

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

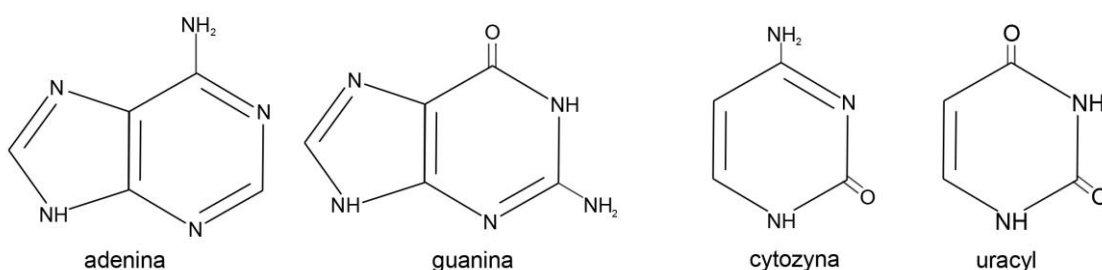
Spis treści

Wstęp teoretyczny	3
RNA – definicja	3
Struktury RNA	3
Metody	6
Afold	6
Carnac	7
CentroidFold	8
CentroidHomfold	10
ContraFold	12
DotKnot	13
HotKnots	13
TurboKnot	16
CMfinder	17
CRWrnaFold	18
CyloFold	19
Mastr	19
MaxExpect	21
PPfold	24
RNAwolf	31
Sfold	32
taveRNA	33
TurboFold	34
Opcje narzędzi	35
Podsumowanie	39
Bibliografia	40

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).



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

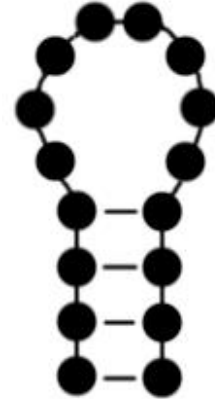
Strukturę pierwszorzędową RNA definiujemy jako ciąg nukleotydów ułożonych jeden po drugim.



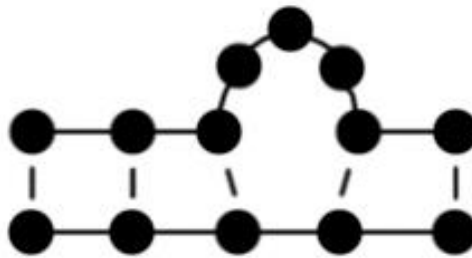
b) Drugorzędowa

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.

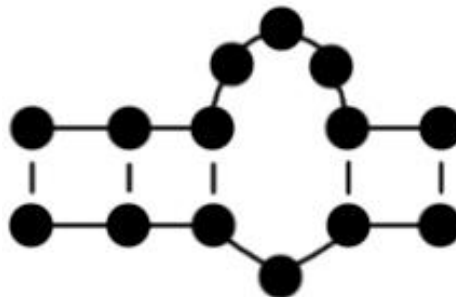
Spinka do włosów (hairpin) - struktura składająca się z części dwuniciowej oraz pętli zewnętrznej.



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



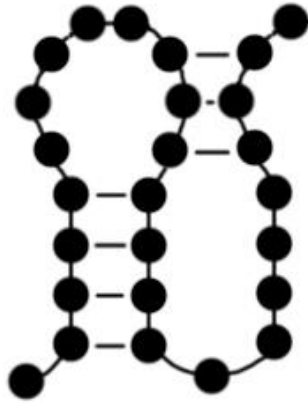
Pętla wewnętrzna (internal loop) - struktura składająca się z niesparowanych nukleotydów na obu niciach pomiędzy dwoma odcinkami podwójnej helisy.



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

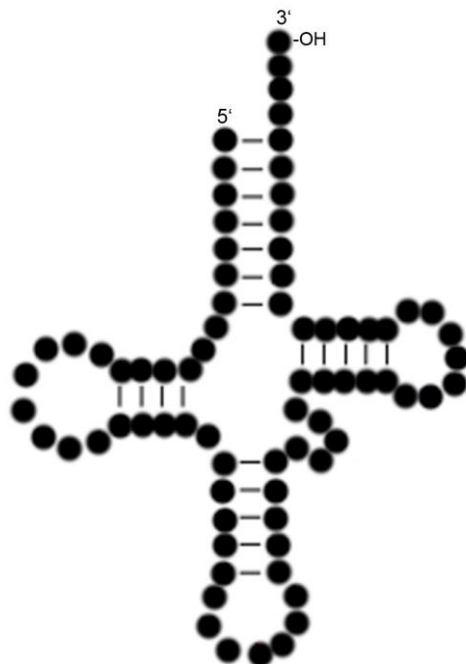


Pseudowęzeł (pseudoknot) - struktura opierająca się na oddziaływaniach pomiędzy nukleotydami wchodzącymi w skład innej struktury, np. spinki do włosów, a innymi nukleotydami.



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.



Metody

Afold

Metoda Afold powstała między 2003 a 2006 rokiem. Autorem narzędzia jest Aleksey Ogurtsov. Służy ona głównie do przewidywania struktury drugorzędowej pętli wewnętrznych RNA dla sekwencji o długości nie większej niż 28 000 nukleotydów. Przetwarzanie tak długiej sekwencji przy użyciu tego algorytmu trwa ok. 28 godzin na komputerze z 2GB RAM.

W algorytmie wykorzystywany jest model najbliższego sąsiada NNM (nearest-neighbor model) z założeniem, że kara dla pętli wewnętrznej zależy od dwóch zmiennych:

$$F(s,d) = f_{\text{Len}}(s) + f_{\text{Diff}}(d),$$

gdzie s jest ilością niesparowanych nukleotydów w pętli, a d jest różnicą w długości pomiędzy dwoma niesparowanymi regionami tworzącymi pętlę. W celu oceny poprawności pętli, algorytm korzysta z listy kandydatów.

Złożoność czasowa algorytmu jest równa $O(M \cdot \log^2 L)$, gdzie L oznacza długość sekwencji. W porównaniu z Mfold oraz ZUCKER, Afold wypada znacznie lepiej. Przede wszystkim jest szybszy już w etapie wypełniania macierzy oraz nie posiada ograniczenia ilości nukleotydów w pętli.[4]

Afold jest dostępny za darmo do pobrania z serwera ftp: <ftp://ftp.ncbi.nlm.nih.gov/pub/ogurtsov/Afold/>. Po przejściu na serwer otrzymujemy widok jak przedstawiony obok.

Mamy do wyboru pliki instalacyjne na system Windows i Linux. Jednakże w obecnej chwili pliki dedykowane dla Windows są niekompletne.



The screenshot shows a web browser window displaying the FTP directory listing for the Afold software package. The address bar shows the URL <ftp://ftp.ncbi.nlm.nih.gov/pub/ogurtsov/Afold/>. The page title is "Indeks – /pub/ogurtsov/Afold/". The table below lists the files and directories available in this directory.

Nazwa	Rozmiar	Data modyfikacji
[katalog główny]		
Linux/		28.01.2006, 00:00:00
RNA_Intern_Loops-Suppl.doc	516 kB	11.04.2006, 00:00:00
RNA_Intern_Loops-Suppl.pdf	360 kB	11.04.2006, 00:00:00
ReadMe.txt	3.4 kB	11.04.2006, 00:00:00
Windows/		04.01.2007, 00:00:00

Ryc.1. Zawartość serwera ftp.

Carnac

Głównym celem metody Carnac jest przewidywanie konserwatywnych elementów struktury drugorzędowej rodzin homologicznych niekodującego RNA.[38] W metodzie tej nie jest wymagane wcześniejsze dopasowanie sekwencji, dlatego idealnie nadaje się do danych posiadających niskie podobieństwo struktur pierwszorzędowych. Trzy podstawowe komponenty metody to: termodynamiczna minimalizacja energii, dopasowanie filogenetyczne oraz konserwacja sekwencji. Dodatkowo wykluczamy wyizolowane pary i pseudowęzły.

Algorytm składa się z trzech kroków:

Krok pierwszy to wyłonienie wszystkich potencjalnych łodyg o niskim poziomie energii swobodnej dla wszystkich sekwencji. Poszukiwanie jest realizowane za pomocą programowania dynamicznego.

Krok drugi polega na analizie wszystkich możliwych par sekwencji i tworzeniu struktury złożonej właśnie z danej pary. Aby dwie sekwencje uznać za parę muszą one spełnić warunki. W obu sekwencjach powinny znajdować się analogiczne dobrze zakonserwowane regiony oraz muszą zawierać co najmniej jedną mutację kompensacyjną.

Krok trzeci to tworzenie grafu macierzystego złożonego na podstawie relacji między najbardziej wiarygodnymi sekwencjami.

Istnieją dwa sposoby na użycie Carnac: pobranie i instalacja oprogramowania na własnym komputerze lub online na stronie: <http://bioinfo.lifl.fr/RNA/carnac/>. Obie wersje są darmowe, bez rejestracji. Dane wejściowe muszą być podane w formacie FASTA, natomiast wynik otrzymujemy w kilku formatach: Connest file (CT), PostScript oraz w formie graficznej. Każdemu procesowi zostaje nadany unikalny numer ID, pod którym zapisywane są operacje. Czas oczekiwania zależy od długości sekwencji. Jeśli długość jest mniejsza niż 300 nukleotydów to zajmuje to kilka sekund, dla większych danych czas wydłuża się do kilku minut.[6]

The screenshot shows the Carnac web interface. At the top, there's a logo for 'Bonsai Software' and the name 'carnac'. Below it are navigation links: 'home', 'web server', 'help', 'examples', and 'retrieve result with an ID'. On the left, there's a sidebar with categories: DNA (YASS, Magnolia, mrepS, ProCARS), HTS (SortMeRNA, CRAC, VidiJ, StORM), RNA (Carnac, RNAfamily, Gardemia, Regliss, CG-seq, RNAspace), Proteins (Path, Protea, ReBLOSUM), and TFM (TFM-Explorer, TFM-Scan, TFM-Pvalue, TFM-CUDA). Below the sidebar, there's a form with the following fields: 'Enter a name for the sequences (optional):', 'Paste your RNA sequences in FASTA format [?]', 'upload a file [Wybierz plik.]' (with a note 'Nie wybrano pliku'), 'Example', 'Disable threshold correction according to GC% [?]', and 'Enter your E-mail address (optional):'. At the bottom of the form are 'Reset' and 'Run CARNAC' buttons. A footer note says 'For questions about carnac or for bug reports, please contact carnac' and the year '2013 - Bonsai bioinformatics'.

Ryc.2. Strona domyślna Carnac.

