

Porównanie narzędzi do wizualizacji struktur drugorzędowych RNA

Ocena jakości wizualizacji

Paulina Gajda, Joanna Gładczak, Damian Rusyn

[Wybierz datę]

Spis treści

Wprowadzenie.....	3
Działanie wykorzystywanych narzędzi.....	4-5
Przepuszczenie struktur przez aplikacje	6-13
RNALogo	6
forna	7
R-CHIE	8-9
RNA-bows	10
PseudoViewer.....	11
Zestawienie otrzymanych wyników	12
Podsumowanie	13
Załączniki	14



www.rcsb.org/pdb/explore/explore.do?structureId=2nbx

Wprowadzenie

Wizualizacje struktur RNA mogą znacząco pomóc w zrozumieniu ich potencjalnych funkcji oraz oszacowaniu wydajności algorytmów ich przewidywania. Wiele ról funkcjonowania RNA może być już analizowanych mając dane drugorzędowe struktury RNA, dzięki opracowanym metodom do wizualizacji tych struktur.

W pracy została użyta seria struktur drugorzędowych RNA o następujących identyfikatorach z bazy Protein Data Bank: 5J01, 2NBX, 5G2Y, 4Y1N, 4R4P, 4Y1I, 4WFM, 4R0D, 4QLN, 4QK9, 4P95, 4C4Q, 4P9R, 4MEH, 3J2C, 4E8K, 4FRN, 4GXY, 4DS6, 3JQ4, 3IZD.

W celu porównania narzędzi do wizualizacji wykorzystano następujące:

1. RILogo (www.rth.dk/resources/rilogo/submit)
2. RNALogo (www.rnalogo.mbc.nctu.edu.tw/)
3. forna (www.nibiru.tbi.univie.ac.at/forna/)
4. XRNA (www.rna.ucsc.edu/rnacenter/xrna/xrna.html)
5. R-CHIE (www.e-rna.org/r-chie/)
6. RNA-bows (www.rna.williams.edu/rnabows/)
7. Pseudoviewer (www.wilab.inha.ac.kr/pseudoviewer/)

Działanie wykorzystywanych narzędzi

1. RLogo:

RLogo tworzy diagram przedstawiający interakcje między dwoma nićmi RNA. Danymi wejściowymi mogą być zarówno dwie pojedyncze sekwencje, jak i dwa dopasowania wielosekwencyjne z objaśnieniem wewnętrznego lub międzymolekularnego parowania zasad pomiędzy tymi dwoma łańcuchami RNA. RLogo wyświetla podobne sekwencje i struktury dla dowolnego RNA dzięki informacjom na temat części wspólnych drugorzędowych struktur tych łańcuchów. RLogo obsługuje cztery różne metody obliczania wspólnych części łańcuchów.

2. RNALogo:

RNALogo prezentuje nowatorską graficzną reprezentację wzorców w dopasowanej sekwencji RNA z jednakową strukturą. Kilka istotnych cech, w tym zachowanie sekwencji, dopasowanie struktur oraz informacje o miejscach sparowanych par zasad, może być ujawnione w ramach dopasowania sekwencji RNA i jednakowej jego struktury drugorzędowej. Narzędzie internetowe jest tak rozwinięte, aby zapewniało szybkie, skuteczne i wysoce konfigurowalne generowanie graficznych przedstawień dopasowań i podobieństw sekwencyjnych. Została również wbudowana galeria RNALogo dla rodzin Rfam RNA. Użytkownicy mogą przeglądać wykresy RNALogo w tej galerii w celu przeprowadzenia dalszych analiz znanych rodzin RNA.

3. forna:

forna to narzędzie do wizualizacji drugorzędowych struktur RNA, które jest funkcjonalne, łatwe w obsłudze i piękne. Pozwala ono na wyświetlanie i edycję drugorzędowej struktury RNA bezpośrednio w przeglądarce bez konieczności instalowania jakiegokolwiek oprogramowania.

4. XRNA:

XRNA to oparty na Javie zestaw narzędzi do tworzenia, objaśniania oraz wyświetlania diagramów drugorzędowych struktur RNA. XRNA dostarcza narzędzia do prostych modyfikacji opublikowanych diagramów cech drugorzędowych, które mogą być zarówno szkicowane ręcznie, jak i generowane automatycznie. Do innych możliwości XRNA możemy zaliczyć grupowanie, numerowanie oraz objaśnianie struktur. Struktury drugorzędowe XRNA mogą być zapisywane w natywnym formacie lub eksportowane jako podskrypty do drukowania i późniejszych modyfikacji w programach takich jak Adobe Illustrator.

