Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2023/2024

Základy praktické bioinformatiky

Téma 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

…

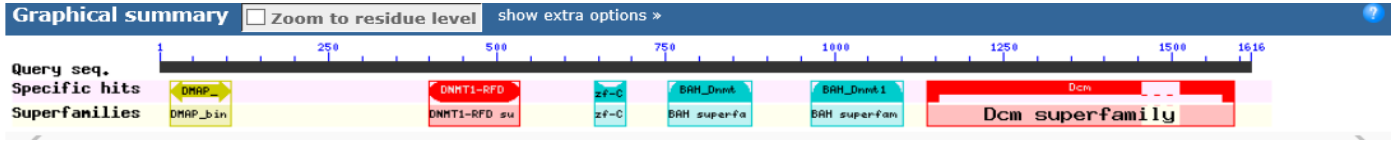
Hledání známých motivů

SEKVENCE➪STRUKTURA➪FUNKCE

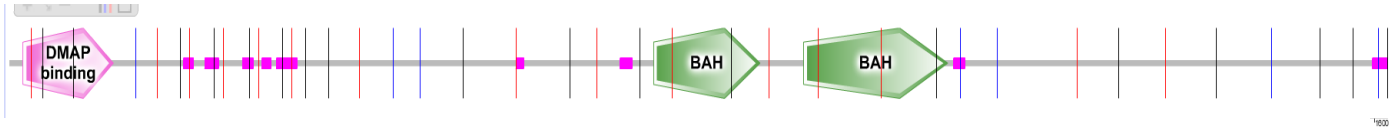
Sekvenční motivy, signální peptidy, místa posttranslačních modifikací ..

- Databáze známých motivů

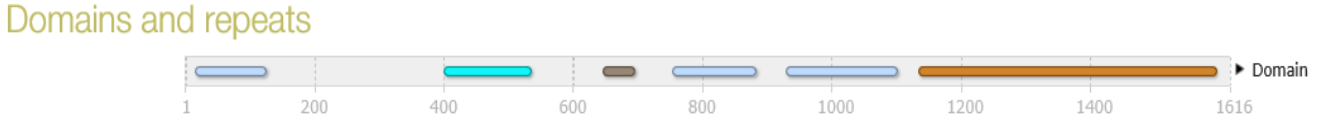
NCBI/CDD



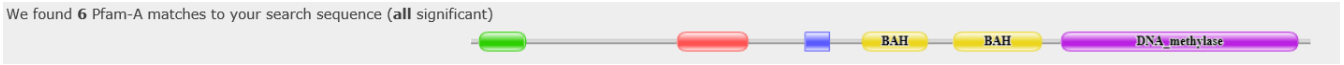
SMART



EMBL/InterPro



Pfam



Pfam

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

EMBL-EBI

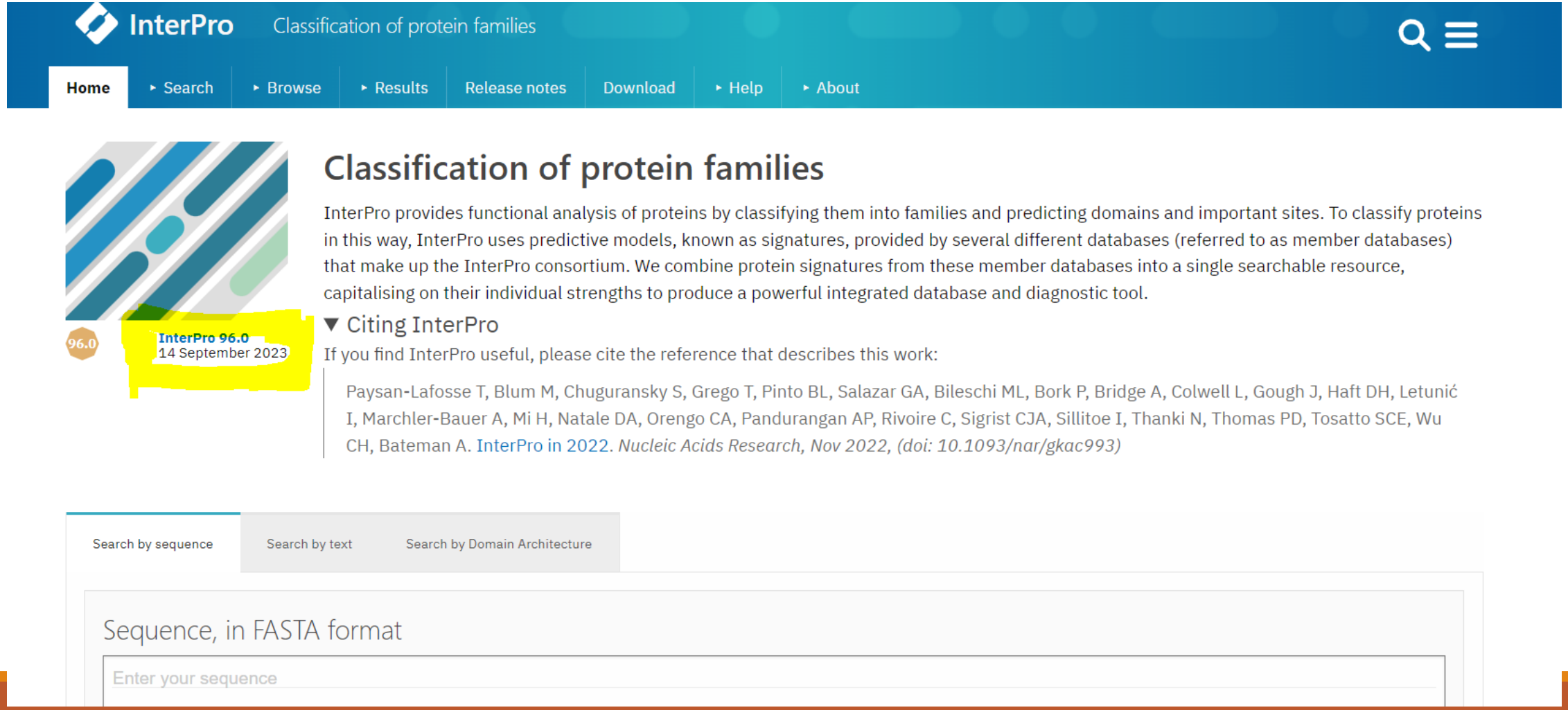
Pfam data and new releases are available through InterPro

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.

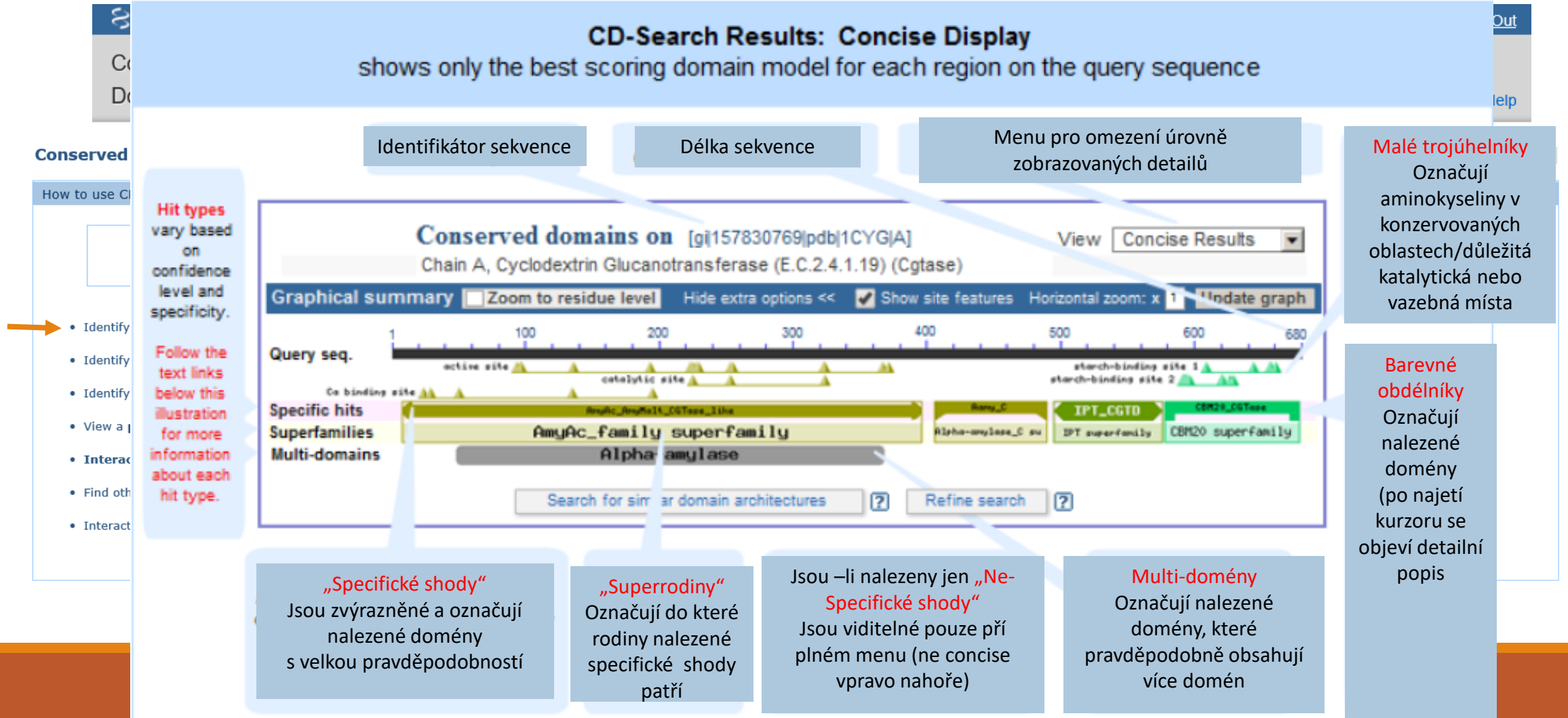
Pfam 36.0 (20 795 entries, 659 clans)

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

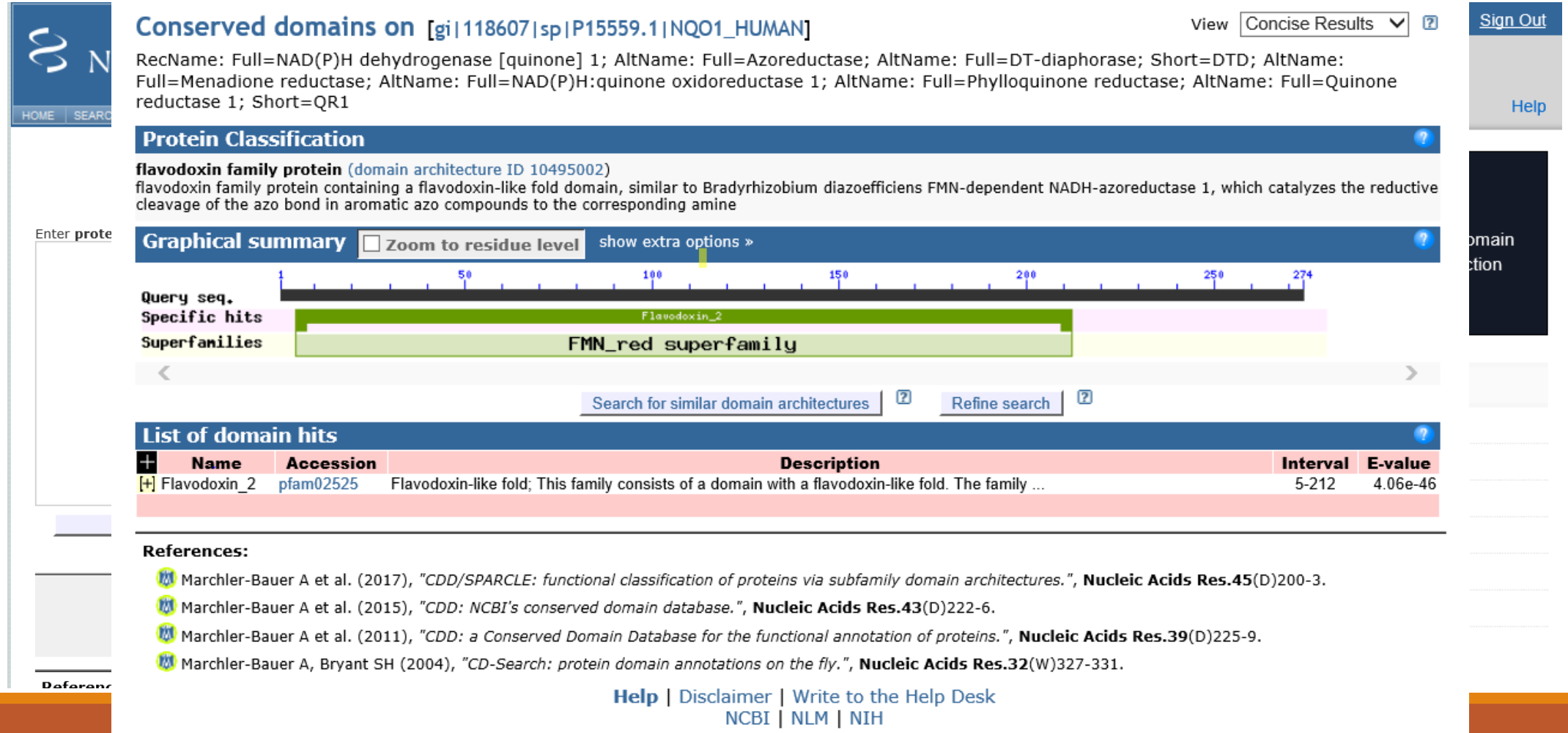
InterPro


InterPro
Classification of protein families
Home
Search
Browse
Results
Release notes
Download
Help
About
96.0
InterPro 96.0
14 September 2023
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)
Search by sequence
Search by text
Search by Domain Architecture
Sequence, in FASTA format
Enter your sequence

Hledání známých motivů-CD (NCBI)



Hledání známých motivů / NQO1



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 »

1 50 100 150 200 250 274

Query seq.
Specific hits: Flavodoxin_2
Superfamilies: FMN_red superfamily

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.

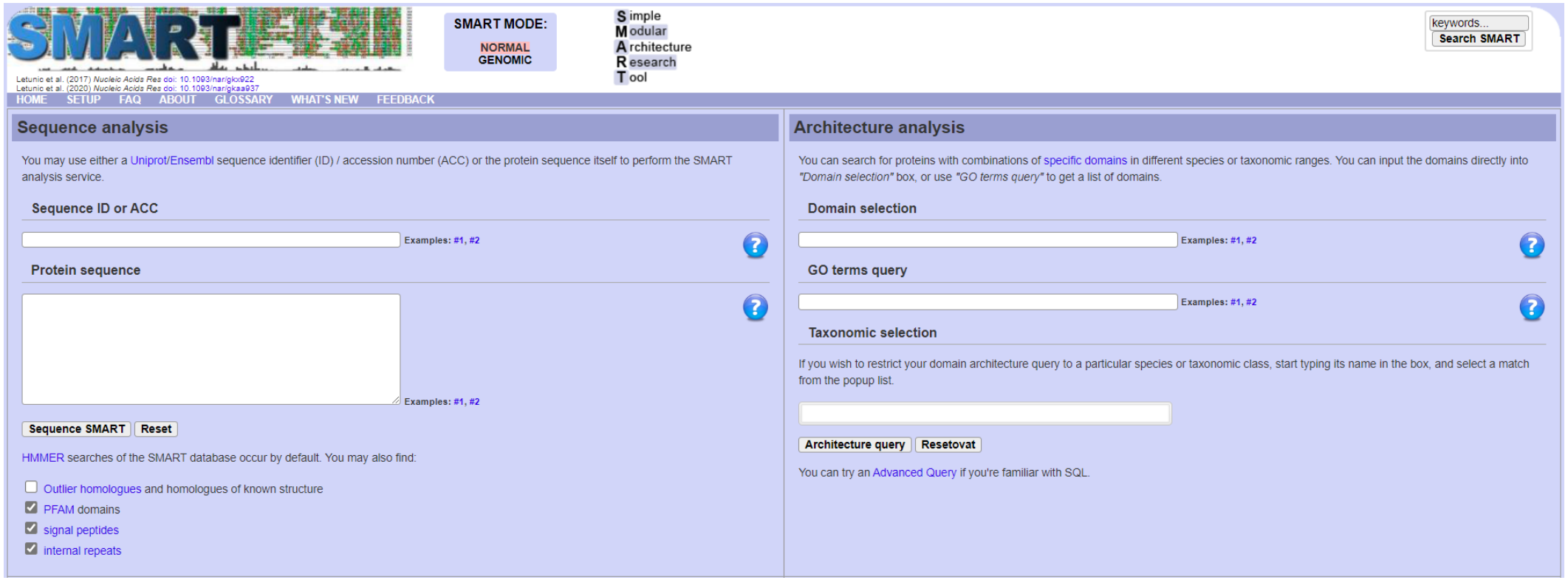
Marchler-Bauer A et al. (2015), *"CDD: NCBI's conserved domain database."*, **Nucleic Acids Res.43**(D)222-6.

