

Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2023/2024

3/10

Proteinová bioinformatika II

Cíle:

Student bude schopen odhalit typické domény a transmembránové úseky v zadané sekvenci proteinu. Bude schopen vyhledat podobné sekvence a na základě podobnosti bude schopen odhadnout funkci neznámého proteinu.

shrnutí - Rešeršní projekt: NQO1

Vyhledejte dostupné informace o NQO1, získejte co nejvíce literárních, sekvenčních, případně i strukturních a dalších údajů o tomto genu/proteinu.

Nalezněte podobné sekvence a porovnejte je na proteinové úrovni.

- **NAD(P)H:chinonoxidoreduktasa / reference**
- **Protein: NP_000894, P15559 / 274 AMK; 30,8 kDa; trypsin štěpí 33x (59AMK nejdelší peptid)...**

„Proteinová bioinformatika“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod...)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce funkce → **Hledání známých motivů**

3D-struktura, vizualizace

Vyhledání a porovnání podobných sekvencí

Evoluční příbuznost sekvencí

...

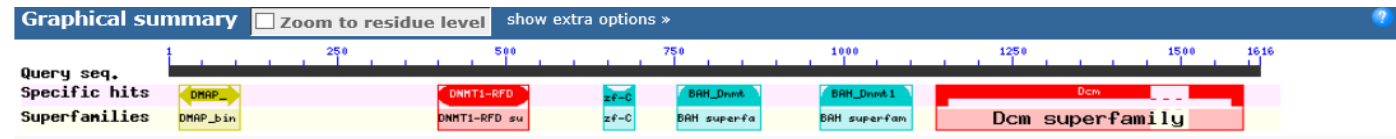
Hledání známých motivů

SEKVENCE ⇌ STRUKTURA ⇌ FUNKCE

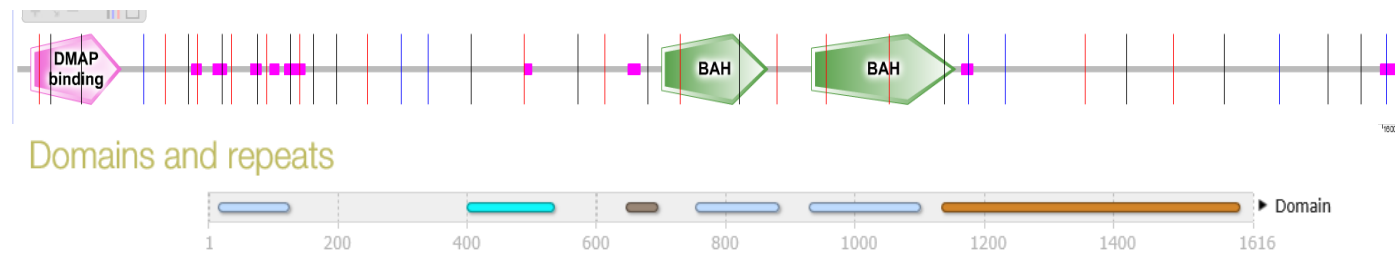
Sekvenční motivy, signální peptidy, místa posttranslačních modifikací ..

- Databáze známých motivů

NCBI/CDD

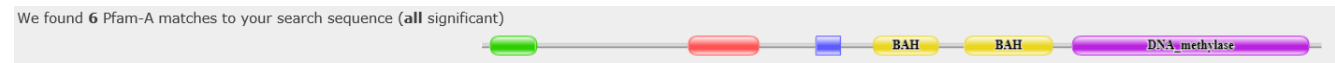


SMART



EMBL/InterPro

Pfam



Pfam



[HOME](#) | [SEARCH](#) | [BROWSE](#) | [FTP](#) | [HELP](#) | [ABOUT](#)



Pfam data and new releases are available through [InterPro](#)

The Pfam website now serves as a **static page with no data updates**. All links below redirect to the closest alternative page in the InterPro website.

Pfam 36.0 (20 795 entries, 659 clans)

The Pfam database is a large collection of protein families, each represented by *multiple sequence alignments* and *hidden Markov models (HMMs)*. [More...](#)

InterPro



InterPro

Classification of protein families



Home

Search

Browse

Results

Release notes

Download

Help

About



Classification of protein families

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium. We combine protein signatures from these member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool.

▼ Citing InterPro

If you find InterPro useful, please cite the reference that describes this work:

Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. [InterPro in 2022](#). *Nucleic Acids Research*, Nov 2022, (doi: 10.1093/nar/gkac993)

96.0

InterPro 96.0
14 September 2023

Search by sequence

Search by text

Search by Domain Architecture

Sequence, in FASTA format

Enter your sequence

Hledání známých motivů-CD (NCBI)

CD-Search Results: Concise Display

shows only the best scoring domain model for each region on the query sequence

Identifikátor sekvence

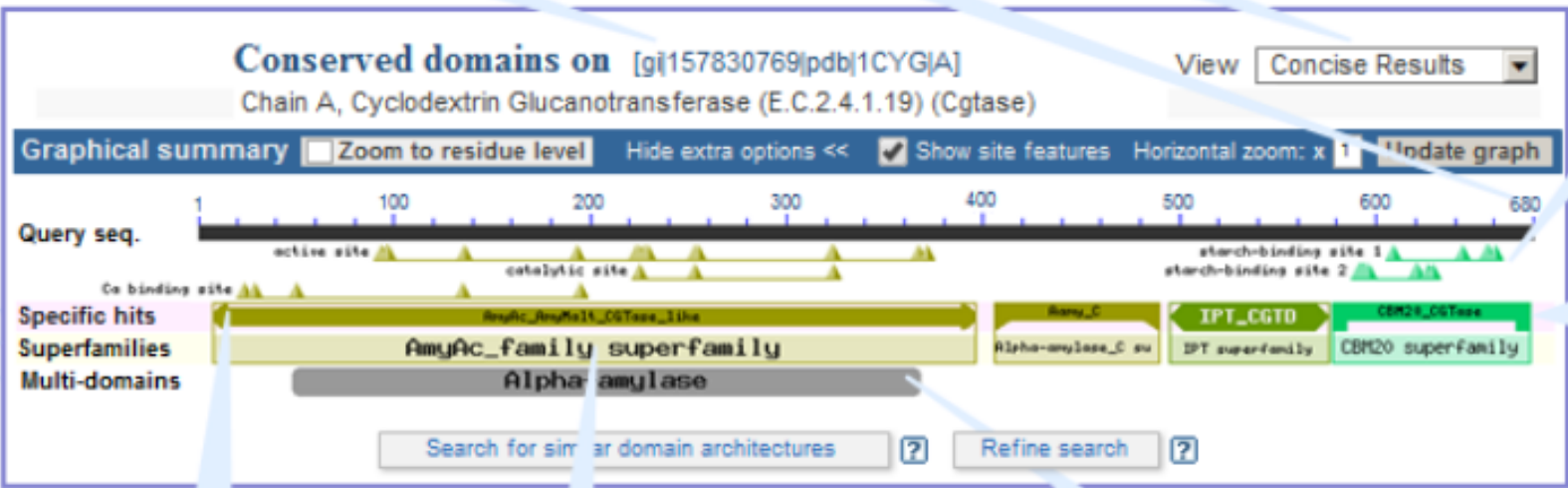
Délka sekvence

Menu pro omezení úrovně zobrazovaných detailů

Malé trojúhelníky
Označují aminokyseliny v konzervovaných oblastech/důležitá katalytická nebo vazebná místa

Hit types vary based on confidence level and specificity.

Follow the text links below this illustration for more information about each hit type.



Barevné obdélníky
Označují nalezené domény (po najetí kurzoru se objeví detailní popis)

„Specifické shody“
Jsou zvýrazněné a označují nalezené domény s velkou pravděpodobností

„Superrodiny“
Označují do které rodiny nalezené specifické shody patří

Jsou –li nalezeny jen „**Ne-Specifické shody**“ Jsou viditelné pouze při plném menu (ne concise vpravo nahoře)

Multi-domény
Označují nalezené domény, které pravděpodobně obsahují více domén

Conserved

How to use C

- Identify
- Identify
- Identify
- View a
- Interac
- Find oth
- Interact

Out
help

Hledání známých motivů / NQO1



Conserved domains on [gi|118607|sp|P15559.1|NQO1_HUMAN]

View Concise Results ?

[Sign Out](#)

RecName: Full=NAD(P)H dehydrogenase [quinone] 1; AltName: Full=Azoreductase; AltName: Full=DT-diaphorase; Short=DTD; AltName: Full=Menadione reductase; AltName: Full=NAD(P)H:quinone oxidoreductase 1; AltName: Full=Phylloquinone reductase; AltName: Full=Quinone reductase 1; Short=QR1

[Help](#)

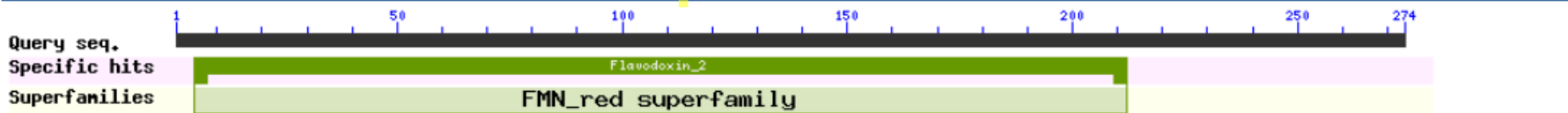
Protein Classification

flavodoxin family protein (domain architecture ID 10495002)

flavodoxin family protein containing a flavodoxin-like fold domain, similar to Bradyrhizobium diazoefficiens FMN-dependent NADH-azoreductase 1, which catalyzes the reductive cleavage of the azo bond in aromatic azo compounds to the corresponding amine

Graphical summary

Zoom to residue level [show extra options](#) ?



Enter prote

[Search for similar domain architectures](#) ?

[Refine search](#) ?

List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	Flavodoxin_2	pfam02525	Flavodoxin-like fold; This family consists of a domain with a flavodoxin-like fold. The family ...	5-212	4.06e-46

References:

- Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", **Nucleic Acids Res.**45(D)200-3.
- Marchler-Bauer A et al. (2015), "CDD: NCBI's conserved domain database.", **Nucleic Acids Res.**43(D)222-6.
- Marchler-Bauer A et al. (2011), "CDD: a Conserved Domain Database for the functional annotation of proteins.", **Nucleic Acids Res.**39(D)225-9.
- Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", **Nucleic Acids Res.**32(W)327-331.

