

Praktická úloha č. 2

**Téma: VÝVOJ LIEKU NA MALÁRIU**

*Pomôcky:*

*Počítač s nainštalovaným programom BioEdit a stabilným pripojením k internetu.*

**Úvod:**

Jednou z oblastí biológie, ktorá sa v modernej dobe dostáva stále viac do popredia je bioinformatika. Bioinformatika sa zaoberá metódami na zhromažďovanie, analýzu a vizualizáciu rozsiahlych súborov biologických dát. Opiera sa najmä o vedomosti z oblasti molekulárnej biológie, genetiky a na druhej strane teoretickej informatiky. Keďže veľká väčšina z vás má pravdepodobne zatiaľ len veľmi malé skúsenosti s týmito oblasťami, budeme sa snažiť vám toho čo najviac vysvetliť v rámci teoretického popisu pri jednotlivých častiach. Taktiež sme si pre vás pripravili slovníček niektorých pojmov, ktoré sa budú používať v tejto úlohe a je vhodné, aby ste im rozumeli. Využívanie internetu na iné účely, ako od vás priamo vyžadujeme v úlohách, je prísne zakázané. Aj keď si v praxi bioinformatici častokrát pomáhajú dohľadávaním informácií, my chceme otestovať aj vaše biologické znalosti a preto, ak budete odhalený pri podvádzaní, môže to viesť k diskvalifikácii z celej súťaže.

Zadanie sme pre vás pripravili aj v papierovej forme, aj ako PDF súbor uložený na ploche v počítači. Je na vás, s ktorou formou sa vám bude pracovať lepšie, avšak pri niektorých úlohách (otvorenie linku, kopírovanie, ...) budete musieť použiť práve súbor z počítača. Naopak, vaše finálne riešenia doplňte do papierovej formy, pretože tú budeme my opravovať.

Bioinformatické nástroje a programy sú v modernej dobe, bohužiaľ, skoro výhradne v angličtine. Aj keď vieme, že väčšina z vás angličtinu ovláda určite dostatočne na to, aby ste nemali problém s takýmito nástrojmi pracovať, nechceme znevýhodniť tých, ktorým by robila problém. Preto sme sa snažili úlohy zostrojiť tak, aby ste jej znalosť potrebovali čo najmenej. Navyše sme pre vás do prehliadača Google Chrome nainštalovali rozšírenie Google Translate, pomocou ktorého si môžete preložiť či už jednotlivé slová alebo celé stránky. Toto rozšírenie nájdete vpravo hore po kliknutí na ikonku:



Pokiaľ narazíte v priebehu súťaže na nejaký technický problém (problém s internetom, nefunkčnosť programu...) alebo problém s prekladom angličtiny, prihláste sa a my vám prideme pomôcť.

### Slovník pojmov:

- **DNA** alebo deoxyribonukleová kyselina je polymér zložený z dvoch polynukleotidových reťazcov, ktoré sa navzájom ovíjajú a vytvárajú dvojitú špirálu. DNA nesie genetické inštrukcie pre vývoj, fungovanie, rast a reprodukciu všetkých známych organizmov a mnohých vírusov. Je uložená v jadre.
- **Nukleotid** – jednotlivé reťazce sú zložené z jednoduchších monomérnych jednotiek nazývaných nukleotidy. Každý nukleotid sa skladá z jednej zo štyroch dusíkatých báz (cytozín [C], guanín [G], adenín [A] alebo tymín [T]), cukru nazývaného deoxyribóza a fosfátovej skupiny.
- **Nukleotidová sekvencia** – poradie nukleotidov, v akom sa nachádzajú na danom vlákne (označené podľa dusíkatých báz, ktoré obsahujú jednotlivé nukleotidy)
- **Zarovnanie (alignment)** - spôsob vzájomného usporiadania sekvencií DNA, RNA alebo proteínu, ktoré slúži na identifikáciu oblastí podobnosti, ktoré môžu byť dôsledkom funkčných, štrukturálnych alebo evolučných vzťahov medzi sekvenciami.
- **Databáza (proteínov)** – súbor informácií (o proteínoch), ktoré boli zhromaždené a sú voľne dostupné pre verejnosť.
- **Kodón** – sekvencia troch nukleotidov, ktoré určujú zaradenie konkrétnej aminokyseliny do polypeptidového reťazca (viac v texte).
- **Aminokyselina** – organická zlúčenina, ktorá predstavuje základnú stavebnú zložku proteínov
- **Polypeptid** – vlákno tvorené aminokyselinami, z ktorého postupnými úpravami vznikne výsledný proteín
- **Ligand** – malá molekula interagujúca s proteínom a spúšťajúca reakciu/zmenu konformácie a pod.

