

Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2023/2024

3/10

„Protein bioinformatics II“

Retrieving protein sequences from databases (Uniprot: FASTA formate)

Computing amino-acids compositions, molecular weight, isoelectric point, and other parameters (SMS)

Prediction of proteases cutting (PeptideCutter)

Predicting elements of protein secondary structure, signal peptide, transmembrane helix

Finding 3-D structure and the domain organization of proteins

Finding all proteins that share a similar sequence and Classifying proteins into families

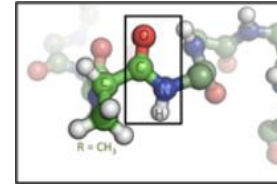
Finding evolutionary relationships between proteins, drawing proteins' family trees

Computing the optimal alignment between two or more protein sequences

...



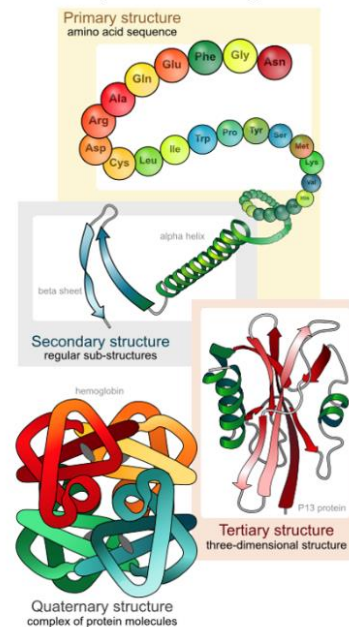
Proteins



20 Aminoacids – primary structure:

(Frederick Sanger-1958 Nobel prize for insulin sequencing)

Secondary structure
Tertiary structure
Quaternary structure

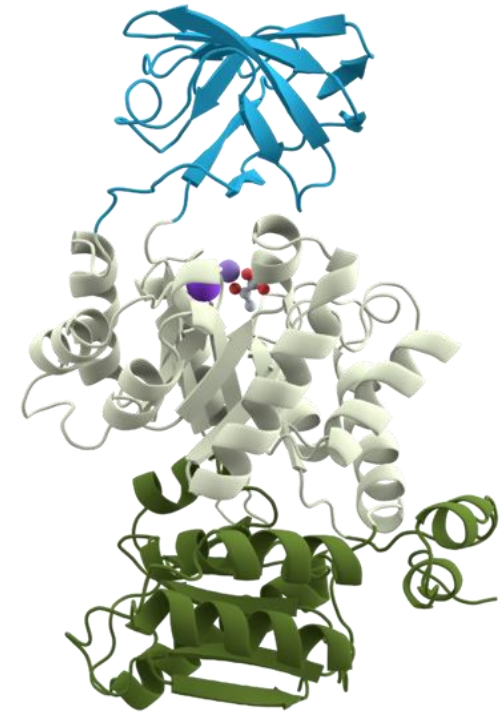
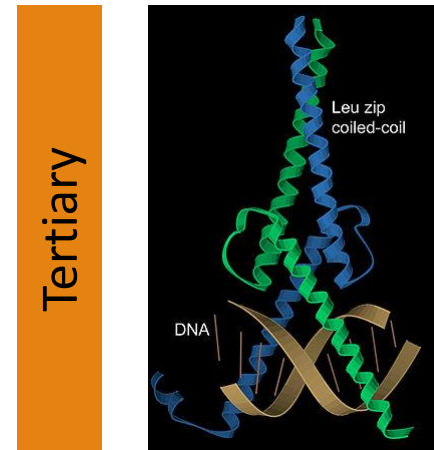
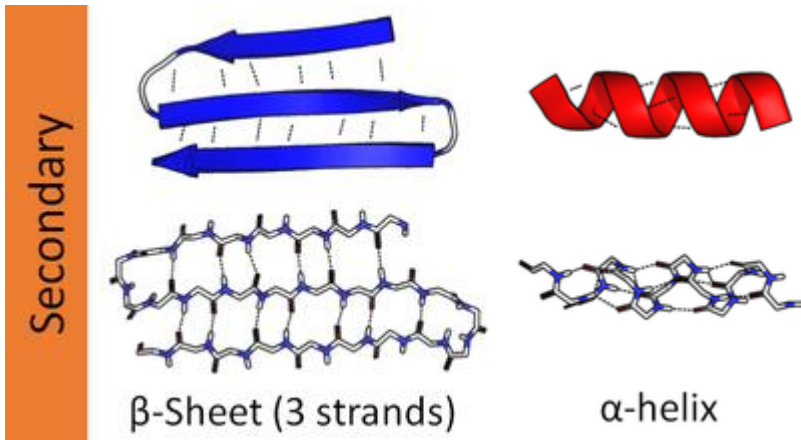


1-letter code	3-letter code	Amino acid	Possible codons
A	Ala	Alanine	GCA, GCC, GCG, GCT
B	Asx	Asparagine or Aspartic acid	AAC, AAT, GAC, GAT
C	Cys	Cysteine	TGC, TGT
D	Asp	Aspartic acid	GAC, GAT
E	Glu	Glutamic acid	GAA, GAG
F	Phe	Phenylalanine	TTC, TTT
G	Gly	Glycine	GGA, GGC, GGG, GGT
H	His	Histidine	CAC, CAT
I	Ile	Isoleucine	ATA, ATC, ATT
K	Lys	Lysine	AAA, AAG
L	Leu	Leucine	CTA, CTC, CTG, CTT, TTA, TTG
M	Met	Methionine	ATG
N	Asn	Asparagine	AAC, AAT
P	Pro	Proline	CCA, CCC, CCG, CCT
Q	Gln	Glutamine	CAA, CAG
R	Arg	Arginine	AGA, AGG, CGA, CGC, CGG, CGT
S	Ser	Serine	AGC, AGT, TCA, TCC, TCG, TCT
T	Thr	Threonine	ACA, ACC, ACG, ACT
V	Val	Valine	GTA, GTC, GTG, GTT
W	Trp	Tryptophan	TGG
X	X	Stop codon	TAA, TAG, TGA
Y	Tyr	Tyrosine	TAC, TAT
Z	Glx	Glutamine or Glutamic acid	CAA, CAG, GAA, GAG

SEQUENCE ⇔ STRUCTURE ⇔ FUNCTION

Protein domain

- region of a protein's polypeptide chain that folds independently from the rest
- forms a compact folded three-dimensional structure
- many proteins consist of several domains