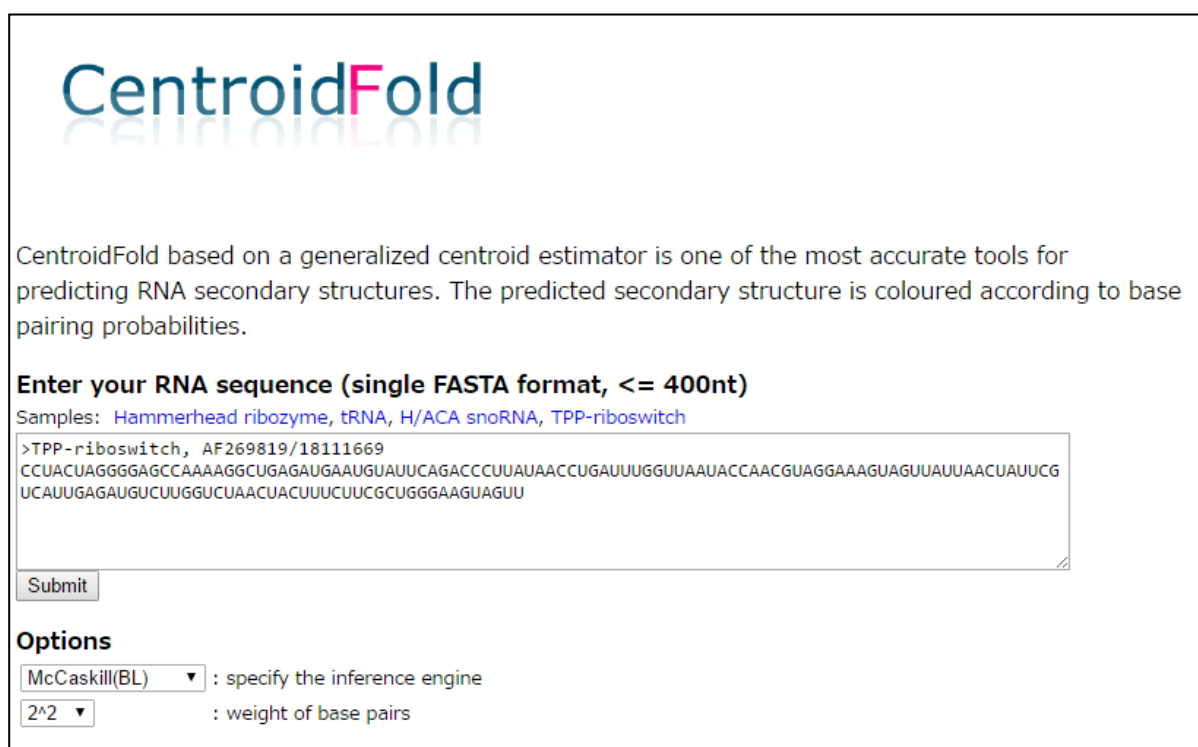
The screenshot shows the results page of the Carnac web interface. At the top, there are navigation links: 'home', 'web server', 'help', 'examples', and 'retrieve result with an ID'. The main content area is titled 'Results for job May_24_2016_21_20_00_31368'. It lists the input files: 'telomerase_1_AF417611/283-441', 'CT file, PS file, JPEO file, bracket notation, list of stems', 'telomerase_2_U15955/50038', 'CT file, PS file, JPEO file, bracket notation, list of stems', 'telomerase_3_2016417612/23192', and 'CT file, PS file, JPEO file, bracket notation, list of stems'. Below this, there's a section 'Alignment [chistat]:' showing sequence alignments for 'telomerase_1', 'telomerase_2', and 'telomerase_3'. The alignments are shown in a grid format with sequence logos. At the bottom, there's a section 'RNAfold structures:' showing the predicted RNA secondary structures for the three sequences.

Ryc.3. Strona wynikowa Carnac.

CentroidFold

Narzędzie CentroidFold może być użyte do przewidywania struktury drugorzędowej RNA jednej sekwencji lub dopasowani wielu sekwencji RNA. Jest to możliwe dzięki opracowaniu algorytmu opartego na γ -centroid estymatorze. Estymator ten to pewnego rodzaju sposób dekodowania bazujący na teorii decyzji statystycznej.[7] Dzięki zastosowaniu estymatora metoda jest bardzo precyzyjna.[39]

CentroidFold dostępny jest wyłącznie online na stronie: <http://rtools.cbrc.jp/centroidfold/>. Jako dane wejściowe akceptowalne są dwa formaty: FASTA dla jednej sekwencji oraz format Clustal W dla dopasowani wielu sekwencji RNA. Nie ma potrzeby zaznaczania jaki format wybrano, ponieważ jest on automatycznie wykrywany.[7] Interfejs jest bardzo prosty i przejrzysty.



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
CCUACUAGGGGAGCCAAAAGGCUGAGAUAAUGUAUUCAGACCCUUAUAACCUGAUUUGGUUAAUACCAACGUAGGAAAGUAGUUUAUUAACUUAUUCG
UCAUUGAGAUUGUCUUGGUCUAAACUACUUUCUUCGCGGGGAAGUAGUU
```

Options

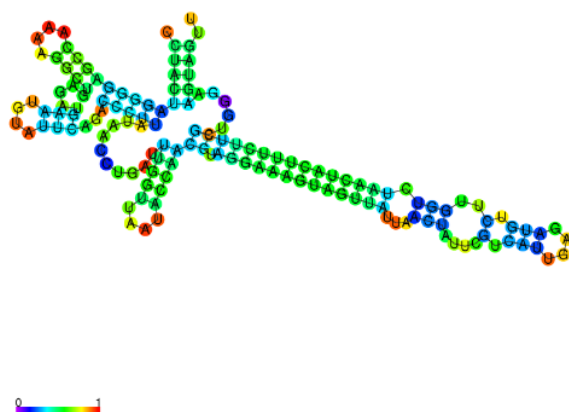
: specify the inference engine

: weight of base pairs

Ryc.4. Strona domyślna CentroidFold.

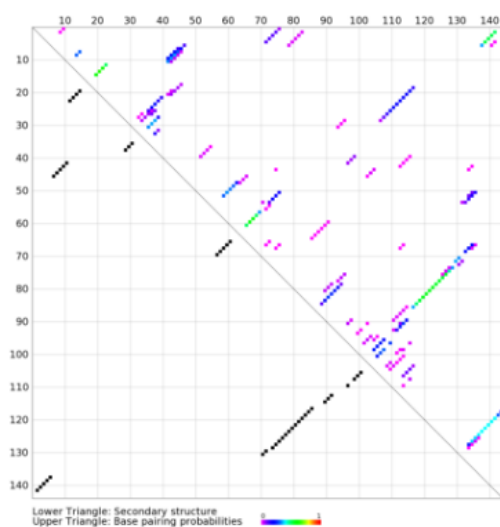
Wyniki otrzymujemy w kilku formatach, w tym także graficznych z odpowiednim kolorowaniem. Jak widać na poniższych ilustracjach mamy do wyboru formaty zapisu do pliku: TEXT, jako obrazek PNG, PDF lub z rozszerzeniem .eps.

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
ACGCUGUCUGUACUUGUAUCAGUACACUGACGAGUCCCUAAAAGGACGAAACAGCGC
.(((((((.(((((((.....)))))).....((((.....)))).....)))))). (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.

CentroidHomfold

Narzędzie pochodzi z tego samego serwera co CentroidFold. Jednakże sposób działania jest znacznie inny. Przy wyznaczaniu struktury drugorzędowej RNA, CentroidHomfold uwzględnia informacje o homologach danej sekwencji.[8] Sekwencje homologiczne pobierane są automatycznie z bazy danych Rfam przy użyciu LAST. Użycie tych dodatkowych informacji daje CentroidHomfold znaczną przewagę nad CentroidFold w dokładności przewidywania struktury. [40] Ponadto narzędzie wykorzystuje również znany nam z poprzedniej metody γ -centroid estymator.

Cele narzędzia możemy podzielić na trzy części. Pierwszy z nich to znalezienie struktury drugorzędowej danej sekwencji RNA. Drugi to znalezienie sekwencji homologicznych do podanej sekwencji oraz trzeci, czyli znalezienie struktur drugorzędowych homologów.

CentroidHomfold dostępny jest online za darmo pod adresem: <http://rtools.cbrc.jp/centroidhomfold/>. Jako dane wejściowe akceptowane są pliki w formacie FASTA. Natomiast wynik otrzymujemy w kilku formatach przedstawionych poniżej.

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](#)

