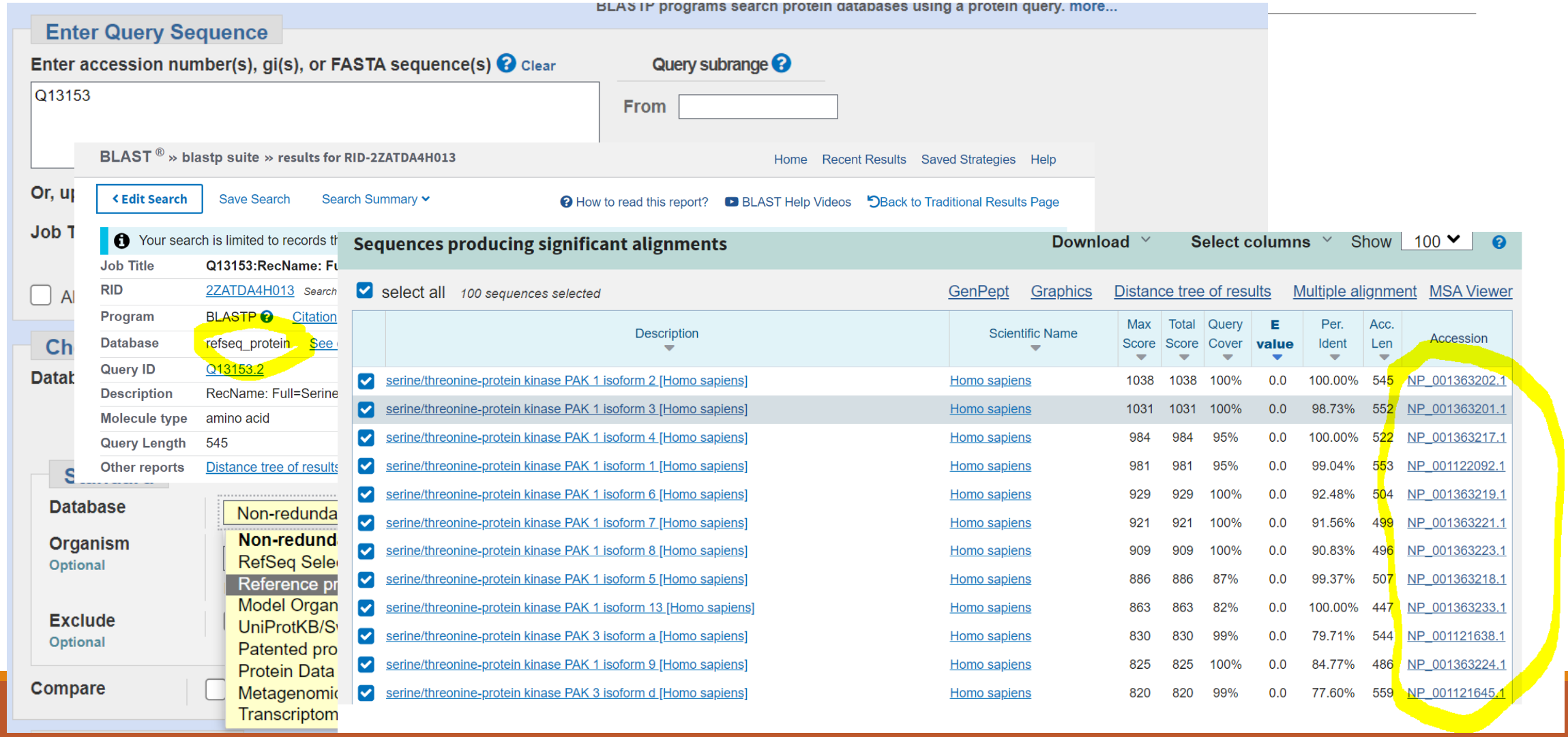
Základy praktické bioinformatiky

PETRA MATOUŠKOVÁ

2022/2023

DÚ3-lidské isoformy
- z „refseq_protein“databáze



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
serine/threonine-protein kinase PAK 1 isoform 2 [Homo sapiens]	Homo sapiens	1038	1038	100%	0.0	100.00%	545	NP_001363202.1
serine/threonine-protein kinase PAK 1 isoform 3 [Homo sapiens]	Homo sapiens	1031	1031	100%	0.0	98.73%	552	NP_001363201.1
serine/threonine-protein kinase PAK 1 isoform 4 [Homo sapiens]	Homo sapiens	984	984	95%	0.0	100.00%	522	NP_001363217.1
serine/threonine-protein kinase PAK 1 isoform 1 [Homo sapiens]	Homo sapiens	981	981	95%	0.0	99.04%	553	NP_001122092.1
serine/threonine-protein kinase PAK 1 isoform 6 [Homo sapiens]	Homo sapiens	929	929	100%	0.0	92.48%	504	NP_001363219.1
serine/threonine-protein kinase PAK 1 isoform 7 [Homo sapiens]	Homo sapiens	921	921	100%	0.0	91.56%	499	NP_001363221.1
serine/threonine-protein kinase PAK 1 isoform 8 [Homo sapiens]	Homo sapiens	909	909	100%	0.0	90.83%	496	NP_001363223.1
serine/threonine-protein kinase PAK 1 isoform 5 [Homo sapiens]	Homo sapiens	886	886	87%	0.0	99.37%	507	NP_001363218.1
serine/threonine-protein kinase PAK 1 isoform 13 [Homo sapiens]	Homo sapiens	863	863	82%	0.0	100.00%	447	NP_001363233.1
serine/threonine-protein kinase PAK 3 isoform a [Homo sapiens]	Homo sapiens	830	830	99%	0.0	79.71%	544	NP_001121638.1
serine/threonine-protein kinase PAK 1 isoform 9 [Homo sapiens]	Homo sapiens	825	825	100%	0.0	84.77%	486	NP_001363224.1
serine/threonine-protein kinase PAK 3 isoform d [Homo sapiens]	Homo sapiens	820	820	99%	0.0	77.60%	559	NP_001121645.1

Základy praktické bioinformatiky

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.

„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

Vyhledání interakcí mezi proteiny

…

Párové porovnání (pairwise alignment)

Globální porovnání – porovná sekvence po celé délce

Lokální porovnání – porovná sekvence částečně v „nejpodobnější“ oblasti (např pro vyhledání podobných domén, odhalení repetitivních sekvencí)

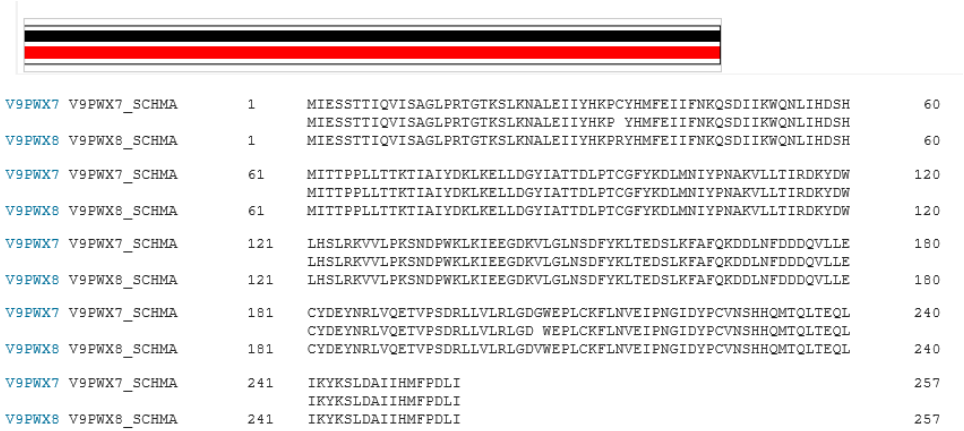
```
seq1   EARDF-NQYYSSIKRSGSIQ
         . : .:::::::::. . .
seq2   LPKLFIDQYYSSIKRTMG-H
```

global sequence alignment

```
seq1   NQYYSSIKRS
       .:::::::::.
seq2   DQYYSSIKRT
```

local sequence alignment

BLAST-součástí výstupu je párové porovnání (optimalizován na rychlost, ne přesnost)



```
V9PWX7 V9PWX7_SCHMA       1    MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKPCYHMFEIIFNKQSDIIKWQNLIHDSH      60
                               MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKP YHMFEIIFNKQSDIIKWQNLIHDSH
V9PWX8 V9PWX8_SCHMA       1    MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKPRYHMFEIIFNKQSDIIKWQNLIHDSH      60

V9PWX7 V9PWX7_SCHMA      61    MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW     120
                               MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW
V9PWX8 V9PWX8_SCHMA      61    MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW     120

V9PWX7 V9PWX7_SCHMA     121    LHSLRKVVLPKSNDPWKLKIEEGDKVLGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLE     180
                               LHSLRKVVLPKSNDPWKLKIEEGDKVLGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLE
V9PWX8 V9PWX8_SCHMA     121    LHSLRKVVLPKSNDPWKLKIEEGDKVLGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLE     180

V9PWX7 V9PWX7_SCHMA     181    CYDEYNRLVQETVPSDRLLVLRLGDGWEPLCKFLNVEIPNGIDYPCVNSHHQMTQLTEQL     240
                               CYDEYNRLVQETVPSDRLLVLRLGD WEPLCKFLNVEIPNGIDYPCVNSHHQMTQLTEQL
V9PWX8 V9PWX8_SCHMA     181    CYDEYNRLVQETVPSDRLLVLRLGDVWEPLCKFLNVEIPNGIDYPCVNSHHQMTQLTEQL     240

V9PWX7 V9PWX7_SCHMA     241    IKYKSLDAIIHMFPDLI                                          257
                               IKYKSLDAIIHMFPDLI
V9PWX8 V9PWX8_SCHMA     241    IKYKSLDAIIHMFPDLI                                          257
```

```
V9PWX7 V9PWX7_SCHMA       1    MIESSTTIQVISAGLPRTGTKSLKNALEIIYHKPCYHMFEIIFNKQSDIIKWQNLIHDSH       60
                               M ESS  + VI AGLPRTGTKSLKNALEIIYHKPCYHM EII  + +DI KWQ L  ++
A0A183QDM9 A0A183QDM9_9TREM1    MSESSNDLLVIGAGLPRTGTKSLKNALEIIYHKPCYHMTEIIIKQHNDIDKWQKLFDEAL       60

V9PWX7 V9PWX7_SCHMA      61    MITTPPLLTTKTIAIYDKLKELLDGYIATTDLPTCGFYKDLMNIYPNAKVLLTIRDKYDW      120
                                +       T  + I D LKE+L  Y A TD+P CGFYK+LMNIYPNAKVLLTIRDKYDW
A0A183QDM9 A0A183QDM9_9TREM61    KMEP-----TNELMINDGLKEILMNYGAVTDVPACGFYKELMNIYPNAKVLLTIRDKYDW      115

V9PWX7 V9PWX7_SCHMA     121    LHSLRKVVLPKSNDPWKLKIEEGDKV----------------------------------      146
                               LHSLRKVVLPKSNDPWKLKIEEGDKV
A0A183QDM9 A0A183QDM9_9TREM116    LHSLRKVVLPKSNDPWKLKIEEGDKVILTIRNKYDWLSSFRQTLMPKSNDSSNRTIDEAD      175

V9PWX7 V9PWX7_SCHMA     147    --LGLNSDFYKLTEDSLKFAFQKDDLNFDDDQVLLECYDEYNRLVQETVPSDRLLVLRLG      204
                                 L L   F K+  DS+K AF+K D + D+D  +L+C+DEYNR V ETVPS+RLL+ +LG
A0A183QDM9 A0A183QDM9_9TREM176    EILKLGPKFIKMAIDSMKLAFRKIDFDIDNDNEMLQCFDEYNRTVIETVPSERLLIRKLG      235

V9PWX7 V9PWX7_SCHMA     205    DGWEPLCKFLNVEIPNGIDYPCVN                                    228
                               DGWEPLC+FLNV++P G+ YP +N
A0A183QDM9 A0A183QDM9_9TREM236    DGWEPLCRFLNVDVPEGVSYPYIN                                    259
```

Nevýhoda-ukáže jen jeden „nejlepší“ alignment

Párové porovnání - Globální

EMBOSS Needle

Protein alignment | Nucleotide alignment | Web services | Help & Documentation

Tools > Pairwise Sequence Alignment > EMBOSS Needle

Service Retirement

We remind you that it is not long until the EBI's Wise2DBA and Promoterwise services are retired on 15th April 2018. Alternatives can be
or BLAT. If you have any concerns, please contact us via support.

