

# Základy praktické bioinformatiky

---

PETRA MATOUŠKOVÁ

2022/2023

4/10

# DÚ3-lidské isoformy

## - z „refseq\_protein“ databáze

BLAST IP programs search protein databases using a protein query. [more...](#)

**Enter Query Sequence**

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

Q13153      From

BLAST® » blastp suite » results for RID-2ZATDA4H013      Home   Recent Results   Saved Strategies   Help

Or, up [← Edit Search](#)   Save Search   Search Summary ▾      [?](#) How to read this report?   [▶ BLAST Help Videos](#)   [↶ Back to Traditional Results Page](#)

Job T **i** Your search is limited to records th

All

**Ch**

**Datat**

**S**

**Database**

**Organism**  
Optional

**Exclude**  
Optional

**Compare**

**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"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 2 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1038	1038	100%	0.0	100.00%	545	<a href="#">NP_001363202.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 3 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1031	1031	100%	0.0	98.73%	552	<a href="#">NP_001363201.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 4 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	984	984	95%	0.0	100.00%	522	<a href="#">NP_001363217.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 1 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	981	981	95%	0.0	99.04%	553	<a href="#">NP_001122092.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 6 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	929	929	100%	0.0	92.48%	504	<a href="#">NP_001363219.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 7 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	921	921	100%	0.0	91.56%	499	<a href="#">NP_001363221.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 8 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	909	909	100%	0.0	90.83%	496	<a href="#">NP_001363223.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 5 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	886	886	87%	0.0	99.37%	507	<a href="#">NP_001363218.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 13 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	863	863	82%	0.0	100.00%	447	<a href="#">NP_001363233.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 3 isoform a [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	830	830	99%	0.0	79.71%	544	<a href="#">NP_001121638.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 1 isoform 9 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	825	825	100%	0.0	84.77%	486	<a href="#">NP_001363224.1</a>
<input checked="" type="checkbox"/>	<a href="#">serine/threonine-protein kinase PAK 3 isoform d [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	820	820	99%	0.0	77.60%	559	<a href="#">NP_001121645.1</a>

**Job Title**    **i** Your search is limited to records th

**Job Title**    Q13153:RecName: Fu

**RID**    [2ZATDA4H013](#)   [Search](#)

**Program**    [BLASTP](#) [?](#)   [Citation](#)

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

**Query ID**    [Q13153.2](#)

**Description**    RecName: Full=Serine

**Molecule type**    amino acid

**Query Length**    545

**Other reports**    [Distance tree of results](#)

**Database**    [Non-redunda](#)

**Organism**    [Non-redund](#)

**Optional**    [RefSeq Sele](#)

**Exclude**    [Reference pr](#)

**Optional**    [Model Organ](#)

**Compare**    [UniProtKB/S](#)

   [Patented pro](#)

   [Protein Data](#)

   [Metagenomic](#)

   [Transcriptom](#)

# 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	MIESSTTIQVISAGLPRGTGKSLKNALEIIYHKPCYHMFEEIFNKQSDIIKWQNLIHDSH	60
V9PWX8	V9PWX8_SCHMA	1	MIESSTTIQVISAGLPRGTGKSLKNALEIIYHKP YHMFEEIFNKQSDIIKWQNLIHDSH	60
V9PWX7	V9PWX7_SCHMA	61	MITTPPLLTTKTIAIYDKLKELDGYIATDLPFCGYKDLNMIYPNKVVLLTIRDKYDW	120
V9PWX8	V9PWX8_SCHMA	61	MITTPPLLTTKTIAIYDKLKELDGYIATDLPFCGYKDLNMIYPNKVVLLTIRDKYDW	120
V9PWX7	V9PWX7_SCHMA	121	LHSLRKVVLPKSNDFWKLKIEEGDKVGLNSDFYKLTEDSLKFAFQKDDLNFDDQVLE	180
V9PWX8	V9PWX8_SCHMA	121	LHSLRKVVLPKSNDFWKLKIEEGDKVGLNSDFYKLTEDSLKFAFQKDDLNFDDQVLE	180
V9PWX7	V9PWX7_SCHMA	181	CYDEYNRLVQETVPSDRLLVLRGQWEPKFLNVEIPNGIDYPCVNSHHQMTQLTEQL	240
V9PWX8	V9PWX8_SCHMA	181	CYDEYNRLVQETVPSDRLLVLRGQWEPKFLNVEIPNGIDYPCVNSHHQMTQLTEQL	240
V9PWX7	V9PWX7_SCHMA	241	IKYKSLDAI IHMFPDLI	257
V9PWX8	V9PWX8_SCHMA	241	IKYKSLDAI IHMFPDLI	257

V9PWX7	V9PWX7_SCHMA	1	MIESSTTIQVISAGLPRGTGKSLKNALEIIYHKPCYHMFEEIFNKQSDIIKWQNLIHDSH	60
A0A183QDM9	A0A183QDM9_9TREM1		M ESS + VI AGLPRGTGKSLKNALEIIYHKPCYHM EII + +DI KWQ L ++	60
V9PWX7	V9PWX7_SCHMA	61	MITTPPLLTTKTIAIYDKLKELDGYIATDLPFCGYKDLNMIYPNKVVLLTIRDKYDW	120
A0A183QDM9	A0A183QDM9_9TREM1		+ T + I D LKE+L Y A TD+P CGFYK+LMNIYPNKVVLLTIRDKYDW	115
V9PWX7	V9PWX7_SCHMA	121	LHSLRKVVLPKSNDFWKLKIEEGDKV-----	146
A0A183QDM9	A0A183QDM9_9TREM1		LHSLRKVVLPKSNDFWKLKIEEGDKV	175
V9PWX7	V9PWX7_SCHMA	147	--LGLNSDFYKLTEDSLKFAFQKDDLNFDDQVLECYDEYNRLVQETVPSDRLLVLRG	204
A0A183QDM9	A0A183QDM9_9TREM1		L L F K+ DS+K AF+K D + D+D +L+C+DEYNR V ETVPS+RLL+ +LG	235
V9PWX7	V9PWX7_SCHMA	205	DGWEPLCKFLNVEIPNGIDYPCVN	228
A0A183QDM9	A0A183QDM9_9TREM2		DGWEPLCFLNVDVPEGVSYPIYIN	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 EBIs [Wise2DBA](#) and [Promoterwise](#) services are retired on 15th April 2018. Alternatives can be found at [Wise2DBA](#) 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 FORM
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  MVSPATRKSLPKVKAMDFITSTAILPLFLGCLGVFGLFRLLQWVR---GK      47
AAI47025.1      1  -----MAMLMPLPLL--LGISGLLFYQEVSRSLWSK      29
AAH09679.1     48  AYLRNAVVVITGATSGLGKCAKVFYAAGAKLVLCGRNGGALEELIRELT      97
AAI47025.1     30  SAVQNKVVVITDAISGLGKECARVFTGGARLVLCGKNWERLENLYDAL-      78
AAH09679.1     98  ASHATKVQTHKPYLVTFDLTDSGAIVAAAAEILQCFGYVDILVNNAGISY     147
AAI47025.1     79  ISVADPSKFTFKLVLLDSDISCVDPVAKEVLDYCGVCDIILNNASVKV     128
AAH09679.1    148  RGTIMDITVDVDKRVMETNYFGPVALTKALLPSMIKRRQGHIVAISSIQG     197
AAI47025.1    129  KGPAAHKISLELDKIMDANYFGPITLTALKLPNMISSRTGQIVLVNNIQG     178
AAH09679.1    198  KMSIPFRSAYAASKHATQAFDCLRAEMEQQVEIEVTVISPGYIHTNLSVN     247
AAI47025.1    179  KFGIPFRITTYAASKHAALGFFDCLRAEVEEYDVVISTVSPTFIR---SYH     225
AAH09679.1    248  AITADGS-----RYGVMDTTTAQGRSPVEVAQDVLAAVGGKKK     284
AAI47025.1    226  VYPEQGWNEASIKWFFFRKLTYGV-----HPVEVAEEVMRTVRRKK     266
AAH09679.1    285  KDVILADLLPSLAVYLRTILAPGLFFSILMASRARKERKSKNS-     325
AAI47025.1    267  QEVFMANPIKRAAVYVRIFFPEFFFAVVACGVKEKLNVPEEG     308
```



# 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 isoform, porovnání vícero výstupů z BLASTu..)

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

Umožňuje fylogenetickou analýzu (fylogenetické stromy - evoluční příbuznost)