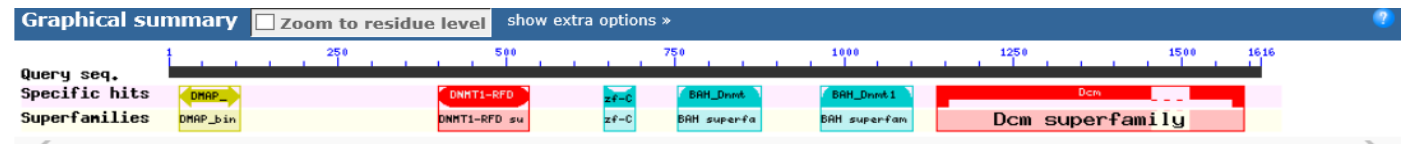


Conserved domain search

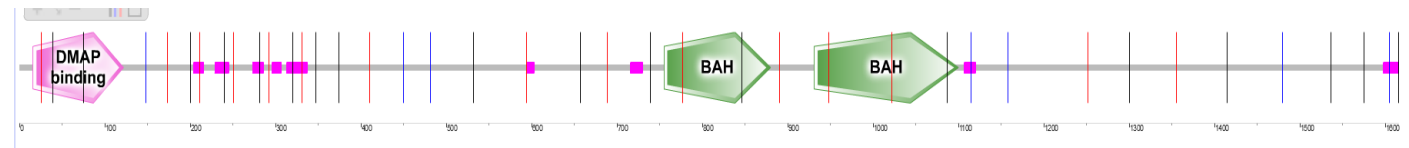
SEQUENCE ⇨ STRUCTURE ⇨ FUNCTION

- Conserved domain databases:

NCBI/CDD

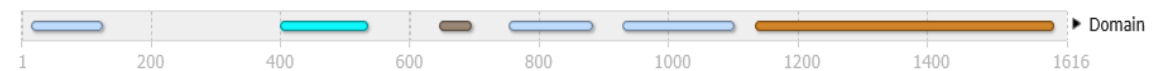


SMART

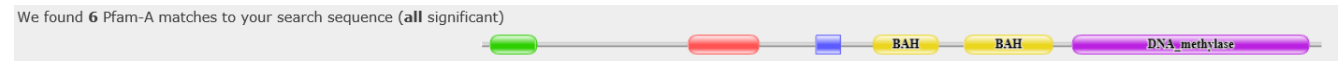


EMBL/InterPro

Domains and repeats



Pfam



Conserved domain search - CD (NCBI)

NCBI Resources ▾ How To ▾ jostovap My NCBI Sign Out

Conserved Domains [Advanced](#) [Help](#)

Conserved Domains and Protein Classification

[OVERVIEW](#) [SEARCH](#) [HOW TO](#) [HELP](#) [NEWS](#) [FTP](#) [PUBLICATIONS](#) [DISCOVER](#)

How to use CDD: examples

This page provides **quick start guides** for some common types of searches.
The **CDD Help document** provides detailed descriptions of the database content, search system, and display formats.
Once records of interest are retrieved, follow Entrez's "Links" to **discover associations among previously disparate data**.

- Identify the putative **function of a protein** sequence.
- Identify a **protein's classification** based on domain architecture.
- Identify the specific **amino acids** in a protein sequence that are putatively **involved in functions such as binding or catalysis**, as mapped from conserved domain annotations to the query sequence.
- View a **protein query sequence embedded within the multiple sequence alignment of a domain model**.
- **Interactively view** the **3D structure** of a conserved domain.
- Find other **proteins with similar domain architecture**.
- Interactively view the **phylogenetic sequence tree** for a conserved domain model of interest, with or without a query sequence embedded.

Conserved domain search - 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

Hit types vary based on confidence level and specificity. Follow the text links below this illustration for more information about each hit type.

Small triangles indicate the amino acids involved in conserved features/sites, such as catalytic and binding sites

Click on the colored bar representing any domain model to view the detailed information for that domain, including a multiple sequence alignment of the proteins used to generate the domain, with your query sequence embedded.

Conserved domains on [gi|157830769|pdb|1CYG|A]
Chain A, Cyclodextrin Glucanotransferase (E.C.2.4.1.19) (Cgtase)

View **Concise Results**

Graphical summary Zoom to residue level Hide extra options << Show site features Horizontal zoom: x 1

Query seq. 1 100 200 300 400 500 600 680

active site catalytic site starch-binding site 1 starch-binding site 2

Ca binding site

Specific hits
Superfamilies
Multi-domains

Alpha amylase

Alpha-amylase_C su

IPT_CGTD
IPT superfamily

CBM20 superfamily

Specific hits are shown in **bright colors** as the top hit type. They represent a high confidence association between the query sequence and a domain model, and therefore a high confidence level in the inferred function of the query protein.

The **superfamily** to which a specific hit belongs is shown beneath it in a similar, **pastel color**.

If CD-Search finds only **non-specific hits** for a region of the query sequence, only the **superfamily** to which the hits belong will be shown in the concise display. The non-specific hits can be viewed in the full display.

Multi-domains are domain models that were computationally detected and are likely to contain multiple single domains. They are typically shown as **grey bars**.

Conserved domain search / NQO1

NCBI Resources How To

HOME SEARCH GUIDE Structure

Search for

Enter protein or nucleotide query as

Submit

Conserved Domains

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

View Concise Results

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

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 »

Query seq. Specific hits Superfamilies

Flavodoxin_2

FMN_red super-family

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)2237-2241.

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)2237-2241.

Conserved domain search - 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

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):

SMART

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

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

SMART MODE:
NORMAL
GENOMIC

Simple
Modular
Architecture
Research
Tool

keywords...
Search SMART

Conserved domain search - SMART

The screenshot displays the SMART website interface, which is used for conserved domain searches. The page is divided into two main sections: "Sequence analysis" and "Architecture analysis".

SMART MODE: NORMAL GENOMIC

SMART: Simple Modular Architecture Research Tool

Search SMART (keywords...)

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.

Conserved domain search - 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.ENSP00000319788, P15559, ENSP00000319788.5, ENSP00000319788, 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.

Conserved domain search - SMART

SMART SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

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

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](#).





Conserved domain search - InterPro


InterPro Classification of protein families

Home Search Browse **Results** Release notes Download Help About

Job ID **i** iprscan5-R20230301-115158-0009-77866479-p1m

Length 274 amino acids

Actions  




Status  finished

Expires **i** Wed Mar 08 2023

Protein family membership

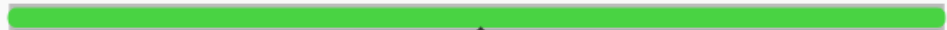
None predicted

Entry matches to this protein **i**

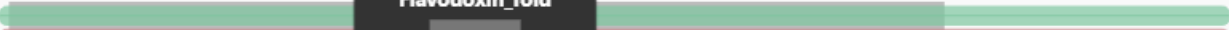
  Options  Export


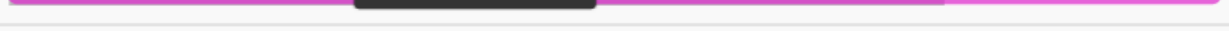
1 20 40 60 80 100 120 140 160 180 200 220 240 260 274