Marchler-Bauer A et al. (2011), *"CDD: a Conserved Domain Database for the functional annotation of proteins."*, **Nucleic Acids Res.39**(D)225-9.

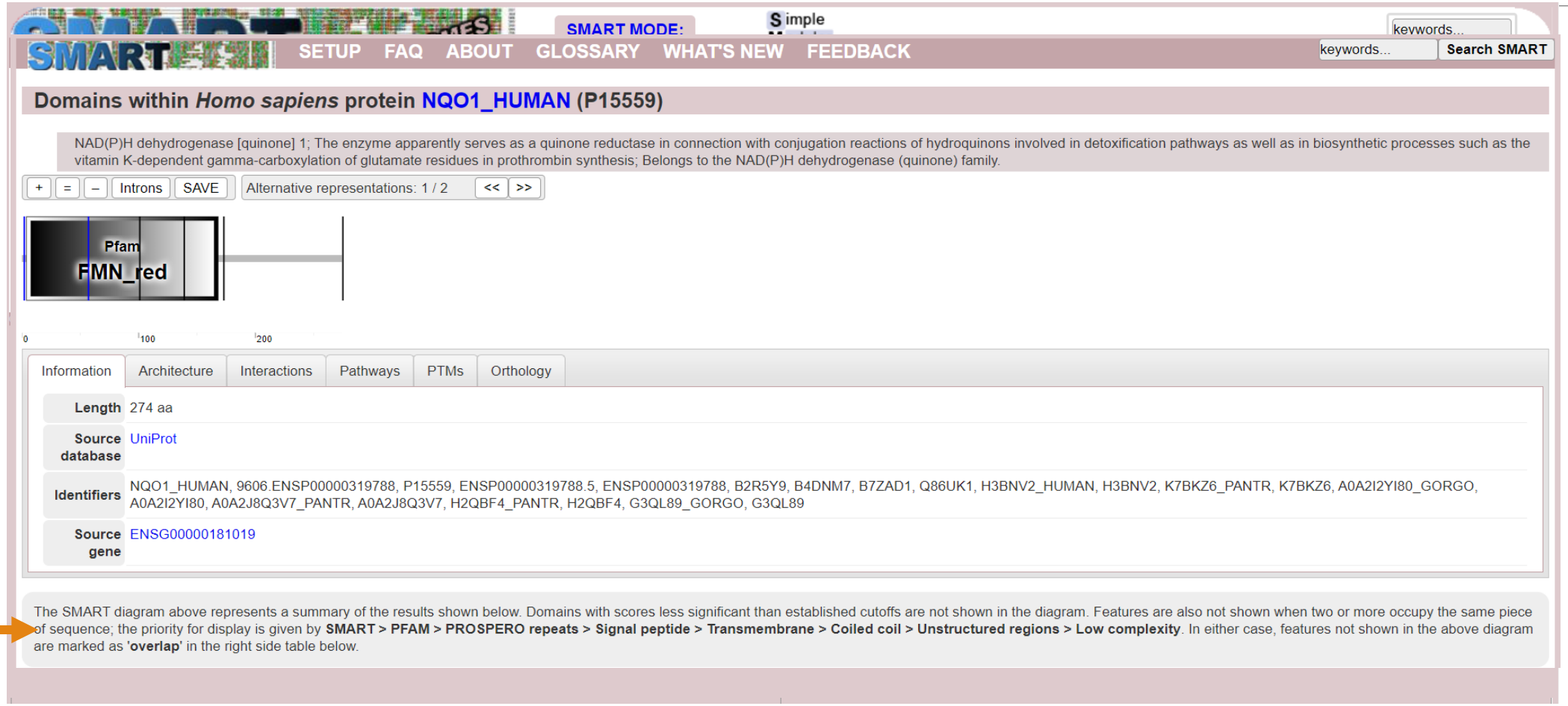
Marchler-Bauer A, Bryant SH (2004), *"CD-Search: protein domain annotations on the fly."*, **Nucleic Acids Res.32**(W)327-331.

Hledání známých motivů-SMART

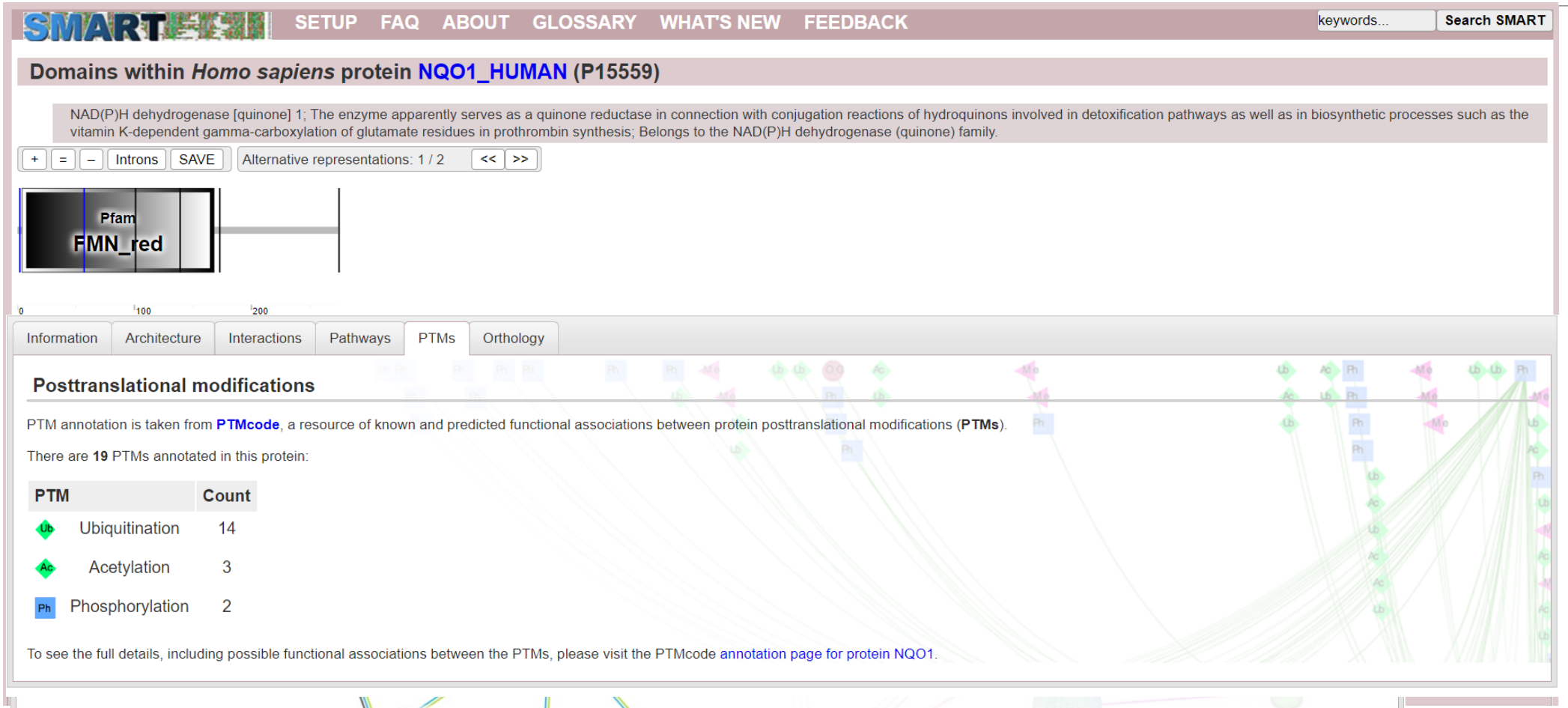
Hledání známých motivů-SMART



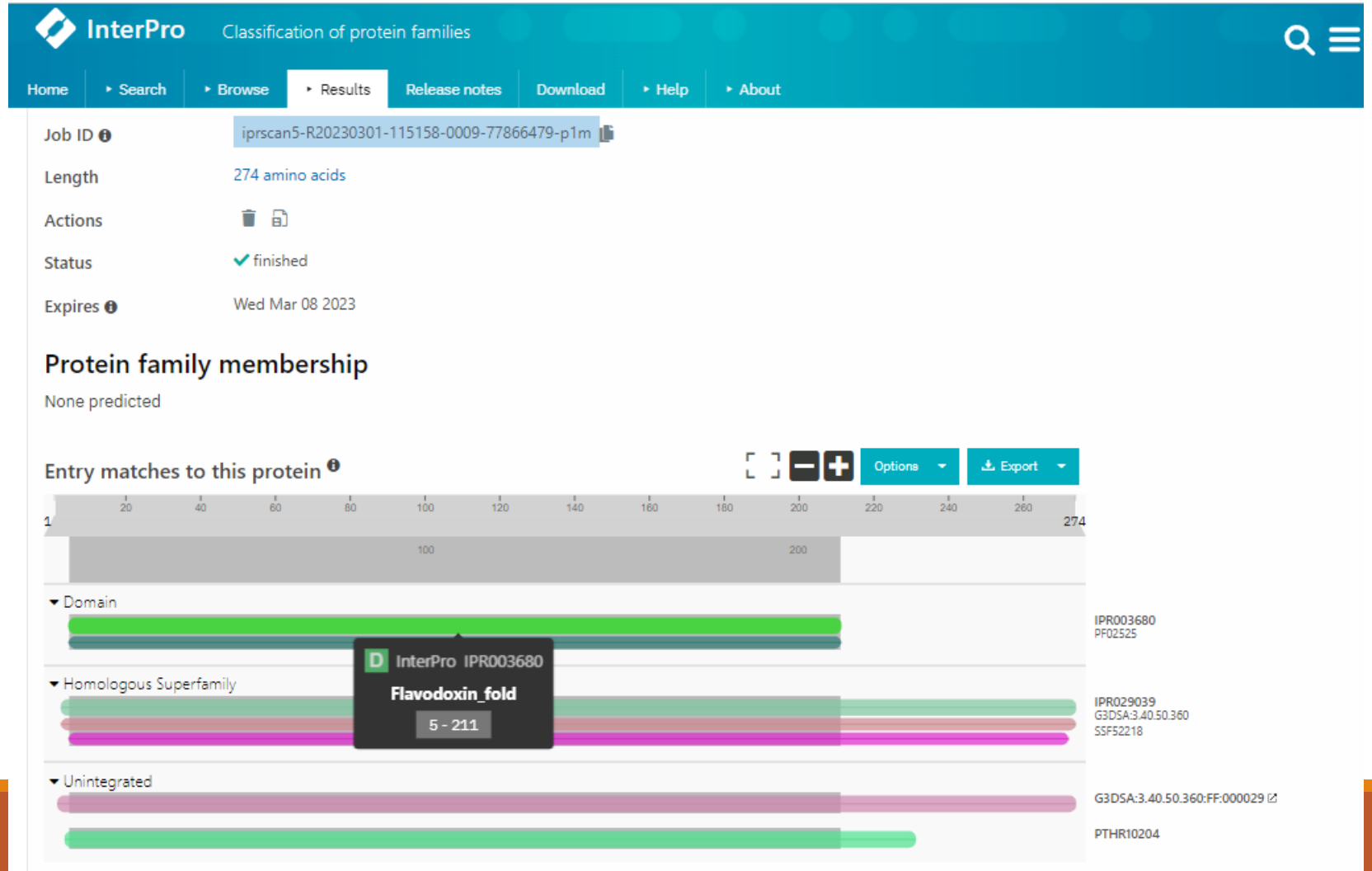
Hledání známých motivů-SMART



Hledání známých motivů-SMART



Hledání známých motivů-InterPro (2023)