```
>A1609485.1/1-127
CUGGAGACUAAGAAAUAAGAGUCCUUGAAAUCAAGCUGACUCUGCUUUUAGCCUCCUAAAUGAAAAGGUAUAAGAACAGGUCUUGUUUGCAAAUA
AAUUCAGACCUACUUAUCUACCAACAGCA
```

Submit

Options

[McCaskill\(BL\)](#) : specify the inference engine for secondary structure

[CONTRAlign](#) : specify the inference engine for pairwise alignment

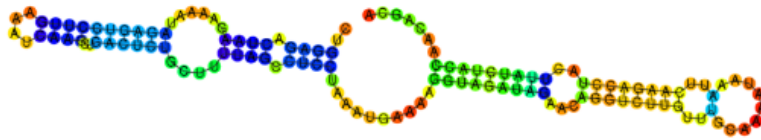
[2³](#) : weight of base pairs

[0.01](#) : E-value for homology search against [Rfam](#)

[30](#) : number of homologous sequences

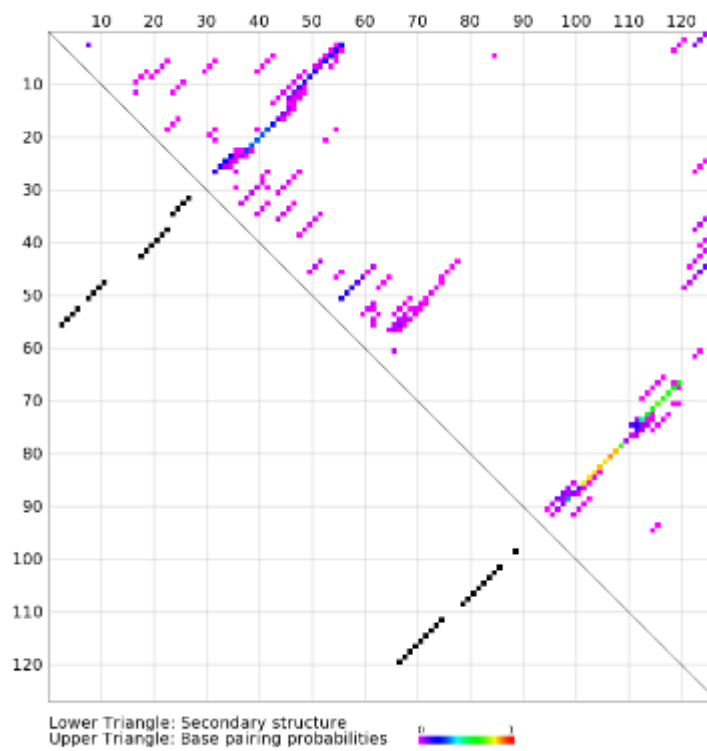
Ryc.5. Strona domyślna CentroidHomfold.

Secondary structure



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

Base pairing probability plot



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

ContraFold

Metoda ContraFold wypełnia lukę pomiędzy metodami probabilistycznymi a termodynamicznymi udowadniając, że procedury uczenia statystycznego mogą stanowić alternatywę dla empirycznego pomiaru parametrów termodynamicznych do przewidywania struktury drugorzędowej RNA. ContraFold oparty jest na warunkowych modelach logarytmiczno-liniowych, probabilistycznych modelach uogólniających stochastyczną gramatykę bezkontekstową (SCFGs) poprzez funkcje wyższych punktacji. Dzięki temu narzędzie osiąga wysoką dokładność predykcji pojedynczego ciągu.[5]

Istnieją dwie opcje użytkowania narzędzia. Po pierwsze, pakiet oprogramowania możemy ściągnąć na swój komputer i zainstalować. Mamy do wyboru dwie wersje – starszą v1.00 lub nowszą v2.00. Obie wersje zaimplementowane są w języku programowania C++ i są darmowe. Jednakże instalacja możliwa jest tylko na systemie Linux.

Drugą opcją jest wersja online na stronie: <http://contra.stanford.edu/contrafold/>. Formatem wejściowym jest FASTA, a wyjściowym BPSEQ. Mamy też możliwość bezpośredniego przesłania wyników na podanego maila. Wartą uwagi jest opcja wyboru pod względem parowania się nukleotydów. Możemy użyć opcji domyślnej typu Watsona-Cricka lub dopuścić parowanie się wszystkich możliwych.[41]

CONTRAFold
CONDitional TRaining for RNA Secondary Structure Prediction

Home CONTRAlign CONTRAFold CONTRAST

[Description | Server | Download | RAF | FAQ | Links]

You may use the CONTRAFold server to predict RNA secondary structures. Input to the server consists of a single FASTA format sequence, up to 1000 characters in length.

For larger sequences, CONTRAFold may be executed on your local machine (see download page). (See the FAQ for a note regarding multiple sequence prediction with CONTRAFold.)

The email server is currently down. Please use the web interface.

Step 1: Choose either of two input methods.

(a) Enter FASTA format sequence (see example)

(b) Upload FASTA format text file:
Wybierz plik Nie wybrano pliku

Step 2: Choose either of two output methods.

☐ Send results to e-mail address:
Type your e-mail address:
Re-type your e-mail address:

☐ Send results to web browser.

Step 3: Select advanced options (optional).

Output format: ☒ Parenthesized ☐ BPSEQ
Base-pairings: ☒ Allow only complementary base pairs (AU, CG, GU) ☐ Allow all possible base pairs

Step 3: Click submit!

Run CONTRAFold! Clear form

Site design by Chuong Do. Artwork by Marina Sirota.
Comments? Questions? Send us feedback.

Ryc.6. Strona domyślna ContraFold.

DotKnot

Znalezienie struktury pseudoowęzłów o minimalnej energii swobodnej jest problemem NP-zupełnym. Jednym ze sposobów na przewidywanie struktur drugorzędowych pseudowęzłów jest podejście heurystyczne polegające na poszukiwaniu w sekwencji potencjalnych kandydatów na pseudowęzły i późniejsza ich weryfikacja. DotKnot wybiera podejrzane regiony i składa je w jeden konstrukt. Metoda skuteczna jest dla długich sekwencji oraz znajduje pseudowęzły z większą dokładnością niż pozostałe algorytmy aktualnie dostępne. DotKnot przewiduje również pseudowęzły typu H, których odcinki mogą być poprzerywane przez wybrzuszenia lub pętle wewnętrzne.[13]

Narzędzie jest dostępne online pod adresem: <http://dotknot.csse.uwa.edu.au/> lub do pobrania na swój komputer. Obie opcje są darmowe.

Formatem wejściowym danych jest sekwencja w formacie FASTA.[42]

The screenshot shows the DotKnot web interface. The header includes 'THE UNIVERSITY OF WESTERN AUSTRALIA' and navigation links: 'Faculty Home', 'School Home', 'Bioinformatics Research Group', and 'KnotSeeker'. The main content area is titled 'DotKnot: pseudoknot prediction including kissing hairpins'. It contains a 'Home' section with links to 'References', 'Manual', 'Download', 'Energy parameters', 'Testset', and 'Contact'. The 'Manual' section is expanded, showing 'Input and Output Format', 'Pseudoknot Types', and 'Local pseudoknots'. The 'Download' section has links for 'Manual [pdf]' and 'Code [DotKnot_1.3.2.tar.gz]'. The 'Energy parameters' section mentions 'DotKnot uses pseudoknot loop entropy parameters published by Song Cao and Shi-Jie Chen in 2006 and 2009'. The 'Testset' section lists 'Testset1' and 'Testset2'. The 'Contact' section provides an email address: 'janasp@csse.uwa.edu.au'. The 'Submission' section has a 'Sequence ID:' field, an 'RNA Sequence (plain format):' text area, and checkboxes for 'Include kissing hairpins', 'Show best local pseudoknots', and 'Show global structure'. There is also a field for 'If you would like to be notified when your computation is done, enter your email address:' and a 'Run DotKnot' button.

Ryc.7. Strona domyślna DotKnot.

Wyniki otrzymujemy w notacji dot-bracket.

Wykryte pseudowęzły są następnie weryfikowane za pomocą technik laboratoryjnych.

The screenshot shows the DotKnot web interface displaying results. The header is the same as in Ryc.7. The main content area is titled 'Your results are here!'. It shows the 'INPUT SEQUENCE:' as 'UGA0260' and '1...75' with a sequence 'GGGCGAAGUGUGACGGGAGCACACAGACUAGUAGUAGGAGGAGUUCGAGUCCUUCUUGUCCACCA'. Below this, it says 'DOTKNOT DETECTED THE FOLLOWING PSEUDOKNOTS:'. It lists 'Position: 16 ... 44', the sequence 'GGGCGACACACAGACUAGUAGUAGG', and the 'Estimated Free energy: -13.8 kcal/mol'. At the bottom, it states 'Your result will be stored here for 7 days.' and provides a link 'Return to the start page for submitting another sequence'.

Ryc.8. Strona wynikowa DotKnot.

HotKnots

HotKnots jest kolejną metodą heurystyczną służącą do przewidywania struktury drugorzędowej RNA z pseudowęzłami lub bez nich. Została ona zaimplementowana przez Jihong Ren w 2005 roku. Narzędzie bada wiele alternatywnych struktur stosując algorytm minimalizacji energii swobodnej, sprawdza potencjalnych kandydatów do tworzenia pseudowęzła i buduje większą strukturę łącząc kandydatów. Tworzone jest to na wzór drzewa, w którym każdy węzeł to jedna substruktura, zwana w tym algorytmie „hotspot”. Wybieranie odpowiednich hotspotów odbywa się dzięki programowaniu dynamicznemu.

Metoda jest dostępna za darmo online na stronie: <http://www.cs.ubc.ca/labs/beta/Software/HotKnots/>. Tam również można pobrać pliki instalacyjne na swój komputer. Mamy do wyboru dwie wersje – starszą 1.0 oraz nowszą 2.0. Wersja 2.0 jest do użycia również online.

Jako dane wejściowe należy wpisać sekwencję ułożoną od 5' do 3'. Ponadto mamy do wyboru 4 modele z różnymi parametrami, na podstawie których zostanie przeprowadzona predykcja. Tutaj również istnieje możliwość przesłania wyników bezpośrednio na podanego w formularzu maila.[43]

[[RNAsoft Home](#) | [PairFold](#) | [CombFold](#) | [RNADesigner](#) | [HotKnots](#)]

HotKnots - Predict RNA secondary structures with pseudoknots

[[About](#) | [Sample Run](#) | [Help](#)]

Input sequence (max. 100 bases) [?]
5' - - 3'

Model and parameters [?]
Dirks & Pierce model, DP09 parameters ▼

The temperature is fixed to 37°C.

Get results ... [?]
☒ as web page
☐ via e-mail to

[[Conditions of use](#)]

Comments, suggestions and feedback are welcome at: [support at rnasoft dot ca](mailto:support@rnasoft.ca)

Ryc.9. Strona domyślna HotKnots.

Natomiast wyniki otrzymujemy w notacji dot-bracket.

[\[RNAsoft Home | PairFold | CombFold | RNA Designer | HotKnots \]](#)

HotKnots - results

[\[New Run | Help \]](#)

Input

Sequence [\[? \]](#):

5'- GGCGCGGCACCGUCCGCGGAACAAACGG -3'

Model and parameters [\[? \]](#):

The model is **Dirks & Pierce**, the set of parameters is **DP09**.
The temperature is fixed to **37°C**.

Output: 9 secondary structures sorted according to MFE:

Structure 1 and its minimum free energy [\[? \]](#):

5'- ..((((((..[[[[]])))).....]]]) -3' -8.02 kcal/mol

Note: You may visualize this structure by copying and pasting the sequence and secondary structure in [Pseudoviewer](#).

Structure 2 and its minimum free energy [\[? \]](#):

5'- ..((((((.....))))). -3' -5.14 kcal/mol

Note: You may visualize this structure by copying and pasting the sequence and secondary structure in [Pseudoviewer](#).

Structure 3 and its minimum free energy [\[? \]](#):

5'- .((((.....[[[[]])).....]]]) -3' -4.61 kcal/mol

Note: You may visualize this structure by copying and pasting the sequence and secondary structure in [Pseudoviewer](#).

Structure 4 and its minimum free energy [\[? \]](#):

5'-(((. [[[]]))]] -3' -3.14 kcal/mol

Note: You may visualize this structure by copying and pasting the sequence and secondary structure in [Pseudoviewer](#).

Structure 5 and its minimum free energy [\[? \]](#):

5'- (((.....))..((((.....)))) -3' -2.03 kcal/mol

Note: You may visualize this structure by copying and pasting the sequence and secondary structure in [Pseudoviewer](#).

Structure 6 and its minimum free energy [\[? \]](#):

5'-(((.....))) -3' -1.77 kcal/mol

Note: You may visualize this structure by copying and pasting the sequence and secondary structure in [Pseudoviewer](#).

Ryc.10. Strona wynikowa HotKnots.

TurboKnot

Metoda służy do przewidywania struktury drugorzędowej RNA, w tym pseudowęzłów zachowanych w obrębie wielu sekwencji. TurboKnot potrafi przewidzieć z dokładnością wynoszącą 80%, co jest bardzo dobrym wynikiem w porównaniu do innych metod. Opiera się na wyborze konformacji o najniższej energii swobodnej z wielu sekwencji.[35]