Pairwise Sequence Alignment (PROTEIN)

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

This is the form for protein sequences. Please go to the nucleotide form if you wish to align DNA or RNA sequences.

STEP 1 - Enter your protein sequences

Enter or paste your first **protein** sequence in any supported format:

STEP 2 - Set your pairwise alignment options

MATRIX	GAP OPEN	GAP EXTEND	OUTPUT FORMA
BLOSUM50	10	0.5	pair

END GAP PENALTY	END GAP OPEN	END GAP EXTEND
false	10	0.5

```
#=======================================
#
# Aligned_sequences: 2
# 1: AAH09679.1
# 2: AAI47025.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 342
# Identity:     138/342 (40.4%)
# Similarity:   196/342 (57.3%)
# Gaps:          51/342 (14.9%)
# Score: 646.5
#
#
#=======================================

AAH09679.1         1 MVSPATRKSLPKVKAMDFITSTAILPLLFGCLGVFGLFRLLQWVR---GK     47
                                          .:..:||||.  ||:.||..:.|.|.   .|
AAI47025.1         1 ------------------MAMLMLPLLL--LGISGLLFIYQEVSRLWSK     29

AAH09679.1        48 AYLRNAVVVITGATSGLGKECAKVFYAAGAKLVLCGRNGGALEELIRELT     97
                      :.::|.|||||.|.|||||||:||:..||:|||||:|...||.|...|
AAI47025.1        30 SAVQNKVVVITDAISGLGKECARVFHTGGARLVLCGKNWERLENLYDAL-     78

AAH09679.1        98 ASHATKVQTHKPYLVTFDLTDSGAIVAAAAEILQCFGYVDILVNNAGISY    147
                      .|.|...:|..|.||..||:|...:...|.|:|.|:|.||||:|||.:..
AAI47025.1        79 ISVADPSKTFTPKLVLLDLSDISCVPDVAKEVLDCYGCVDILINNASVKV    128

AAH09679.1       148 RGTIMDTTVDVDKRVMETNYFGPVALTKALLPSMIKRRQGHIVAISSIQG    197
                      :|.....:::::||::|:.|||||:.|||||||:||.||.|.||.:::|||
AAI47025.1       129 KGPAHKISLELDKKIMDANYFGPITLTKALLPNMISRRTGQIVLVNNIQG    178

AAH09679.1       198 KMSIPFRSAYAASKHATQAFFDCLRAEMEQYEIEVTVISPGYIHTNLSVN    247
                      |..||||:.|||||||...||||||||:|:|::.::.:||.:|.   |.:
AAI47025.1       179 KFGIPFRTTYAASKHAALGFFDCLRAEVEEYDVVISTVSPTFIR---SYH    225

AAH09679.1       248 AITADGS-------------RYGVMDTTTAQGRSPVEVAQDVLAAVGKKK    284
                      .....|:             .|||         .||||||:|:..|.:||
AAI47025.1       226 VYPEQGNWEASIWKFFFRKLTYGV---------HPVEVAEEVMRTVRRKK    266

AAH09679.1       285 KDVILADLLPSLAVYLRTLAPGLFFSLMASRARKERKSKNS-    325
                      ::|.:|:.:|..|||:||..|..||:::|...:::.....
AAI47025.1       267 QEVFMANPIPKAAVYVRTFFPEFFFAVVACGVKEKLNVPEEG    308
```

Párové porovnání - Lokální

LALIGN

Protein alignment | Nucleotide alignment | Web services | Help & Documentation | Feedback

Tools > Pairwise Sequence Alignment > LALIGN

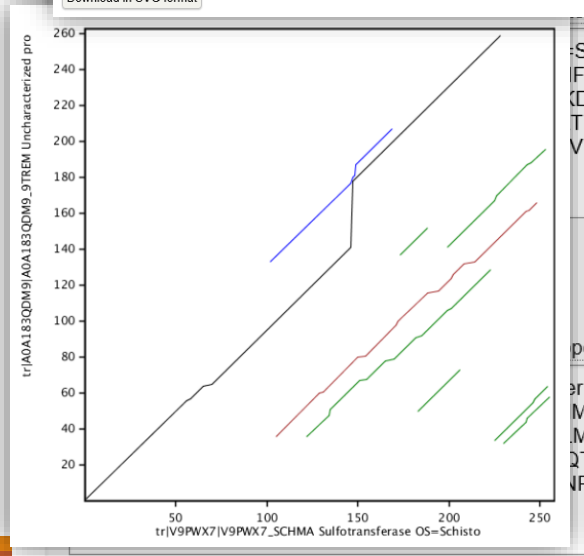
Pairwise Sequence Alignment

LALIGN finds internal duplications by calculating non-intersecting local alignments of **protein** or nucleotide sequences.

Results for job lalign-I20180327-155818-0289-6701829-p2m

Alignment | Visual Output | Submission Details

Download in SVG format



d format:

-Schistosoma mansoni GN=SULT-OR PE=2 SV=1
IFEIIFNKQSDIIKWQNLIHDSH
CDLMNIYPNAKVLLTIRDKYDW
TEDSLKFAFQKDDLNFDDDQVLLE
VEIPNGIDYPCVNSHHQMTQLTEQL

pported format:

erized protein OS=Schistosoma rodhaini PE=4 SV=1
MTEIIIKQHNDIDKWQKLFDEAL
MNIYPNAKVLLTIRDKYDWLHSLR
QTLMPKSNDSSNRTIDEADEILKL
NRTVIETVPSERLLIRKLGDGWEP

Or, upload a file: Procházet...

Results for job lalign-I20180327-155818-0289-6701829-p2m

Alignment | Visual Output | Submission Details

View Alignment File

```
>>AAI47025.1 Dehydrogenase/reductase (SDR family) member  (308 aa)
 Waterman-Eggert score: 840; 253.1 bits; E(1) <  6.6e-72
45.0% identity (76.3% similar) in 300 aa overlap (24-319:5-301)

               30        40        50        60        70        80
AAH096 ILPLLFGCLGVFGLFRLLQWVR---GKAYLRNAVVVITGATSGLGKECAKVFYAAGAKLV
       .::::.  ::. ::. . : :    .:. ..: ::::: : :::::::::.::...::.::
AAI470 MLPLLL--LGISGLLFIYQEVSRLWSKSAVQNKVVVITDAISGLGKECARVFHTGGARLV
            10          20        30        40        50        60

               90       100       110       120       130       140
AAH096 LCGRNGGALEELIRELTASHATKVQTHKPYLVTFDLTDSGAIVAAAAEILQCFGYVDILV
       :::.:   ::.:   :  : :   .:  : :: .::.: . .  .: :.:.:.: :::::.
AAI470 LCGKNWERLENLYDALI-SVADPSKTFTPKLVLLDLSDISCVPDVAKEVLDCYGCVDILI
            70        80        90       100       110       120

              150       160       170       180       190       200
AAH096 NNAGISYRGTIMDTTVDVDKRVMETNYFGPVALTKALLPSMIKRRQGHIVAISSIQGKMS
       ::::.. .:     ....::..:.::::::.:::::::.::.:: :.:: ...::::..
AAI470 NNASVKVKGPAHKISLELDKKIMDANYFGPITLTKALLPNMISRRTGQIVLVNNIQGKFG
            130       140       150       160       170       180

              210       220       230       240       250
AAH096 IPFRSAYAASKHATQAFFDCLRAEMEQYEIEVTVISPGYIHT-NLSVNAITADGSRYGVM
       :::::.:::::::. .:::::::::.:.:.. ....:: .:.. ..  .  . ..: .  .
AAI470 IPFRTTYAASKHAALGFFDCLRAEVEEYDVVISTVSPTFIRSYHVYPEQGNWEASIWKFF
            190       200       210       220       230       240

     260       270       280       290       300       310
AAH096 DTTTAQGRSPVEVAQDVLAAVGKKKKDVILADLLPSLAVYLRTLAPGLFFSLMASRARKE
          . :  :::::..:. .: .::..:.:. .:. :::.::. : .::...:  ....
AAI470 FRKLTYGVHPVEVAEEVMRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFFAVVACGVKEK
            250       260       270       280       290       300

>--
 Waterman-Eggert score: 34; 14.4 bits; E(1) <  0.99
42.1% identity (63.2% similar) in 19 aa overlap (289-307:74-92)

      290       300
AAH096 LADLLPSLAVYLRTLAPGL
       : : : :.:   .:..: :
AAI470 LYDALISVADPSKTFTPKL
            80        90
```

Vyzkoušejte si porovnat 2 sekvence

Porovnejte váš (lidský) protein s homologním proteinem z myši (Mus musculus). (DÚ)

Vyzkoušejte oba programy:

- Jaká je identita/similarita těchto dvou proteinů?
  → globální porovnání
- Jak dlouhé je nejlepší lokální porovnání těchto dvou sekvencí?
  → lokální porovnání

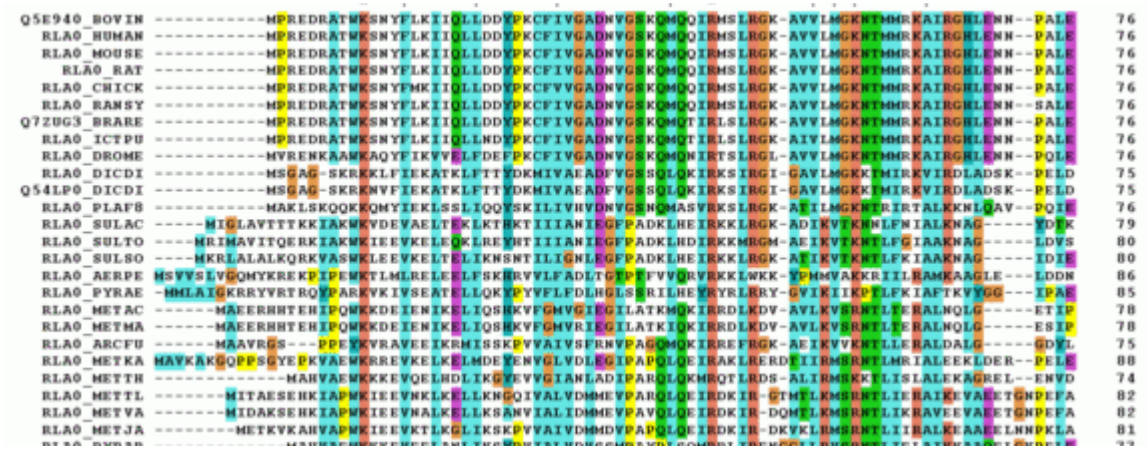
Mnohonásobné porovnání (multiple sequence alignment)

= přepsání sekvencí, tak aby stejné aminokyseliny či oblasti byly srovnány pod sebou ve sloupcích

Získání sekvencí (např: porovnání různých isoforem, porovnání vícero výstupů z BLASTu..)