▼ Domain



 IPR003680 PF02525


▼ Homologous Superfamily

 IPR029039 G3DSA:3.40.50.360 SSF52218

▼ Unintegrated

 G3DSA:3.40.50.360:FF:000029 

 PTHR10204

InterPro IPR003680
Flavodoxin_fold
5 - 211

Practical part

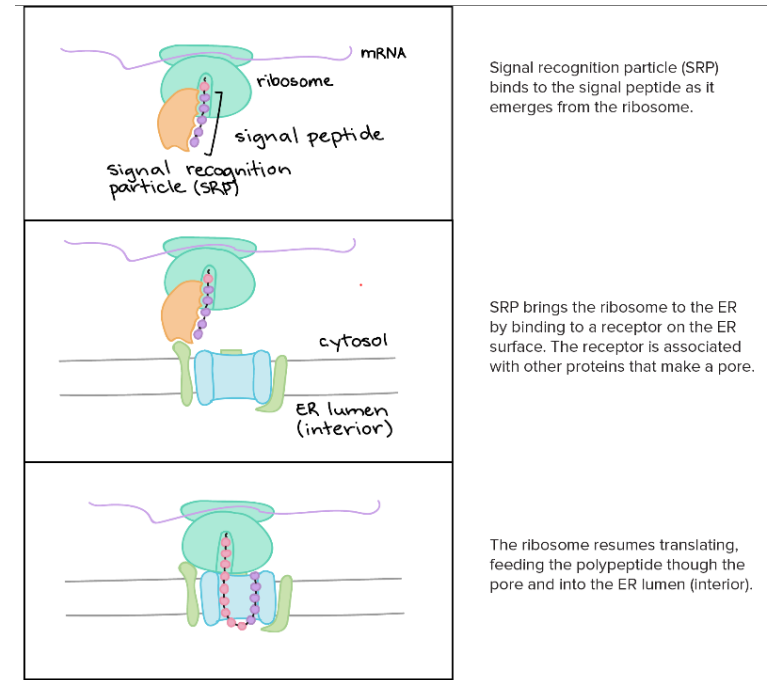
Try
CD / SMART/ InterPro
search

Find domains in your sequence



ER signal peptide prediction

Endoplasmic reticulum signal peptide: 15-60 amino acids on protein N-terminus



Signal peptides

SignalP

DTU Health Tech

Research Publications Education Collaboration Services and Products News About

SignalP 6.0 is based on a [transitional 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|
```



Signal peptides

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

```
MVVAATVAAAMLLWAAACAQQEQDFYDFKAVNIRGKLVLSLEKYRGSVSLVVNVA  
SECGFT  
DQHYRALQQLQRLDLPHHFNVLAFFPCNQFGQQEPDSNKEIESFARRTYSVSFFM  
FSKIIV  
TGIGAHPAFKYLAQTSGKEPTWNEWKYLVA PDGKVVGAWDFTVSVEEVRPQITA  
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

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:

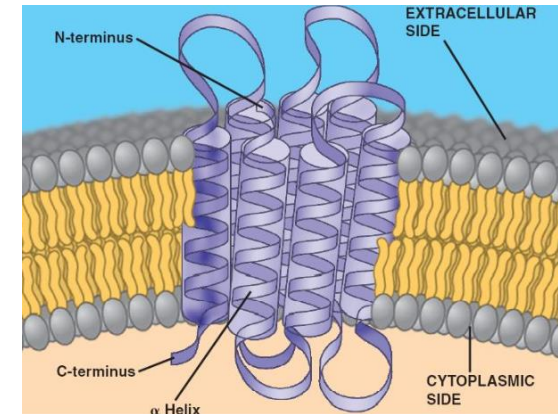
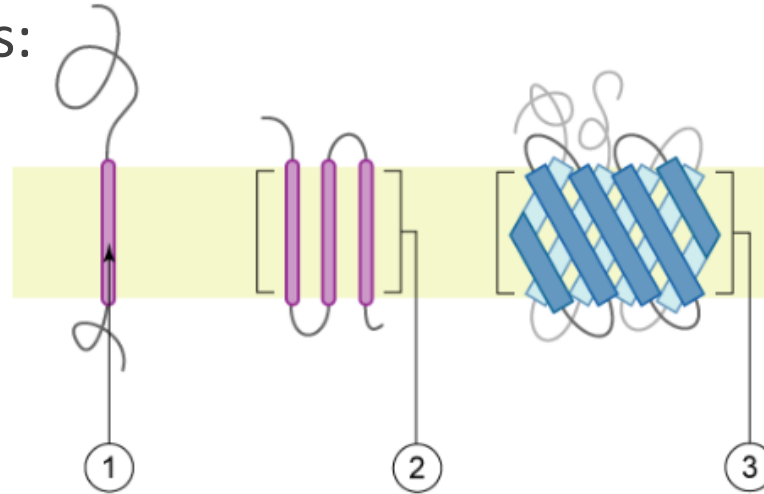
Name	Start ▲	End	E-value
signal peptide	1	18	N/A
Pfam:UDPGT	19	523	8e-64

Practical part

search for signal peptide in
your sequence

Prediction of transmembrane helices

Transmembrane proteins:



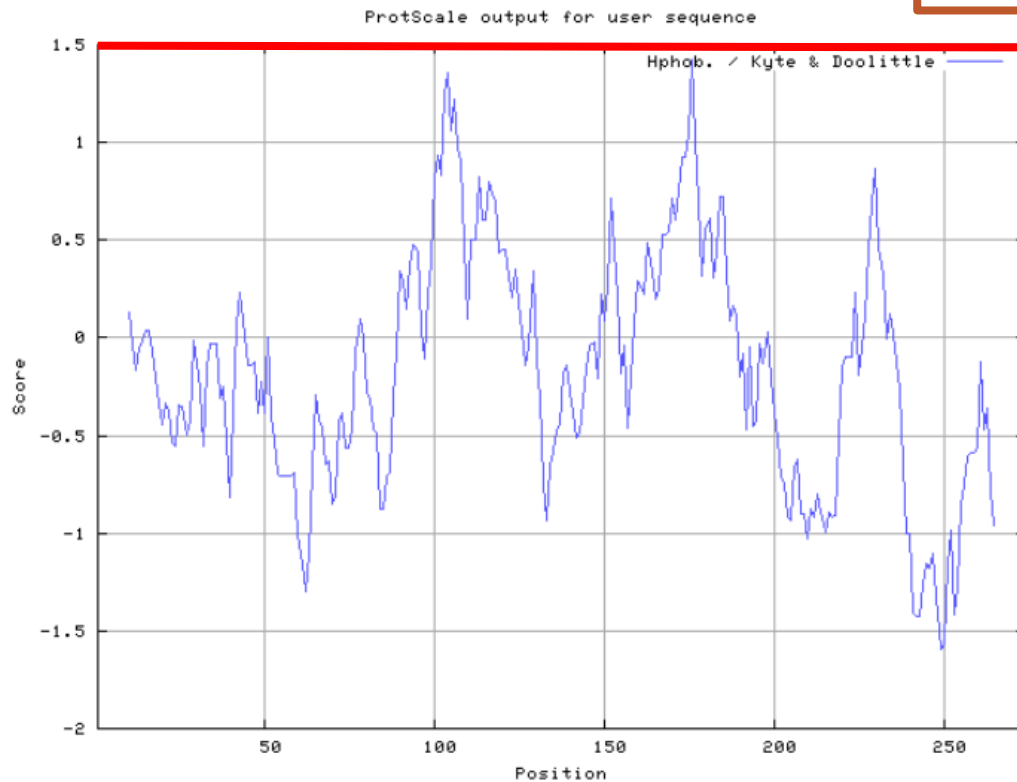
Amino acid Hydrophobicity

- various programs – different algorithms – different results
- Topological predictions (estimation of in and out topology)

