Bioinfo_10_opakování

2022

Stáhněte z databáze sekvenci NM_008255.2

- Co a z jakého organismu tato sekvence kóduje?
- Jak dlouhý protein je touto sekvencí kódován?
- Bude CDS štěpena enzymy XhoI, NdeI, EcoRI ?
- Manuálně navrhněte primery pro amplifikaci sekvence **3. exonu** genu
- Ověřte přítomnost obou vašich navržených primerů v celé „CDS“ sekvenci pomocí programu „Multalin“

1) Sekvence z databáze

Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr), transcript variant 1, mRNA

NCBI Reference Sequence: NM_008255.2

Myší HMGCR

2)Délka proteinu

- Zobrazení CDS:

```
ORIGIN
        1 gacgatcctt ccttattggc ggcccgctgg cggcctggag cgtgcgtaag cgcagttcct
       61 tccgcccggg gctccgttgg ctggagacgg cagctgggcc ggcttggcct ccattgagat
      121 ccggaggatc caaggactgt gaggctacaa tgttgtcaag acttttccgg atgcatggcc
      181 tcttcgtggc ctcccacccc tgggaagtta ttgtgggaac agtgacactt acaatctgta
      241 tgatgtccat gaacatgttc accggcaaca acaagatctg tggctggaat tatgagtgcc
      301 ctaaatttga agaggacgtg ctgagcagcg acatcatcat cctgacgata acgcggtgca
      361 tcgccatcct gtacatttac ttccagttcc agaacctacg gcagcttggg tccaagtaca
      421 ttctgggtat tgctggcctc ttcacaattt tctcaagctt tgtcttcagt acagtcgtca
      481 ttcatttcct cgacaaagaa ctgacaggct taaatgaagc tttgcccttt tttctgctct
      541 tgattgacct ttctagagcg agtgcattag caaagtttgc cctcagttca aattcacagg
      601 atgaagtaag ggaaaatata gctcgtggaa tggcaatcct gggccccaca ttcactcttg
      661 acgctcttgt ggaatgcctt gtgattggag ttggcaccat gtcaggcgtc cgccagctgg
      721 agatcatgtg ctgcttcggc tgcatgtcag tgttggccaa ctactttgtg ttcatgacat
      781 tcttcccggc ctgtgtgtcg ctggtcctag agctttctcg tgaaagccgt gagggtcgtc
      841 caatttggca gctcagccat tttgccagag ttttagaaga agaagaaaat aaaccaaacc
      901 ccgtaaccca aagggtcaag atgattatgt ctttaggctt ggtccttgtt cacgctcata
      961 gtcgctggat agctgatcct tctcctcaga acagcacagc ggagcaggct aaggtttcct
     1021 tgggtctgga cgaagatgtg tccaagagaa ttgaaccaag tgtttctctc tggcagtttt
     1081 atctctccaa gatgatcagc atggacattg agcaagtgat taccctgagt ttagccttcc
     1141 ttttggctgt caagtatatt ttctttgaac aagcagagac agaatcgaca ctgtcattaa
     1201 aaaatcctat cacgtctccc gtcgtgacct caaagaaagc tcaagacaac tgttgtagac
     1261 gtgagcctct gcttgtgaga aggaaccaga agctttcgtc agtagaggag gatccaggag
     1321 cgaaccaaga gagaaaagtt gaggttataa aaccattagt ggtggaagca gagactacaa
     1381 gcagagctac gtttgtgctt ggggcttctg tagccagccc tccatccgcc ctggggacac
     1441 aggagcctgg aattgaactc cccatcgagc cacgacctaa tgaagaatgt ctgcagatac
     1501 tggagaatgc agagaaaggt gcaaagttcc ttagtgatgc agagatcatc cagttggtca
     1561 atgccaagca tatcccagct tacaaattgg aaactctaat ggaaactcat gaacgtggtg
     1621 tgtctattcg ccggcagctc ctctccacaa agcttgcaga gccgtcctct ctacagtacc
     1681 tgccttacag agattataat tattccttgg tgatgggagc ttgctgtgag aatgtgatcg
     1741 gatatatgcc catccctgtt ggagtggcag ggcctctgtg cctggatggg aaggagtacc
     1801 aggtgccgat ggcaacgacg gaaggctgtc ttgtggccag cactaacaga ggctgcagag
     1861 ccataagcct tggtggaggt gccagcagcc gggtccttgc agatgggatg actcgaggcc
     1921 cagtggtgcg tcttccacgt gcttgtgact ctgcagaagt aaagacctgg cttgaaacac
     1981 ctgaagggtt tgcagtgata aaggaggcct ttgatagcac cagcagattt gctcgtctac
     2041 agaaactcca cgtgacgatg gcaggacgca acctctatat ccgtttccag tccaggacgg
     2101 gggacgccat gggcatgaac atgatctcta agggtacgga gaaagcactg ctgaagcttc
     2161 aggagttctt tcctgacatg cagattctgg cagtcagtgg gaactattgc accgacaaga
     2221 agcctgctgc cataaattgg atcgaaggac gaggaaagac tgtggtttgt gaagccgtca
     2281 ttccagccaa ggtggtgaga gaggtgttaa agacaactac ggaagctatg gttgacgtaa
     2341 acattaacaa gaatcttgtg ggctcggcca tggctgggag cataggcggc tacaacgccc
     2401 acgcagcaaa cattgtcact gctatctaca tcgcatgtgg ccaggatgca gcacagaatg
     2461 tggggagttc aaactgtatt actttaatgg aagccagtgg tcccacaaat gaagacttat
     2521 atatcagctg caccatgcca tcgatagaga taggaaccgt gggtggtggg accaaccttc
     2581 tacctcagca agcctgcctg cagatgctag gtgttcaagg agcatgcaaa gacaatcctg
     2641 gagaaaacgc acggcagctt gcccgaattg tatgtggcac tgtgatggct ggtgagctgt
     2701 ccttgatggc agccttggca gcaggacatc ttgtcagaag tcacatggtt cacaacagat
     2761 caaagataaa tttacaagat cttcaaggaa cgtgcaccaa gaaggcagct tgaggatcct
     2821 gacactgaac tgaagcgcgg gcattgggtt ctcaaggact aacatgcaat ctgtgaatta
```

NM_008255 : 1 segment

```
150..2813
/gene="Hmgcr"
/gene_synonym="HMG-CoAR; Re
/EC_number=" 1.1.1.34 "
/note="isoform 1 is encod  by transcript variant 1;
HMG-CoA reductase; 3-hydroxy-3-methylglutaryl-CoA
reductase"
/codon_start=1
/product="3-hydroxy-3-methylglutaryl-coenzyme A reductase
isoform 1"
/protein_id=" NP_032281.2 "
/db_xref="CCDS: CCDS26706.1 "
/db_xref="GeneID: 15357 "
/db_xref="MGI: MGI:96159 "
/translation="MLSRLFRMHGLFVASHPWEVIVGTVTLTICMMSMNMFTGNNKIC
GWNYECPKFEEDVLSSDIIILTITRCIAILYIYFQFQNLRQLGSKYILGIAGLFTIFS
SFVFSTVVIHFLDKELTGLNEALPFFLLLIDLSRASALAKFALSSNSQDEVRENIARG
MAILGPTFTLDALVECLVIGVGTMSGVRQLEIMCCFGCMSVLANYFVFMTFFPACVSL
VLELSRESREGRPIWQLSHFARVLEEEENKPNPVTQRVKMIMSLGLVLVHAHSRWIAD
PSPQNSTAEQAKVSLGLDEDVSKRIEPSVSLWQFYLSKMISMDIEQVITLSLAFLLAV
KYIFFEQAETESTLSLKNPITSPVVTSKKAQDNCCRREPLLVRRNQKLSSVEEDPGAN
QERKVEVIKPLVVEAETTSRATFVLGASVASPPSALGTQEPGIELPIEPRPNEECLQI
LENAEKGAKFLSDAEIIQLVNAKHIPAYKLETLMETHERGVSIRRQLLSTKLAEPSSL
QYLPYRDYNYSLVMGACCENVIGYMPIPVGVAGPLCLDGKEYQVPMATTEGCLVASTN
RGCRAISLGGGASSRVLADGMTRGPVVRLPRACDSAEVKTWLETPEGFAVIKEAFDST
SRFARLQKLHVTMAGRNLYIRFQSRTGDAMGMNMISKGTEKALLKLQEFFPDMQILAV
SGNYCTDKKPAAINWIEGRGKTVVCEAVIPAKVVREVLKTTTEAMVDVNINKNLVGSA
MAGSIGGYNAHAANIVTAIYIACGQDAAQNVGSSNCITLMEASGPTNEDLYISCTMPS
IEIGTVGGGTNLLPQQACLQMLGVQGACKDNPGENARQLARIVCGTVMAGELSLMAAL
AAGHLVRSHMVHNRSKINLQDLQGTCTKKAA"
```

3-hydroxy-3-methylglutaryl-coenzyme A reductase isoform 1 [Mus musculus]

NCBI Reference Sequence: NP_032281.2

```
LOCUS       NP_032281                887 aa            linear   ROD 02-JAN-2022
DEFINITION  3-hydroxy-3-methylglutaryl-coenzyme A reductase isoform 1 [Mus
            musculus].
ACCESSION   NP_032281 XP_127496
```

Sequence Manipulation Suite:

Protein Stats

Protein Stats returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each resi
sequences.

