

POLITECHNIKA WROCŁAWSKA,
WYDZIAŁ PODSTAWOWYCH PROBLEMÓW TECHNIKI
WSTĘP DO BIOINFORMATYKI

ZADANIE NR 2
DOPASOWANIE GLOBALNE PAR SEKWENCJI

Termin oddania: **07.04.2021 r., godz. 23:59**

Przygotuj poprawnie działający program do ilościowego globalnego porównania par sekwencji kodujących białka lub DNA za pomocą algorytmu Needlemana-Wunscha i wykonaj za jego pomocą przykładowe analizy. Program powinien kontrolować poprawność danych wejściowych, generować prawidłowe wyniki, a także być zwięzły, czytelny (opatrzone komentarzami) i zgodny z zasadami stylu używanego języka programowania. Odrębne zadania powinny być wykonywane przez odrębne funkcje. Do programu dołączamy sprawozdanie.

PROGRAM wymagania:

1. Możliwość wczytania sekwencji na 3 różne sposoby:
 - ręczne,
 - z pliku w formacie FASTA,
 - zdalne (z bazy danych NCBI).
2. Możliwość wyboru schematu punktowania zgodności i niezgodności liter oraz kar za przerwy;
3. Wygenerowanie ścieżki optymalnego dopasowania
 - jej wyświetlenie razem z parametrami programu, punktacją, długością dopasowania procentem pozycji identycznych i przerw
 - i zapisem do pliku w formacie:

```
# 1: seq1
# 2: seq2
# Mode: distance
# Match: 0
# Mismatch: 1
# Gap: 1
# Score: 4
# Length: 7
# Identity: 3/7 (43%)
# Gaps: 3/7 (43%)
GC-AGTC
| | |
-CTAG-G
```
4. Graficzne wyświetlenie tablicy punktacji wraz z optymalną ścieżką oraz jej zapis do pliku;
5. Opracowanie interfejsu użytkownika z poziomu linii komend (wiersza poleceń), z zestawem testów (przykładów użycia) **albo** graficznego.

Sprawozdanie ma zawierać:

1. Schemat blokowy algorytmu dopasowania globalnego.
2. Analizę złożoności obliczeniowej czasowej i pamięciowej.
3. Porównanie przykładowych par sekwencji ewolucyjnie
 - powiązanych,
 - niepowiązanych.