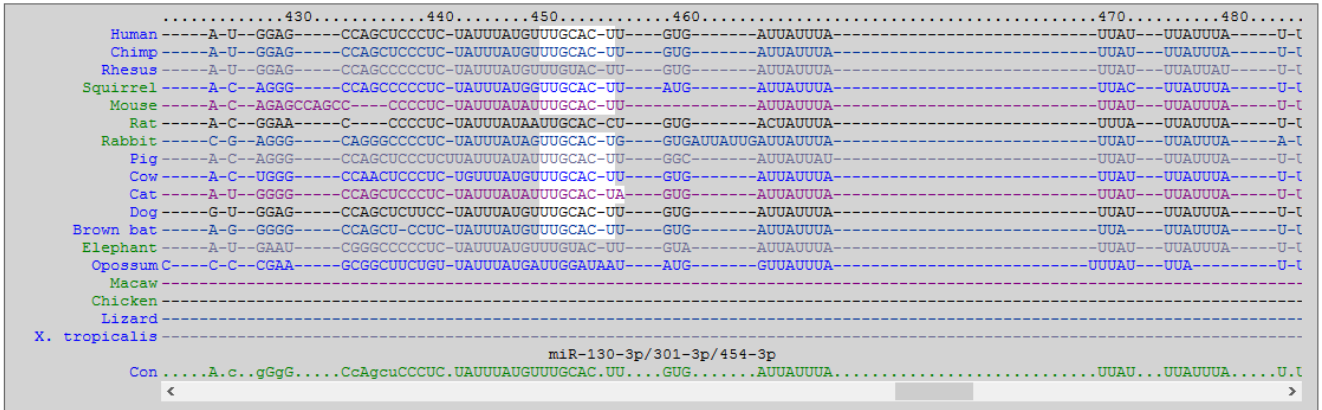
- 8mer
- 7mer-m8
- 7mer-A1
- non-canonical

Sites with lower probability of preferential conservation

- 8mer
- 7mer-m8
- 7mer-A1
- non-canonical

- [Download SVG image of miRNA sites]
- [View table of miRNA sites]
- [View human genome browser (hg19)]

[Show all species]



Conserved

	Predicted consequential pairing of target region (top) and miRNA (bottom)	Site type	Context++ score	Context++ score percentile	Weighted context++ score	Conserved branch length	P _{CT}
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGUUUGCAGC... 3' UGGGAUAUUCGUUAUAACGUGAU	7mer-m8	-0.36	98	-0.36	3.438	0.49
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGUUUGCAGC... 3' UACGGGAAAUAUGUAACGUGAC	7mer-m8	-0.38	98	-0.38	3.438	0.49
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGU-UUGCACUU... 3' UACGGGAAAUAUGUAACGUGAC	7mer-m8	-0.36	98	-0.36	3.438	0.49
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGUUUGCAGC... 3' AGCCGUAGAUGUAACGUGAC	7mer-m8	-0.35	98	-0.35	3.438	0.49
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGUUUGCAGC... 3' CGAAACUGUUAUGUAACGUGAC	7mer-m8	-0.34	98	-0.34	3.438	0.49
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGUUUGCAGC... 3' CGAAACUGUUAUGUAACGUGAC	7mer-m8	-0.34	98	-0.34	3.438	0.49
Position 451-457 of TNF 3' UTR	5' ...UCCUCUUAUUUAUGUUUGCAGC... 3' UUCCUUUUGUAACGUGAC	7mer-m8	-0.30	96	-0.30	3.438	0.49