5. R-CHIE:

R-chie pozwala wykonać diagramy łukowe drugorzędowych struktur RNA, co znacznie ułatwia porównanie dwóch struktur. Umożliwia również wyświetlenie dopasowania par zasad w kolorze, odpowiadających dopasowań wielosekwencyjnych czy też informacji korelacji sekwencji. R-Chie jest zasilane przez R4RNA - pakiet R, który jest dostępny do pobrania i użytku lokalnego.

6. RNAbows:

RNAbows oferuje 4 rodzaje łuków (RNAbows) do wizualizacji obliczeniowych funkcji:

1. Pary zasad są łączone łukami, których grubość i odcień są proporcjonalne do ich prawdopodobieństwa.
2. Całość funkcji kolorowana jest na czarno z czerwonym wyróżnieniem par MFE (Minimum Free Energy).
3. Porównywanie dwóch dominujących grup zgięć po wcześniejszym podzieleniu funkcji.
4. Wyróżnienie różnic pomiędzy zgięciami dwóch sekwencji tej samej długości.

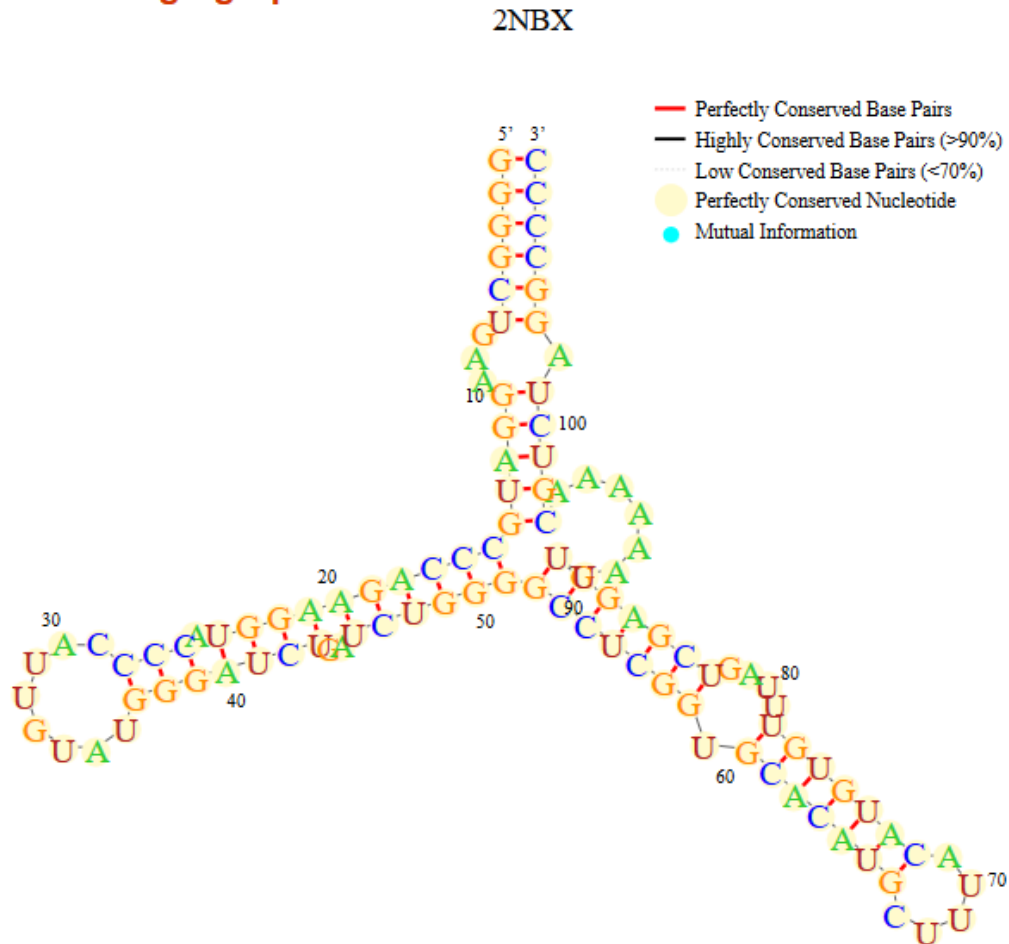
7. PseudoViewer:

Wizualizacja struktur drugorzędowych i pseudowęzłów RNA jest podstawowym narzędziem systemów bioinformatycznych służących do analizy struktur RNA. Jednakże wiele systemów bioinformatycznych wykorzystuje rozmaite dane strukturalne oraz niezgodne elementy oprogramowania, zatem integracja składników oprogramowania w system może być utrudniona przez ich wzajemną niezgodność. W PseudoViewer opracowano nową usługę internetową XML i aplikację internetową do wizualizacji struktur drugorzędowych z pseudowęzłem. Wyniki badań pokazują, że serwis internetowy PseudoViewer i aplikacja są przydatne do rozwiązywania wielu problemów niekompatybilności elementów oprogramowania, a także do wizualizacji dużych struktur drugorzędowych RNA z elementem dowolnego typu pseudowęzła.

Przepuszczenie struktur przez aplikacje

RNAlogo

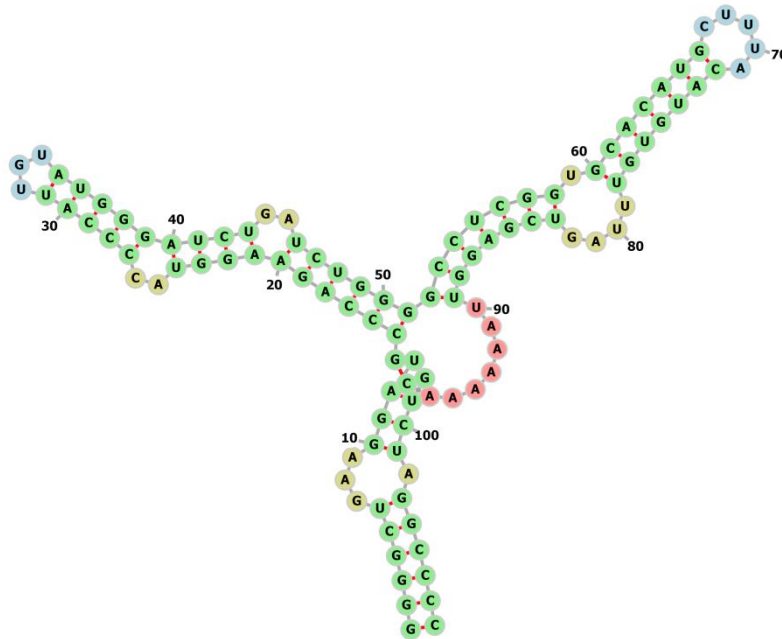
► The RNAlogo graph:



Przykładowy wynik. Struktura 2NBX

1. Są możliwe 4 formaty wyświetlenia (SVG, JPG, PNG, TIF, PDF).
2. W formacie SVG są 3 możliwości wyświetlenia struktury (pokazane na przykładzie 5J01). Fragmenty niektórych struktur są zasłonięte przez legendę, ale po 1 kliknięciu legenda znika.
3. Im struktura bardziej skomplikowana tym obrazek staje się mniej czytelny.
4. W aplikacji brak możliwości przybliżenia grafiki. Można natomiast przy wprowadzaniu danych ustalić wielkość obrazka podając ilość pixeli.
5. Możliwość sprawdzenia dokładnej wartości bitów (The detail **bits** value)
6. Oprócz grafiki wizualizującej wygląd struktury, znajdziemy jeszcze logo sekwencji (The SequenceLogo) oraz consensus structure.
7. Niektóre sekwencje są zbyt długie. Długość sekwencji nie może przekraczać 600 par zasad (3J2C - 928pz, 3JQ4 - 2881pz, 4R0D - 623pz, 5G2Y - 705pz)

forna



Przykładowy wynik. Struktura 2NBX

1. Istnieją cztery możliwości kolorowania

Pozycja - nukleotydy kolorowane są w zależności od ich pozycji w cząsteczce, tj.

te znajdujące się bliżej końca 5' na zielono, w środku na żółto, bliżej końca 3' na czerwono.