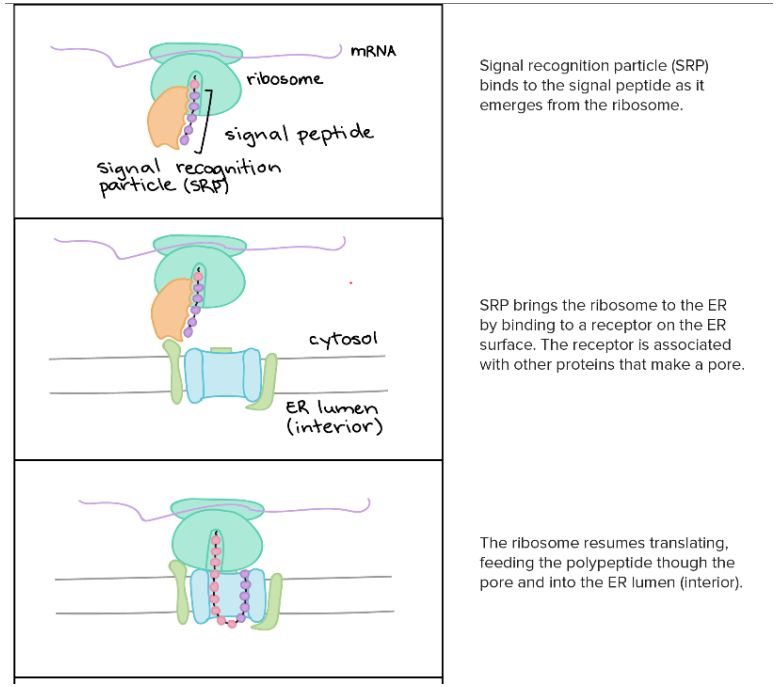


Praktická ukázka

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

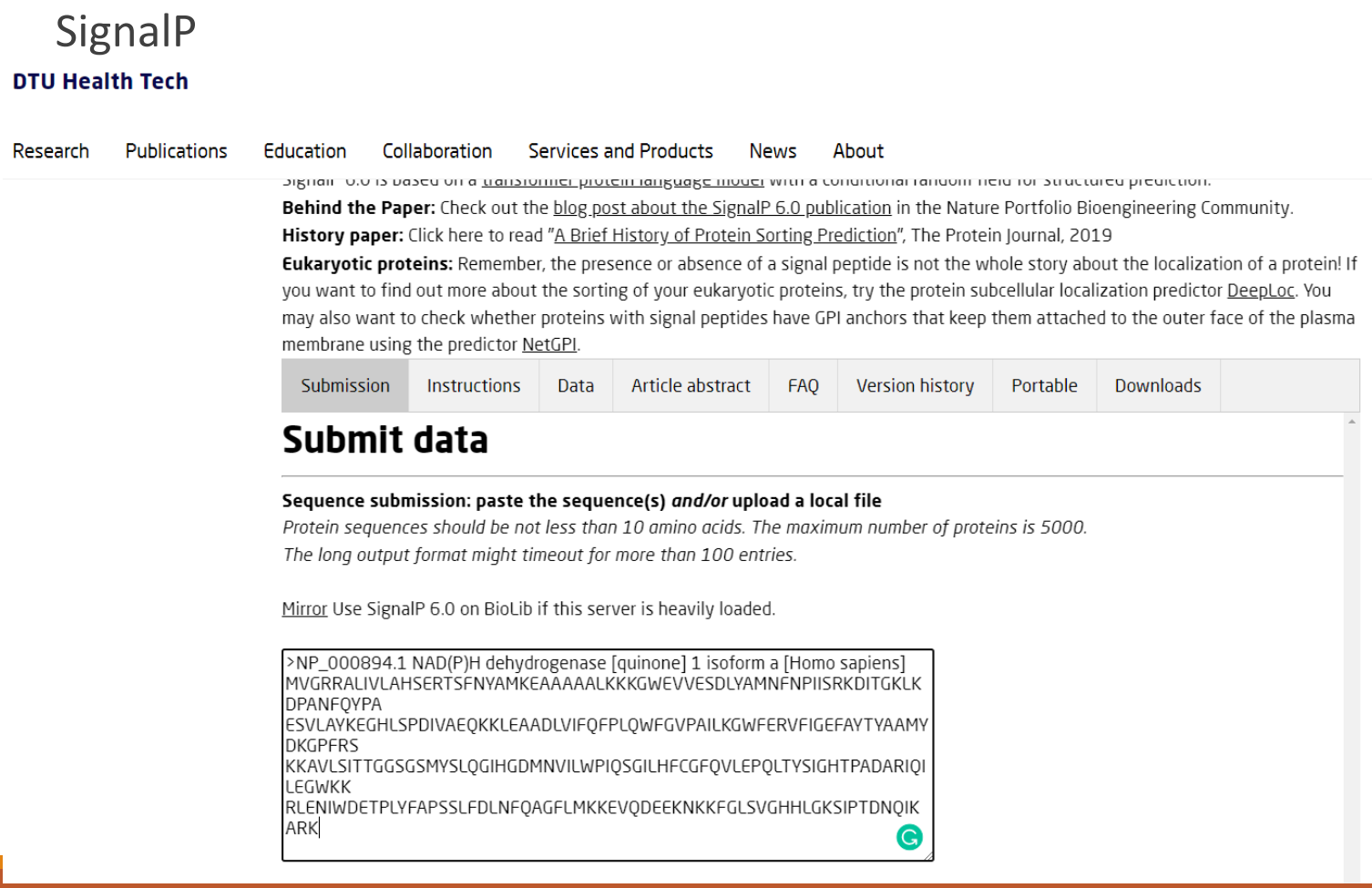
Hledání signálních peptidů

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



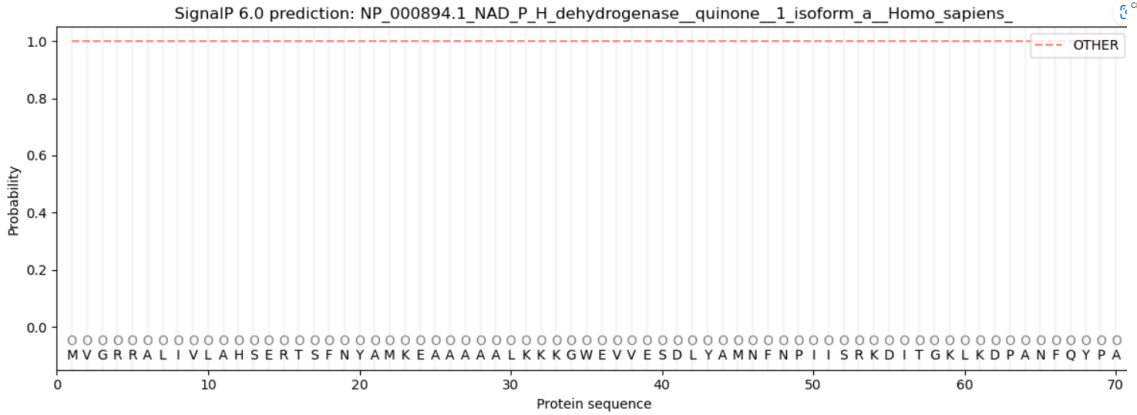
https://www.khanacademy.org/science/biology/gene-expression-central-dogma#central-dogma-transcription

Hledání signálních peptidů

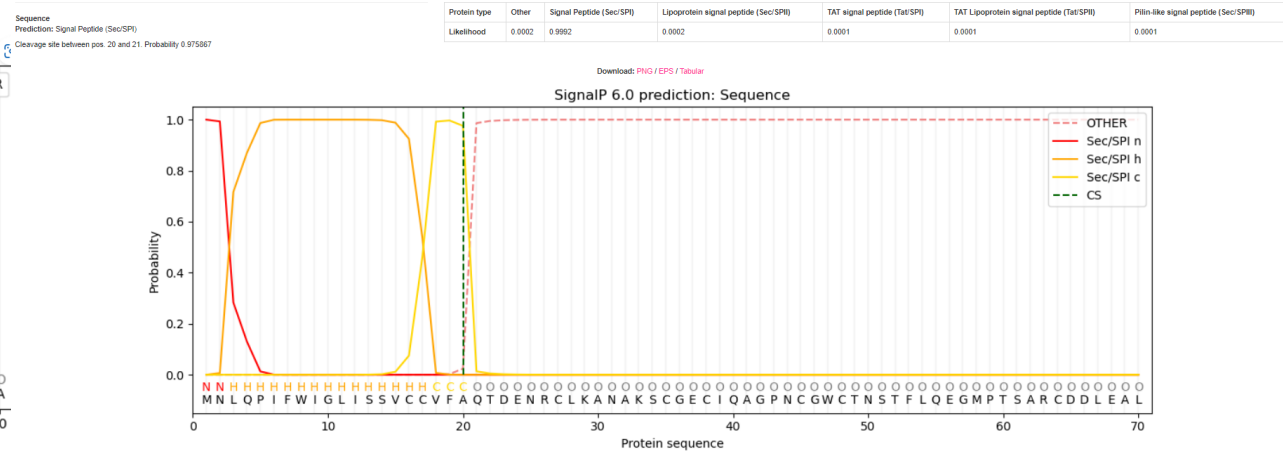


Hledání signálních peptidů

SignalP

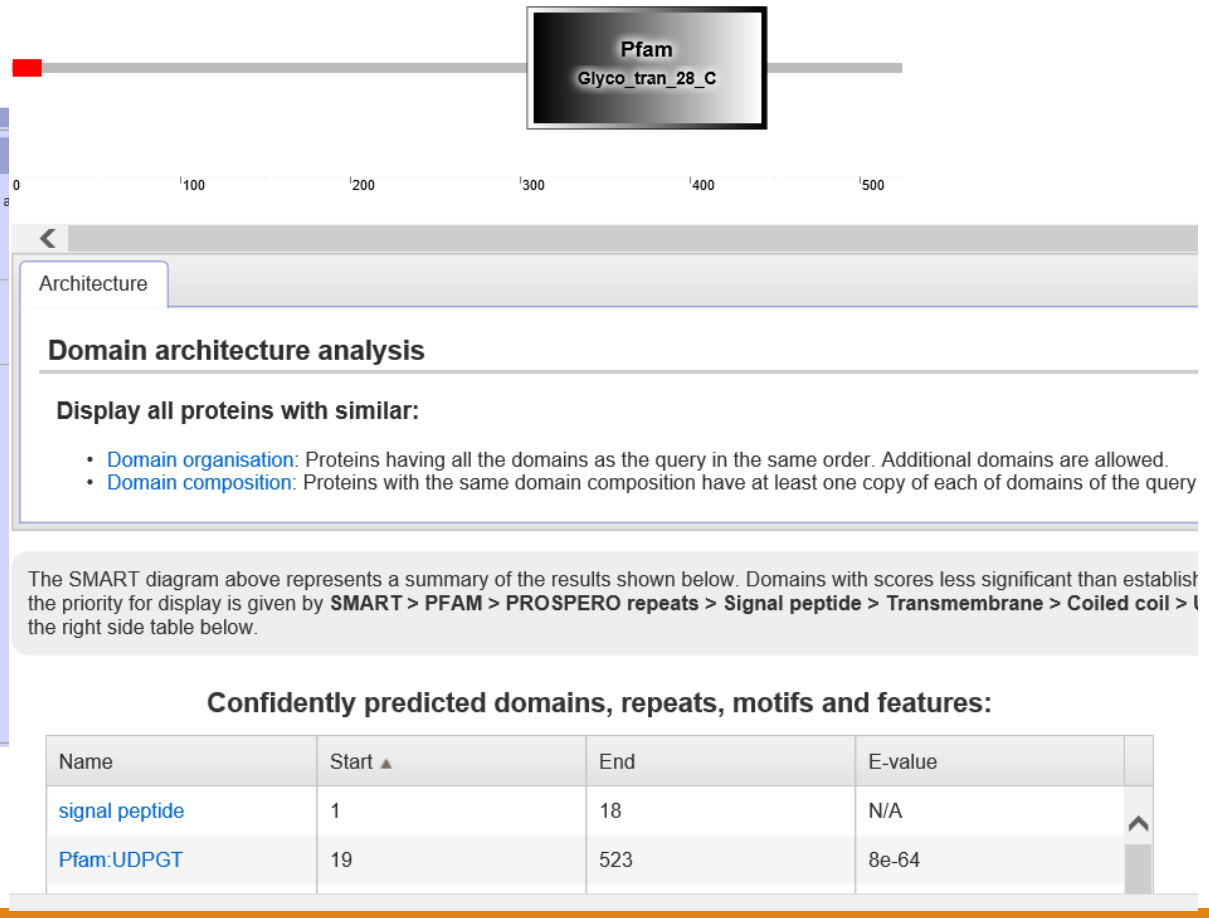


protein nemá signální peptid



protein MÁ signální peptid
(s určitou pravděpodobností)

Hledání signálních peptidů



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

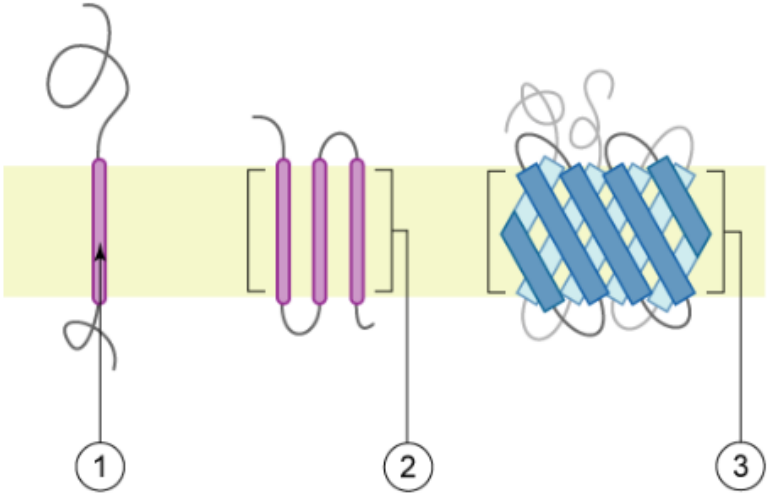
…

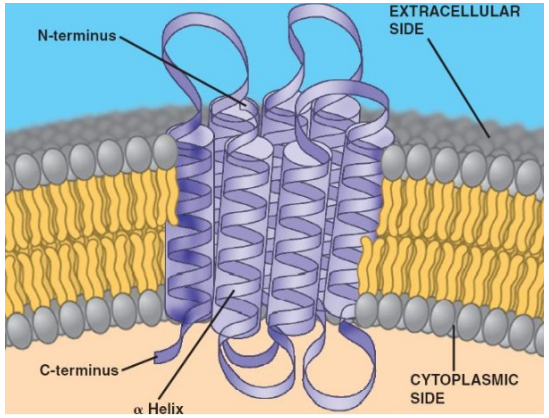
Predikce transmembránových úseků

Hydrofobicita aminokyselin

- různé programy – různé algoritmy výpočtů - různé výstupy

➢Předpověď topologického uspořádání (odhad, které části budou v cytoplazmě a které venku)

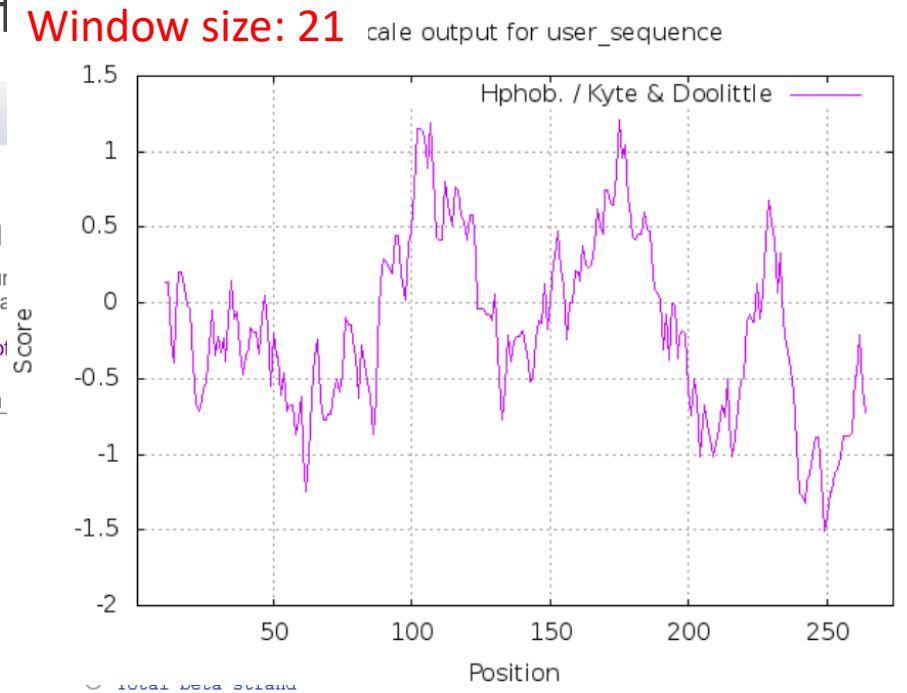




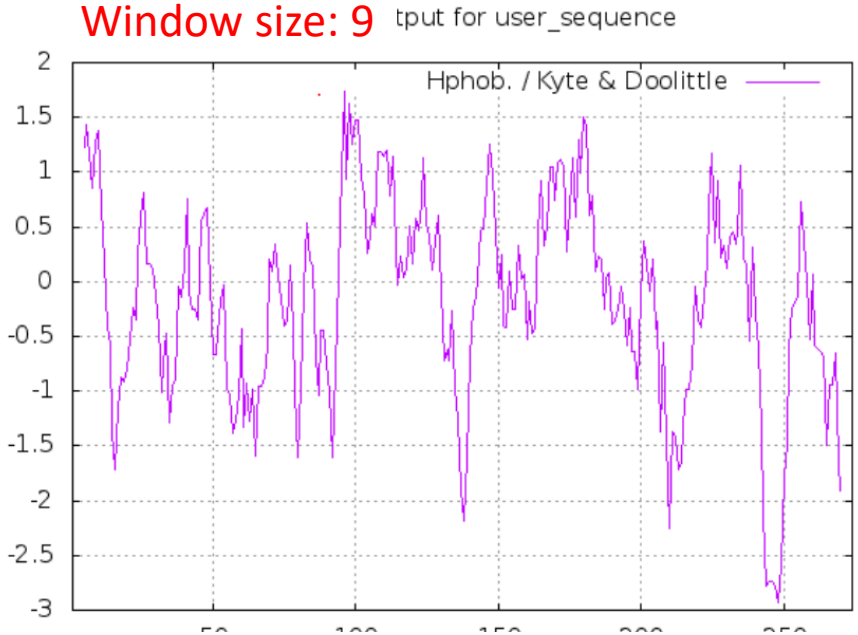
Predikce transmembránových úseků

Profil - Hydrof

Window size: 21



Window size: 9



ExPASy Bioinformatics Resource Portal

ProtScale

ProtScale [Reference / Documentation]

An **amino acid scale** is defined by a nur
but many other scales exist which are ba

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
- Parallel beta-strand
- A.A. comp. in Swiss-Prot

Window size: 9

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

Window size: 19 (21)

