# Sprawozdanie nr 2

#### Analiza sekwencji i struktury drugorzędowej białek

Małgorzata Stęperska 151546, Adam Dachtera 147890

Legenda: do uzupełnienia, nie jestem pewna

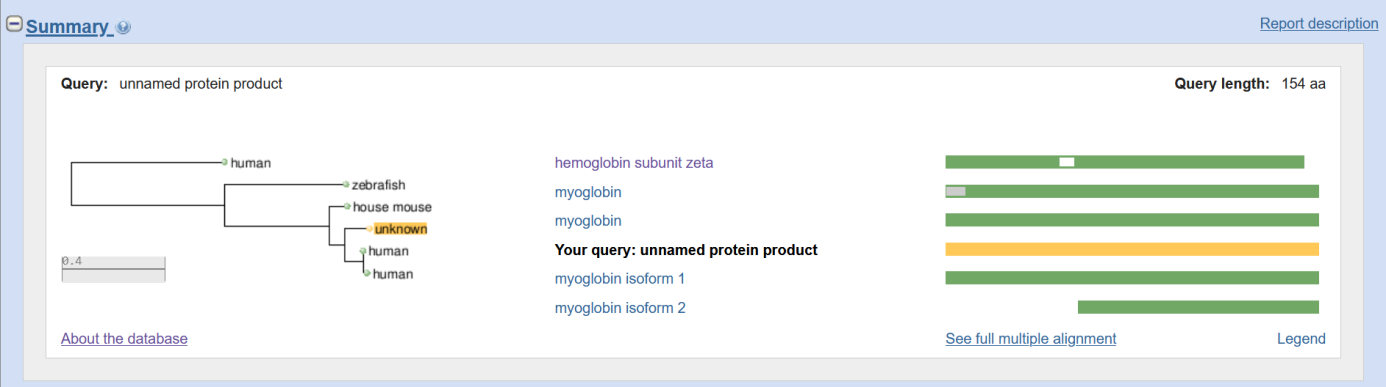
II. Algorytm *wpisać algorytm  
Wkleić link*

