

Zadanie 1.

Dane są dwie sekwencja DNA:

pierwsza:

ATGAAAACTGATCGTAGCT

AGCTAGCTAGCTAGTCGATGCTG, i druga:

ATGGCCAATCGTAGCTAGTCG

ATGCATGCTAGCTGATCGATGC.

Oblicz stopień podobieństwa sekwencji, przyjmując, że identyczne nukleotydy na odpowiadających pozycjach liczymy jako 1, a różne jako zero.

Zadanie 2.

Znajdź za pomocą `NeedlemanWunschSimilarity` z opcją `SimilarityRules` (lub wyszukaj i wprowadź ręcznie) podobieństwo wszystkich możliwych par pojedynczych aminokwasów (oznaczonych jednoliterowo) wg. macierzy BLOSUM62. Powtórz obliczenia dla macierzy PAM250 używając `SmithWatermanSimilarity`. Porównaj wyniki z danymi podręcznikowymi.

Zadanie 3.

Które z sekwencji DNA zawartych w pliku <http://ribes.if.uj.edu.pl/bioinfo/z5z3.txt>

są do siebie najbardziej podobne wg. kryterium z Zad. 1? Które z nich są najbardziej odmienne?

Zadanie 3.

Znajdź optymalne ułożenie (ang. *alignment*) **bez przerw** dwóch sekwencji DNA:

pierwsza:

AGCCTTACGAAATATGTAATCGAAGATT

AATATACTACCTGCGCTAGTCGAGCGTC

GAAGAGGAGCAGTGTAAGCGGAGCGCG

ATGGAATAGATGCGTGGGTGTTCCAAGG

ACAGATGCGGTGTTTGGCCCCCTAC

druga:

ACACACACACACCAACAGCCTTACGTAA

TATGTAATCGAAGATTAATATACCCTAC

Określenie „optymalne ułożenie” należy rozumieć następująco: ustawiamy sekwencje jedna nad drugą przesunięte o wybraną ilość pozycji, a następnie obliczamy stopień podobieństwa jak w Zad. 1. Optymalnym nazywany to ułożenie, dla którego „jakość dopasowania” przyjmuje maksymalną wartość.

UWAGA: w celu wyrobienia sobie intuicji proponuję najpierw wyznaczyć optymalne ustawienie „optycznie” używając np. `Manipulate`, a następnie zweryfikować wynik obliczeniowo; w praktyce metody te stosuje się równocześnie.