Struktura - w zależności od typu struktury

- zielony - stems (kanoniczne helisy)
- czerwony - multiloops (węzły)
- żółty - interior loops
- niebieski - hairpin loops
- pomarańczowy - niesparowane regiony 5' i 3'

Sekwencja - w zależności od typu nukleotydu

żółty - adenina | zielony - cytozyna | niebieski - uracyl | czerwony - guanina

Własne ustawienia

2. Możliwość pobrania otrzymanych wyników w trzech formatach: SVG, PNG, JSON.

3. Struktury mogą być automatycznie wczytywane i wyświetlane z plików PDB

4. Bardzo dobra czytelność nawet przy długiej i skomplikowanej strukturze - możliwość wielokrotnego przybliżania bez najmniejszej utraty jakości.

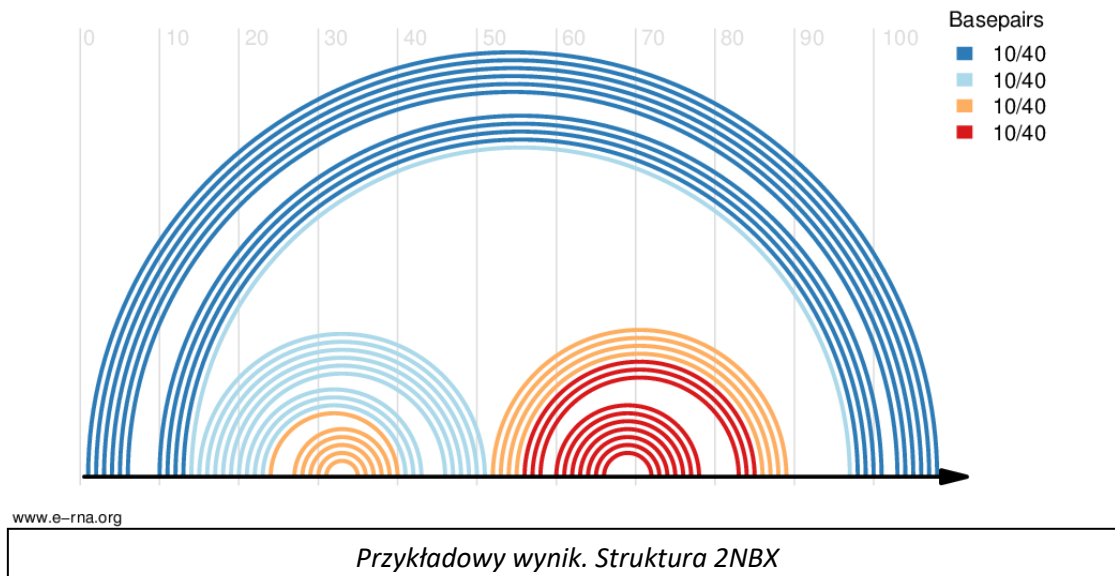
5. Błędy:

4qk9 -nie czytuje X, prosi o sprawdzenie:

UCGGCCCGAACCCGUCAGCUAACCUUGUAAGCGUGGAAAGAGGX

3JQ4 - błąd (0), niezdefiniowano

R-CHIE



1. Dane wejściowe:

struktura drugorzędowa w formacie dot-bracket, connect, bpseq lub helix.

2. Opcje przetwarzania danych wejściowych:

- Uporządkowanie par zasad według wartości malejących lub rosnących
- Odfiltrowanie par zasad z wartości: "mniejszych od...", "większych od..." lub "mniejszych od... i większych od..."

3. Opcje wyglądu:

Ilość możliwych grup kolorowania: 1, 2, 3, 4, 5, 6, 7 lub 8

Ilość palet kolorów: 36

Różne rodzaje grupowania:

- równa liczba par zasad
- równa wartość odległości
- wspólna zlogarytmowana wartość przerwy: $1e^k$, gdzie $k \in \{-7, -6, -5, -4, -3, -2, -1, 0\}$
- konkretna liczba par zasad
- konkretna wartość odstępów
- brak grup pseudowęzłów
- kropkowo – nawiasowy typ struktury: (...), <...>, [...], {...}, A...a, B...b, C...c, D...d
- zakres kowariancji par zasad: przedziały z zakresu [2.00; -2.00] podzielone równo na ich ilość
- zakres konserwacji par zasad: przedziały z zakresu [1.00; 0.00] podzielone równo na ich ilość
- procent kanonicznych par zasad: przedziały z zakresu [1.00; 0.00] podzielone równo na ich ilość