### Praktická časť:

Malária je veľmi vážnym ochorením s vysokou prevalenciou najmä v rovníkovej Afrike, prenášaným komármi rodu *Anopheles*. Toto ochorenie je spôsobené prvokmi rodu *Plasmodium*, vážny priebeh ochorenia je asociovaný s infekciou najmä druhom *P. falciparum*. V nasledujúcej úlohe pomocou bioinformatických metód zistíte informácie o proteíne z *P. falciparum*, ktorý môže byť cieľom pre antimalariká. Ide o enzým **dihydroorotát-dehydrogenázu (DHODH)**. Následne identifikujete dôležité miesta tohto enzýmu, na ktoré možno pri vývoji liečiva cieľiť a nakoniec otestujete rôzne látky, ktoré môžu mať antimalarický potenciál a vyhodnotíte ich interakciu s enzýmom.

Aby sme mohli s proteínom DHODH pracovať, potrebujeme jeho sekvenciu. Tú získame z niektorej z mnohých databáz proteínov, medzi najznámejšie z nich patrí UniProt:

<https://www.uniprot.org>

V tejto databáze však okrem sekvencie nájdeme aj množstvo užitočných informácií o našom proteíne.

### Úloha 1

Vyhľadajte v databáze UniProt proteín DHODH z *P. falciparum* (použite kľúčové slová). Vyhľadávanie by vám malo ponúknuť niekoľko výsledkov. Ďalej budete pracovať s proteínom, ktorý je kódovaný génom s názvom **PFF0160c**.

- a) Napíšte identifikačný kód (Accession) tohto proteínu.
  
- b) Aká je dĺžka tohto proteínu (počet aminokyselín)?

## PO VYPRACOVANÍ TEJTO ÚLOHY PRIVOLAJTE DOZOR NA KONTROLU!

### Úloha 2

Pomocou informácií v databáze UniProt odpovedzte na nasledujúce otázky:

- 2.1 Koľko dvojitych väzieb uhlík-uhlík obsahuje oxidovaný produkt katalytickej reakcie tohto enzýmu?
  
- 2.2 Aká molekula je kofaktorom tohto enzýmu?
  
- 2.3 V ktorej časti bunky je lokalizovaný tento enzým?
  - a) jadro
  - b) membrána mitochondrie
  - c) lyzozóm
  - d) cytoplazma
- 2.4 Bola zistená štruktúra tohto proteínu? Ak áno, akou metódou?

### Úloha 3

Zamyslite sa, ako môže inhibícia tohto enzýmu prispieť k liečbe malárie a vyberte správnu možnosť (opäť vám môžu pomôcť informácie v UniProte).

- a) DHODH sa priamo zúčastňuje prenosu protónov na vnútornej membráne mitochondrie, jeho inhibícia by mala za dôsledok inhibíciu tvorby ATP.
- b) DHODH zabezpečuje tvorbu pyrimidínových báz, jeho inhibícia by negatívne ovplyvnila replikáciu a transkripciu genómu parazita.
- c) DHODH zabezpečuje tvorbu pyrimidínových báz, jeho inhibícia by negatívne ovplyvnila najmä pohyb parazita prostredníctvom aktínových filamentov.

Dozvedeli ste sa niečo o našom proteíne, teraz je čas aplikovať nejaké bioinformatické nástroje. Stiahnite aminokyselinovú sekvenciu proteínu vo FASTA formáte (tento formát sa v bioinformatike používa veľmi často, skladá sa z názvu sekvencie v prvom riadku za znakom „>“ a následne zo samotnej sekvencie uvedenej v ďalšom riadku). Môžete tak urobiť priamo v



A aby ste si zobrazili iba zhody sekvencií, ktoré budú vyznačené bodkou, kliknite na túto ikonku:



(Pozn.: Všimnite si, že ľavý stĺpec s názvami sekvencií sa dá v programe BioEdit zväčšovať a zmenšovať potiahnutím za lištu.)