Referenc

Hledání známých motivů-SMART

The screenshot shows the SMART website interface. At the top left is the SMART logo with a protein structure visualization. Below it are two citations: Letunic et al. (2017) *Nucleic Acids Res* doi: 10.1093/nar/gkx922 and Letunic et al. (2020) *Nucleic Acids Res* doi: 10.1093/nar/gkaa937. To the right of the logo is a 'SMART MODE:' header box with 'NORMAL' highlighted in red and 'GENOMIC' in blue. Further right is a vertical menu with 'Simple', 'Modular', 'Architecture', 'Research', and 'Tool' options. At the top right is a search bar with 'keywords...' and a 'Search SMART' button. Below the logo is a navigation menu with 'HOME', 'SETUP', 'FAQ', 'ABOUT', 'GLOSSARY', 'WHAT'S NEW', and 'FEEDBACK'. The main heading is 'Select your default SMART mode'. The text explains that SMART can be used in 'normal' or 'genomic' modes, with 'Normal SMART' using Swiss-Prot, SP-TREMBL, and stable Ensembl proteomes, and 'Genomic SMART' using proteomes of completely sequenced genomes. It notes that the protein database in Normal SMART has significant redundancy. A section titled 'Different color schemes are used to easily identify the mode you're in.' shows two preview boxes: 'Normal mode' (blue background) and 'Genomic mode' (red background). The 'Normal mode' box is circled in orange. Below this is the instruction 'Click on the images above to select your default mode.' and a note that information about the selected mode is stored in a browser cookie. At the bottom, another SMART logo and 'SMART MODE:' header are shown, with 'NORMAL' circled in red and 'GENOMIC' in blue. The 'SETUP' link in the navigation menu is also circled in red.

Letunic et al. (2017) *Nucleic Acids Res* doi: 10.1093/nar/gkx922
Letunic et al. (2020) *Nucleic Acids Res* doi: 10.1093/nar/gkaa937

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

Select your default SMART mode

You can use SMART in two different modes: **normal** or **genomic**. The main difference is in the underlying protein database used. In **Normal SMART**, the database contains Swiss-Prot, SP-TREMBL and stable Ensembl proteomes. In **Genomic SMART**, only the proteomes of completely sequenced genomes are used; Ensembl for metazoans and Swiss-Prot for the rest. The complete list of genomes in Genomic SMART [is available here](#).

The protein database in Normal SMART has significant redundancy, even though identical proteins are removed. If you use SMART to explore domain architectures, or want to find exact domain counts in various genomes, consider switching to **Genomic** mode. The numbers in the domain annotation pages will be more accurate, and there will not be many protein fragments corresponding to the same gene in the architecture query results. Remember you are exploring a limited set of genomes, though.

Different color schemes are used to easily identify the mode you're in.

Normal mode	Genomic mode
SMART MODE: NORMAL GENOMIC	SMART MODE: NORMAL GENOMIC
Simple Modular Architecture Research Tool	Simple Modular Architecture Research Tool

Click on the images above to select your default mode.

Information about your selected mode is stored in a browser cookie. If you for whatever reason don't want/can't use cookies, access SMART [through this page](#).

You can easily change modes later, by clicking on the links in the 'SMART MODE' header box, or in your personal preference settings ('SETUP' link in the menu):

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5857-5864
Letunic et al. (2004) *Nucleic Acids Res* **32**, D142-D144

Hledání známých motivů-SMART

SMART
Letunic et al. (2017) *Nucleic Acids Res* doi: 10.1093/nar/gkx922
Letunic et al. (2020) *Nucleic Acids Res* doi: 10.1093/nar/gkaa937

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

SMART MODE:
NORMAL
GENOMIC

Simple
Modular
Architecture
Research
Tool

keywords...
Search SMART

Sequence analysis

You may use either a [Uniprot/Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to perform the SMART analysis service.

Sequence ID or ACC

Examples: #1, #2 ?

Protein sequence

Examples: #1, #2 ?

Sequence SMART **Reset**

HMMER searches of the SMART database occur by default. You may also find:

- Outlier homologues and homologues of known structure
- PFAM domains
- signal peptides
- internal repeats

Architecture analysis

You can search for proteins with combinations of [specific domains](#) in different species or taxonomic ranges. You can input the domains directly into "Domain selection" box, or use "GO terms query" to get a list of domains.

Domain selection

Examples: #1, #2 ?

GO terms query

Examples: #1, #2 ?

Taxonomic selection

If you wish to restrict your domain architecture query to a particular species or taxonomic class, start typing its name in the box, and select a match from the popup list.

Architecture query **Resetovat**

You can try an [Advanced Query](#) if you're familiar with SQL.

Hledání známých motivů-SMART

SMART MODE: Simple

keywords... Search SMART

Domains within *Homo sapiens* protein **NQO1_HUMAN** (P15559)

NAD(P)H dehydrogenase [quinone] 1; The enzyme apparently serves as a quinone reductase in connection with conjugation reactions of hydroquinons involved in detoxification pathways as well as in biosynthetic processes such as the vitamin K-dependent gamma-carboxylation of glutamate residues in prothrombin synthesis; Belongs to the NAD(P)H dehydrogenase (quinone) family.

+ = - Introns SAVE Alternative representations: 1 / 2 << >>

0 100 200

Information Architecture Interactions Pathways PTMs Orthology

Length 274 aa

Source database UniProt

Identifiers NQO1_HUMAN, 9606.ENSPO0000319788, P15559, ENSPO0000319788.5, ENSPO0000319788, B2R5Y9, B4DNM7, B7ZAD1, Q86UK1, H3BNV2_HUMAN, H3BNV2, K7BKZ6_PANTR, K7BKZ6, A0A2I2Y180_GORGO, A0A2I2Y180, A0A2J8Q3V7_PANTR, A0A2J8Q3V7, H2QBF4_PANTR, H2QBF4, G3QL89_GORGO, G3QL89

Source gene ENSG00000181019

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by **SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity**. In either case, features not shown in the above diagram are marked as **'overlap'** in the right side table below.

Hledání známých motivů-SMART

SMART SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK keywords... Search SMART

Domains within *Homo sapiens* protein NQO1_HUMAN (P15559)

NAD(P)H dehydrogenase [quinone] 1; The enzyme apparently serves as a quinone reductase in connection with conjugation reactions of hydroquinons involved in detoxification pathways as well as in biosynthetic processes such as the vitamin K-dependent gamma-carboxylation of glutamate residues in prothrombin synthesis; Belongs to the NAD(P)H dehydrogenase (quinone) family.

+ = - Introns SAVE Alternative representations: 1 / 2 << >>

Pfam FMN_red

Information Architecture Interactions Pathways PTMs Orthology

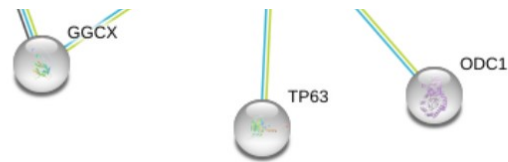
Posttranslational modifications

PTM annotation is taken from [PTMcode](#), a resource of known and predicted functional associations between protein posttranslational modifications (PTMs).

There are **19** PTMs annotated in this protein:

PTM	Count
Ubiquitination	14
Acetylation	3
Phosphorylation	2

To see the full details, including possible functional associations between the PTMs, please visit the [PTMcode annotation page for protein NQO1](#).



Hledání známých motivů-InterPro (2023)

The screenshot displays the InterPro interface for a specific protein job. The top navigation bar includes the InterPro logo and the text "Classification of protein families". The main navigation menu has options for Home, Search, Browse, Results (selected), Release notes, Download, Help, and About. The job details section shows the Job ID as "iprscan5-R20230301-115158-0009-77866479-p1m", a length of 274 amino acids, a status of "finished", and an expiration date of "Wed Mar 08 2023".

The "Protein family membership" section indicates "None predicted". Below this, the "Entry matches to this protein" section shows a sequence alignment from position 1 to 274. A tooltip for the "Flavodoxin_fold" motif (InterPro IPR003680, residues 5-211) is overlaid on the alignment. The alignment shows several matches, including a full match to IPR003680 (PF02525) and partial matches to IPR029039 (G3DSA:3.40.50.360, SSF52218) and PTHR10204.

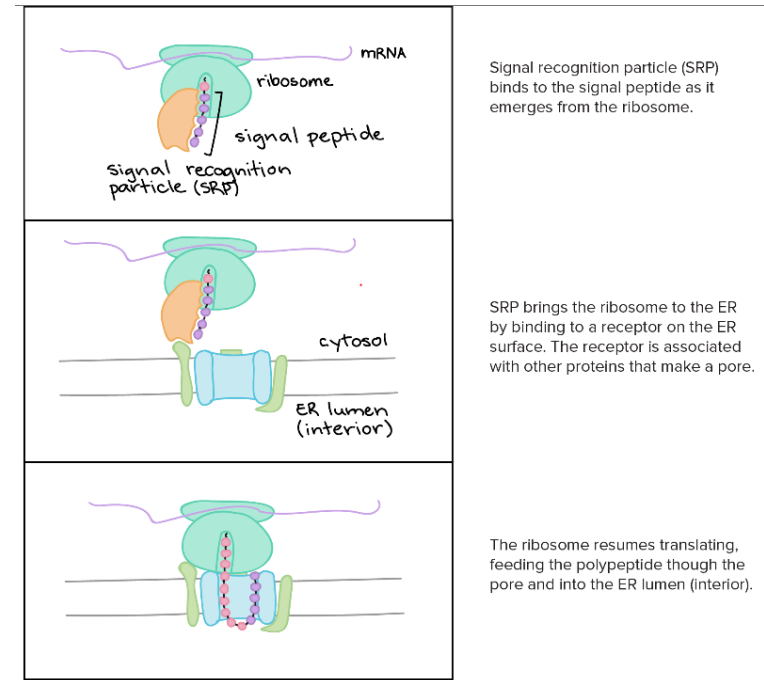
Category	Match	Residues
Domain	IPR003680 PF02525	1-211
Homologous Superfamily	IPR029039 G3DSA:3.40.50.360 SSF52218	1-211
Unintegrated	G3DSA:3.40.50.360:FF:000029	1-211
Unintegrated	PTHR10204	1-211

Praktická ukázka

- otevřete si soubor/stránku s vaší sekvencí
- otevřete si odkazy (každý v novém okně) a hledejte sekvenční motivy
- porovnejte výstupy z jednotlivých programů

Hledání signálních peptidů

Signální peptid pro ER: 15-60 aminokyselin na N-konci proteinů



Hledání signálních peptidů

SignalP

DTU Health Tech

Research Publications Education Collaboration Services and Products News About

SignalP 6.0 is based on a [transformer protein language model](#) with a conditional random field for structured prediction.

Behind the Paper: Check out the [blog post about the SignalP 6.0 publication](#) in the Nature Portfolio Bioengineering Community.

History paper: Click here to read "[A Brief History of Protein Sorting Prediction](#)", The Protein Journal, 2019

Eukaryotic proteins: Remember, the presence or absence of a signal peptide is not the whole story about the localization of a protein! If you want to find out more about the sorting of your eukaryotic proteins, try the protein subcellular localization predictor [DeepLoc](#). You may also want to check whether proteins with signal peptides have GPI anchors that keep them attached to the outer face of the plasma membrane using the predictor [NetGPI](#).

Submission

Instructions

Data

Article abstract

FAQ

Version history

Portable

Downloads

Submit data

Sequence submission: paste the sequence(s) and/or upload a local file

Protein sequences should be not less than 10 amino acids. The maximum number of proteins is 5000.

The long output format might timeout for more than 100 entries.

Mirror: Use SignalP 6.0 on BioLib if this server is heavily loaded.

```
>NP_000894.1 NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDI TGK LK
DPANFQYPA
ESVLAYKEGHLSPDIVAEQKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMY
DKGPFRS
KKAVLSITGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQI
LEGWKK
RLENIWDETPLYFAPSSFLDLNFQAGFLMKKEVQDEEKKKFGLSVGHHLGKSIPTDNQIK
ARK|
```



Hledání signálních peptidů

SMART

Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-5864
Letunic et al. (2014) *Nucleic Acids Res* doi: 10.1093/nar/gku949

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

Sequence analysis

You may use either a [Uniprot/Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to perform the SMART service.