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

```
RGCRAISLGGGASSRVLADGMTRGPVVRLPRACDSAEVKTWLETPEGFAVIKEAFDST
SRFARLQKLHVTMAGRNLYIRFQSRTGDAMG
SGNYCTDKKPAAINWIEGRGKTVVCEAVIPA
MAGSIGGYNAHAANIVTAIYIACGQDAAQNV
IEIGTVGGGTNLLPQQACLQMLGVQGACKDN
AAGHLVRSHMVHNRSKINLQDLQGTCTKKAA
```

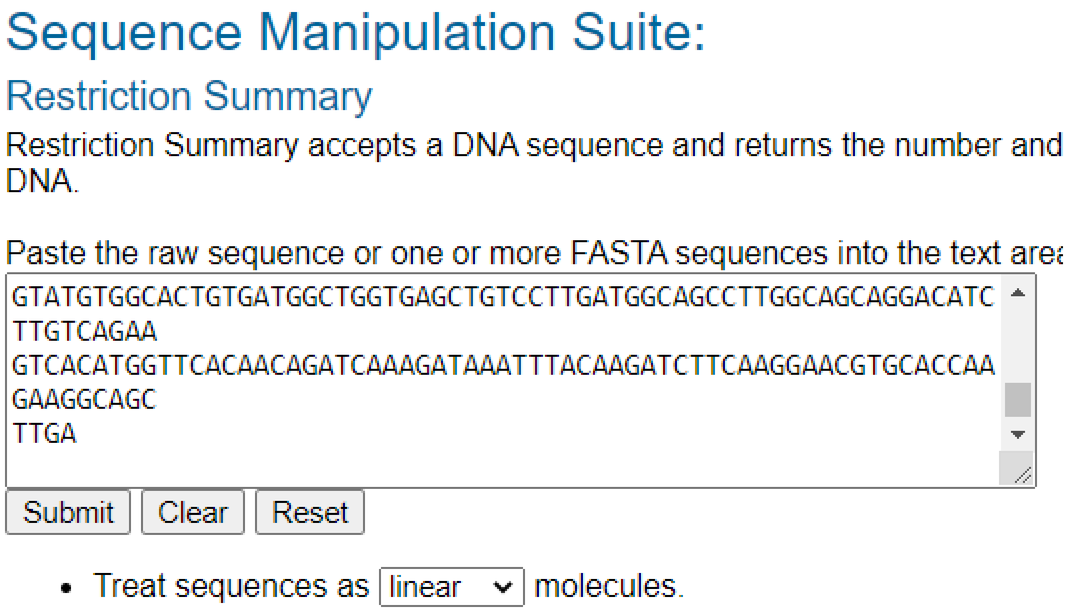
*This page requires JavaScript. See
*You can mirror this page or use it o

Protein Stats results

Results for 887 residue sequence "Untitled" starting "MLSRLFRMHG"

Pattern:	Times found:	Percenta
A	76	8.57

3) štěpení



Sequence Manipulation Suite:

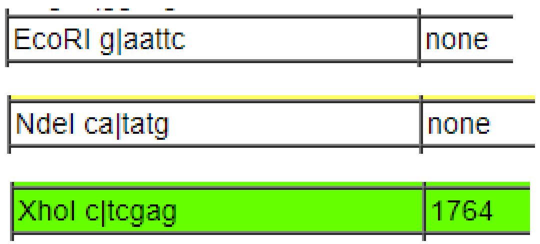
Restriction Summary

Restriction Summary accepts a DNA sequence and returns the number and
DNA.

Paste the raw sequence or one or more FASTA sequences into the text are

```
GTATGTGGCACTGTGATGGCTGGTGAGCTGTCCTTGATGGCAGCCTTGGCAGCAGGACATC
TTGTCAGAA
GTCACATGGTTCACAACAGATCAAAGATAAATTTACAAGATCTTCAAGGAACGTGCACCAA
GAAGGCAGC
TTGA
```

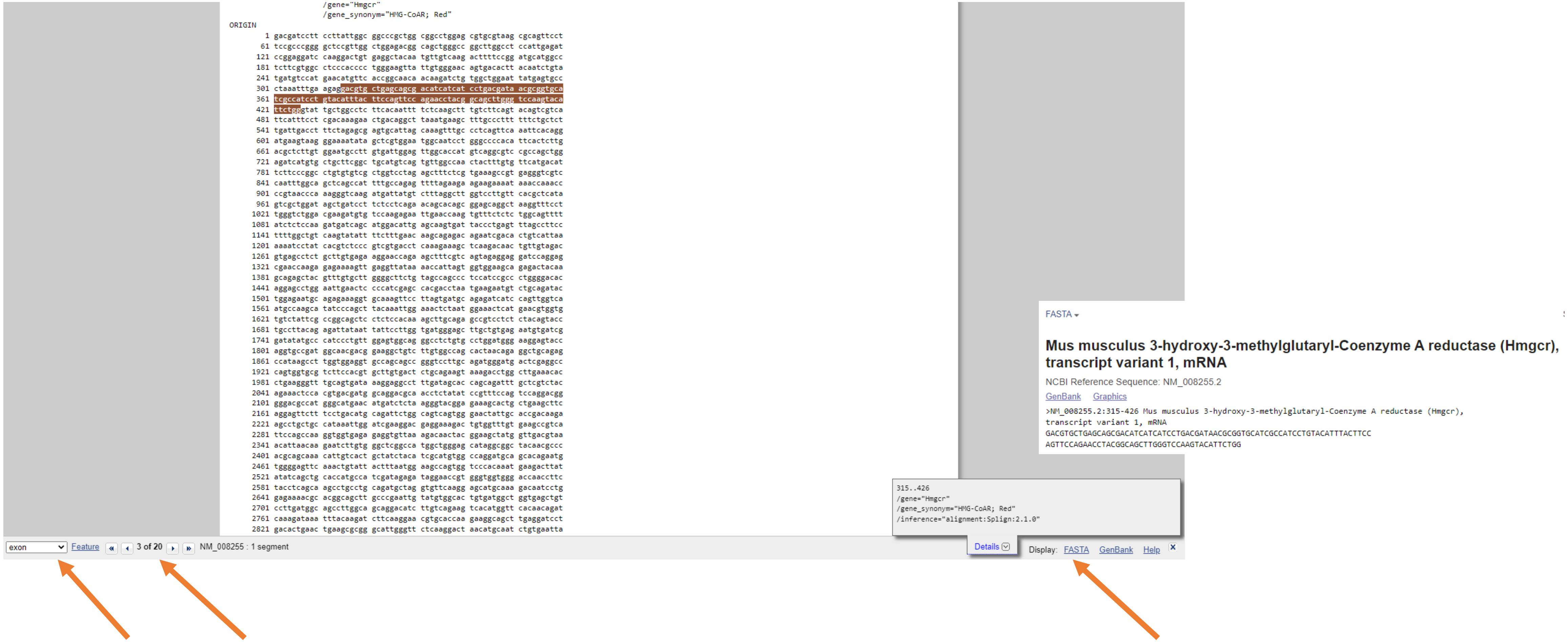
- Treat sequences as linear molecules.