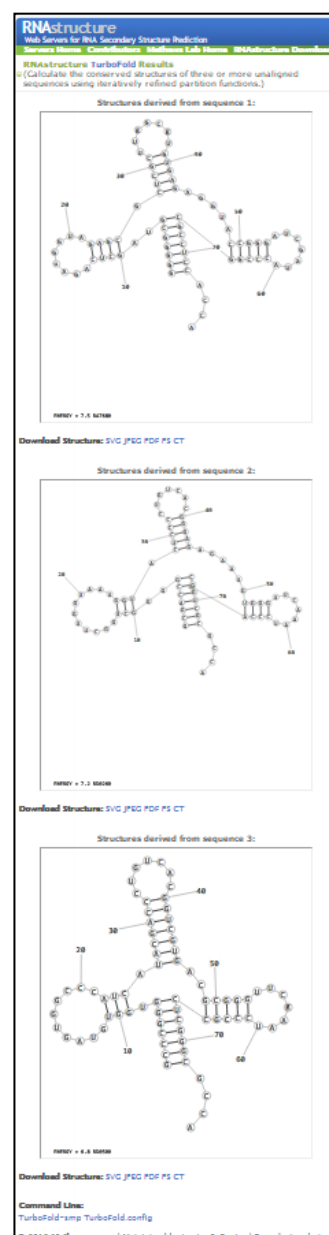
TurboKnot jest częścią większego oprogramowania RNAstructure, jest to jeden z kilku podprogramów. Aby pobrać go na swój komputer należy pobrać całą paczkę oprogramowania ze strony RNAstructure: <http://rna.urmc.rochester.edu/RNAstructure.html>. Natomiast jeśli chcemy go użyć online jako samodzielnego programu należy udać się pod adres: <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/TurboFold/TurboFold.html>.

Plik wejściowy to co najmniej trzy sekwencje w formacie FASTA w jednym pliku. Dodatkowo mamy możliwość wyboru jednego z trzech trybów oraz wiele opcji (temperatura, min długość helixy, itp.). Plik wyjściowy możemy pobrać w formatach: SVG, PDF, JPG, PostScript, CT.[44]

Ryc.11. Strona domyślna TurboKnot (z lewej).

Ryc.12. Strona wynikowa TurboKnot (z prawej).

The screenshot shows the 'RNAstructure TurboFold' web interface. It includes a navigation bar with links like 'Home', 'Contact Us', 'Help', and 'Download'. The main content area is titled 'Welcome to the TurboFold Web Server'. It contains instructions on how to use the server, a section for 'Select Sequences Input' with a file upload button and a text area for FASTA format input, and a section for 'Select Folding Mode' with radio buttons for 'Maximum Expected Accuracy' and 'Pseudoknots'. There are also sections for 'Select TurboFold Options' with various input fields for parameters like 'Temperature', 'TurboFold Gamma', 'TurboFold Iterations', 'Maximum Expected Accuracy Options', 'Pseudoknots Mode Options', and 'Optional Data'. A 'Submit Query' button is at the bottom.



CMfinder

CMfinder to nowa metoda pozwalająca na przewidywanie motywów RNA w niedopasowanych sekwencjach. Do opisu motywów wykorzystywany jest model kowariancji i framework Bayesa. Może być również wykorzystywany do poszukiwania homologów aby udoskonalić modele struktur na podstawie znalezionych sekwencji homologicznych. Dzięki temu uzyskano bardzo dokładne modele kowariancji znanych motywów RNA na podstawie niewielkiej liczby powiązanych sekwencji, które znaleziono u odległych homologów.[9]

Narzędzie możemy używać lokalnie na swoim komputerze. Aktualnie dostępna jest wersja 0.2, którą możemy ściągnąć ze strony: <http://bio.cs.washington.edu/CMfinderWeb/CMfinderDownload.pl>. Przed pobraniem konieczne jest wypełnienie formularza z danymi. Alternatywą jest wersja online dostępna pod adresem: <http://bio.cs.washington.edu/CMfinderWeb/CMfinderInput.pl>. Sekwencję podajemy w formacie FASTA, a wynik otrzymujemy zapakowane w archiwum. Niestety metoda ma ograniczenie do 60 sekwencji o długości maksymalnej do 500 nukleotydów.[45]

University of Washington
Computer Science & Engineering

News: CMfinder software [download](#) is available now.

You need help adjusting the parameters of CMfinder? Please read the [manual](#). Questions? [send mail](#)

You can run CMfinder using 2 sets of parameters at one time

First configuration:

[Number of stem-loops](#) 1 ▾
[Number of motifs](#) 3 <10
[Minimum length of motif](#) 30 >15
[Maximum length of motif](#) 100 <150
[Number of Candidates](#) 40 <100

Second configuration:

[Number of stem-loops](#) 2 ▾
[Number of motifs](#) 3 <10
[Minimum length of motif](#) 30 >15
[Maximum length of motif](#) 100 <150
[Number of Candidates](#) 40 <100

Paste Sequences (*) in FastA Format
Please limit your dataset to
4-60 sequences with length < 500bp

Or Upload a FastA file (*): Nie wybrano pliku

Post process options: ☐ [Remove Redundant motifs](#)

Output compression Options: ☒ zip archive ☐ gzipped tar ball

Your email address where results will be sent to

Press to submit information, or to default parameters.

Ryc.13. Strona domyślna CMfinder.

CRWrnafold

CRWrnafold jest pewnego rodzaju bazą zawierającą zróżnicowaną kolekcję modeli struktur drugorzędowych przewidywanych na podstawie analizy porównawczej dla różnych grup filogenetycznych – 5S, 16S, 23S, tRNA, I i II grupa intronów.

Obecnie jest dostępnych ok 1092 struktur drugorzędowych z dokładnością około 97%, natomiast sekwencji jest ponad 54 000.[9]

Baza dostępna jest online pod adresem: <http://www.rna.ccbb.utexas.edu/DAT/3C/Structure/index.php> lub możemy ją pobrać na swój komputer. Nie musimy pobierać od razu całej bazy, możemy wybrać to, co nas interesuje i pobrać tylko konkretne fragmenty.[46]

THE UNIVERSITY OF TEXAS AT AUSTIN

COMPARATIVE RNA WEB SITE AND PROJECT THE GUTELL LAB

1. CAR 2. SAE 3. DAT 4. SIM 5. FAM 6. AI

This Page: | [Facebook](#) | [Twitter](#) | [Google+](#) | [LinkedIn](#) | [Share](#) | 19 | [Like](#) | [105](#) the Comparative RNA Web (CRW) Site

Welcome to the Comparative RNA Web (CRW) Site.

Recent CRW Site Publications (Complete List »)

- 2015: (Pub #132) R3D-2-MSA: the RNA 3D Structure-to-Multiple Sequence Alignment Server. *Nucleic Acids Research*, 43(W1):W15-W23. [[PM](#) | [PMC](#) | [DOI](#)]
- 2015: (Pub #131) rRNA -- the evolution of that magic molecule. *RNA*, 21:627-629. [[PM](#) | [PMC](#) | [DOI](#)]
- 2015: (Pub #130) RNACentral: an international database of ncRNA sequences. *Nucleic Acids Research*, 43(D1):D123-D129. [[PM](#) | [PMC](#) | [DOI](#)]
- 2014: (Pub #129) Multiple entries in: *Concise Encyclopedia of Bioinformatics and Computational Biology*, 2nd Edition, Hancock J.M. and Zvelebil M.J. (eds.). Wiley, Hoboken, New Jersey.
- 2014: (Pub #128) Ten Lessons with Carl Woese about RNA and Comparative Analysis. *RNA Biology*, 11(3):254-272. [[PM](#) | [PMC](#) | [DOI](#)]
- 2014: (Pub #127) Introduction to Special Carl Woese Issue in *RNA Biology*. *RNA Biology*, 11(3):170-171. [[PM](#) | [PMC](#) | [DOI](#)]
- 2014: (Pub #126) Helix Capping in RNA Structure. *PLOS One*, 9(4):e93664. [[PM](#) | [PMC](#) | [DOI](#)]

Table of Contents

1. Comparative Analysis of RNA (CAR) »
General discussion about Comparative Analysis of RNA, Structure models for reference RNA molecules.
2. Structure and Evolution (SAE) »
Analysis of the Structure and Evolution of Comparative RNA Data.
3. RNA Data (DAT) »
Comparative RNA Data: alignments, structure model diagrams, information about structures, metadata, and other less analyzed data.
4. Supplementary Information for Manuscripts (SIM) »
Online appendices to published papers. Currently the Gutell Lab research focuses on five major categories: RNA Structure, Motifs, Folding, rRNA-based Phylogeny, and Lateral Gene Transfer.
5. RNA Families (FAM) »
Data organized by RNA Families (under development).
6. Ancillary Information (AI) »
Information related to the Gutell Lab and the Comparative RNA Web Site.

Contact Information | Guidelines for Data Usage | Acknowledgements

Recent CRW Site Updates (More Updates »)

- 2014-present: Ongoing collaboration with the Leontis-Zirbel group at BGSU.
- R3D-2-MSA Server (website at BGSU, database at UT-Austin)
- 2015-02-01: Poster at Gordon Research Conference (Converging Disciplines for the Advancement of Inter-RNA Interactions; website).
- 2015-05-09: R3D-2-MSA manuscript (*Nucleic Acids Research*, 43(W1):W15-W23.
- 2015-05-26: Poster at the RNA Society Meeting (Madison 2015; website).
- 2016-01-25: Added manuscript appendix (Pub #132)

2016-01-26: Added appendices for RNAMotif, an RNA Secondary Structure Definition and Search Algorithm, cAT, the Comparative Analysis Toolkit, and Covariation Analysis.

2014-09-10: RNACentral press release (2015-01-20: See also (Pub #130)).

2013-07-08: Added 1E. Modeling the Comparative Structure of Ribosomal RNA to the CRW database.

CRW Site: Site Overview | Site Map | Help/Questions/Comments | Technical Information and Assistance

U. of Texas: UT Home | CCBB | IB | CNS | ICES | ICMB

RRNA R-RNA R-RNA TEXAS RRNA

© 2008-2014 The Gutell Lab, The University of Texas at Austin

Ryc.14. Strona domyślna CRWrnafold.

CyloFold

Metoda CyloFold służąca do przewidywania struktur drugorzędowych z uwzględnieniem pseudowęzłów. Główną jej zaletą jest to, że algorytm nie ma ograniczeń pod względem złożoności pseudowęzłów.

Algorytm początkowo tworzy listę wszystkich możliwych helis z ponad 3 pb. Helisy mogą się parować zgodnie z teorią Watsona-Cricka lub GU-wobble. Predykcja odbywa się poprzez wybór najlepiej punktowanych struktur po 50 krokach symulacji. Wynik jest ustalany jako suma wkładu energii swobodnej z już wprowadzonych helis.

Dostęp do metody możliwy jest tylko online na stronie: <https://cylofold.ncifcrf.gov/cylofold-0.1/sequenceJob/create>. Główny algorytm zaimplementowany jest w C++, natomiast serwer internetowy zaimplementowano przy użyciu Grails frameworku, który jest oparty na Groovy.[11] Aby wygenerować obraz przewidywanej struktury uruchamiany jest program VARNA. Wyniki są tymczasowo przechowywane w relacyjnej bazie danych. Sekwencję wkleja się w formacie FASTA lub Raw i trzeba zaznaczyć który format się wybiera.[47]

Ryc.15. Strona domyślna CyloFold (powyżej).

70				
1	G	8	2	0
2	U	1	3	0
3	C	2	4	0
4	U	3	5	0
5	A	4	6	0
6	C	5	7	0
7	C	6	8	0
8	U	7	9	0
9	A	8	10	0
10	U	9	11	0
11	C	10	12	0
12	G	11	13	0
13	G	12	14	0
14	G	13	15	0
15	C	14	16	0
16	U	15	17	0
17	A	16	18	0
18	A	17	19	0
19	G	18	20	0
20	G	19	21	0
21	A	20	22	0
22	G	21	23	0
23	C	22	24	0
24	C	23	25	0
25	G	24	26	0
26	U	25	27	0
27	U	26	28	0
28	U	27	29	0
29	G	28	30	0
30	C	29	31	0
31	G	30	32	0
32	A	31	33	0

Ryc.15. Strona wynikowa CyloFold (powyżej).

Ryc.16. Strona wynikowa CyloFold (powyżej).

MASTR

Algorytm MASTR (Multiple Aligment of STrustural RNAs), jak sama nazwa wskazuje, służy do przewidywania struktur drugorzędowych dla dopasowań wielu sekwencji RNA. Do tego celu używa Markowskich łańcuchów Monte Carlo, iteracyjnie poprawia dopasowanie sekwencji i przewidywanie struktury dla zbioru sekwencji RNA. Wykorzystywane są również ukryte modele Markova (HMM) oraz suma par. Metoda wykazuje wysoką dokładność i wydajność obliczeniową.