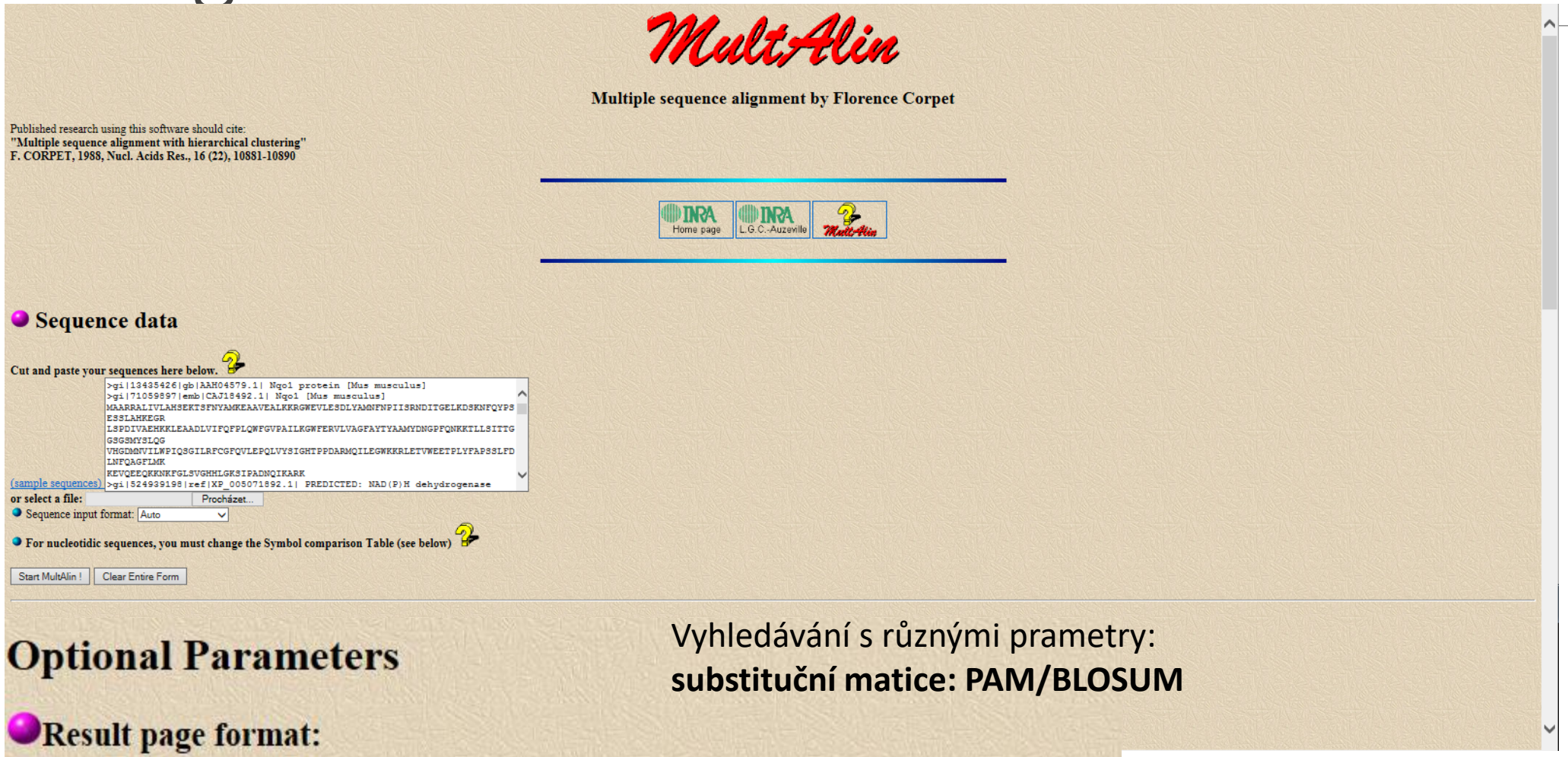
Vlastní porovnání (různé programy, různé matice, různé „výstupy“…)

Umožňuje fylogenetickou analýzu (fylogenetické stromy - evoluční příbuznost)



Alignment: MultAlin

http://multalin.toulouse.inra.fr/multalin/



Vyhledávání s různými prametry:
substituční matice: PAM/BLOSUM

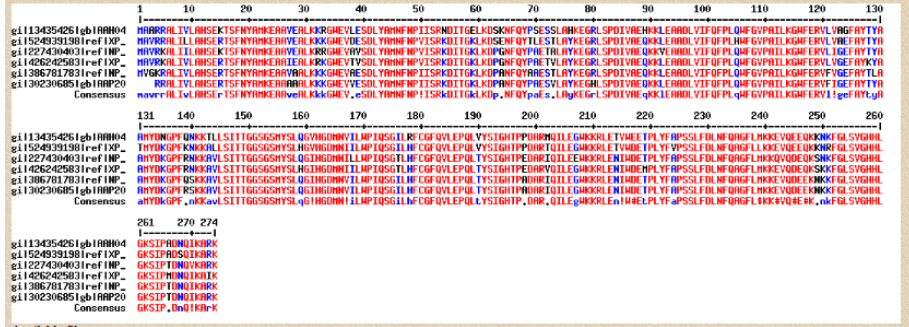
MultAlin

Multalin result page

Back | Home page | L.G.C.-Auzeville | MultAlin

Go directly to Alignment

```
Multalin version 5.4.1
Copyright I.N.R.A. France 1989, 1991, 1994, 1996
Published research using this software should cite
Multiple sequence alignment with hierarchical clustering
F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890
Symbol comparison table: blosum62
Gap weight: 12
Gap length weight: 2
Consensus levels: high=90% low=50%
Consensus symbols:
 ! is anyone of IV
 $ is anyone of LM
 % is anyone of FY
 # is anyone of NDQEBZ

 MSF:   274    Check:   0        .
 Name: gi|13435426|gb Len:   274  Check: 6705  Weight:  1.23
 Name: gi|524939198|r Len:   274  Check: 6867  Weight:  1.23
 Name: gi|227430403|r Len:   274  Check: 6661  Weight:  0.89
 Name: gi|426242583|r Len:   274  Check: 5108  Weight:  0.89
 Name: gi|386781783|r Len:   274  Check: 4019  Weight:  0.89
 Name: gi|30230685|gb Len:   274  Check: 4190  Weight:  0.89
 Name: Consensus      Len:   274  Check: 4506  Weight:  0.00

//
```



Available files:
-Sequence Input file
-Cluster file
-Results as a fasta file
-Results as a text page (msf)

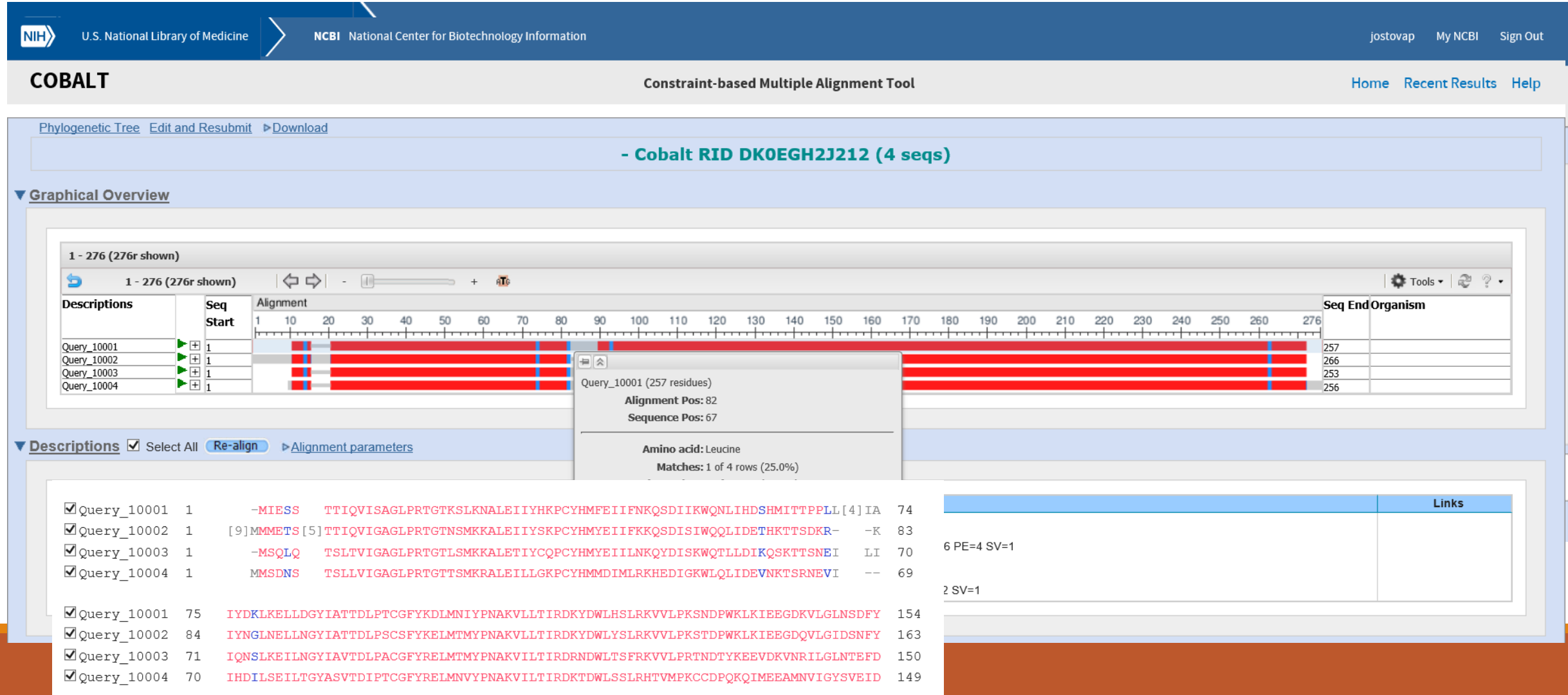
-Results as postscript page(s) with ESPript (protein only)
-Alignment and tree description (rfd) Get a better view of your protein family : phylogenetic tree, pruned tree and subtrees, summarised coloured alignment and subalignments.
- Results as an html page (needs to enable style sheets)

-Results as a text page with colour indications (need a text editor)
-Results as a gif image