```
Q5E940_BOVIN -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_HUMAN -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_MOUSE -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_RAT -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_CHICK -----MPREDRATWKSNYFMKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_RANSY -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
Q7ZUG3_BRARE -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0 ICTPU -----MPREDRATWKSNYFLKIIQLDDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_DROME -----MVENKAAWKAQYFIKYYLFDVFKCFIVGADNVGKQMDQIRMSLRGK-AVVLGCKNTMMRKAIRGHLNN--PALE 76
RLA0_DICDI -----MSAG-SKREKLFEEKATKLFITTDKMIYAEADVFG-SQLKIRKSTIRGI-GAVLMCKNTMIRKVIHDLADSK--FELD 75
Q54LP0_DICDI -----MSAG-SKRENVFEEKATKLFITTDKMIYAEADVFG-SQLKIRKSTIRGI-GAVLMCKNTMIRKVIHDLADSK--FELD 75
RLA0_PLAFL -----MAKLSKQKQKQMYEEKLSSLIQQYSKILIVHVDVWGMMASVYKSLRGK-AVVLGCKNTMIRKVIHDLADSK--FELD 76
RLA0_SULAC -----MIGLAVYTTTKIAKWKYDEVAELT-KLTKTKTIIIAHIEGFPADKLEIRKSLRGK-ADIKVYKNTLPMIAKKNAG--FDK 79
RLA0_SULTO -----MRIMAVITQERKIAKWKIEEYKLEKLRDHTIIAHIEGFPADKLDIRKMMRGG-AEIKVYKNTLPMIAKKNAG--LDVS 80
RLA0_SULSO -----MKRLALALKQRKVASWKLLEEVKLETELKNSNTILIGHLEGFADKLEIRKSLRGK-AEIKVYKNTLPMIAKKNAG--IDTE 80
RLA0_AERPE MSVVSIVGOMYKREKIPENKTLMLRELELFSKRRVFLADLTGTFVVDVYRKKLWKK-VMMVAKKRITLHAKKAGLE--LDDN 86
RLA0_PYRAE -MHLAIGKRRYVRTQYFARKVKIYSEATELLOKVFYVFLFDLHLSRILHEVRYRLRRY-GVIKIIPFLFKIAFTKVYGG--IPAE 85
RLA0_METAC -----MAEERHHTEHFQKKDEIENKELIQSHKVFQMVRIEGILATKMDKIRRDLDV-AVLKVRNTLSEHALWQLG--ETIP 78
RLA0_METMA -----MAEERHHTEHFQKKDEIENKELIQSHKVFQMVRIEGILATKMDKIRRDLDV-AVLKVRNTLSEHALWQLG--ESIF 78
RLA0_ARCFU -----MAAVRGS--PPEKYVRAVEIKRMISSEVVAIVSFRNVFAGMDKIRREFRGK-AEIKVYKNTLSEHALDALG--GDYL 75
RLA0_METKA HAVKAGQPPSGYE-PKVAEKKRREYKELKLMDEYENYGLVDLEGIPAPQLOEIRAKLRERD-IIRMRHTLMRVALEEKIDER--PELE 88
RLA0_METH -----MAHVAEKKKKEVQELHDLIKRYEYVGIANLADIPAPQLOEKMRQTLRDS-ALIRMRHTLISLALAKKREEL--ENVY 74
RLA0_METTL -----MITAESHKIAFHKIEEVNKLKLLKNGQIVAVDMMYVPAQLOEIRDKIR-ETMELKMRHTLIEHAKIYVALETGNPEFA 82
RLA0_METVA -----MIDAKSEHKIAFHKIEEVNKLKLLKSNVIALDMMYVPAQLOEIRDKIR-DQMLKMRHTLIEHAKIYVALETGNPEFA 82
RLA0_METJA -----METKYKAVYAEKIEEVKTIKGLIKSKYVYVAVDMMYVPAQLOEIRDKIR-DKFKLHMRHTLIRALIEHAKIYVALETGNPEFA 81
```



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

# Alignment: MultAlin

## MultAlin

Multiple sequence alignment by Florence Corpet

Published research using this software should cite:  
"Multiple sequence alignment with hierarchical clustering"  
F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890



### Sequence data

Cut and paste your sequences here below. ?

```
>gi|13435426|gb|AAH04579.1| Nqo1 protein [Mus musculus]
>gi|71059897|emb|CAJ18492.1| Nqo1 [Mus musculus]
MAARRALIVLAHSERTSPNYAMKRAAVALKRRGWVLESPLYAMNPNFIISRNDITGELKDSKNFQYPS
ESLAKKEGR
LSPDIVAEHKKLEAADLVIPQPLQWPGVPAILKGFVFLVAGFAYTYAAMYDNGFPQMKKILLISITG
GSGSMYSLQG
VHGDMVILNFIQSGILRFGFQVLEPOLVYSIGHTFPDARMQILEGWKKRLETVWEETPLVFAPSSLFD
LNFQAGFLMK
KEVQEEQKKNKFGLSVGHHLGKSI PADNQIKARK
>gi|524939198|ref|XP_005071892.1| PREDICTED: NAD(P)H dehydrogenase
```

or select a file: Procházet...

Sequence input format: Auto

For nucleotidic sequences, you must change the Symbol comparison Table (see below) ?

Start MultAlin! Clear Entire Form

## Optional Parameters

### Result page format:

The sequence alignment will be displayed as a coloured image

Vyhledávání s různými parametry:  
substituční matice: PAM/BLOSUM

# MultAlin

## Multalin result page



[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. CORFET, 1988, Nucl. Acids Res., 16 (22), 10881-10890  
Symbol comparison table: biosum62  
Gap weight: 12  
Gap length weight: 2  
Consensus levels: high=90% low=50%  
Consensus symbols:  
! is anyone of IV  
\$ is anyone of IM  
\* is anyone of FV  
# is anyone of NDQEBZ  
  
MSF: 274 Check: 0  
Name: gi1134354261gb|IABM04 Len: 274 Check: 4705 Weight: 1.23  
Name: gi15249391981ref|XP\_ Len: 274 Check: 6867 Weight: 1.23  
Name: gi12274304031ref|NP\_ Len: 274 Check: 6661 Weight: 0.89  
Name: gi14262425031ref|XP\_ Len: 274 Check: 6108 Weight: 0.89  
Name: gi13867817031ref|NP\_ Len: 274 Check: 4019 Weight: 0.89  
Name: gi1302306851gb|IABP20 Len: 274 Check: 4190 Weight: 0.89  
Name: Consensus Len: 274 Check: 4506 Weight: 0.00

