

Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk



– rozprawa doktorska –

**Modelowanie przestrzennych struktur RNA
w badaniach ewolucji na poziomie molekularnym
i projektowaniu nanocząstek RNA**

– Marcin Biesiada –

Praca wykonana
w Zakładzie Struktury i Funkcji Retrotranspozonów

Promotor: dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

Poznań 2023

Serdeczne podziękowania

Mojej Promotor, dr hab. Katarzynie Pachulskiej-Wieczorek, prof. ICHB PAN
za nieocenioną pomoc udzieloną w trakcie przygotowywania rozprawy

Mojej promotor pomocniczej, dr Katarzynie J. Purzyckiej
cierpliwość, wyrozumiałość motywację do krytycznego myślenia i inspirację

Prof. dr. hab. Ryszardowi W. Adamiakowi
za możliwość rozpoczęcia wspaniałej przygody w świecie RNA

Wszystkim pracownikom Zakładu Struktury i Funkcji Retrotranspozonów ICHB PAN
oraz Instytutu Informatyki Politechniki Poznańskiej
za długie wspólne dyskusje i otrzymaną pomoc

Nigdy nie rezygnuj z marzeń tylko dlatego, że ich realizacja wymaga dużo czasu.

Czas i tak upłynie

Earl Nightingale

Ważne żeby dobrze się bawić tym co robisz

Katarzyna J. Purzycka

Pracę dedykuję mojej żonie Gosi oraz córce Weronice

SPIS TREŚCI

1. WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ.....	6
2. WYKAZ POZOSTAŁYCH PRAC NAUKOWYCH.....	7
3. STRESZCZENIE.....	8
4. ABSTRACT.....	9
5. WPROWADZENIE.....	10
5.1. Struktura RNA.....	10
5.1.1. Nukleotydy RNA.....	10
5.1.2. Struktura pierwszorzędowa RNA.....	11
5.1.3. Struktura drugorzędowa RNA.....	12
5.1.4. Struktura trzeciorzędowa RNA.....	14
5.2. Przewidywanie struktur przestrzennych RNA.....	15
5.2.1. Modelowanie porównawcze RNA.....	16
5.2.2. Modelowanie RNA oparte o składnie fragmentów.....	16
5.2.3. Modelowanie <i>de novo</i> cząsteczek RNA.....	16
5.2.4. Modelowanie ab inito.....	17
5.2.5. Modelowanie zintegrowane RNA.....	17
5.2.6. Modelowanie oparte o uczenie maszynowe.....	17
5.3. Zasada działania systemu RNAComposer.....	18
5.4. Wielowymiarowy aspekt struktury RNA w rozwoju nowoczesnych terapii.....	19
5.5. Wyzwania w zakresie badań nad strukturą RNA.....	20
6. CELE NAUKOWE PRZEDSTAWIONEGO CYKLU PRAC.....	22
7. SKRÓTOWY OPIS PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ.....	23
7.1. Automated RNA 3D Structure Prediction with RNAComposer.....	23
7.1.1. Przewidywanie struktur RNA.....	23
7.1.2. RNAComposer - webservice.....	23
7.1.3. Przykłady przewidywania – ryboprzełącznik lizynowy.....	24

7.1.4. Wnioski	26
7.2. RNAComposer and RNA 3D structure prediction for nanotechnology.....	28
7.2.1. Nanotechnologia RNA	28
7.2.2. Schemat modelowania nanostruktur RNA.....	28
7.2.3. RNA prokapsydu bakteriofaga phi29	30
7.2.4. Struktura nanokwadratu RNA o programowalnych funkcjach	30
7.2.5. Wnioski	32
7.3. rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles.....	33
7.3.1. Siódmy segment ekspansyjny	33
7.3.2. Zwój sygnowy siódmego segmentu ekspansyjnego.....	35
7.3.3. Modelowanie struktur RNA siódmego segmentu ekspansyjnego.....	35
7.3.4. Metoda modelowania struktury 2D i 3D RNA.....	36
7.3.5. Walidacja metody.....	37
7.3.6. Przykłady przewidzianych modeli RNA ES7.....	38
7.3.7. Modelowanie kompleksu 28S oraz 5.8S rRNA dużej podjednostki rybosomalnej	40
7.3.8. Wnioski	41
8. PODSUMOWANIE	42
9. WYKAZ STOSOWANYCH SKRÓTÓW I NOMENKLATURA	44
10. LITERATURA.....	46
11. ZAŁĄCZNIKI	53
ZAŁĄCZNIK NR 1. Życiorys naukowy	53
ZAŁĄCZNIK NR 2. Prace naukowe wchodzące w skład rozprawy doktorskiej.....	58
ZAŁĄCZNIK NR 3. Oświadczenia autorów określające wkład w powstanie prac naukowych wchodzących w skład rozprawy doktorskiej	100

1. WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

Rozprawę doktorską stanowi cykl powiązanych tematycznie publikacji, wymienionych poniżej. Prace te zostały opublikowane w renomowanych czasopismach z listy JCR.

Sumaryczny współczynnik wpływu (IF, ang. impact factor)* przedstawionego cyklu prac naukowych wynosi 23+, a liczba punktów MNiSW** – 370.

W każdej z prac jestem pierwszym autorem.

[1] Biesiada M; Purzycka KJ; Szachniuk M; Blazewicz J; Adamiak RW[§]. *Automated RNA 3D structure prediction with RNAComposer*. **Methods Mol Biol.** 2016; 1490:199-215

IF₂₀₁₆ 1,010*, pkt MNiSW 70

[2] Biesiada M; Pachulska-Wieczorek K; Adamiak RW; Purzycka KJ[§]. *RNAComposer and RNA 3D structure prediction for nanotechnology*. **Methods.** 2016, 103:120-7

IF₂₀₁₆ 3,748*, pkt MNiSW 100

[3] Biesiada M; Hu MY; Williams LD; Purzycka KJ[§]; Petrov AS[§]. *rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles*. **Nucleic Acids Res.** 2022, 50(18): p. 10717-10732

IF₂₀₂₁ 19,326*, pkt MNiSW 200

[§] autor do korespondencji

* IF na podstawie danych w bazie Scopus na platformie SCI Journal

** pkt MNiSW na podstawie ujednoliconego wykazu czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego z dnia 18.02.2021

Kopie powyższych prac naukowych wraz z materiałami dodatkowymi umieszczono w załączniku nr 2. Oświadczenia, dotyczące wkładu autorów w powstanie poszczególnych prac naukowych, znajdują się w załączniku nr 3.