Sequence ID or ACC

Examples: #1, #2

Protein sequence

```
MVRAATVAAAWLLWAAACAQQEQDFYDFKAVNIRGKLVLSLEKYRGSVSLVNVNVA  
SECGFT  
DQHYRALQQLQRLDLPHHFNVLAFFPCNFGQQEPDSNKEIESFARRTYSVSFFPM  
FSKIIV  
TGIGAHFPAFKYLAQTSGKEPTWNEFKYLVAPDGKVVGAWDFTVSVEEVRPQITA  
LVRKLI  
LLKREDL
```

Examples: #1, #2

Sequence SMART Reset

HMMER searches of the SMART database occur by default. You may also find:

- Outlier homologues and homologues of known structure
- PFAM domains
- signal peptides
- internal repeats

Simple Modular Architecture Research Tool

SMART MODE: NORMAL GENOMIC

Pfam Glyco_tran_28_C

0 100 200 300 400 500

Architecture

Domain architecture analysis

Display all proteins with similar:

- [Domain organisation](#): Proteins having all the domains as the query in the same order. Additional domains are allowed.
- [Domain composition](#): Proteins with the same domain composition have at least one copy of each of domains of the query

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > I the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
signal peptide	1	18	N/A
Pfam:UDPGT	19	523	8e-64

Praktická ukázka

-podívejte se, jestli váš protein obsahuje signální peptid

„Proteinová bioinformatika II“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod...)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce funkce, hledání známých motivů a signálních peptidů

3D-struktura, vizualizace

Vyhledání a porovnání podobných sekvencí

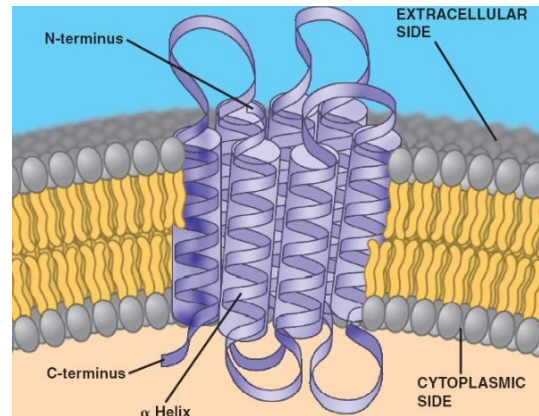
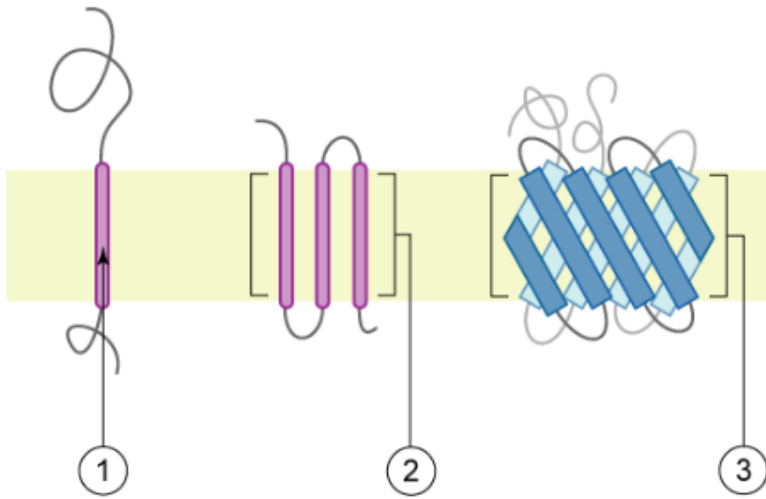
Evoluční příbuznost sekvencí

...

Predikce transmembránových úseků

Hydrofobicita aminokyselin

- různé programy – různé algoritmy výpočtů - různé výstupy
- Předpověď topologického uspořádání (odhad, které části budou v cytoplazmě a které venku)



Predikce transmembránových úseků

Profil - Hydrof **Window size: 21** cale output for user_sequence



ProtScale

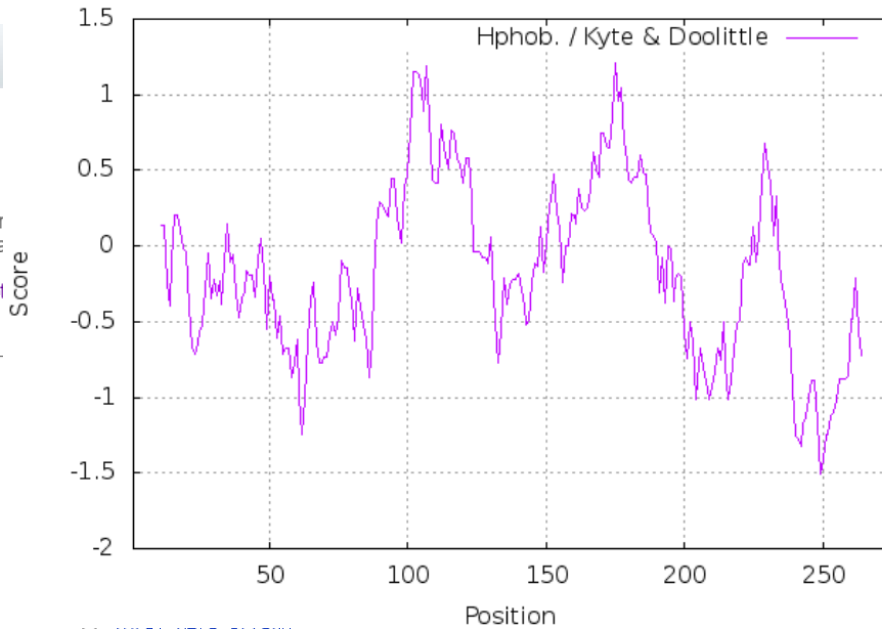
ProtScale [Reference / Documentation]

An amino acid scale is defined by a number but many other scales exist which are based on different properties

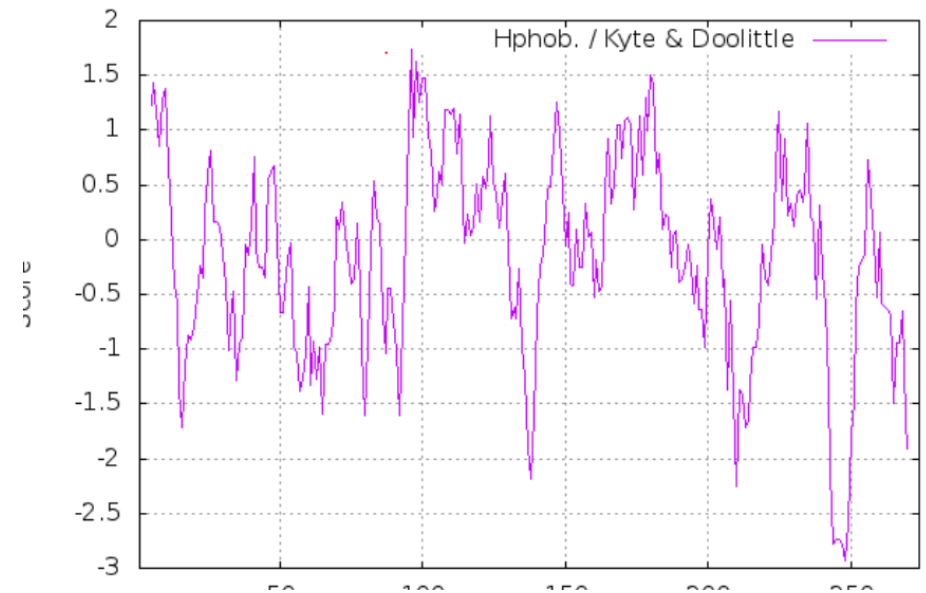
Enter a UniProtKB/Swiss-Prot or UniProt accession number

Or you can paste your own sequence in

- Hphob. / Chothia
- Ratio hetero end/side
- Average flexibility
- beta-sheet / Chou & Fasman
- alpha-helix / Deleage & Roux
- beta-turn / Deleage & Roux
- alpha-helix / Levitt
- beta-turn / Levitt
- Antiparallel beta-strand
- A.A. composition
- Relative mutability



Window size: 9 tput for user_sequence



Window size: 9

Window size: 19 (21)

Relative weight of the window edges compared to the window center (in %): 100

Weight variation model (if the relative weight at the edges is < 100%): linear exponential

Do you want to normalize the scale from 0 to 1? yes no

If you need more information about how to set these parameters, please click [here](#).

Submit Reset

Predikce transmembránových úseků

Profil - Hydrof **Window size: 21** scale output for **Cut off: 1.5 (1.6)**



ProtScale

ProtScale [Reference / Documentation]

An amino acid scale is defined by a number but many other scales exist which are based on different properties.

Enter a UniProtKB/Swiss-Prot or UniProt

Or you can paste your own sequence in

- Hphob. / Chothia
- Ratio hetero end/side
- Average flexibility
- beta-sheet / Chou & Fasman
- alpha-helix / Deleage & Roux
- beta-turn / Deleage & Roux
- alpha-helix / Levitt
- beta-turn / Levitt
- Antiparallel beta-strand
- A.A. composition
- Relative mutability

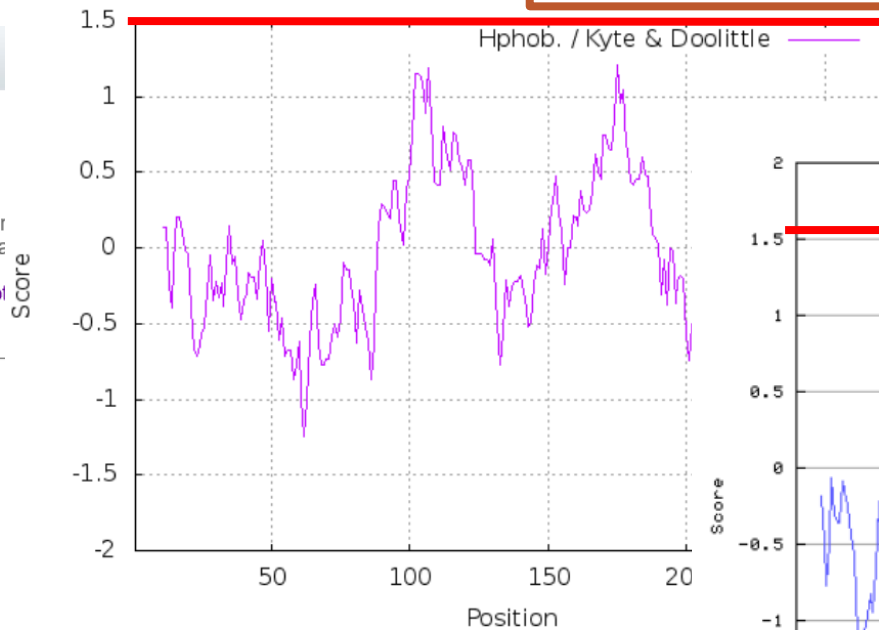
Window size:

Relative weight of the window edges compared to the window center (in %):