```
//
1 10 20 30 40 50 60 70 80 90 100 110 120 130
g1134354261gb|IABM04 HARRRRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVLESOLYAHNFPISRNDDTTELKDSKNQYPSSESLAKREGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA
g15249391981ref|XP_ HARRRRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVLESOLYAHNFPVTSRNDITLKRKESNQTLLSTLWYKESGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA
g12274304031ref|NP_ HAYPKRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVWYSDLYAHNFPVTSRNDITLKRKESNQTLLSTLWYKESGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA
g14262425031ref|XP_ HAYPKRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVVSOLYAHNFPVTSRNDITLKRKESNQTLLSTLWYKESGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA
g13867817031ref|NP_ HAYPKRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVLESOLYAHNFPVTSRNDITLKRKESNQTLLSTLWYKESGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA
g1302306851gb|IABP20 HARRRRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVLESOLYAHNFPISRNDDTTELKDSKNQYPSSESLAKREGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA
Consensus
narrfRLIVLASEKTSFYNAHKEHAIIVELKRRGAEVLESOLYAHNFPISRNDDTTELKDSKNQYPSSESLAKREGALSPDIYVHAKLEARDLVTFQPLQAFGVPALLKGAERYLVGFRITTYA

131 140 150 160 170 180 190 200 210 220 230 240 250 260
g1134354261gb|IABM04 HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL
g15249391981ref|XP_ HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL
g12274304031ref|NP_ HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL
g14262425031ref|XP_ HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL
g13867817031ref|NP_ HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL
g1302306851gb|IABP20 HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL
Consensus
HAYTKGPFRRKQVLSITLGGSSGSHSLQIIGAGRWLLNPITQSGLLNF CGFQVLEPQLYSIGIHPRRKQIILEGAKRLEIADLPLIYFAPSSLFALNFQAGFLMKKQVQDEKSRKFLGSVGHIL