2. WYKAZ POZOSTAŁYCH PRAC NAUKOWYCH

Biesiada M; Kourkoumelis N; Kubicki M; Owczarzak AM; Balas V; Hadjidakou SK. Fundamental chemistry of iodine. *The reaction of di-iodine towards thiourea and its methyl-derivative: formation of aminothiazoles and aminothiadiazoles through dicationic disulfides*. **Dalton Trans.** 2014; 43(12):4790-806.

IF₂₀₁₄ 4.221, pkt MNiSW 140

Pachulska-Wieczorek K; Błaszczuk L; Gumna J; Nishida Y; Saha A; Biesiada M; Garfinkel D; Purzycka KJ. *Characterizing the functions of Ty1 Gag and the Gag-derived restriction factor p22/p18*. **Mob Genet Elements.** 2016, 6(2):e1154637.

IF₂₀₁₆ -, pkt MNiSW 70

Pachulska-Wieczorek K; Błaszczuk L; Biesiada M; Adamiak RW; Purzycka KJ. *The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag*. **Retrovirology.** 2016; 13:18.

IF₂₀₁₆ 3,891, pkt MNiSW 100

Miao Z; Adamiak RW; Antczak M; Batey RT; Becka AJ; Biesiada M; Boniecki MJ; Bujnicki J; Chen SJ; Cheng CY; Chou FC; Ferré-D'Amaré AR; Das R; Dawson WK; Feng D; Dokholyan NV; Dunin-Horkawicz S; Geniesse C; Kappel K; Kladwang W; Krokhotin A; Łach GE; Major F; Mann TH; Magnus M; Pachulska-Wieczorek K; Patel DJ; Piccirilli JA; Popenda M; Purzycka KJ; Ren A; Rice GM; Santalucia J Jr; Sarzynska J; Szachniuk M; Tandon A; Trausch JJ; Tian S; Wang J; Weeks KM; Williams B 2nd; Xiao Y; Xu X; Zhang D; Zok T; Westhof E. *RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme*. **RNA.** 2017; 23(5):655-672.

IF₂₀₁₇ 3,229 pkt MNiSW 140

Błaszczuk L; Biesiada M; Saha A; Garfinkel DJ; Purzycka KJ. *Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression*. **Viruses.** 2017; 9(4). pii: E74.

IF₂₀₁₇ 4,067, pkt MNiSW 100

Gómez Ramos LM; Degtyareva NN; Kovacs NA; Holguin SY; Jiang L; Petrov AS; Biesiada M; Hu MY; Purzycka KJ; Arya DP; Williams LD. *Eukaryotic Ribosomal Expansion Segments as Antimicrobial Targets*. **Biochemistry.** 2017; 56(40):5288-5299.

IF₂₀₁₇ 3,004, pkt MNiSW 100

3. STRESZCZENIE

Głównym celem mojej rozprawy doktorskiej było badanie procesów ewolucji cząsteczek RNA oraz projektowanie funkcjonalnych nanocząstek RNA przy użyciu zaawansowanych metod obliczeniowych przewidywania struktury przestrzennej RNA. Badania zostały wykonane z wykorzystaniem pełnych możliwości aplikacji RNAComposer. RNAComposer bazuje na uczeniu maszynowym i modelowaniu poprzez składanie fragmentów, które przebiega bardzo szybko oraz w sposób wysoce zautomatyzowany.

Na poziomie molekularnym ewolucja dotyczy przede wszystkim biologicznych makromolekuł jak RNA, DNA czy białka. Szczegółowe badanie mechanizmów ewolucyjnych pozwala zrozumieć zmiany strukturalno-funkcjonalne makromolekuł zachodzące w czasie. Mechanizmy ewolucji molekularnej nie zostały dotychczas jednoznacznie określone.

W ramach mojej pracy doktorskiej zbudowałem modele struktury przestrzennej najbardziej zmiennego ewolucyjnie fragmentu rybosomalnego RNA określanego jako siódmy segment ekspansyjny - ES7 (ang. *Expansion Segment 7*). Przewidywałem modele ES7 pochodzących z różnych organizmów i różniących się między sobą sekwencją, rozmiarem i poziomem skomplikowania struktury przestrzennej. Jednak wspólną ich cechą był zakonserwowany na poziomie struktury przestrzennej trzon cząsteczki, tzw. zwój sygnaturowy. W pracy tej potwierdziliśmy model akrecyjny ewolucji cząsteczek RNA na poziomie molekularnym i zaproponowaliśmy model wzrostu struktury cząsteczek rRNA.

Nanotechnologia jest prężnie rozwijającą się dziedziną nauki o ogromnym potencjale aplikacyjnym. Aktualnie kilka cząsteczek RNA jest stosowanych jako środki terapeutyczne, a kilkanaście kolejnych jest w fazie testów. Rozwój metod pozwalających na efektywne projektowanie „szytych na miarę” struktur RNA pozwoliłoby na typowanie kolejnych nanocząstek o dedykowanych właściwościach.

W mojej pracy doktorskiej przedstawiłem możliwości projektowania nanocząstek RNA na przykładzie struktury nanokwadratu RNA. Struktura nanokwadratu została zaprojektowana tak, aby mogła tworzyć przestrzenne sieci molekularne poprzez wprowadzoną do cząstki RNA strukturę spinki. Spinka ta umożliwiła oddziaływanie pomiędzy cząsteczkami RNA dzięki parowaniu się wolnych reszt nukleotydowych w pętli apikalnej.

4. ABSTRACT

The main objective of my Ph.D. dissertation was to investigate the evolutionary processes of RNA molecules and design functional RNA nanoparticles using advanced computational methods for predicting RNA spatial structure. These studies were performed using the full capabilities of the RNAComposer application. RNAComposer is based on machine learning and fragment assembly modeling, which proceeds very fast and in a highly automated manner.

At the molecular level, evolution primarily concerns biological macromolecules such as RNA, DNA, and proteins. Detailed study of evolutionary mechanisms allows for understanding the structural and functional changes of macromolecules over time. Molecular evolution mechanisms have not been definitively determined thus far.

As part of my Ph.D. thesis, I constructed models of the spatial structure for the most evolutionarily variable fragment of ribosomal RNA, known as the Expansion Segment 7 (ES7). I predicted ES7 models from various organisms, differing in sequence, size, and level of structural complexity. However, a common feature among them was the conserved core structure, referred to as the signature fold. In this study, we confirmed an accretionary model of RNA molecule evolution at the molecular level and proposed a model for the growth of rRNA molecule structures.