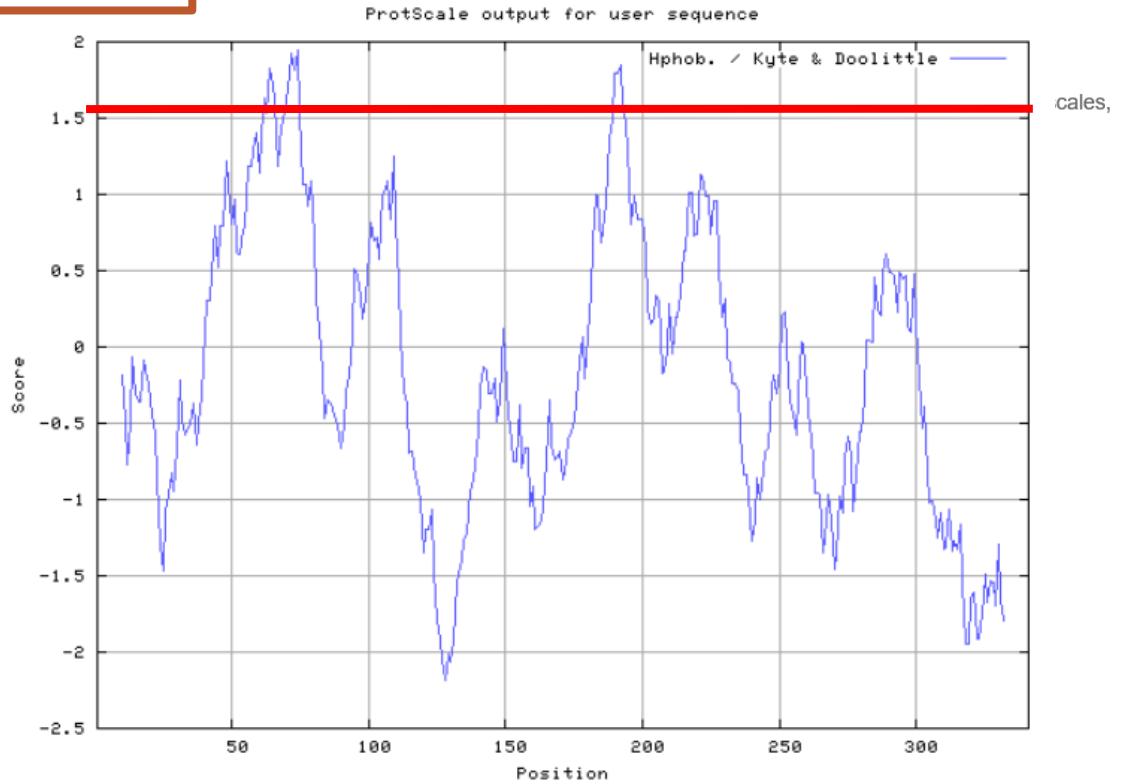
Prediction of transmembrane helices

Profile of amino acids hydrofobicity

Cut off: 1.5 (1.6)