261 270 274
g1134354261gb|IABM04 GKSIPRNDIKRKK
g15249391981ref|XP_ GKSIPRDSIDRKR
g12274304031ref|NP_ GKSIPRNDIKRKK
g14262425031ref|XP_ GKSIPRNDIKRKK
g13867817031ref|NP_ GKSIPRNDIKRKK
g1302306851gb|IABP20 GKSIPRNDIKRKK
Consensus
GKSIPRNDIKRKK
```

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 (frd) 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



## COBALT

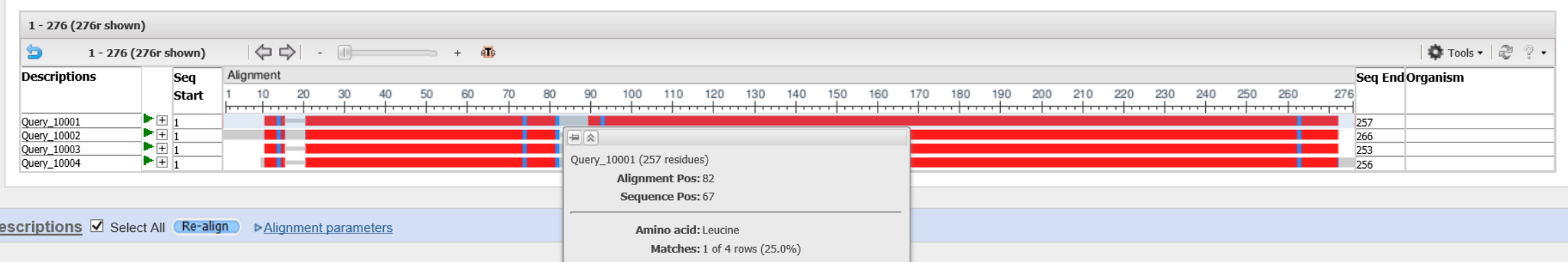
### Constraint-based Multiple Alignment Tool

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

[Phylogenetic Tree](#) [Edit and Resubmit](#) [Download](#)

### - Cobalt RID DK0EGH2J212 (4 seqs)

#### Graphical Overview



**Descriptions**  Select All [Re-align](#) [Alignment parameters](#)

<input checked="" type="checkbox"/>	Query_10001	1	-MIESS	TTIQVISAGLPRTGTKSLKNALEIIYHKPCYHMFEEIFNKQSDIIKWQLIHDSHMITTPPLL[4]IA	74
<input checked="" type="checkbox"/>	Query_10002	1	[9]MMETS[5]	TTIQVIGAGLPRTGTNSMKKALEIISKPCYHMYEIIFFKKQSDISIWQQLIDEHKTTSDKR-	-K 83
<input checked="" type="checkbox"/>	Query_10003	1	-MSQLQ	TSLTVIGAGLPRTGTLISMKKALETIYCQPCYHMYEIIILNKQYDISKWQTLDDIKQSKTTSNEI	LI 70
<input checked="" type="checkbox"/>	Query_10004	1	MMSDNS	TSLLVIGAGLPRTGTTSMKRALEILLGKPCYHMMDIMLRKHEDIGKWLQLIDEVNKTSRNEVI	-- 69
<input checked="" type="checkbox"/>	Query_10001	75	IYDKLKELLDGYIATTDLPFCGYFDLNMNIYPNAKVLLTIRDKYDWLHSLRQVLPKSNDPWKLKIEEGDKVLGGLNSDFY	154	
<input checked="" type="checkbox"/>	Query_10002	84	IYNGLNELLNGYIATTDLPSCSFYKELMTMYPNKAVLLTIRDKYDWLHSLRQVLPKSTDPWKLKIEEGDQVLGIDSNFY	163	
<input checked="" type="checkbox"/>	Query_10003	71	IQNSLKEILNGYIAVTDLPACGFYRELMTMYPNKAVILTIRDRNDWLTSFRKVVLPRTNDTYKEEVDKVNRIILGLNTEFD	150	
<input checked="" type="checkbox"/>	Query_10004	70	IHDILSEILTGYASVTDIPTCGFYRELNMVYPNAKVILTIRDKTDWLSLRHTVMPKCCDPQKQIMEEAMNVIGYSVEID	149	

Links
6 PE=4 SV=1
2 SV=1

# Alignment: Clustal Omega

## Clustal Omega

[Input form](#)[Web services](#)[Help & Documentation](#)[Feedback](#)[Share](#)

[Tools](#) > [Multiple Sequence Alignment](#) > [Clustal Omega](#)

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

### STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

Submit

Or, [upload a file:](#)

[Procházet...](#)

# Alignment: Clustal Omega

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

Alignments

Download Alignments

CLUSTAL O(1.2.4) multiple

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

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

(\*) konzervovány  
(:) aminokyseliny  
(.) aminokyseliny s podobnou velikostí NEBO hydrofobicitou

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

Alignments

Download Alignments

CLUSTAL O(1.2.4) multiple

Mus MAARRALIVLAI  
Rattus MAVRRALIVLAI  
Homo MVGRRALIVLAI  
Bos MAVRKALIVLAI  
Sus MAVRKALIIILAI  
\* . \*: \*\*

Mus KDSKNFQYPSEI  
Rattus KDSENFQYPVEI  
Homo KDPANFQYPAEI  
Bos KDPGNFQYPAEI  
Sus KDPGNFQYPAEI  
\*\* \*\*\*\*\* \*

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

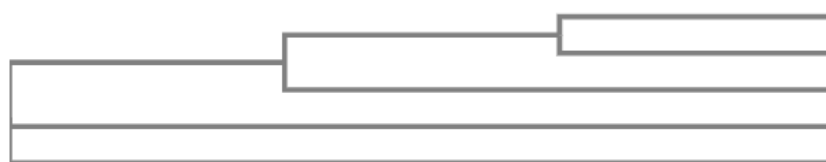
Alignments Result Summary Guide Tree Phylogenetic Tree Results Viewers Submission Details

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

# Vyzkoušejte si porovnat sekvence uložené v minulém úkolu

---

Vyzkoušejte víc programů!

# „pokročilejší“ fylogenetické porovnávání

Méthodes et Algorithmes pour la Bio-informatique LIRMM

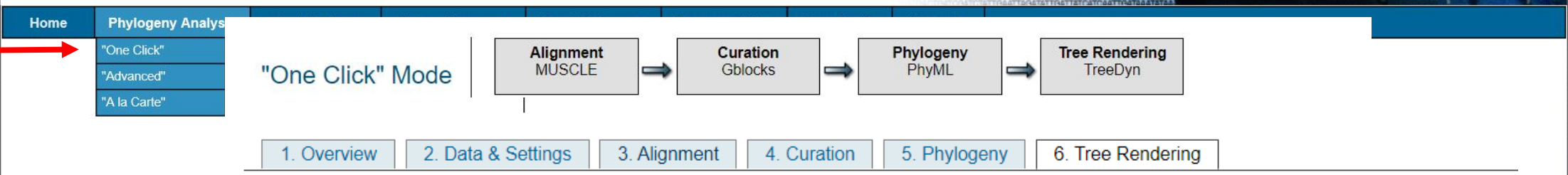
Information Genomique et Structurale

Home Phylogeny Analysis Blast Explorer Online Programs Your Workspace Documentation Downloads Contacts

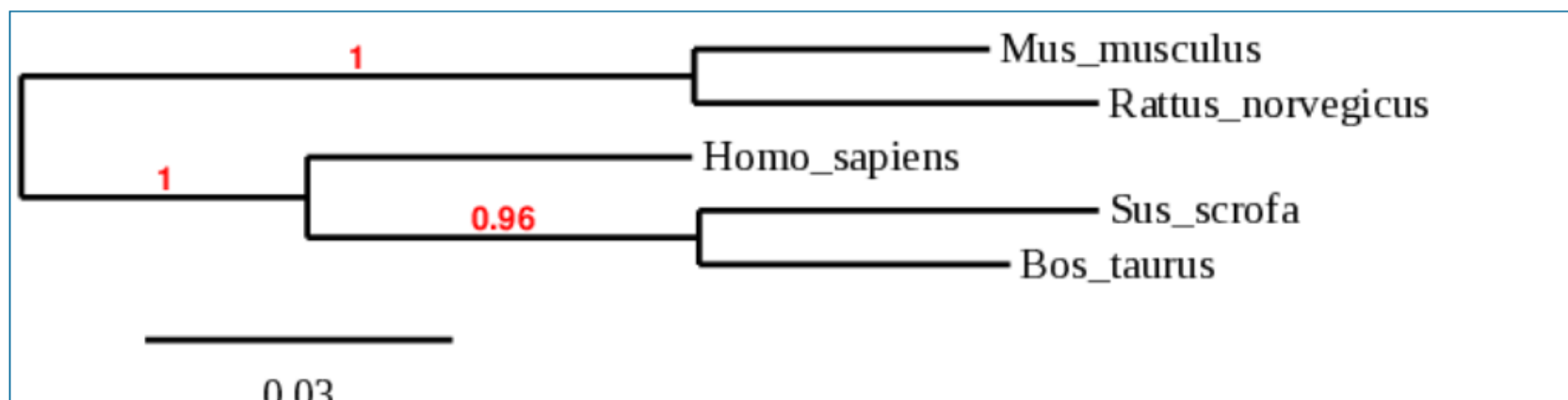
**Phylogeny.fr**  
Robust Phylogenetic Analysis For The Non-Specialist



# „pokročilejší“ fylogenetické porovnávání



## Tree Rendering results




# 3-D struktura proteinů: PDB


RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

**RCSB PDB** An Information Portal to 128330 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations





14 Structures 2 Unreleased Structures 10 Citations 12 Ligands

## Search Parameter:

Refine Search Save Search to MyPDB

Text Search for: nqo1 and TAXONOMY is just Homo sapiens (human)

## Refinements



Currently showing 1 - 14 of 14

Displaying 25 Results

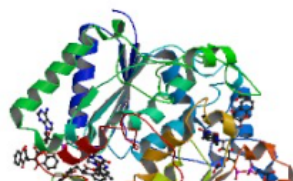
ORGANISM  
Homo sapiens only (14)

UNIPROT MOLECULE NAME  
NAD(P)H dehydrogenase [qu ... (13)  
Ribosyldihyronicotinamid ... (1)  
[Refine Query](#)

TAXONOMY

View: Detailed Reports: Select a Report Sort: Release Date: Newest to Oldest Download Files

UKÁZKA



5FUQ

CRYSTAL STRUCTURE OF THE H80R VARIANT OF NQO1 BOUND TO DICOUMAROL

[Medina-Carmona, E.](#), [Fuchs, J.E.](#), [Gavira, J.A.](#), [Salido, E.](#), [Palomino-Morales, R.](#), [Mesa-Torres, N.](#), [Timson, D.L.](#), [Ray, A.J.](#)

Download File View File

# 3-D struktura proteinů: PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PROTEIN

Structure Summary **3D View** Annotations Sequence Sequence Similarity Structure Similarity Experiment

PDB-101

## 2F1O

Crystal Structure of NQO1 with Dicoumarol

Display Files Download Files