EcoRI g|aattc	none
NdeI ca|tatg	none
XhoI c|tcgag	1764

4) Manuální primery

- 3.exon-sekvence



```
                     /gene="Hmgcr"
                     /gene_synonym="HMG-CoAR; Red"
ORIGIN
        1 gacgatcctt ccttattggc ggcccgctgg cggcctggag cgtgcgtaag cgcagttcct
       61 tccgcccggg gctccgttgg ctggagacgg cagctgggcc ggcttggcct ccattgagat
      121 ccggaggatc caaggactgt gaggctacaa tgttgtcaag acttttccgg atgcatggcc
      181 tcttcgtggc ctcccacccc tgggaagtta ttgtgggaac agtgacactt acaatctgta
      241 tgatgtccat gaacatgttc accggcaaca acaagatctg tggctggaat tatgagtgcc
      301 ctaaatttga agaggacgtg ctgagcagcg acatcatcat cctgacgata acgcggtgca
      361 tcgccatcct gtacatttac ttccagttcc agaacctacg gcagcttggg tccaagtaca
      421 ttctgggtat tgctggcctc ttcacaattt tctcaagctt tgtcttcagt acagtcgtca
      481 ttcatttcct cgacaaagaa ctgacaggct taaatgaagc tttgcccttt tttctgctct
      541 tgattgacct ttctagagcg agtgcattag caaagtttgc cctcagttca aattcacagg
      601 atgaagtaag ggaaaatata gctcgtggaa tggcaatcct gggccccaca ttcactcttg
      661 acgctcttgt ggaatgcctt gtgattggag ttggcaccat gtcaggcgtc cgccagctgg
      721 agatcatgtg ctgcttcggc tgcatgtcag tgttggccaa ctactttgtg ttcatgacat
      781 tcttcccggc ctgtgtgtcg ctggtcctag agctttctcg tgaaagccgt gagggtcgtc
      841 caatttggca gctcagccat tttgccagag ttttagaaga agaagaaaat aaaccaaacc
      901 ccgtaaccca aagggtcaag atgattatgt ctttaggctt ggtccttgtt cacgctcata
      961 gtcgctggat agctgatcct tctcctcaga acagcacagc ggagcaggct aaggtttcct
     1021 tgggtctgga cgaagatgtg tccaagagaa ttgaaccaag tgtttctctc tggcagtttt
     1081 atctctccaa gatgatcagc atggacattg agcaagtgat taccctgagt ttagccttcc
     1141 ttttggctgt caagtatatt ttctttgaac aagcagagac agaatcgaca ctgtcattaa
     1201 aaaatcctat cacgtctccc gtcgtgacct caaagaaagc tcaagacaac tgttgtagac
     1261 gtgagcctct gcttgtgaga aggaaccaga agctttcgtc agtagaggag gatccaggag
     1321 cgaaccaaga gagaaaagtt gaggttataa aaccattagt ggtggaagca gagactacaa
     1381 gcagagctac gtttgtgctt ggggcttctg tagccagccc tccatccgcc ctggggacac
     1441 aggagcctgg aattgaactc cccatcgagc cacgacctaa tgaagaatgt ctgcagatac
     1501 tggagaatgc agagaaaggt gcaaagttcc ttagtgatgc agagatcatc cagttggtca
     1561 atgccaagca tatcccagct tacaaattgg aaactctaat ggaaactcat gaacgtggtg
     1621 tgtctattcg ccggcagctc ctctccacaa agcttgcaga gccgtcctct ctacagtacc
     1681 tgccttacag agattataat tattccttgg tgatgggagc ttgctgtgag aatgtgatcg
     1741 gatatatgcc catccctgtt ggagtggcag ggcctctgtg cctggatggg aaggagtacc
     1801 aggtgccgat ggcaacgacg gaaggctgtc ttgtggccag cactaacaga ggctgcagag
     1861 ccataagcct tggtggaggt gccagcagcc gggtccttgc agatgggatg actcgaggcc
     1921 cagtggtgcg tcttccacgt gcttgtgact ctgcagaagt aaagacctgg cttgaaacac
     1981 ctgaagggtt tgcagtgata aaggaggcct ttgatagcac cagcagattt gctcgtctac
     2041 agaaactcca cgtgacgatg gcaggacgca acctctatat ccgtttccag tccaggacgg
     2101 gggacgccat gggcatgaac atgatctcta agggtacgga gaaagcactg ctgaagcttc
     2161 aggagttctt tcctgacatg cagattctgg cagtcagtgg gaactattgc accgacaaga
     2221 agcctgctgc cataaattgg atcgaaggac gaggaaagac tgtggtttgt gaagccgtca
     2281 ttccagccaa ggtggtgaga gaggtgttaa agacaactac ggaagctatg gttgacgtaa
     2341 acattaacaa gaatcttgtg ggctcggcca tggctgggag cataggcggc tacaacgccc
     2401 acgcagcaaa cattgtcact gctatctaca tcgcatgtgg ccaggatgca gcacagaatg
     2461 tggggagttc aaactgtatt actttaatgg aagccagtgg tcccacaaat gaagacttat
     2521 atatcagctg caccatgcca tcgatagaga taggaaccgt gggtggtggg accaaccttc
     2581 tacctcagca agcctgcctg cagatgctag gtgttcaagg agcatgcaaa gacaatcctg
     2641 gagaaaacgc acggcagctt gcccgaattg tatgtggcac tgtgatggct ggtgagctgt
     2701 ccttgatggc agccttggca gcaggacatc ttgtcagaag tcacatggtt cacaacagat
     2761 caaagataaa tttacaagat cttcaaggaa cgtgcaccaa gaaggcagct tgaggatcct
     2821 gacactgaac tgaagcgcgg gcattgggtt ctcaaggact aacatgcaat ctgtgaatta
```

```
315..426
/gene="Hmgcr"
/gene_synonym="HMG-CoAR; Red"
/inference="alignment:Splign:2.1.0"
```

FASTA

Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr), transcript variant 1, mRNA

NCBI Reference Sequence: NM_008255.2

```
>NM_008255.2:315-426 Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr),
transcript variant 1, mRNA
GACGTGCTGAGCAGCGACATCATCATCCTGACGATAACGCGGTGCATCGCCATCCTGTACATTTACTTCC
AGTTCCAGAACCTACGGCAGCTTGGGTCCAAGTACATTCTGG
```

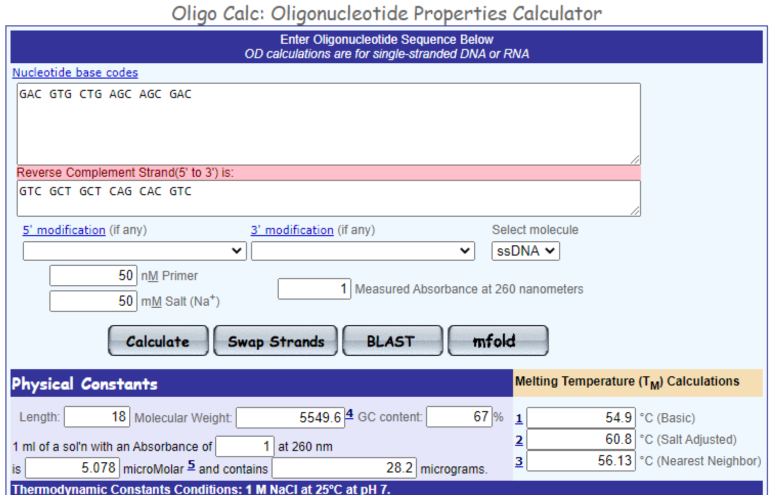
4) Manuální primery

- F_primer: `GACGTGCTGAGCAGCGAC`

Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr), transcript variant 1, mRNA

NCBI Reference Sequence: NM_008255.2

```
>NM_008255.2:315-426 Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr),
transcript variant 1, mRNA
GACGTGCTGAGCAGCGACATCATCATCCTGACGATAACGCGGTGCATCGCCATCCTGTACATTTACTTCC
AGTTCCAGAACCTACGGCAGCTTGGGTCCAAGTACATTCTGG
```



Oligo Calc: Oligonucleotide Properties Calculator

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

```
GAC GTG CTG AGC AGC GAC
```

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

```
GTC GCT GCT CAG CAC GTC
```

Select molecule: ssDNA

50 nM Primer

50 mM Salt (Na^+)

1 Measured Absorbance at 260 nanometers

Physical Constants

Length: 18 Molecular Weight: 5549.6 [4] GC content: 67 %

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 5.078 microMolar [5] and contains 28.2 micrograms.

Melting Temperature (T_M) Calculations

1 54.9 °C (Basic)
2 60.8 °C (Salt Adjusted)
3 56.13 °C (Nearest Neighbor)

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

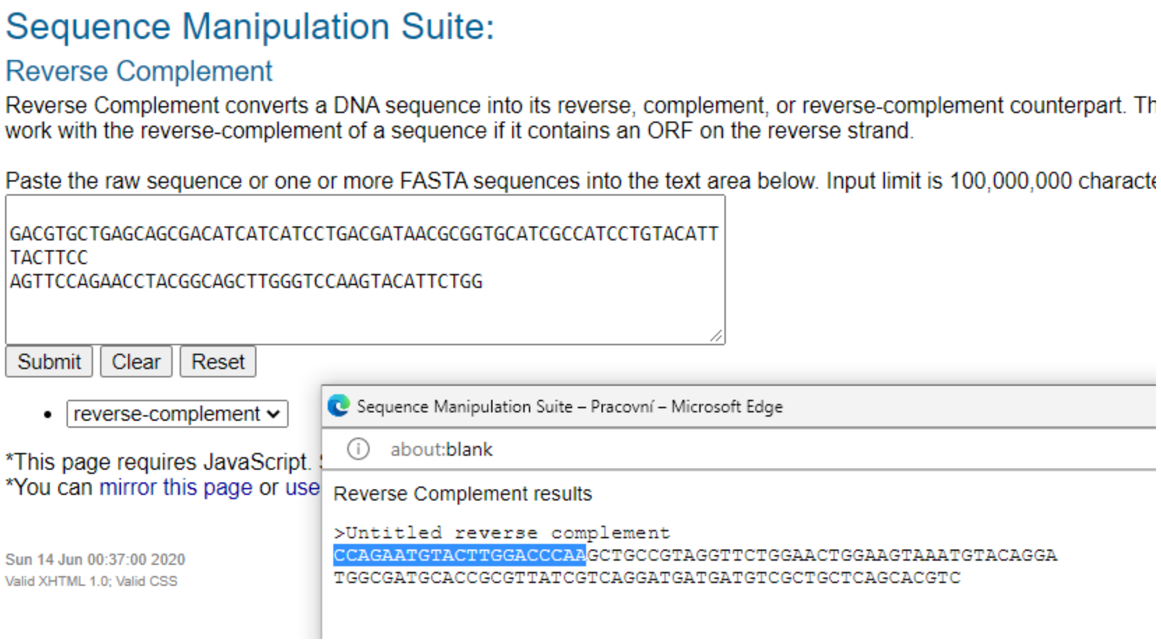
4) Manuální primery

- **R_primer:** CCAGAATGTACTTGGACCCAA

Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr), transcript variant 1, mRNA