Technika optymalizacji inspirowana jest fizycznym procesem wyżarzania, w którym to powolne chłodzenie materiału powoduje powstawanie struktury krystalicznej. Ogólnie im mniejsza energia tym układ stabilniejszy. Analogicznie jest tym algorytmie, tworzy się symulację by uniknąć minimów lokalnych i uzyskać minimum globalne na końcu procesu. Symulowane wyżarzanie może być stosowane do zminimalizowania kosztów funkcji i do wielokrotnego dopasowania.[16]

Metoda jest dostępna w wersji online na stronie: <http://servers.binf.ku.dk/mastr/> oraz jako kod źródłowy do pobrania z: <http://servers.binf.ku.dk/mastr/download.php>. Formatem wejściowym jest FASTA, a wynikowym PostScript.[48]

Welcome to the **MASTR** server

Multiple Alignment of STructural RNAs

This webserver performs multiple alignment and secondary structure prediction on a set of structural RNA sequences using the MASTR program. It is hosted by [The Bioinformatics Center](#) at the [University of Copenhagen](#) in Denmark. Questions, bugs etc. should be emailed to Stinus.Lindgreen.

To use the MASTR webserver, simply upload your RNA sequences (i.e. at least 2) in **FASTA format** either as a file or by pasting them into the box below. You will receive an email when the results are ready.

If you use this service, please add a reference to our paper:
S. Lindgreen, P.P. Gardner and A. Krogh (2007): *MASTR: Multiple alignment and structure prediction of non-coding RNAs using simulated annealing*, *Bioinformatics*, **23**(24):3304-11 [[Link to paper](#)]

You can download the source code and various data files [here](#).

Dataset with RNA sequences (i.e. at least 2) in Fasta format. [Wybierz plik](#) | Nie wybrano pliku

Advanced parameters:

Gap open probability	<input type="text" value="0.5"/>
Gap extension probability	<input type="text" value="0.74"/>
Base pairing null probability	<input type="text" value="0.23"/>
Scaling factor of basepairing probability	<input type="text" value="1.5"/>
Scaling factor of log-likelihood	<input type="text" value="1"/>
Scaling factor of covariation	<input type="text" value="0.6"/>
Covariation cut-off	<input type="text" value="0.25"/>
Iteration dependence on alignment size (#nuc1)	<input type="text" value="700"/>
Iteration dependence on alignment length	<input type="text" value="2500"/>

Email notification (when done):

Ryc.17. Strona domyślna MASTR.

MaxExpect

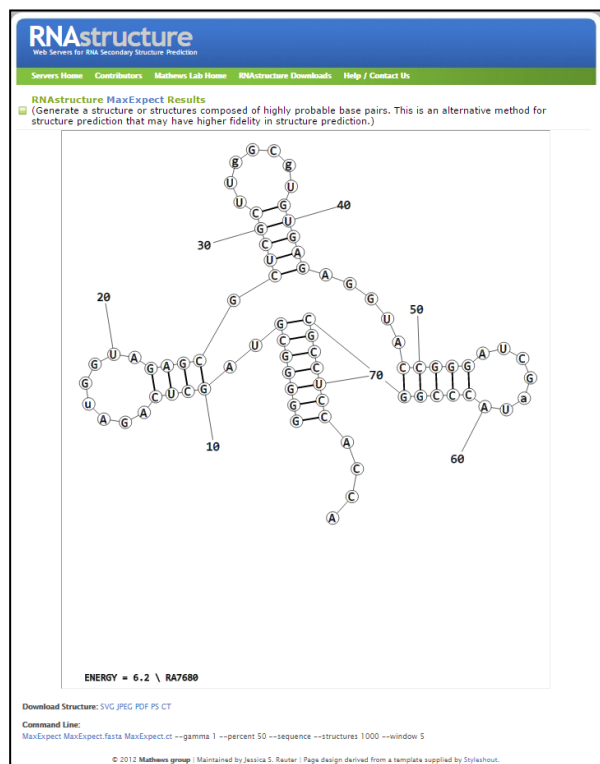
Metoda służąca do przewidywania struktury drugorzędowej RNA w oparciu o maksymalizację oczekiwanej dokładności parowania się zasad. Algorytm przewiduje zarówno optymalną strukturę oraz nieoptymalne, które mogą posłużyć jako alternatywa. Średnia dodatnia wartość predykcijna (PPV) dla MaxExpect jest o 2% większa w porównaniu do algorytmów korzystających z minimalizacji energii swobodnej.[17]

Narzędzie należy do oprogramowania RNAstructure. Dostęp online pod adresem: <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/MaxExpect/MaxExpect.html>. Jak również można pobrać całą paczkę wraz z danym narzędziem.

Jako dane wejściowe akceptowany jest format FASTA. Po wykonaniu się zapytania otrzymujemy wyniki w kilku formatach: FASTA, SVG, JPEG, PDF, PostScript oraz CT.[49]

The screenshot shows the 'Welcome to the MaxExpect Web Server' page. It includes a navigation bar with links like 'Servers Home', 'Contributors', 'Mathews Lab Home', 'RNAstructure Downloads', and 'Help / Contact Us'. The main content area contains instructions and a form for submitting a query. The form has sections for 'Select Sequence Input' (with a 'Submit Query' button), 'Select Nucleic Acid Type' (DNA or RNA), and 'Select MaxExpect Options' (with fields for 'Maximum W Energy Difference', 'Maximum Number of Structures', 'Window Size', and 'Gamma'). At the bottom, there is a 'Designate Address to Email Results' section and a 'Submit Query' button.

Ryc.18. Strona domyślna MaxExpect.



Ryc.19. Strona wynikowa MaxExpect.

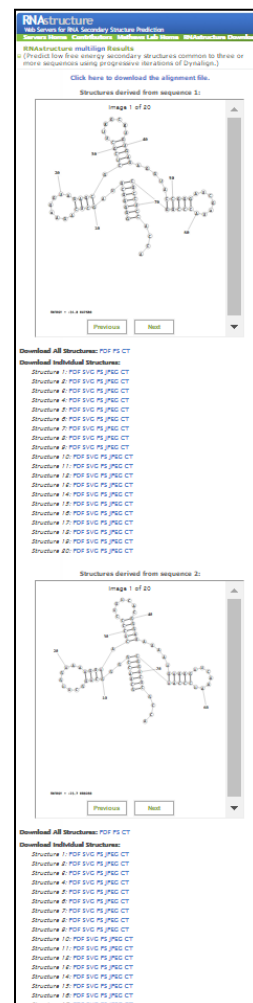
Multialign

Multialign jest metodą służącą do znajdowania najniższej energii swobodnej struktury drugorzędowej wspólnej dla wielu sekwencji RNA. Algorytm opiera się o program Dynalign, który w tym przypadku pomaga stopniowo konstruować konserwatywną strukturę dla każdej z par w ciągu obliczeń. Parowanie się zasad uznawane jest tylko wtedy, gdy danym zestawie Dynalign znajduje minimum energii swobodnej. Takim sposobem Multialign poprawia dokładność przewidywania, wyłączając konkurencyjne niepoprawne pary zasad. Dzięki takiemu rozwiązaniu algorytm ma złożoność liniową w stosunku do liczby sekwencji. Został on przetestowany na bardzo wielu zestawach danych sekwencji o znanych strukturach i okazało się, że aktualnie jest on jednym z najlepszych i najdokładniejszych algorytmów. Uruchomienie w Multialign sekwencji powyżej 1500 nukleotydów nie stanowi problemu.[18]

Narzędzie należy do grupy oprogramowania RNAstructure. Wersja online na: <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/multilign/multilign.html>. Można również ściągnąć pliki instalacyjne dostępne z pełnym oprogramowaniem RNAstructure. Plik wejściowy w formacie FASTA, a wyniki w kilku formatach.[50]

The screenshot shows the RNAstructure Multilign Web Server interface. It includes a header with the RNAstructure logo and navigation links. The main content area is titled 'Welcome to the Multilign Web Server' and contains instructions for using the tool. There are sections for 'Select Sequences Input' (with a text area and a 'Submit Query' button), 'Select Nucleic Acid Type' (with radio buttons for DNA and RNA), and 'Select Multilign Options' (with various input fields for parameters like Number of Structures, Structure Window Size, and Temperature). At the bottom, there is a 'Submit Query' button and a copyright notice.

Ryc.20. Strona domyślna Multialign.



Ryc.21. Strona wynikowa Multialign.

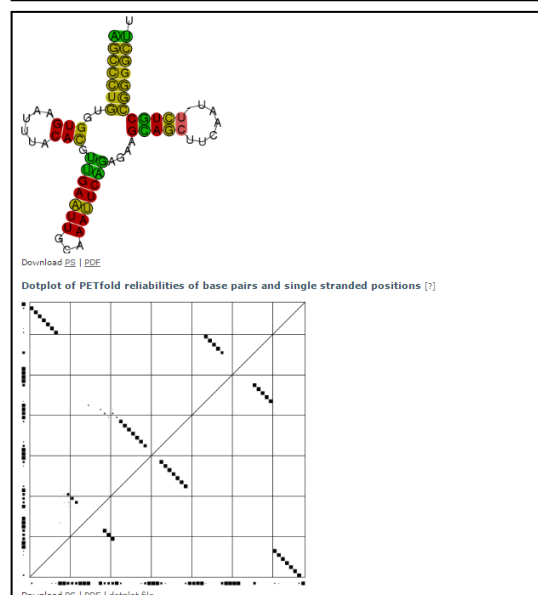
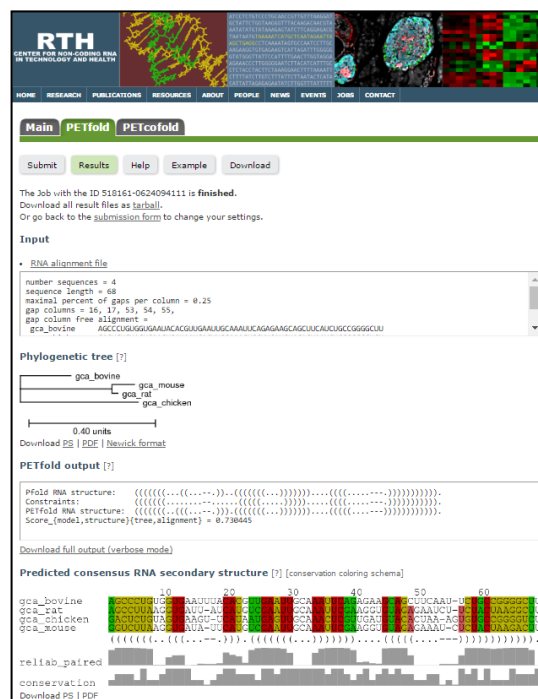
PETfold

Powstała w 2008 roku metoda łączy podejście wykorzystujące energię swobodną z podejściem opartym na ewolucji do przewidywania struktury drugorzędowej RNA. Wdrożono identyfikację par zasad, które z wysokim prawdopodobieństwem są konserwowane lub energetycznie korzystne.

Algorytm został przetestowany na podstawie 46 dobrze poznanych zestawów z Rfam. Wydajność PETfold jest lepsza w porównaniu do RNAalifold.[21]

Narzędzie jest dostępne online za darmo na stronie: <http://rth.dk/resources/petfold/submit.php> oraz jest możliwość pobrania na własny komputer. Mamy do wyboru 4 wersje: v1.0, v1.1, v2.0pre oraz najnowszą v2.0 wydaną w 2013 roku.[51]

Ryc.22. Strona domyślna PETfold (powyżej).



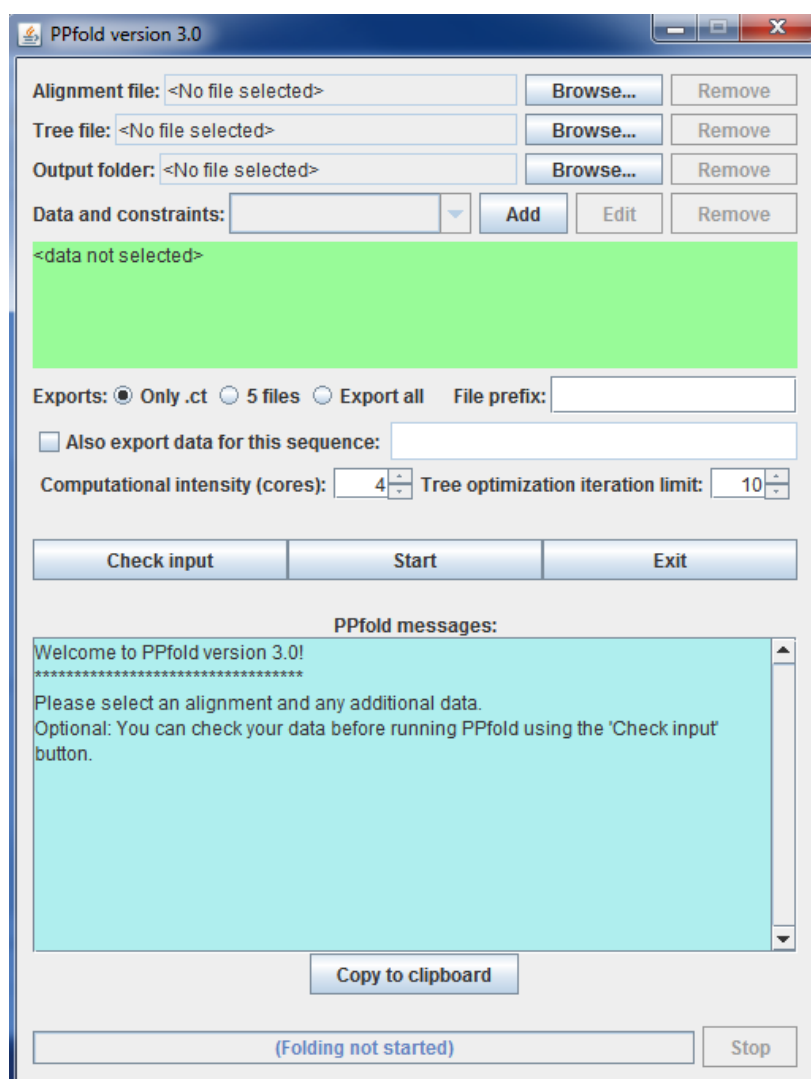
Ryc.23. Strona wynikowa PETfold (obok).

PPfold

Metoda PPfold to udoskonalona metoda Pfold, zaimplementowana w Java 6.0. Służy do przewidywania struktury drugorzędowej dopasowań sekwencji RNA. PPfold łączy w sobie probabilistyczną gramatykę bezkontekstową z modelem ewolucyjnym. Ponadto można wykorzystać dane eksperymentalne z próbkowania do predykcji struktury.[52]

Można uznać PPfold jako pierwszy algorytm równoległe porównujący struktury RNA. Za jego pomocą możemy przewidywać struktury długich transkryptów, ponieważ jest wielowątkowe.[23]

PPfold jest darmowym narzędziem do ściągnięcia na swój komputer ze strony: <http://www.daimi.au.dk/~compbio/pfold/downloads.html>. Można go używać jako samodzielna aplikacja lub jako wtyczka do CLC Workbenches. Najnowsza wersja 3.1.1. została wydana w 2014 roku. Starsze wersje nadal są dostępne.



Ryc.24. Wygląd okna aplikacji PPfold.

ProbKnot

ProbKnot służy do przewidywania struktury drugorzędowej RNA zawierającej pseudowęzły. Algorytm opiera się na obliczonych prawdopodobieństwach parowania się zasad. Po pierwsze przewidywane są prawdopodobieństwa parowania się zasad za pomocą funkcji partycji, która nie uwzględnia pseudowęzłów. Wykonuje się to o czasie $O(N^3)$. Kolejnym krokiem jest złożenie struktur o najwyższej oczekiwanej dokładności o czasie $O(N^2)$ bez użycia programowania dynamicznego, gdzie N oznacza długość sekwencji.

Podczas testów na sekwencji o długości 700 nukleotydów, ProbKnot wypadł najlepiej z testowanych programów (m.in. HotKnots i MaxExpect).[24]

Narzędzie jest częścią zestawu oprogramowania RNAstructure i jest dostępny za darmo online pod adresem: <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/ProbKnot/ProbKnot.html>. Dane wejściowe akceptowane są w formacie FASTA. Dodatkowo można zmieniać liczbę iteracji. Zwiększenie kroków iteracji może skutkować zwiększeniem dokładności. Jeśli zostanie przewidziana struktura zawierająca pseudowęzły to będzie wyświetlana jako struktura kołowa. Natomiast jeśli nie zostaną znalezione pseudowęzły to zostanie przedstawiona jako struktura radialna.[53]

Ryc.25. Strona domyślna ProbKnot.



Ryc.26. Strona wynikowa ProbKnot.

RNAalifold

Aktualnie istnieją dwie wersje RNAalifold – starsza z 2002 roku, mniej efektywna oraz nowsza z 2008 roku. Nowsza metoda wprowadza bardziej racjonalne postępowanie z lukami w dopasowaniach oraz zastępuje uproszczony model kowariancji na macierze punktacji RIBOSUM. Te zmiany nie naruszają wydajności obliczeniowej algorytmu. Metoda może stanowić konkurencję dla metod opartych na SCFGs, modelu najbliższego sąsiada lub maksymalnej oczekiwanej dokładności.[26]

Narzędzie jest częścią oprogramowania ViennaRNA. Najnowszą wersję 2.2 można pobrać za darmo z: <http://www.tbi.univie.ac.at/RNA/>. Aplikacja działa tylko na systemie Linux. Dostępna jest również wersja online pod adresem: <http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi>.

RNAalifold WebServer 1 Enter Input Parameters 2 View Results

[Home|New Job|Help]

Welcome to the RNAalifold web server. It will predict a consensus secondary structure of a set of aligned sequences. Current limits are 3000 nt and 300 sequences for an alignment.

Simply paste or upload your alignment(s) below and click *Proceed*. Accepted alignment formats are CLUSTAL W and FASTA (will be detected automatically). To get more information on the meaning of the options click the symbols. You can test the server using [this sample alignment](#).

Paste your alignment(s) here: [clear]

