

## Przykładowa wiązka zadań z wykorzystaniem narzędzi bioinformatycznych

Plik „6C1Y.fasta” w podkatalogu „Sekwencje” zawiera sekwencję aminokwasową części jednego z czynników transkrypcyjnych człowieka, dla którego struktura przestrzenna została zdeponowana w bazie Protein Data Base pod numerem dostępu 6C1Y.

Korzystając z programu *tblastn* dostępnego na stronie <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, wyszukaj sekwencje nukleotydowe kodujące białko o podobnej sekwencji aminokwasowej. Aby ograniczyć przeszukiwanie do pełnej długości RNA, wybierz bazę danych (ang. *database*) „Reference RNA sequences (refseq\_rna)”.

Spośród znalezionych wpisów do bazy danych wybierz ten przypisany do *Homo sapiens*. Wyświetl przyrównanie, a następnie, klikając na identyfikator (ang. *Sequence ID*), wyświetl wpis do bazy danych.

### Pytanie 1: Część struktury jakiego białka została zdeponowana w bazie danych Protein Data Bank pod numerem identyfikatora „6C1Y”?

Na samym dole wpisu do bazy danych znajduje się sekwencja mRNA kodująca to białko. Skopiuj ją, utwórz nowy plik w programie ApE i wklej ją do pustego okna. Klikając „ORFs > Find next” znajdź najdłuższą otwartą ramkę odczytu w obrębie tego mRNA.

### Pytanie 2: Z ilu pozycji nukleotydowych składa się najdłuższa otwarta ramka odczytu?

Przeprowadź translację *in silico* znalezionej otwartej ramki odczytu i zapisz sekwencję aminokwasową tego białka w nowym pliku w formacie FASTA (nazwij ten plik „znalezione\_bialko.fasta”).

### Pytanie 3: Ile aminokwasów zawiera białko kodowane przez znaleziony w bazie danych mRNA?

Dodaj sekwencję znalezionego białka (z pliku „znalezione\_bialko.fasta”) do sekwencji tego samego białka z czterech różnych organizmów, które znajdują się w pliku „inne.fasta” (podkatalog „Sekwencje”). Utwórz przyrównanie w programie ClustalX: (1) otwórz nowo utworzony plik zawierający w sumie 5 sekwencji aminokwasowych (File > Load Sequences), (2) utwórz przyrównanie (Alignment > Do Complete Alignment) i (3) zachowaj wynik w nowym pliku (File > Save Sequences as...), nazywając go „inne\_znalezione”. Zaznacz „NEXUS format”, aby przyrównanie zostało zapisane w formacie NEXUS.

Wczytaj sekwencję „inne\_znalezione.nxs” do programu PAUP\* (File > Open...). Następnie oblicz drzewo filogenetyczne (Analysis > Neighbor Joining/UPGMA) – wybierz algorytm UPGMA oraz opcję zapisania drzewa filogenetycznego w formacie NEXUS (Save to treefile).

**Pytanie 4:** Który z niżej wymienionych organizmów jest najbliższym spokrewnionym z człowiekiem według drzewa obliczonego na podstawie powyższych sekwencji aminokwasowych?

- A) *Aotus nancymaae*.
- B) *Canis lupus*.
- C) *Cebus imitator*.
- D) *Propithecus coquereli*.

**Odpowiedzi**

Pytanie 1: methyl-CpG binding protein 2 (MECP2).

Pytanie 2: 1497 par zasad (pz).

Pytanie 3: 498.

Pytanie 4: A.