

# Implementacja algorytmu Nussinov do obliczania struktury drugorzędowej DNA i RNA

---

Jerzy Balcerzak, Oleg Druzhynets

June 10, 2014

## 1 CEL PROJEKTU

Celem projektu była implementacja programu obliczającego strukturę drugorzędową cząsteczki DNA lub RNA wykorzystując algorytm Nussinov.

## 2 UŻYTE TECHNOLOGIE

Stworzony program został całkowicie przygotowany w języku Python w standardowej wersji 2.7. Wybór padł właśnie na ten język programowania przede wszystkim ze względu na jego obiektowy charakter, przejrzystość składni oraz bogactwo bibliotek.

Duża część kodu głównego algorytmu oraz modułu obsługi wejścia została przygotowywana w duchu metodyki Test Driven Development (TDD). Świetnym narzędziem wykorzystanym podczas prac okazał się więc framework pyUnit.

Ponadto, do testów gotowego już programu wykorzystano: ... Wizualizacja otrzymanych wyników została oparta o bibliotekę networkx.

## 3 OSIĄGNIĘTA FUNKCJONALNOŚĆ APLIKACJI

W ramach projektu stworzono prostą w obsłudze aplikację konsolową (nussinovCalculator.py). Jej funkcjonalność można opisać następująco:

- Pomoc dla użytkownika - wywołanie programu z parametrem -h wyświetla pulę obsługiwanych opcji oraz ich domyślne parametry.

- Wejście - wywołanie programu z opcją -i oraz podaną nazwą pliku (np. ./nussinov-Calculator.py -i "NazwaPliku.txt") spowoduje odczytanie z pliku ciągów nukleotydów (można podać ich wiele w kilku liniach) i przeprowadzenie algorytmu nussinov dla każdego z nich. Program można wywołać bez tej opcji (wtedy domyślnie odczytywane są dane z pliku "input.txt"), jak również podać ciąg nukleotydów prosto do konsoli używając opcji -r (np. -r GGGAAAACCC)
- Wyjście - domyślnie program wypisuje wynik na konsolę oraz zwraca plik "output.txt" w postaci listy par indeksów nukleotydów które będą połączone w strukturze drugorzędowej. Używając opcji -o można wskazać plik do którego dane mają być zapisane (np. -o outputs/out1.txt).
- Parametryzacja algorytmu - przy uruchamianiu programu użytkownik może dodatkowo wprowadzić:
  - minimalną ilość nukleotydów w pętli używając opcji -m (domyślnie: -m 4),
  - oraz wartości w macierzy energii (opcja -e) w formacie listy list (np. -e [[0,0,0,2],[0,0,3,0],[0,3,0,3],[2,0,1,0]] ). Przy czym kolejne kolumny i wiersze odnoszą się do nazw nukleotydów w kolejności alfabetycznej - A, C, G, U.
- Wizualizacja

## 4 OPIS ZASTOSOWANEGO ALGORYTMU

Prezentowany program rozwiązuje następujące zadanie: "Określ strukturę drugorzędową dla danej sekwencji nukleotydów". Do rozwiązania tego problemu użyto algorytmu zaproponowanego przez profesor Nussinov [1][2] opierającego swoje działanie na metodyce programowania dynamicznego. Ogólnikowo, przebieg algorytmu przedstawiają się następująco:

- Przygotowanie macierzy Nussinov oraz uzupełnienie jej wartości - implementacja tej części w głównej mierze została zainspirowana informacjami przekazanyymi w wykładzie dr Roberta Nowaka [3])
- Przejście zwrotne po macierzy w celu znalezienia par nukleotydów, tzw. Traceback. Istnieją różne możliwości implementacji tej procedury - w projekcie została wykorzystana jej elegancja wersja rekurencyjna (przedstawiona w materiałach udostępnionych przez uniwersytet w Tübingen [4])

## 5 PRZEPROWADZONE TESTY

## 6 PORÓWNANIE Z DZIAŁANIA Z APLIKACJI

W celu dodatkowego potwierdzenia skuteczności naszego programu porównaliśmy jego działanie z komercyjnie dostępnym narzędziem: RNA Secondary Structure Prediction and Analysis on the web [5]

Porównywanie wyników predykcji struktury drugorzędowej sekwencji nukleotydów RNA pokazało, iż zaimplementowany przez nas algorytm Nussinov działa poprawnie. W porównaniu do wyżej wymienionego odpowiednika działa szybciej nasz program działa szybciej ponieważ nie przeprowadza znacznie prostszą analizę danych (przeprowadza jedynie predykcję struktury drugorzędowej). Przygotowana przez nas wizualizacja wyników również pozostawia wiele do życzenia z porównaniem do swojego odpowiednika.

## 7 WIZUALIZACJA WYNIKÓW

## 8 WNIOSKI

Jakieś wnioski

### Literatura i źródła

- [1] Nussinov, Ruth; Pieczenik, George; Griggs, Jerrold R.; Kleitman, Daniel J. (1 July 1978). "Algorithms for Loop Matchings". *SIAM Journal on Applied Mathematics* 35 (1): 68–82. doi:10.1137/0135006.
- [2] Nussinov, R; Jacobson, AB (Nov 1980). "Fast algorithm for predicting the secondary structure of single-stranded RNA.". *Proceedings of the National Academy of Sciences of the United States of America* 77 (11): 6309–6313. doi:10.1073/pnas.77.11.6309. PMC 350273. PMID 6161375.
- [3] Dr Robert Nowak. Slajdy do wykładu "Metody Bioinformatyki (MBI)". Politechnika Warszawska, <http://staff.elka.pw.edu.pl/~rnowak2/dyd/mbi2014L/index.html>
- [4] Prof. Dr. Daniel Huson, Dr. Christoph Dieterich . "Script for Algorithms in Bioinformatics I Lecture". The University of Tübingen, [http://ab.inf.uni-tuebingen.de/teaching/ws06/albi1/script/protstruct2d\\_20Dec2006.pdf](http://ab.inf.uni-tuebingen.de/teaching/ws06/albi1/script/protstruct2d_20Dec2006.pdf)
- [5] RNA Secondary Structure Prediction and Analysis on the web <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>
- [6] RNA STRAND v2.0 - The RNA secondary STRucture and statistical ANalysis Database, <http://www.rnasoft.ca/strand/>