```
Seq1      UCCCUCCUGGCCCA
Seq2      UUCAGCAUAGCCCA
Seq3      ACCGAGCGGCGCUA
Seq5      UCCCGGCCAACGCA
Seq4      CCCAGAAGAGAGUA
```

Show constraint folding

Or upload a file: Nie wybrano pliku

RNAalifold version

- ☒ new RNAalifold with RIBOSUM scoring (Bernhart SH et al. 2008)
- ☐ new RNAalifold (Bernhart SH et al. 2008)
- ☐ old RNAalifold (Hofacker IL et al. 2002)

Fold algorithms and basic options

- ☒ minimum free energy (MFE) and partition function
- ☐ minimum free energy (MFE) only
- ☐ output "most informative sequence" instead of simple consensus
- ☐ no GU pairs at the end of helices
- ☒ avoid isolated base pairs

Show advanced options

Output options

- ☒ interactive RNA secondary structure plot
- ☒ RNA secondary structure plots with reliability annotation (Partition function folding only)
- ☒ Mountain plot

Notification via e-mail upon completion of the job (optional):

Ryc.27. Strona domyślna RNAalifold.

[\[Home\]](#) [\[New Job\]](#) [\[Help\]](#)**Results for minimum free energy prediction**

The optimal secondary structure in dot-bracket notation with a minimum free energy of **-55.94** (-21.78 plus -34.16 from covariance contributions) kcal/mol is given below.