U
his
ific
,
N



TMHMM

CBS >> CBS Prediction Servers >> TMHMM

TMHMM Server v. 2.0
Prediction of transmembrane helices in proteins

TMHMM result
[HELP](#) with output formats

no TM helix

```
# 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 posterior probabilities for WEBSEQUENCE

Submission
Submission of a local file in **FASTA** form
Procházet...

OR by pasting sequence(s) in **FASTA** form

Output format:
 Extensive, with graphics
 Extensive, no graphics
 One line per protein

Other options:
 Use old model (version 1)

Submit Clear

Restrictions:
At most 10,000 sequences and 4,000,000 a

Confidentiality:
The sequences are kept confidential and wi

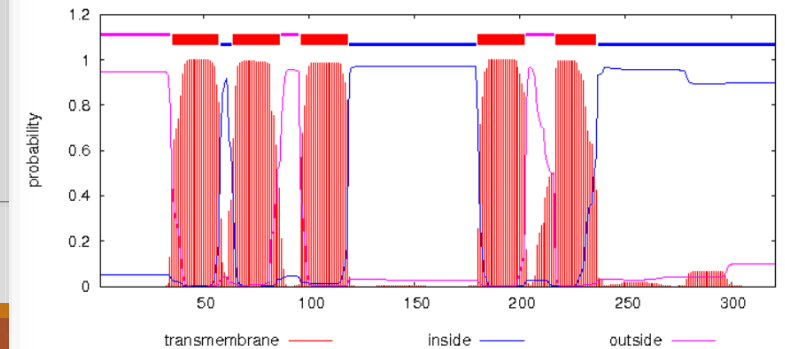
TMHMM result

[HELP](#) with output formats

five predicted TM helices

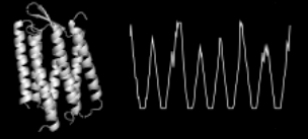
```
# WEBSEQUENCE Length: 321
# WEBSEQUENCE Number of predicted TMHs: 5
# WEBSEQUENCE Exp number of AAs in TMHs: 108.47546
# WEBSEQUENCE Exp number, first 60 AAs: 21.62676
# WEBSEQUENCE Total prob of N-in: 0.05151
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE TMHMM2.0 outside 1 34
WEBSEQUENCE TMHMM2.0 TMhelix 35 57
WEBSEQUENCE TMHMM2.0 inside 58 63
WEBSEQUENCE TMHMM2.0 TMhelix 64 86
WEBSEQUENCE TMHMM2.0 outside 87 95
WEBSEQUENCE TMHMM2.0 TMhelix 96 118
WEBSEQUENCE TMHMM2.0 inside 119 179
WEBSEQUENCE TMHMM2.0 TMhelix 180 202
WEBSEQUENCE TMHMM2.0 outside 203 216
WEBSEQUENCE TMHMM2.0 TMhelix 217 236
WEBSEQUENCE TMHMM2.0 inside 237 321
```

TMHMM posterior probabilities for WEBSEQUENCE



TOPCONS

TOPCONS



TOPCONS

TOPCONS

New query

Batch WSDL API

Download

References

News

Server status

Example results

Old TOPCONS

Help

Results

- Submitted: 2018-03-05 15:59:14
- Status: **Finished**
- Waiting time: 0 sec
- Running Time: 0 sec

Results of your prediction with jobid: **rst_BKOIKK**

Zipped folder of your result can be found in [rst_BKOIKK.zip](#)

Dumped prediction in one text file can be found in [query.result.txt](#)

The sequence(s) you submitted can be found in [query.raw.fa](#)

Predicted topologies and predicted ΔG values:

Method	Inside	Outside	TM-helix (IN->OUT)	TM-helix (OUT->IN)	Signal peptide
TOPCONS	0	1	1	1	1
OCTOPUS	0	1	1	1	1
Philius	0	1	1	1	1
PolyPhobius	0	1	1	1	1
SCAMPI	0	1	1	1	1
SPOCTOPUS	0	1	1	1	1

PDB-homology: *****No homologous TM proteins detected*****

© Arne Elofsson

Your recent jobs:

Queued	Running	Finished	Failed
0	0	5	0

Results

- Submitted: 2018-03-05 15:59:14
- Status: **Finished**
- Waiting time: 1 sec
- Running Time: 28 sec

Results of your prediction with jobid: **rst_BKOIKK**

Zipped folder of your result can be found in [rst_BKOIKK.zip](#)

Dumped prediction in one text file can be found in [query.result.txt](#)

The sequence(s) you submitted can be found in [query.raw.fa](#)

Predicted topologies and predicted ΔG values:

Method	Inside	Outside	TM-helix (IN->OUT)	TM-helix (OUT->IN)	Signal peptide
TOPCONS	1	1	1	1	1
OCTOPUS	1	1	1	1	1
Philius	1	1	1	1	1
PolyPhobius	1	1	1	1	1
SCAMPI	1	1	1	1	1
SPOCTOPUS	1	1	1	1	1

© Arne Elofsson

Consensus prediction of membrane proteins and signal peptides

Please paste your amino acid sequences in [FASTA](#) format. Allowed characters: "ABCDEFGHIJKLMNPQRSTUVWXYZ*", " ". (Sequences should be no shorter than 10 amino acids)

Alternatively, upload a text file in FASTA format upto 1000 amino acids. [Procházet...](#)

Job name (optional):

Email (recommended for batch submissions):

Force run (do not use cached results):

[Submit](#) [Clear](#) [Generate example input](#)

© Arne Elofsson

New query

Batch WSDL API

Download

References

News

Server status

Example results

Old TOPCONS

Help

Your recent jobs:

Queued	Running	Finished	Failed
0	0	25	0

Phobius



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

Phobius

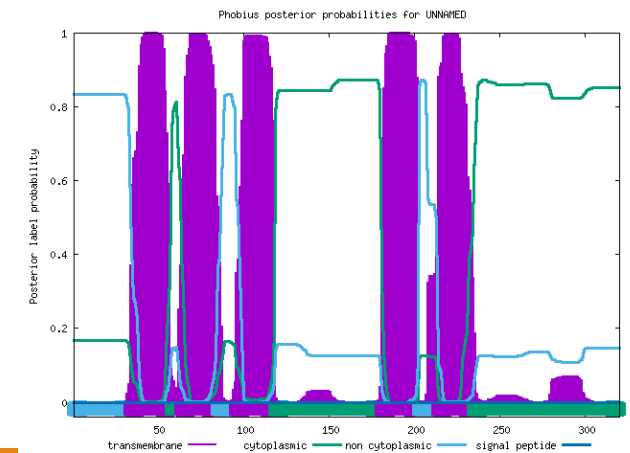
A combined transmembrane topology and signal peptide predictor

Phobius prediction

Prediction of UNNAMED