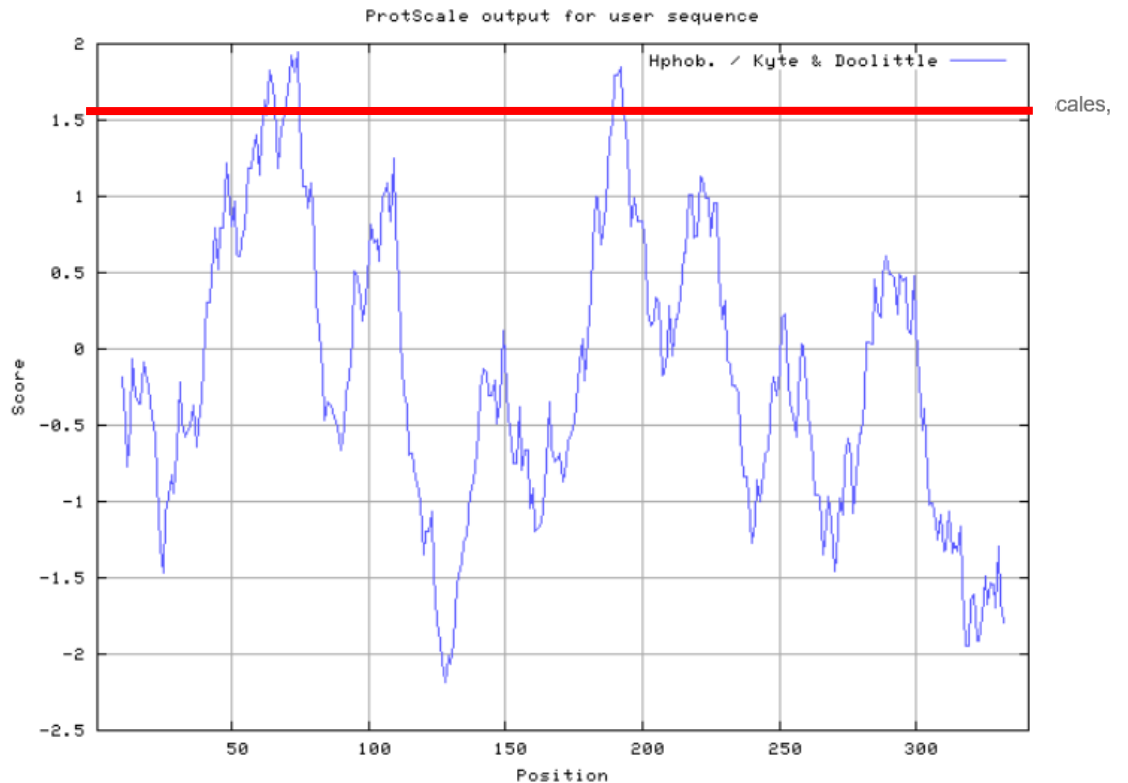
Weight variation model (if the relative weight at the edges is < 100%): linear exponential

Do you want to normalize the scale from 0 to 1? yes no

If you need more information about how to set these parameters, please click [here](#).



Window size: 19 (21)



[Home](#) | [Contact](#)

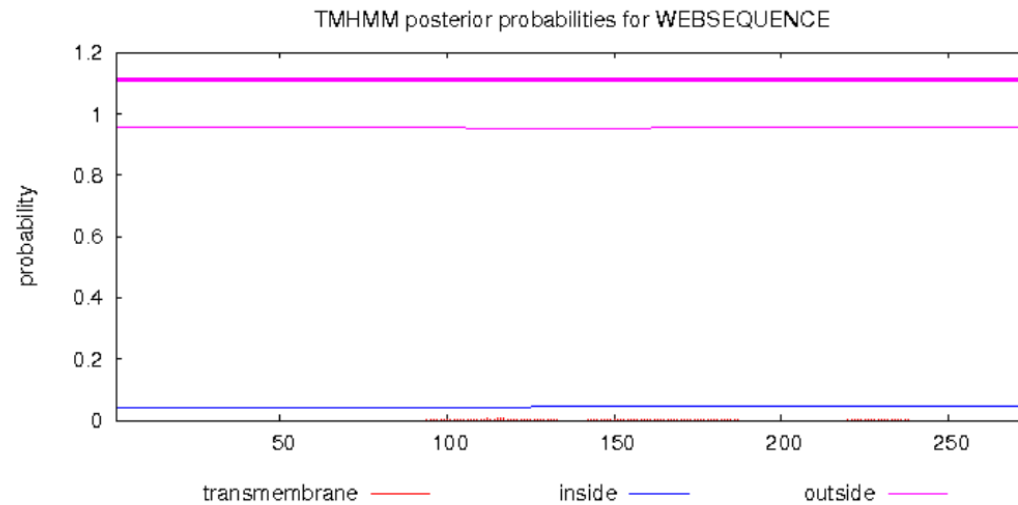
TMHMM

Desaturasa AM158251 (lišaj tabákový)

TMHMM result

[HELP](#) with output formats

```
# WEBSEQUENCE Length: 274
# WEBSEQUENCE Number of predicted TMHs: 0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.20324
# WEBSEQUENCE Exp number, first 60 AAs: 0
# WEBSEQUENCE Total prob of N-in: 0.04315
WEBSEQUENCE    TMHMM2.0    outside    1    274
```



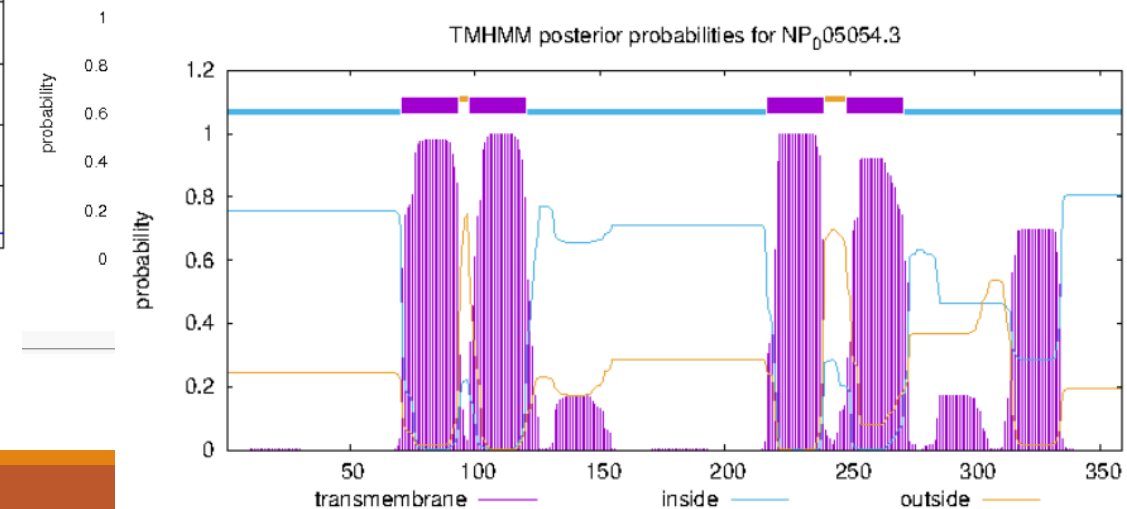
TMHMM result

[HELP](#) with output formats

TMHMM result

```
# WEBSEQUENCE
# WEBSEQUENCE
# WEBSEQUENCE
# WEBSEQUENCE
# WEBSEQUENCE # NP_005054.3 Length: 359
# WEBSEQUENCE # NP_005054.3 Number of predicted TMHs: 4
WEBSEQUENCE # NP_005054.3 Exp number of AAs in TMHs: 104.54198
WEBSEQUENCE # NP_005054.3 Exp number, first 60 AAs: 0.03484
WEBSEQUENCE # NP_005054.3 Total prob of N-in: 0.75464
WEBSEQUENCE NP_005054.3 TMHMM2.0 inside 1 70
WEBSEQUENCE NP_005054.3 TMHMM2.0 TMhelix 71 93
WEBSEQUENCE NP_005054.3 TMHMM2.0 outside 94 97
WEBSEQUENCE NP_005054.3 TMHMM2.0 TMhelix 98 120
WEBSEQUENCE NP_005054.3 TMHMM2.0 inside 121 216
WEBSEQUENCE NP_005054.3 TMHMM2.0 TMhelix 217 239
WEBSEQUENCE NP_005054.3 TMHMM2.0 outside 240 248
WEBSEQUENCE NP_005054.3 TMHMM2.0 TMhelix 249 271
WEBSEQUENCE NP_005054.3 TMHMM2.0 inside 272 359
```

Desaturasa lidská SCD



Phobius



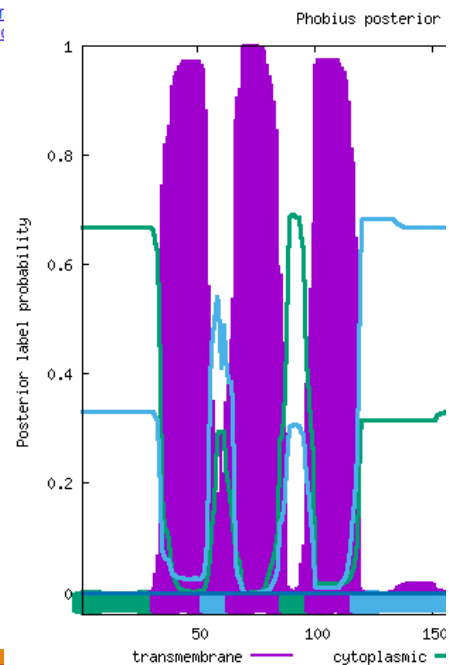
Phobius prediction

Prediction of UNNAMED

ID	UNNAMED		
FT	TOPO_DOM	1	33
FT	TRANSMEM	34	54
FT	TOPO_DOM	55	65
FT	TRANSMEM	66	88
FT	TOPO_DOM	89	99
FT	TRANSMEM	100	118
FT	TOPO_DOM	119	180
FT	TRANSMEM	181	202
FT	TOPO_DOM	203	213
FT	TRANSMEM	214	234
FT	TOPO_DOM	235	344
//			

A combined transmembrane prediction

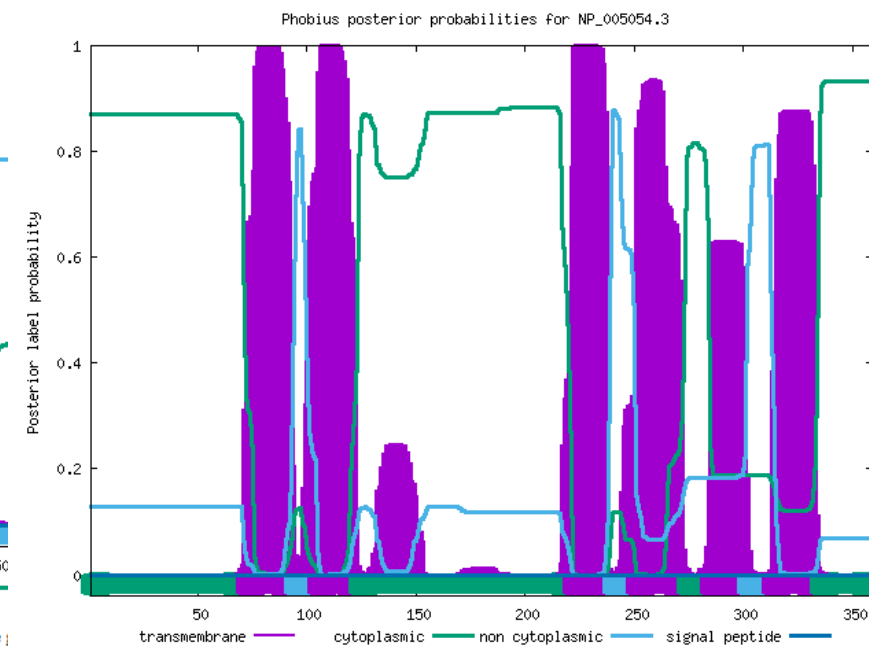
[Normal prediction](#) [Constrained prediction](#)



