

Zadanie 1.

Dla dwóch wybranych sekwencji aminokwasowych s i c narysować wykres zawierający elementy macierzy $M_{ij} = w(s_i, c_j)$, gdzie waga w to np. elementy macierzy jednostkowej, BLOSUM62, PAM250, lub innej określającej podobieństwo pojedynczych aminokwasów. Narysować elementy macierzy korzystając z **DotPlot**, **ArrayPlot**, **MatrixPlot**, **Raster** i **ReliefPlot**.

WSKAZÓWKA: *Do porównania można użyć np: sekwencji białkowe **ProteinData["HOXA1**"]***

Zadanie 2.

Wyszukać w bazie danych PDB (www.pdb.org) białko, które zawiera podsekwencje (ang. *motif*) aminokwasowe LEEL oraz RRRR rozdzielone niesprecyzowaną ilością dowolnych aminokwasów.

Pobrać rekord, wczytać do Mathematici sekwencję za pomocą **Import** i zaznaczyć kolorem lub pogrubieniem wyszukany motyw.

UWAGA: *treść została zmieniona aby zredukować ilość znalezionych rekordów do 9*

Zadanie 3.

Wyszukać w bazie danych PDB białka zawierające motyw, na który składa się: LE, dowolny aminokwas, pomiędzy 1 a 10 dowolnych aminokwasów oraz RRRR.

Zapisać podaną przez PDB listę identyfikatorów znalezionych sekwencji do pliku, wczytać do Mathematici wszystkie sekwencje i wykonać wykres typu „dot matrix” dla wszystkich możliwych par. Ocenic wizualnie stopień podobieństwa.

Zbiór wyszukanych sekwencji podzielić na podzbiory zawierające podobne elementy za pomocą **FindClusters** z podobieństwem mierzonym poprzez **EditDistance** używając metody grupowania hierarchicznego (**Agglomerate**). Dokonać kilku dopasowań sekwencji wewnątrz grup i porównać w wcześniejszymi ocenami wizualnymi.

UWAGA: *aby uniknąć wielokrotnego długotrwałego pobierania rekordów z PDB, raz pobrane dane należy zapisać na dysku (Z:), a eksperymenty rozpocząć od możliwie małego podzbioru danych; można też przygotować listę adresów i wcześniej pobrać całe rekordy używając np. TotalCommandera pod Windows lub polecenia wget pod Linuksem*