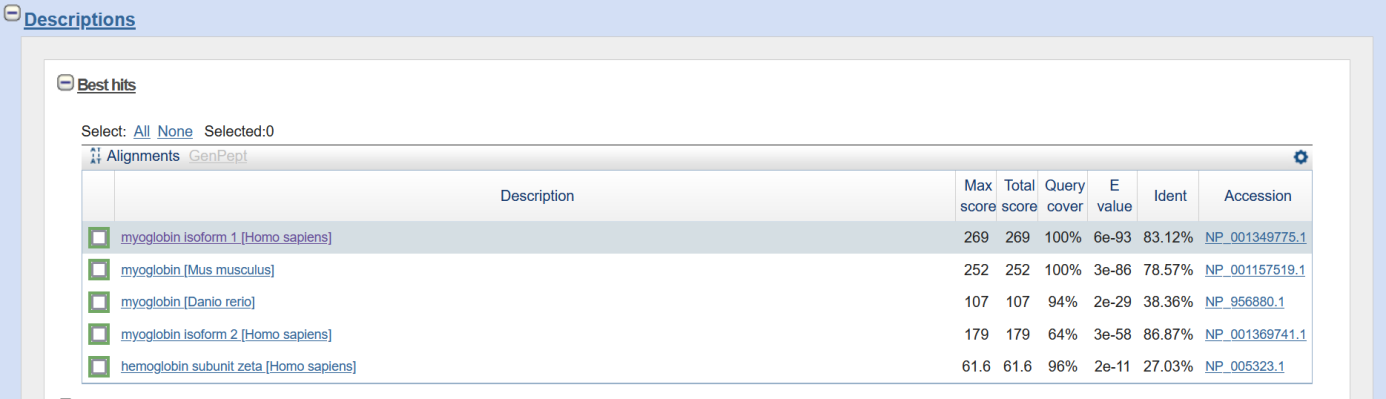
1. BLAST

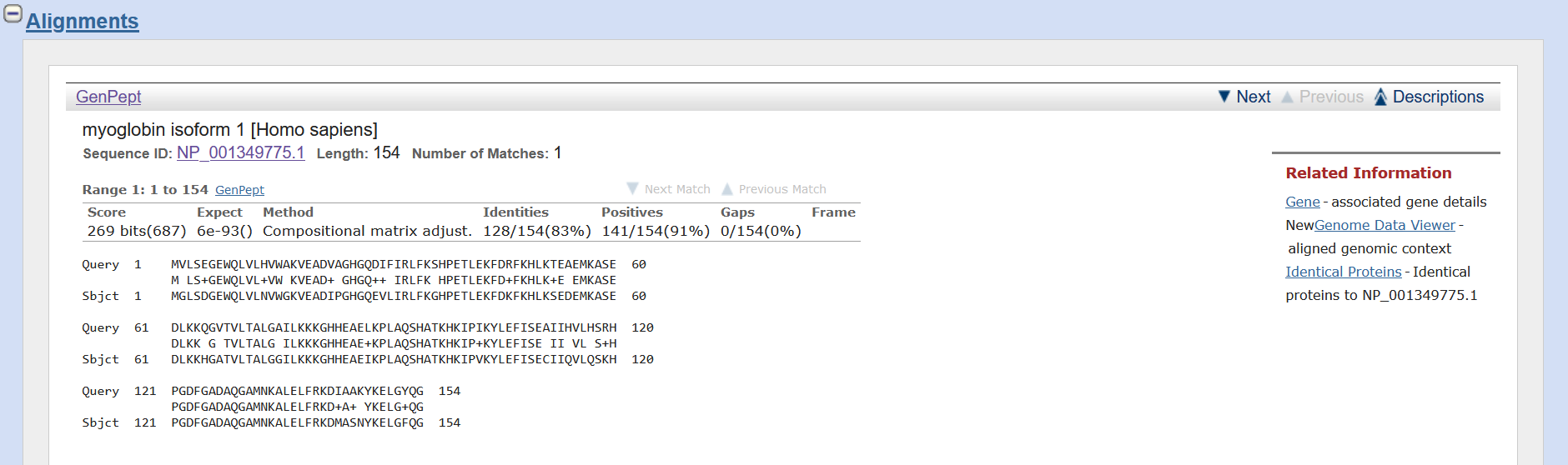
W miejscu wyszukiwania wpisaliśmy „myoglobin”, dzięki czemu znaleźliśmy strukturę 1MCY (sperm whale myoglobin (mutant with initiator met and with his 64 replaced by gln, leu 29 replaced by phe)).

Następnie przy pomocy narzędzia blastp wyszukaliśmy sekwencji homologicznych do sekwencji z pliku .fasta dla struktury 1MCY pobranego z bazy PDB.

**Wyniki i ich interpretacja**





Wynikiem działania narzędzia jest pięć najlepszych dopasowań. Jeśli jest taka możliwość, każde dopasowanie pochodzi z innego organizmu. Rezultaty przedstawione są w postaci drzewa filogenetycznego oraz graficznej (na schemacie po lewej stronie). Dodatkowo SmartBLAST pokazuje dopasowania z bazy konserwatywnych domen. Odniesienia w środkowej części przenoszą do sekcji *alignments,* gdziemożemy uzyskać informacje jakie?

Ogólny graficzny zarys wyników (po lewej) należy odczytywać następująco: na zielono zaznaczone są dopasowania z landmark database, dopasowania z nieredundantnej bazy białek określa kolor niebieski, sekwencja wprowadzona na wejściu zaznaczona jest kolorem żółtym. The presentation grays out a range of a subject sequence to indicate that BLAST did not align the query there. It represents a deletion according to the multiple alignment by leaving a range on a graphical overview bar white. Biała część słupka wskazuje obszar delecji wykryty zgodnie z MSA, szara – brak dopasowania (zarówno przez blastp jaki i multiple aligment)

W sekcji *descriptions* w formie tabelarycznej zebrane zostały najważniejsze informacje o dopasowaniach.

Max score – Najwyższy wynik dopasowania. Im wyższy wynik, tym lepsze dopasowanie.

Total score – Suma wyników dopasowania. Im wyższy wynik, tym lepsze dopasowanie.

Max score – najwyższy score dopasowania.

Total score – suma scorców z dopasowania z wszystkihc segmentów sekwencji. Im wyższy wartość, tym lepsze dopasowanie.