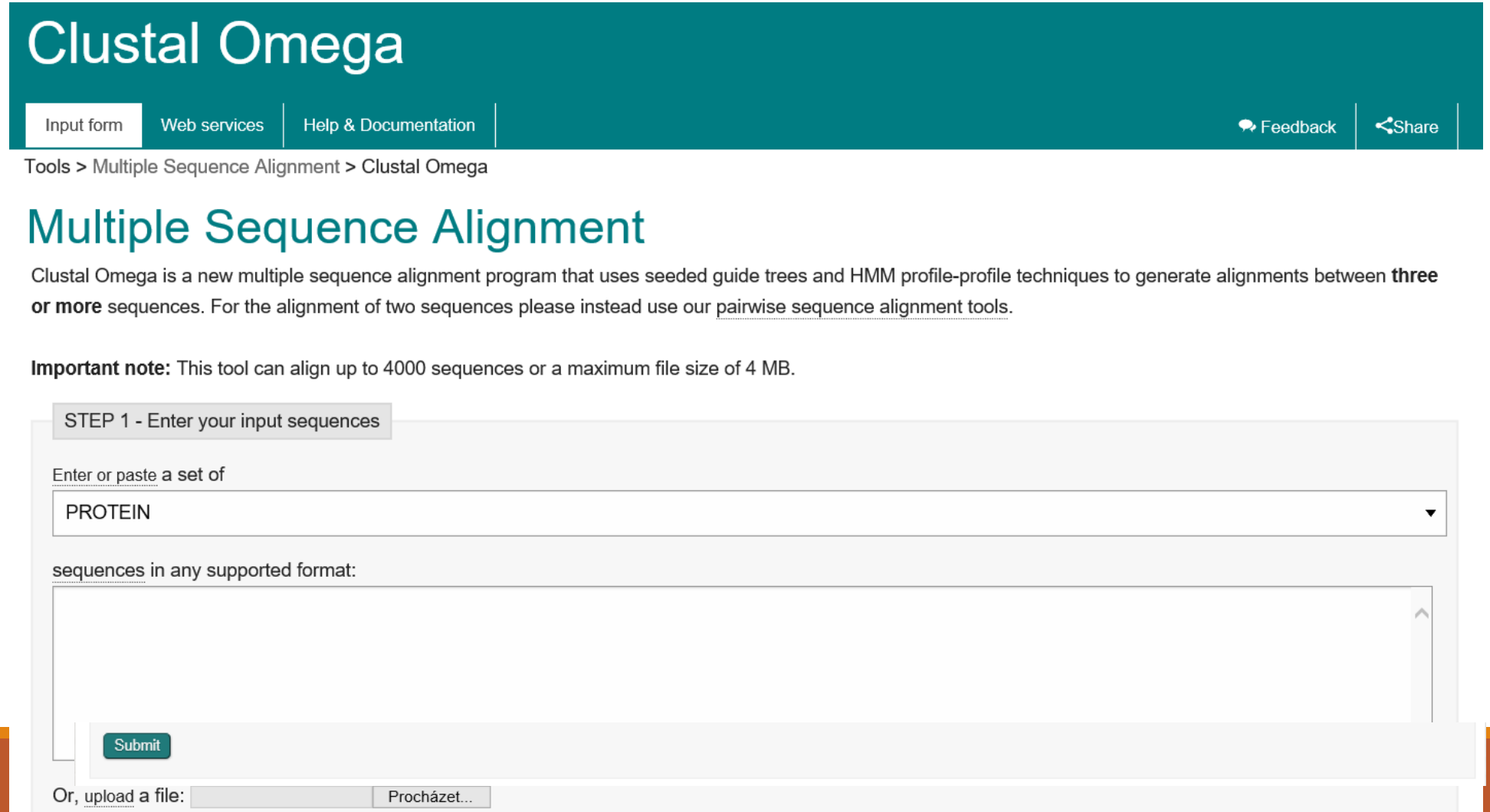
Table 9-7 Patterns of Conservation in Multiple Sequence Alignments

Amino Acid	*Characteristic*
W,Y, F	It is common to find conserved tryptophans. Tryptophan is a large hydrophobic residue that sits deep in the core of proteins. It plays an important role in their stability and is therefore difficult to mutate. When tryptophan mutates, it is usually replaced by another aromatic amino acid, such as phenylalanine or tyrosine. Patterns of conserved aromatic amino acids constitute the most common signatures for recognizing protein domains.
G, P	It is common to find conserved columns with a glycine or a proline in a multiple alignment. These two amino acids often coincide with the extremities of well-structured beta strands or alpha helices. (For more on these structures, see Chapter 11.)
C	Cysteines are famous for making C-C (disulphide) bridges. Conserved columns of cysteines are rather common and usually indicate such bridges. Columns of conserved cysteines with a specific distance provide a useful signature for recognizing protein domains and folds.
H, S	Histidine and serine are often involved in catalytic sites, especially those of proteases. Conserved histidine or a conserved serine are good candidates for being part of an active site.
K, R, D, E	These charged amino acids are often involved in ligand binding. Highly conserved columns can also indicate a salt bridge inside the core of the protein.
L	Leucines are rarely very conserved unless they're involved in protein-protein interactions such as a leucine zipper.

Alignment: NCBI/COBALT



Alignment: Clustal Omega



Alignment: Clustal Omega

Results for job clustalo-I20200329-181722-0961-32853360-p2m

Alignments | Re

Download Alignr

CLUSTAL O(1.2.4) mu

```
Mus       MAARRAL
Rattus    MAVRRAL
Homo      MVGRRAL
Bos       MAVRKAL
Sus       MAVRKAL
          *. *:**

Mus       KDSKNFQ
Rattus    KDSENFQ
Homo      KDPANFQ
Bos       KDPGNFQ
Sus       KDPGNFQ
          **  ***
```

Results for job clustalo-I20200329-181722-0961-32853360-p2m

Alignments | Result

Download Alignmen

CLUSTAL O(1.2.4) multip

```
Mus       MAARRALIVLA
Rattus    MAVRRALIVLA
Homo      MVGRRALIVLA
Bos       MAVRKALIVLA
Sus       MAVRKALIILA
          *. *:***:**

Mus       KDSKNFQYPSE
Rattus    KDSENFQYPVE
Homo      KDPANFQYPAE
Bos       KDPGNFQYPAE
Sus       KDPGNFQYPAE
          **  ***** *
```

Results for job clustalo-I20200329-181722-0961-32853360-p2m

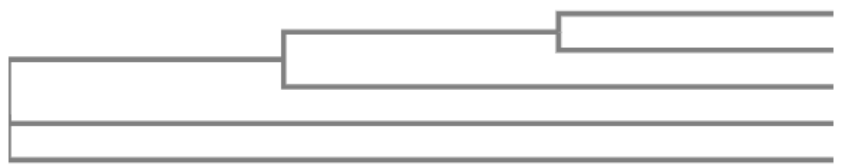
Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details

Download Phylogenetic Tree Data

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ◉ Cladogram ○ Real



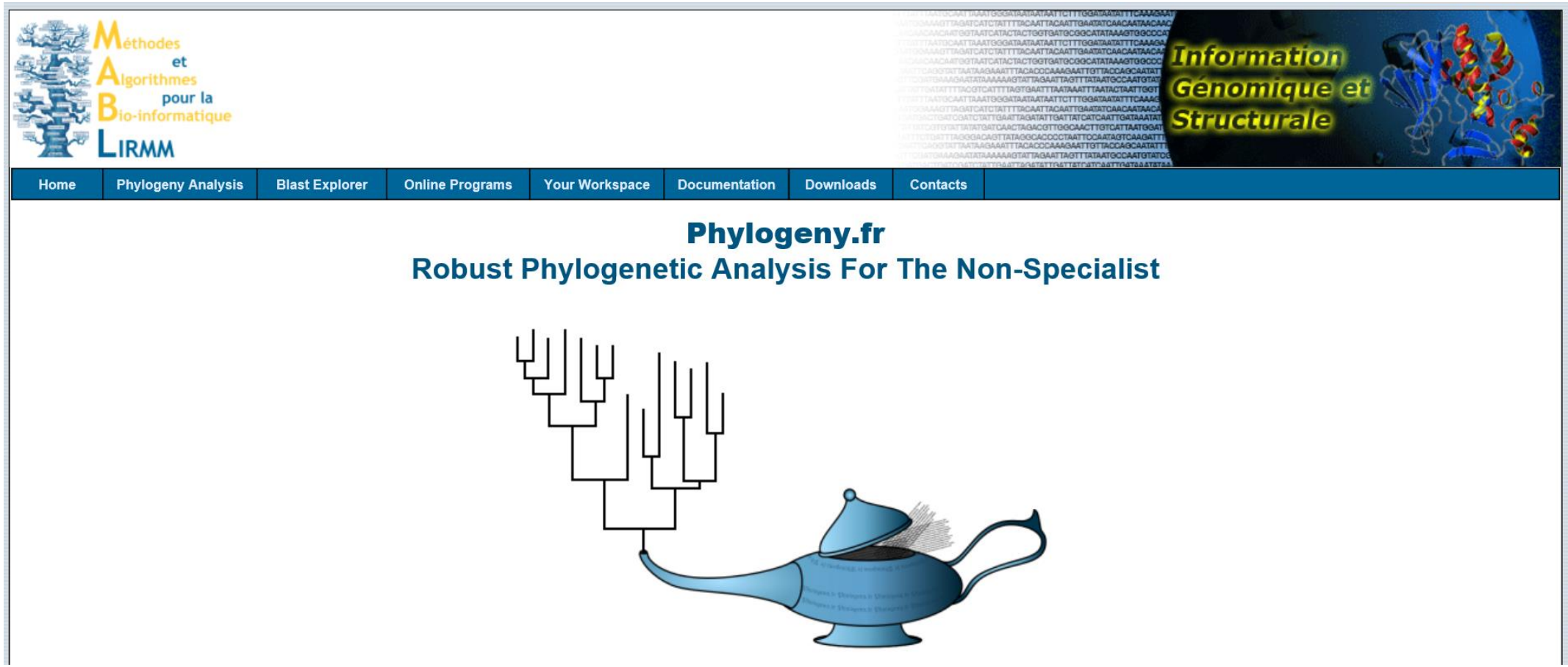
Mus 0.02555
Rattus 0.0365
Homo 0.04197
Bos 0.02828
Sus 0.03741

(*) konzervov
(:) aminokyse
(.) aminokyseliny s podobnou velikosti NEBO hydrofobicitou