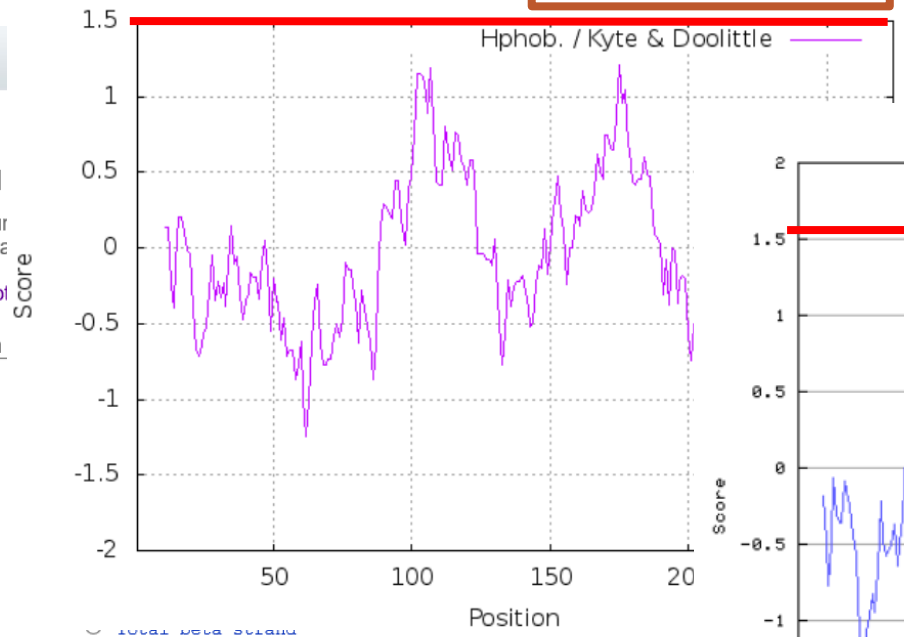
Predikce transmembránových úseků

Profil - Hydrof

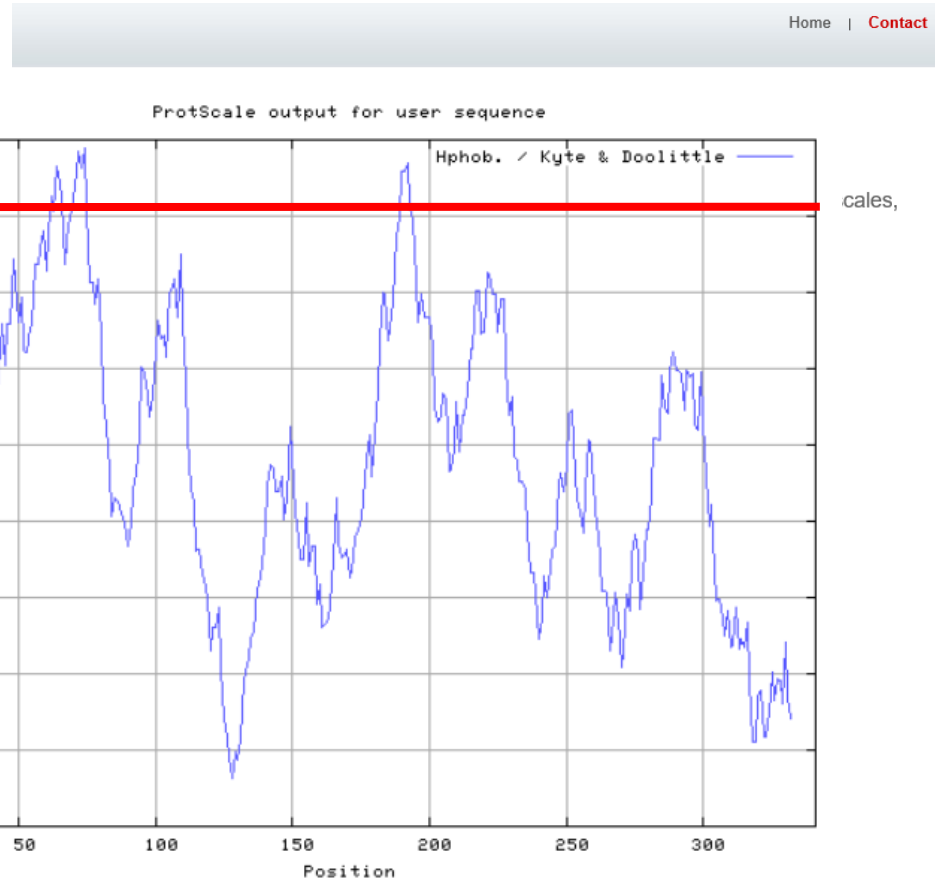
Window size: 21

cale output fo

Cut off: 1.5 (1.6)



Window size: 19 (21)



