

**POLITECHNIKA WARSZAWSKA**  
**WYDZIAŁ ELEKTRONIKI I TECHNIK INFORMACYJNYCH**



**Uczenie Maszynowe w Bioinformatyce**

Sequence Logo

Kacper Kubicki 293556

Data realizacji projektu:

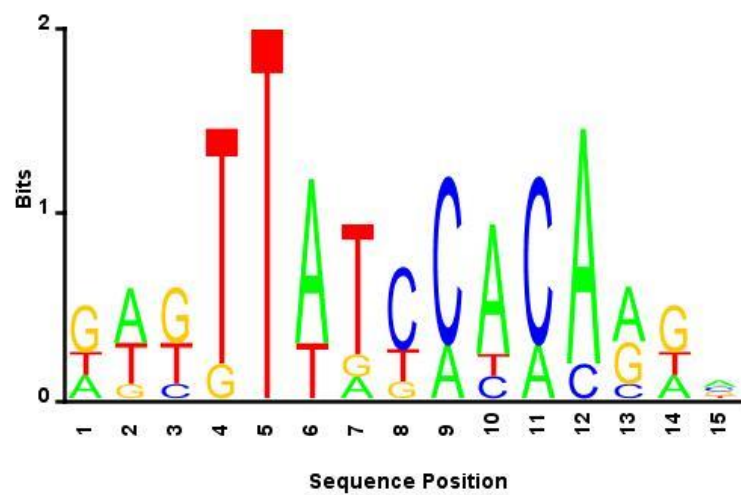
16.06.2021r.

## 1. Wstęp

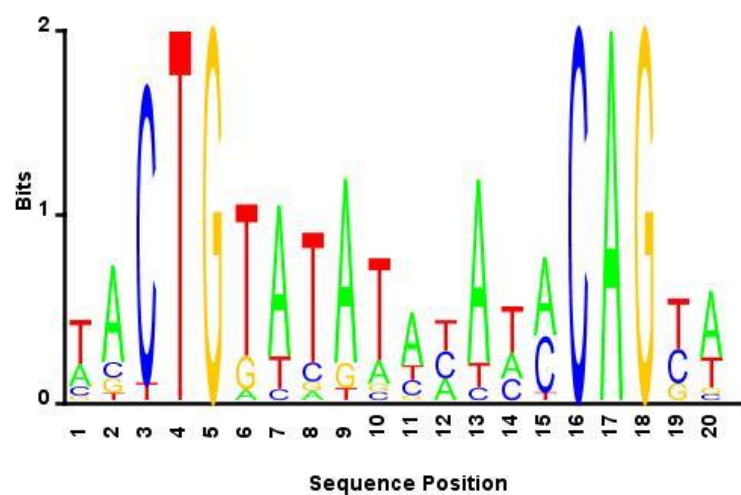
W bioinformatyce logo sekwencji jest graficzną reprezentacją konserwacji sekwencji nukleotydów (w nici DNA/RNA) lub aminokwasów (w sekwencjach białek). Logo sekwencji jest tworzone ze zbioru dopasowanych sekwencji i przedstawia sekwencję konsensusową i różnorodność sekwencji. Logo sekwencji są często używane do przedstawiania cech sekwencji, takich jak miejsca wiązania białek w DNA lub jednostki funkcjonalne w białkach. Logo sekwencji składa się ze stosu liter w każdej pozycji. Względne rozmiary liter wskazują ich częstotliwość w sekwencjach. Całkowita wysokość liter przedstawia zawartość informacyjną pozycji w bitach. Aby utworzyć logo sekwencji, pokrewne sekwencje DNA, RNA lub białek lub sekwencje DNA, które mają wspólne konserwowane miejsca wiązania, są dopasowywane tak, aby najbardziej konserwowane części tworzyły dobre wyrównania. Logo sekwencji można następnie utworzyć z zachowanego dopasowania wielu sekwencji. Logo sekwencji pokaże, jak dobrze reszty są zachowane w każdej pozycji: im wyższa liczba reszt, tym wyższe będą litery, ponieważ tym lepsza jest konserwacja w tej pozycji. Różne reszty w tej samej pozycji są skalowane zgodnie z ich częstotliwością. Wysokość całego stosu reszt to informacja mierzona w bitach. Logo sekwencji można wykorzystać do przedstawienia konserwatywnych miejsc wiązania DNA, w których wiążą się czynniki transkrypcyjne.

## 2. Wyniki

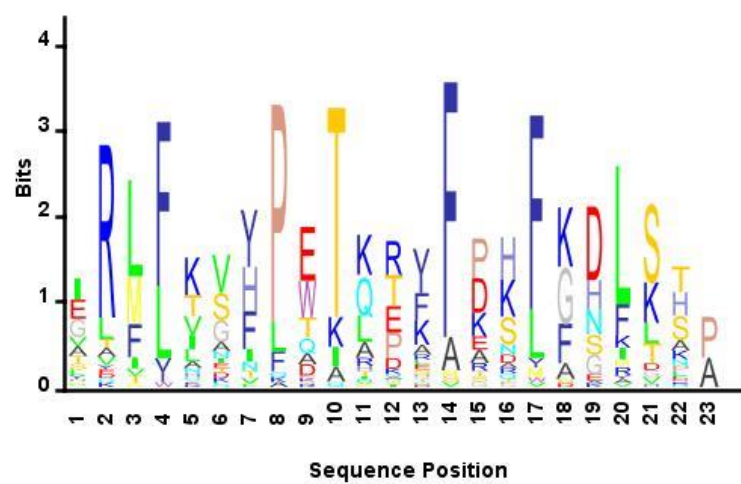
Przy wykorzystaniu środowiska MATLAB R2021a odczytane zostały następujące pliki o rozszerzeniu fasta: nukleotydy1, nukleotydy2, aminokwasy1, aminokwasy2. Poniżej na ich podstawie zwizualizowane zostały sekwencje nukleotydów (rys. 1 i 2) oraz aminokwasów (rys. 3 i 4) przy wykorzystaniu Sequence Logo.



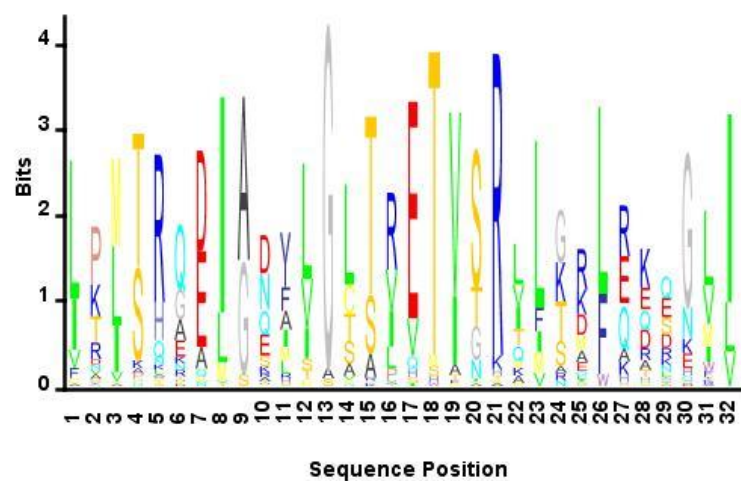
Rysunek 1. Sequence Logo sekwencji zawartych w pliku nukleotydy1.fasta



Rysunek 2. Sequence Logo sekwencji zawartych w pliku nukleotydy2.fasta



Rysunek 3. Sequence Logo sekwencji zawartych w pliku aminokwasy1.fasta



Rysunek 4. Sequence Logo sekwencji zawartych w pliku aminokwasy2.fasta