NCBI Reference Sequence: NM_008255.2

```
>NM_008255.2:315-426 Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Hmgcr),
transcript variant 1, mRNA
GACGTGCTGAGCAGCGACATCATCATCCTGACGATAACGCGGTGCATCGCCATCCTGTACATTTACTTCC
AGTTCCAGAACCTACGGCAGCTTGGGTCCAAGTACATTCTGG
```



Sequence Manipulation Suite:

Reverse Complement

Reverse Complement converts a DNA sequence into its reverse, complement, or reverse-complement counterpart. Th
work with the reverse-complement of a sequence if it contains an ORF on the reverse strand.

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

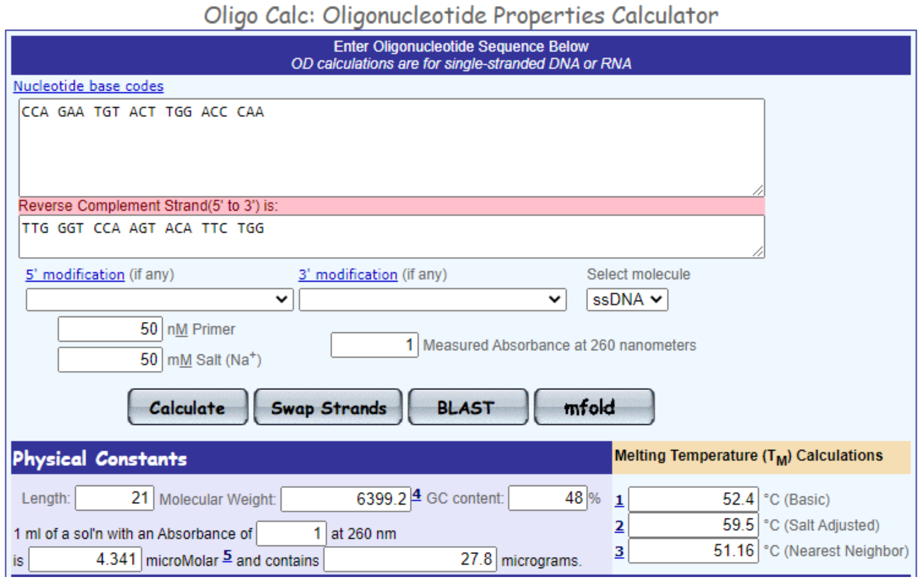
```
GACGTGCTGAGCAGCGACATCATCATCCTGACGATAACGCGGTGCATCGCCATCCTGTACATT
TACTTCC
AGTTCCAGAACCTACGGCAGCTTGGGTCCAAGTACATTCTGG
```

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

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

Reverse Complement results

```
>Untitled reverse complement
CCAGAATGTACTTGGACCCAAGCTGCCGTAGGTTCTGGAACTGGAAGTAAATGTACAGGA
TGGCGATGCACCGCGTTATCGTCAGGATGATGATGTCGCTGCTCAGCACGTC
```



Oligo Calc: Oligonucleotide Properties Calculator

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

```
CCA GAA TGT ACT TGG ACC CAA
```

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

```
TTG GGT CCA AGT ACA TTC TGG
```

Select molecule: ssDNA

50 nM Primer

50 mM Salt (Na$^+$)

1 Measured Absorbance at 260 nanometers

Physical Constants

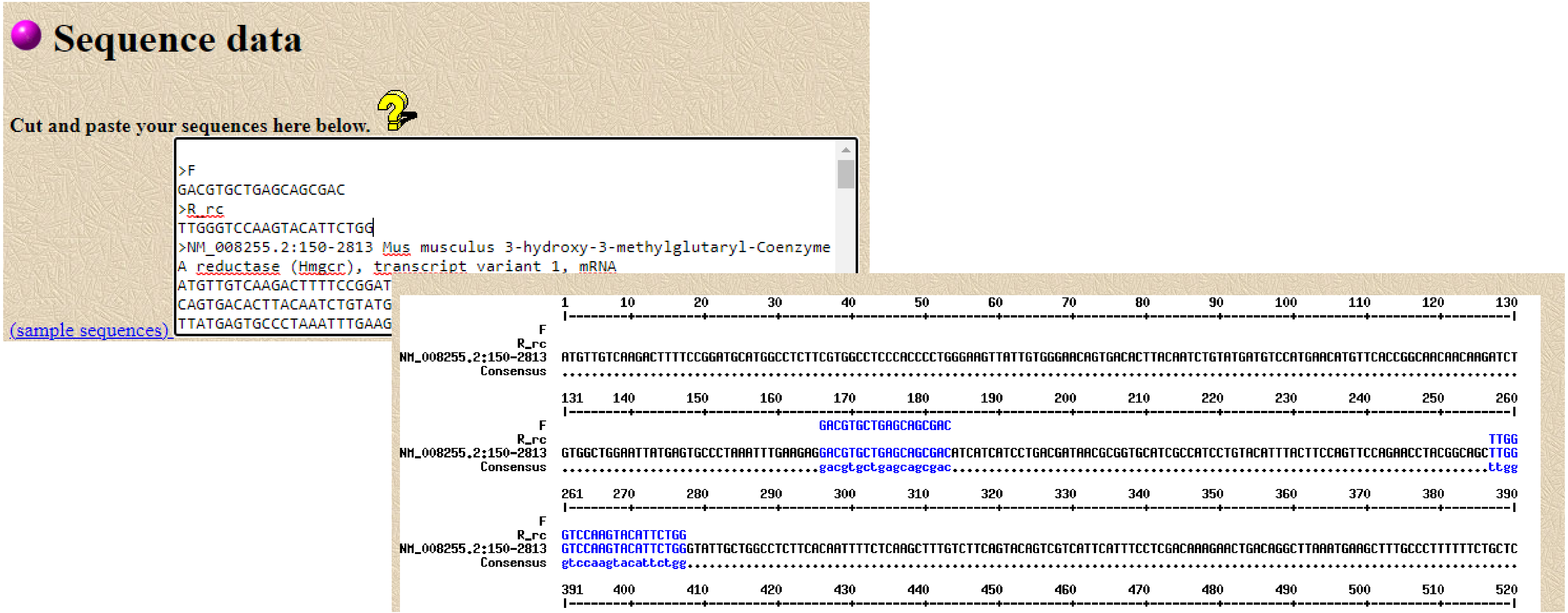
Length: 21 Molecular Weight: 6399.2[4] GC content: 48 %

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 4.341 microMolar[5] and contains 27.8 micrograms.