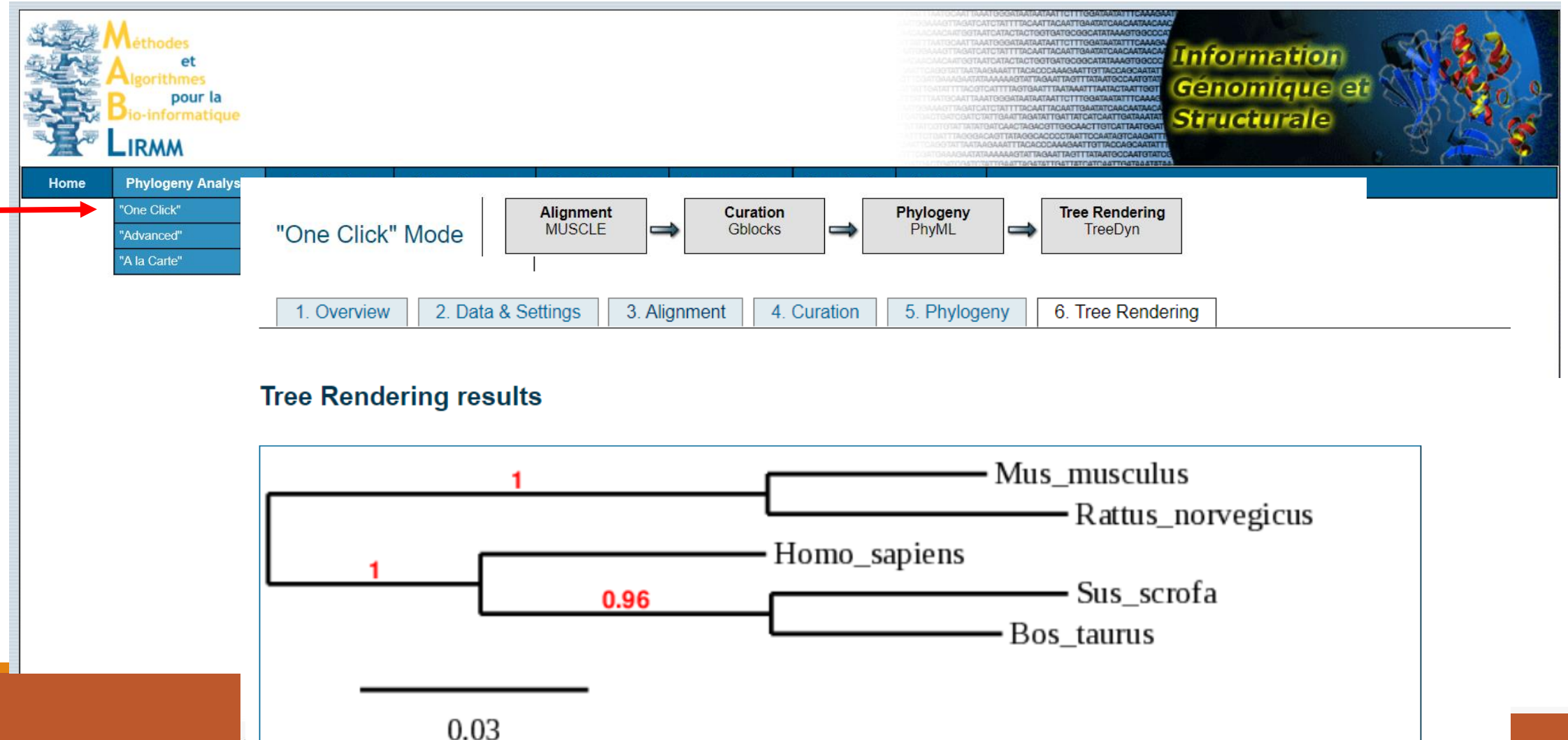
Vyzkoušejte si porovnat sekvence uložené v minulém úkolu

Vyzkoušejte víc programů!

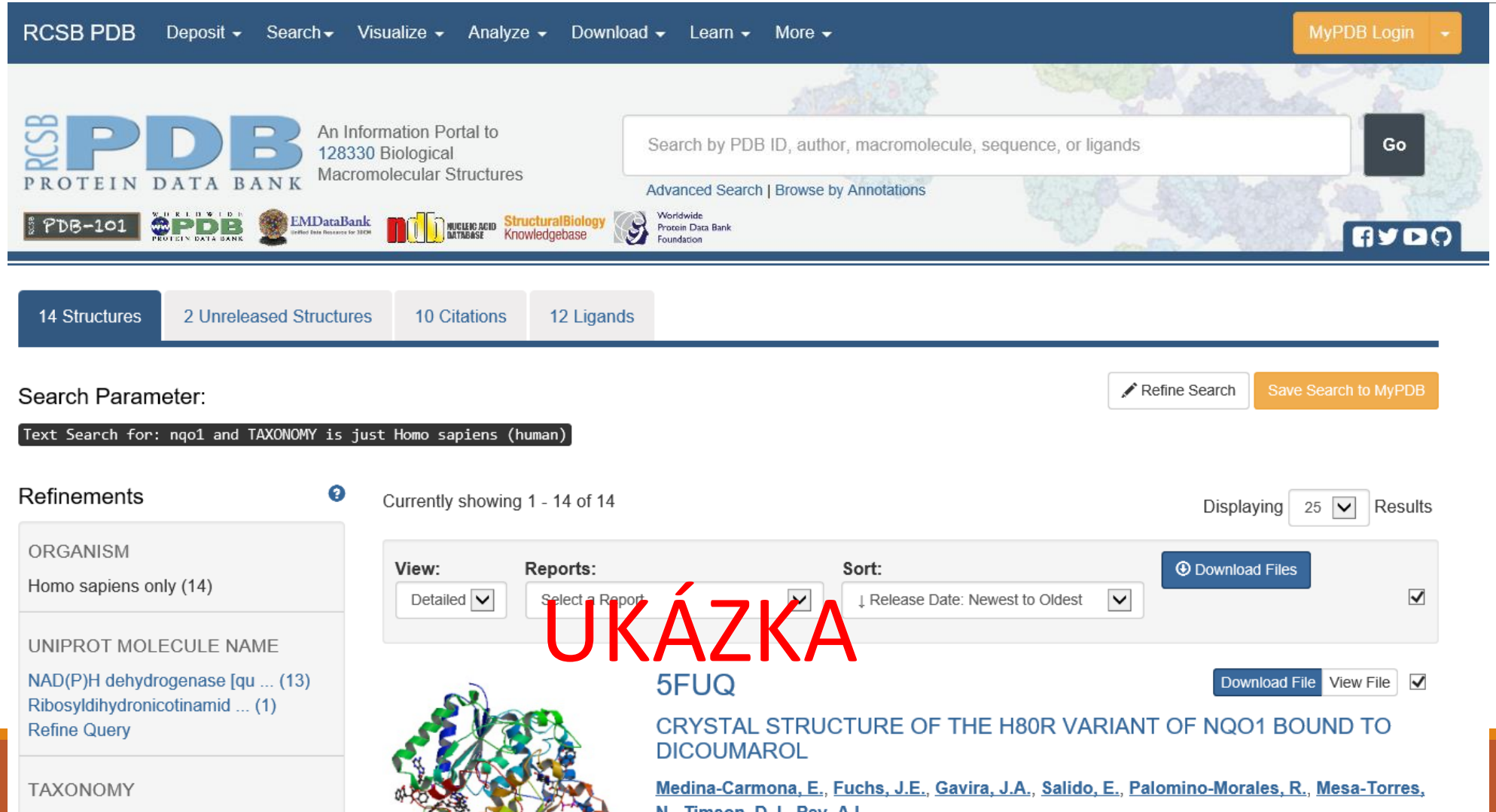
„pokročilejší“ fylogenetické porovnávání



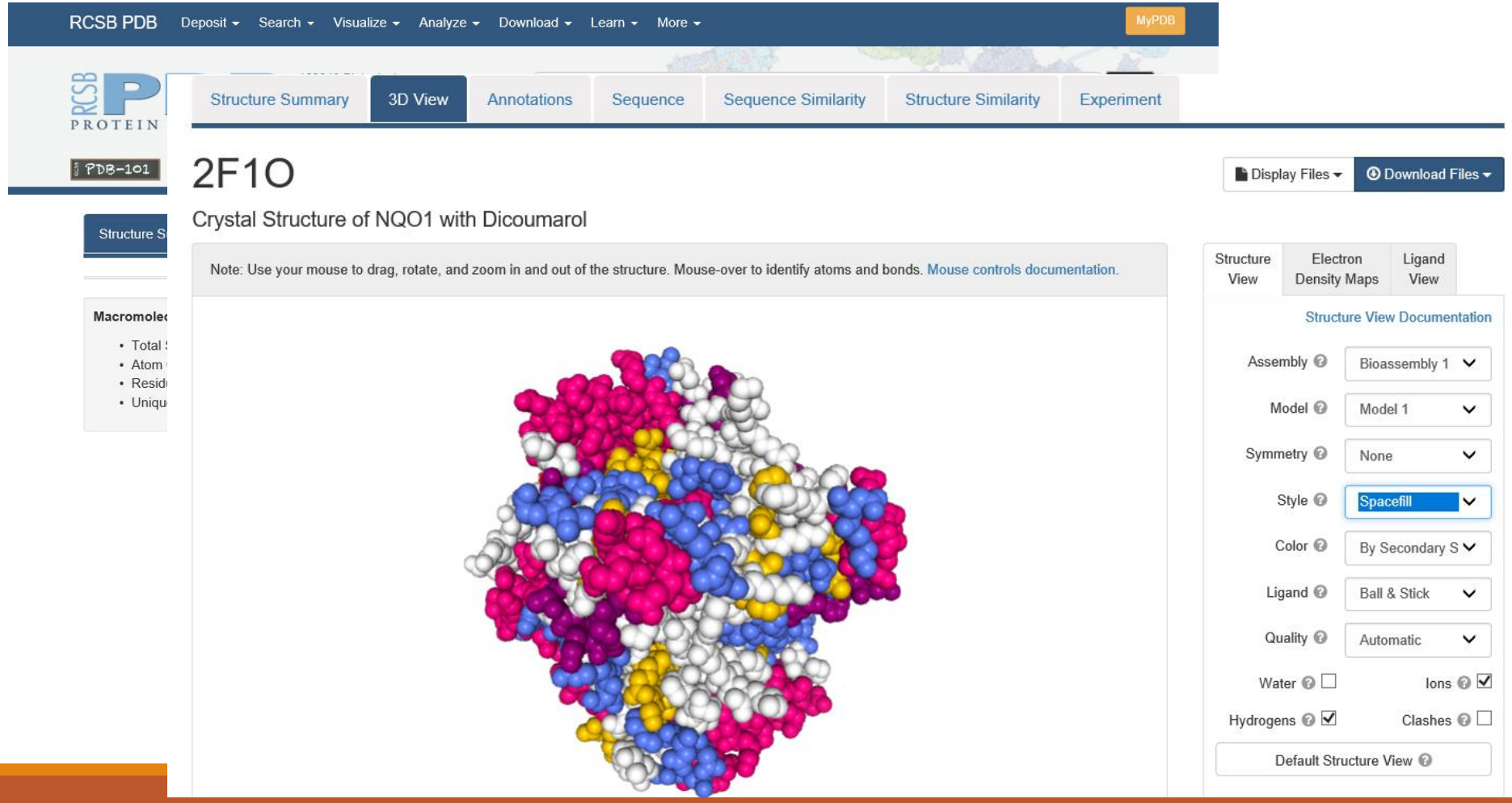
„pokročilejší“ fylogenetické porovnávání