Nanotechnology is a rapidly advancing field with tremendous application potential. Currently, several RNA molecules are used as therapeutic agents, and several more are undergoing testing. The development of methods that enable the efficient design of custom-tailored RNA structures would allow for the characterization of other nanoscale particles with dedicated properties.

In my Ph.D. thesis, I presented the possibilities of designing RNA nanoparticles using the example of an RNA nano-square structure. The nano-square structure was designed to form spatial molecular networks through the introduced RNA hairpin structure. This additional hairpin allows RNA - RNA interaction between molecules through the loop – loop pairing.

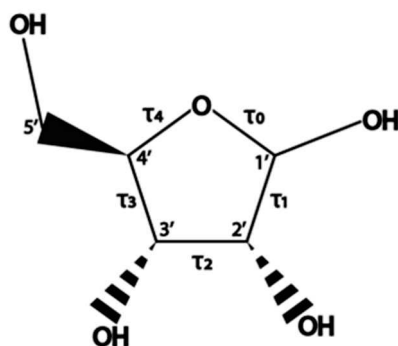
5. WPROWADZENIE

Struktura RNA

RNA – kwasy rybonukleinowe, obok DNA, tłuszczów, białek i węglowodanów zalicza się do jednej z pięciu grup makrocząsteczek kontrolujących funkcjonowanie organizmów żywych. RNA jest biopolimerem w którym powtarzającą się jednostką (monomerem) jest rybonukleotyd. Na architekturę RNA ma wpływ sekwencja reszt nukleotydowych (struktura pierwszorzędowa), parowanie się reszt nukleotydowych (struktura drugorzędowa) oraz przestrzenne ułożenie poszczególnych reszt względem siebie (struktura trzeciorzędowa). Zdolność nici RNA do zwijania się w stabilne struktury przestrzenne ma fundamentalne znaczenie dla ich funkcjonowania, między innymi ma kluczowe znaczenie w procesach transkrypcji, translacji, transportu, czy katalizy.

5.1.1. Nukleotydy RNA

Nukleotyd to podstawowy mer kwasów nukleinowych. Każdy rybonukleotyd składa się reszty pierścienia rybozy, reszty zasady azotowej oraz reszty kwasu fosforowego. RNA zbudowany jest z czterech różnych rybonukleotydów różniących się resztami zasad azotowych. Zasady azotowe połączone z resztą rybozy tworzą rybonukleozydy: adenozyne, guanozyne, urydynie i cytydynie. Rybonukleozydy połączone resztami fosforanu tworzą łańcuch RNA [1]. Reszta fosforanu przyjmuje strukturę tetraedru [2]. Ryboza występuje w postaci pięcioczłonowego pierścienia furanozowego. Ustalono, że najbardziej korzystnymi energetycznie są konformacje pierścienia w którym heterocykliczny atom tlenu oraz atomy C1' i C4' są współpłaszczyznowe, natomiast C2' i C3' nie leżą w płaszczyźnie pierścienia. Konformacja C2'- i C3'-endo jest konformacją w której atomy C2' lub C3' leżą po tej samej stronie płaszczyzny pierścienia co atom C5'. Konformacja C2'- i C3'-egzo to konformacja w której atomy C2' lub C3' leżą po przeciwnej stronie płaszczyzny pierścienia niż atom C5' [1]. Przestrzenne rozmieszczenie atomów reszty cukrowej można również opisać za pomocą kątów torsyjnych τ [2] (Rysunek 1).



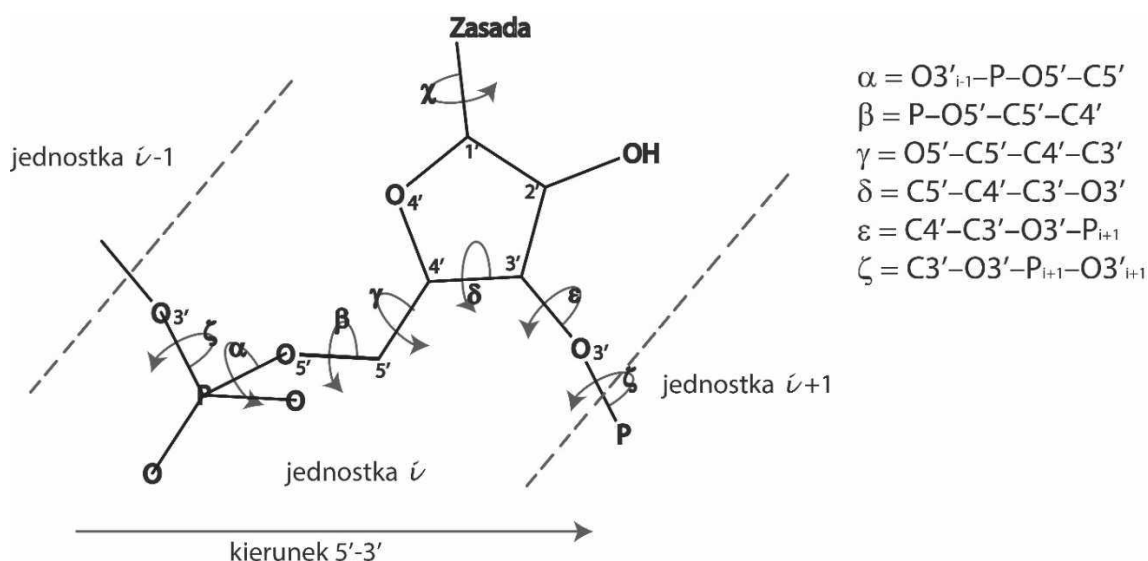
Rysunek 1. Reszta rybozy z zaznaczonymi kątami dwuściennymi τ .

Do zasad azotowych kwasu rybonukleinowego zaliczamy adeninę, cytozynę, guaninę i uracyl. Cytozyna i uracyl są pochodnymi pirymidyny, czyli heterocyklicznego związku złożonego z sześciocząłkowego pierścienia z dwoma atomami azotu w pozycji 1 i 3. Z kolei adenina i guanina są pochodnymi puryny, czyli kolejnego heterocyklicznego związku złożonego z pirymidyny połączonej z pierścieniem imidazolowym w bicykliczną formę. Atomy heterocyklicznych zasad azotowych leżą w jednej płaszczyźnie, ponieważ zasady te są związkami aromatycznymi [3]. Zasady azotowe połączone są z resztą cukrową wiązaniem N- glikozydowym, które powstaje pomiędzy atomem C1' reszty rybozy a atomem N1 zasady pirymidynowej lub N9 zasady purynowej. Z kolei reszta rybozy łączy się w pozycji C5' z resztą kwasu ortofosforowego wiązaniem estrowym [4].