Melting Temperature (T_M) Calculations

52.4 °C (Basic)
59.5 °C (Salt Adjusted)
51.16 °C (Nearest Neighbor)

5) porovnání



Sequence data

Cut and paste your sequences here below.

```
>F
GACGTGCTGAGCAGCGAC
>R_rc
TTGGGTCCAAGTACATTCTGG
>NM_008255.2:150-2813 Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme
A reductase (Hmgcr), transcript variant 1, mRNA
ATGTTGTCAAGACTTTTCCGGAT
CAGTGACACTTACAATCTGTATG
TTATGAGTGCCCTAAATTTGAAG
```

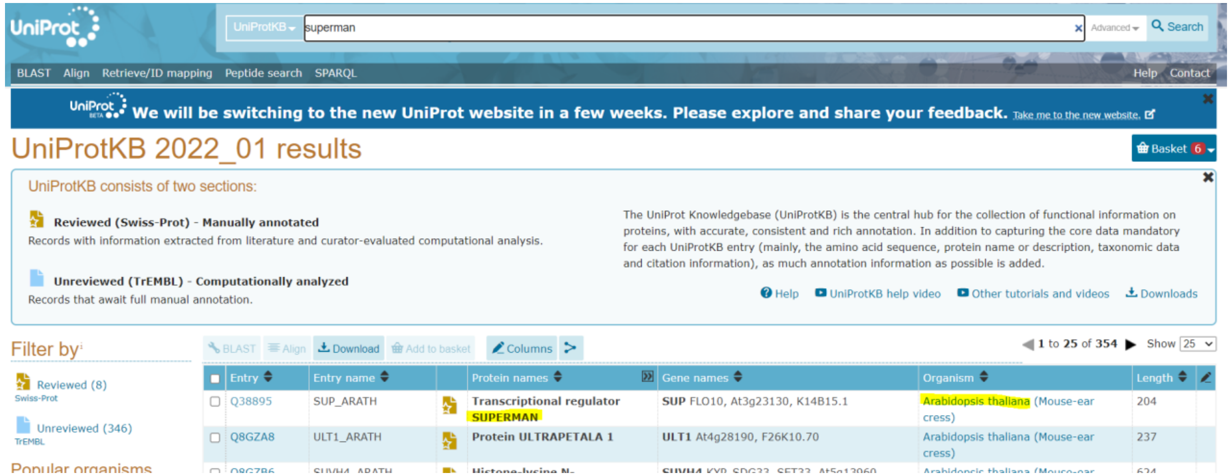
(sample sequences)

```
                      1       10        20        30        40        50        60        70        80        90       100       110       120       130
                      |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
                   F
                R_rc
NM_008255.2:150-2813  ATGTTGTCAAGACTTTTCCGGATGCATGGCCTCTTCGTGGCCTCCCACCCCTGGGAAGTTATTGTGGGAACAGTGACACTTACAATCTGTATGATGTCCATGAACATGTTCACCGGCAACAACAAGATCT
           Consensus  ..................................................................................................................................

                      131    140       150       160       170       180       190       200       210       220       230       240       250       260
                      |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
                   F                                     GACGTGCTGAGCAGCGAC
                R_rc                                                                                                                                TTGG
NM_008255.2:150-2813  GTGGCTGGAATTATGAGTGCCCTAAATTTGAAGAGGACGTGCTGAGCAGCGACATCATCATCCTGACGATAACGCGGTGCATCGCCATCCTGTACATTTACTTCCAGTTCCAGAACCTACGGCAGCTTGG
           Consensus  ...................................gacgtgctgagcagcgac.........................................................................ttgg

                      261    270       280       290       300       310       320       330       340       350       360       370       380       390
                      |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
                   F
                R_rc  GTCCAAGTACATTCTGG
NM_008255.2:150-2813  GTCCAAGTACATTCTGGGTATTGCTGGCCTCTTCACAATTTTCTCAAGCTTTGTCTTCAGTACAGTCGTCATTCATTTCCTCGACAAAGAACTGACAGGCTTAAATGAAGCTTTGCCCTTTTTTCTGCTC
           Consensus  gtccaagtacattctgg.................................................................................................................

                      391    400       410       420       430       440       450       460       470       480       490       500       510       520
                      |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
```

Najděte proteinovou sekvenci s názvem “SUPERMAN” z *Arabidopsis thaliana*

- Jaký je přístupový kód a funkce tohoto proteinu?
- Kolikrát by tato sekvence byla štepena pepsinem (pH1.3)?
- Má sekvence nějaké pravděpodobné transmembránové helixy?
- Navrhněte primery pro detekci tohoto genu (tak aby amplifikační produkt nebyl delší než 500bp).
- Existuje nějaký lidský homolog SUPERMANa(u)?

1)Superman *Arabidopsis thaliana*



We will be switching to the new UniProt website in a few weeks. Please explore and share your feedback. Take me to the new website.

UniProtKB 2022_01 results

UniProtKB consists of two sections:

Reviewed (Swiss-Prot) - Manually annotated
Records with information extracted from literature and curator-evaluated computational analysis.

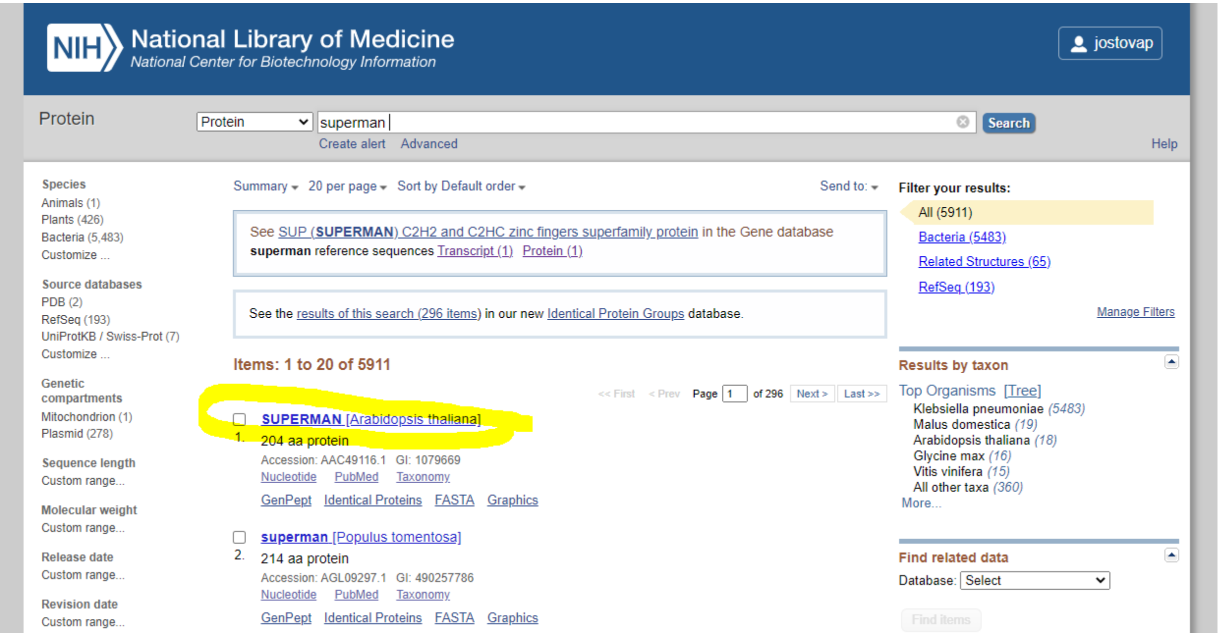
Unreviewed (TrEMBL) - Computationally analyzed
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Help · UniProtKB help video · Other tutorials and videos · Downloads

☐	Entry	Entry name	Protein names	Gene names	Organism	Length
☐	Q38895	SUP_ARATH	**Transcriptional regulator SUPERMAN**	**SUP** FLO10, At3g23130, K14B15.1	Arabidopsis thaliana (Mouse-ear cress)	204
☐	Q8GZA8	ULT1_ARATH	**Protein ULTRAPETALA 1**	**ULT1** At4g28190, F26K10.70	Arabidopsis thaliana (Mouse-ear cress)	237
☐	Q8GZB6	SUVH4_ARATH	**Histone-lysine N-**	**SUVH4** KYP, SDG33, SET33, At5g13060	Arabidopsis thaliana (Mouse-ear	624

Nebo:



See SUP (**SUPERMAN**) C2H2 and C2HC zinc fingers superfamily protein in the Gene database
superman reference sequences Transcript (1) Protein (1)

See the results of this search (296 items) in our new Identical Protein Groups database.