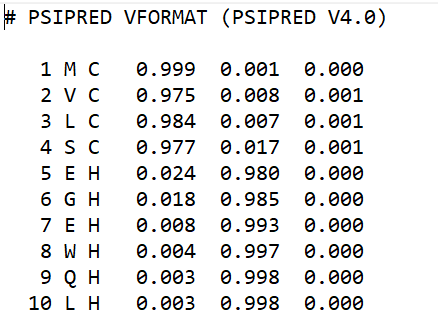
Coverage – określa procentową ilość sekwencji badanej, która pasuje/ pokrywa się z sekwencją z bazy NCBI. Wysoki coverage wskazuje, że duża część sekwencji badanej pasuje do sekwencji docelowej.

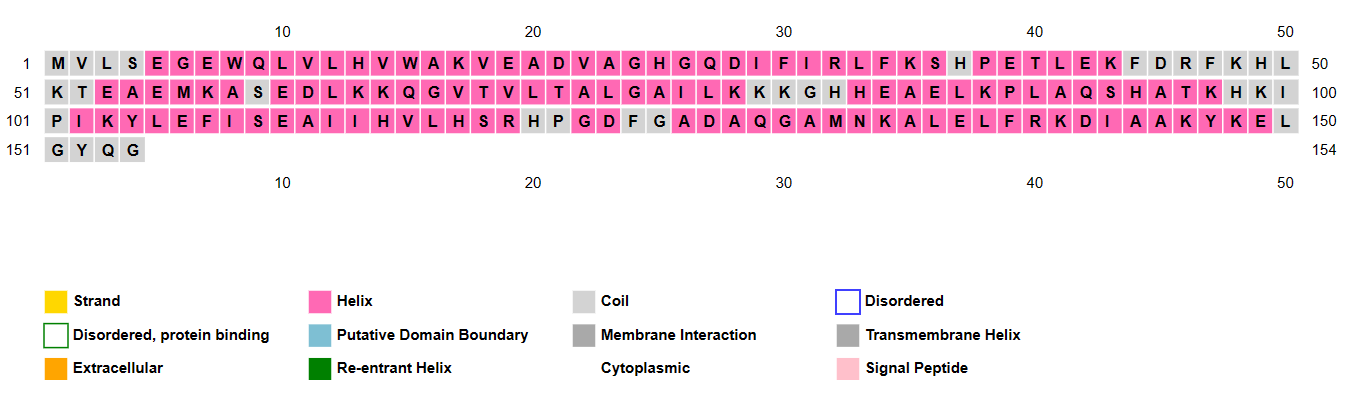
e-value – oczekiwana liczba pozytywnie fałszywych wyników w porównaniu do bazy danych. Im wartość bliższa 0, tym wynik porównania sekwencji jest bardziej istotny.

ident – % zasad, które są identyczne z genomem referencyjnym.

Wnioski co do naszych dopasowań.

III. STRUKTURA DRUGORZĘDOWA BIAŁEK

Pierwsza kolumna to indeks aminokwasu, kolejna to jednoliterowe oznaczenie aminokwasu. W kolumnie trzeciej znajduje się informacja o strukturze drugorzędowej. Ostatnie trzy kolumny to kolejno prawdopodobieństwo struktury: coil, helix, sheet.

Zgodnie z wynikami struktura drugorzędowa badanego białka składa się z naprzemiennie występujących: helisy i coil. Prawdopodobieństwa tych struktur drugorzędowych w środku sekwencji aminokwasowej dość mocno spadają ( momentami nawet <50%). Poza tym utrzymują się na dość wysokim poziomie (min. 89%).   
Poniżej znajduje się schemat, który graficznie przedstawia strukturę drugorzędową.

