

# Základy praktické Bioinformatiky

---

PETRA MATOUŠKOVÁ

2022/2023

10/10

# Základy praktické bioinformatiky

Téma 10/10

---

## Opakování

**Cíle:**

Student bude dobře připraven na zkoušku 😊

# Zkouškové otázky 2021

---

# Zkoušková písemka:

---

2 úlohy, každá 5 bodů

Hodnocení: max celkem 10 bodů

9-10b: 1

7-8b: 2

5-6b: 3

**Verze 4.5.2021**

A: Stáhněte si sekvenci `NM_002085.5`

- 1) Co tato sekvence kóduje?
- 2) Naleznete isoformy příslušného enzymu a porovnejte je.
- 3) Navrhnete primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoRI.
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.

B: Pracujte s následující sekvencí peptidu:

```
EKHSGGFIFPPSYVFPVVMSELTQMTFMERVKNMIVLYFDWFEEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLIIRNSWNFQFPYPLLPNVDFVGGGLHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRFDGNKPDITLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI
```

- 1) Identifikujte příslušný protein, zapište přístupový kód referenční sekvence.
- 2) Jaká je molekulová hmotnost tohoto peptidu?
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

# Stáhněte si sekvenci NM\_002085.5

---

- 1) Co tato sekvence kóduje?
- 2) Nalezněte isoformy příslušného enzymu a porovnejte je.
- 3) Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoR.I
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.

# Stáhněte si sekvenci NM\_002085.5

---

- 1) Co tato sekvence kóduje? → [NCBI](#)
- 2) Nalezněte isoformy příslušného enzymu a porovnejte je.
- 3) Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoR.I
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.

# Stáhněte si sekvenci NM\_002085.5

---

1)

NUCLEOTIDE SEQUENCE

**Homo sapiens glutathione peroxidase 4 (GPX4), transcript variant 1, mRNA**

Homo sapiens

851 bp mRNA sequence

NM\_002085.5

[FASTA](#) [Gene](#)

# Stáhněte si sekvenci NM\_002085.5

---

- 1) Co tato sekvence kóduje? → NCBI
- 2) Nalezněte **isoformy** příslušného enzymu a porovnejte je. → UNIPROT
- 3) Navrhněte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoR.I
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.



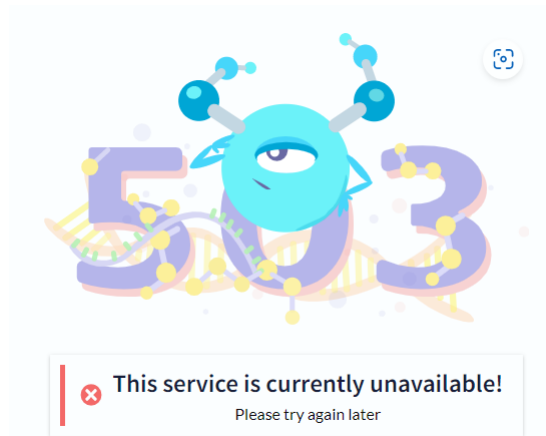
Stáhněte si sekvenci NM\_002085.5

---

# Stáhněte si sekvenci NM\_002085.5

2) nefunguje-li UNIPROT

alternativa: NCBI (úvodní karta)



GENE

[GPX4 – glutathione peroxidase 4](#)

*Homo sapiens (human)*

Also known as: GPx-4, GSHPx-4, MCSP, PHGPx, SMDS, snGPx, sn

Gene ID: 2879

[RefSeq transcripts \(4\)](#) [RefSeq proteins \(4\)](#) [RefSeqGene \(1\)](#)

[Orthologs](#) [Genome Data Viewer](#) [BLAST](#) [Down](#)

Was this helpful?

