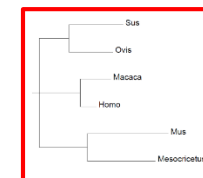
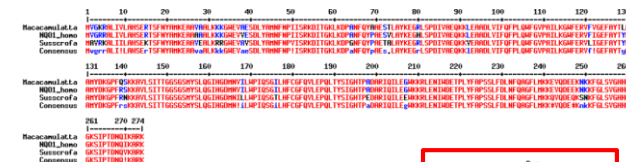


Proteinová bioinformatika I-III

Shrnutí a procvičení

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



Př1: DHRS7

- Vyhledejte v databázi NCBI sekvence těchto dvou lidských enzymů DHRS7B (**AAH09679.1**) a DHRS7C (**AAI47025.1**) a proved'te párové porovnání.
- Jaká je identita těchto dvou proteinů?

Př1: DHRS7

- Vyhledejte v databázi NCBI sekvence těchto dvou lidských enzymů DHRS7B (AAH09679.1) a DHRS7C (AAI47025.1) a proveďte párové porovnání.
- Jaká je identita těchto dvou proteinů?

```
>>AAH09679.1 Dehydrogenase/reductase (SDR family) member (325 aa)
Waterman-Eggert score: 840; 246.0 bits; E(1) < 9e-70
45.0% identity (76.3% similar) in 300 aa overlap (5-301:24-319)
zde je požadovaná informace o identitě sekvencí (u
prvního názevu porovnání)
10 20 30 40 50 60
AAI470 MLPLLL--LGISGLLFYQEVSRWSKSAVQNKVVVITDAISGLGKECARVFHTGGARLV
AAH096 ILPLLFGCLGVFGLFRLLQWVR--GKAYLRNAVVVITGATSGLGKECAKVFYAAGAKLV
30 40 50 60 70 80

70 80 90 100 110 120
AAI470 LCGKNWERLENLYDALI-SVADPSKTFTPKLVLLDLSDISCPDVAKEVLDYGVCDILI
AAH096 LCGRNGGALEELIRELTASHATKVQTHKPYLVTFDLTDSGAIVAAAEEILQCFGYVDILV
90 100 110 120 130 140

130 140 150 160 170 180
AAI470 NNASVKVKGPAHKISLELDKKIMDANYFGPITLTKALLPNMISRRGTGQIVLVNNIQGKFG
AAH096 NNAGISYRGTIMDTTVDDKRVMETNYFGPVALTKALLPSMIKRRQGHIVAISISQGMKS
150 160 170 180 190 200

190 200 210 220 230 240
AAI470 IPFRTTYAASKHAALGFDFCLRAEVEEYDVVISTVSPTFIRSYHVYPQGNWEASIWKFF
AAH096 IPFRSAYAASKHATQAFDFCLRAEMEQQEIEVTVISPGYIHT-NLSVNAITADGSRYGVM
210 220 230 240 250

250 260 270 280 290 300
AAI470 FRKLTYGVHPVEVAEEVMRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFFAVVACGVKEK
AAH096 DTTTAQGRSPVEVAQDVLAAVGGKKKDVILADLLPSLAVYLRITLAPGLFFSLMASRARKE
260 270 280 290 300 310
```

Př2: isoformy NQO1

- Stáhněte si z Uniprot databáze tři sekvence isoforem lidské NQO1 a porovnejte je. Odpovídá vaše porovnání popisu jednotlivých isoforem?

Př2: isoformy NQO1

- Stáhněte si z Uniprot databáze tři sekvence isoforem lidské NQO1 a porovnejte je. Odpovídá vaše porovnání popisu jednotlivých isoforem?
 - **Nápověda:** Uniprot ID: P15559, vlevo oddíl sekvence (3), stáhněte fasta formáty a porovnejte pomocí např. multalin, cobalt..popis je u každé isoformy zvlášť (př. i2:140-173: Missing.)

Př2: isoformy NQO1

- Stáhněte si z Uniprot databáze tři sekvenční isoformy lidské NQO1 a porovnejte je. Odpovídá vaše porovnání popisu jednotlivých isoform?
- Nápověda:** Uniprot ID: P15559, vlevo oddíl sekvenční (3), stáhněte fasta formáty a porovnejte pomocí např. multalin, cobalt..popis je u každé isoformy zvlášť (př. i2:140-173: Missing.)

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
sp P15559 NQO1_HUMAN	MYGRRALIVLAHSERTSFNYAMKEAAAAALKKKGMEVVEDLYAMNFNPIISRKDITGKLDKPNFQYPAESVLAAYKEGHLSPDIVAEQKKLEADLVIFQFPLQWFGVPAILKGMFERVFIGEFAYTYA													
sp P15559-2 NQO1_HUM	MYGRRALIVLAHSERTSFNYAMKEAAAAALKKKGMEVVEDLYAMNFNPIISRKDITGKLDKPNFQYPAESVLAAYKEGHLSPDIVAEQKKLEADLVIFQFPLQWFGVPAILKGMFERVFIGEFAYTYA													
sp P15559-3 NQO1_HUM	MYGRRALIVLAHSERTSFNYAMKEAAAAALKKKGMEVVEDLYAMNFNPIISRKDITGKLDKPNFQYPAESVLAAYKEGHLSPDIVAEQKKLEADLVIFQFPLQWFGVPAILKGMFERVFIGEFAYTYA													
Consensus	MYGRRALIVLAHSERTSFNYAMKEAAAAALKKKGMEVVEDLYAMNFNPIISRKDITGKLDKPNFQYPAESVLAAYKEGHLSPDIVAEQKKLEADLVIFQFPLQWFGVPAILKGMFERVFIGEFAYTYA													