Odpovedzte na nasledujúce otázky:

- Ktorý koniec proteínu parazita (N alebo C) by mohlo potenciálne liečivo selektívne rozpoznávať?
- Na ktorom úseku (okrem koncového) je v sekvencií ľudského enzýmu pravdepodobne rozsiahla delécia? (Pozor, udávajte pozíciu vzhľadom na celý alignment (zarovnanie), NIE na niektorú zo sekvencií, napr. na *P. falciparum*.)
- Aký je najdlhší konzervovaný úsek (jeho sekvencia) medzi týmito dvomi sekvenciami?
- Prečo algoritmus označuje aminokyseliny V – valín, L – leucín a I – izoleucín ako podobné? Prečo tak označuje aminokyseliny D – aspartát a E – glutamát?

**Súbory so stiahnutými sekvenciami si nechajte otvorené alebo si ich skopírujte do textového súboru. Budete ich potrebovať v ďalších častiach úlohy.**

Už sme sa niečo dozvedeli o podobnosti nášho enzýmu s jeho ľudským variantom, teraz sa pokúsime zistiť, ktoré časti jeho sekvencie sú konzervované naprieč viacerými druhmi a na základe toho identifikujeme, kam by mohlo naše liečivo cieľiť.

### Úloha 5

Okrem *P. falciparum* u človeka spôsobuje maláriu aj niekoľko ďalších druhov tohto parazita, jedným z nich je *P. knowlesi*. Vyhľadajte v prehliadači, ktoré ďalšie tri druhy *Plasmodium* spôsobujúce maláriu u človeka a stiahnite z UniProtu proteínové sekvencie ich enzýmu DHODH (zvoľte vždy prvý výsledok, ktorý vám UniProt zobrazí po zadaní do vyhľadávacieho okienka „*plasmodium* [druhovú meno] *dhodh*“) týchto troch druhov (nemusíte vyhľadávať sekvenciu *P. knowlesi*).

- I. Napíšte názvy troch vybratých druhov a ku každému Accession kód sekvencie, ktorú ste zvolili.
  
- II. Importujte vybrané FASTA sekvencie do BioEditu rovnako ako v predchádzajúcich úlohách, vyberte opäť požitím *Ctrl* + klik sekvencie vybratých druhov a druhu *P. falciparum*, ktorú ste už používali, neoznačujte však ľudskú sekvenciu. Urobte zarovnanie (*alignment*) rovnako ako v predchádzajúcej úlohe a vypracujte nasledujúce podúlohy:
  - a) Napíšte pozíciu troch najdlhších konzervovaných oblastí naprieč všetkými sekvenciami. (Pozor, udávajte pozíciu vzhľadom na celé zarovnanie (alignment), NIE na niektorú zo sekvencií, napr. na *P. falciparum*.)
  
  - b) Sekvencia ktorého druhu obsahuje najmenej medzier (*gaps*) v zarovnaní (*alignmente*) v porovnaní s ostatnými?

### Úloha 6

Pridajte k sekvenciám aj sekvenciu ľudského enzýmu, urobte zarovnanie (*alignment*) všetkých sekvencií. Vyberte správne možnosti:

- a) Dva z troch úsekov sekvencie, ktoré boli konzervované (z predchádzajúcej úlohy) naprieč rôznymi druhmi *Plasmodium*, sa zhodujú s daným úsekom ľudskej sekvencie.
- b) Najväčšiu podobnosť (pozor, podobnosť, NIE zhodu) spomedzi v predchádzajúcej úlohe vybratých úsekov s daným úsekom ľudskej sekvencie vykazuje najkratší z úsekov.
- c) Čím je sledovaný úsek sekvencie, na ktorú nie je vyvíjaný selekčný tlak, dlhší, tým vyššia je teoretická pravdepodobnosť, že sa bude medzi rôznymi druhmi líšiť.
- d) Najvýhodnejšie je pri výbere lieku cieľiť na najkratší z vybratých konzervovaných úsekov proteínu.

Na odhad možného cieľa liečiva nám môže poslúžiť znalosť 3D štruktúry tohto proteínu. Priamo zo sekvencie ju dokáže odhadnúť niekoľko nástrojov, vy použijete [Swiss-model](#).