Orientację reszt zasad heterocyklicznych względem reszty cukrowej można opisać za pomocą kątów dwuściennych opartych na wiązaniu glikozydowym. Jeden z kątów dwuściennych określony jest między płaszczyzną zasady heterocyklicznej a średnią płaszczyzną pierścienia furanozowego. Kąt ten w rybonukleotydach jest ostry. Kolejnym kątem jest kąt χ określony przez atomy O4'-C1'-N9-C4 w purynach i O4'-C1'-N1-C2 w pirymidynach. Kąt ten przyjmuje wartości od 50 – 80° dla konformacji syn i 180-280° w konformacji anti [3, 5].

5.1.2. Struktura pierwszorzędowa RNA

Struktura pierwszorzędowa RNA określona jest poprzez sekwencję nukleotydów. Ze względu na to, że jednostki nukleotydowe zawierają różne reszty zasad azotowych, RNA jest heteropolimerem. Sekwencja reszt nukleotydowych w cząsteczkach RNA jest zakodowana w sekwencji DNA, co stanowi element centralnego dogmatu biologii molekularnej [6]. Jednak RNA może również być nośnikiem informacji genetycznej, jak u retrovirusów lub retrotranspozonów, które w cyklu replikacyjnym wykorzystują proces odwrotnej transkrypcji do syntezy DNA na matrycy genomowego RNA [7].

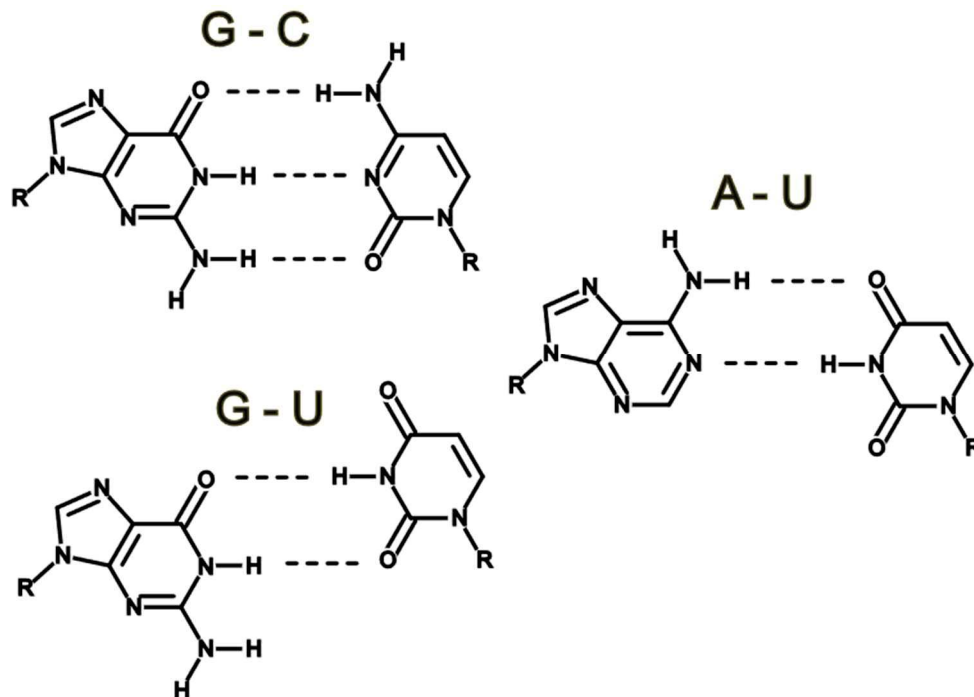


Rysunek 2. Wiązanie fosfodiesterowe z zaznaczonymi kątami dwuściennymi.

RNA tworzy biopolimer poprzez łączenie nukleotydów za pomocą wiązania fosfodiesterowego. Wiązanie fosfodiesterowe tworzy się pomiędzy resztą kwasu ortofosforowego wchodzącego w skład jednego nukleotydu z atomem O3' kolejnej jednostki nukleotydowej. Tak jak w przypadku pierścienia cukrowego konformację łańcucha fosfodiesterowego można opisać za pomocą kątów torsyjnych [8] (α , β , γ , δ , ϵ , ζ ; Rysunek 2).

5.1.3. Struktura drugorzędowa RNA

Struktura drugorzędowa opisuje parowanie się zasad nukleotydowych. Reszty zasad azotowych mogą oddziaływać ze sobą poprzez wiązania wodorowe. Najbardziej trwałe oddziaływania występują pomiędzy resztami par kanonicznych (Rysunek 3) guanozyny i cytozyny (G-C) i adenozyny i urydyny (A-U) oraz tzw. parą chwiejną (ang. *wobble*) guanozyny i urydyny (G-U) [9].



Rysunek 3. Kanoniczne pary zasad Watsona i Cricka G-C, A-U oraz wobble G-U.

W topologii struktury drugorzędowej RNA wyróżnia się motywy strukturalne do których należą pętle, rejony helikalne, pseudowęzły i niesparowane fragmenty końcowe [10]. Rejon helikalny jest rejonem, w którym reszty nukleotydowe parują się. Rejony te wyznaczają granice dla pętli i niesparowanych fragmentów końcowych. Niesparowane fragmenty końcowe są to jednoniciowe fragmenty na końcach 5' lub 3' cząsteczki RNA. Pętle są jednoniciowymi fragmentami wewnątrz łańcucha cząsteczki RNA. Wyróżnia się pętle:

- apikalne – rejony jednoniciowe ograniczone przez jeden rejon helikalny
- wewnętrzne – rejony jednoniciowe ograniczone przez dwa rejony helikalne
- wewnętrzne jednostronne (wybrzuszenia) – szczególny przypadek pętli wewnętrznej w którym niesparowane reszty nukleotydowe występują tylko w jednej nici

- wieloramienne - rejony jednoniciowe ograniczone przez trzy lub więcej rejonów helikalnych.

Pseudowęzły są specyficzną kombinacją helis i pętli. Powstają w wyniku sparowania jednoniciowego fragmentu pętli z innym fragmentem jednoniciowym [11].

Zapis kropkowo-nawiasowy struktury drugorzędowej RNA

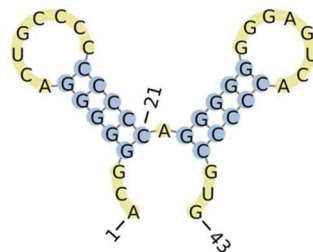
Najbardziej powszechnym zapisem struktury drugorzędowej, oprócz wizualizacji graficznej, jest zapis kropkowo-nawiasowy [12, 13]. Podstawowy zapis kropkowo-nawiasowy składa się z ciągu kropek i nawiasów okrągłych. Zaczynając od końca 5'-OH łańcucha RNA, każdej reszcie nukleotydowej przypisuje się jeden znak:

- „.” kropkę w przypadku niesparowanej reszty,
- „(„nawias okrągły otwierający w przypadku sparowanej reszty rozpoczynającej parę,
- „)” nawias okrągły zamykający w przypadku sparowanej reszty kończącej parę.