6. Charakterystyka nazwa kolumn wykorzystywana do opisu aminokwasów wchodzących w skład analizowanej struktury  
RESIDUE –  
AA – jednoliterowy kod aminokwasowy, dla mostków disiarczkowych cysteiny używane są kolejne małe litery alfabetu, dla obu reszt.  
STRUCTURE – informacje o strukturze wtórnej białek na podstawie analizy współrzędnych atomów. W wyniku można otrzymać oznaczenia takie jak: H dla α-helis, E dla β-kartek.  
BP1, BP2 – numery pierwszej i drugiej reszty aminokwasowej pomiędzy którymi występują mostki wodorowe.  
ACC – dostępność powierzchniowego obszaru dla danego aminokwasu w analizowanym fragmencie białka (mierzona w Å2). Dla aminokwasów wewnętrznych wartość jest niska, ponieważ aminokwasy nie mają dostępu do otaczającej je wody ani innych czynników środowiskowych. Aminokwasy na powierzchni będą miały wyższą wartość ACC ze względu na większy dostęp do otoczenia.   
N-H-->O – odległość między atomem azotu w jednym aminokwasie a atomem tlenu w innym (związane z mostkami wodorowymi).  
O-->H-N – ta kolumna zawiera informacje o atomie tlenu w jednym aminokwasie, który połączony jest mostkiem wodorowym z atomem wodoru związanym z atomem azotu w innym aminokwasie.   
N-H-->O – analogicznie do poprzednich dwóch kolumn; mostki wodorowe, w których atom azotu jest donorem wodoru a atom tlen - akceptorem  
O-->H-N – atom azotu jest donorem wodoru, atom tlenu – akceptorem.   
TCO – określa chwilowy stopień skręcenia α-helis w analizowanym fragmencie białka.  
KAPPA – kąt pomiędzy trzema atomami Cα reszt I-2, I, I+2 (I – dowolna reszta). Używane do definiowania zakręcenia α-helisy.   
ALPHA – kąt torsyjny zdefiniowany na czterech atomach Cα reszt (I-1, I, I+1, I+2). Używane do definiowania chiralności.   
PHI, PSI – kąty torsyjne aminokwasów w danej sekwencji.  
X-CA, Y-CA, Z-CA – współrzędne atomu Cα aminokwasu w danej sekwencji.

7. Znaczenie symboli opisujących strukturę drugorzędową białek:

C – coil  
H – α-helix  
T– β-turn   
S – bend  
G – 310 -helix  
I – π-helix  
E – β-strand  
B – β-bridge

C. Porównaj ze sobą powyższe struktury drugorzędowe (2D) białek, a mianowicie

przewidzianą na podstawie znanej sekwencji (A) z wyekstrahowaną z referencyjnej struktury

3D (B) poprzez wyznaczenie wartości współczynnika identyczności (SSI). Czy porównywane

struktury 2D są do siebie podobne? Odpowiedź uzasadnij

Seq – badana sekwencja aminokwasowa

SS – predykcja DSSP

PSI – predykcja PSIPRED

Na zielono zaznaczone zostały niedopasowania

|  |  |  |
| --- | --- | --- |
| Seq: | MVLSEGEWQL VLHVWAKVEA DVAGHGQDIF IRLFKSHPET LEKFDRFKHL |  |
| SS: | **ccccHHHHHH** **HHHHHHHHGG** **GHHHHHHHHH** **HHHHHHcGGG GGGcTTTTTc** |  |
| PSI: | **CCCCHHHHHH HHHHHHHHHH HHHHHHCHHH HHHCCCCCCC CCHHHHHHCC** |  |
|  |  |  |
|  |  |  |
| Seq: | KTEAEMKASE DLKKQGVTVL TALGAILKKK GHHEAELKPL AQSHATKHKI |  |
| SS: | **cSHHHHHHcH** **HHHHHHHHHH** **HHHHHHHHTT** **TccHHHHHHH** **HHHHHHTScc** |  |
| PSI: | **CCHHHHHHCH HHHHHHHHHH HHHHHHHHCC CCHHHHHHHH HHHHHHHCCC** |  |
|  |  |  |
|  |  |  |
| Seq: | PIKYLEFISE AIIHVLHSRH PGDFGADAQG AMNKALELFR KDIAAKYKEL |  |
| SS: | **cHHHHHHHHH** **HHHHHHHHHc** **GGGcSHHHHH** **HHHHHHHHHH** **HHHHHHHHHH** |  |
| PSI: | **CHHHHHHHHH HHHHHHHHHC CHHCCHHHHH HHHHHHHHHH HHHHHHHHHC** |  |
|  |  |  |
|  |  |  |
| Seq: | GYQG |  |
| SS: | **Tccc** |  |
| PSI: | PSI:CCCC |

Dopasowania: 154-32 = 122

Wysoki współczynnik identyczności oznacza, że większość predykcji z obu metod pokrywa się. Jak odnosi się to do naszych wyników. Narzędzia dają podobne wyniki