Phobius prediction

Prediction of NP_005054.3

ID	NP_005054.3		
FT	TOPO_DOM	1	71
FT	TRANSMEM	72	93
FT	TOPO_DOM	94	104
FT	TRANSMEM	105	123
FT	TOPO_DOM	124	221
FT	TRANSMEM	222	239
FT	TOPO_DOM	240	250
FT	TRANSMEM	251	273
FT	TOPO_DOM	274	284
FT	TRANSMEM	285	301
FT	TOPO_DOM	302	312
FT	TRANSMEM	313	334
FT	TOPO_DOM	335	359
//			



Normal prediction

Paste your protein sequence here in Fasta format:

Or: Select the sequence file you wish to use Nevybrán žádný soubor

Select output format:

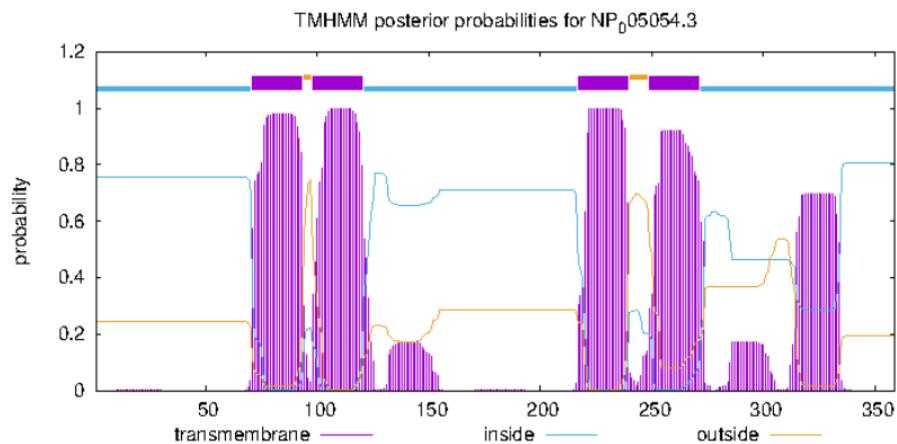
- Short
- Long without Graphics
- Long with Graphics

The probability data used in the plot is found [here](#), and the

Predikce transmembránových úseků (SCD)

TMHMM result

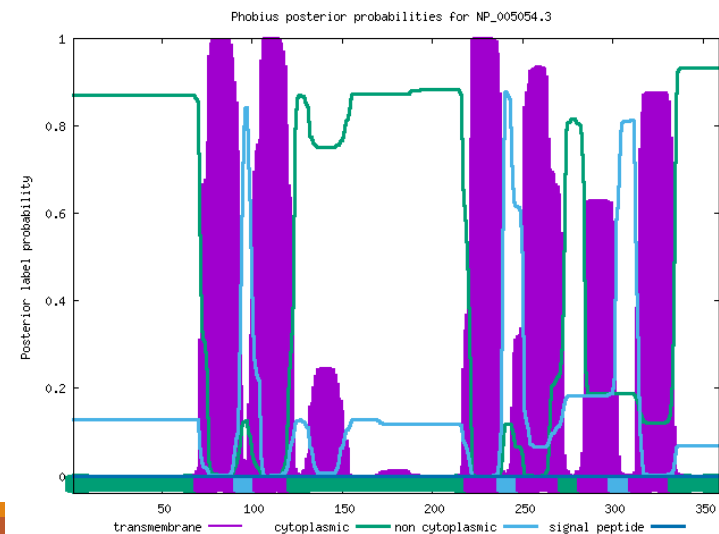
```
# NP_005054.3 Length: 359
# NP_005054.3 Number of predicted TMHs: 4
# NP_005054.3 Exp number of AAs in TMHs: 104.54198
# NP_005054.3 Exp number, first 60 AAs: 0.03484
# NP_005054.3 Total prob of N-in: 0.75464
NP_005054.3 TMHMM2.0 inside 1 70
NP_005054.3 TMHMM2.0 TMhelix 71 93
NP_005054.3 TMHMM2.0 outside 94 97
NP_005054.3 TMHMM2.0 TMhelix 98 120
NP_005054.3 TMHMM2.0 inside 121 216
NP_005054.3 TMHMM2.0 TMhelix 217 239
NP_005054.3 TMHMM2.0 outside 240 248
NP_005054.3 TMHMM2.0 TMhelix 249 271
NP_005054.3 TMHMM2.0 inside 272 359
```



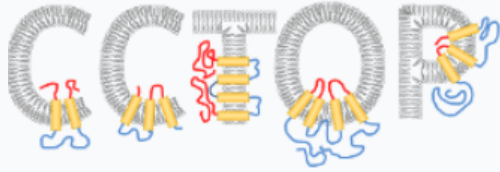
Phobius prediction

Prediction of NP_005054.3

```
ID NP_005054.3
FT TOPO_DOM 1 71 CYTOPLASMIC.
FT TRANSMEM 72 93
FT TOPO_DOM 94 104 NON CYTOPLASMIC.
FT TRANSMEM 105 123
FT TOPO_DOM 124 221 CYTOPLASMIC.
FT TRANSMEM 222 239
FT TOPO_DOM 240 250 NON CYTOPLASMIC.
FT TRANSMEM 251 273
FT TOPO_DOM 274 284 CYTOPLASMIC.
FT TRANSMEM 285 301
FT TOPO_DOM 302 312 NON CYTOPLASMIC.
FT TRANSMEM 313 334
FT TOPO_DOM 335 359 CYTOPLASMIC.
//
```



CCTOP

[⚡ Submit](#)[📖 Manual](#)[? About](#)[⬇ Standalone](#)[☰ MyJobs](#)

Submission form

Please do not submit sequences of complete proteomes! CCTOP prediction results for most of them can be downloaded from [TmAlphaFold Database](#).

Job Name

Sequences

Please provide input sequence(s) in FASTA format! The sequence(s) should be at least 30 amino acids long.

Or upload a file:

Please note that files are only validated after upload. Amino acid sequences must be longer than 30 and shorter than 5000 amino acids. Only

TOPCONS

TOPCONS

New query

Batch WSDL API

Download

References

Results

- Submitted: 2020-03-27 18:40:32 UTC
- Status: **Finished**
- Waiting time: 0 sec
- Running Time: 5 mins

Results of your prediction with ibid: **rst_ob13aA**

Predicted signal peptide and TM-helix positions (position starting from 1):

```

TOPCONS      ***No signal peptide nor TM-regions predicted***
OCTOPUS      ***No signal peptide nor TM-regions predicted***
Philius      ***No signal peptide nor TM-regions predicted***
PolyPhobius  ***No signal peptide nor TM-regions predicted***
SCAMPI       ***No signal peptide nor TM-regions predicted***
SPOCTOPUS   ***No signal peptide nor TM-regions predicted***
PDB-homology ***No homologous TM proteins detected***
    
```

Queued 0
Running 0
Finished 1
Failed 0

© Arne Elofsson

SciLifeLab NBS

OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
PDB-homology ***No homologous TM proteins detected***

New query

Batch WSDL API

Download

References

News

Server status

Example results

Old TOPCONS

Help

Your recent jobs:

Queued 0
Running 0
Finished 25
Failed 0

Results

- Submitted: 2018-03-27 18:40:32 UTC
- Status: **Finished**
- Waiting time: 1 sec
- Running Time: 28 s

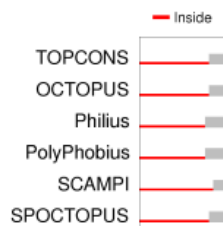
Results of your prediction with ibid: **rst_ob13aA**

Zipped folder of your result

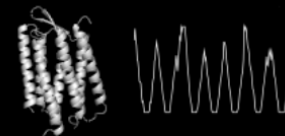
Dumped prediction in one file

The sequence(s) you submitted

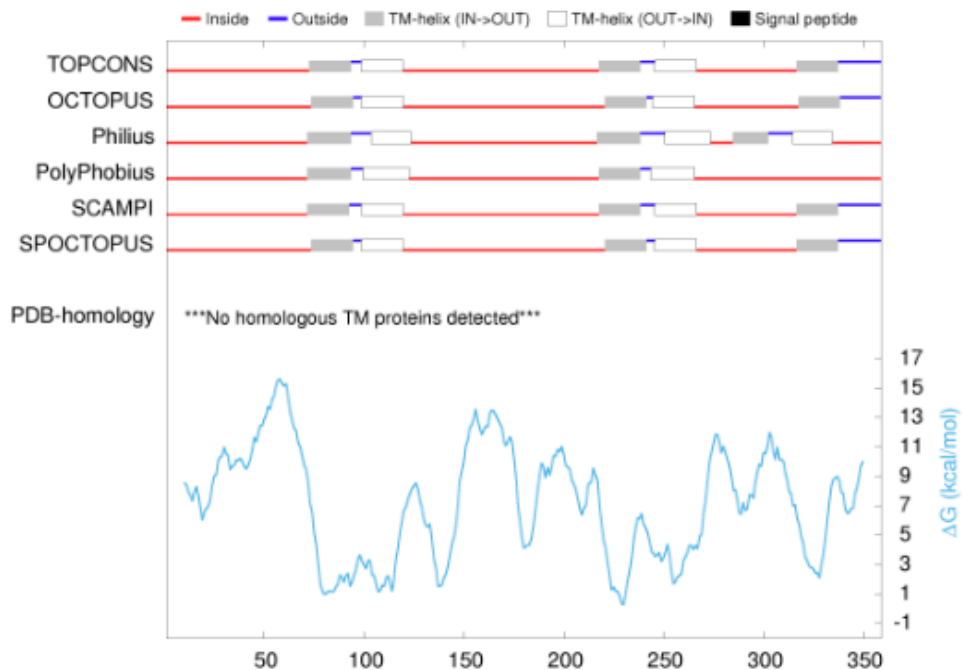
Predicted topologies and predicted ΔG values:



TOPCONS

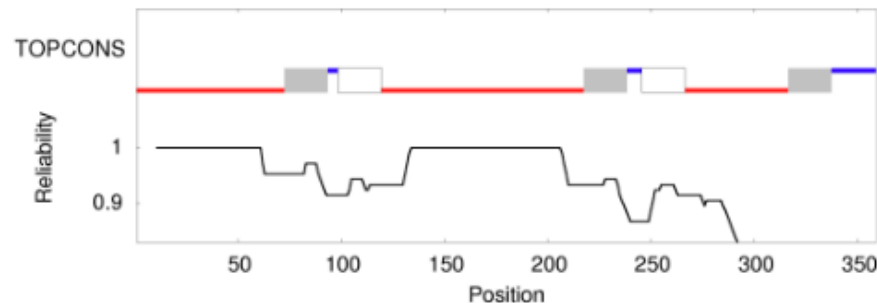


Predicted topologies and predicted ΔG values:



[High-resolution image](#)

Consensus prediction (TOPCONS):



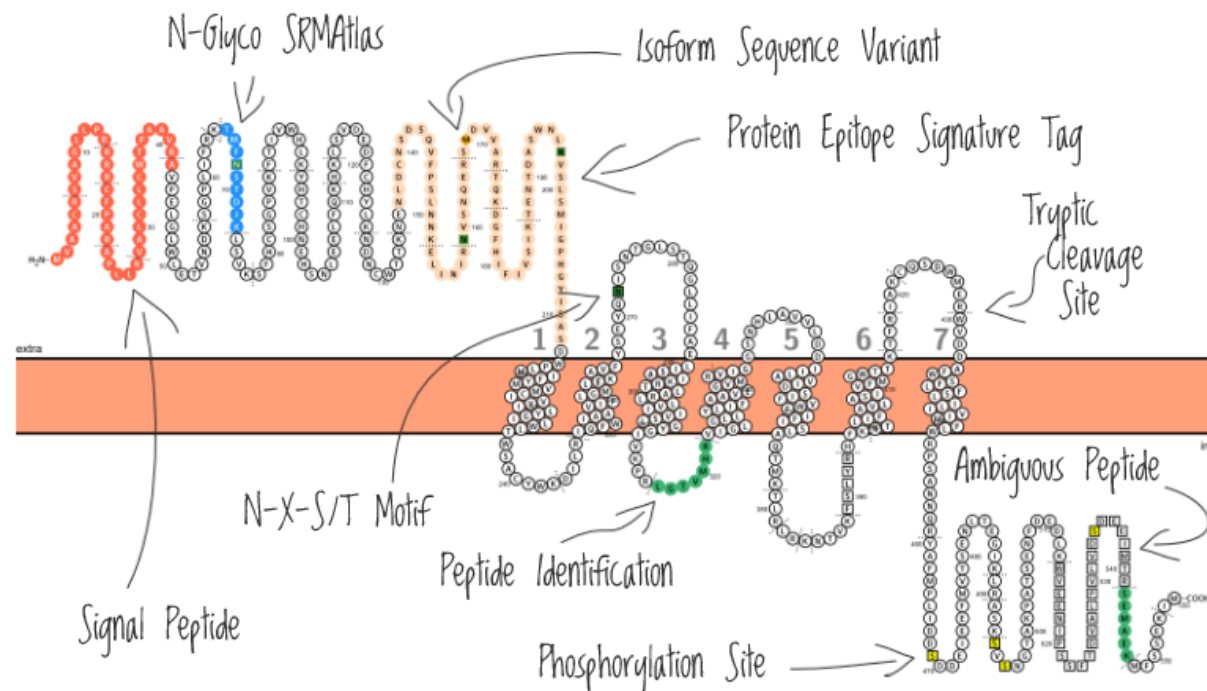
PROTTER-obrázek!

-pracuje s daty z UniProt
(nepočítá sám)

PROTTER

version 1.0 | help | manual | Wollscheid Lab

Welcome to Protter — the open-source tool for visualization of proteoforms and interactive integration of annotated and predicted sequence features together with experimental proteomic evidence!



start **PROTTER**

PROTTER-obrázek!

-pracuje s daty z UniProt

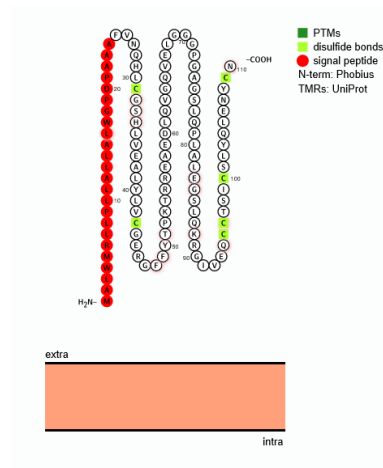
(nepočítá sám)

-zadává se přístupový kód:

UGT1A6 (P19224)

Desaturasa (O00767)

(prepro) insulin (P01308)



The screenshot shows the PROTTER web interface. The top navigation bar includes tabs for 'protein', 'topology', 'styles', 'misc.', 'help', and 'manual'. The main content area is titled 'by accession' and 'by sequence'. A red arrow points to the 'by accession' tab. Below the tabs, there is a form to enter a UniProt protein accession number, with 'submit' and 'load example' buttons. Below the form, there are buttons for 'enter a list of proteins' and 'load a proteomics result file'. The bottom section of the interface shows a protein structure diagram with a legend for PTMs, variants, disulfide bonds, signal peptide, N-term, and TMRs. The diagram shows a protein chain with a red signal peptide, a yellow disulfide bond, and a green PTM. The protein is shown in a 3D ribbon representation. Below the protein is a bar chart showing the protein's topology across a membrane, with 'Luminal' and 'Cytoplasmic' regions labeled. The bar chart shows four transmembrane domains (I, II, III, IV) and a signal peptide. The protein is shown in a 3D ribbon representation.

PROTTER-obrázek!

PROTTER 1 protein 2 topology 3 styles 4 misc. ? help manual

by accession
by sequence

or, alternatively you can:

Luminal
Cytoplasmic
H₂N
COOH

1 2 3 4

■ PTMs
◆ variants
— disulfide bonds
● signal peptide
● N-term: UniProt
● TMRs: UniProt

PROTTER 1 protein 2 topology 3 styles 4 misc. ? help manual

by accession

by sequence

extra
intra
H₂N
COOH

1 2 3 4 5 6

■ N-glyco motif
● signal peptide
● N-term: Phobius
● TMRs: Phobius

Predikce transmembránových úseků

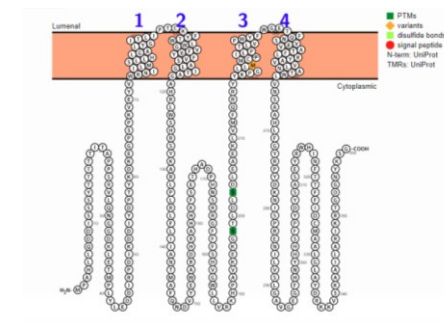
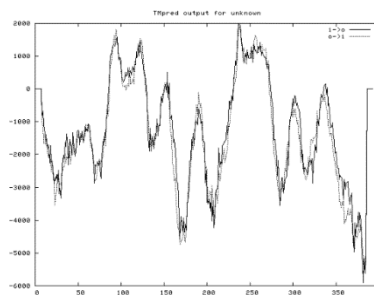
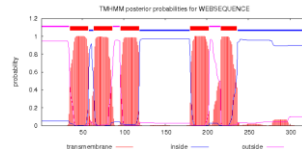
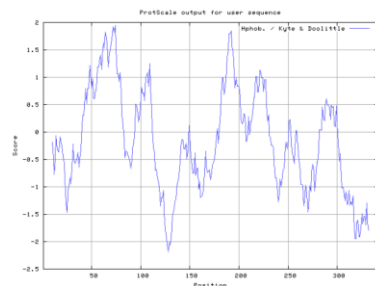
ProtScale

TMHMM

Phobius...

TOPCONS

CCTOP



➤ **Vyzkoušet více programů!**

Praktická ukázka

- otevřete si soubor/stránku s vaší sekvencí
- otevřete si odkazy (každý v novém okně) a zjistěte zda je to rozpustný protein, případně kolik má předpovězených transmembránových úseků

„Proteinová bioinformatika“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod...)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce sekundární struktury → Hledání známých motivů

3D-struktura, vizualizace

Vyhledání a porovnání podobných sekvencí

Evoluční příbuznost sekvencí

...

Hledání podobnosti / porovnávání (alignment)

Podobnost (similarity) x homologie

BLAST: Basic Local Alignment and Search Tool

Hledá podobné sekvence z vybraných databáze pomocí krátkých „slov“ (sekvencí vytvořených ze zadané sekvence), s využitím „substituční matice“, která udává míru důležitosti rozdílných aminokyselin v zadané sekvenci a vyhledané sekvenci

„Leucin je podobnější Isoleucinu než Histidinu“

BLAST je součástí téměř všech dostupných databází (včetně genomových).

NCBI/BLAST

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>



U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

[jostovap](#) [My NCBI](#) [Sign Out](#)

BLAST[®]

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

Magic-BLAST 1.2.0 released

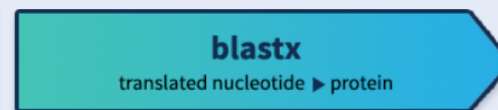
A new version of the BLAST RNA-seq mapping tool is now available.
Mon, 27 Feb 2017 14:00:00 EST

[More BLAST news...](#)

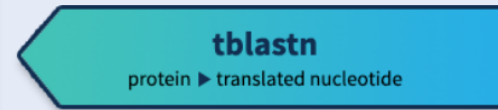
Web BLAST



Nucleotide BLAST
nucleotide ▶ nucleotide



blastx
translated nucleotide ▶ protein



tblastn
protein ▶ translated nucleotide



Protein BLAST
protein ▶ protein

BLAST Genomes

Search

[Human](#)

[Mouse](#)

[Rat](#)

[Microbes](#)

NCBI/BLAST

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI Welcome jostovap. [Sign Out]

NCBI/ BLAST/ blastp suite **Standard Protein BLAST**

blastn blastp blastx tblastn tblastx

Enter accession number(s), gi(s) or FASTA sequence(s) [is using a protein query. more...](#) [Reset page](#) [Bookmark](#)

```
MAARRALIVLAHSEKTSFNAMKEAAVEALKKRGWEVLESDLYAMNENPIISRNDITGELKDSKNFQ
YPS
ESSLAHKEGRLSPDIVAEHKKLEAADLVIFQFPLQWFGVPAILKGFERVLVAGFAYTYAAMYDNGP
FQN
KKTLLSITGGSGSMYSLQGVHGMNVILWPIQSGILRFQGFQVLEPQLVYSIGHTPPDARMQILEG
```

From To

Or, upload file

Job Title
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

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

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query
Enter an Entrez query to limit search

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 using Blastp (protein-protein BLAST) Show results in a new window

[Algorithm parameters](#)

NCBI/BLAST



BLAST® » blastp suite » results for RID-ZYHJKTA1013

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

Job Title NP_005054.3 stearoyl-CoA desaturase [Homo...]
RID [ZYHJKTA1013](#) Search expires on 03-02 19:34 pm [Download All](#)
Program BLASTP [Citation](#)
Database nr [See details](#)
Query ID lc|Query_51026
Description NP_005054.3 stearoyl-CoA desaturase [Homo sapiens]
Molecule type amino acid
Query Length 359
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

Descriptions **Graphic Summary** Alignments Taxonomy

hover to see the title click to show alignments Show Conserved Domains Alignment Scores < 40 40 - 50 50 - 80 80 - 200 >= 200

100 sequences selected Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. 1 70 140 210 280 350
Specific hits
Superfamilies