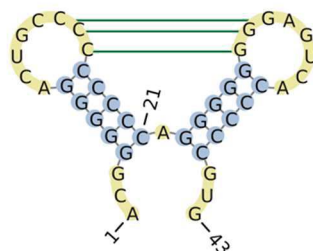
Jednakże podstawowy zapis kropkowo-nawiasowy nie uwzględnia elementów struktury jakimi są pseudowęzły. Notacje pseudowęzłów wyraża się za pomocą nawiasów wyższego rzędu w zależności od porządku ich parowania [14]:

- nawiasy kwadratowe „[]” dla pseudowęzłów tworzących się w pierwszej kolejności,
- nawiasy klamrowe „{ }” dla pseudowęzłów tworzących się w drugiej kolejności,
- nawiasy ostrokątne „< >” dla pseudowęzłów tworzących się w trzeciej kolejności.

A) Sekwencja: ACGGGGGGACUGCCCCCCCCAGGGGGGGGGAGUCACCCCGUG
Struktura 2D: ... (((((((.....)))))).. (((((((.....))))))..



B) Sekwencja: ACGGGGGGACUGCCCCCCCCAGGGGGGGGGAGUCACCCCGUG
Struktura 2D: ... (((((((..... [[[]]])))). (((((((.....))))))..



Rysunek 4. Przykłady graficznych reprezentacji struktury drugorzędowej RNA oraz ich zapisu kropkowo-nawiasowego, A) zapis struktury z dwiema pętlami apikalnymi, B) zapis struktury z dwiema pętlami apikalnymi oraz dodatkowym pseudowęzłem.

W przypadku występowania pseudowęzłów o wyższym porządku parowania stosuje się litery w kolejności alfabetycznej. Wielkie litery rozpoczynają, a małe kończą parowanie reszt nukleotydowych.

Zapis kropkowo-nawiasowy struktury drugorzędowej jest wykorzystywany aby usprawnić i uprościć zapis graficzny (Rysunek 4) oraz ułatwić jego wprowadzenie do oprogramowania do przewidywania struktur trzeciorzędowych [15, 16], czy wizualizacji struktur drugorzędowych [17, 18].

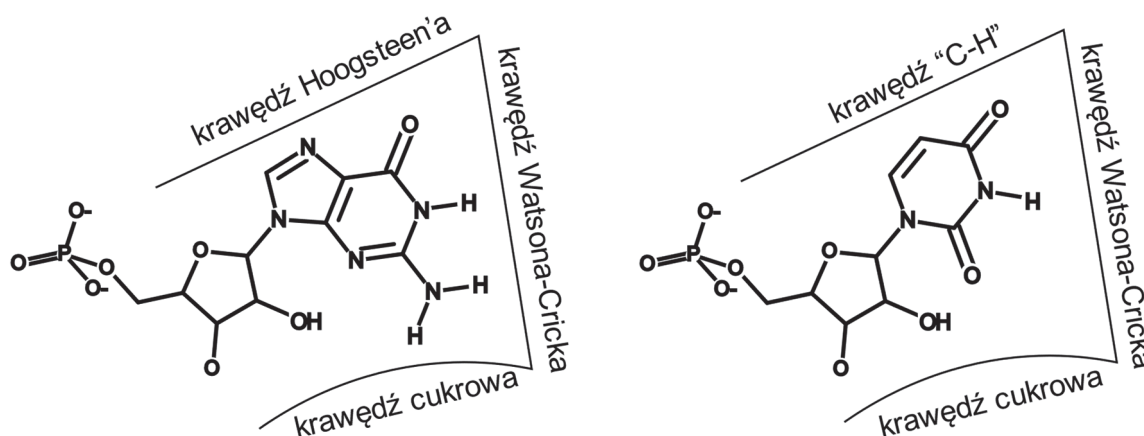
5.1.4. Struktura trzeciorzędowa RNA

Rejony sparowane tworzą proste, nieznacznie zagięte prawoskrętne helisy w formie A [19, 20], a w szczególnych przypadkach lewoskrętne helisy w formie Z [21]. Obecność pętli wewnętrznych wprowadza lub jest wynikiem zmiany kierunku przebiegu fragmentów helikalnych. Oprócz oddziaływań wodorowych zaangażowanych w parowanie się reszt nukleotydowych, helisa RNA utrzymywana jest poprzez oddziaływania warstwowe pomiędzy pierścieniami aromatycznymi typu π [22]. Oddziaływania pomiędzy pierścieniami aromatycznymi typu π często są rozciągnięte na fragmenty jednoniciowe.

Istotną grupę oddziaływań w strukturze przestrzennej stanowią oddziaływania niekanoniczne, charakteryzujące się występowaniem interakcji pomiędzy zasadami reszt nukleotydowych innymi niż parowania Watsona-Cricka i wobble. Rodzaje oddziaływań niekanonicznych zostały szeroko opisane poprzez Leontisa i Westhofa [23].

Wyróżnili oni trzy krawędzie reszty nukleotydowej w RNA, istotne dla struktury przestrzennej tej cząsteczki (Rysunek 5). Należą do nich krawędź:

- Watsona-Cricka
- cukrowa
- Hoogsteen'a w przypadku reszt purynowych lub „C-H” w przypadku reszt pirymidynowych.



Rysunek 5. Krawędzie reszty purynowej (lewa strona) i pirymidynowej (prawa strona) zaangażowane w tworzenie oddziaływań wodorowych.

Każda z tych krawędzi może oddziaływać z którąkolwiek z krawędzi w znajdującej się w pobliżu reszcie nukleotydu. Ponadto oddziaływanie może następować we względnej orientacji reszt nukleotydowych w formie *cis* lub *trans* co daje 12 podstawowych typów oddziaływań niekanonicznych.

Te oddziaływania są ważne w stabilizacji struktury przestrzennej cząsteczki, kształtowaniu wewnątrzcząsteczkowych oddziaływań dalekiego zasięgu, a także umożliwiają oddziaływania z innymi RNA oraz białkami, przez co są one kluczowe dla spełnianych przez RNA funkcji.