Structure View Electron Density Maps Ligand View

Structure View Documentation

Assembly Bioassembly 1

Model Model 1

Symmetry None

Style **Spacefill**

Color By Secondary S

Ligand Ball & Stick

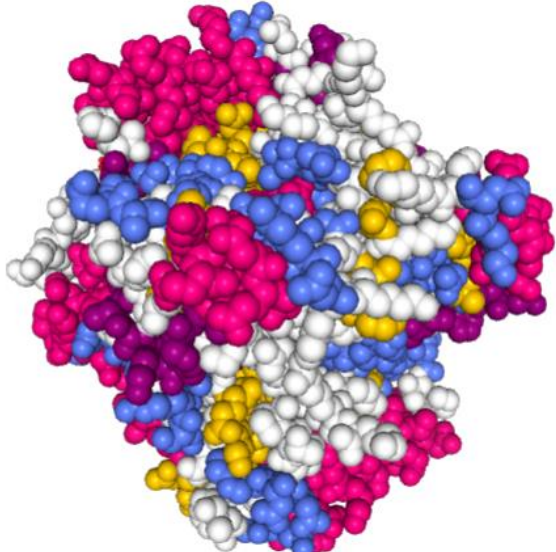
Quality Automatic

Water  Ions

Hydrogens  Clashes

Default Structure View

Note: Use your mouse to drag, rotate, and zoom in and out of the structure. Mouse-over to identify atoms and bonds. [Mouse controls documentation](#).

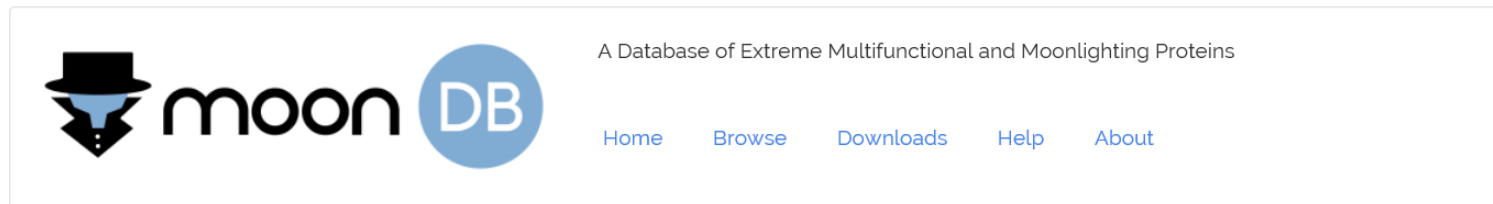


# Vyzkoušejte si PDB databázi

---

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

# Multifunkční proteiny: moondb

[Search Protein](#)[Browse MoonDB](#)

## Welcome to MoonDB v2.0

A Database of Extreme Multifunctional and Moonlighting Proteins

MoonDB is a database containing predicted **Extreme Multifunctional (EMF) proteins** (i.e. proteins with several unrelated functions), as well as a set of manually curated moonlighting proteins. **Moonlighting proteins** are a subclass of multifunctional proteins.

EMF proteins were detected through the [MoonGO pipeline](#), which combines network topological information and protein annotations. This approach is described on:

Ribeiro D.M., Briere G., Bely B., Spinelli L., Brun C. (2018) "[MoonDB 2.0: an updated database of extreme multifunctional and moonlighting proteins](#)" *Nucleic Acids Research*, <https://doi.org/10.1093/nar/gky1039>

# Fosforylace proteinů: PhosphoSite



Powered by Cell Signaling Technology

[Feedback](#) | [Contact](#) | [Login](#)

[Home](#) [Search](#) [Browse](#) [Tools](#) [Downloads](#) [About](#)

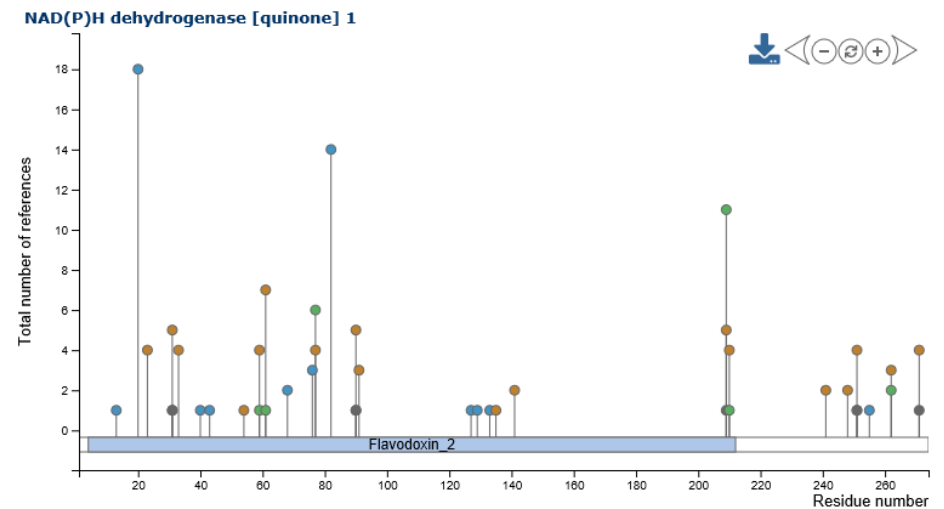
PhosphoSitePlus® provides comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, acetylation, and more. The web use is free for everyone including commercial.

### 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 datbáze: Brenda

## EC Browser

- 1 Oxidoreductases (9651 organisms)
- 2 Transferases (6622 organisms)
- 3 Hydrolases (10604 organisms)
- 4 Lyases (5111 organisms)
- 5 Isomerases (2083 organisms)
- 6 Ligases (1547 organisms)
- 7 Translocases (966 organisms)

go to...  
HOME  
Classic view  
1987-2019  
The Comprehensive  
BRENDA  
Enzymes (900 organisms)

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
  - $A - B + H_2O \rightarrow AOH + HB$
  - **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)

## EC class 7

These enzymes catalyse the movement of ions or molecules across membranes or their separation within membranes, the reaction is designated as a transfer from side 1 to side 2 because the designations in and out, which had previously been used, can be ambiguous. The subclasses designate the types of components transferred and the sub-sub-classes indicate the reaction processes that provide the driving force for the translocation.

contains  
delete search field start search

## 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




Version: 11.0 LOGIN | REGISTER

 Search Download Help My Data

There are several matches for 'NQO1'.  
Please select one from the list below and press Continue to proceed. [<- BACK](#) [CONTINUE ->](#)

organism	protein
<input checked="" type="checkbox"/> <b>Homo sapiens</b>	<b>NQO1</b> - 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
<input type="checkbox"/> <b>Homo sapiens</b>	TCF7L1 - Transcription factor 7-like 1; Participates in the Wnt signaling pathway. Binds to DNA and acts as a repressor in the absence of CTNNB1, and as an activator in its presence. Necessary for the terminal differentiation of epidermal cells, the formation of keratohyalin granules and the development of the barrier function of the epidermis (By similarity). Down-regulates <b>NQO1</b> , leading to increased mitomycin c resistance; TCF/LEF transcription factor family [a.k.a. <i>TCF3</i> , <i>Hs.516297</i> , <i>transcription factor 7 like 1</i> ]

---

© STRING CONSORTIUM 2020	<b>ABOUT</b>	<b>INFO</b>	<b>ACCESS</b>	<b>CREDITS</b>
 SIB - Swiss Institute of Bioinformatics	Content	Scores	Versions	Funding
 CPR - Novo Nordisk Foundation Center Protein Research	References	Use scenarios	APIs	Datasources
 EMBL - European Molecular Biology Laboratory	Contributors	FAQs	Licensing	Partners
	Statistics	Cookies/Privacy	Usage	Software

---

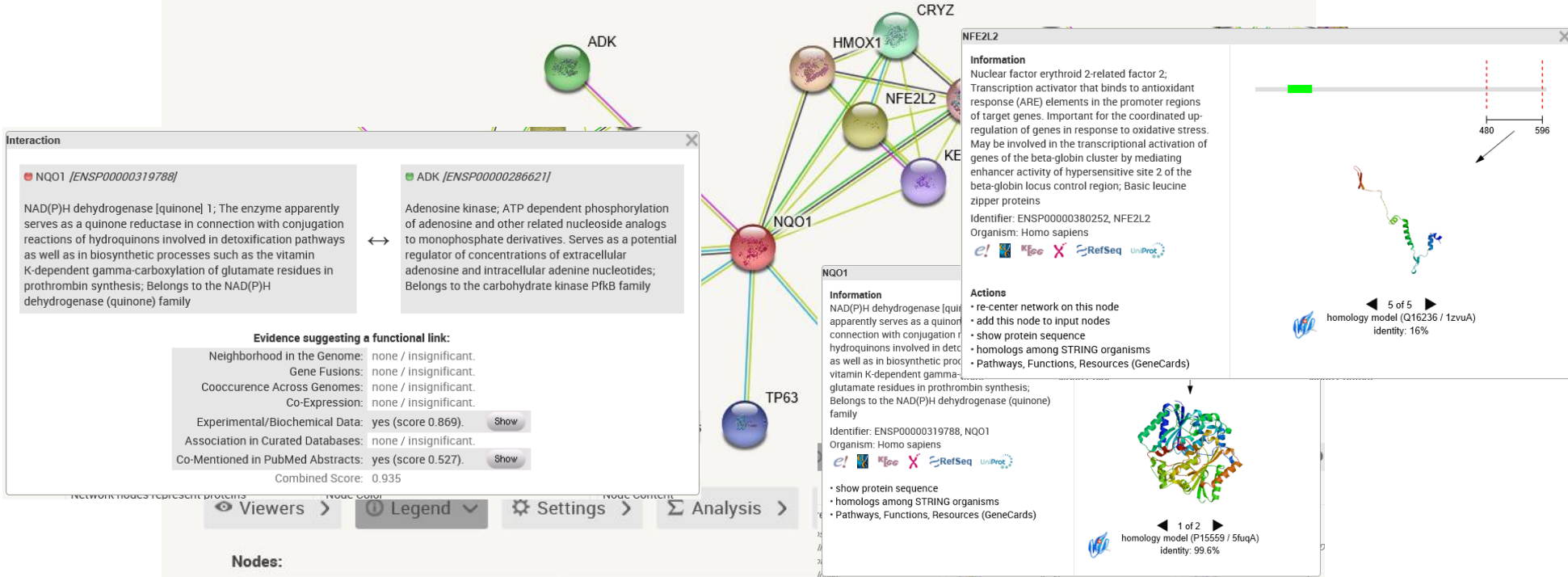
<input type="checkbox"/> <b>Balaenoptera acutorostrata</b>	<b>NQO1</b> - NAD(P)H dehydrogenase [quinone] 1
--	---



# Interakce proteinů



Search Download Help My Data



**Nodes:**

Network nodes represent proteins

*splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.*

**Node Color**

colored nodes: query proteins and first shell of interactors

white nodes: second shell of interactors

**Node Content**

empty nodes: proteins of unknown 3D structure

filled nodes: some 3D structure is known or predicted

# 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 ukázkou.
- 5) Je váš protein enzym? Jaké má enzymové číslo (EC)?

# Příklad řešení

DÚ4

