

Uniwersytet Jagielloński
Wydział Matematyki i Informatyki
INSTYTUT INFORMATYKI I MATEMATYKI KOMPUTEROWEJ
Studia dzienne

Nr indeksu: 1104482

Marta Szynczewska

Zbiór i ocena jakości i aktualności narzędzi do przewidywania struktury drugorzędowej RNA

Opracowano zgodnie z Ustawą o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. 1994 nr 24 poz. 83) wraz z nowelizacją z dnia 25 lipca 2003 r. (Dz.U. 2003 nr 166 poz. 1610) oraz z dnia 1 kwietnia 2004 r. (Dz.U. 2004 nr 91 poz. 869)

Kraków 2016

Spis treści

Wstęp teoretyczny	3
RNA – definicja	3
Struktury RNA	3
Metody	5
Afold	5
Carnac	5
CentroidFold	7
CentroidHomfold	9
ContraFold	10
DotKnot	11
HotKnots	11
TurboKnot	11
Opcje narzędzi	12
Porównanie	15
Podsumowanie	16
Bibliografia	17

Wstęp teoretyczny

RNA – definicja

RNA, zwane inaczej kwasem rybonukleinowym, to długa liniowa makrocząsteczka polinukleotydowa. Nukleotydy połączone są wiązaniami fosfodiesterowymi 3' → 5'. Każdy nukleotyd zbudowany jest z jednostki cukrowej, przynajmniej jednej reszty fosforanowej oraz zasady azotowej. W przypadku RNA resztą cukrową jest ryboza oraz wyróżniamy cztery podstawowe zasady azotowe: adeninę oznaczaną literą A, guaninę (G), cytozynę (C) i uracyl (U).

(wzory strukturalne zasad azotowych i rybozy)

Cząsteczki RNA występują głównie w postaci pojedynczej nici, jednakże łańcuch RNA może się zwinąć i stworzyć strukturę *spinki do włosów* o budowie dwuniciowej helisy. W tych strukturach adenina tworzy pary z uracylem, a guanina z cytozyną.

Struktury RNA

a) Pierwszorzędowa

Strukturę pierwszorzędową RNA definiujemy jako ciąg nukleotydów ułożonych jeden po drugim.

(grafika przedstawiająca sekwencję nukleotydów)

b) Drugorzędowa

Strukturę drugorzędową RNA definiujemy jako położenie ciągu nukleotydów na płaszczyźnie. Strukturami, jakie możemy zauważyć są m.in. spinki do włosów, wybrzuszenia, pętle wewnętrzne, węzły, pseudowęzły oraz odcinki dwuniciowe.

Spinka do włosów (hairpin) - struktura składająca się z części dwuniciowej oraz pętli zewnętrznej.

(rysunek)

Wybrzuszenie (bulge loop) - struktura składająca się z przynajmniej dwóch niesparowanych nukleotydów tylko na jednej z nici.

Pętla wewnętrzna (internal loop) - struktura składająca się z niesparowanych nukleotydów na obu niciach pomiędzy dwoma odcinkami podwójnej helisy.

(rysunek)

Węzeł (junction) - rozgałęzienie przynajmniej trzech odcinków podwójnej helisy.

(rysunek)

Pseudowęzeł (pseudoknot) - struktura opierająca się na oddziaływaniach pomiędzy nukleotydami wchodzącymi w skład innej struktury, np. spinki do włosów, a innymi nukleotydami.

(rysunek)

Często zdarza się, że zasady nie tworzą idealnych par typu Watsona-Cricka. Uracyl może tworzyć parę z guaniną, ponieważ występują wiązania wodorowe pomiędzy N3 uracylu i C6 guaniny oraz C2 uracylu i N1 guaniny. Jednakże para guaniny i cytozyny jest od niej silniejsza.

Najbardziej znanym przykładem struktury drugorzędowej RNA jest tRNA, które wyglądem przypomina liść koniczyny.

(struktura tRNA)

Metody

Afold

- serwer ftp: <ftp://ftp.ncbi.nlm.nih.gov/pub/ogurtsov/Afold/>
- metoda szybciej konstruuje pętle wewnętrzne niż Mfold i ZUKER,
- powstanie: 2003-2006r.
- autor: Aleksey Ogurtsov



Carnac

- strona internetowa: <http://bioinfo.lifl.fr/RNA/carnac/>
- metoda pozwalająca na przewidywanie konserwatywnych elementów struktury drugorzędowej rodzin homologicznych niekodującego RNA,
- metoda opiera się na termodynamicznej minimalizacji energii

CentroidFold

- strona internetowa: <http://rtools.cbrc.jp/centroidfold/>
- metoda bardzo precyzyjna, ponieważ oparta jest na uogólnionym estymatorze.