Distribution of the top 100 Blast Hits on 100 subject sequences

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

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> stearoyl-CoA desaturase [Homo sapiens]	Homo sapiens	748	748	100%	0.0	100.00%	3 9	NP_005054.3
<input checked="" type="checkbox"/> stearoyl-CoA desaturase [Homo sapiens]	Homo sapiens	747	747	100%	0.0	99.72%	3 9	AAD29870.1
<input checked="" type="checkbox"/> stearoyl-CoA desaturase variant [Homo sapiens]	Homo sapiens	746	746	100%	0.0	99.72%	3 6	BAD92219.1
<input checked="" type="checkbox"/> stearoyl-CoA desaturase variant [Homo sapiens]	Homo sapiens	744	744	100%	0.0	99.72%	3 9	BAD96582.1

E-value (expectancy)

Odkazy

NCBI/BLAST

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view

Pairwise



Restore defaults

Download

100 sequences selected



Download

GenPept Graphics

Next Previous Descriptions

stearoyl-CoA desaturase [Homo sapiens]

Sequence ID: [NP_005054.3](#) Length: 359 Number of Matches: 1

[See 2 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 359 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
748 bits(1931)	0.0	Compositional matrix adjust.	359/359(100%)	359/359(100%)	0/359(0%)

Query 1 MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPPLYLEDDIRPDIKDDIYDPTYK 60

MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPPLYLEDDIRPDIKDDIYDPTYK

Sbjct 1 MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPPLYLEDDIRPDIKDDIYDPTYK 60

Query 61 DKEGSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH 120

DKEGSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH

Sbjct 61 DKEGSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH 120

Query 121 RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARHRAHKKFSETHADPHNSRRGFFFS 180

RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARHRAHKKFSETHADPHNSRRGFFFS

Sbjct 121 RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARHRAHKKFSETHADPHNSRRGFFFS 180

Query 181 HVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKPGLLMCMFILPTLVPWYFWGE 240

HVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKPGLLMCMFILPTLVPWYFWGE

Related Information

[Gene](#) - associated gene details

[Genome Data Viewer](#) - aligned genomic context

[Identical Proteins](#) - Identical proteins to NP_005054.3

→upravit uložené sekvence, tak aby v názvu byl jenom organismus

NCBI/BLAST

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download

select all 5 sequences selected

- GenBank (complete)
- FASTA (complete)
- FASTA (aligned)
- GenBank (complete)
- Hit Table (text)
- Hit Table (CSV)

Description

- [stearoyl-CoA desaturase \[Homo sapiens\]](#)
- [stearoyl-CoA desaturase \[Homo sapiens\]](#)
- [stearoyl-CoA desaturase variant \[Homo sapiens\]](#)
- [stearoyl-CoA desaturase variant \[Homo sapiens\]](#)
- [acyl-CoA desaturase \[Gorilla gorilla gorilla\]](#)
- [SCD isoform 1 \[Pongo abelii\]](#)
- [stearoyl-CoA desaturase \[Pongo abelii\]](#)
- [SCD protein \[Homo sapiens\]](#)
- [acyl-CoA desaturase \[Pan troglodytes\]](#)
- [acyl-CoA desaturase \[Hylobates moloch\]](#)
- [stearoyl CoA desaturase \[Homo sapiens\]](#)
- [acyl-CoA desaturase \[Nomascus leucogenys\]](#)

*seqdump (1).txt - Poznámkový blok

```
Soubor Úpravy Formát Zobrazení Nápověda
>Homo sapiens
MPAHLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDEGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCMCFILPTLVPWYFWGE
TFQNSVVFVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LARIKRTGDGNYKSG
>Gorilla gorilla gorilla
MPAHLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDEGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCMCFILPTLVPWYFWGE
TFQNSVVFVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LARIKRTGDGNYKSG
>Pan troglodytes
MPAHLQDDITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDEGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCMCFILPTLVPWYFWGE
TFQNSVVFVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LARIKRTGDGNYKSG
>Camelus ferus
MPAHLQEEISSSYTTTTITAPPSRVLQNGGDKLEKTPLYLEEDIRPEMKDDIYDPSYQDKEGPKPKVYVWRNIILMG
LLHLGALYGITLIPTCKFYTFQWVLFYIIISALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVFEWARDHRAH
HKFSETDADPHNSRRGFFFSHVGWLLVRKHPAVKEKGGLLDLSDLKAEKLVMPQRRYYKPGILLMCFIMPTLVPWYFWGE
TFQHSYLYLATFLRYAVVLNVTWLVNSAAHLYGYRYPYDKTINPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINPT
TFPIDCMAALGLAYDRKGVSKAAI LAKVKRTGDGSYKSG
>Ovis aries
MPAHLQEEISSSYTTTTITAPPSRVLQNGGDKLEKTPLYLEEDIRPEMRDDIYDPTYQDKEGPKPKLEYVWRNIILMG
LLHLGALYGITLIPTCKIYTLWVLFYVVISALGITAGVHRLWSHRTYKARLPRLRFLIIANTMAFQNDVFEWARDHRAH
HKFSETDADPHNSRRGFFFSHVGWLLVRKHPAVREKATLDLSDLRAEKLVMPQRRYYKPGVLLLCFILTLPWYFWLWGE
TFQNSLFFATFLRYAVVLNATWLVNSAAHMYGYRYPYDKTINPRENI LVS LGAVGEGGFHNYHTFPYDYSASEYRWHINPT
TFPIDCMAAIGLAYDRKGVSKAAV LARMKRTGEEYSYKSG
>gi|13435426|gb|AAH04579.1| Nqo1 pr
MAARRALIVLHSEKTSFNAMKEAAVEALKKRGW
LSPDIVAEHKKLEAADLVIFQFPLQWFGVPAILKKG
VHGMNVILWPIQSGILHFCGFQVLEPQLVYSIGH
>gi|524939198|ref|XP_005071892.1| P
MAVRRALIVLHSEKTSFNAMKEAAVEALKKRGW
LSPDIVAEQKLEAADLVIFQFPLHWFQVPAILKKG
VHGMNIIWPIQSGILHFCGFQVLEPQLVYSIGHTPPDARTQILEGWKKRLETWDETPLYFVPSLFDLNFQAGFLKKEVQEEQKNRFGLSVGHHLGKSIPTDQVQKARK
>gi|227430403|ref|NP_001153085.1| NAD(P)H dehydrogenase [quinone] 1 [Sus scrofa]
MAVRKALIVLHSEKTSFNAMKEAAVEALKRRGWEVAVSDLYAMNPNVIRSKDITGKLDKDPGNFQYPAETALAYKEGR
LSPDIVAEQKKEAADLVIFQFPLQWFGVPAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQ
IHGMNIIWPIQSGILHFCGFQVLEPQLTYSIGHTPEDARTQILEEWKKRLENWDETPLYFAPSSLFDLNFQAGFLMKKQVQDEQKSNKFGLSVGHHLGKSIPTDQVQKARK
>gi|386781783|ref|NP_001247927.1| NAD(P)H dehydrogenase [quinone] 1 [Macaca mulatta]
MVGKRALIVLHSEKTSFNAMKEAAVAALKKKGWEVAVSDLYAMNPNVIRSKDITGKLDKDPANFQYAAESTLAYKEGR
LSPDIVAEQKLEAADLVIFQFPLQWFGVPAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQ
IHGMNIIWPIQSGILHFCGFQVLEPQLTYSIGHTPADARTQILEGWKKRLENWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKFGLSVGHHLGKSIPTDQVQKARK
>gi|426242583|ref|XP_004015151.1| PREDICTED: NAD(P)H dehydrogenase [quinone] 1 [Ovis aries]
MAVRKALIVLHSEKTSFNAMKEAAEALKRRGWEVTVSDLYAMNPNVIRSKDITGKLDKDPGNFQYPAETVLAAYKEGR
LSPDIVAEQKLEAADLVIFQFPLQWFGVPAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQ
IHGMNIIWPIQSGILHFCGFQVLEPQLTYSIGHTPADARVQILEGWKKRLENWDEMPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKFGLSVGHHLGKSIPTDQVQKARK
>gi|30230685|gb|AAP20940.1| NAD(P)H dehydrogenase, quinone 1 [Homo sapiens]
RRALIVLHSEKTSFNAMKEAAALKKKGWEVSDLYAMNPNVIRSKDITGKLDKDPANFQYPAESVLAAYKEGHLSP
DIVAEQKLEAADLVIFQFPLQWFGVPAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQIGIH
DMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKKRLENWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKFGLSVGHHLGKSIPTDQVQKARK
```

NCBI/BLAST-podobné sekvence u jednoho org.

Standard Protein BLAST

[blastn](#) **[blastp](#)** [blastx](#) [tblastn](#) [tblastx](#)

[Reset page](#) [Bookmark](#)

Enter Query Sequence

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

```
MVGRRALIVLAHSEKTSFNAMKEAAAAALKKKGWEVVESDIYAMNENPIISRKDIKGL
KDPANFOYPAESVLAYKEGHLSPDIVAEQKLEAADLVIFQFPLQWFGVPAILKGFERV
FIGEFAYTYAAMYDKGPFRRSKKAVLSITTTGGSGSMYSLOGIHGDMNVILWPIQSGILHFC
GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
KEVODEEKNNKFGLSVGHHLGKSIPTDNOIKARK
```

From
To

BLAST results will be displayed in a new format by default
You can always switch back to the Traditional Results page.

Or, upload file Nevybrán žádný soubor

Job Title
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database



Organism
Enter organism common name, binomial, or tax id. Only

Exclude Models (XM/XP) Non-redundant RefSeq prot

Program Selection

Algorithm blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)

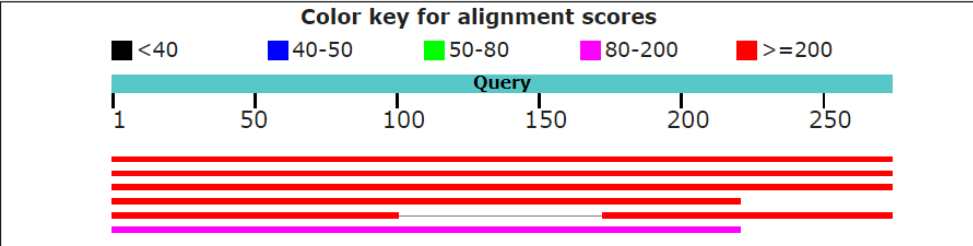
Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. 
Specific hits 

Distribution of the top 7 Blast Hits on 6 subject sequences
Mouse over to see the title, click to show alignments

Color key for alignment scores

■ <40	■ 40-50	■ 50-80	■ 80-200	■ >=200
-------	---------	---------	----------	---------



Uniprot/BLAST

<https://www.youtube.com/watch?v=UPaConHNP7E>

BLAST

How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g. P00750 or A4_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

[Help](#) [BLAST help video](#) [Other tutorials and videos](#) [Downloads](#)

Protein sequence, Nucleotide sequence or UniProt identifier

Target database¹ UniProtKB E-Threshold¹ 10 Matrix¹ Auto Filtering¹ None Gapped¹ yes Hits¹ 250

Run BLAST in a separate window.