```
>>sp|Q64669|MQO1_MOUSE_NAD(P)H_dehydrogenase_[quinone]_1_(274_aa)
Warning: Expect score: 1626; 421.3 bits; E(1) < 1.1e-132
86.5% identity (97.8% similar) in 274 aa overlap (1-274:1-274)
```

i:86,5% Porovni probhlo v celém rozsahu obou proteinů.

1)

```

      10      20      30      40      50      60
sp|F15 MVGRRALIVLAHSERTSIFYAMKEAAAALEKKGWVVESDLYAMNFWPIISRKIDITGKL
      .:.....:.....:.....:.....:.....:.....:
sp|Q64 MAARRALIVLAHSEKTSIFYAMKEAAVERLEKKGWVLESDDLYAMNFWPIISRNDITGEL
      10      20      30      40      50      60

      70      80      90     100     110     120
sp|F15 KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
      : :.....:.....:.....:.....:.....:.....:
sp|Q64 KDSKNFCYPSSESLAYKEGRSPDIVAEHKKLEAADLVIQFPLQWFGVPAILKGFPERV
      70      80      90     100     110     120

      130     140     150     160     170     180
sp|F15 FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
      ... :.....:.....:.....:.....:.....:.....:
sp|Q64 LVAGFAYTYAAMYDNGPFQNKRTLLSITTGSGSGMSYLGQVHGDMNVILWPIQSGILRPF
      130     140     150     160     170     180

      190     200     210     220     230     240
sp|F15 GFQVLEPQLTYSIGHTPADARIQILEGWKKRLEINWDETPLYFAPSSLEFDLNFQAGFLMK
      : :.....:.....:.....:.....:.....:.....:
sp|Q64 GFQVLEPQLVYSIGHTFPDARMQILEGWKKRLETVMEETPLYFAPSSLEFDLNFQAGFLMK
      190     200     210     220     230     240

      250     260     270
sp|F15 KEVQDEEKKKKFGLSVGHHLGKSIPTDNIKARK
      : :.....:.....:.....:.....:.....:.....:
sp|Q64 KEVQDEKKNKFGLSVGHHLGKSIPTDNIKARK
      250     260     270
```

2)

```

      1      10      20      30      40      50      60      70      80      90     100     110     120     130
Macaca MYARRALIVLAHSERTSIFYAMKEAAAALEKKGWVVESDLYAMNFWPIISRKIDITGEL
Sus MYARRALIVLAHSERTSIFYAMKEAAAALEKKGWVVESDLYAMNFWPIISRNDITGEL
Mus MYARRALIVLAHSERTSIFYAMKEAAVERLEKKGWVLESDDLYAMNFWPIISRNDITGEL
Alligator MYARRALIVLAHSERTSIFYAMKEAAVERLEKKGWVLESDDLYAMNFWPIISRNDITGEL
Consensus MYARRALIVLAHSERTSIFYAMKEAAVERLEKKGWVLESDDLYAMNFWPIISRNDITGEL