4. Dodatkowe ustawienia: ukryta skala, ukryta legenda.

5. Sześć dostępnych typów wykresów:

- **Single Structure** - wizualizacja skomplikowanych par zasad jako łuki o liniowej sekwencji, kolorowanie par zasad przez wartość
- **Double Structures** - nakreślenie dwóch struktur w tej samej sekwencji dla łatwego porównania
- **Overlapping Structures** - nałożenie na siebie dwóch struktur w celu podświetlenia podobieństw i różnic w tej samej sekwencji
- **Single Covariance** - wyświetlanie wielu sekwencji przyrównanych ze strukturą, kolorowanie dopasowań nukleotydów według statusu pary zasad
- **Double Covariances** - wyświetlanie wielu dopasowań sekwencji dla dwóch struktur dla łatwego porównania struktur i/lub przyrównań
- **Overlapping Covariances** - wyróżnienie podobieństw i różnic dwóch struktur używając informacji dla znanych jak i nowych porównań par zasad

6. Ustawienia dla przepuszczonych struktur drugorzędowych:

Single Structure

Ilość grup kolorowania: 4

Przy grupowaniu konkretną liczbą par zasad wybrano ilości równe kolejno: 2,4,6,8

Błędy po wprowadzeniu struktur:

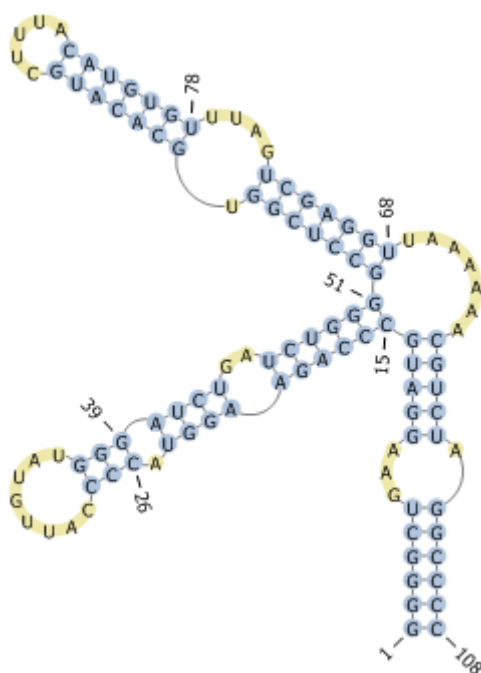
- Równe wartości przerw, wspólna zlogarytmowana wartość przerw, konkretna wartość odstępów: Operacja nie zakończyła się niepowodzeniem, bądź wprowadzono błędne dane wejściowe.
- Zakres kowariancji par zasad, Zakres kowariancji par zasad, procent kanonicznych par zasad.
- Struktury, które nie uległy załadowaniu mimo długiego czasu oczekiwania: 3J2C (działa jedynie 3J2C:M), 3JQ4 (działa jedynie 3JQ4:B), 4Y1I.

RNA-bows



1. Są możliwe 3 sposoby wyświetlenia graficznego (zapis grafiki w PDF, SVG, EPS).
2. Są możliwe 2 sposoby opisu (pairs, bot.pairs).
3. W sposobie AllPairs oraz AllPairsMFE:
 - a) zostaje obliczone G - częściowa wolna energia funkcji [kcal/mol].
 - b) grubość linii każdego łuku jest proporcjonalna do prawdopodobieństwa stworzenia pary zasad pomiędzy dwoma clusterami. Ciemniejsze i cieńsze łuki reprezentują pary zasad z wyższym prawdopodobieństwem połączenia.
4. W sposobie Clusters:
 - a) zostaje obliczone p - prawdopodobieństwo zgrupowania.
 - b) czerwone łuki opisują pary unikalne dla górnego clusteru.
 - c) niebieskie łuki dla dolnego clusteru.
 - d) czarne łuki oznaczają pary charakterystyczne dla obu grup.
 - e) grubość lini każdego łuku jest proporcjonalna do prawdopodobieństwa stworzenia pary zasad pomiędzy tymi dwoma grupami.

PseudoViewer



Przykładowy wynik. Struktura 2NBX

1. Dostępne formaty: EPS, PNG, GIF, SVG, Bracket view.
2. Wyświetlane są 2 diagramy - pokazane na przykładzie 3IZD.
3. Danymi wejściowymi musi być sekwencja i struktura w postaci kropkowo-nawiasowej.
4. Nawet białka o skomplikowanej strukturze są przedstawione w sposób czytelny.

Zestawienie otrzymanych wyników

Podsumowanie

Załączniki