

Zadanie polega na odczytaniu danych z pliku typu FASTA, który zawiera sekwencję nukleotydów, następnie należy odnaleźć przedziały, w których początkiem zawsze jest trójka nukleotydów kodująca kodon START natomiast końcem jedna z trzech trójek kodujących kodon STOP. Ostatnim krokiem jest zamiana pozostałych trójek znajdujących się wewnątrz przedziału na odpowiadające im trzyliterowe skróty aminokwasów.

Do programu powinniśmy wprowadzać tylko i wyłącznie dane z plików typu FASTA składających się jedynie z nukleotydów DNA, ponieważ pliki FASTA mogą również zawierać sekwencję nukleotydów RNA jak i sekwencję składających się z jednoliterowych skrótów aminokwasów. Program oczekuje odpowiedniego formatowania danych, które jest specyficzne dla plików typu FASTA tj. nagłówek który zawsze rozpoczyna się od znaku ">" oraz sekwencji nukleotydów znajdujących się bezpośrednio pod nim.

Dane przetworzone przez program zostają zapisane w oddzielnym pliku w takim samym porządku co dane pobrane, to znaczy że zawsze otrzymamy komplet składający się z nagłówka opisującego dany gen i jego sekwencję oraz samej sekwencji, jednak zamiast zamiast sekwencji składającej się z liter ATCG otrzymamy trzyliterowe skróty aminokwasów zaczynających się od słowa start a kończących się na słowie stop.

Założenie programu jest proste: pobrać dane wejściowe i przetworzyć je na skróty aminokwasów jednak aby to wykonać program odczytuje kod linijka po linijce i na podstawie pierwszego odczytanego znaku z danej linijki decyduje co wykonać. Jeżeli natrafia na linie nagłówkową mają za zadanie jedynie przepisać do pliku bez żadnej edycji natomiast w innym wypadku ma przepisać wszystkie jej litery do tablicy w celu dalszej obróbki.

Posiadając całą tablicę składającą się z liter możemy przejść do wyszukiwania trójki kodującej kodon start, w tym celu sprawdzamy pozycje po pozycji litery ponieważ jeżeli pierwsza trójka składająca się z pozycji 0,1,2 nie koduje startu to nie możemy od razu przejść do następnej trójki 3,4,5 ponieważ kodon start może znajdować się na pozycjach 1,2,3.

Dopiero gdy natrafimy na kodon start możemy zacząć dzielić tablicę na trójki i przypisywać im odpowiednie skróty aminokwasów zapisując je do pliku. Program pracuje w pętli wypisując trójki aminokwasów aż do momentu natrafienia na kodon stop i zapisaniu słowa stop do pliku oraz powraca do wyszukiwania następnego trójki oznaczającej start.

Łukasz Michał