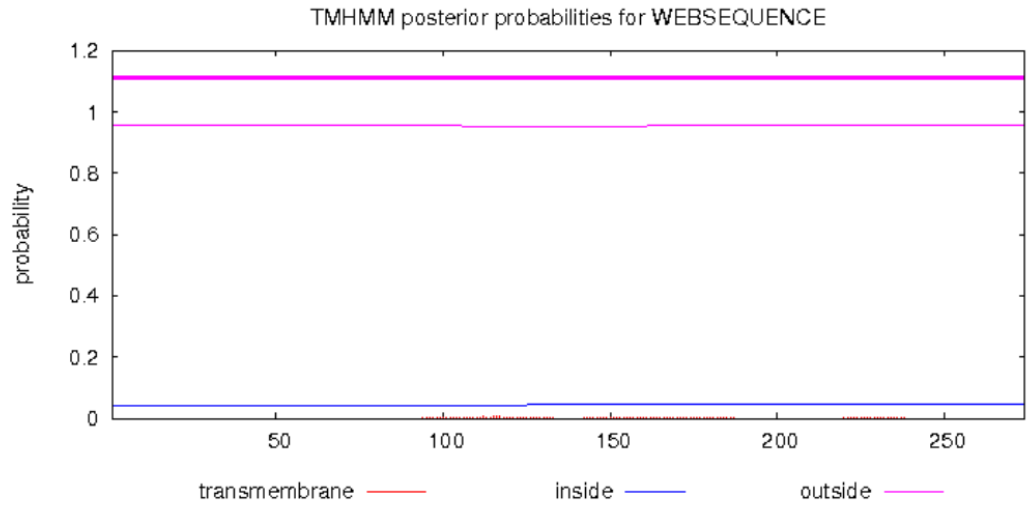
TMHMM

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
```



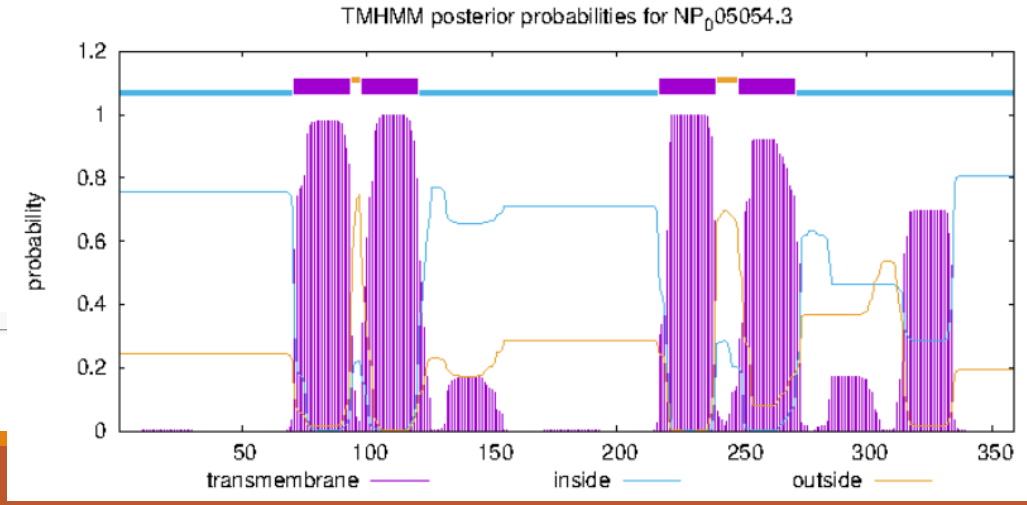
Desaturasa lidská SCD

TMHMM result

HELP with output formats

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

A combined transme

Normal prediction Constr predi

Normal prediction

Paste your protein sequence here in Fasta format:

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

Select output format:

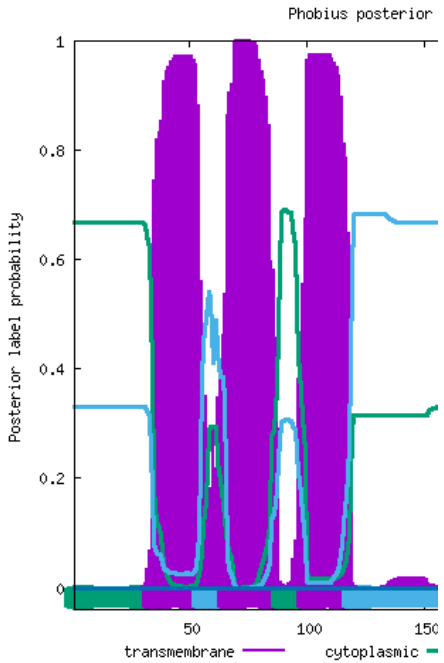
- ○ Short
- ○ Long without Graphics
- ◉ Long with Graphics

Odeslat Resetovat

Phobius prediction

Prediction of UNNAMED

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

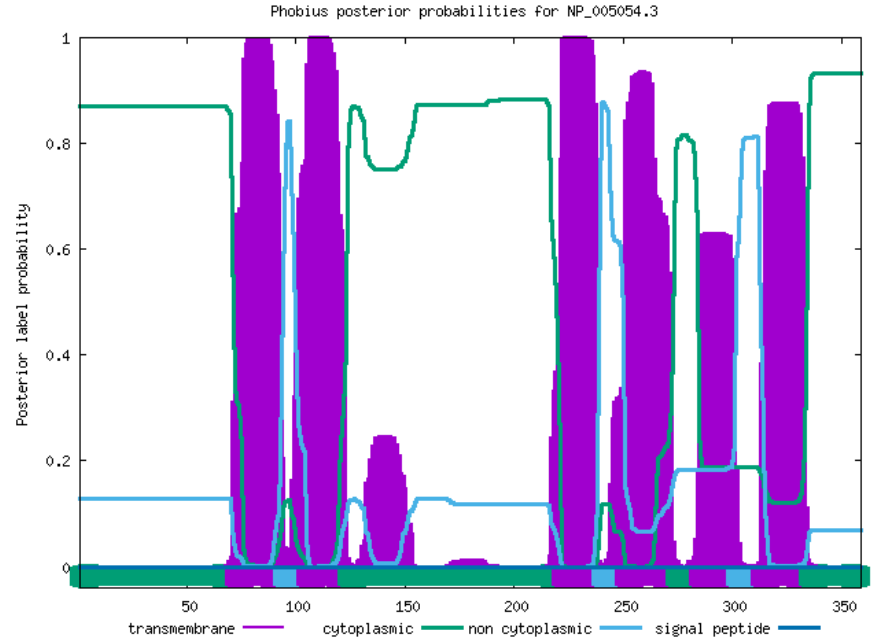


The probability data used in the plot is found here, and the

Phobius prediction

Prediction of NP_005054.3

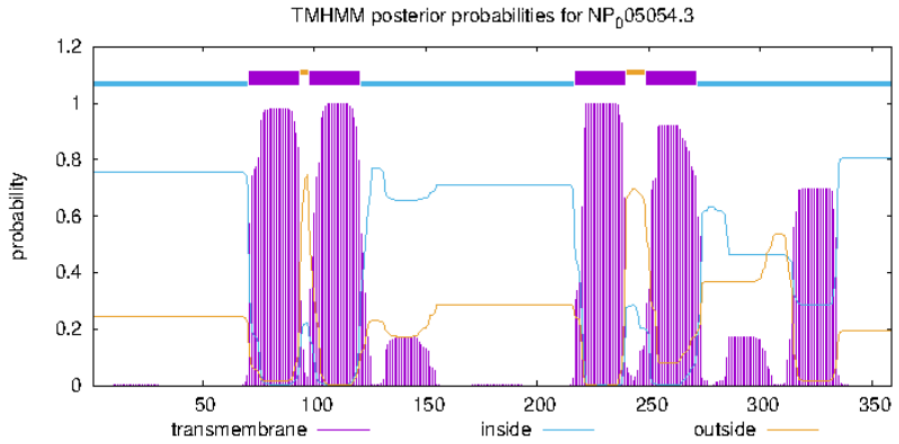
```
ID   NP_005054.3
FT   TOPO_DOM      1     71       CYTOPLASMIC.
FT   TRANSMEM     72     93
FT   TOPO_DOM     94    104       NON CYTOPLASMIC.
FT   TRANSMEM    105    123
FT   TOPO_DOM    124    221       CYTOPLASMIC.
FT   TRANSMEM    222    239
FT   TOPO_DOM    240    250       NON CYTOPLASMIC.
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.
//
```



Predikce transmembránových úseků (SCD)

TMHMM result

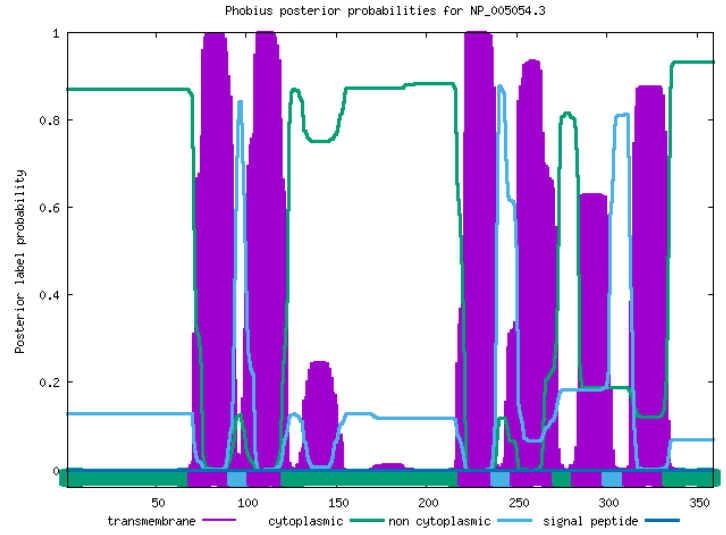
```
# NP_005054.3 Length: 359
# NP_005054.3 Number of predicted TMHs:  4
# NP_005054.3 Exp number of AAs in TMHs: 104.54198
# NP_005054.3 Exp number, first 60 AAs:  0.03484
# NP_005054.3 Total prob of N-in:        0.75464
NP_005054.3     TMHMM2.0        inside      1    70
NP_005054.3     TMHMM2.0        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
```



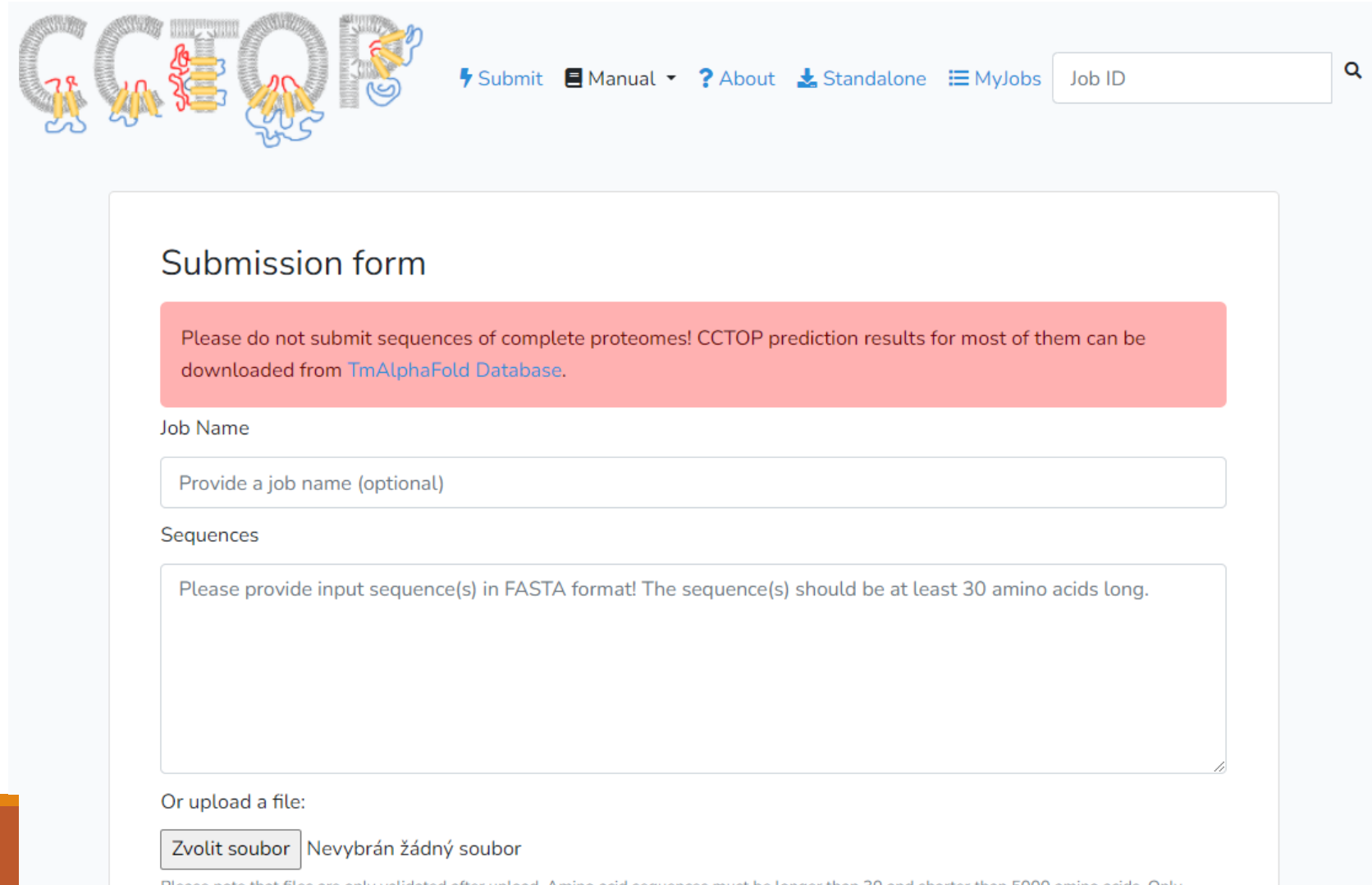
Phobius prediction

Prediction of NP_005054.3

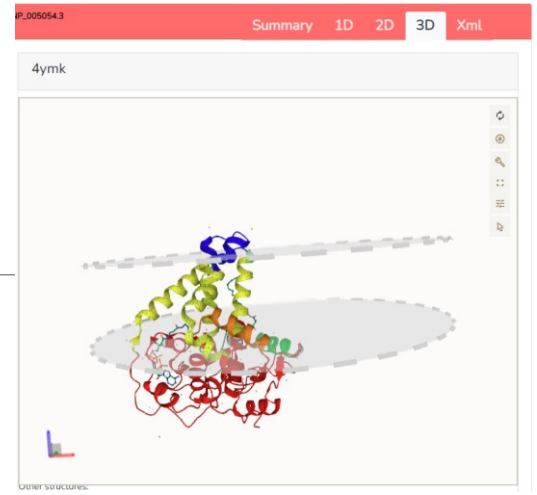
```
ID   NP_005054.3
FT   TOPO_DOM      1     71       CYTOPLASMIC.
FT   TRANSMEM     72     93
FT   TOPO_DOM     94    104       NON CYTOPLASMIC.
FT   TRANSMEM    105    123
FT   TOPO_DOM    124    221       CYTOPLASMIC.
FT   TRANSMEM    222    239
FT   TOPO_DOM    240    250       NON CYTOPLASMIC.
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.
//
```

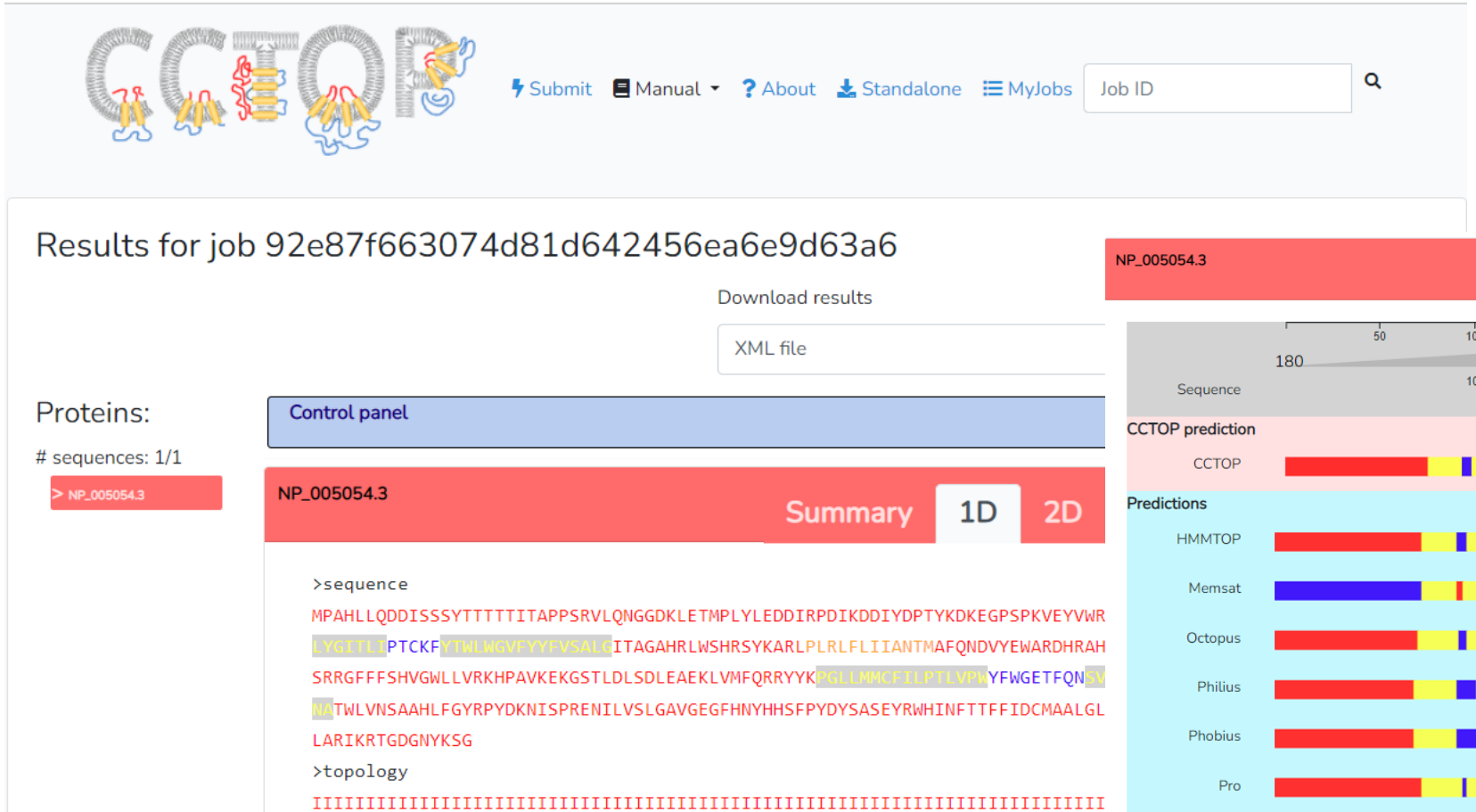


CCTOP


Submit
Manual
About
Standalone
MyJobs
Job ID
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
Provide a job name (optional)
Sequences