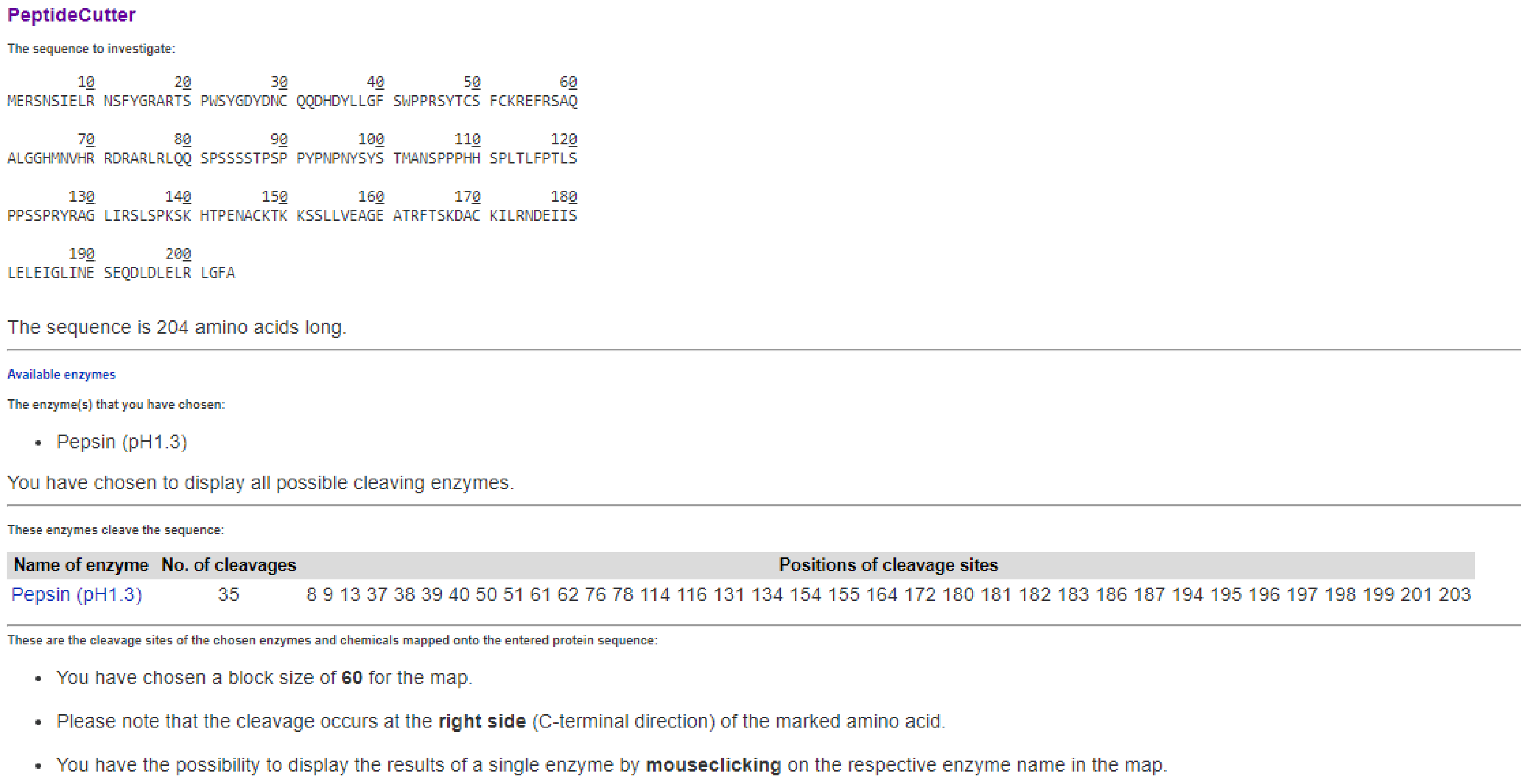
Items: 1 to 20 of 5911

1. ☐ **SUPERMAN** [Arabidopsis thaliana]
204 aa protein
Accession: AAC49116.1 GI: 1079669
Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

2. ☐ **superman** [Populus tomentosa]
214 aa protein
Accession: AGL09297.1 GI: 490257786
Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

2)Štěpení pepsinem

- 35x



PeptideCutter

The sequence to investigate:

```
        10         20         30         40         50         60
MERSNSIELR NSFYGRARTS PWSYGDYDNC QQDHDYLLGF SWPPRSYTCS FCKREFRSAQ

        70         80         90        100        110        120
ALGGHMNVHR RDRARLRLQQ SPSSSSTPSP PYPNPNYSYS TMANSPPPHH SPLTLFPTLS

       130        140        150        160        170        180
PPSSPRYRAG LIRSLSPKSK HTPENACKTK KSSLLVEAGE ATRFTSKDAC KILRNDEIIS

       190        200
LELEIGLINE SEQDLDLELR LGFA
```

The sequence is 204 amino acids long.

Available enzymes

The enzyme(s) that you have chosen:

- Pepsin (pH1.3)

You have chosen to display all possible cleaving enzymes.

These enzymes cleave the sequence:

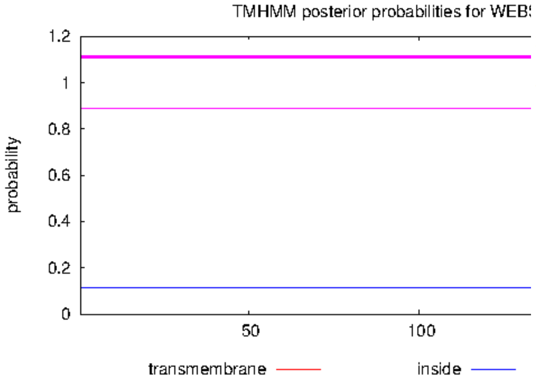
Name of enzyme	No. of cleavages	Positions of cleavage sites
Pepsin (pH1.3)	35	8 9 13 37 38 39 40 50 51 61 62 76 78 114 116 131 134 154 155 164 172 180 181 182 183 186 187 194 195 196 197 198 199 201 203

These are the cleavage sites of the chosen enzymes and chemicals mapped onto the entered protein sequence:

- You have chosen a block size of **60** for the map.
- Please note that the cleavage occurs at the **right side** (C-terminal direction) of the marked amino acid.
- You have the possibility to display the results of a single enzyme by **mouseclicking** on the respective enzyme name in the map.

3)TM helixy?

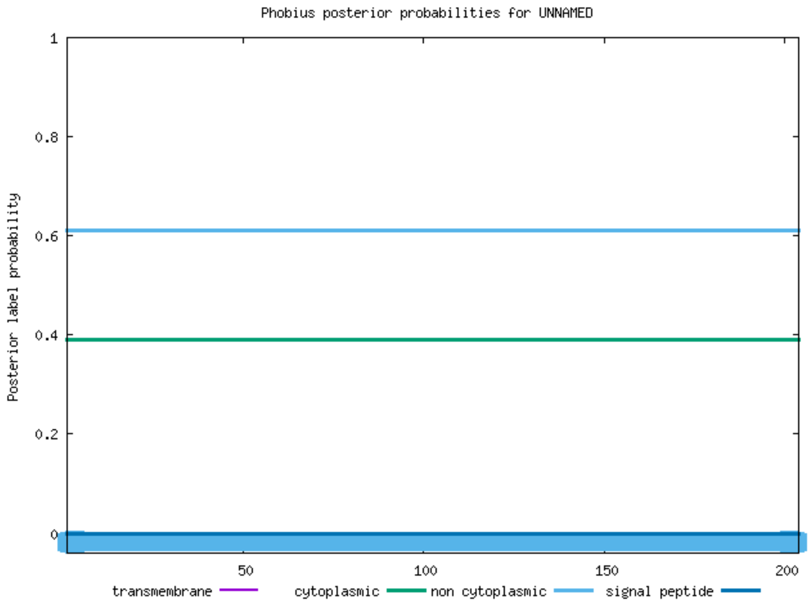
- ne



Phobius prediction

Prediction of UNNAMED

```
ID   UNNAMED
FT   TOPO_DOM      1     204       NON CYTOPLASMIC.
//
```



Results for job af3a5213b773d883a54ae65f50d2a077

Download results

XML file

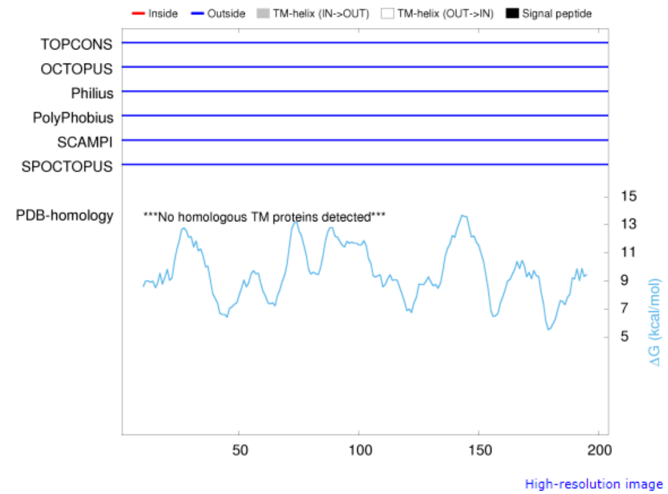
Proteins:

sequences: 1/1

Code: SUP
Evidence: notTm
Reliability: 99
#Tm: 0
Cross references:

