# Bioinformatyka – zadanie 1

## Opis problemu

Badanym problemem jest problem uliniowienia sekwencji. Podane zadanie zostało wykonane z użyciem dwóch algorytmów Needlemana-Wunscha oraz Watermana-Smitha. Oba algorytmy zostały zaimplementowane w języku Python. Działanie podanych algorytmów zostało sprawdzone podczas porównania dwóch sekwencji z użyciem kilku różnych funkcji scorujących (scoring function).

## Metody

Dla podanego eksperymentu zostały użyte dwa algorytmy stosowane do problemu dopasowania sekwencji :

* Algorytm Needlemana-Wunscha – algorytm stosowany do znalezienia optymalnego uliniowienia globalnego. Podany algorytm stosuje ideę programowania dynamicznego do rozwiązania problemu dopasowania sekwencji. Przy użyciu tablicy, porównywane są mniejsze pod problemy, które pomagają uzyskać najlepszy wynik dla całych sekwencji.
* Algorytm Smitha-Watermana – algorytm bardzo podobny do algorytmu Needlemana-Wunscha. Główną różnicą pomiędzy algorytmami jest fakt, że algorytm Smitha-Watermana wybiera maksimum z wykonywanych działań oraz 0, co sprawia że w macierzy na, której pracuje znajdują się wyłącznie dodatnie wartości. Podany algorytm znajduje rozwiązania lokalne.

## Eksperymenty

Powyższe algorytmy zostały przetestowane dla poniższych problemów :

1. Porównanie dwóch genów homologicznych, zostały użyte geny :
   1. APOE człowieka - <https://www.ncbi.nlm.nih.gov/gene/348>
   2. APOE makaka królewskiego - <https://www.ncbi.nlm.nih.gov/gene/714623>
2. Porównanie sekwencji białkowej insuliny :
   1. ludzkiej - <https://www.ncbi.nlm.nih.gov/gene/3630>
   2. chomika - <https://www.ncbi.nlm.nih.gov/gene/100760716>

Eksperymenty zostały przeprowadzone dla trzech różnych funkcji scorujących :

* Funkcja, która w znacznym stopniu nagradza za dobre dopasowanie :
  + 10 punktów za dobre dopasowanie
  + -1 punkt za złe dopasowanie
  + -1 punkt za przerwę w sekwencji
* Funkcja, która karze za złe dopasowanie
  + 2 punkty za dobre dopasowanie
  + -6 punktów za złe dopasowanie
  + -2 punkty za przerwę w sekwencji
* Funkcja, która karze za przerwy w sekwencji
  + 2 punkty za dobre dopasowanie
  + -2 punkty za złe dopasowanie
  + -5 punkty za przerwę w sekwencji

## Wnioski