Please provide input sequence(s) in FASTA format! The sequence(s) should be at least 30 amino acids long.
Or upload a file:
Zvolit soubor
Nevybrán žádný soubor

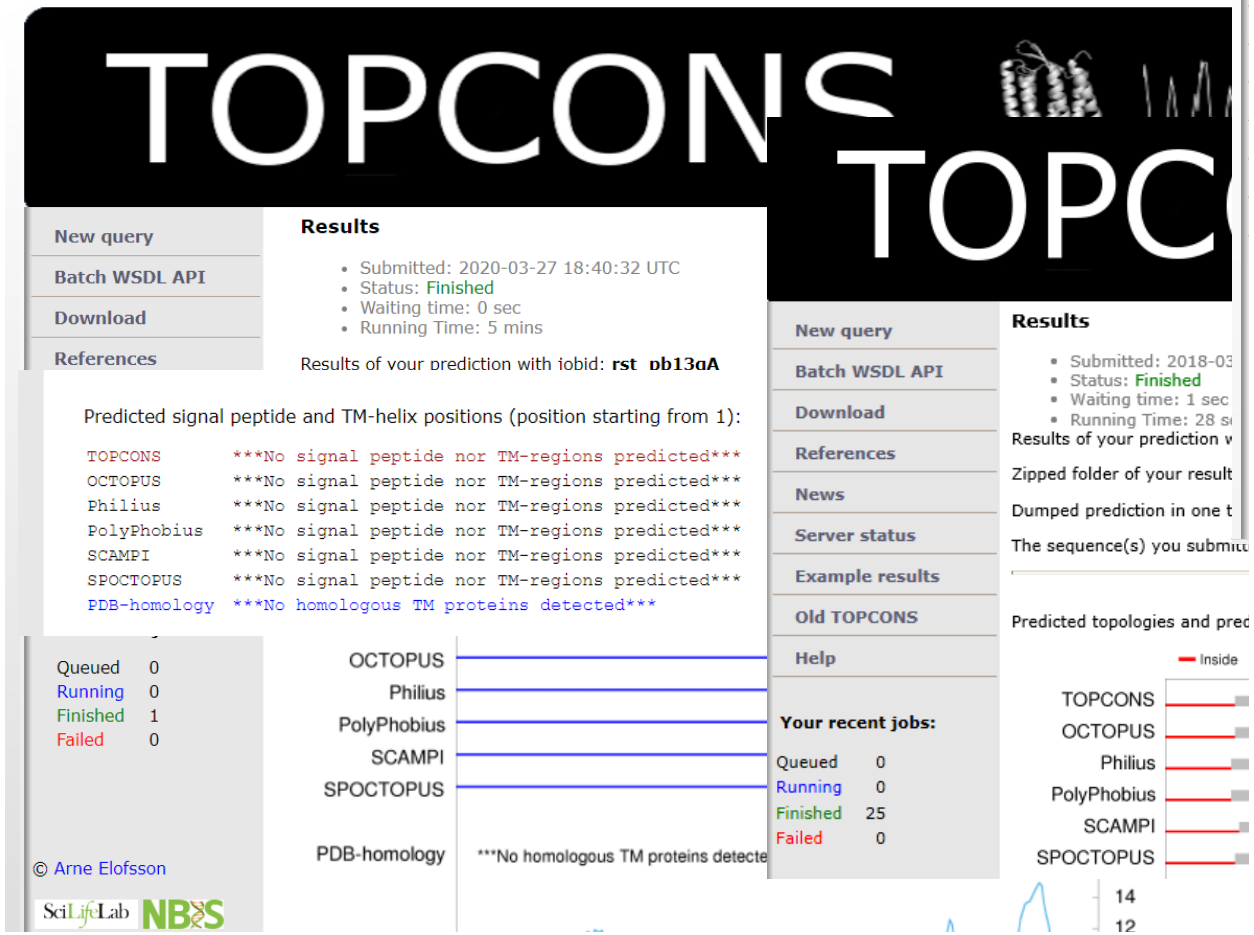
CCTOP

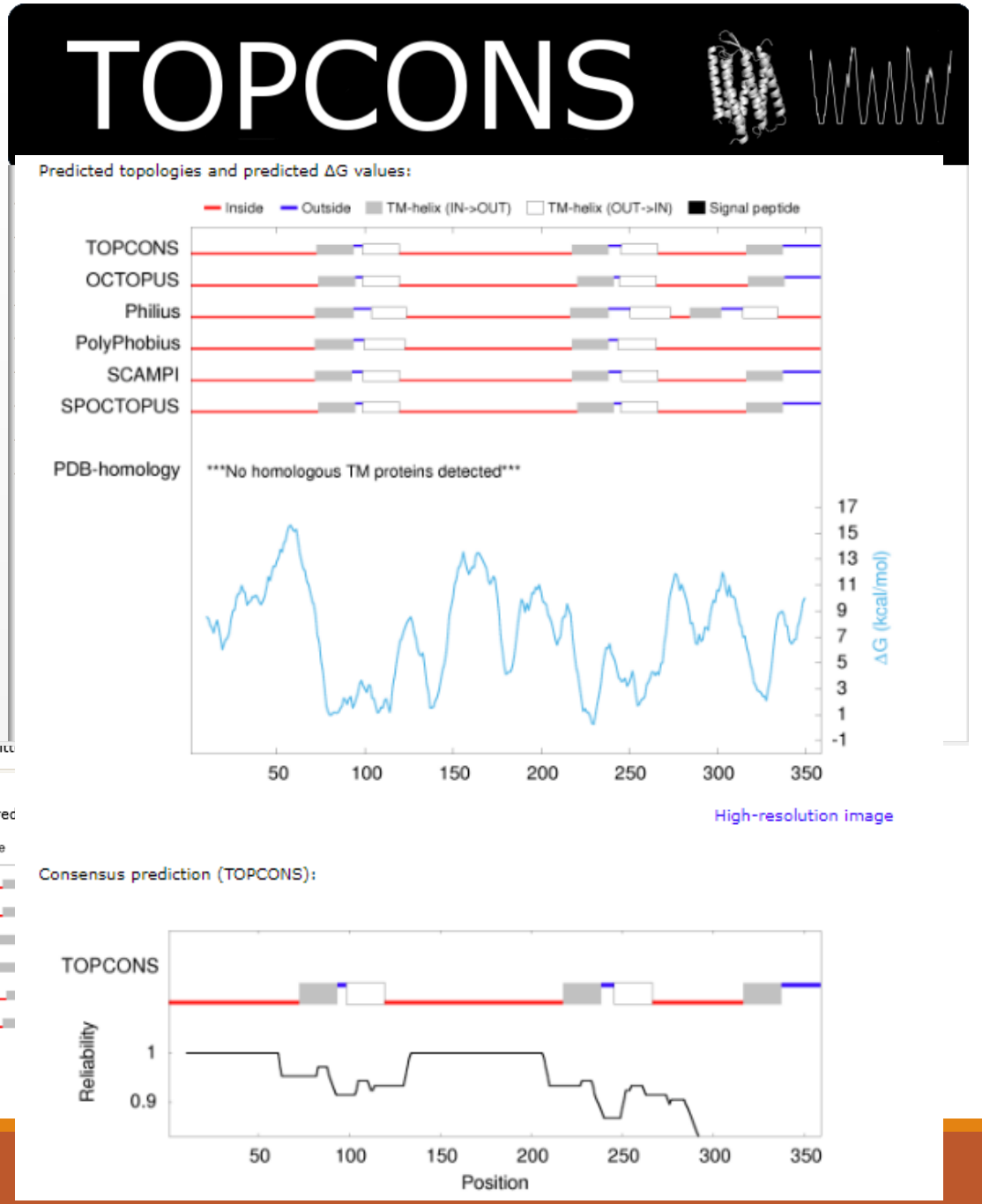

Summary
1D
2D
3D
Xml
4ymk


Submit
Manual
About
Standalone
MyJobs
Job ID
Results for job 92e87f663074d81d642456ea6e9d63a6
Download results
XML file
Proteins:
sequences: 1/1
> NP_005054.3
Control panel
NP_005054.3
Summary
1D
2D
>sequence
MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKEGPSPKVEYVWR
LYGITLIPTCKFVTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAH
SRRGFFFSHVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKPGLLMMCFILPTLVPWYFWGETFQNSV
NATWLVNSAAHLFGYRPYDKNISPRENILVSLGAVGEGFHNYHHSFPYDYSASEYRWHINFTTFFIDCMAALGL
LARIKRTGDGNYKSG
>topology
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
MMMMMMMOOOOOMMMMMMMMMMMMMMMIIIIIIIIIIIIIIIIIILLLLLLLLLLLIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
MMMMMMMMMMMMMMMMOOOOOOOOOMMMMMMMMMMMMMM
MMIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIII


NP_005054.3
Summary
1D
2D
3D
Xml
50
100
150
200
250
300
350
180
180
Sequence
100
200
300
CCTOP prediction
CCTOP
Predictions
HMMTOP
Memsat
Octopus
Philius
Phobius
Pro
Prodiv

TOPCONS

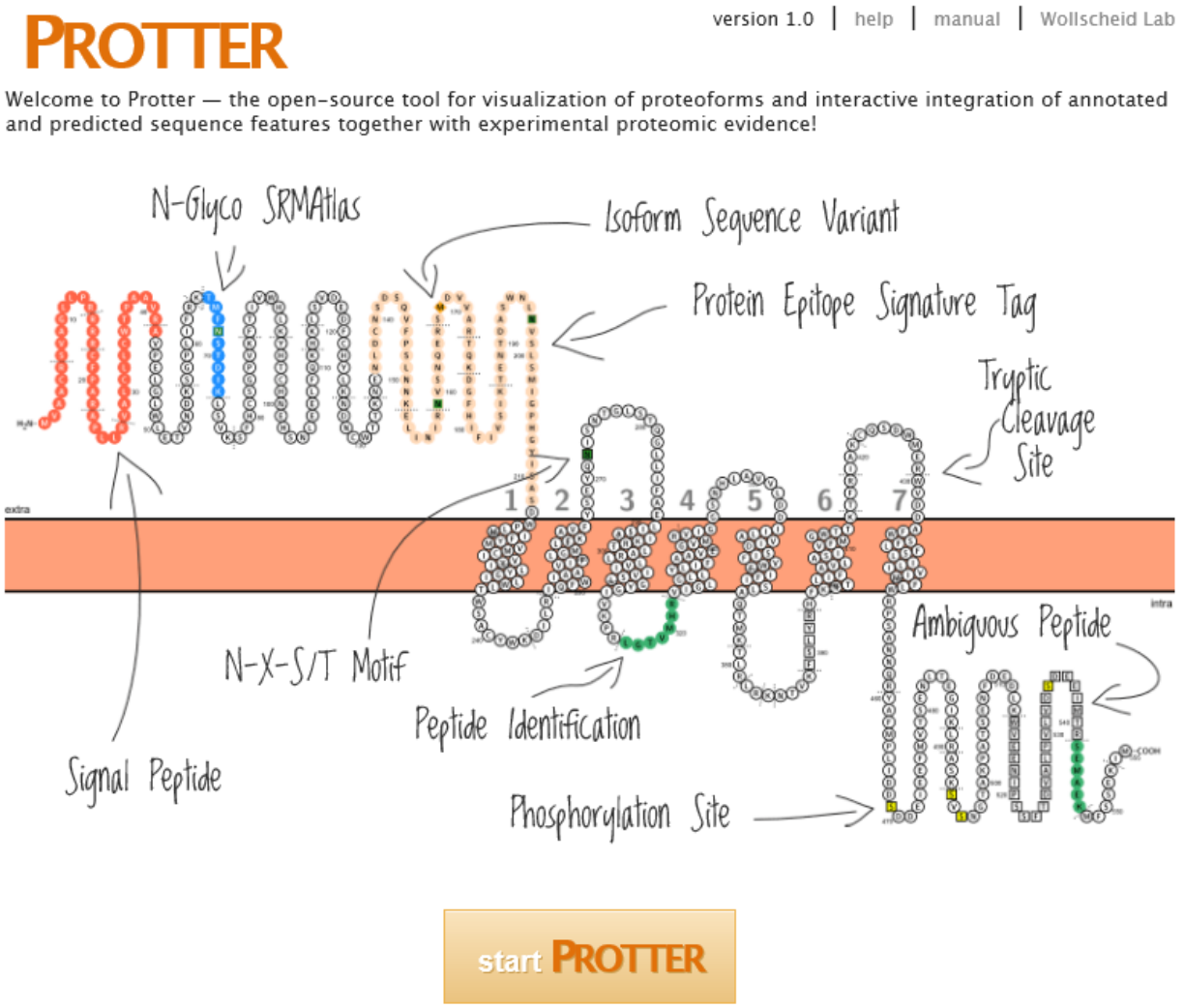

TOPCONS
New query
Batch WSDL API
Download
References
Results
Submitted: 2020-03-27 18:40:32 UTC
Status: Finished
Waiting time: 0 sec
Running Time: 5 mins
Predicted signal peptide and TM-helix positions (position starting from 1):
TOPCONS ***No signal peptide nor TM-regions predicted***
OCTOPUS ***No 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
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
PDB-homology
© Arne Elofsson
SciLifeLab NBIS
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
Status: Finished
Waiting time: 1 sec
Running Time: 28 s
Results of your prediction
Zipped folder of your result
Dumped prediction in one t
The sequence(s) you submit
Predicted topologies and pred
Inside
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
14
12


TOPCONS
Predicted topologies and predicted ΔG values:
Inside
Outside
TM-helix (IN->OUT)
TM-helix (OUT->IN)
Signal peptide
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
PDB-homology
No homologous TM proteins detected
17
15
13
11
9
7
5
3
1
-1
ΔG (kcal/mol)
50
100
150
200
250
300
350
High-resolution image
Consensus prediction (TOPCONS):
TOPCONS