CentroidFold

CentroidFold based on a generalized centroid estimator is one of the most accurate tools for predicting RNA secondary structures. The predicted secondary structure is coloured according to base pairing probabilities.

Enter your RNA sequence (single FASTA format, <= 400nt)

Samples: [Hammerhead ribozyme](#), [tRNA](#), [H/ACA snoRNA](#), [TPP-riboswitch](#)

```
>TPP-riboswitch, AF269819/18111669
CCUACUAGGGGAGCCAAAGGCUGAGAUAAUGUAUUCAGACCCUUAUAACCUGAUUUGGUUAAUACCAACGUAGGAAAGUAGUUUAUUAACUAUUCG
UCAUUGAGAUUGUCUUGGCUAACUACUUCUUCGCGGGGAAGUAGUU
```

Submit

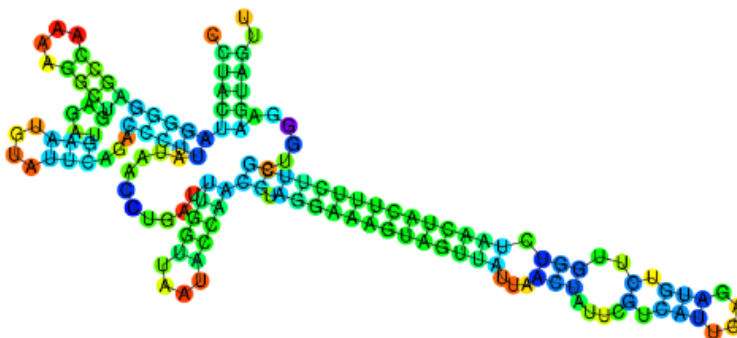
Options

McCasikill(BL) : specify the inference engine

2^2 : weight of base pairs

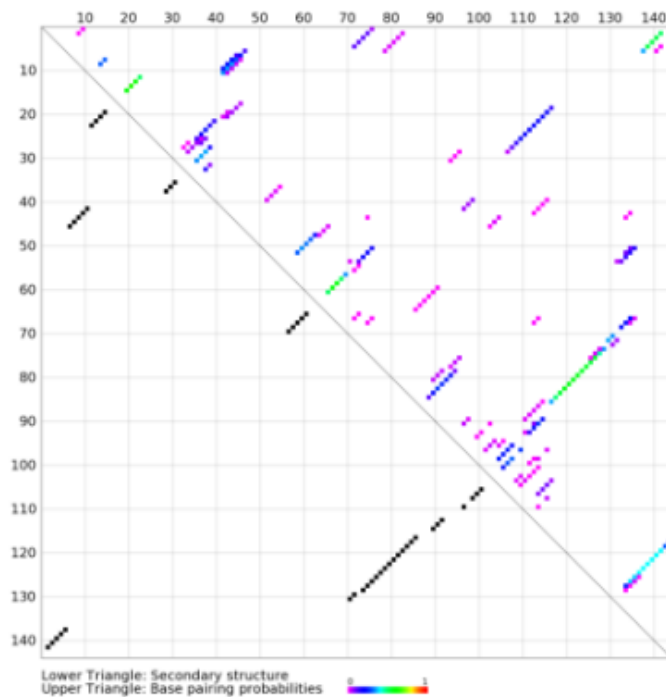
- strona wynikowa:

Secondary structure



Download: [TEXT](#) [PNG](#) [PDF](#) [EPS](#)

Base pairing probability plot



Download: [TEXT](#) [PNG](#) [PDF](#) [EPS](#)

Download:TEXT (Secondary structure) is a dot-bracket format with a fasta-like header line, indicating a secondary structure.

```
>sequence_name
ACGCUGUCUGUACUUGUAUCAGUACACUGACGAGUCCUAAAGGACGAAACAGCGC
.(((((((.(((((((.....)))))).....((((.....)))).....))))). (g=4,th=0.2,e=-9.14)
```

In this format, each dot represents an unpaired base, opening and closing brackets represent a base pair.