## Úloha 7

Otvorte odkaz, stlačte *Start Modeling*. Vložte sekvenciu DHODH vo FASTA formáte z *P. falciparum* do okna, ktoré vám ponúka Swiss-model a kliknite na *Build model*. Počkajte, kým nástroj vytvorí niekoľko modelov (môže to trvať niekoľko minút, ak je to dlhšie, privolajte dozor) a odpovedzte na nasledujúce otázky:

7.1 Tento nástroj vytvára modely na základe podobnosti so štruktúrami v databáze. Akou metódou bola získaná štruktúra, na základe ktorej bol vytvorený model, ktorý pokrýva našu sekvenciu najlepšie?

- a) röntgenová (X-ray) kryštalografia
- b) počítačové modelovanie (AlphaFold)
- c) kryoelektrónová mikroskopia
- d) cirkulárny dichroizmus

7.2 Model vám ponúka množstvo možností, veľmi dôležité je vedieť, ako veľmi si je nástroj istý svojím modelom. Aká aminokyselina a s akou pravdepodobnosťou (spoľahlivosťou) je podľa modelu 1 na pozícii 151?

7.3 Na ktorej pozícii je podľa modelu 2 cysteín (C, Cys), ktorý interaguje s produktom (substrátom) enzymatickej reakcie DHODH?

- a) 175
- b) 276
- c) 321
- d) 512

7.4 Z akého organizmu pochádza templát pre model, ktorý pokrýva najmenšiu časť našej sekvencie?

## Úloha 8

Najdlhšia konzervovaná časť sekvencie naprieč štyrmi druhmi *Plasmodium* podľa modelu 1 tvorí:

- a) Beta-barrel
- b) Alfa-helix
- c) Beta-otočku (*beta-turn*) a alfa-helix

Dostávame sa do finále, už poznáme konzervované časti sekvencie nášho enzýmu, poznáme jeho štruktúru a je načas pozrieť sa na jeho interakcie s potenciálnymi liečivami.

## Úloha 9

Na overenie interakcií ligandov (malých molekúl interagujúcich s proteínom ovplyvňujúcich jeho funkcie) s našim proteínom použijeme nástroj [PLIP](#). Ten potrebuje na vyhodnotenie štruktúru z databázy štruktúr [PDB](#). Bohužiaľ, nemožno použiť model, ktorý ste si v predchádzajúcej úlohe vytvorili. Identifikačný kód štruktúry enzýmu DHODH, ktorú použijete je 6I4B. Tento kód vložte do okna, ktoré ponúka PLIP a stlačte *Analyze*. Dostanete ligandy,

ktoré interagujú s našou štruktúrou. Dva z nich by vás nemali prekvapiť. Pre jednoduchšiu orientáciu v modeli si ho viete zobraziť v 3D kliknutím na tlačidlo v pravej dolnej časti okienka, kde je model zobrazený.

[Click for 3D-View](#)

Odpovedzte na nasledujúce otázky (na vyhľadávanie informácií o ligandoch odporúčame použiť databázu chemických látok [PubChem](#), kde môžete nájsť všetky potrebné informácie na správne vypracovanie úlohy, takisto môžete použiť tabuľku so vzorcami aminokyselín, ktorú ste dostali spolu so zadaním, a môžete použiť aj vyhľadávanie v prehliadači):

9.1 Aká je molekulová hmotnosť tretieho, pre vás zatiaľ neznámeho ligandu?

9.2 Čo a prečo by pravdepodobne spôsobila zámena fenylalanínu za alanín na pozícii 188 v sekvencii enzýmu?

- a) došlo by k porušeniu  $\pi$ -interakcie ligandu s proteínom, pretože alanín nemá aromatický bočný reťazec
- b) k zmene interakcie by pravdepodobne nedošlo, pretože alanín a fenylalanín sú si štruktúrne veľmi podobné
- c) pravdepodobne by táto zámena zapríčinila posilnenie interakcie ligandu s proteínom, pretože bočný reťazec alanínu je schopný tvoriť s aromatickým jadrom ligandu kovalentnú väzbu

9.3 Aký atóm ligandu interaguje vodíkovými väzbami s arginínom?

- a) kyslík
- b) vodík
- c) dusík
- d) síra

9.4 Mala by väčší efekt na interakciu ligandu s proteínom zámena valínu za leucín na pozícii 172 alebo cysteínu za glutamát na pozícii 175? Prečo?

Táto úloha bola poslednou v tejto praktickej úlohe.

**Priestor pre spätnú väzbu:**

Zdala sa vám úloha ťažká? Čo sa vám na nej páčilo/nepáčilo?