Ważnym elementem reszty nukleotydu oddziałującym z krawędziami opisanymi przez Leontisa i Westhofa jest reszta fosforanowa, która pełni podobne funkcje w tworzeniu struktury cząsteczki RNA jak krawędzie reszty nukleozydu [24].

W strukturze trzeciorzędowej obserwuje się swoiste motywy strukturalne, których nie można odczytać ze struktury drugorzędowej. Przykładami takich motywów są platformy adenozyne, tryplety, zamki rybozowe czy kwadrupleksy guanozyny [25-28].

- Platforma adenozynowa jest to zazwyczaj para adenozyn występujących kolejno obok siebie w łańcuchu cząsteczki RNA, które oddziałują ze sobą w płaszczyźnie pierścieni purynowych poprzez oddziaływania niekanoniczne, naśladując parę zasad oraz dopasowując się do oddziaływań warstwowych pierścieni aromatycznych w rejonie helikalnym.
- Tryplety RNA stanowią trzy reszty nukleotydu oddziałujące ze sobą poprzez krawędzie reszty zasady azotowej w sposób niekanoniczny. Tryplety RNA są często spotykane w pętlach sarcynowo-rycynowych, skręconym zwrocie (ang. *kink-turn*) czy w pętlach typu C [28].
- Zamek rybozowy jest to szereg oddziaływań typu wiązań wodorowych, pomiędzy grupami 2'-hydroksylowymi rybozy, pochodzącymi z dwóch regionów łańcucha cząsteczki RNA ułożonych względem siebie w sposób antyrównoległy.
- W tworzeniu struktur kwadrupleksu zaangażowane są cztery nici cząsteczki RNA bogate w reszty guanozyny. Reszty guaniny tworzą kwartety ułożone planarnie, oddziałujące ze sobą wiązaniami wodorowymi typu Hoogsteen'a. Kwartety z kolei ustawiają się jedne nad drugimi, stabilizowane przez oddziaływania pomiędzy pierścieniami aromatycznymi typu π oraz umiejscowiony między nimi jon metalu.

Przewidywanie struktur przestrzennych RNA

Struktura cząsteczek RNA jest bardzo istotna dla ich funkcji, dlatego badanie struktury jest istotne dla zrozumienia procesów w których biorą udział cząsteczki RNA. Mimo to w bazie danych struktur przestrzennych kwasów nukleinowych Nucleic Acid Database (ndb) w 2013 roku było zdeponowanych zaledwie około 2300 struktur cząsteczek RNA podczas gdy w bazie danych Rfam było obecnych około 90 000 000 różnych sekwencji RNA [29, 30]. Obecnie liczba rozwiązanych struktur przestrzennych cząsteczek RNA przekroczyła 6500, co i tak stanowi jedynie niewielki ułamek znanych sekwencji. Ze względu na trudności w oznaczeniu struktur przestrzennych RNA metodami eksperymentalnymi rozwija się metody obliczeniowe ułatwiające przewidywanie modeli cząsteczek RNA.

Wyróżnia się kilka głównych podejść w modelowaniu struktury przestrzennej cząsteczek RNA: modelowanie porównawcze, modelowanie oparte o składanie fragmentów, modelowanie *de novo*, czy modelowanie *ab initio*. Często modelowanie oparte na komputerowych obliczeniach nie jest wystarczające i do modelowania skomplikowanych struktur RNA wykorzystuje się modelowanie zintegrowane będące hybrydą jednej z metod obliczeniowych i informacji uzyskanych z metod eksperymentalnych [31]. Na uwagę zasługuje również intensywnie rozwijające się ostatnio modelowanie oparte na uczeniu maszynowym / głębokim uczeniu.

5.1.5. Modelowanie porównawcze RNA

Do modelowania porównawczego potrzebne jest sekwencja cząsteczki przewidywanej oraz cząsteczka homologiczna o znanej strukturze przestrzennej. Często w tego typu metodach jako daną wejściową używa się topologii struktury drugorzędowej. Zakłada się, że cząsteczki RNA o zbliżonej sekwencji tworzą podobne struktury przestrzenne [32]. Model cząsteczki przewidywanej powstaje w wyniku modyfikacji struktury homologu poprzez wymianę niezgodnych reszt nukleotydowych, dodanie reszt brakujących lub usunięcie reszt nadmiarowych. Przewidywanie to nie wymaga dużych mocy obliczeniowych i zwykle przebiega bardzo szybko, również nie jest ograniczone co do wielkości cząsteczki. Za to jest zależne do dostępności cząsteczki homologicznej i od jakości jej struktury przestrzennej. W tej metodzie rejony cząsteczki wybiegające poza strukturę homologu mogą być modelowane niedokładnie. Do modelowania porównawczego można wykorzystać program ModeRNA [32] czy RNABuilder [33].

5.1.6. Modelowanie RNA oparte o składanie fragmentów

Fundamentem modelowania polegającego na składaniu fragmentów jest baza danych zawierająca przestrzenne elementy struktury RNA [34]. Elementy te służą do przypisywania ich do sekwencji przewidywanej cząsteczki RNA lub jej struktury drugorzędowej. Modelowana cząsteczka RNA jest dzielona na wiele mniejszych części według określonego schematu. Przestrzenne elementy wybrane z bazy są dobierane do wydzielonych części strukturalnych, a następnie składane w finalny model cząsteczki RNA. Ten typ modelowania również nie wymaga dużych mocy obliczeniowych i właściwie nie jest ograniczony co do wielkości cząsteczki. Jakość przewidywanej struktury jest ściśle zależna od jakości struktury referencyjnej. Skomplikowane konstrukcje przestrzenne mogą wymagać wprowadzenia dodatkowych informacji, jak więzy poprzez przestrzeń czy określenie wartości kątów torsyjnych. Do tej grupy można zaliczyć RNAComposer [15], 3dRNA [35] lub VFoldLA [36].

5.1.7. Modelowanie *de novo* cząsteczek RNA

Modelowanie *de novo* opiera się na przewidywaniu struktury 3D RNA na podstawie sekwencji wykorzystując algorytmy obliczeniowe [37]. Algorytmy te bazują na próbkowaniu przestrzeni konformacyjnej zgodnie z przyjętymi założeniami. Ilość obliczeń jakie trzeba

wykonać jest zależna od liczby elementów oddziałujących ze sobą, dlatego przewidywanie to jest obliczeniowo kosztowne i relatywnie wolne, dlatego ograniczone do małych cząsteczek (poniżej 100 reszt nukleotydowych [38]). W modelowaniu *de novo*, aby uprościć obliczenia i skrócić czas przewidywania stosuje się zamiast reprezentacji pełnoatomowych – reprezentacje kulkowe [39, 40]. W reprezentacji kulkowej – kulka jest kwaziatomem reprezentującym grupę atomów tworzących resztę nukleotydową lub resztę zasady azotowej, fosforanowej, czy cukrowej. Przykładami programów wykorzystujących podejście *de novo* są IsRNA1 [39], FARFAR [41], oxRNA [42], czy SimRNA [43].