```
66GCCCCGAGCAGUGGA__AGAGCACAUGCCUCCAACCA GAUGUCCGGGUGCAUCCAGCCGAGCCCA
(((((((.....))))).(((.....)))).....((((.....)))))))))
```

You can download the minimum free energy (MFE) structure in [\[Vienna Format\]](#) [\[Ct Format\]](#).

You can get detailed information about each base pair [here](#).

You may look at the **structure annotated alignment** [\[EPS\]](#) [\[PDF\]](#) [\[IMAGE\]](#) [CONVERTER\]](#).

Results for thermodynamic ensemble prediction

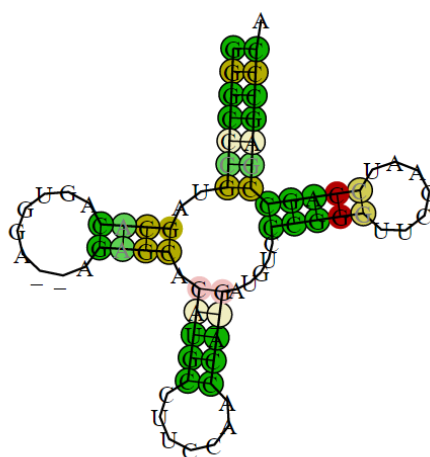
The free energy of the thermodynamic ensemble is **-56.69** kcal/mol.

The frequency of the MFE structure in the ensemble is **29.87** %.

You may look at the **dot plot** containing the base pair probabilities [\[EPS\]](#) [\[PDF\]](#) [\[IMAGE\]](#) [CONVERTER\]](#).

Graphical output

You may look at the interactive drawing of the MFE structure below.

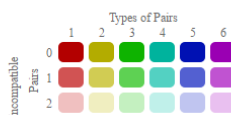


Sequence display options

- ☒ Sequence with conservation annotation
- ☐ Plain Sequence
- ☐ No Sequence

Other display options

- ☒ Conservation annotation
- ☐ Base-pair probabilities
- ☐ Positional entropy
- ☐ None

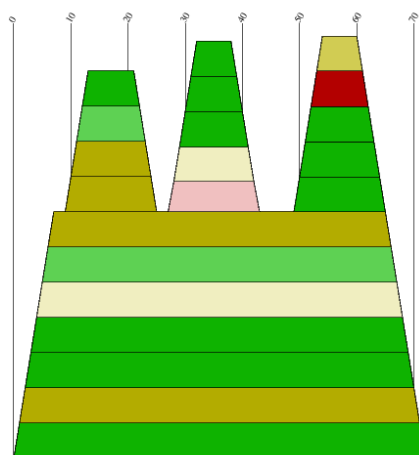
**Image description**

structure drawing with conservation annotation
 structure drawing encoding base-pair probabilities
 structure drawing encoding positional entropy

Download options

[\[EPS\]](#) [\[PDF\]](#) [\[IMAGE\]](#) [CONVERTER\]](#)
[\[EPS\]](#) [\[PDF\]](#) [\[IMAGE\]](#) [CONVERTER\]](#)
[\[EPS\]](#) [\[PDF\]](#) [\[IMAGE\]](#) [CONVERTER\]](#)

Here you find a [mountain plot](#) representation of the MFE structure, the thermodynamic ensemble of RNA structures, and the centroid structure. Additionally we present the positional entropy for each position. Download as [\[EPS\]](#) [\[PDF\]](#) [\[IMAGE\]](#) [CONVERTER\]](#).



An equivalent RNAalifold command line call would have been