```
ID UNNAMED
FT TOPO_DOM 1 33 NON CYTOPLASMIC.
FT TRANSHEM 34 57
FT TOPO_DOM 58 63 CYTOPLASMIC.
FT TRANSHEM 64 84
FT TOPO_DOM 85 95 NON CYTOPLASMIC.
FT TRANSHEM 96 118
FT TOPO_DOM 119 180 CYTOPLASMIC.
FT TRANSHEM 181 202
FT TOPO_DOM 203 213 NON CYTOPLASMIC.
FT TRANSHEM 214 234
FT TOPO_DOM 235 320 CYTOPLASMIC.
//
```

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



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

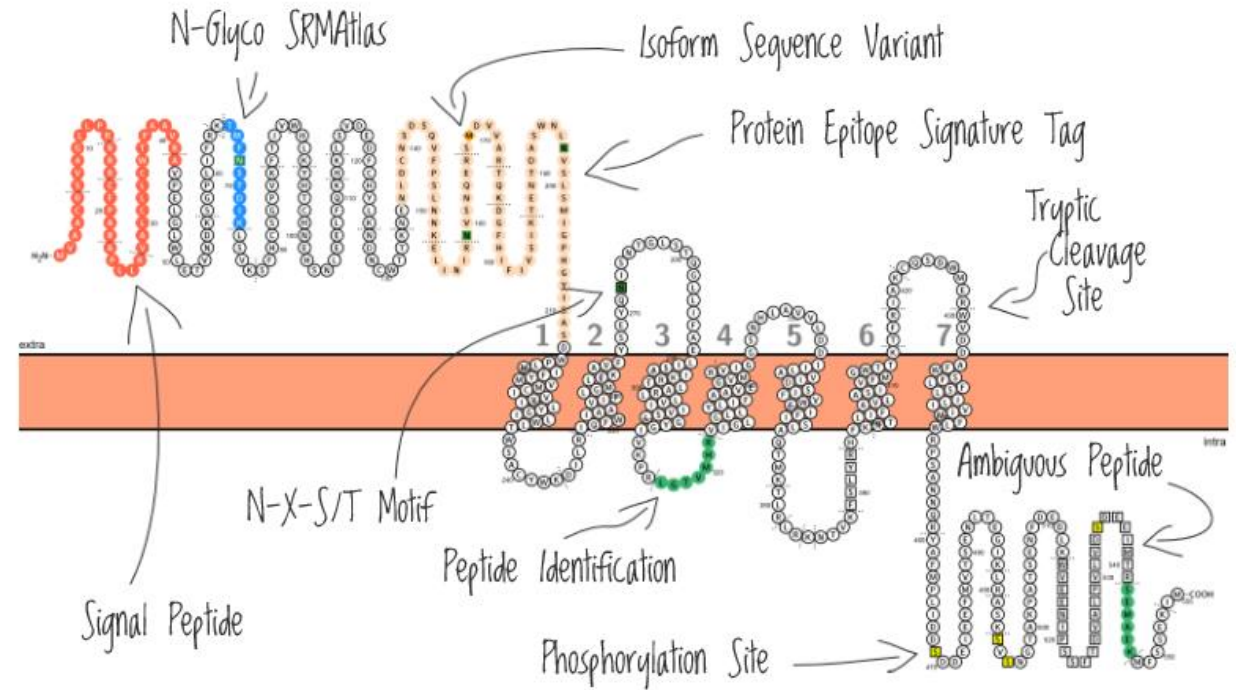
PROTTER-figure!

-creates figure from the UniProt data

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-figure!

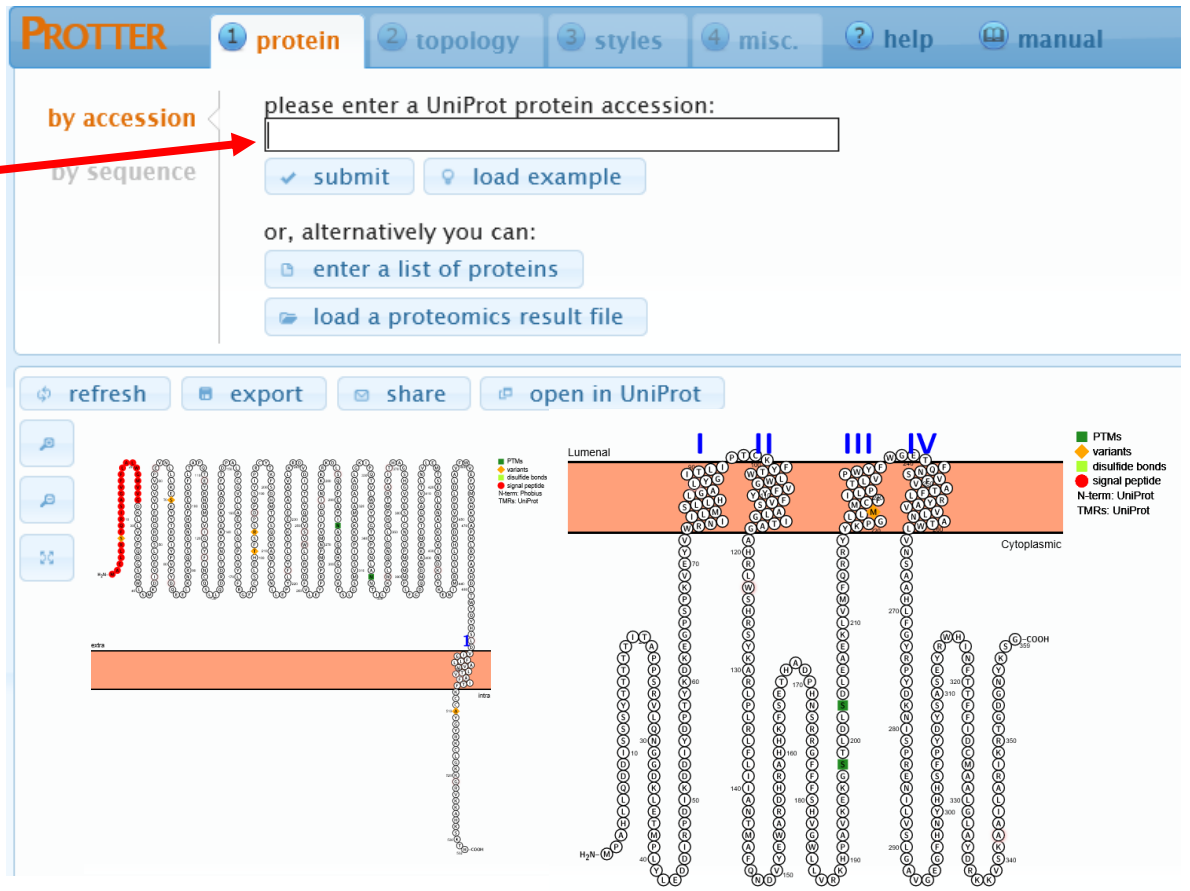
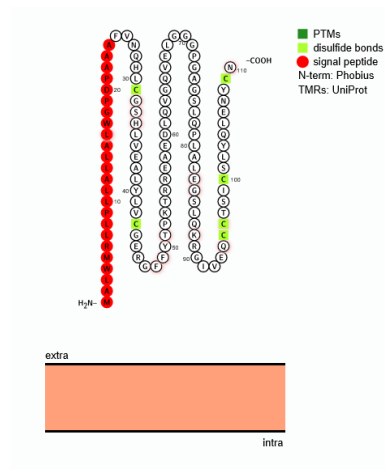
-creates figure from the UniProt data

-uses UniProt ID:

UGT1A6 (P19224)

Desaturase (O00767)

(prepro) insulin (P01308)



PROTTER-figure!

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

by accession
by sequence

or, alternatively you can:

Luminal
1 2 3 4
Cytoplasmic
H₂N
C-COOH

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

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

by accession

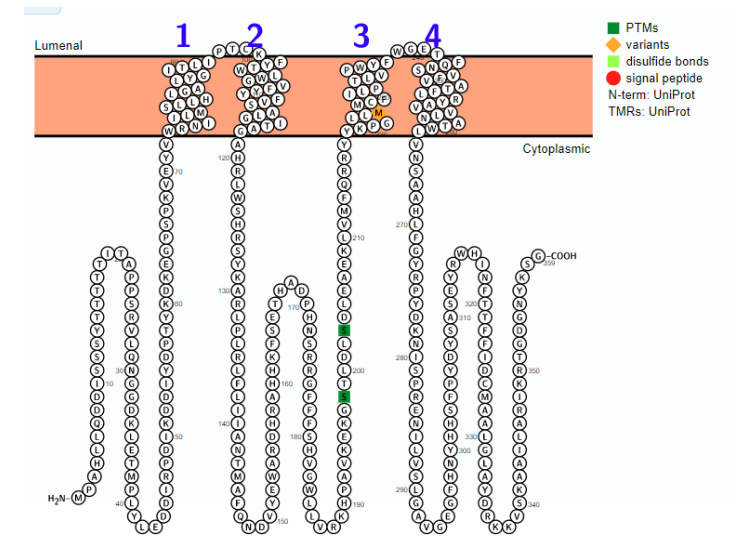
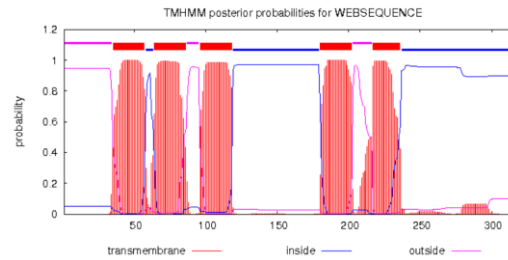
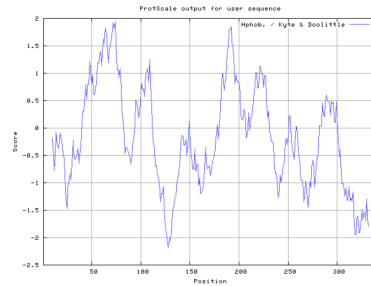
by sequence

extra
1 2 3 4 5 6
intra
H₂N
C-COOH

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

Prediction of transmembrane helices

ProtScale
 TMHMM
 TOPCONS



➤ Always try more programs!

Practical part

Try more programs.

Does your sequence have any TMHs?
and/or signal peptide?

„Protein bioinformatics I“

Retrieving protein sequences from databases (Uniprot: FASTA formate)

Computing amino-acids compositions, molecular weight, isoelectric point, and other parameters (SMS)

Prediction of proteases cutting (PeptideCutter)

Predicting elements of protein secondary structure, signal peptide, transmembrane helix

Finding 3-D structure and the domain organization of proteins

Finding all proteins that share a similar sequence and Classifying proteins into families

Finding evolutionary relationships between proteins, drawing proteins' family trees

Computing the optimal alignment between two or more protein sequences

...

Searching for similar sequences

Similarity x Homology

BLAST: Basic Local Alignment and Search Tool

Finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

Similarity matrix:

„Leucine is more similar to Isoleucine than Histidine“

The BLOSUM62 similarity matrix

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
A	4	-2	0	-2	-1	-2	0	-2	-1	-1	-1	-1	-2	-1	-1	-1	1	0	0	-3	-1	-2	-1
B	-2	6	-3	6	2	-3	-1	-1	-3	-1	-4	-3	1	-1	0	-2	0	-1	-3	-4	-1	-3	2
C	0	-3	9	-3	-4	-2	-3	-3	-1	-3	-1	-1	-3	-3	-3	-1	-1	-1	-2	-1	-2	-4	-4
D	-2	6	-3	6	2	-3	-1	-1	-3	-1	-4	-3	1	-1	0	-2	0	-1	-3	-4	-1	-3	2
E	-1	2	-4	2	5	-3	-2	0	-3	1	-3	-2	0	-1	2	0	0	-1	-2	-3	-1	-2	5
F	-2	-3	-2	-3	-3	6	-3	-1	0	-3	0	0	-3	-4	-3	-3	-2	-2	-1	1	-1	3	-3
G	0	-1	-3	-1	-2	-3	6	-2	-4	-2	-4	-3	0	-2	-2	-2	0	-2	-3	-2	-1	-3	-2
H	-2	-1	-3	-1	0	-1	-2	8	-3	-1	-3	-2	1	-2	0	0	-1	-2	-3	-2	-1	2	0
I	-1	-3	-1	-3	0	-4	-3	4	-3	2	1	-3	-3	-3	-2	-1	3	3	-1	-1	-3	-1	-3
K	-1	-1	-3	-1	1	-3	-2	-1	-3	5	-2	-1	0	-1	1	2	0	-1	-2	-3	-1	-2	1
L	-1	-4	-1	-4	-3	0	-4	-3	2	-2	4	2	-3	-3	-2	-2	-2	-1	1	-2	-1	-1	-3
M	-1	-3	-1	-3	-2	0	-3	-2	1	-1	2	5	-2	-2	0	-1	-1	-1	-1	-1	-1	-1	-2
N	-2	1	-3	1	0	-3	0	1	-3	0	-3	-2	6	-2	0	0	1	0	-3	-4	-1	-2	0
P	-1	-1	-3	-1	-1	-4	-2	-2	-3	-1	-3	-2	-2	7	-1	-2	-1	-1	-2	-4	-1	-3	-1
Q	-1	0	-3	0	2	-3	-2	0	-3	1	-2	0	0	-1	5	1	0	-1	-2	-2	-1	-1	2
R	-1	-2	-3	-2	0	-3	-2	0	-3	2	-2	-1	0	-2	1	5	-1	-1	-3	-3	-1	-2	0
S	1	0	-1	0	0	-2	0	-1	-2	0	-2	-1	1	-1	0	-1	4	1	-2	-3	-1	-2	0
T	0	-1	-1	-1	-1	-2	-2	-2	-1	-1	-1	-1	0	-1	-1	-1	1	5	0	-2	-1	-2	-1
V	0	-3	-1	-3	-2	-1	-3	-3	3	-2	1	1	-3	-2	-2	-3	-2	0	4	-3	-1	-1	-2
W	-3	-4	-2	-4	-3	1	-2	-2	-3	-3	-2	-1	-4	-4	-2	-3	-3	-2	-3	11	-1	-2	3
X	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Y	-2	-3	-2	-3	-2	3	-3	2	-1	-2	-1	-1	-2	-3	-1	-2	-2	-2	-1	2	-1	7	-2
Z	-1	2	-4	2	5	-3	-2	0	-3	1	-3	-2	0	-1	2	0	0	-1	-2	-3	-1	-2	5

NCBI/BLAST

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