5.1.8. Modelowanie *ab initio*

W modelowaniu *ab initio* wykorzystuje się tylko prawa fizyki. Danymi wejściowymi do modelowania są sekwencja RNA oraz parametry układu takie jak stężenie jonów, pH, temperatura [44]. Metoda ta pozwala na modelowanie bez żadnych dodatkowych wiązań. Jest ona jednak bardzo kosztowna obliczeniowo, czasochłonna i ogranicza się tylko do niewielkich systemów (RNA o łańcuchach zdecydowanie krótszych niż w modelowaniu *de novo*) o dokładnie zdefiniowanych parametrach. *Ab initio* można modelować za pomocą programu Hire-RNA [45].

5.1.9. Modelowanie zintegrowane RNA

Danymi wejściowymi do modelowania zintegrowanego jest sekwencja RNA oraz dane uzyskane z eksperymentów obrazowania struktur RNA. Ten typ modelowania często ma zastosowanie w przypadku udokładniania struktur RNA niskiej rozdzielczości lub struktur 3D częściowo rozwiązanych, pochodzących, w obu przypadkach z oznaczeń metodami NMR, krystalograficznymi, czy mikroskopii elektronowej [46]. W tych przewidywaniach korzysta się z danych pozyskanych z eksperymentów mających na celu uzyskanie częściowych informacji o strukturze RNA. Wykorzystuje się różne metody, takie jak metoda SHAPE (reakcja selektywnej acylacji reszty cukrowej w pozycji 2'-OH analizowana poprzez reakcję odwrotnej transkrypcji) [47], która umożliwia określenie parowania reszt nukleotydowych. Wykorzystuje się również metodę mapowania rodnikami hydroksylowymi [48], która pozwala na określenie ogólnego kształtu cząsteczki poprzez mapowanie obszarów ekspozowanych na oddziaływanie z rozpuszczalnikiem. Używane są również metody map mutacyjnych oraz MOHCA (analiza wielopunktowego rozerwania wiązań rodnikami hydroksylowymi) [49], które umożliwiają określenie oddziaływania reszt nukleotydowych w przestrzeni trójwymiarowej. Dane takie mogą być informacją bazową procesu przewidywania [48] lub wprowadzać ograniczenia na formowaną strukturę [50]. Jakość struktury przewidywanej jest ściśle zależna od jakości wprowadzonych danych eksperymentalnych.

5.1.10. Modelowanie oparte o uczenie maszynowe

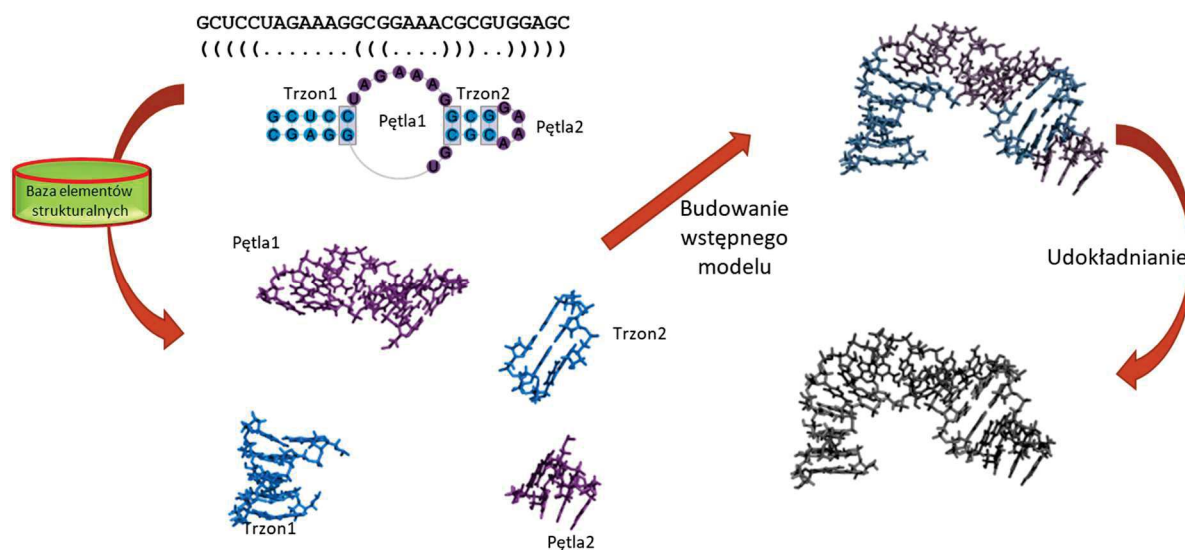
Modelowanie oparte o nauczanie maszynowe może być podkategorią każdego z opisanych powyżej typów modelowania. Wyróżnia ją to, że warunki przewidywania struktur

przestrzennych nie są zdefiniowane w programie do przewidywania. Do przewidywania wprowadza się zbiór znanych struktur przestrzennych RNA po czym program je analizuje i na ich podstawie uczy się zależności w nich występujących. Po wprowadzeniu sekwencji przewidywanej cząsteczki przewiduje jej strukturę na podstawie wyuczonych zależności. Metoda ta wydaje się być bardzo obiecująca ponieważ zależności są wielopoziomowe i mogą być ściślej dopasowywane do przewidywanej cząsteczki, zależnie od wprowadzonego zbioru wejściowego [51]. Jednak w przypadku skomplikowanych struktur RNA może dojść do przeuczenia się programu, nadinterpretacji danych wejściowych i otrzymania zaburzonych modeli wyjściowych. Do modelowania w oparciu o uczenie maszynowe można wykorzystać ARES [52] lub E2Efold-3D [53]

Zasada działania systemu RNAComposer

Zasada działania RNAComposer jest oparta na obserwacji właściwości strukturalnych wielu cząsteczek RNA. W różnych cząsteczkach RNA występują pewne elementy strukturalne, które są analogiczne. Dzieląc strukturę cząsteczki RNA na mniejsze fragmenty zwiększamy prawdopodobieństwo znalezienia danego fragmentu w innej cząsteczce.