Items: 4  
Selected: 4

- [phospholipid hydroperoxide glutathione peroxidase isoform A precursor \[Homo sapiens\]](#)  
1. 197 aa protein  
Accession: NP\_002076.2 GI: 75709200  
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [phospholipid hydroperoxide glutathione peroxidase isoform B precursor \[Homo sapiens\]](#)  
2. 227 aa protein  
Accession: NP\_001034936.1 GI: 90903238  
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [phospholipid hydroperoxide glutathione peroxidase isoform C \[Homo sapiens\]](#)  
3. 234 aa protein  
Accession: NP\_001034937.1 GI: 90903240  
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [phospholipid hydroperoxide glutathione peroxidase isoform D \[Homo sapiens\]](#)  
4. 170 aa protein  
Accession: NP\_001354761.1 GI: 1540583210  
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

RefSeq Sequences

Multalin:

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
NP_002076.2	MSLGLRLKPLLCGALRPPGLAGTNCASRDHRCARSHIEFSAKDIDGHAYNLDRYRGVCIYVTVASQGTETVNYTQLVDLHARYEEGG													
NP_001034937.1	NGRRAGSGPGRRRQRCQSRGRRRPPRRRKPACRRRRRRRKRKPCPSRLRPEIHECPKSQDPCASRDHRCARSHIEFSAKDIDGHAYNLDRYRGVCIYVTVASQGTETVNYTQLVDLHARYEEGG													
NP_001354761.1	MSLGLRLKPLLCGALRPPGLAGTNCASRDHRCARSHIEFSAKDIDGHAYNLDRYRGVCIYVTVASQGTETVNYTQLVDLHARYEEGG													
NP_001034936.1	MSLGLRLKPLLCGALRPPGLAGTNCASRDHRCARSHIEFSAKDIDGHAYNLDRYRGVCIYVTVASQGTETVNYTQLVDLHARYEEGG													
Consensus	MSLGLRLKPLLCGALRPPGLAGTNCASRDHRCARSHIEFSAKDIDGHAYNLDRYRGVCIYVTVASQGTETVNYTQLVDLHARYEEGG													
NP_002076.2	LRLRFFPCNFGKDEPGSNEETKEFRAGYVYKFDNFSCIKVNGDRAHPLKANKIQPKGKGLGARLKHNFTKFLIDKNGCVYKRYGPEEPLVIEKDLPHYF													
NP_001034937.1	LRLRFFPCNFGKDEPGSNEETKEFRAGYVYKFDNFSCIKVNGDRAHPLKANKIQPKGKGLGARLKHNFTKFLIDKNGCVYKRYGPEEPLVIEKDLPHYF													
NP_001354761.1	LRLRFFPCNFGKDEPGSNEETKEFRAGYVYKFDNFSCIKVNGDRAHPLKANKIQPKGKGLGARLKHNFTKFLIDKNGCVYKRYGPEEPLVIEKDLPHYF													
NP_001034936.1	LRLRFFPCNFGKDEPGSNEETKEFRAGYVYKFDNFSCIKVNGDRAHPLKANKIQPKGKGLGARLKHNFTKFLIDKNGCVYKRYGPEEPLVIEKDLPHYF													
Consensus	LRLRFFPCNFGKDEPGSNEETKEFRAGYVYKFDNFSCIKVNGDRAHPLKANKIQPKGKGLGARLKHNFTKFLIDKNGCVYKRYGPEEPLVIEKDLPHYF													
263	I-I													
NP_002076.2														
NP_001034937.1														
NP_001354761.1														
NP_001034936.1														
RHS														
...														

# Stáhněte si sekvenci NM\_002085.5

---

- 1) Co tato sekvence kóduje? → NCBI
- 2) Nalezněte **isoformy** příslušného enzymu a porovnejte je. → UNIPROT
- 3) **Navrhněte primery**, tak aby bylo **kódující sekvenci** tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoR.I → NCBI / CDS
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.