Reliability
1
0.9
50
100
150
200
250
300
350
Position

PROTTER-obrázek!

-pracuje s daty z UniProt

(nepočítá sám)



PROTTER-obrázek!

-pracuje s daty z UniProt

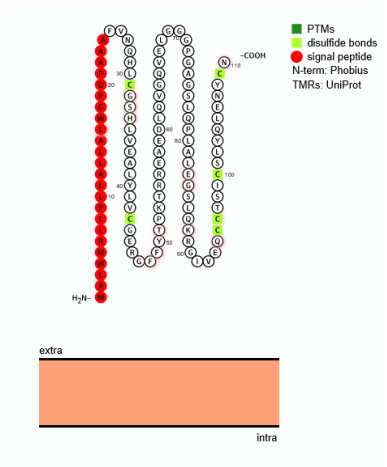
(nepočítá sám)

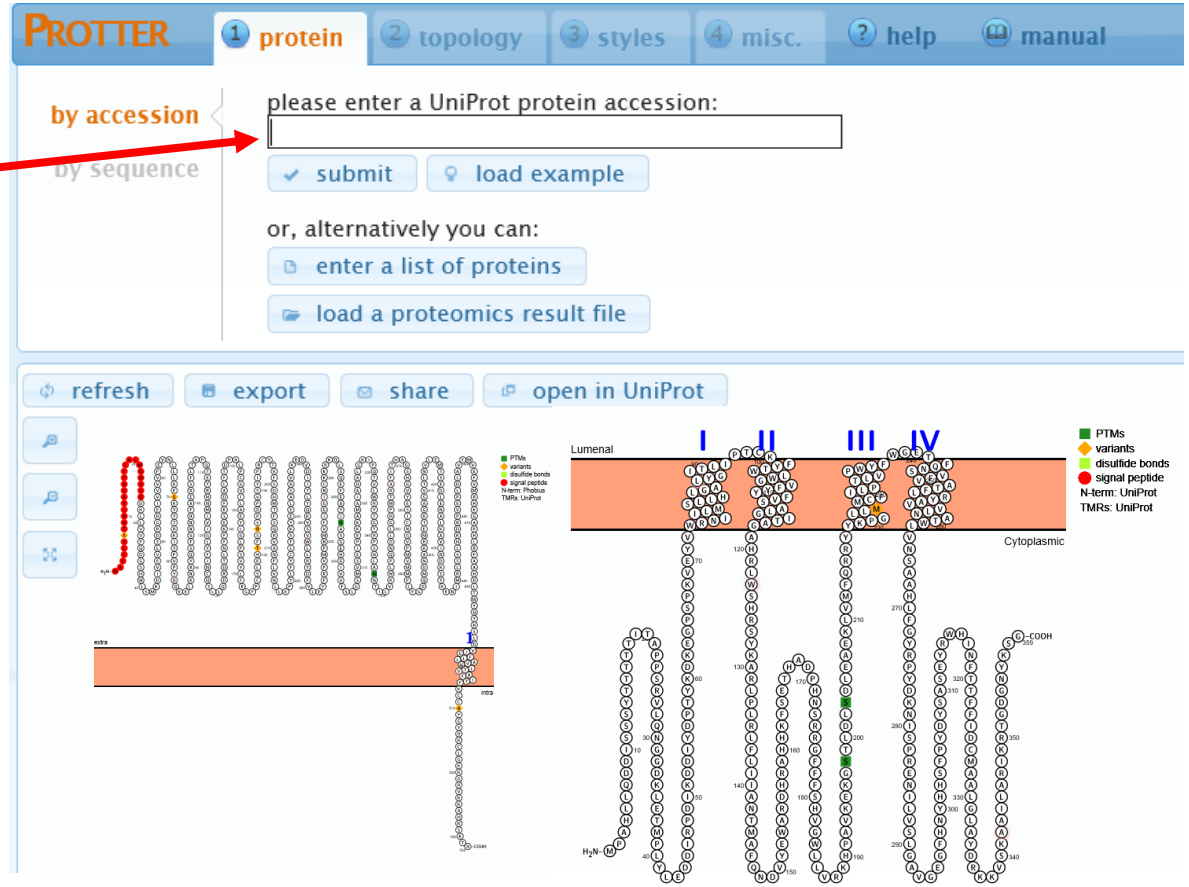
-zadává se přístupový kód:

UGT1A6 (P19224)

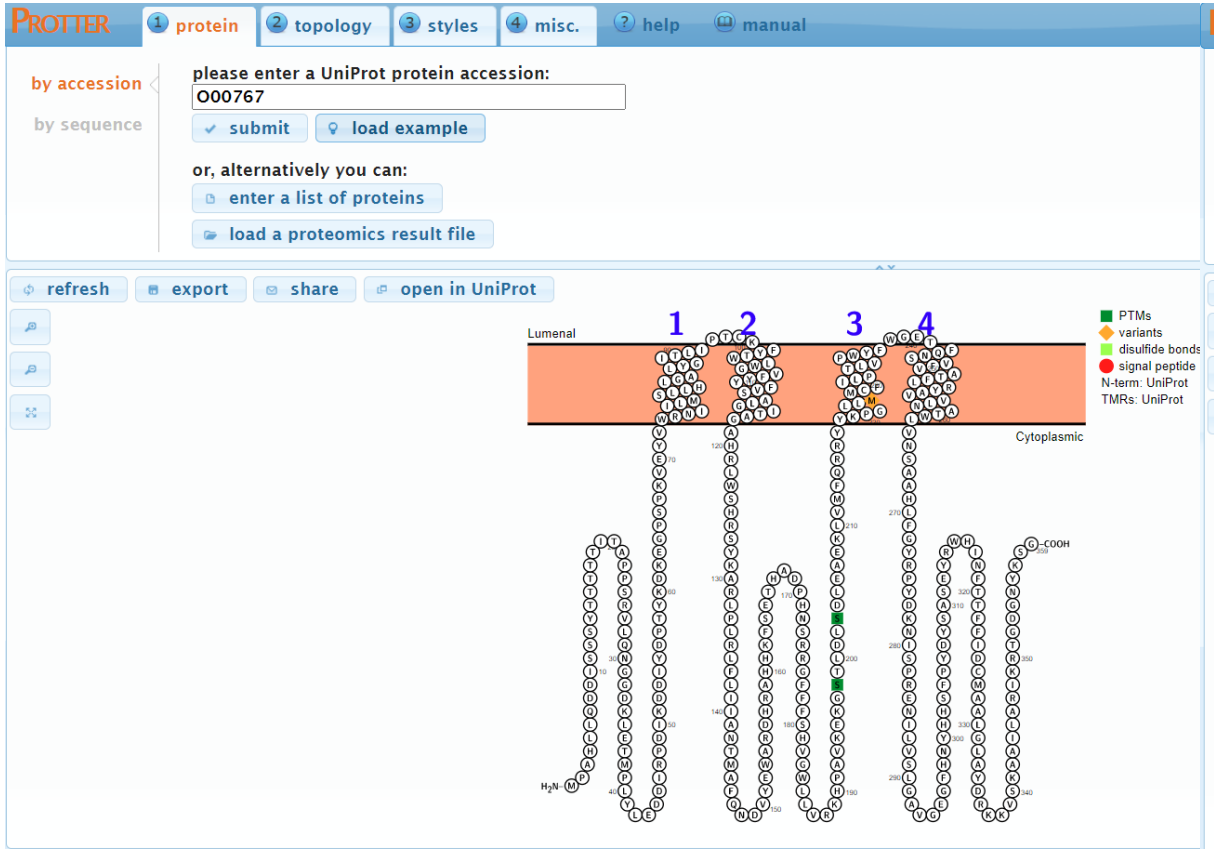
Desaturasa (O00767)

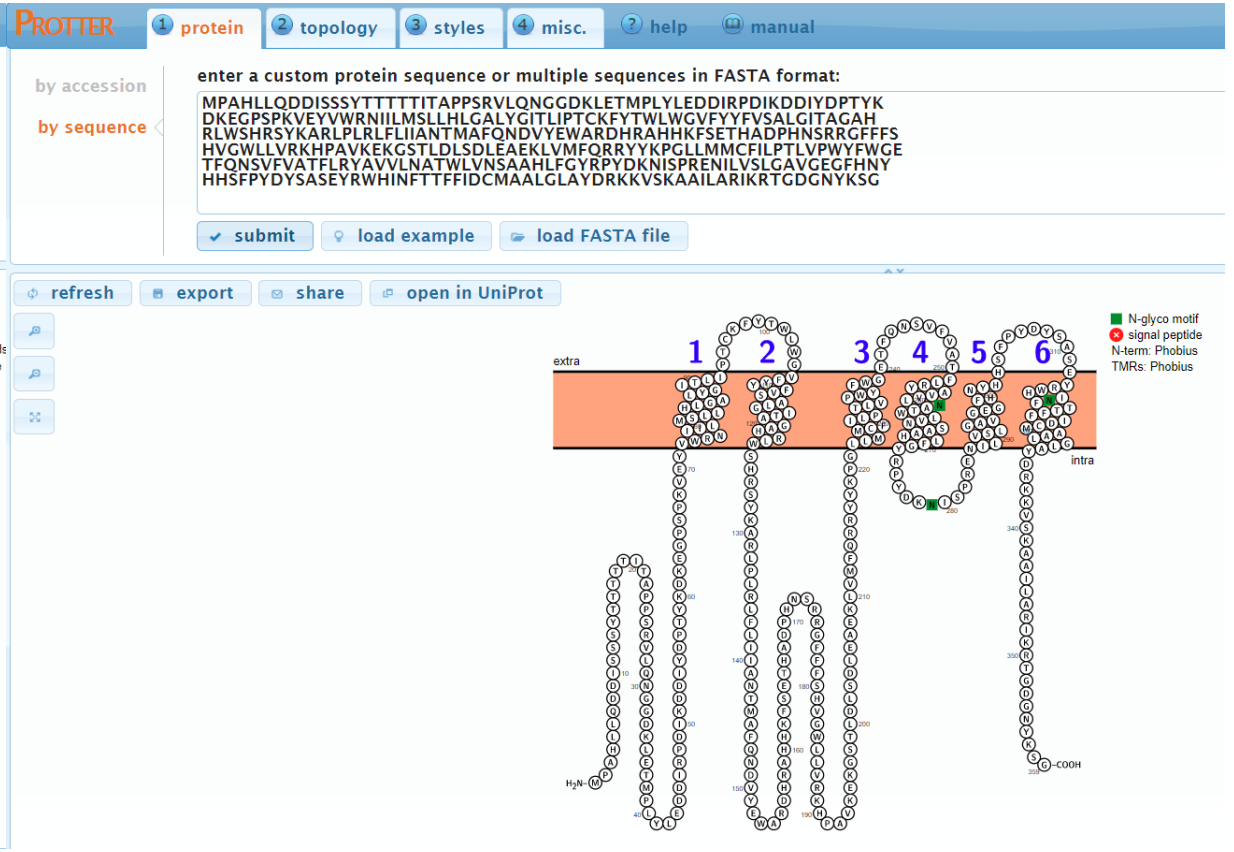
(prepro) insulin (P01308)





PROTTER-obrázek!





Predikce transmembránových úseků

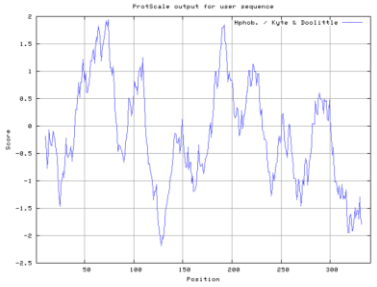
ProtScale

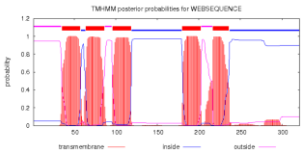
TMHMM

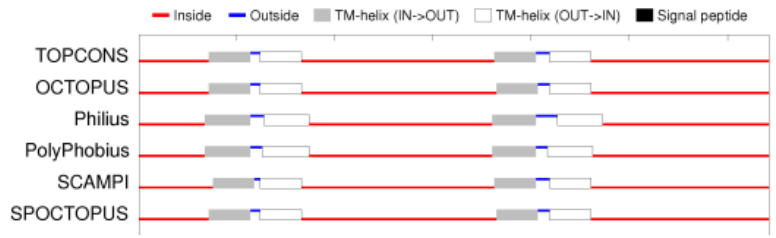
Phobius…

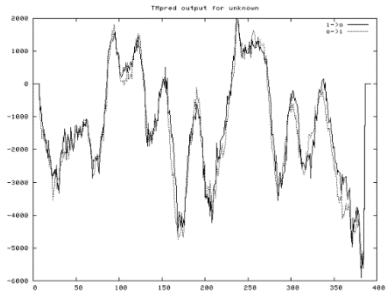
TOPCONS

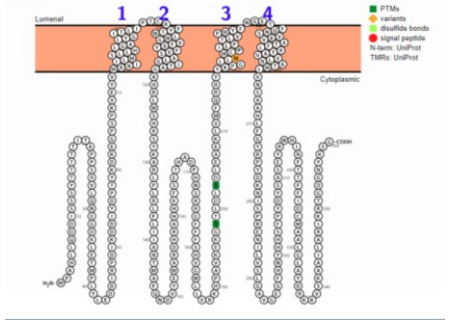
CCTOP











- **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

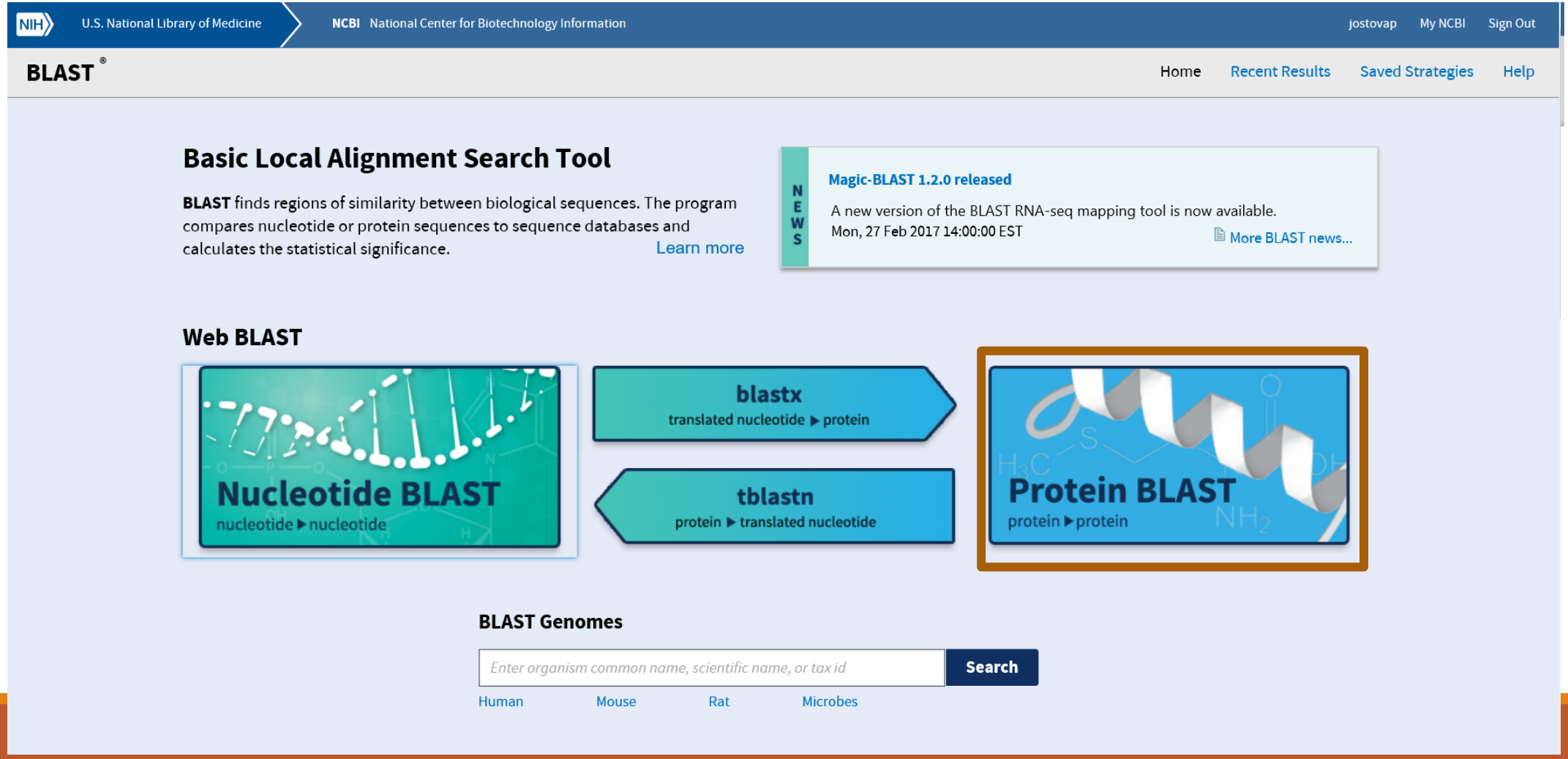
Hledá podobné sekvence z vybraných databáze pomocí krátkých „slov“ (sekvencí vytvořených ze zadané sekvence), s využitím „substituční matice“, která udává míru důležitosti rozdílných aminokyselin v zadané sekvenci a vyhledané sekvenci

„Leucin je podobnější Isoleucinu než Histidinu“

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

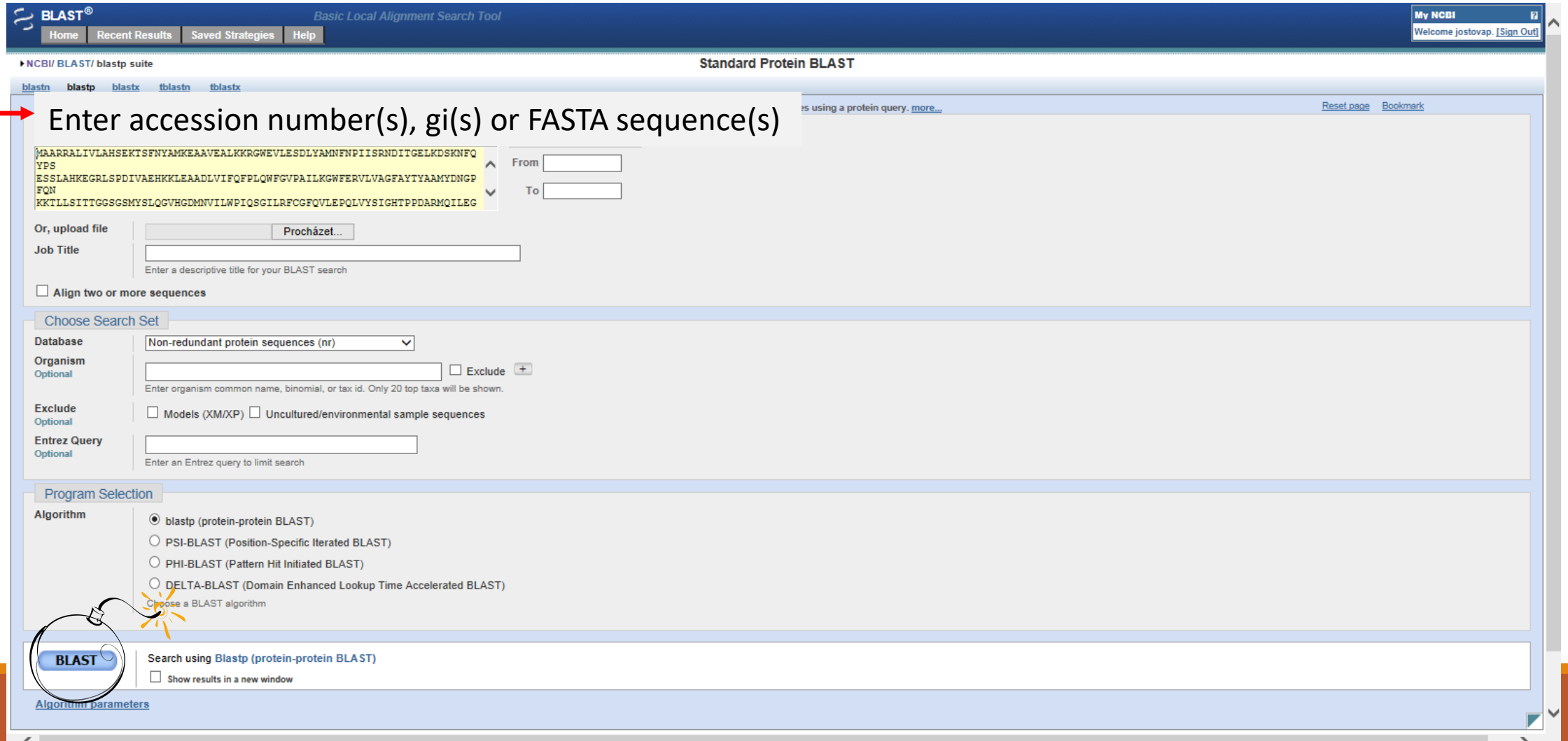
NCBI/BLAST

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

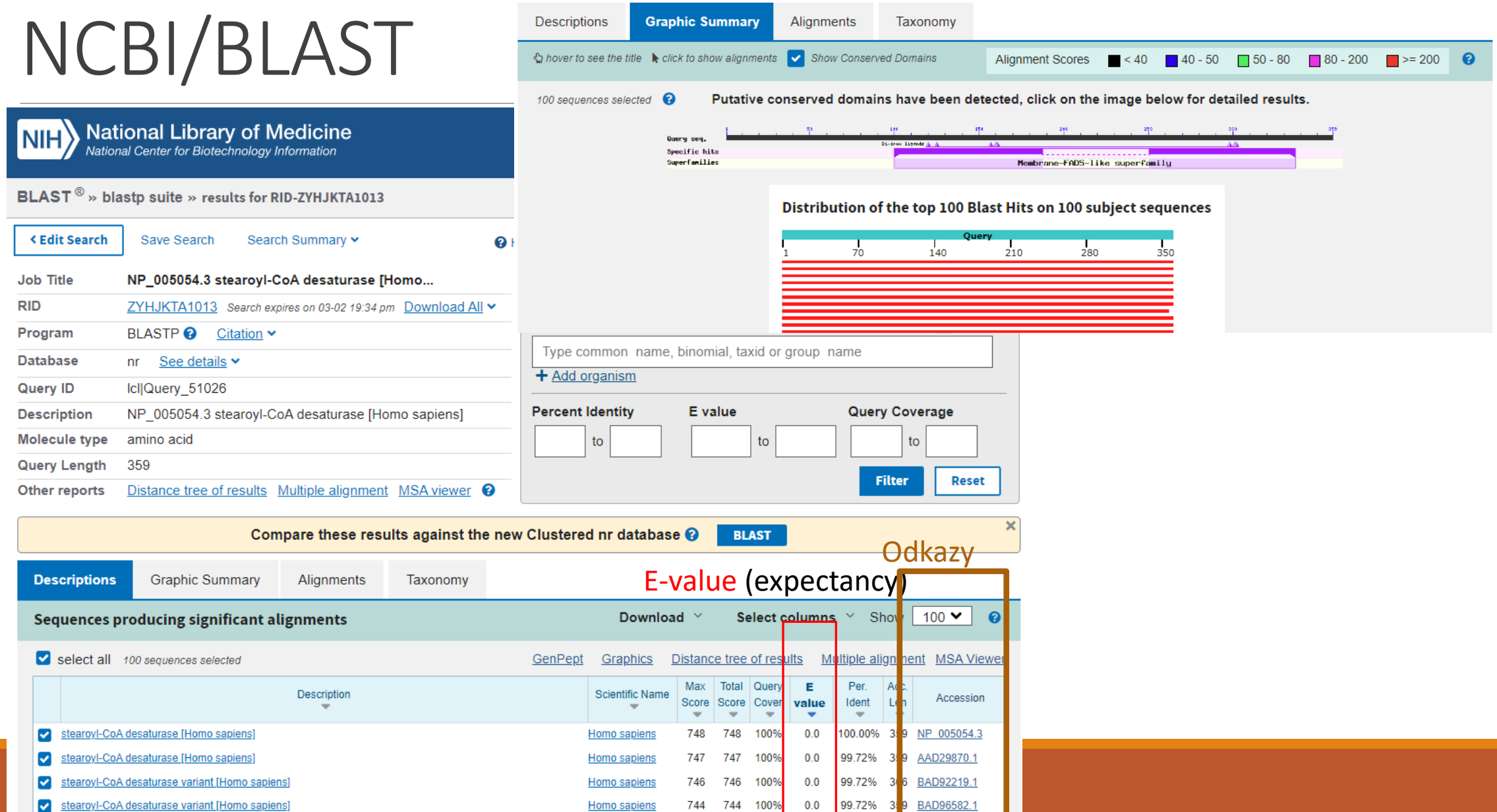


http://www.youtube.com/watchBLAST

NCBI/BLAST



NCBI/BLAST



Odkazy

E-value (expectancy)

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    MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYK  60
            MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYK
Sbjct  1    MPAHLLQDDISSSYTTTTTITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYK  60

Query  61   DKEGPSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH  120
            DKEGPSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH
Sbjct  61   DKEGPSPKVEYVWRNIILMSLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAH  120

Query  121  RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHHKFSETHADPHNSRRGFFFS  180
            RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHHKFSETHADPHNSRRGFFFS
Sbjct  121  RLWSHRSYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHHKFSETHADPHNSRRGFFFS  180

Query  181  HVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKPGLLMMCFILPTLVPWYFWGE  240
            HVGWLLVRKHPAVKEKGSTLDLSDLEAEKLVMFQRRYYKPGLLMMCFILPTLVPWYFWGE
```