3-D struktura proteinů: PDB



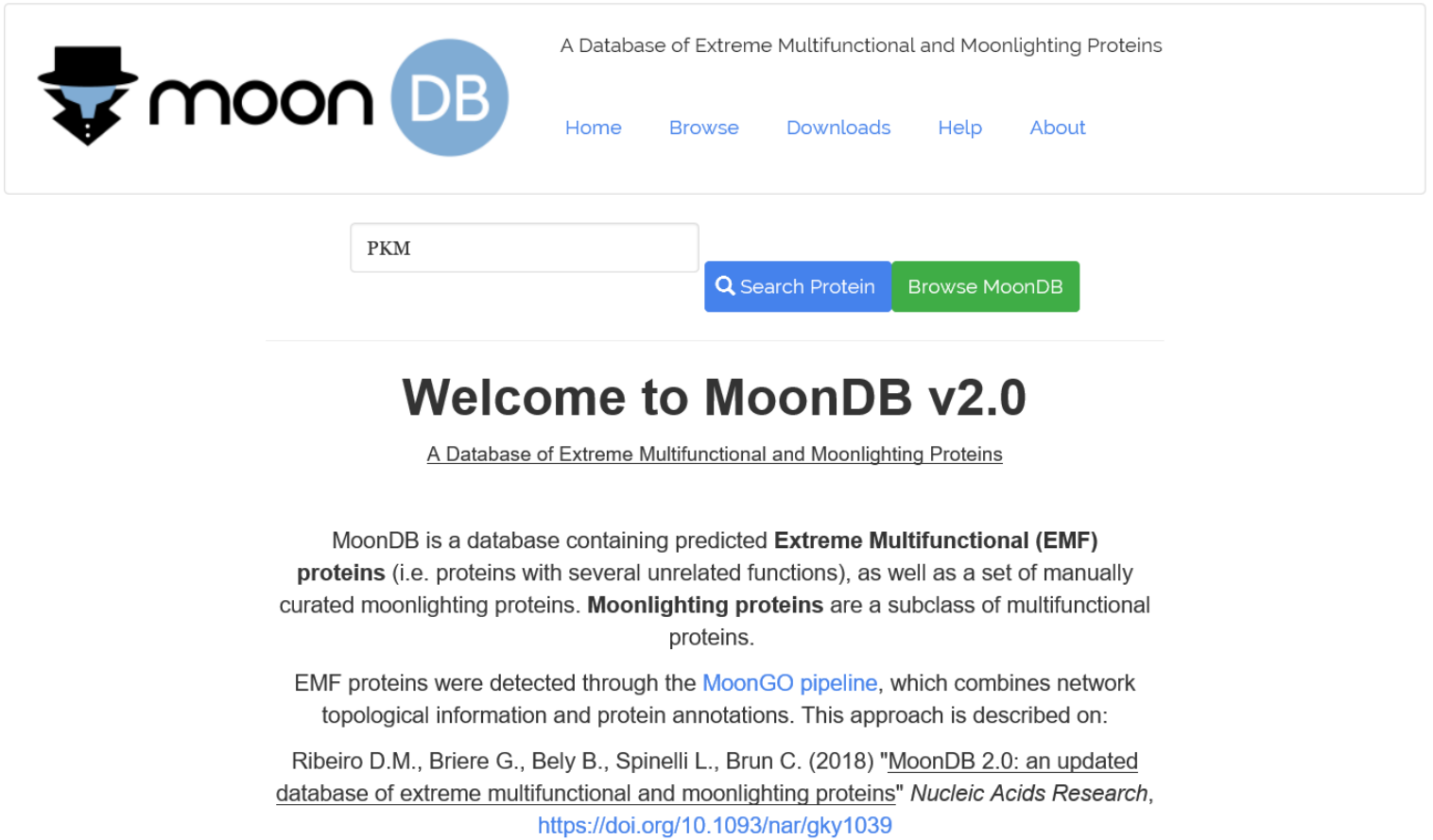
3-D struktura proteinů: PDB



Vyzkoušejte si PDB databázi

(Nemá-li váš protein vyřešenou struktury, vyzkoušejte NQO1)

Multifunkční proteiny: moondb



Fosforylace proteinů:PhosphoSite

PhosphoSitePlus®

Powered by Cell Signaling Technology

Feedback | Contact | Login

Home Search Browse Tools Downloads About

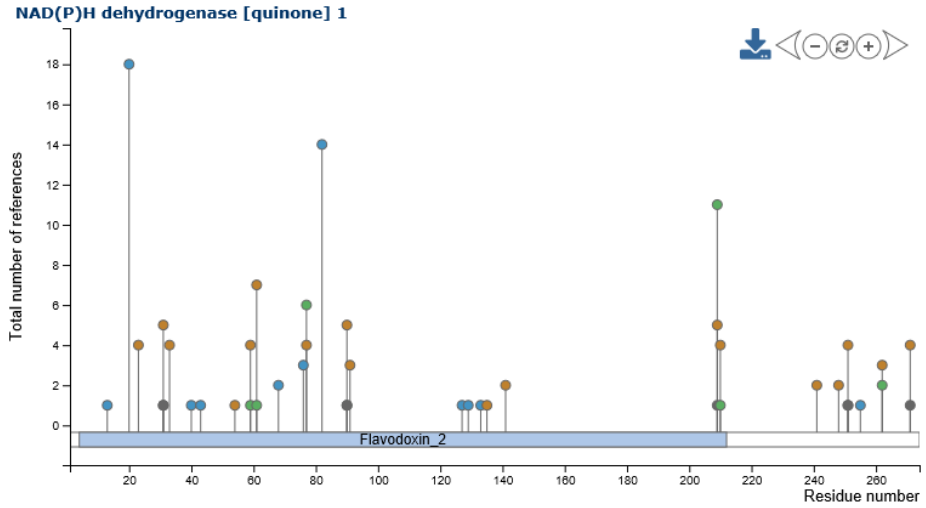
PhosphoSitePlus® provides comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, acetylation, and more. The web use is free for everyone including commercial.

Protein or Substrate Search

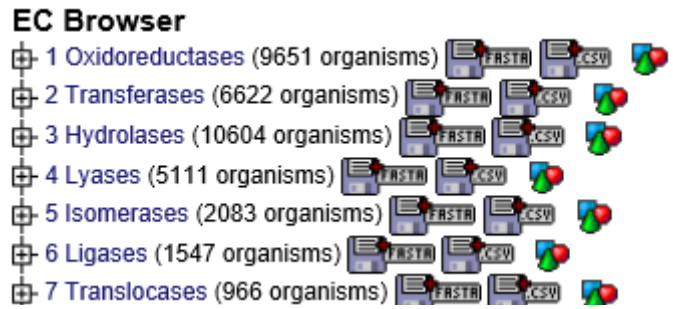
Protein Name | Search

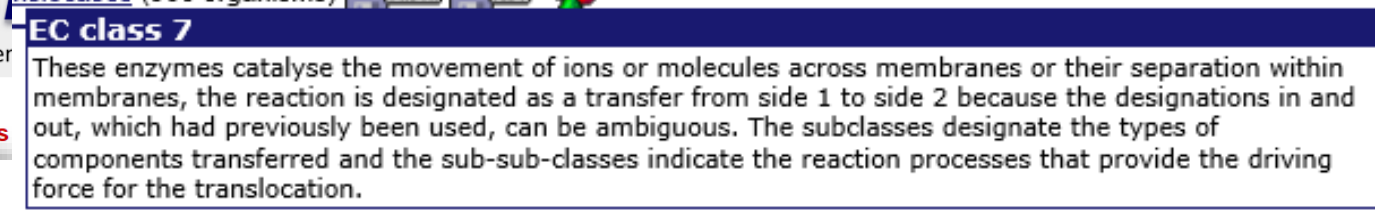
Alternative Search Options

- Protein, Sequence, or Reference Search
- Site Search
- Comparative Site Search
- Browse MS2 Data by Disease
- Browse MS2 Data by Cell Line
- Browse MS2 Data by Tissue



Enzymová databáze: Brenda





A new class EC 7, Translocases, is

Rozdělení do tříd

- 1. OXIDOREDUKTASY – oxidačně-redukční děje
  - donor + akceptor → oxidovaný donor + redukovaný akceptor
  - **Systematický název**: donor : akceptor-oxidoreduktasa
- 2. TRANSFERASY – přenos funkčních skupin
  - donor_SK + akceptor → donor + akceptor_SK
  - **Systematický název** : donor : akceptor_skupinatransferasa
- 3. HYDROLASY – hydrolýza
  - A - B + H_2O → AOH + HB
  - **Systematický název**: substrát (skupina) hydrolasa
- 4. LYASY – eliminace skupin za vzniku dvojných vazeb
  - substrát 1 (+ substrát2) → produkt1 + produkt2 (malý)
  - **Systematický název**: substrát1 (:substrát 2)_ produkt2lyasa
- 5. ISOMERASY – izomerace
  - **Systematický název**: substráttyp
- 6. LIGASY – tvorba vazeb spojená s hydrolýzou ATP
  - substrát1 + substrát2 + A(G) TP → substrát1_substrát2 + ADP + Pi
  - substrát1 + substrát2 + ATP → substrát1_substrát2 + AMP + PPi
  - **Systematický název**: substrát1 :substrát2_ligasa (tvořící ADP/AMP)

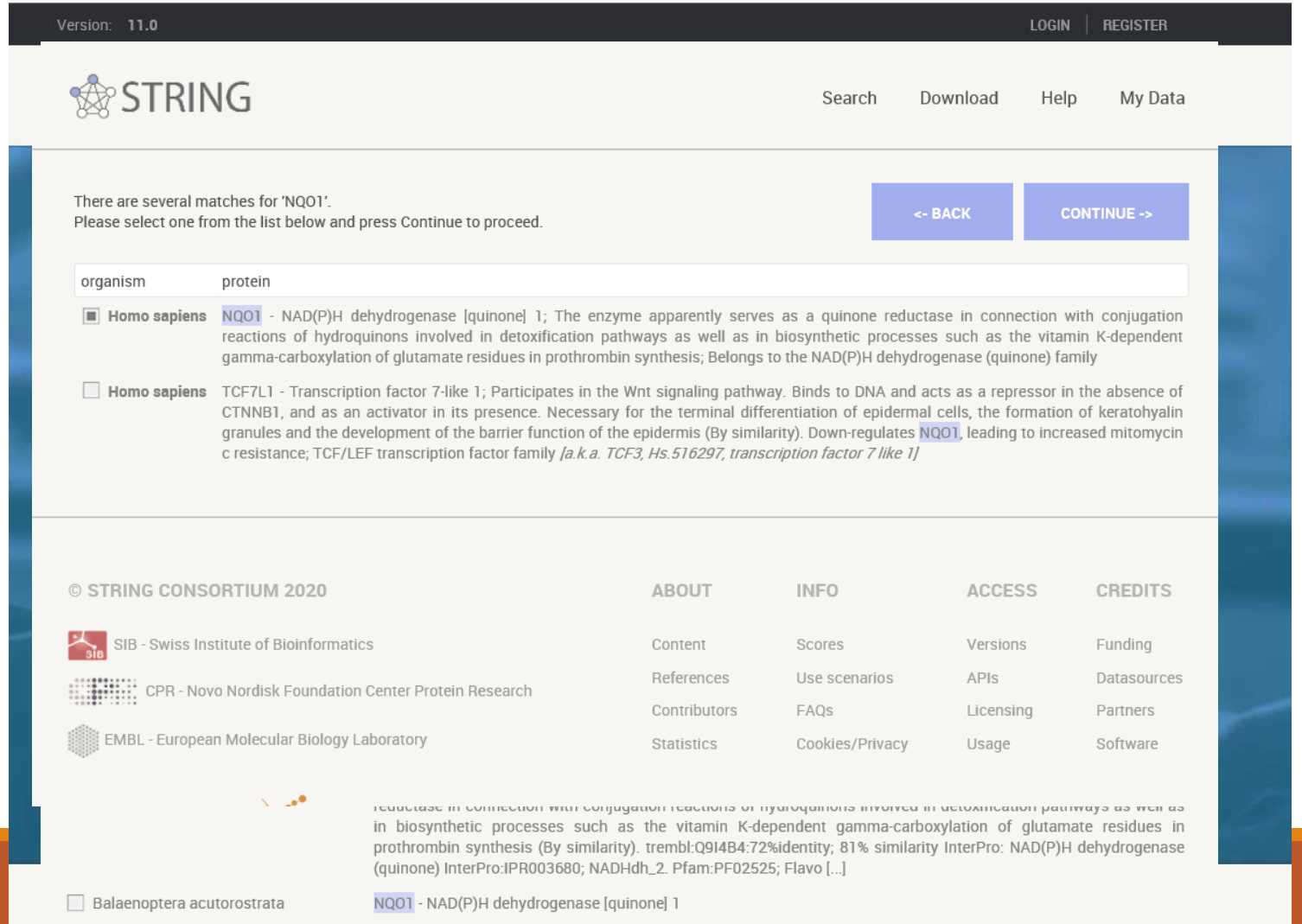
Příklad systematického názvosloví

- E.C. Enzyme Classification
- 1 Oxidoreduktasa
  - 1.1 působící na CH-OH skupiny donoru
    - 1.1.1 s NAD nebo NADP jako akceptorem
      - 1.1.1.1 alkohol dehydrogenasa ☺