[Clear](#) [Run BLAST](#)

Tools

BLAST
Align
Retrieve/ID mapping
Peptide search

Core data

Protein knowledgebase (UniProtKB)
Sequence clusters (UniRef)
Sequence archive (UniParc)
Proteomes

Supporting data

Literature citations
Taxonomy
Keywords
Subcellular locations
Cross-referenced databases
Diseases

Information

About UniProt
Help
FAQ
UniProtKB manual
Technical corner
Expert biocuration

Uniprot/BLAST

UniProt BLAST Align Peptide search ID mapping SPARQL Tool results Advanced | List Search

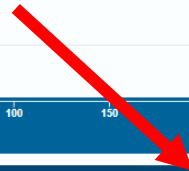
Blast parameters
Identity: 53.4, Score: 214, E-Value: 2.8e-15

BLAST 249 results found in UniProtKB

Overview Taxonomy Hit Distribution Text Output Input Parameters API Request

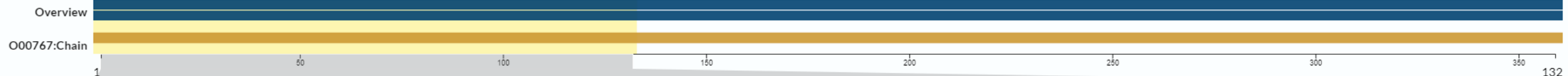
BLAST Align Map IDs Download Add Customize columns Resubmit

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	Score	Identity
<input type="checkbox"/> O00767	SCD_HUMAN	Stearoyl-CoA desaturase[...]	SCD, FADSS, SCD1, SCDOS	Homo sapiens (Human)	359 AA	100%	1946
<input type="checkbox"/> G3QFK6	G3QFK6_GORGO	Stearoyl-CoA desaturase	SCD	Gorilla gorilla gorilla (Western lowland gorilla)	359 AA	99.2%	1936
<input type="checkbox"/> A0A6D2WW66	A0A6D2WW66_PONAB	SCD isoform 1[...]	SCD, CR201_G0023498	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	359 AA	99.2%	1931
<input type="checkbox"/> A0A2R9AD04	A0A2R9AD04_PANPA	Stearoyl-CoA desaturase	SCD	Pan paniscus (Pygmy chimpanzee) (Bonobo)	359 AA	98.9%	1924
<input type="checkbox"/> H2Q2F3	H2Q2F3_PANTR	Stearoyl-CoA desaturase[...]	SCD	Pan troglodytes (Chimpanzee)	359 AA	98.9%	1924
<input type="checkbox"/> H2NBA9	H2NBA9_PONAB	Stearoyl-CoA desaturase	SCD	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	404 AA	99.1%	1882



O00767 · Stearoyl-CoA desaturase · Homo sapiens

Highlight properties Select annotation View: Overview Wrapped




Query: MPAHLLQDDI SSSYTTTTT I TAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKKEGSPKVEYVWRNII LMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARL131
 Match:O00767 MPAHLLQDDI SSSYTTTTT I TAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKKEGSPKVEYVWRNII LMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARL131

Vyzkoušejte si BLAST

Uložte si pět sekvencí z různých organismů ve FASTA formátu (do poznámkového bloku).

- identifikátor ve fasta formátu upravte jen na organismus (vhodné pro porovnávání v příští lekci)

→upravit uložené sekvence, tak aby v názvu byl jenom organismus



```
*seqdump (1).txt - Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
>Homo sapiens
MPAHLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPPLYLEDDIRPDIKDDIYDPTYKDKGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLFLIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFHVWGLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCFILPTLVPEWYFWGE
TFQNSVFEVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENIVSLGAVGEGPHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKVKVSKAAI LARIKRTGDGNYKSG
>Gorilla gorilla gorilla
MPAHLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPPLYLEDDIRPDIKDDIYDPTYKDKGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLFLIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFHVWGLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCFILPTLVPEWYFWGE
TFQNSVFEVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENIVSLGAVGEGPHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKVKVSKAAI LARIKRTGDGNYKSG
>Pan troglodytes
MPAHLQDDITAPPSRVLQNGGDKLETMPPLYLEDNIRPDIKDDIYDPTYKDKGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLFLIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFHVWGLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCFILPTLVPEWYFWGE
TFQNSVFEVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENIVSLGAVGEGPHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKVKVSKAAI LARIKRTGDGNYKSG
>Camelus ferus
MPAHLQDEISSSYTTTTITAPPSRVLQNGGDKLEKTPPLYLEEDIRPEMKDDIYDPSYQDKGPKKVVYVWRNIILMG
LLHLGALYGITLIPTCKFYTFWVLFYYIISALGITAGAHRLWSHRSYKARLPRLVFLIANTMAFQNDVFewardHRAH
HKFSETDADPHNSRRGFFFHVWGLLVRKHPAVKEKGGLLDLSDLKAEKLVMPQRRYYKPGIILMCFIMPVLVPEWYFWGE
TFQHSLYLATFLRYAVVLNVTWLVNSAAHLYGYRYPYDKTINPRENIVSLGAVGEGPHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKVKVSKAAI LAKVKRTGDGSYKSG
>Ovis aries
MPAHLQDEISSSYTTTTITAPPSRVLQNGGGKLEKTPPLYLEEDIRPEMRDDIYDPTYQDKGPKKLEYVWRNIILMG
LLHLGALYGITLIPTCKIYTFWVLFYYVVISALGITAGVHRLWSHRTYKARLPRLVFLIANTMAFQNDVFewardHRAH
HKFSETDADPHNSRRGFFFHVWGLLVRKHPAVREKGTLDLSDLRAEKLVMPQRRYYKPGVLLLCFILPTLVPEWYLWGE
TFQNSLFFATFLRYAVVLNATWLVNSAAHMYGYRYPYDKTINPRENIVSLGAVGEGPHNYHHTFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKVKVSKAAV LARMKRTGEESEYKSG
```


DÚ3: analýza proteinů, podobnost

Pracujte s „vaším“ genem/proteinem/enzymem (př. NQO1)

- 1) Obsahuje váš protein nějaké typické motivy? (vytřihněte výstup z jednoho programu)
- 2) Je to transmembránový protein? (-||-)
- 3) Má signální peptid? (-||-)
- 4) Vyberte pět podobných sekvencí (vyhledaných BLASTem) z jiných organismů – uložte si sekvence
- 5) Vyhledejte podobné referenční sekvence jen u Homo sapiens – kolik sekvencí bylo nalezeno?

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

-počet se ukáže po zaškrtnutí „ALL“

➤ vše zpracujte do bloku OneNote

P-II: DÚ3-příklad řešení

DÚ3



TMMOD
Annotation NON-TM PROTEIN
Length 274
Number of predicted TMs 0
Exp number of AAs in TMs 0.000000
Exp number, first 60 AAs 0.000000
Total prob of N-in 0.493689 274
outside 1
[show posterior probabilities](#)

- 1)
- 2)
- 3)
- 4)

```
>Macaca mulatta
MVGKRALIVLASEKTSFNFMGEAAVAALPKKQWVAESDLYAMNFMFIISKDIIGKLEKDPANFQYAAESTLAYEGR
LSEDTVLEKQKLEALDLVTPQFLQWTFVALLKQWFERVUGSEFAYTAAADTQKGFQKQKAWLSITIGSSGSMVSLQG
IHGIMNVLWFIQSGILRFGCFQVLEPQLTYSIGRTFADARIQILEGKQKRLNIMDETFYFAPSSLPDLNFQAGFLMK
KEUQDEKQKQKGLSVGKHLGWSIFDINQKARK

>Sus scrofa
MAGKRALIVLASEKTSFNFMGEAAVAALPKKQWVAESDLYAMNFMFVISKDIIGKLEKDPANFQYFAETLAYEGR
LSEDTVLEKQKLEALDLVTPQFLQWTFVALLKQWFERVUGSEFAYTAAADTQKGFQKQKAWLSITIGSSGSMVSLQG
IHGIMNVLWFIQSGILRFGCFQVLEPQLTYSIGRTFADARIQILEGKQKRLNIMDETFYFAPSSLPDLNFQAGFLMK
KEUQDEKQKQKGLSVGKHLGWSIFDINQKARK

>Bos taurus
MAGKRALIVLASEKTSFNFMGEAAVAALPKKQWVAESDLYAMNFMFVISKDIIGKLEKDPANFQYFAETLAYEGR
LSEDTVLEKQKLEALDLVTPQFLQWTFVALLKQWFERVUGSEFAYTAAADTQKGFQKQKAWLSITIGSSGSMVSLQG
IHGIMNVLWFIQSGILRFGCFQVLEPQLTYSIGRTFADARIQILEGKQKRLNIMDETFYFAPSSLPDLNFQAGFLMK
KEUQDEKQKQKGLSVGKHLGWSIFDINQKARK

>Mus musculus
MAGKRALIVLASEKTSFNFMGEAAVAALPKKQWVAESDLYAMNFMFIISKDIIGKLEKDPANFQYFAETLAYEGR
LSEDTVLEKQKLEALDLVTPQFLQWTFVALLKQWFERVUGSEFAYTAAADTQKGFQKQKAWLSITIGSSGSMVSLQG
IHGIMNVLWFIQSGILRFGCFQVLEPQLTYSIGRTFADARIQILEGKQKRLNIMDETFYFAPSSLPDLNFQAGFLMK
KEUQDEKQKQKGLSVGKHLGWSIFDINQKARK

>Alligator sinensis
MAGKRALIVLASEKTSFNFMGEAAVAALPKKQWVAESDLYAMNFMFVISKDIIGKLEKDPANFQYFAETLAYEGR
LSEDTVLEKQKLEALDLVTPQFLQWTFVALLKQWFERVUGSEFAYTAAADTQKGFQKQKAWLSITIGSSGSMVSLQG
IHGIMNVLWFIQSGILRFGCFQVLEPQLTYSIGRTFADARIQILEGKQKRLNIMDETFYFAPSSLPDLNFQAGFLMK
KEUQDEKQKQKGLSVGKHLGWSIFDINQKARK
```

Nalezeno 6 sekvencí: tři *isoformy* a tři "jiné" sekvence

Description	Max score	Total score	E value	Ident	Accession
<input type="checkbox"/> NC0231:Mus musculus [accession:1] (Mus musculus)	587	587	100%	100%	NC_023109.1
<input type="checkbox"/> NC0231:Mus musculus [accession:2] (Mus musculus)	481	481	100%	100%	NC_023109.1
<input type="checkbox"/> NC0231:Mus musculus [accession:3] (Mus musculus)	541	541	100%	100%	NC_023109.1
<input type="checkbox"/> G04894:Macaca mulatta [accession:1] (Macaca mulatta)	225	225	80%	34/73	NC_023109.1
<input type="checkbox"/> NC0231:Mus musculus [accession:1] (Mus musculus)	219	424	74%	24/71	NC_023109.1
<input type="checkbox"/> G04894:Macaca mulatta [accession:2] (Macaca mulatta)	184	184	80%	34/50	NC_023109.1

Příště:

Téma 4/10

Proteinová bioinformatika III

Cíle:

Student bude schopen porovnat sekvence dvou či více proteinů, určit míru identity, zobrazit jednoduchý fylogenetický strom vybraných proteinů a najít 3D strukturu a interakční partnery proteinu...