

Úkol č. 1: analyzujte následující sekvenci:

MSSQFHGLAIGNGNSNYLQVLGLANITDTAYLTDWQDSGGNWHAGFALPVPSDYPKGHFFQLTTGVGNSN
YLQVLGAGEDGNPYLVSQDGSQKWHGGMPLPKPSGYSGGPLVTGIGNSNYLQVIGARVESSPYLVAWQD
NNGNWHAGMPLPNPSGYAGGFQQLATGNGNDHFLQVVGVGNDGNAYLVTWQNAQQQWSPGFALPKPSGYS
GTFTQLATGVGNGNFLQVLGIGTDGNAYLVAWQDNGGNWHHPGFALPKPSGYNGTFAKLVTGIGNSNYLQV
FGIGSNGVAYLVSQDSSGGNWHGGLTLPQPSGYNGSFSQLAAGNGNSHYLQVVGTDAQGNVYLVSWQDSE
GKWHAGFELPRAS

predpoved funkce

predpoved struktury

predpoved vazebneho/ných miest/a

predikce fyzikalne-chemických vlastnosti proteinu

ma sekvence proteinu nejaka sva specifika?

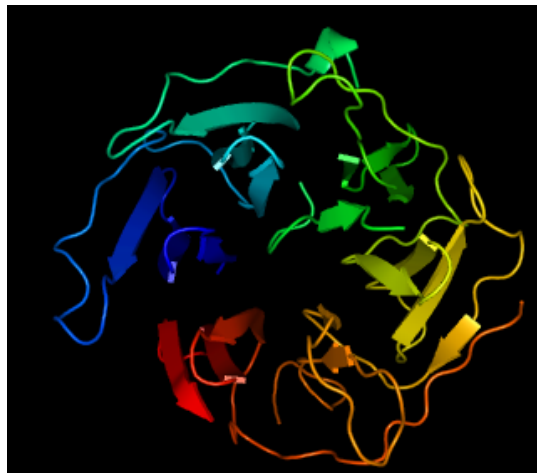
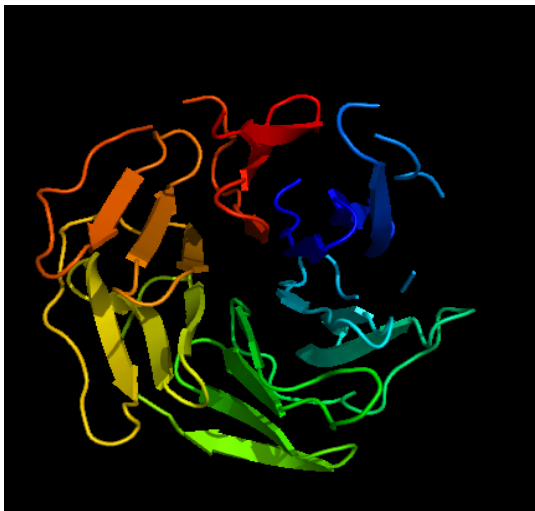
Úkol č. 2 :

Mám neznámý hypotetický protein, který je kódován následující DNA sekvencí.

```
ATGTATCCGTTTTTCGATAACCCGAATTATACCAATACCTATGCCACCAATGAAGATTTTGTGTGTCC
GTATTTCTGGATTATTATAACAACAGCCAGGATGATTATAAAAACCTCCGTGGCGAGAACTATGATT
TTGAAGATACCGAAGAAAATATTGAAAATCGCAATATTGAAGAAACCGAGTATGAAGGTCTGTTTCGT
GCATGGAATCCGTGGAATAATCTGGGTGGTAATATTACCAGCGGTCTGGGTGCAAGCAGCTGGGCAGC
AAATCGTATTGACCTGTTTGCACGTGGTTCGTGGTGGTGAAGTATTTCATAAATGGTTTTGATAATGGCA
AATGGAACCTATTGGGAAAACCTGGGTGGTATTCTGACCAGCAGCCCGAAAGCAGTTAGCTGGGGTTTT
AATCGTATTGATGTTGTTTGTCTGGCACCAGATAACGCAATGTATCATAAATGGTGGGATGGTAGCAG
CTGGTCAGGTTTTGAAAATCTGGGTGGCCAGCTGACCAGCGCACCGACCATTTGTAGCTGGGCACCGA
ATCGTCTGGATTGTTTTGCCCGTGGTACAGATAATCAGCTGCACCACAAATGGTGGGACGGTTCAAGT
TGGAGCCAGTGGGAAGCACTGGGTGGTAGCCTGACCTCAGGTCCGGGTGCAGTGAAGCTGGGGTCCGAA
CCGCATTGATGTTTTTGCAGCGTGGTTCGTAATAACACCCTGATCCATAAATGGTGGAACGGCACCAGTT
GGTCACAGTGGGAAGATCTGGGAGGTTTTTCTGACCTCAGCACCCTGTGCAAGCAGCCGTGGTTCAGAAT
CGCATCGATGTGTTTGCACGCGGACGCAATAATCGTCTGATGTATAAATATTGGGATGGTAGTCGTTG
GAGCGATTGGACCTTCTGCAGGGTTATCTGACCAGCGAACCGGTTAGCGTTAGCCGTAATAGCAGCA
GCATTAATGTGTTTGCAAAAGGTCCGCGTGAAAATGTGATTGAACGCATTTATAGCTAA
```