# Stáhněte si sekvenci NM\_002085.5

---

Kódující sekvence:

## **Homo sapiens glutathione peroxidase 4 (GPX4), transcript variant 1, mRNA**

NCBI Reference Sequence: NM\_002085.5

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

```
>NM_002085.5:51-644 Homo sapiens glutathione peroxidase 4 (GPX4), transcript variant 1,  
mRNA
```

```
ATGAGCCTCGGCCGCCTTTGCCGCCTACTGAAGCCGGCGCTGCTCTGTGGGGCTCTGGCCGCGCCTGGCC  
TGGCCGGGACCATGTGCGCGTCCCAGGACGACTGGCGCTGTGCGCGCTCCATGCACGAGTTTTCCGCCAA  
GGACATCGACGGGCACATGGTTAACCTGGACAAGTACCGGGGCTTCGTGTGCATCGTCACCAACGTGGCC  
TCCCAGTGAGGCAAGACCGAAGTAACTACACTCAGCTCGTCGACCTGCACGCCGATACGCTGAGTGTG  
GTTTGGCGATCCTGGCCTTCCCCTGTAACAGTTCGGGAAGCAGGAGCCAGGGAGTAACGAAGAGATCAA  
AGAGTTCGCCGCGGGCTACAACGTCAAATTCGATATGTTTCAGCAAGATCTGCGTGAACGGGGACGACGCC  
CACCCGCTGTGGAAGTGGATGAAGATCCAACCAAGGGCAAGGGCATCCTGGGAAATGCCATCAAGTGGGA  
ACTTACCAAGTTCCTCATCGACAAGAACGGCTGCGTGGTGAAGCGCTACGGACCCATGGAGGAGCCCCCT  
GGTATAGAGAAGGACCTGCCCCACTATTTCTAG
```

# Stáhněte si sekvenci NM\_002085.5

F: TTT GGA TCC ATG AGC CTC GGC C

R: TTT GAA TTC CTA GAA ATA GTG GGG CAG

Oligo Calc: Oligonucleotide Properties Calculator

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

Nucleotide base codes

TTT **GGA TCC** ATG AGC CTC GGC C

Reverse Complement Strand(5' to 3') is:  
GGC CGA GGC TCA TGG ATC CAA A

5' modification (if any) 3' modification (if any) Select molecule  
50 nM Primer 1 Measured Absorbance at 260 nanometers  
50 mM Salt (Na<sup>+</sup>)

Calculate Swap Strands BLAST mfold

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

Length: 22 Molecular Weight: 6702.4<sup>4</sup> GC content: 59%  
1 ml of a sol'n with an Absorbance of 1 at 260 nm  
is 4.6 microMolar<sup>5</sup> and contains 30.8 micrograms.

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.  
RlnK 33.404 cal/(°K\*mol) deltaH 187.5 Kcal/mol  
deltaG 31.2 Kcal/mol deltaS 487.8 cal/(°K\*mol)

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

Oligo Calc: Oligonucleotide Properties Calculator

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

Nucleotide base codes

TTT **GAA TTC** CTA GAA ATA GTG GGG CAG

Reverse Complement Strand(5' to 3') is:  
CTG CCC CAC TAT TTC TAG GAA TTC AAA

5' modification (if any) 3' modification (if any) Select molecule  
50 nM Primer 1 Measured Absorbance at 260 nanometers  
50 mM Salt (Na<sup>+</sup>)

Calculate Swap Strands BLAST mfold

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

Length: 27 Molecular Weight: 8378.5<sup>4</sup> GC content: 41%  
1 ml of a sol'n with an Absorbance of 1 at 260 nm  
is 3.268 microMolar<sup>5</sup> and contains 27.4 micrograms.

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.  
RlnK 33.404 cal/(°K\*mol) deltaH 213.3 Kcal/mol  
deltaG 33.6 Kcal/mol deltaS 563.2 cal/(°K\*mol)

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

# Stáhněte si sekvenci NM\_002085.5

---

- 1) Co tato sekvence kóduje? → NCBI
- 2) Nalezněte **isoformy** příslušného enzymu a porovnejte je. → UNIPROT
- 3) **Navrhněte primery**, tak aby bylo **kódující sekvenci** tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoR.I → NCBI / CDS
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu. → Restriction summary
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.