Download:TEXT (Base pairing probability plot) contains base-pairing probabilities more than "weight of base pairs" in blank delimited format.

```
pos1 nt1 pair-pos11:prob11 pair-pos12:prob12 ...
pos2 nt2 pair-pos21:prob21 pair-pos22:prob22 ...
;
```

pos1,pos2 are sequence positions in 1-based coordinate.

nt1,nt2 are nucleotides.

pair-pos11,pair-pos12 ... are sequence positions of pairing partners in 1-based coordinate.

pron11,prob12 ... are base pairing probabilities.

- jak widać na powyższych ilustracjach mamy do wyboru wiele formatów zapisu do pliku: TEXT, jako obrazek PNG, PDF lub z rozszerzeniem .eps.

CentroidHomfold

- strona internetowa: <http://rtools.cbrc.jp/centroidhomfold/>
- metoda automatycznie pobiera sekwencje homologiczne do sekwencji docelowej.
- sekwencje homologiczne pobierane są z bazy Rfam,
- pobieranie sekwencji homologicznych daje CentroidHomfold przewagę nad CentroidFold w dokładności przewidywania struktury.

CentroidHomfold

CentroidHomfold predicts RNA secondary structures by employing automatically collected homologous sequences of the target. Homologous sequences are collected from [Rfam](#) using [LAST](#). If homologous sequences are available, CentroidHomfold can predict secondary structures for the target sequence more accurately than [CentroidFold](#) using homologous sequence information with the probabilistic consistency transformation for base-pairing probabilities.

Enter your RNA sequence (single FASTA format, <= 400nt)

Samples: [Hammerhead ribozyme](#), [tRNA](#), [H/ACA snoRNA](#), [TPP-riboswitch](#)

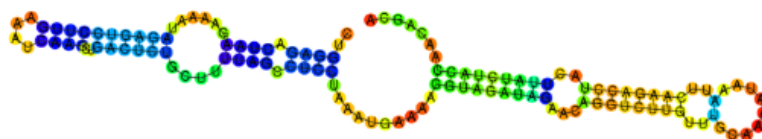
```
>A7609485.1/1-127
CUGGAGACUAGAAAAUAGAGUCCUUGAAAUAAGCUGACUCUGCUUUUAGCCUCCUAAAUGAAAAGGUAGAUAGAACAGGUCUUGUUUGCAAAUA
AAUUAAGACCUACUUAUCUACCAACAGCA
```

Submit

Options

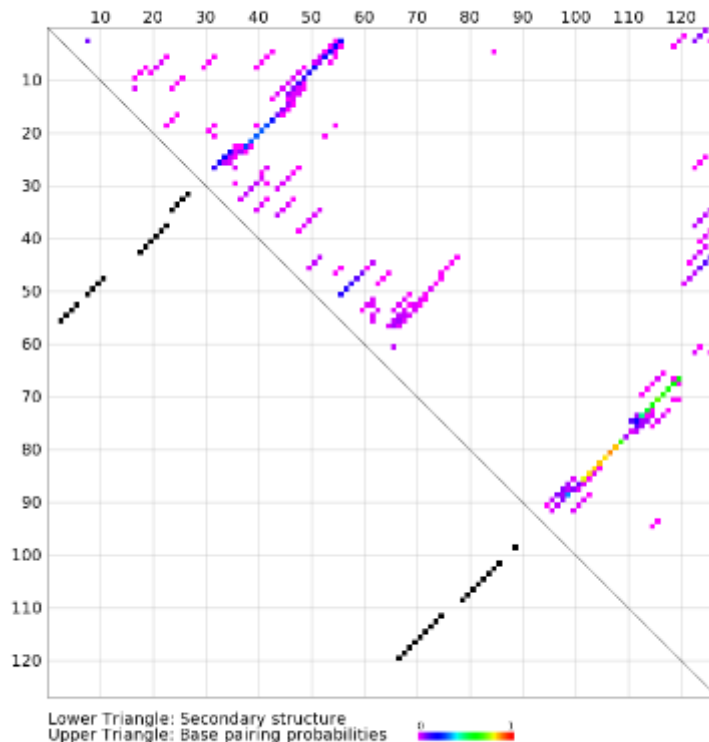
- : specify the inference engine for secondary structure
- : specify the inference engine for pairwise alignment
- : weight of base pairs
- : E-value for homology search against [Rfam](#)
- : number of homologous sequences