```
RNAalifold -p -r -d2 --noLP --color --aln < youralignment.aln > youralignment.out
```

RNA Sampler

RNA Sampler jest metodą przewidującą strukturę drugorzędową motywów RNA w grupie pokrewnych sekwencji. Algorytm znajduje wspólną strukturę pomiędzy dwoma sekwencjami na podstawie probabilistycznego próbkowania. Porównując parami sprawdza wspólne struktury konserwowane wśród różnych sekwencji. Program posiada trzy główne zalety. Przede wszystkim ma możliwość wykrywania pseudowęzłów. Po drugie, nie wymaga wcześniejszego przygotowania strukturalnego. Po trzecie, nie jest konieczna obecność globalnych podobieństw sekwencji, ale można z niej korzystać.[27, 54]

Program został zaimplementowany w C przez Xing Xu, Yongmei Ji oraz Gary Stormo. Ostatnia aktualizacja odbyła się w 2004 roku. Aktualnie można pobrać pliki na swój komputer ze strony: <http://stormo.wustl.edu/RNASampler/index.html>. Do poprawnego działania RNA Sampler należy wcześniej zainstalować pakiet Vienna RNA, Contrafold, Perla oraz ClustalW.

[illegible]

Ryc.29. Przykładowy wynik działania RNA Sampler.

RNAshapes

RNASLOpt

RNAwolf

- strona internetowa: <http://www.tbi.univie.ac.at/software/rnawolf/rnawolf.html>
- strona internetowa: <http://hackage.haskell.org/package/RNAwolf>

Sfold

- strona internetowa: <http://sfold.wadsworth.org/cgi-bin/index.pl>

taveRNA

TurboFold

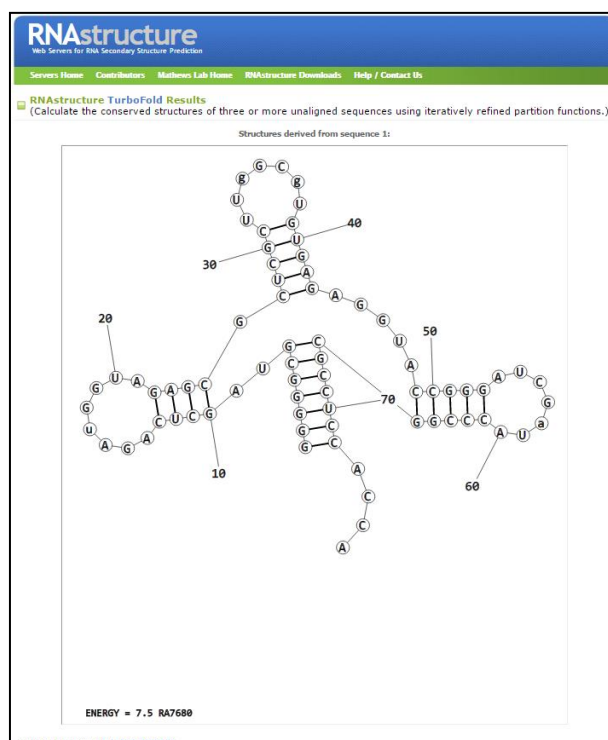
TurboFold to metoda iteracyjno-probabilistyczna służąca do przewidywania struktury drugorzędowej dla wielu sekwencji RNA, która wykorzystuje informacje z modelu termodynamicznego. W przeciwieństwie do innych algorytmów, TurboFold nie wymaga bliskiej relacji badanych sekwencji, dlatego wykorzystywany jest w predykcji struktur homologów, które się znacząco oddaliły.

Dane wejściowe do programu to zestaw co najmniej trzech homologicznych sekwencji RNA. Muszą być one w formacie FASTA. Natomiast na wyjściu otrzymujemy średnie prawdopodobieństwa parowania się zasad dla każdej sekwencji.

Program jest częścią oprogramowania RNAstructure, które można pobrać. Dostępna jest też wersja online pod adresem: <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/TurboFold/TurboFold.html>.

The screenshot shows the TurboFold web interface. At the top is the RNAstructure logo and navigation links. The main heading is "Welcome to the TurboFold Web Server". Below this is a section for "Select Sequences Input:" with a text area for FASTA sequences. There are radio buttons for "OR" and "AND" options. Below the text area is a "Submit Query" button. The "Optional Data" section includes a "Threshold" input field. The "Designate Address to Email Results" section has an "Address to Email Results" input field.

Ryc. .Strona domyślna TurboFold.



Ryc. .Strona wynikowa TurboFold.

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 server) lub lokalnie na własnym komputerze	Linux, Windows	FASTA	CT, PostScript, JPEG, dot-bracket, list of stems, archiwum .zip	elementy struktury drugorzędowej rodzin homologicznych niekodującego RNA		do 300 nt (kilka s) do 2000 nt (kilka min)
CentroidFold	darmowe oprogramowanie dostępne dla wszystkich	online	-	FASTA	TEXT, PNG, PDF, EPS	wszystkie elementy struktury drugorzędowej		do 400 nt
CentroidHomfold	darmowe oprogramowanie dostępne dla wszystkich	online	-	FASTA	TEXT, PNG, PDF, EPS	wszystkie elementy struktury drugorzędowej		do 400 nt
ContraFold	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux	FASTA	BPSEQ	wszystkie elementy struktury drugorzędowej	C++	do 1000 nt

Nazwa metody	Dostępność	Dostęp	Systemy operacyjne, na których działa	Input	Output	Struktury	Język	Ograniczenia
DotKnot	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Windows, Linux, Mac OS-X	sekwencja RNA w formacie FASTA	dot-bracket	pseudowęzły	Python	do 1000 nt dla pseudowęzłów typu H do 400 nt dla kissing harpins
HotKnots	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Windows, Linux, Mac OS-X	sekwencja RNA w formacie FASTA lub tekstowym	dot-bracket	pseudowęzły	C/C++	do 500 nt lokalnie do 100 nt online
TurboKnot	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Windows, Linux, Mac OS-X	FASTA	SVG, JPEG, PDF, CT, PostScript	pseudowęzły	C/C++, Java	
CMFinder	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Windows, Linux, Mac OS-X	FASTA	archiwum .zip na podany adres mailowy	motywy		do 60 sekwencji po 500 nt
CRWrnaFold	darmowe oprogramowanie dostępne dla wszystkich	online	-	wybór konkretnych informacji	archiwum .zip	spinki, pętle wewnętrzne	C#	brak
CyloFold	darmowe oprogramowanie dostępne dla wszystkich	online	-	FASTA lub Raw Sequence	JPG, dot-bracket, CT	pseudowęzły	C++, Groovy	do 300 nt
Mastr	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux	FASTA	PostScript	odcinki non-coding RNAs	C++	od 70 do 250 nt

Nazwa metody	Dostępność	Dostęp	Systemy operacyjne, na których działa	Input	Output	Struktury	Język	Ograniczenia
MaxExpect	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux, Windows	FASTA	FASTA, SVG, JPEG, PDF, PostScript, CT	wszystkie elementy struktury drugorzędowej		do 3000 nt
Multialign	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux, Windows	FASTA	SVG, JPEG, PDF, PostScript, CT	wspólna struktura drugorzędowa o najniższej energii dla kilku sekwencji	C++	powyżej 1500 nt
PETfold	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	dowolny	FASTA, Newick, notacja dot-bracket	PostScript, PDF, Newick, dotplot file	wszystkie elementy struktury drugorzędowej	Perl, C	
PPfold	darmowe oprogramowanie dostępne dla wszystkich	lokalnie na własnym komputerze	Linux	FASTA, Newick	CT, pik .seq	wszystkie elementy struktury drugorzędowej	Java 6.0	do ok. 9000 nt
ProbKnot	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Windows, Linux, Mac OS-X	FASTA	SVG, JPEG, PDF, PostScript, CT	pseudowęzły	C++, Java	do ok 3000 nt
RNAalifold	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux	FASTA lub Clustal W	Vienna Format, CT, PDF, EPS, JPG	wszystkie elementy struktury drugorzędowej		
RNA Sampler	darmowe oprogramowanie dostępne dla wszystkich	lokalnie na własnym komputerze	Linux, Windows	FASTA	własny format	wszystkie elementy struktury drugorzędowej	C	

Nazwa metody	Dostępność	Dostęp	Systemy operacyjne, na których działa	Input	Output	Struktury	Język	Ograniczenia
RNAshapes	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux, Windows	FASTA	dot-bracket	wszystkie elementy struktury drugorzędowej	C	do 1500 nt
RNASLOpt	darmowe oprogramowanie dostępne dla wszystkich	lokalnie na własnym komputerze	Linux	sekwencja RNA	optymalne reprezentacje w notacji dot-bracket wraz z ich energiami	wszystkie elementy struktury drugorzędowej		do 250 nt
RNAwolf	darmowe oprogramowanie dostępne dla wszystkich	lokalnie na własnym komputerze	Linux	sekwencja RNA		triplety i niekanoniczne pary zasad	Haskell	
Sfold	darmowe oprogramowanie dostępne dla wszystkich	online	-	FASTA, GenBank	PostScript, PDF, TXT, archiwum .zip	wszystkie elementy struktury drugorzędowej		do 200 nt online, do 5000 nt lokalnie
taveRNA	darmowe oprogramowanie dostępne dla wszystkich	online lub niektóre z narzędzi możliwe do pobrania	Linux	FASTA	CT, PostScript, GIF	wszystkie elementy struktury drugorzędowej	C/C++	do 150 nt dla Stacked Pair Model, do 500 nt dla Loop Models
TurboFold	darmowe oprogramowanie dostępne dla wszystkich	online lub lokalnie na własnym komputerze	Linux, Windows	FASTA	SVG, JPEG, PDF, PostScript, CT	wszystkie elementy struktury drugorzędowej		

Podsumowanie

Bibliografia

- [1] Lubert Stryer *Biochemia* Wydawnictwo Naukowe PWN (1999)
- [2] Suzanne Clancy, Ph.D. *Chemical Structure of RNA* Nature Education (2008)
- [3] Ronny Lorenz, Michael T. Wolfinger, Andrea Tanzer, Ivo L. Hofacker
Predicting RNA secondary structures from sequence and probing data Elsevier (2016)
- [4] 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 Bioinformatics (2006)
- [5] Woods, D.A., Batzoglou, S.
CONTRAFold: RNA Secondary Structure Prediction without Energy-Based Models. Bioinformatics (2006)
- [6] Helene Touzet, Olivier Perriquet
CARNAC: folding families of related RNAs (2004)
- [7] Kengo Sato, Michiaki Hamada, Kiyoshi Asai, Toutai Mituyama
CENTROIDFOLD: a web server for RNA secondary structure prediction (2009)
- [8] Michiaki Hamada, Koichiro Yamada, Kengo Sato, Martin C. Frith, Kiyoshi Asai
CentroidHomfold-LAST: accurate prediction of RNA secondary structure using automatically collected homologous sequences (2011)
- [9] Zizhen Yao, Zasha Weinberg, Walter L. Ruzzo
CMfinder - a covariance model based RNA motif finding algorithm Bioinformatics (2005)
- [10] David P. Gardner, Pengyu Ren, Stuart Ozer, Robin R. Gutell
Statistical Potentials for Hairpin and Internal Loops Improve the Accuracy of the Predicted RNA Structure (2011)
- [11] Eckart Bindewald¹, Tanner Kluth, Bruce A. Shapiro
CyloFold: secondary structure prediction including pseudoknots (2010)
- [12] Kengo Sato, Yuki Kato, Tatsuya Akutsu, Kiyoshi Asai, Yasubumi Sakakibara
DAFS: simultaneous aligning and folding of RNA sequences via dual decomposition Bioinformatics (2012)
- [13] Jana Sperschneider, Amitava Datta
DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model (2010)
- [14] Jihong Ren, Baharak Rastegardi, Anne Condon, Holger H. Hoos

HotKnots: Heuristic prediction of RNA secondary structures including pseudoknots
Bioinformatics (2005)

[15] Kengo Sato, Yuki Kato, Michiaki Hamada, Tatsuya Akutsu, Kiyoshi Asai
IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming Bioinformatics

[16] Stinus Lindgreen, Paul P. Gardner, Anders Krogh
MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing Bioinformatics (2007)

[17] Zhi John Lu, Jason W. Gloor, David H. Mathews
Improved RNA secondary structure prediction by maximizing expected pair accuracy
Bioinformatics (2009)

[18] Zhenjiang Xu, David H. Mathews
Multalign: an algorithm to predict secondary structures conserved in multiple RNA sequences Bioinformatics (2010)

[19] Hisanori Kiryu, Yasuo Tabei, Taishin Kin, Kiyoshi Asai
Murlet: a practical multiple alignment tool for structural RNA sequences
Bioinformatics (2007)

[20] Yasuo Tabei, Kiyoshi Asai
A local multiple alignment method for detection of non-coding RNA sequences
Bioinformatics (2009)

[21] Stefan E. Seemann, Jan Gorodkin, Rolf Backofen
Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments (2008)

[22] Jens Reeder, Peter Steffen, Robert Giegerich
pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows (2007)

[23] Zsuzsanna Sükösd, Bjarne Knudsen, Morten Værum, Jørgen Kjems, Ebbe S Andersen
Multithreaded comparative RNA secondary structure prediction using stochastic context-free grammars BMC Bioinformatics (2011)

[24] Stanislav Bellaousov, David H. Mathews
ProbKnot: Fast prediction of RNA secondary structure including pseudoknots
Bioinformatics (2010)

[25] Xiaomin Ying, Hong Luo, Jingchu Luo, Wuju Li
RDfolder: a web server for prediction of RNA secondary structure (2004)

[26] Stephan H Bernhart, Ivo L Hofacker, Sebastian Will, Andreas R Gruber, Peter F Stadler
RNAalifold: improved consensus structure prediction for RNA alignments

BMC Bioinformatics (2008)

[27] Xing Xu, Yongmei Ji, Gary D. Stormo
RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment Bioinformatics (2007)

[28] Peter Steffen, Björn Voß, Marc Rehmsmeier, Jens Reeder, Robert Giegerich
RNAshapes: an integrated RNA analysis package based on abstract shapes Bioinformatics (2005)

[29] Yuan Li, Shaojie Zhang
Finding stable local optimal RNA secondary structures Bioinformatics (2011)

[30] Jessica S Reuter, David H Mathews
RNAstructure: software for RNA secondary structure prediction and analysis BMC Bioinformatics (2010)

[31] Christian Höner zu Siederdissen, Stephan H. Bernhart, Peter F. Stadler Ivo L. Hofacker
A folding algorithm for extended RNA secondary structures Bioinformatics

[32] Junilda Spirollari, Jason T.L. Wang, Kaizhong Zhang, Vivian Bellofatto2, Yongkyu Park, Bruce A. Shapiro
Predicting Consensus Structures for RNA Alignments Via Pseudo-Energy Minimization (2009)

[33] Ye Ding, Chi Yu Chan, Charles E. Lawrence
Sfold web server for statistical folding and rational design of nucleic acids (2004)

[34] Cagri Aksay, Raheleh Salari, Emre Karakoc, Can Alkan, S. Cenk Sahinalp
taveRNA: a web suite for RNA algorithms and applications (2007)

[35] Arif O Harmanci, Gaurav Sharma, David H Mathews
TurboFold: Iterative probabilistic estimation of secondary structures for multiple RNA sequences BMC Bioinformatics (2011)

[36] Matthew G. Seetin, David H. Mathews
TurboKnot: rapid prediction of conserved RNA secondary structures including pseudoknots Bioinformatics (2012)

[37] <http://iimcb.genesilico.pl/comparna/methods/>

[38] <http://bioinfo.lifl.fr/RNA/carnac/>

[39] <http://rtools.cbrc.jp/centroidfold/>

[40] <http://rtools.cbrc.jp/centroidhomfold/>

[41] <http://contra.stanford.edu/contrafold/>

[42] <http://dotknot.csse.uwa.edu.au/>

[43] <http://www.cs.ubc.ca/labs/beta/Software/HotKnots/>

[44]

<http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/TurboFold/TurboFold.html>

[45] <http://bio.cs.washington.edu/CMfinderWeb/CMfinderInput.pl>

- [46] <http://www.rna.cccb.utexas.edu/DAT/3C/Structure/index.php>
- [47] <https://cylofold.ncifcrf.gov/cylofold-0.1/sequenceJob/create>
- [48] <http://servers.binf.ku.dk/mastr/>
- [49] <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/MaxExpect/MaxExpect.html>
- [50] <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/multalign/multalign.html>
- [51] <http://rth.dk/resources/petfold/submit.php>
- [52] <http://www.daimi.au.dk/~compbio/pfold/downloads.html>
- [53] <http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/ProbKnot/ProbKnot.html>
- [54] <http://stormo.wustl.edu/RNASampler/index.html>
- [55]
- [56]