# Stáhněte si sekvenci NM\_002085.5

---

Jeden ne, Jeden ano

BamHI g gatcc	288
EcoRI g aatc	none

# Stáhněte si sekvenci NM\_002085.5

---

- 1) Co tato sekvence kóduje? → NCBI
- 2) Nalezněte **isoformy** příslušného enzymu a porovnejte je. → UNIPROT
- 3) **Navrhněte primery**, tak aby bylo **kódující sekvenci** tohoto genu možné vložit do plazmidu mezi restriční endonukleázy BamHI a EcoR.I → NCBI / CDS
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu. → Restriction summary
- 5) Přeložte tuto sekvenci do proteinu a **identifikujte 100. aminokyselinu**, zapište mutaci této AMK na alanin.



# Stáhněte si sekvenci NM\_002085.5

Translate: `>rf 1 Untitled`  
MSLGRLCRLLPALLCGALAAPGLAGTMCASRDDWRCARSMHEFSAKDIDGHMVNL DKYR  
GFVCIVTNVASQ\*GKTEVNYTQLVDLHARYAECGLRILAFPCNQFGKQEPGSNEEIKEFA  
AGYNVKFDMFSKICVNGDDAHPLWKWMKIQPKGKILGNAIKWNFTKFLIDKNGCVVKRY  
GPMEEPLVIEKDLPHYF\*

→ F

→ P

Sekvence z databáze:

MSLGRLCRLLPALLCGALAAPGLAGTMCASRDDWRCARSMHEF  
SAKDIDGHMVNL DKYR GFVCIVTNVASQUGKTEVNYTQLVDL  
NQFGKQEPGSNEEIKEFAAGYNVKFDMFSKICVNGDDAHPLW  
IKWNFTKFLIDKNGCVVKRYGPMEEPLVIEKDLPHYF

## Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or ranges. The residues can be specified as numbers, as ranges, 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

```
MSLGRLCRLLPALLCGALAAPGLAGTMCASRDDWRCARSMHEF
SAKDIDGHMVNL DKYR GFVCIVTNVASQUGKTEVNYTQLVDLHARYAECGLRILAFPC
NQFGKQEPGSNEEIKEFAAGYNVKFDMFSKICVNGDDAHPLWKWMKIQPKGKILGNA
IKWNFTKFLIDKNGCVVKRYGPMEEPLVIEKDLPHYF
```

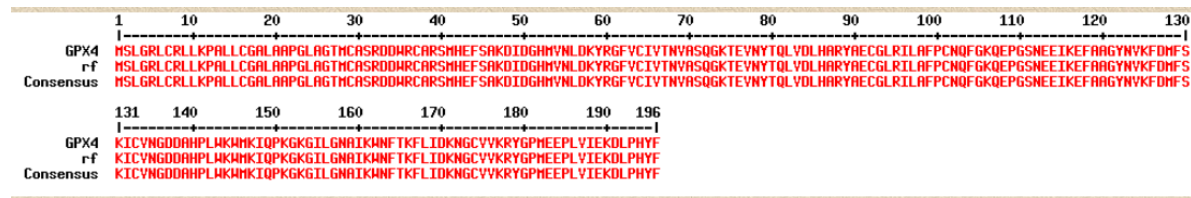
Sequence Manipulation Suite - Pracovní - Microsoft Edge

Enter the residue number, a range, or a range

Range Extractor Protein results

100  
>results for 196 residue sequence "Untitled" starting "MSLGRLCRLLP" p

Porovnání:



Mutace: F100A  
(Nebo P100A-správně)

# Zkoušková písemka:

---

2 úlohy, každá 5 bodů

Verze 4.5.2021

Hodnocení: max celkem 10 bodů

9-10b: 1

7-8b: 2

5-6b: 3

A: Stáhněte si sekvenci NM\_002085.5

- 1) Co tato sekvence kóduje?
- 2) Nalezňte isoformy příslušného enzymu a porovnejte je.
- 3) Navrhňte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy BamHI a EcoR.I
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, запиšte mutaci této AMK na alanin.