- strona wynikowa:
Secondary structure



Download: [TEXT](#) [PNG](#) [PDF](#) [EPS](#)

Base pairing probability plot



Download: [TEXT](#) [PNG](#) [PDF](#) [EPS](#)

ContraFold

- strona internetowa: <http://contra.stanford.edu/contrafold/>
- zaimplementowany w C++,
- metoda oparta warunkowych modelach logarytmiczno-liniowych, probabilistycznych modelach uogólniających stochastyczną gramatykę bezkontekstową (SCFGs) poprzez funkcje wyższych punktacji,
- metoda osiąga wysoką dokładność predykcji pojedynczego ciągu,
- metoda łączy techniki probabilistyczne z fizycznymi. Jest to odpowiedź na lukę pomiędzy metodami probabilistycznymi a termodynamicznymi udowadniając, że procedury uczenia statystycznego mogą stanowić skuteczną alternatywę dla empirycznego pomiaru parametrów termodynamicznych do przewidywania struktury drugorzędowej RNA,
- wyniki można przesłać sobie bezpośrednio na wcześniej podanego maila lub otworzyć je w przeglądarce od razu po wykonaniu,
- mamy dwie opcje do wyboru pod względem parowania się nukleotydów. Możemy wybrać opcję domyślną typu Watsona-Cricka lub dopuścić parowanie się wszystkich możliwych.

DotKnot

- strona internetowa: <http://dotknot.csse.uwa.edu.au/>

HotKnots

- strona internetowa: <http://www.cs.ubc.ca/labs/beta/Software/HotKnots/>
- metoda heurystyczna,

TurboKnot

- strona internetowa: <http://rna.urmc.rochester.edu/>
- dostępny jako część RNAstructure

Opcje narzędzi

Nazwa metody	Dostępność	Dostęp	Systemy operacyjne, na których działa	Input	Output	Struktury	Język	Ograniczenia
Afold	darmowe oprogramowanie dostępne dla wszystkich	przez serwer ftp	Linux, Windows	FASTA lub format Genbank	plik „Afold.res”	pętle wewnętrzne	C/C++	do 28 000 nt, min. 2GB RAM
Carnac	darmowe oprogramowanie dostępne dla wszystkich	online (web serwer) lub lokalnie na komputerze	Linux, Windows	FASTA	CT, PostScript, JPEG, bracket notation, list of stems, archiwum .zip	elementy struktury drugorzędowej rodzin homologicznych niekodującego RNA		do 300 nt (kilka sekund) do 2000 nt (kilka min)
CentroidFold	darmowe oprogramowanie dostępne dla wszystkich	online	Linux, Windows	FASTA	TEXT, PNG, PDF, EPS	wszystkie elementy struktury drugorzędowej		do 400 nt
CentroidHomfold	darmowe oprogramowanie dostępne dla wszystkich	online	Linux, Windows	FASTA	TEXT, PNG, PDF, EPS	wszystkie elementy struktury drugorzędowej		do 400 nt
ContraFold	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na komputerze	Linux	sekwencja RNA w formacie FASTA	BPSEQ		C++	do 1000 nt

Nazwa metody	Dostępność	Dostęp	Systemy operacyjne, na których działa	Input	Output	Struktury	Język	Ograniczenia
DotKnot								
HotKnots							C/C++	do 500 nt
TurboKnot								
RNAstructure			Windows, Linux, Mac OS-X				C++ oraz JAVA	

Nazwa metody	Dostępność	Dostęp	Systemy operacyjne, na których działa	Input	Output	Struktury	Język	Ograniczenia

Porównanie

Podsumowanie

Bibliografia

Lubert Stryer *Biochemia* Wydawnictwo Naukowe PWN (1999)

Suzanne Clancy, Ph.D. *Chemical Structure of RNA* Nature Education (2008)

Aleksey Y. Ogurtsov, Svetlana A. Shabalina, Alexey S. Kondrashov, Mikhail A. Roytberg
Analysis of internal loops within the RNA secondary structure in almost quadratic time (2006)

Do, C.B., Woods, D.A., and Batzoglou, S. (2006) CONTRAfold: RNA Secondary Structure Prediction without Energy-Based Models. *Bioinformatics*