Prohledávání databáze NCBI pomocí BLASTu ukazuje na nejvyšší podobnosti s jinými hypotetickými proteiny, sacharidy-vázající proteiny či sialidasy. Phyre jako nejlepší modely generuje s templáty:

Fucose-specific lectin from *Aleuria Aurantia* , PDB code: 1 OFZ (vlevo) a NanI sialidase from *Clostridium perfringens*, PDB code: 2VK7 (vpravo)



Rozhodněte, zda se jedná o lektin či enzym . Oba proteiny se liší počtem a lokalizací vazebných míst. Analyzujte sekvenci pomocí nástrojů, které uznáte za vhodné. Identifikujte potenciální vazebná místa a zaměřte se na potenciální konzervované aminokyseliny, které Vám mohou pomoci.

Evoluční konzervované proteiny je nejlépe hledat v sérii proteinů, při sekvenčním přiložení pouze dvou sekvencí se pravděpodobnost náhodné shody umístění aminokyselin zvyšuje. Naznačená předpovězená struktura naznačuje, že protein může obsahovat repetice.

Úkol 3:

Identifikujte 4 arabinosyltransferasy z různých organismů (ale opravdu odlišných). Získejte jejich sekvence.

Předpovězte jejich fyzikálně chemické vlastnosti

Vyrobte plot hydrofobicity pro každý protein a identifikujte potenciální transmembránové úseky v sekvencích.

Proveďte vícenásobné sekvenční přiložení a ukažte potenciální homologní oblasti.

vygenerujte si sekvenční logo svých sekvencí: <http://weblogo.berkeley.edu/>

Úkol 4:

Vyhledejte 20-30 sekvenčních homologů proteinu „hypothetical protein [Chromobacterium violaceum] NCBI Reference Sequence: WP_011135296.1“ z databáze 20 - 30 proteinových sekvencí s alespoň 30% homologií. Vyberte v celém rozsahu s různým zastoupením organismů (pokud budou ze stejného druhu, pak berte s odlišnou homologií). Podle potřeby zvětšete počet vrácených sekvencí (zvýšit limit ze standardních 100 na 500)

Můžete si upravit jejich názvy do přijatelnější podoby pro snazší práci.

Připojte vybrané sekvence ve FASTA formátu

Provedte vícenásobné sekvenční přiložení pomocí CLUSTALW2 (EBI)

Vytvořte dvě fylogenetické analýzy (v kroku 3 ClustaW2) - jednu za pomoci UPGMA metody, druhou použitím Neighbor Joining method.

- 1) Popište výsledný alignment a připojte grafické výstupy MSA a fylogenetické analýzy.
- 2) Zhodnoťte, zda jsou stromy biologicky a taxonomicky validní (porovnejte se standardními taxonomiemi na „NCBI Taxonomy Page“.
- 3) Mají oba stromy stejnou topologii? Mají oba stromy stejnou délku větví? Pokud ne, popište odlišnosti a řekněte proč si myslíte, že se oba stromy odlišují.
- 4) Jsou odlišnosti taxonomicky či biologicky relevantní?
- 5) Odhadněte, zda stromy ukazují na paralogní evoluci? Které spoje by mohly být ortologní a které paralogní?