B: Pracujte s následující sekvencí peptidu:

```
EKHSGGFIFFPSYVFPVVMSELTQMTFMERVKNMIVLYFDWFEEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLIIRNSWNFOFPYPLLPNVDFVGGGLHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRFDGNKPDITLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI
```

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence.
- 2) Jaká je molekulová hmotnost tohoto peptidu?
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYFDWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLIRNSWNFQFPYPLLPNVDFVGGHLCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRFDGNKPDTLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI

# Pracujte s následující sekvencí peptidu:


---

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence.
- 2) Jaká je molekulová hmotnost tohoto peptidu?
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYFDWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLIRNSWNFQFPYPLLPNVDFVGGHLCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIQKVLWRFDGNKPDTLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI

# Pracujte s následující sekvencí peptidu:

---

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence.  NCBI / BLASTp
- 2) Jaká je molekulová hmotnost tohoto peptidu?
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

# Pracujte s následující sekvencí peptidu:

## BLASTp

BLAST® » blastp suite » results for RID-4DD7SN7Z01N

Home Recent Results Saved Strategies Help

[< Edit Search](#) Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title **Protein Sequence**

RID [4DD7SN7Z01N](#) Search expires on 04-26 00:57 am [Download All](#) ▾

Program **BLASTP** [Citation](#) ▾

Database **nr** [See details](#) ▾

Query ID [Ic|Query\\_573122](#)

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](#)

Compare these results against the new Clustered nr database ? [BLAST](#)

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download ▾ Select columns ▾ Show 100 ▾ ?

select all 0 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input type="checkbox"/>	<a href="#">UGT2B7 [synthetic construct]</a>	<a href="#">synthetic construct</a>	431	431	100%	7e-148	100.00%	529	<a href="#">AIC62044.1</a>
<input type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B7 isoform X1 [Pan paniscus]</a>	<a href="#">Pan paniscus</a>	431	431	100%	9e-148	100.00%	529	<a href="#">XP_003805577.2</a>
<input type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B7 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	431	431	100%	1e-147	100.00%	529	<a href="#">XP_526600.2</a>
<input type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B7 isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	431	431	100%	1e-147	100.00%	529	<a href="#">NP_001065.2</a>
<input type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B7 isoform X2 [Pan paniscus]</a>	<a href="#">Pan paniscus</a>	430	430	100%	3e-147	99.51%	529	<a href="#">XP_054967441.1</a>
<input type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase (EC 2.4.1.17) [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	429	429	100%	5e-147	99.51%	529	<a href="#">AAA36793.1</a>

### UDP-glucuronosyltransferase 2B7 isoform 1 precursor [Homo sapiens]

NCBI Reference Sequence: NP\_001065.2

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

Job Title **Protein Sequence**

RID [4DCP9K3201N](#) Search expires on 04-26 00:48 am [Download All](#) ▾

Program **BLASTP** [Citation](#) ▾

Database **refseq\_protein** [See details](#) ▾

Query ID [Ic|Query\\_436865](#)

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](#)

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download ▾ Select columns ▾ Show 100 ▾ ?

BLAST® » blastp suite » results for RID-4DD2E536016

Home Recent Results Saved Strategies Help

[< Edit Search](#) Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title **Protein Sequence**

RID [4DD2E536016](#) Search expires on 04-26 00:54 am [Download All](#) ▾

Program **BLASTP** [Citation](#) ▾

Database **refseq\_select\_prot** [See details](#) ▾

Query ID [Ic|Query\\_13208](#)

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](#)

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download ▾ Select columns ▾ Show 100 ▾ ?



select all 100 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B7 isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	431	431	100%	1e-148	100.00%	529	<a href="#">NP_001065.2</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B10 isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	408	408	100%	2e-139	92.16%	528	<a href="#">NP_001066.1</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B4 isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	405	405	100%	1e-138	92.16%	528	<a href="#">NP_069962.2</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B11 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	397	397	100%	4e-135	88.73%	529	<a href="#">NP_001064.1</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B28 isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	394	394	100%	4e-134	88.73%	529	<a href="#">NP_444267.1</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B15 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	380	380	100%	2e-128	84.31%	530	<a href="#">NP_001067.2</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B17 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	380	380	100%	2e-128	84.31%	530	<a href="#">NP_001068.1</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B31-like isoform X1 [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	360	360	100%	6e-121	79.90%	530	<a href="#">XP_003751392.1</a>
<input checked="" type="checkbox"/>	<a href="#">UDP-glucuronosyltransferase 2B10 precursor [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	356	356	100%	3e-119	79.41%	532	<a href="#">NP_001178605.1</a>

EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYFDWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLI RNSWNFQFPYPLLPNVDFVGGHLCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRF DGNKPD T LGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI

# Pracujte s následující sekvencí peptidu:

---

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence.  NCBI / BLASp
- 2) Jaká je molekulová hmotnost **tohoto peptidu**?  SMS: Protein molecular Weight
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

# Pracujte s následující sekvencí peptidu:

---

## Protein Molecular Weight

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

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

```
EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYDFWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLIRNSWNFQFPYLLPNVDFVGGHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRFDGKPDTLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI
```

Submit

Clear

R

• Add  copy

\*This page requires

\*You can mirror this

Sequence Manipulation Suite – Pracovní – Microsoft Edge

about:blank

Protein Molecular Weight results

**Results for 204 residue sequence "Untitled" starting "EKHSGGFIFP"**

**23.36 kDa**

EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYFDWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLI RNSWNFQFPYPLLPNVDFVGG LHC KPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRF DGNKPD TLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI

# Pracujte s následující sekvencí peptidu:

---

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence. → NCBI / BLASTp
- 2) Jaká je molekulová hmotnost tohoto peptidu? → SMS: Protein molecular Weight
- 3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky? → Různé (signalP, SMART; TMHMM, ...)
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?



# Pracujte s následující sekvencí peptidu:

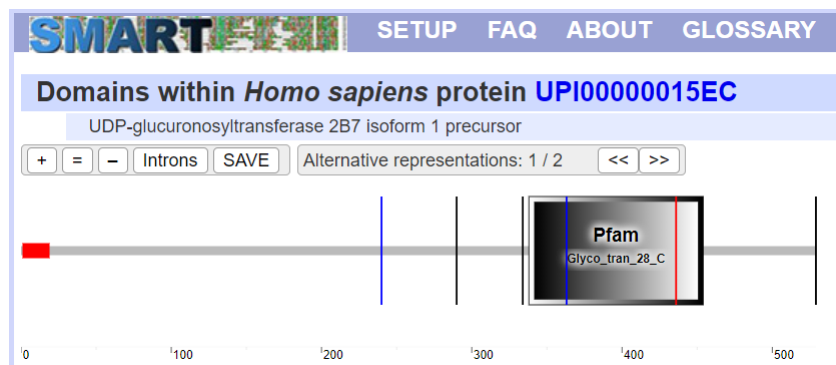
Signální peptid obsahuje.