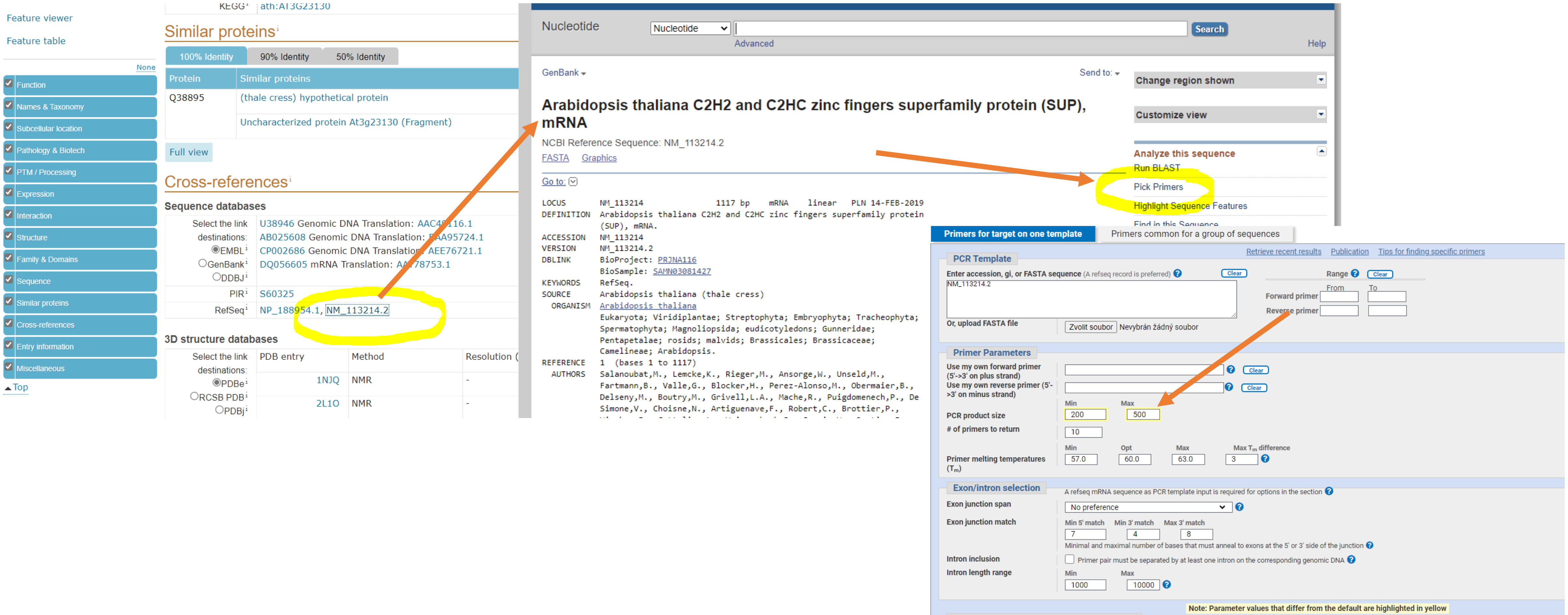
All: #visitors: 2364671, #jobs: 545034, #seqs: 1763306 .:|:. Last week: 247 jobs .:|:. Current load: 0 jobs

Predicted topologies and predicted ΔG values:



4)Primery pro **detekci** genu (mám protein)

- Nutná (pouze) specificita! – použít „pick primers“



KEGG: ath:AT3G23130

Similar proteins

Protein	Similar proteins
Q38895	(thale cress) hypothetical protein
	Uncharacterized protein At3g23130 (Fragment)

Cross-references

Sequence databases

Select the link destinations: ● EMBL ○ GenBank ○ DDBJ

U38946 Genomic DNA Translation: AAC49116.1
AB025608 Genomic DNA Translation: BAA95724.1
CP002686 Genomic DNA Translation: AEE76721.1
DQ056605 mRNA Translation: AAY78753.1

PIR: S60325

RefSeq: NP_188954.1, NM_113214.2

3D structure databases

Select the link destinations: ● PDBe ○ RCSB PDB ○ PDBj

PDB entry	Method	Resolution (
1NJQ	NMR	-
2L1O	NMR	-

Arabidopsis thaliana C2H2 and C2HC zinc fingers superfamily protein (SUP), mRNA

NCBI Reference Sequence: NM_113214.2

```
LOCUS       NM_113214               1117 bp    mRNA    linear   PLN 14-FEB-2019
DEFINITION  Arabidopsis thaliana C2H2 and C2HC zinc fingers superfamily protein
            (SUP), mRNA.
ACCESSION   NM_113214
VERSION     NM_113214.2
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Camelineae; Arabidopsis.
REFERENCE   1  (bases 1 to 1117)
  AUTHORS   Salanoubat,M., Lemcke,K., Rieger,M., Ansorge,W., Unseld,M.,
            Fartmann,B., Valle,G., Blocker,H., Perez-Alonso,M., Obermaier,B.,
            Delseny,M., Boutry,M., Grivell,L.A., Mache,R., Puigdomenech,P., De
            Simone,V., Choisne,N., Artiguenave,F., Robert,C., Brottier,P.,
```

Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred): NM_113214.2
Or, upload FASTA file: Nevybrán žádný soubor
Range: Forward primer From / To; Reverse primer From / To

Primer Parameters

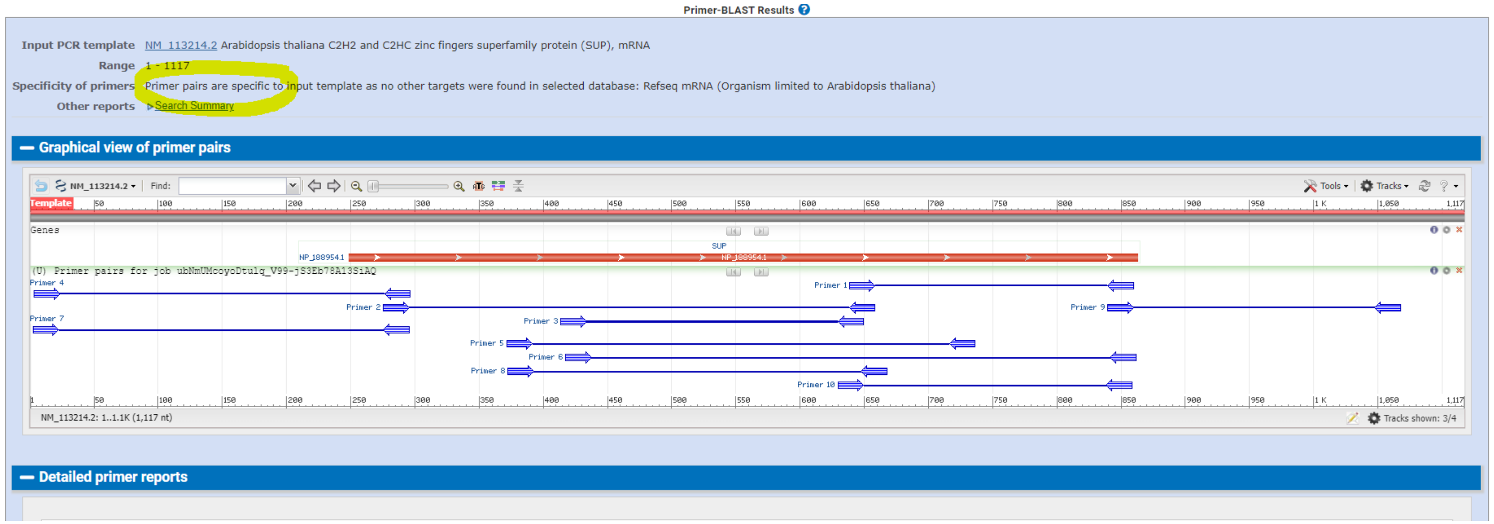
Use my own forward primer (5'->3' on plus strand)
Use my own reverse primer (5'->3' on minus strand)
PCR product size: Min 200, Max 500
# of primers to return: 10
Primer melting temperatures (T_m): Min 57.0, Opt 60.0, Max 63.0, Max T_m difference 3

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section
Exon junction span: No preference
Exon junction match: Min 5' match 7, Min 3' match 4, Max 3' match 8
Minimal and maximal number of bases that must anneal to exons at the 5' or 3' side of the junction
Intron inclusion: ☐ Primer pair must be separated by at least one intron on the corresponding genomic DNA
Intron length range: Min 1000, Max 10000

Note: Parameter values that differ from the default are highlighted in yellow

4)Primery pro **detekci** genu



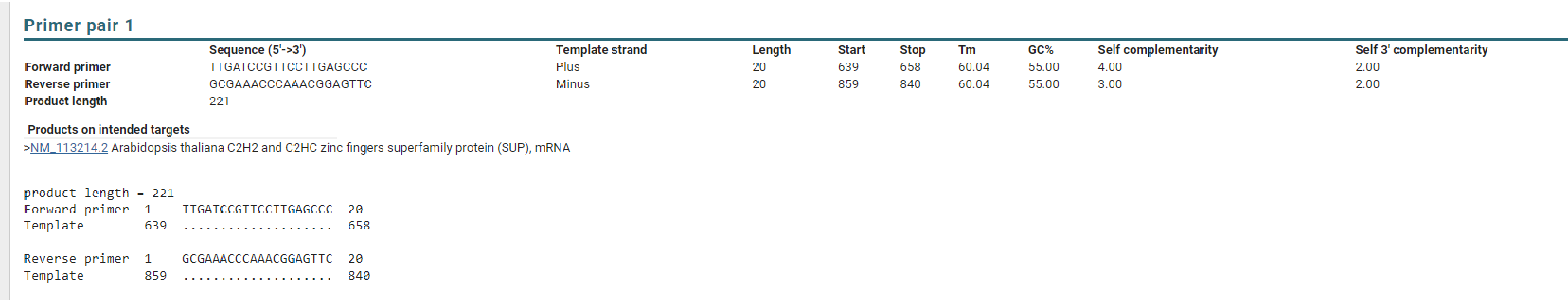
Primer-BLAST Results

Input PCR template NM_113214.2 Arabidopsis thaliana C2H2 and C2HC zinc fingers superfamily protein (SUP), mRNA
Range 1 - 1117
Specificity of primers Primer pairs are specific to input template as no other targets were found in selected database: Refseq mRNA (Organism limited to Arabidopsis thaliana)
Other reports ▷Search Summary