The screenshot shows the NCBI BLAST website interface. At the top left is the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". On the top right, there is a user profile icon labeled "jostovap". Below the header is a navigation bar with "BLAST" and links for "Home", "Recent Results", "Saved Strategies", and "Help".

The main content area features a "Basic Local Alignment Search Tool" section with a description: "BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance." A "Learn more" link is provided. To the right is a "NEWS" box with the headline "ElasticBLAST 1.0.0 is Now available!" and a sub-headline: "ElasticBLAST version 1.0.0 has support for faster cheaper disks at AWS and better supports Kubernetes on GCP!". The date "Mon, 09 Jan 2023" and a "More BLAST news..." link are also present.

Below the news box is the "Web BLAST" section, which contains three main options:

- Nucleotide BLAST**: nucleotide → nucleotide (represented by a DNA double helix icon).
- blastx**: translated nucleotide → protein (represented by a blue arrow pointing right).
- tblastn**: protein → translated nucleotide (represented by a blue arrow pointing left).
- Protein BLAST**: protein → protein (represented by a protein ribbon structure icon, which is highlighted with a red border).

At the bottom is the "BLAST Genomes" section, featuring a search input field with the placeholder text "Enter organism common name, scientific name, or tax id" and a "Search" button. Below the input field are four tabs: "Human", "Mouse", "Rat", and "Microbes".

NCBI/BLAST

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

NCBI Resources How To jostovap My NCBI Sign Out

NCBI National Center for Biotechnology Information

All Databases Search

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.


[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [NCBI News](#)

Get Started

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

NCBI YouTube channel

Learn how to get the most out of NCBI tools and databases with video tutorials on the NCBI YouTube Channel.



Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[PubMed Health](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI Announcements

NCBI Video: Submitting manuscripts on NIHMS

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
ESSLAHKEGRLSPDIVAEHKKLEAADLVIFQFPLQWFGVPAILKGFVERVLVAGFAYTYAAMYDNGP
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)

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 Icl|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.**

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)

Links

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 MPAHLLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYK 60

MPAHLLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYK

Sbjct 1 MPAHLLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYK 60

Query 61 DKEGSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH 120

DKEGSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH

Sbjct 61 DKEGSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH 120

Query 121 RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHKKFSETHADPHNSRRGFFFS 180

RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHKKFSETHADPHNSRRGFFFS

Sbjct 121 RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHKKFSETHADPHNSRRGFFFS 180

Query 181 HVGWLLVRKHPAVKEKGSTLDLSLEAEKLVMFQRRYYKPGLLMCFILPTLVPWYFWGE 240

HVGWLLVRKHPAVKEKGSTLDLSLEAEKLVMFQRRYYKPGLLMCFILPTLVPWYFWGE

Related Information

[Gene](#) - associated gene details

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

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

→change sequences (FASTA) names into organism only

NCBI/BLAST

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download

select all 5 sequences selected

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
MPAHLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKEGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCMCFILPTLVPWYFWGE
TFQNSVVFVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LARIKRTGDGNYKSG
>Gorilla gorilla gorilla
MPAHLQDDISSSYTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKEGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCMCFILPTLVPWYFWGE
TFQNSVVFVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LARIKRTGDGNYKSG
>Pan troglodytes
MPAHLQDDITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKEGSPKVEYVWRNIILMS
LLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVYEWARHRAH
HKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMPQRRYYKPGLLMCMCFILPTLVPWYFWGE
TFQNSVVFVATFLRYAVVLNATWLVNSAAHLFGYRYPYDKNISPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LARIKRTGDGNYKSG
>Camelus ferus
MPAHLQEEISSSYTTTTITAPPSRVLQNGGDKLEKTPLYLEEDIRPEMKDDIYDPSYQDKEGPKPKVYVWRNIILMG
LLHLGALYGITLIPTCKFYTFQVWLFYYIIISALGITAGAHRLWSHRSYKARLPRLRFLIIANTMAFQNDVFEWARDHRAH
HKFSETDADPHNSRRGFFFSHVGWLLVRKHPAVKEKGGLLDSDLKAEKLVMPQRRYYKPGILLMCFIMPTLVPWYFWGE
TFQHSLYLATFLRYAVVLNVTWLVNSAAHLGYRYPYDKTINPRENI LVS LGAVGEGGFHNYHHSFPYDYSASEYRWHINFT
TFPIDCMAALGLAYDRKGVSKAAI LAKVKRTGDGSYKSG
>Ovis aries
MPAHLQEEISSSYTTTTITAPPSRVLQNGGDKLEKTPLYLEEDIRPEMRDDIYDPTYQDKEGPKPKLEYVWRNIILMG
LLHLGALYGITLIPTCKIYTLWVLFYYVISALGITAGVHRLWSHRTYKARLPRLRFLIIANTMAFQNDVFEWRSRDHRAH
HKFSETDADPHNSRRGFFFSHVGWLLVRKHPAVREKGGATLDLSDLRAEKLVMPQRRYYKPGVLLLCFILTLPVWYLWGE
TFQNSLFFATFLRYAVVLNATWLVNSAAHMYGYRYPYDKTINPRENI LVS LGAVGEGGFHNYHTFPYDYSASEYRWHINFT
TFPIDCMAAIGLAYDRKGVSKAAV LARMKRTGEEYSYKSG
>gi|13435426|gb|AAH04579.1| Nqo1 pr
MAARRALIVLAHSEKTSFNAMKEAAVEALKKRGW
LSPDIVAEHKKLEAADLVIFQFPLQWFGVPAAILKKG
VHGMNVILWPIQSGILHFCGFQVLEPQLVYSIGH
>gi|524939198|ref|XP_005071892.1| P
MÄVRRALIVLAHSEKTSFNAMKEAAVEALKKRGW
LSPDIVAEQKKLEAADLVIFQFPLHWFQVPAAILKKG
VHGMNIIWPIQSGILHFCGFQVLEPQLVYSIGHTPPDARTQILEGWKKRLETVWDETPLYFVPSLFDLNFQAGFLKKEVQEEQKKNRGLSVGHHLGKSIPTDQVQKARK
>gi|227430403|ref|NP_001153085.1| NAD(P)H dehydrogenase [quinone] 1 [Sus scrofa]
MÄVRKALIVLAHSEKTSFNAMKEAAVEALKRRGWEVAVSDLYAMNPNVISRDKITGKLDKDPGNFQYPAETALAYKEGR
LSPDIVAEQKKVEAADLVIFQFPLQWFGVPAAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQ
IHGMNIIWPIQSGILHFCGFQVLEPQLTYSIGHTPEDARTQILEEWKKRLENWDETPLYFAPSSLFDLNFQAGFLMKKQVQDEQKSNKGLSVGHHLGKSIPTDQVQKARK
>gi|386781783|ref|NP_001247927.1| NAD(P)H dehydrogenase [quinone] 1 [Macaca mulatta]
MVGKRALIVLAHSEKTSFNAMKEAAVAALKKKGWEVAVSDLYAMNPNVISRDKITGKLDKDPANFYAAESTLAYKEGR
LSPDIVAEQKKLEAADLVIFQFPLQWFGVPAAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQ
IHGMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARTQILEGWKKRLENWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKFFGLSVGHHLGKSIPTDQVQKARK
>gi|426242583|ref|XP_004015151.1| PREDICTED: NAD(P)H dehydrogenase [quinone] 1 [Ovis aries]
MÄVRKALIVLAHSEKTSFNAMKEAAEALKRRGWEVTVSDLYAMNPNVISRDKITGKLDKDPGNFYPAETVLAAYKEGR
LSPDIVAEQKKLEAADLVIFQFPLQWFGVPAAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQ
IHGMNIIWPIQSGILHFCGFQVLEPQLTYSIGHTPADARVQILEGWKKRLENWDEMPLYFAPSSLFDLNFQAGFLMKKEVQDEQKSKFFGLSVGHHLGKSIPTDQVQKARK
>gi|30230685|gb|AAP20940.1| NAD(P)H dehydrogenase, quinone 1 [Homo sapiens]
RRALIVLAHSEKTSFNAMKEAAALKKKGWEVSDLYAMNPNVISRDKITGKLDKDPANFYPAESVLAAYKEGHLSP
DIVAEQKKLEAADLVIFQFPLQWFGVPAAILKKGWFERVLEGEFAYTYAAMYDYGPFRRNKAVALSITTTGSGSMYSLQIGIH
DMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKKRLENWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKFFGLSVGHHLGKSIPTDQVQKARK
```

NCBI/BLAST (reference proteins)

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](#)

MAARRALIVLAHSEKTSFNAMKEAAVEALKKRGWEVLES...
YPS
ESSLAHKEGRLSPDIVAEHKKLEADLVIQFPLQWFGVPAILKGFVRLVAGFAYTYAAMYDNGP
FQN
KKTLLSITGGSGSMYSLQGVHGMNVILWPIQSGILRFQGFQVLEPQLVYSIGHTPPDARMQILEG

From
To

Or, upload file

Job Title
Enter a descriptive title for your BLAST search

Align two or more sequences

Standard

Database

Organism exclude

Exclude 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)

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

[Algorithm parameters](#)

Practical part

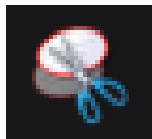
Try BLAST.

Download five similar sequences from
different organisms.

Homework 3

- 1) How many times will be the whole sequence cut by trypsin?
- 2) Does your sequence have a typical domain?
- 3) Does your sequence have transmembrane helix?
- 4) Does your sequence have ER retention signal?
- 5) Find and download five similar sequences.

E.g use „výstřížky“



„snipping tool“

- Compile in „one note“ (or word, or pdf)
- Submit via Moodle