TM úsek obsahuje

## UDP-glucuronosyltransferase 2B7 isoform 1 precursor [Homo sapiens]

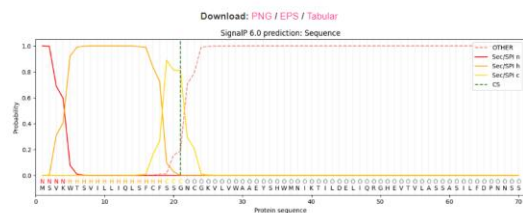
NCBI Reference Sequence: NP\_001065.2

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

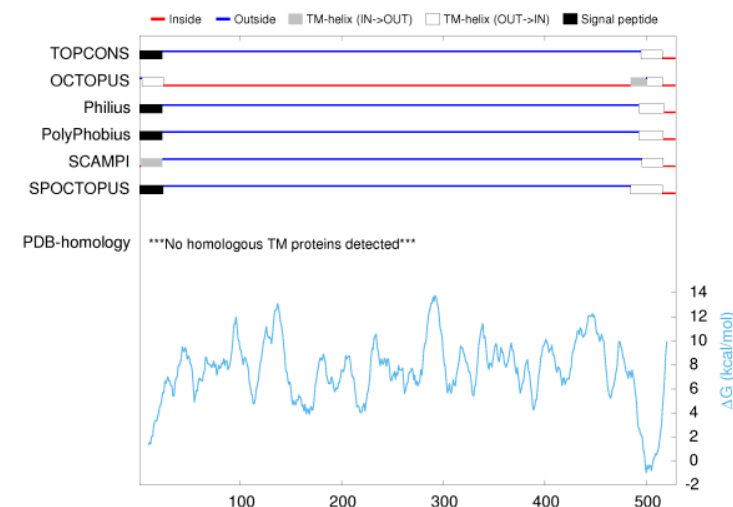


Sequence  
**Prediction:** Signal Peptide (Sec/SPI)  
 Cleavage site between pos. 21 and 22.  
 Probability 0.812238

Protein type	Other	Signal Peptide (Sec/SPI)	Lipoprotein signal peptide (Sec/SPII)	TAT signal peptide (Tat/SPI)	TAT Lipoprotein signal peptide (Tat/SPII)	Pilin-like signal peptide (Sec/SPIII)
Likelihood	0.0003	0.9991	0.0002	0.0001	0.0001	0.0001

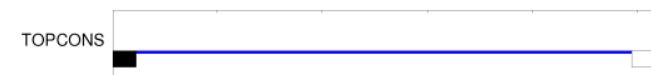


Predicted topologies and predicted  $\Delta G$  values:



[High-resolution image](#)

Consensus prediction (TOPCONS):



EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYFDWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLI RNSWNFQFPYPLLPNVDFVGGGLHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRF DGNKPD T LGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI

# Pracujte s následující sekvencí peptidu:

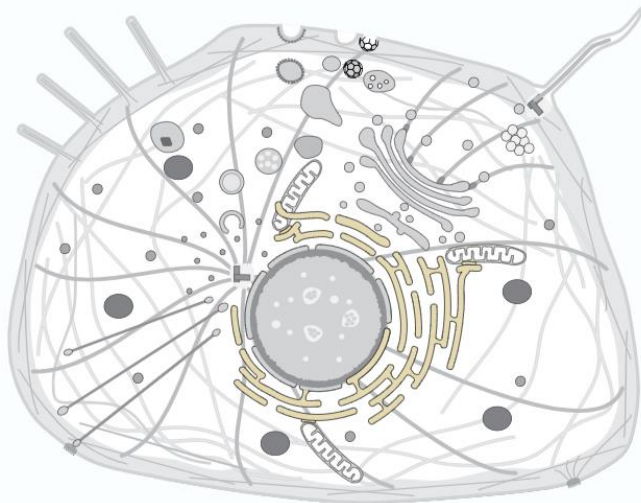
---

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence. → NCBI / BLASTp
- 2) Jaká je molekulová hmotnost tohoto peptidu? → SMS: Protein molecular Weight
- 3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky? → Různé (signalP, SMART; TMHMM, ...)
- 4) Kde je tento protein v buňce lokalizován? → UNIPROT
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

# Pracujte s následující sekvencí peptidu:

## Subcellular Location<sup>i</sup>

UniProt Annotation   GO Annotation



📍 **Endoplasmic reticulum membrane** 📄 1 Publication : Single-pass membrane protein 📄 Sequence Analysis

EKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVLYFDWFEIFDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLI RNSWNFQFPYPLLPNVDFVGGGLHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRF DGNKPD T LGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI

# Pracujte s následující sekvencí peptidu:

---

- 1) Identifikujte příslušný protein, запиšte přístupový kód referenční sekvence. → NCBI / BLASTp
- 2) Jaká je molekulová hmotnost tohoto peptidu? → SMS: Protein molecular Weight
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky? → Různé (signalP, SMART; TMHMM, ...)
- 4) Kde je tento protein v buňce lokalizován? → UNIPROT
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen? → Gene/ Gene Expression

# Pracujte s následující sekvencí peptidu:

## V ledvinách a játrech