Related Information

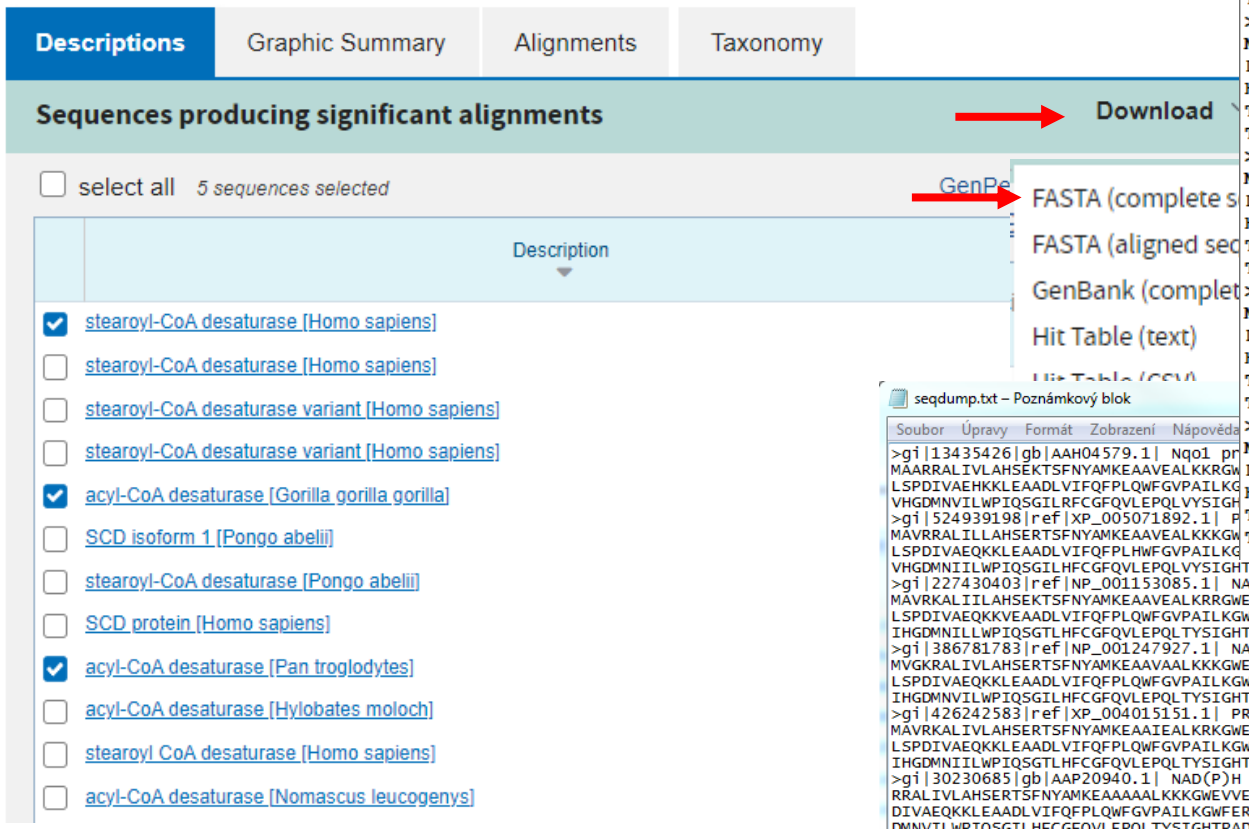
Gene - associated gene details

Genome Data Viewer - aligned genomic context

Identical Proteins - Identical proteins to NP_005054.3

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

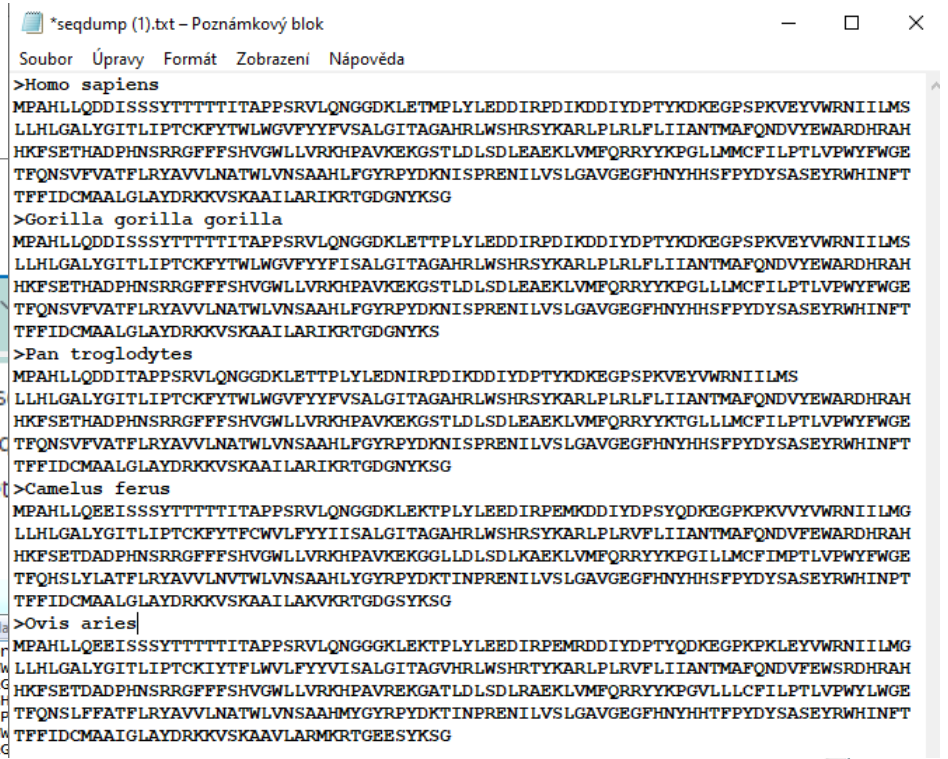
NCBI/BLAST



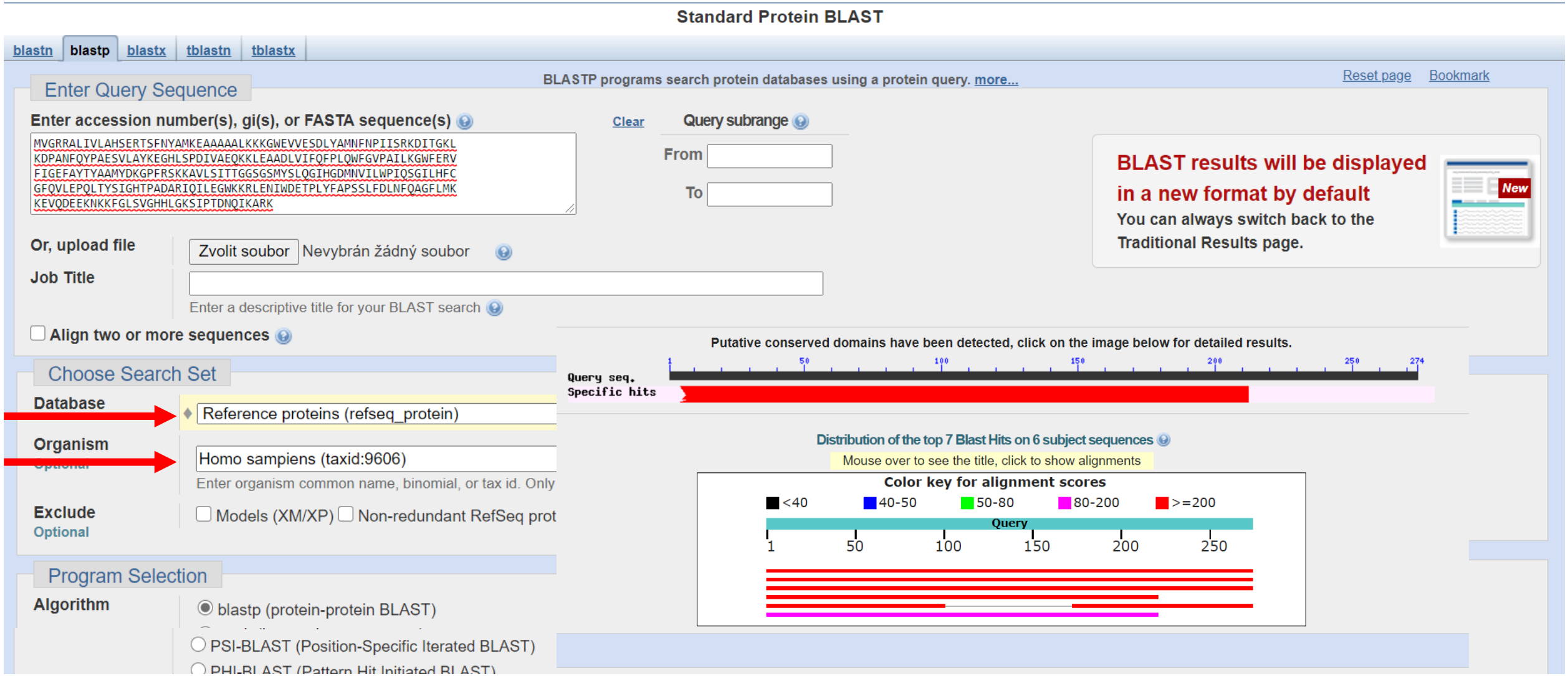
√

√

√

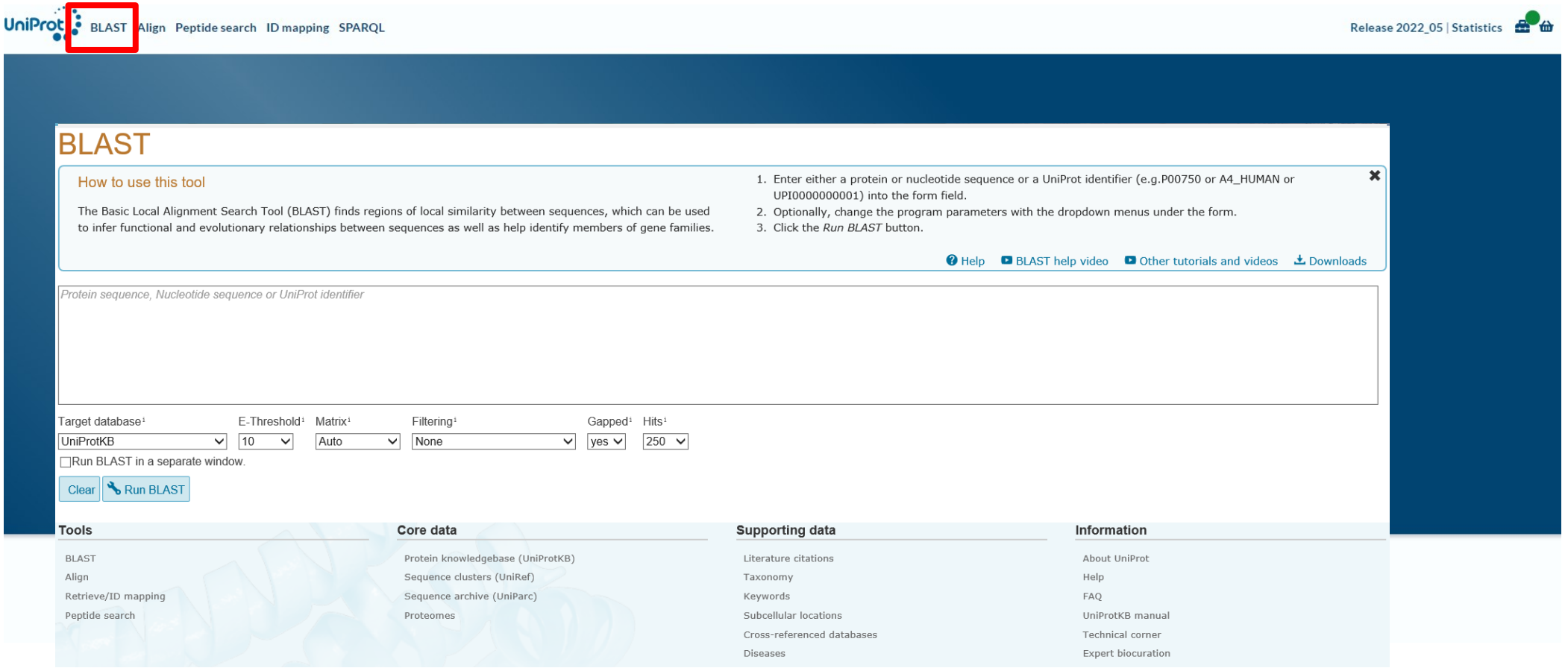


NCBI/BLAST-podobné sekvence u jednoho org.

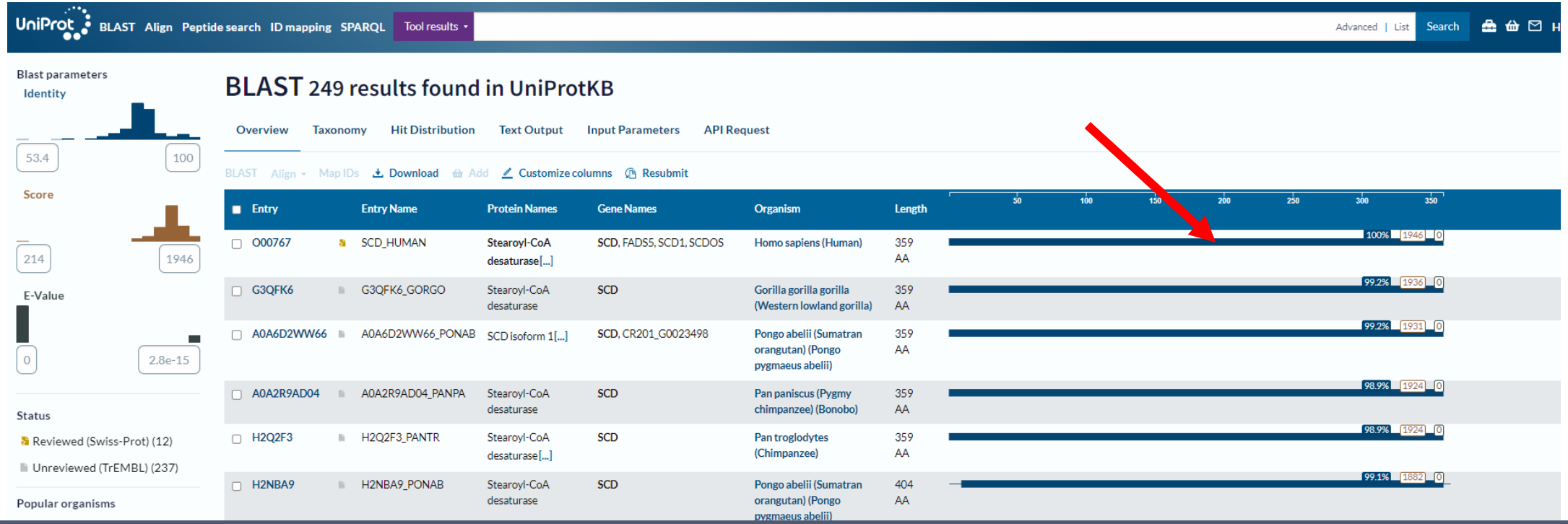


Uniprot/BLAST

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



Uniprot/BLAST

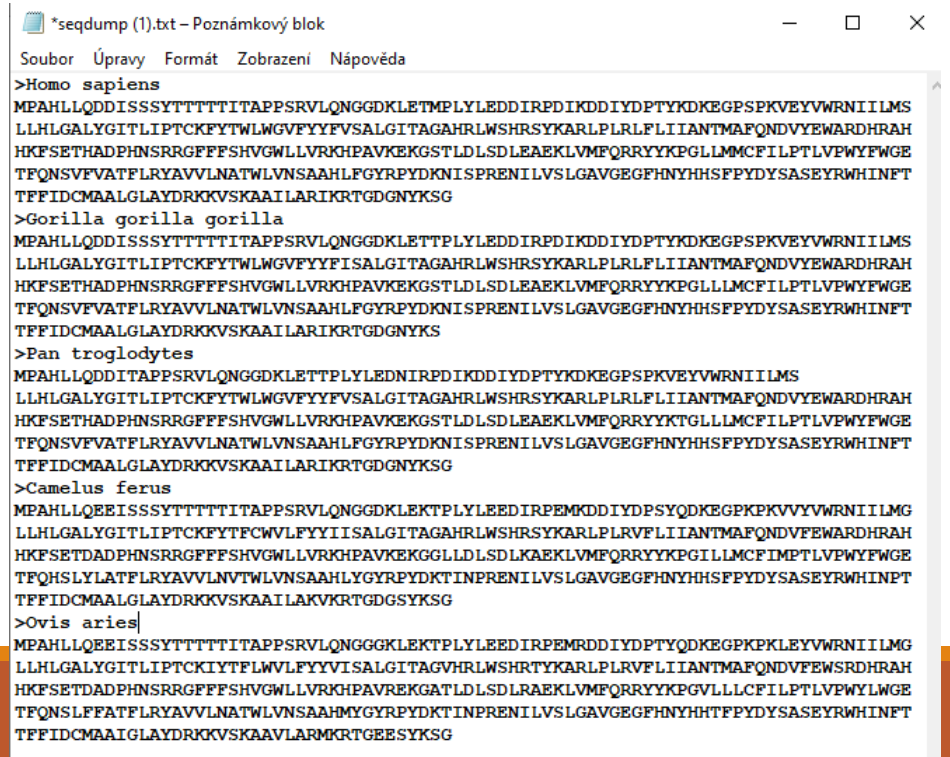


Vyzkoušejte si BLAST

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



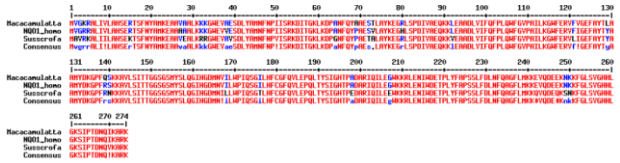
shrnutí I - Rešeršní projekt: NQO1

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)…
- **Flavodoxinová doména** 

, N-terminální signální peptid
- **Cytosolický/Extracelulární (Rozpustný) protein**
- **Podobné proteiny vyhledány, FASTA soubory uloženy, sekvence porovnány**



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? (-II-)
3) Má signální peptid? (-II-)
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

DÚ1 - OneNote

Soubor Domů Vložení Kreslit Historie Revize Zobrazení

Můj poznámkový blok — Rychlé poznámky | paper challenge | Nový oddíl 1 | bioinformatika

DÚ3

1)

2)

TMMOD

# Annotation	NON-TM PROTEIN
# Length	274
# Number of predicted TMHs	0
# Exp number of AAs in TMHs	0.000000
# Exp number, first 60 AAs	0.000000
# Total prob of N-in	0.493689
outside	1 274

show posterior probabilities

3)

>Macaca mulatta
MVGKRALIVLAHSERTSFNYAMKEAAVAALKKKGWEVAESDLYAMNFNPIISRKDITGKLKDPANFQYAAESTLAYKEGR
LSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFVGEFAYTLAAMYDKGPFQSKGAVLSITTGGSGSMYSLQG
IHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMKK
KEVQDEEKNKKPGLSVGHHLGKSIPTDNQIKARK

>Sus scrofa
MAVRKALIILAHSEKTSFNYAMKEAAVEALKRRGWEVAVSDLYAMNFNPVISRKDITGKLKDPGNFQYPAETALAYKEGR
LSPDIVAEQKKVEAADLVIFQFPLQWFGVPAILKGWFERVLIGEFAYTYAAMYDKGPFRNKGAVLSITTGGSGSMYSLQG
IHGDMNILLWPIQSGTLHFCGFQVLEPQLTYSIGHTPEDARIQILEEWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMKK
KQVQDEQKSNKKPGLSVGHHLGKSIPTDNQVKARK

>Bos taurus
MAVRKALIVLAHSERTSFNYAMKEAAIEALKRRGWEVTVSDLYAMNFNPVISRKDITGKLKDPGNFQYPAETVLAYKEGR
LSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVLIGEFAYKYAAMYDKGPFRNKGAVLSITTGGSGSMYSLHG
IHGDMNIILWPIQSGTLHFCGFQVLEPQLTYSIGHTPEDARVQILEGWKKRLENIWDEMPLYFAPSSLFDLNFQAGFLMKK
KEVQDEQKSKKPGLSVGHHLGKSIPMDNQIKAIK

>Mus musculus
MAARRALIVLAHSEKTSFNYAMKEAAVEALKKRGWEVLESDLYAMNFNPIISRNDITGELKDSKNFQYPSESSLAYKEGR
LSPDIVAEHKKLEAADLVIFQFPLQWFGVPAILKGWFERVLVAGFAYTYAAMYDNGPFQNKGTLLSITTGGSGSMYSLQG
VHGDMNVILWPIQSGILRFCGFQVLEPQLVYSIGHTPPDARMQILEGWKKRLETVWEETPLYFAPSSLFDLNFQAGFLMKK
KEVQEEQKGNKPGLSVGHHLGKSIPADNQIKARK

>alligator sinensis
MAAKGALIVLAHAEKTSFNHAMKDAAVDALQKKGWSVAVSDLYAMKFNPVQSREDITGKPKDPNNFNYAVEMGLAMKEGR
LSSDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFDRVFTGGFAYSYASMYDQGPFQKKGKVVLSFTTGGAGSMYTPNG
INGDVNILLWPIQNGTLYFCGFQVLEPQLIAYSIGHTPEDVRSQILMGWKERLGSIWEEKPLSVVFSSSFEMSFKGGFKLK
AETQEQQKDQKYGLSVGQHLGKAIPFDNQVKAQRK

4) Nalezeno 6 sekvencí: tři isoformy a tři "jiné" sekvence

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…