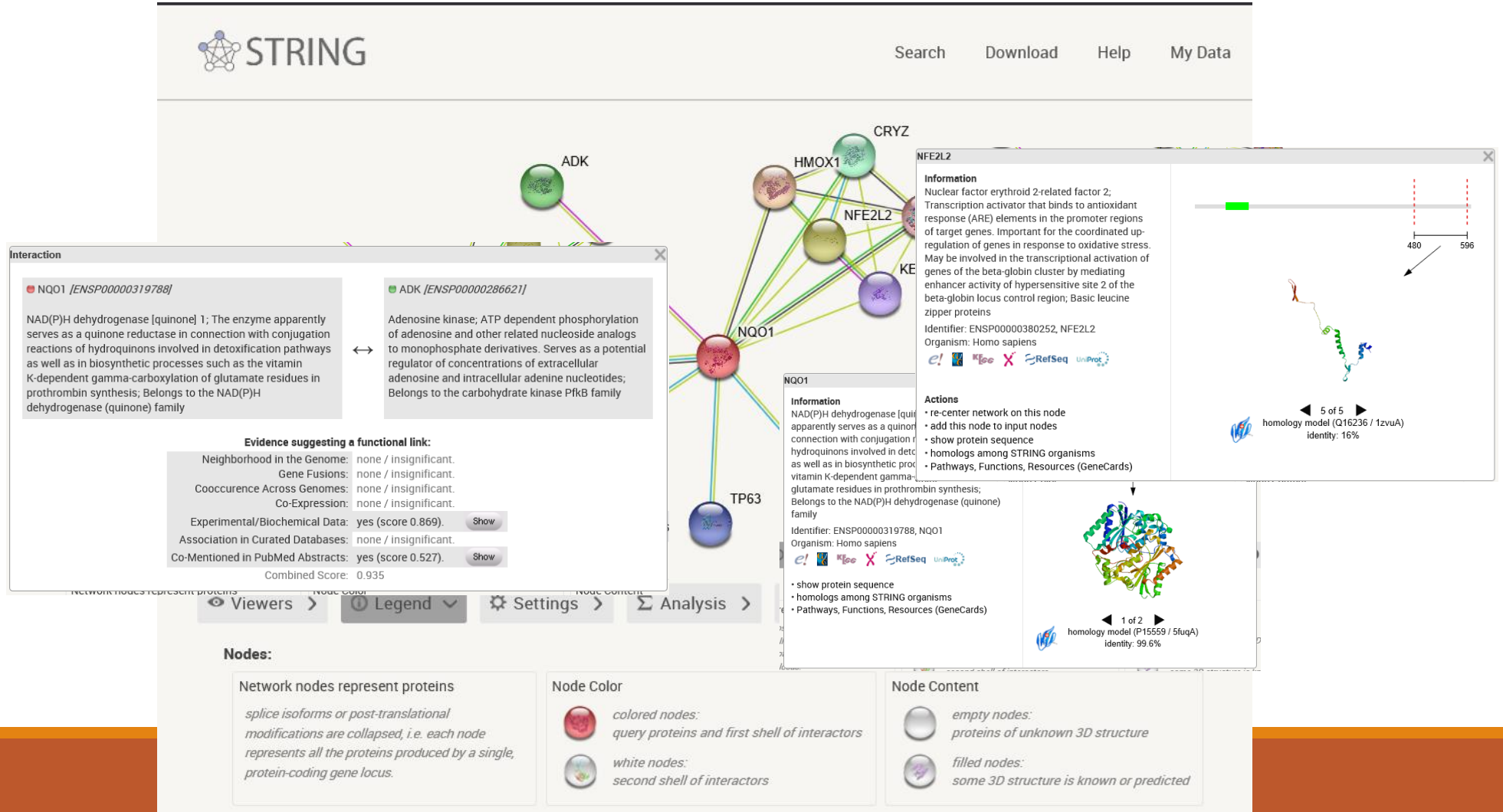
EC 7 – translokázy: Transport látek, nejčastěji přes biologické membrány;

některé translokázy vyžadují ATP; tato třída byla nově zavedena v r. 2018.

Interakce proteinů



Interakce proteinů



Podívejte se do vybraných databází

Je váš protein multifunkční?

Má typická fosforylační místa?

Má nějaké interakční partnery?

Je váš protein enzym? Jaké má enzymové číslo (E.C.x.x.x.x)? (DÚ)

DÚ4: porovnávání proteinů, 3D struktury

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

1) Porovnejte „váš“ protein se „stejným“ proteinem z myši – párové porovnání. Jaká je identita těchto dvou sekvencí? Proběhlo porovnání celé délky sekvence?
2) Vytvořte mnohonásobné porovnání všech sekvencí z minulého úkolu (DÚ3) + vystřihněte „evoluční strom“
3) Má váš protein nějaké isoformy? Porovnejte je…
4) Byla určena 3D struktura vašeho proteinu? Vystřihněte jednu na ukázku.
5) Je váš protein enzym? Jaké má enzymové číslo (EC)?

Příklad řešení

DÚ4

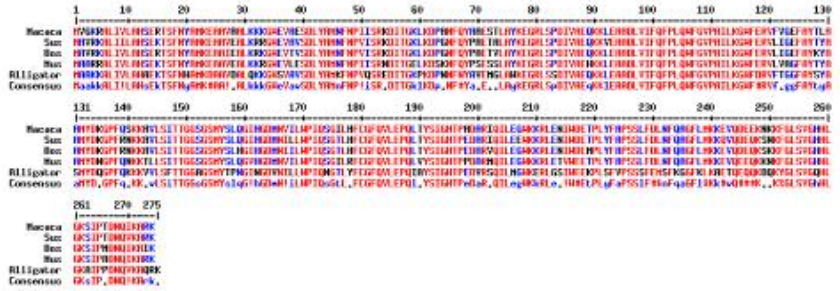
1)

```
>>sp|Q64669|NQO1_MOUSE NAD(P)H dehydrogenase [quinone] 1  (274 aa)
 Waterman-Eggert score: 1626;  421.3 bits; E(1) <  1.1e-122
 86.5% identity (97.8% similar) in 274 aa overlap (1-274:1-274)

                 10        20        30        40        50        60
sp|P15 MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
       :.:::::::::::.:::::::::::. :::::.::::.::::::::::::::::::.:::.:
sp|Q64 MAARRALIVLAHSEKTSFNYAMKEAAVEALKKRGWEVLESDLYAMNFNPIISRNDITGEL
                 10        20        30        40        50        60

                 70        80        90       100       110       120
sp|P15 KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
       ::  :::::.:: :::::::.::::::::.:::::::::::::::::::::::::::::::
sp|Q64 KDSKNFQYPSESSLAYKEGRLSPDIVAEHKKLEAADLVIFQFPLQWFGVPAILKGWFERV
                 70        80        90       100       110       120

                130       140       150       160       170       180
sp|P15 FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC
       ... ::::::::::.:::..::.:::::::::::::::::.:::::::::::::::::.::
sp|Q64 LVAGFAYTYAAMYDNGPFQNKKTLLSITTGGSGSMYSLQGVHGDMNVILWPIQSGILRFC
                130       140       150       160       170       180

                190       200       210       220       230       240
sp|P15 GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
       :::::::::.::::::: :::.:::::::::::..:.:::::::::::::::::::::::
sp|Q64 GFQVLEPQLVYSIGHTPPDARMQILEGWKKRLETVWEETPLYFAPSSLFDLNFQAGFLMK
                190       200       210       220       230       240

                250       260       270
sp|P15 KEVQDEEKNKKPGLSVGHHLGKSIPTDNQIKARK
       ::::.:.:..:::::::::::::::.:::::::
sp|Q64 KEVQEEQKKNKPGLSVGHHLGKSIPADNQIKARK
                250       260       270
```

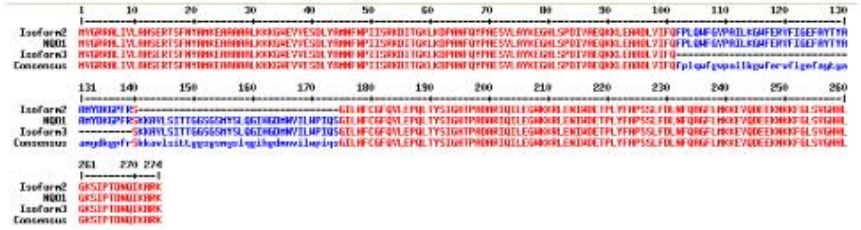
i:86,5%

Porovní proběhlo v celém rozsahu obou proteinů.

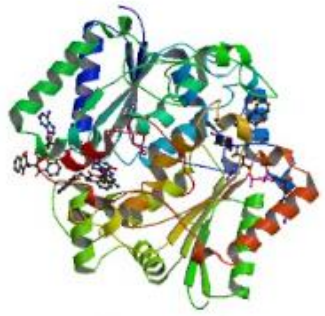
2)



3)



4)



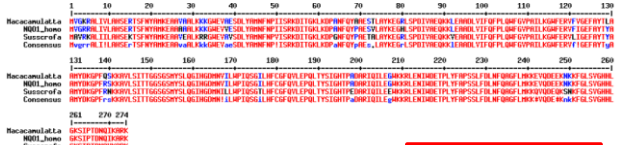
shrnutí I - Rešeršní projekt: NQO1

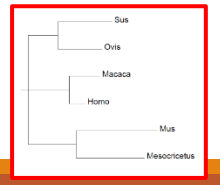
Vyhledejte dostupné informace o NQO1, získejte co nejvíce literárních, sekvenčních, případně i strukturních a dalších údajů o tomto genu/proteinu.

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
- **Rozpustný protein**
- **Podobné proteiny vyhledány, FASTA soubory uloženy**
- **Sekvence porovnány (evoluční strom)**
- **3D struktura**





Procvičování (proteinová část)

Pracujte s následující sekvencí peptidu (Moodle-procvičování)

```
VTNLFILNLAISDLLVGIFCMPITLLDNIIAGWPFGNTMCKISGLVQGISVAASVFTLVA
IAVDRFQCVVYPFKPKLTIKTAFVIIMIIWVLAITIMSPSAVMLHVQEEKYYRVRLNSQN
KTSPVYWCREDWPNQEMRKIYTTVLFANIYLAPLSLIVIMYGRIGISLFRAAVPHTGRKN
QEQWHVVSRKKQKIIKMLLIVALLFILSWLPLWTLMMLSDYADLSPNELQIINIYIYPFA
HWLAFGNSSVNPIIYGFFNENFRRGFQEAFQLQLCQKRAKPMEAYALKAKSHVLINTSNQ
```

1) Identifikujte příslušný protein, zapište přístupový kód referenční sekvence.

2) Jaká je molekulová hmotnost tohoto peptidu?