      131     140     150     160     170     180     190     200     210     220     230     240     250     260
Macaca KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
Sus KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
Mus KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
Alligator KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
Consensus KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV

      261     270     274
Macaca FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
Sus FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
Mus FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
Alligator FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
Consensus FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
```



3)

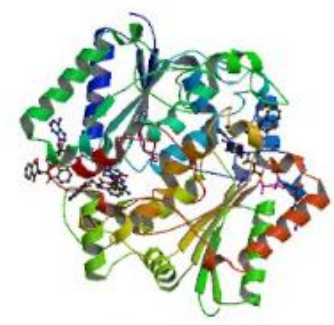
```

      1      10      20      30      40      50      60      70      80      90     100     110     120     130
Isafar2 MYARRALIVLAHSERTSIFYAMKEAAAALEKKGWVVESDLYAMNFWPIISRKIDITGEL
NBD1 MYARRALIVLAHSERTSIFYAMKEAAAALEKKGWVVESDLYAMNFWPIISRNDITGEL
Isafar3 MYARRALIVLAHSERTSIFYAMKEAAVERLEKKGWVLESDDLYAMNFWPIISRNDITGEL
Consensus MYARRALIVLAHSERTSIFYAMKEAAVERLEKKGWVLESDDLYAMNFWPIISRNDITGEL

      131     140     150     160     170     180     190     200     210     220     230     240     250     260
Isafar2 KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
NBD1 KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
Isafar3 KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV
Consensus KDFANFQYPAESVLAAYKRGHLSPDIVAEQKKELAADLVIPQFPLQWFGVPAILKGFPERV

      261     270     274
Isafar2 FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
NBD1 FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
Isafar3 FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
Consensus FIGEFAYTYAAMYDKGPFASKKAVALSITTGSGSGMSYLGQIHGDMNVILWPIQSGILRPF
```


4)



# shrnutí I - Rešeršní projekt: NQO1

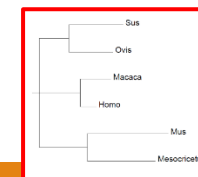
Vyhledejte dostupné informace o NQO1, získejte co nejvíce literárních, sekvenčních, případně i strukturních a dalších údajů o tomto genu/proteinu.

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



```
1 10 20 30 40 50 60 70 80 90 100 110 120 130
MacacaMulatta 1 10 20 30 40 50 60 70 80 90 100 110 120 130
NQO1_Homo 1 10 20 30 40 50 60 70 80 90 100 110 120 130
Sus 1 10 20 30 40 50 60 70 80 90 100 110 120 130
Ovis 1 10 20 30 40 50 60 70 80 90 100 110 120 130
Mus 1 10 20 30 40 50 60 70 80 90 100 110 120 130
Meiocricetus 1 10 20 30 40 50 60 70 80 90 100 110 120 130
Consensus 1 10 20 30 40 50 60 70 80 90 100 110 120 130
131 140 150 160 170 180 190 200 210 220 230 240 250 260
MacacaMulatta 131 140 150 160 170 180 190 200 210 220 230 240 250 260
NQO1_Homo 131 140 150 160 170 180 190 200 210 220 230 240 250 260
Sus 131 140 150 160 170 180 190 200 210 220 230 240 250 260
Ovis 131 140 150 160 170 180 190 200 210 220 230 240 250 260
Mus 131 140 150 160 170 180 190 200 210 220 230 240 250 260
Meiocricetus 131 140 150 160 170 180 190 200 210 220 230 240 250 260
Consensus 131 140 150 160 170 180 190 200 210 220 230 240 250 260
261 270 274
MacacaMulatta 261 270 274
NQO1_Homo 261 270 274
Sus 261 270 274
Ovis 261 270 274
Mus 261 270 274
Meiocricetus 261 270 274
Consensus 261 270 274
```



# Procvičování (proteinová část)

---

Pracujte s následující sekvencí peptidu (Moodle-procvičování)

```
VTNLFILNLAISDLLVGIFCMPITLLDNI IAGWPFNGNTMCKISGLVQGISVAASVFTLVA  
IAVDRFQC VVYPFKPKLTIKTA FVIIMI IWVLAITIMSPSAVMLHVQEEKYYRVRLNSQN  
KTS PVYWCRE DWPNQEMRKIYTTVLFANIYLAPLSLIVIMYGRIGISLFRAAVPHTGRKN  
QEQWHVVSRRKQKI IKMLLIVALLFILSWLPLWTLMLSDYADLSPNELQIINIYIYPFA  
HWLAFGNSSVNP I IYGFFNENFRRGFQEAFLQLCQKRAKPM EAYALKAKSHVLINTSNQ
```

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

# řešení

## →BLASTp

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

The screenshot displays the BLASTp web interface. The top navigation bar includes 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx', with 'blastp' selected. The main section is titled 'Standard Protein BLAST'. Under 'Enter Query Sequence', a text box contains the query sequence: QEQVHVSRKKQKIKMLLIVALLFILSWLPLWTLMLSDYADLSPNELQINIY IYFPA HWLAFGNSSVNPPIYGFNENFRRGFEAFQLQLCQKRAKPMEAAYALKAKS HVLINTSNQ. Below this, there are options to upload a file or enter a job title. The 'Choose Search Set' section shows 'Standard databases' selected, with 'Non-redundant protein sequences (nr)' chosen as the database. The 'Sequences producing significant alignments' table is visible at the bottom, showing four results for Homo sapiens, with the first result highlighted in yellow.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
neuropeptide FF receptor 2 isoform 1 [Homo sapiens]	Homo sapiens	617	617	100%	0.0	100.00%	420	NP_004876.3
neuropeptide FF receptor 2 isoform 3 [Homo sapiens]	Homo sapiens	617	617	100%	0.0	100.00%	423	NP_001138228.1
RecName: Full=Neuropeptide FF receptor 2; AltName: Full=G-protein coupled receptor 74; AltName: Full=G...	Homo sapiens	615	615	100%	0.0	100.00%	522	Q9Y5X5.2
G-protein-coupled receptor 74 [Homo sapiens]	Homo sapiens	614	614	100%	0.0	99.67%	408	AAK58513.1

# řešení

## 2) Jaká je molekulová hmotnost tohoto peptidu?

### → SMS Suite: Protein Molecular Weight

2) 34,42kDa

**SMS** Sequence Manipulation Suite:  
Protein Molecular Weight

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

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 200,000,000 characters.

