

Konstancja Gałat
Karolina Stelmaszczyk
grupa 3
ćwiczenie modeller

1. Wyszukiwanie struktur podobnych do danego białka

- environ

Klasa environ zawiera większość informacji o środowisku MODELLER, takich jak funkcja energii oraz biblioteki parametrów i topologii. Zwykle jest to pierwsza klasa używana w skrypcie MODELLER, ponieważ zapewnia metody tworzenia innych głównych klas.

- pdb_95.pir

Plik jest bazą sekwencji aminokwasowych kilkunastu różnych białek - tworzą one alignment.

- sequence_db

Klasa sequence_db przechowuje bazę danych sekwencji. Taka baza danych jest podobna do dopasowania wielu sekwencji, ale zawiera mniej informacji pomocniczych (na przykład żadna sekwencja nie może zawierać przerw). Wymaga to mniej pamięci niż prawdziwe wyrównanie, a zatem jest bardziej odpowiednie dla dużych baz danych sekwencji. Taką bazę danych można przeskanować w poszukiwaniu dopasowań do sekwencji wejściowej za pomocą `sequence_db.search()`, użyć do zbudowania profili sekwencji za pomocą `profile.build()` lub przefiltrować według zadanych kryteriów za pomocą `sequence_db.filter()`.

`minmax_db_seq_len=(30, 4000)` oznacza długość wyszukiwanych sekwencji.

- Nowy format

Plik zostaje przepisany na format bin. Jest to spowodowane tym, że odczyt pliku binarnego jest znacznie szybszy niż odczyt pliku PIR czy FASTA.

- Macierz podobieństwa

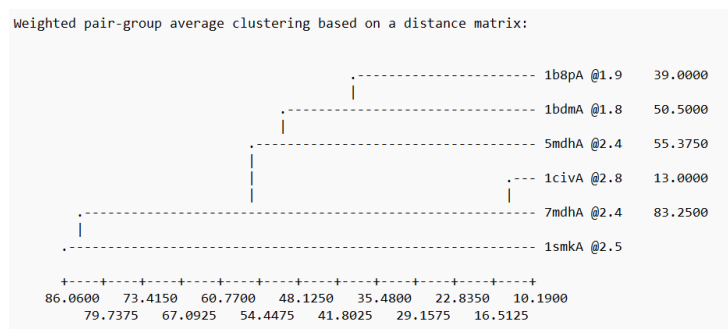
Program używa macierzy podobieństw Blosum62. E-value to prawdopodobieństwo, że sekwencje zostały uznane za homologiczne, mimo że takie nie są - im e-value mniejsze tym wiarygodniejszy wynik.

- Białka podobne

Znaleziono 30 białek podobnych.

2. Wybór wzorca do modelowania homologicznego

Potencjalnie najlepszymi kandydatami do modelowania homologicznego będą te o najniższej wartości e-value, o największej homologii oraz o najwyższej rozdzielczości.



Rysunek 1: Dendrogram zaproponowanych modeli

Wybieramy 1bdmA ze względu na to, że identyczność tej sekwencji jest najbardziej zbliżona do sekwencji kwerendy i ma wysoki współczynnik krystalograficzny R (wysoką rozdzielczość).

3. Dopasowanie białka i wzorca

4. Budowa modelu

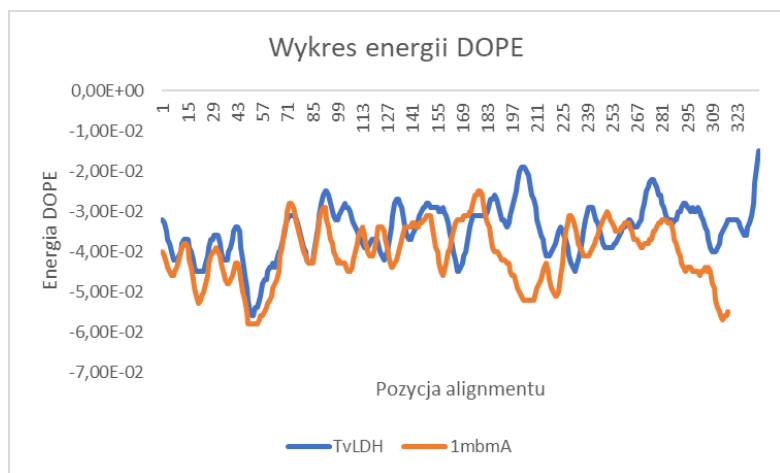
>> Summary of successfully produced models:

Filename	molpdf	DOPE score	GA341 score
TvLDH.B99990001.pdb	1819.52161	-37730.20703	1.00000
TvLDH.B99990002.pdb	1749.18665	-37702.49609	1.00000
TvLDH.B99990003.pdb	1976.34973	-37818.56250	1.00000
TvLDH.B99990004.pdb	1855.88367	-37706.48828	1.00000
TvLDH.B99990005.pdb	1826.46118	-37670.34375	1.00000

Rysunek 2 Wartości energii DOPE dla uzyskanych modeli.

Plik log zawiera podsumowanie wszystkich zbudowanych modeli - zostało utworzonych 5 modeli.

5. Ocena modelu



Wykres 1: Wykres energii DOPE

Znaczące różnice w kształcie wykresów świadczące o słabym dopasowaniu w danym miejscu i wskazujące na słabą jakość modelu w tym obszarze występują. Są 2 duże obszary, w których występuje niezgodność.