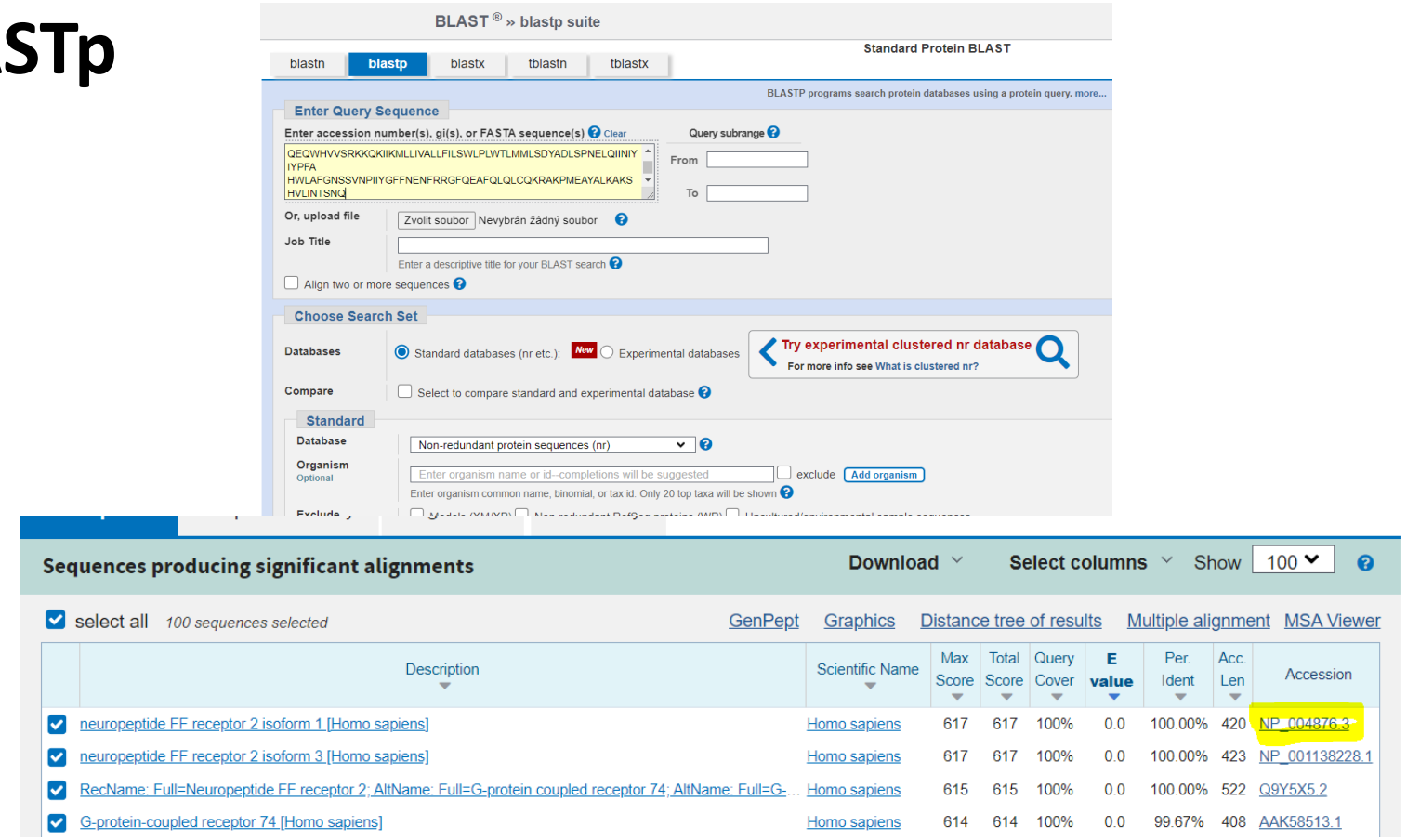
3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky?

4) Kde je tento protein v buňce lokalizován?

řešení

1) Identifikujte příslušný protein, zapište přístupový kód referenční sekvence.

→BLASTp

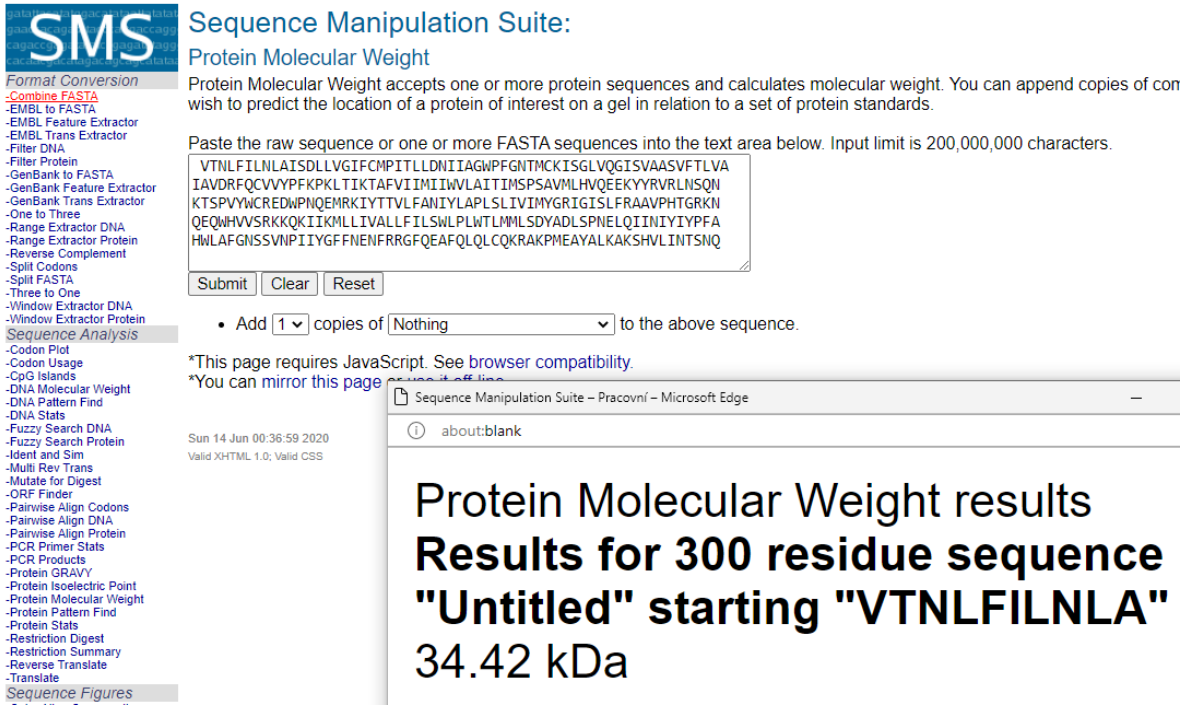


řešení

2) Jaká je molekulová hmotnost tohoto peptidu?

→ SMS Suite: Protein Molecular Weight

2) 34,42kDa

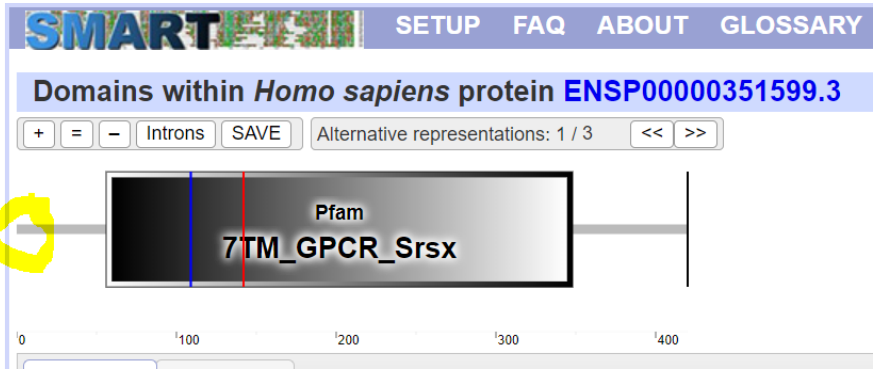


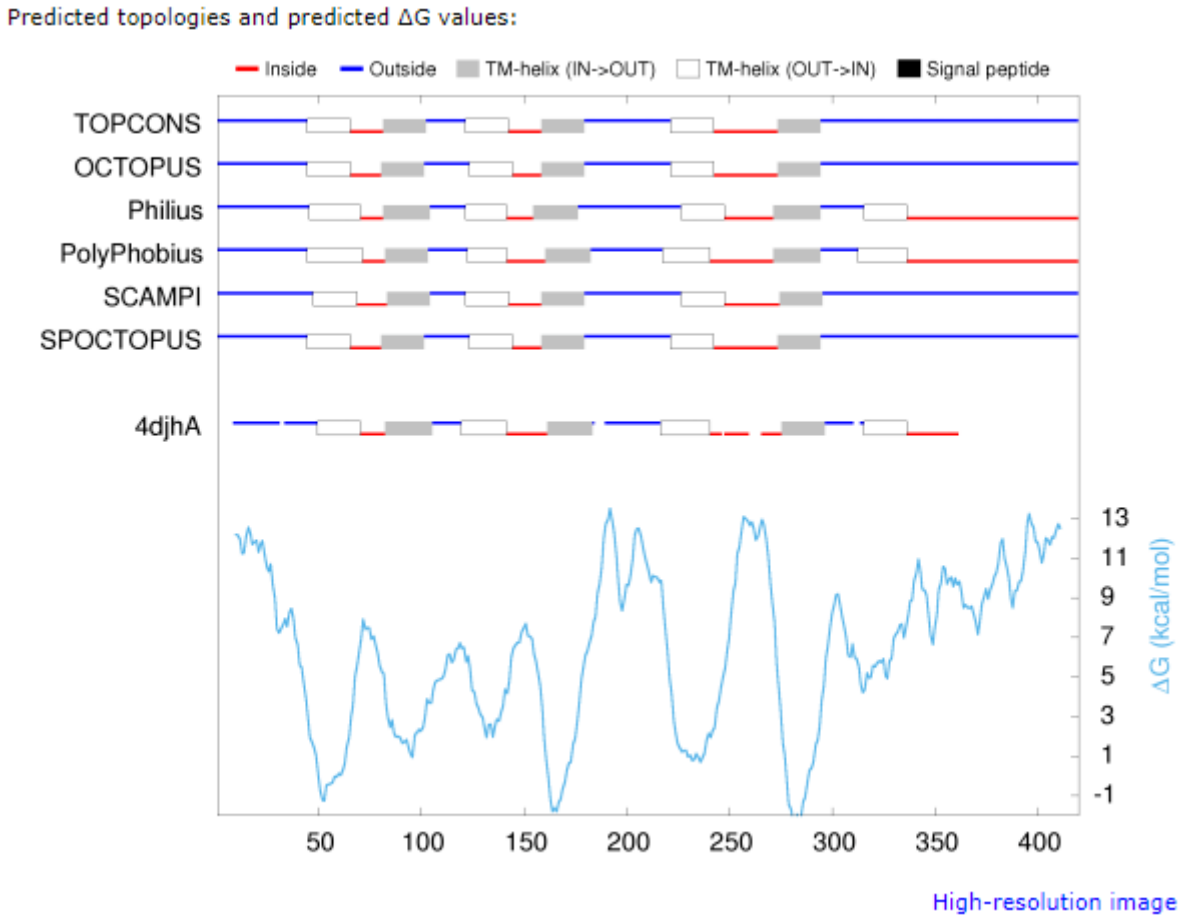
řešení

3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky?

→ různé programy (SignalP, TMHMM, topcons…)

3) ne, ano (6xTM)





řešení

4) Kde je tento protein v buňce lokalizován?

→ UNIPROT (Q9Y5X5 · NPFF2_HUMAN)

4) Na membráně buňky