```
VTNLFILNLAISDLLVGFICMPITLLDNIAGWPPFGNTMCKISGLVQGISVAASVFTLVA  
IAVDRFQCVVYFPKPLTIKTAFAVIMIIWLAITIMSPSAVMLHVQEEKYYRRLNLSQ  
KTSPPVYWCREDWPNQEMRKIYTTVL FANIYLAPLSLIVIMYGRIGISLFRAAVPHTRKN  
QEQWHVYSRKKQKIKMLLIVALLFLSWLPLWTLMLSDYADLSPNELQIINIYIYFPA  
HMLAFGNSSVNP IYGF FNENFRGFQEAFLQLCQKRAKMEAYALKAKSHVLINTSNQ
```

Submit Clear Reset

- Add 1 copies of Nothing to the above sequence.

\*This page requires JavaScript. See [browser compatibility](#).  
\*You can mirror this page [on your own server](#).

Sequence Manipulation Suite - Pracovní - Microsoft Edge  
about:blank

### Protein Molecular Weight results

## Results for 300 residue sequence "Untitled" starting "VTNLFILNLA"

### 34.42 kDa

Sun 14 Jun 00:36:59 2020  
Valid XHTML 1.0; Valid CSS

**Format Conversion**

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Spill FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

**Sequence Analysis**

- Codon Plot
- Codon Usage
- CpG Islands
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim
- Multi Rev Trans
- Mutate for Digest
- ORF Finder
- Pairwise Align Codons
- Pairwise Align DNA
- Pairwise Align Protein
- PCR Primer Stats
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point
- Protein Molecular Weight
- Protein Pattern Find
- Protein Stats
- Restriction Digest
- Restriction Summary
- Reverse Translate
- Translate

**Sequence Figures**

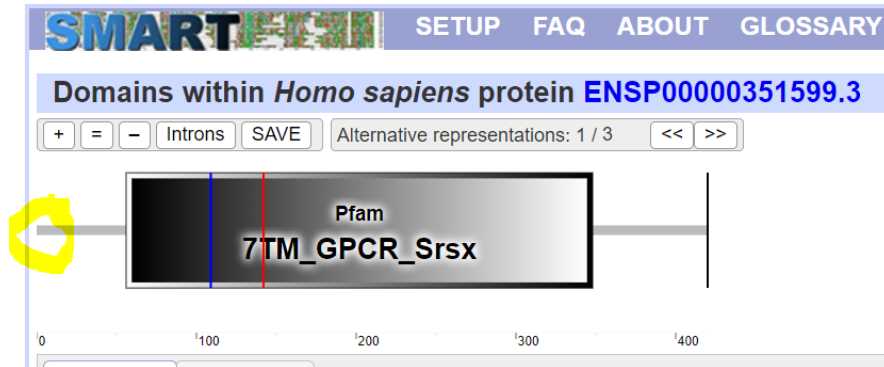


# řešení

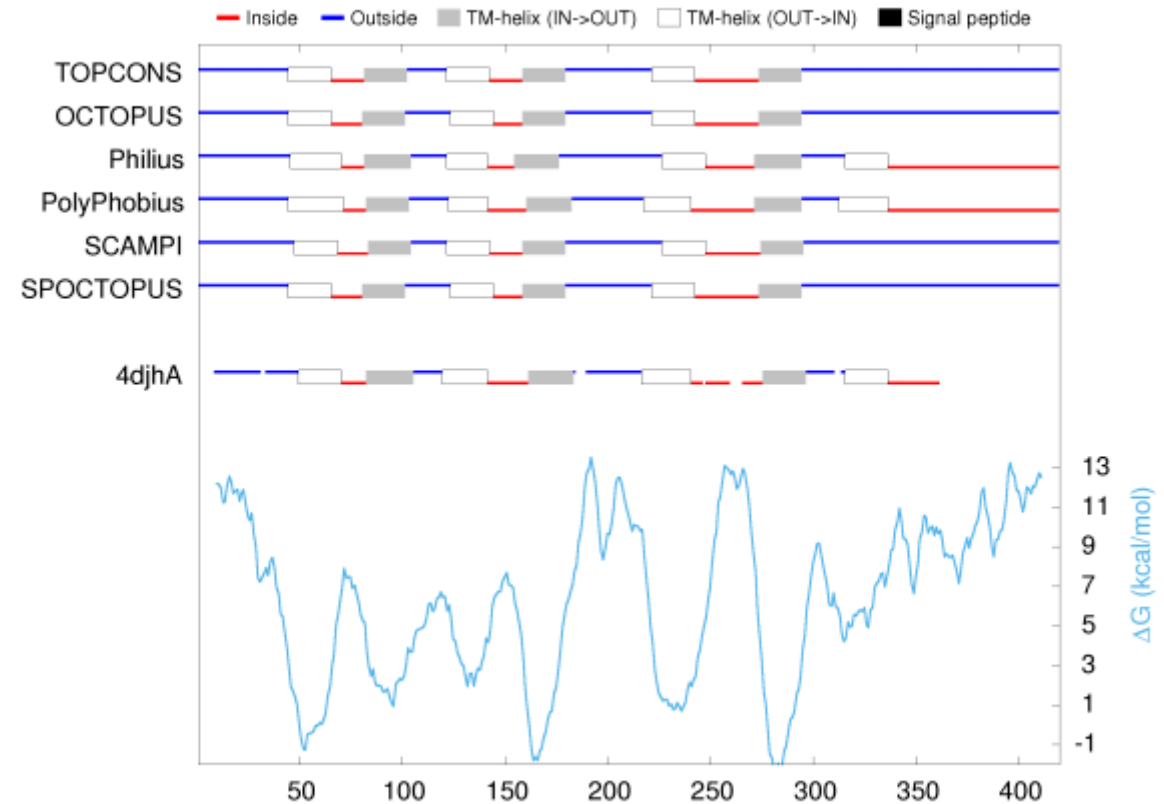
3) Obsahuje **celý identifikovaný protein** signální peptid nebo transmembránové úseky?

→ různé programy (SignalP, TMHMM, topcons...)

3) ne, ano (6xTM)



Predicted topologies and predicted  $\Delta G$  values:



High-resolution image

# řešení

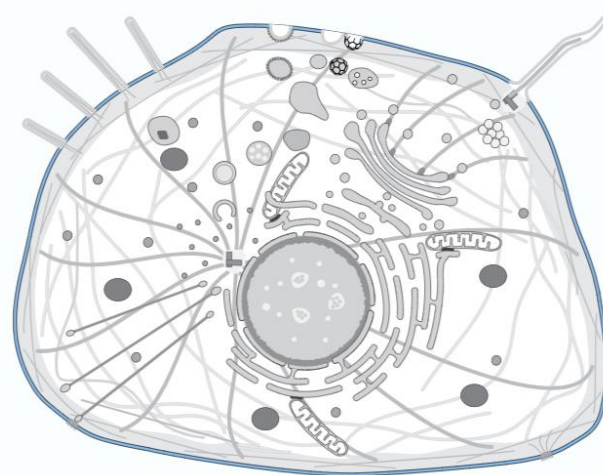
## 4) Kde je tento protein v buňce lokalizován?

→ **UNIPROT (Q9Y5X5 · NPFF2\_HUMAN)**

4) Na membráně buňky

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

UniProt Annotation    GO Annotation



📍 **Cell membrane** ; Multi-pass membrane protein