	131	140	150	160	170	180	190	200	210	220	230	240	250	260
sp P15559 NQO1_HUMAN	ANYDKGPFSSKKAVLSITTTGGSGSMYSLQGIHGDMNVILWPIQSGLHFCGQVLEPQLTYSIGHTPADARIQILEGAKKRLNIWDETPLYFAPSSSLFDLNFQAGFLMKKEVQDEEKNNKFGLSVGHHL													
sp P15559-2 NQO1_HUM	ANYDKGPFSSKKAVLSITTTGGSGSMYSLQGIHGDMNVILWPIQSGLHFCGQVLEPQLTYSIGHTPADARIQILEGAKKRLNIWDETPLYFAPSSSLFDLNFQAGFLMKKEVQDEEKNNKFGLSVGHHL													
sp P15559-3 NQO1_HUM	ANYDKGPFSSKKAVLSITTTGGSGSMYSLQGIHGDMNVILWPIQSGLHFCGQVLEPQLTYSIGHTPADARIQILEGAKKRLNIWDETPLYFAPSSSLFDLNFQAGFLMKKEVQDEEKNNKFGLSVGHHL													
Consensus	ANYDKGPFSSKKAVLSITTTGGSGSMYSLQGIHGDMNVILWPIQSGLHFCGQVLEPQLTYSIGHTPADARIQILEGAKKRLNIWDETPLYFAPSSSLFDLNFQAGFLMKKEVQDEEKNNKFGLSVGHHL													

	261	270	274
sp P15559 NQO1_HUMAN	GKSIPTDNQIKARK		
sp P15559-2 NQO1_HUM	GKSIPTDNQIKARK		
sp P15559-3 NQO1_HUM	GKSIPTDNQIKARK		
Consensus	GKSIPTDNQIKARK		

Isoform 2 (identifier: **P15559-2**) [UniParc] [FASTA](#) [Added to basket](#)

The sequence of this isoform differs from the canonical sequence as follows:

140-173: Missing.

Př3: sulfotransferasa

- Pomocí proteomických nástrojů předpovězte fyzikálně-chemické vlastnosti sulfotransferasy (**sulfotransferase**) z krevničky střešní (**Schistosoma mansoni**)
- Předpovězte isoelektrický bod tohoto proteinu.
- Spočítejte molekulovou hmotnost.
- Kolik cysteinových zbytků tento protein obsahuje?

Př3: sulfotransferasa

- Pomocí proteomických nástrojů předpovězte fyzikálně-chemické vlastnosti sulfotransferasy (**sulfotransferase**) z krevničky střešní (**Schistosoma mansoni**)
- Předpovězte isoelektrický bod tohoto proteinu.
- Spočítejte molekulovou hmotnost.
- Kolik cysteinových zbytků tento protein obsahuje?
 - Náповěda: Uniprot-G4VLE5

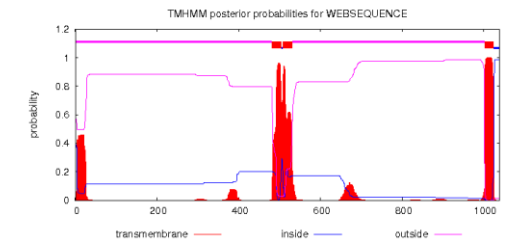
Př4:identifikace neznámé sekvence (moodle)

- Pomocí predikce domén určete o jakou sekvenci by se mohlo jednat.
- Ověřte tuto skutečnost pomocí programu BLAST.
- Z jakého organismu tato sekvence pochází?
- Obsahuje nějaké transmembránové úseky?

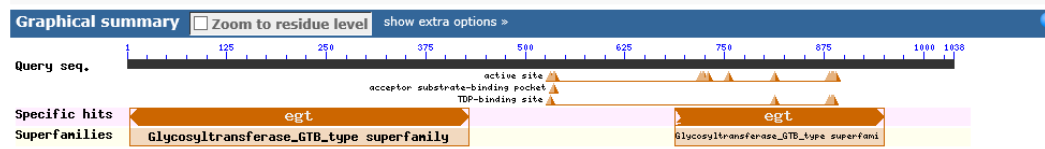
Př4:identifikace neznámé sekvence (moodle)

- Pomocí predikce domén určete o jakou sekvenci by se mohlo jednat.
- Ověřte tuto skutečnost pomocí programu BLAST.
- Z jakého organismu tato sekvence pochází?
- Obsahuje nějaké transmembránové úseky?

```
# WEBSSEQUENCE Length: 1038
# WEBSSEQUENCE Number of predicted TMs: 3
# WEBSSEQUENCE Exp number of AAs in TMs: 72.47098
# WEBSSEQUENCE Exp number, first 60 AAs: 9.92926
# WEBSSEQUENCE Total prob of N-in: 0.43176
# WEBSSEQUENCE TMHMM2.0 outside 1 481
# WEBSSEQUENCE TMHMM2.0 TMhelix 482 504
# WEBSSEQUENCE TMHMM2.0 inside 505 508
# WEBSSEQUENCE TMHMM2.0 TMhelix 509 531
# WEBSSEQUENCE TMHMM2.0 outside 532 1001
# WEBSSEQUENCE TMHMM2.0 TMhelix 1002 1024
# WEBSSEQUENCE TMHMM2.0 inside 1025 1038
```



3x TM



Doména: 2x glykosyltransferasa

☐ [UDP-glucuronosyl UDP-glucosyltransferase domain containing protein \[Haemonchus contortus\]](#)

[organismus]

2152 2152 100% 0.0 100% [CDJ89503.1](#)