— Graphical view of primer pairs

— Detailed primer reports

- Např:



Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TTGATCCGTTCCTTGAGCCC	Plus	20	639	658	60.04	55.00	4.00	2.00
Reverse primer	GCGAAACCCAAACGGAGTTC	Minus	20	859	840	60.04	55.00	3.00	2.00
Product length	221								

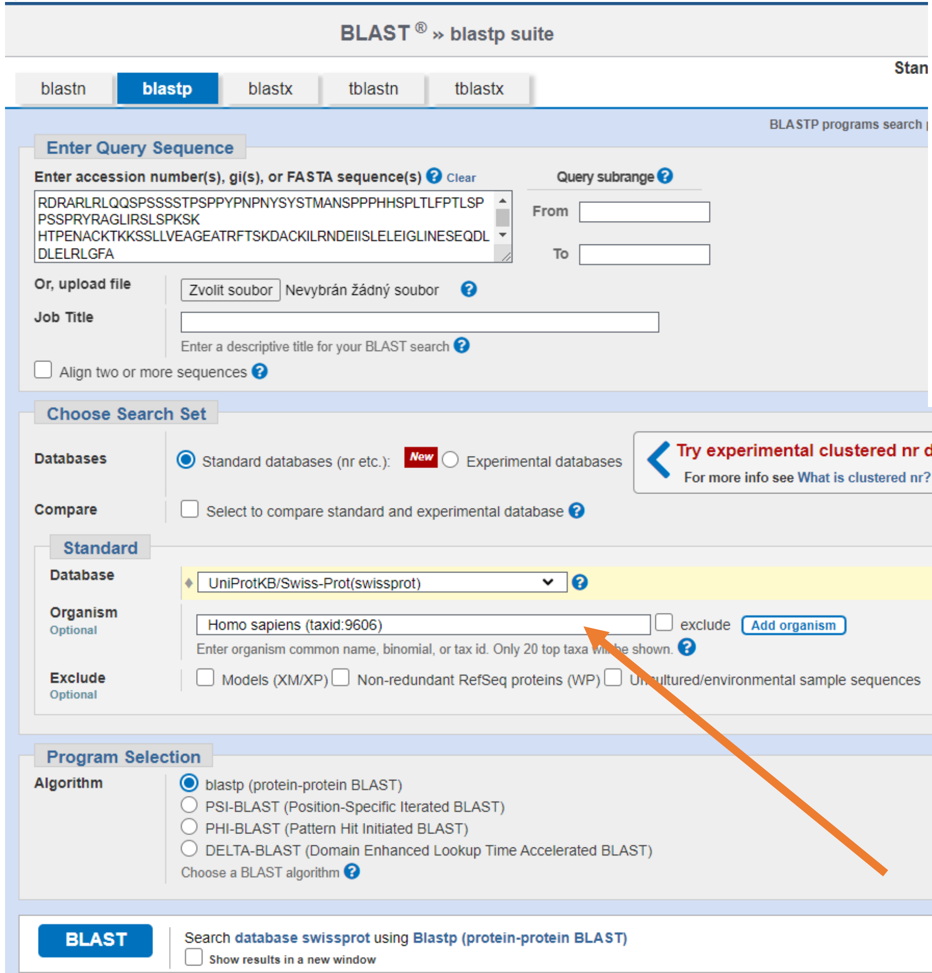
Products on intended targets

>NM_113214.2 Arabidopsis thaliana C2H2 and C2HC zinc fingers superfamily protein (SUP), mRNA

```
product length = 221
Forward primer  1    TTGATCCGTTCCTTGAGCCC  20
Template        639  ....................  658

Reverse primer  1    GCGAAACCCAAACGGAGTTC  20
Template        859  ....................  840
```

5)Lidský homolog?

- BLASTp



BLAST ® » blastp suite

BLASTP programs search

Enter Query Sequence

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

RDRARLRLQQSPSSSSTPSPPYPNPNYSYSTMANSPPPHHSPLTLFPTLSP
PSSPRYRAGLIRSLSPKSK
HTPENACKTKKSSLLVEAGEATRFTSKDACKILRNDEIISLELEIGLINESEQDL
DLELRLGFA

Query subrange: From / To

Or, upload file Zvolit soubor Nevybrán žádný soubor

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Databases ● Standard databases (nr etc.): New ○ Experimental databases

Try experimental clustered nr database
For more info see What is clustered nr?

Compare ☐ Select to compare standard and experimental database

Standard

Database UniProtKB/Swiss-Prot(swissprot)

Organism Optional: Homo sapiens (taxid:9606) ☐ exclude Add organism

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

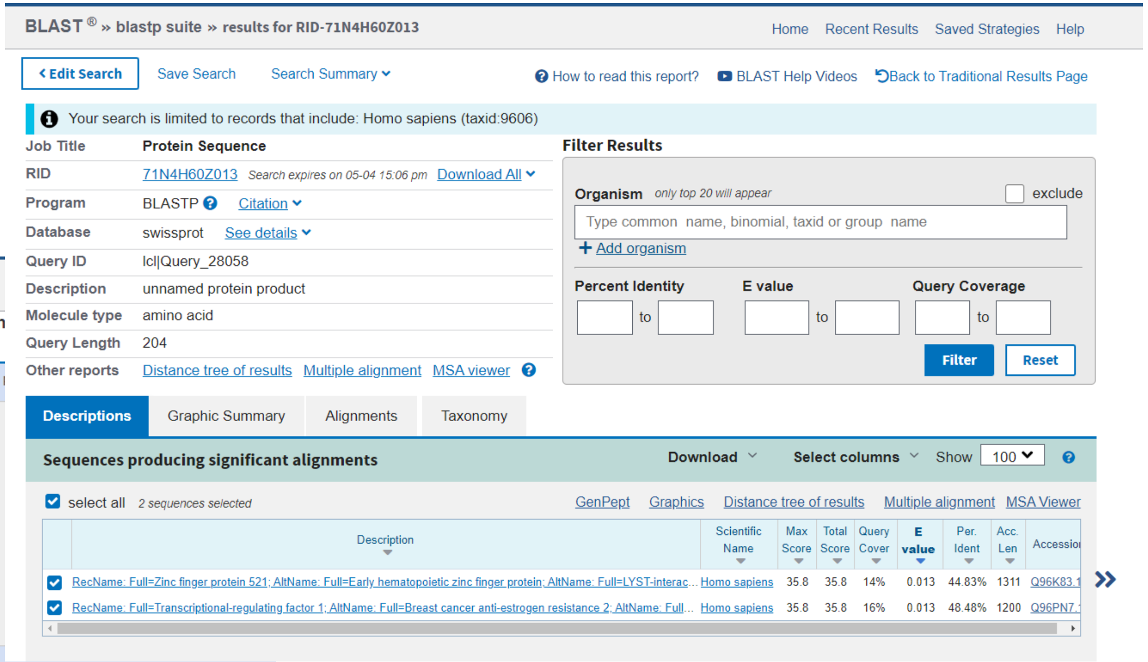
Exclude Optional: ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm
● blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)
○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST Search database swissprot using Blastp (protein-protein BLAST)
☐ Show results in a new window



BLAST ® » blastp suite » results for RID-71N4H60Z013

Your search is limited to records that include: Homo sapiens (taxid:9606)

Job Title	Protein Sequence
RID	71N4H60Z013 Search expires on 05-04 15:06 pm Download All
Program	BLASTP Citation
Database	swissprot See details
Query ID	lcl|Query_28058
Description	unnamed protein product
Molecule type	amino acid
Query Length	204
Other reports	Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism only top 20 will appear ☐ exclude
Type common name, binomial, taxid or group name
Add organism
Percent Identity ... to ... E value ... to ... Query Coverage ... to ...
Filter Reset

Sequences producing significant alignments Download Select columns Show 100

☑ select all 2 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accessio
☑	RecName: Full=Zinc finger protein 521; AltName: Full=Early hematopoietic zinc finger protein; AltName: Full=LYST-interac...	Homo sapiens	35.8	35.8	14%	0.013	44.83%	1311	Q96K83.1
☑	RecName: Full=Transcriptional-regulating factor 1; AltName: Full=Breast cancer anti-estrogen resistance 2; AltName: Full...	Homo sapiens	35.8	35.8	16%	0.013	48.48%	1200	Q96PN7.

Nalezeny dvě podobné sekvence