RNAComposer przewiduje struktury przestrzenne na podstawie sekwencji i topologii struktury drugorzędowej (Rysunek 6) [15]. Topologia struktury drugorzędowej może być wprowadzona jako dana wejściowa ręcznie lub może zostać uzyskana za pomocą zaimplementowanych w systemie skryptów do przewidywania struktury drugorzędowej RNAstructure [54], CONTRAfold [55], RNAfold [56], CentroidFold [57], ContextFold [58] lub IPknot [59]. Algorytm RNAComposer dzieli strukturę drugorzędową na fragmenty sparowane poprzez kanoniczne pary oraz rejony jednoniciowe i pętle zamknięte przez pary kanoniczne. Pary zamykające elementów są częściami wspólnymi sąsiadujących elementów strukturalnych. Następnie system automatycznie wyszukuje reprezentacje przestrzenne dla elementów struktury drugorzędowej, wykorzystując słownik oparty o bazę elementów strukturalnych RNA FRABASE [34]. RNA FRABASE jest zbiorem elementów strukturalnych zbudowanym w oparciu o struktury RNA zdeponowane w RCSB PDB [60], ponadto baza stale się powiększa, ponieważ jest na bieżąco uaktualniana. Dobór reprezentacji przestrzennych następuje według kryteriów zgodności struktury drugorzędowej oraz podobieństw sekwencji i rodzaju reszt (puryna, pirymidyna), wartości rozdzielczości cząsteczki z którego pochodzi element i wartości energii wyszukanego elementu. Zgodność struktury drugorzędowej jest obligatoryjna, lecz w przypadku braku zgodności sekwencji, reszty nukleotydowe są wymieniane na właściwe. W przypadku braku elementu wykazującego zgodność struktury drugorzędowej, szukany element jest generowany jako prawoskrętna helisa A. Wyszukane elementy przestrzenne są składane we wstępny model struktury poprzez nałożenie wspólnych par zamykających. Model ten jest udokładniany w przestrzeni kątów torsyjnych (w oparciu o program CYANA [61]) oraz w przestrzeni kartezyjskiej (w oparciu o pole siłowe CHARMM [62]) do uzyskania finalnego pełnoatomowego modelu cząsteczki RNA (Rysunek 6). Poprawność działania systemu RNAComposer została przetestowana na dużym zbiorze cząsteczek RNA [15].



Rysunek 6. Schemat działania programu RNAComposer.

Wielowymiarowy aspekt struktury RNA w rozwoju nowoczesnych terapii

Od odkrycia RNA w 1921 przez Phoebusa Levene'a [63] było ono przez długi okres czasu uważane jedynie za nośnik informacji między DNA i białkami. Dzisiaj znanych jest dużo więcej istotnych funkcji jakie pełni kwas rybonukleinowy w komórkach. Dla przykładu interesujące funkcje pełnią niewielkie RNA, np. miRNA i siRNA, które mogą regulować ekspresję genów, poprzez modulowanie translacji białek lub stabilności cząsteczek mRNA [64]. Istotną grupę stanowią katalityczne RNA, tzw. rybozomy, których najpopularniejszym przedstawicielem jest rybosomalny RNA transferazy peptydylowej, enzymu wchodzącego w skład rybosomów, który jest bezpośrednio zaangażowany w katalizowanie tworzenia wiązań peptydowych. Do tej grupy należą również inne rybozomy, katalizujące cięcie i ligację cząsteczek RNA, jak snRNA, które odgrywają istotną rolę w procesie dojrzewania pre-mRNA, samowycinające się introny grupy I, rybozomy typu hammerhead, hairpin, delta oraz RNaza P [65]. Ponadto RNA stanowi genom różnych wirusów, wliczając w to wirusy wywołujące choroby zakaźne jak AIDS, zapalenie wątroby typu C, grypę, gorączkę krwotoczną czy COVID-19 [66, 67]. RNA niektórych patogenicznych bakterii w kontakcie z organizmem człowieka może prowadzić do np. czerwonki, zapalenia płuc czy ciężkich zatruc pokarmowych [68, 69]. Takie RNA wirusowe i bakteryjne są obiecującym celem dla nowoczesnych leków, a informacje o ich strukturze znacznie zwiększają szanse na sukces podczas projektowania leków nacelowanych na RNA.

Struktura drugorzędowa, a raczej zdolność RNA do tworzenia kanonicznych par zasad jest obecnie najpowszechniej wykorzystywaną cechą cząstek RNA w terapiach dostępnych na rynku. Leki oparte o antysensowe oligonukleotydy RNA wykorzystuje się w walce z rdzeniowym zanikiem mięśni [70]. Natomiast małe interferujące RNA używane są do wyciszania genów produkujących dysfunkcyjne białko prealbuminy, którego ekspresja prowadzi do polineuropatii [71]. W terapiach wykorzystuje się również właściwość RNA

do kodowania białek. Powstają nowoczesne szczepionki oparte o mRNA, jak ta przeciwko COVID-19 [72]. W tym przypadku zadaniem mRNA jest kodowanie wirusowego białka odpowiedzialnego za kontakt z receptorem na powierzchni komórki gospodarza, co indukuje odpowiedź immunologiczną. Jednakże to struktura trzeciorzędowa RNA posiada bardziej wyrafinowane właściwości, które będą wykorzystywane w przyszłości do projektowania nowych leków opartych na RNA czy transporterów leków w postaci nanocząstek RNA. Już teraz poszukuje się niewielkich cząsteczek, ligandów, które mogą wiązać się lub wiążą się ze specyficznymi strukturami 3D patogennych RNA jak w przypadku SARS-CoV2 [73, 74] czy wirusa zapalenia wątroby typu C [75, 76].

Wyzwania w zakresie badań nad strukturą RNA

Dawniej RNA postrzegano jako jednoniciowy polimer, którego struktura była pomijana, lub przedstawiana w postaci bliżej nieokreślonej nitki „spaghetti” [77]. Kolejne lata badań nad strukturą RNA doprowadziły do identyfikacji struktury pierwszorzędowej, drugorzędowej i trzeciorzędowej RNA [78]. Jednak w kontekście rozpoznawania funkcji zależnych od RNA wydaje się być to podziałem niewystarczającym. Pomiędzy sekwencją a strukturą przestrzenną brakuje opisu ważnych oddziaływań pomiędzy resztami nukleotydów [79]. Każda reszta nukleotydu RNA posiada takie same zdolności do oddziaływań. Część reszty nukleotydowej odpowiedzialna jest za oddziaływania elektrostatyczne, pierścienie aromatyczne uczestniczą w wiązaniach warstwowych, a donory i akceptory wodoru tworzą wiązania wodorowe. Topologia struktury drugorzędowej przedstawia tylko parowanie się reszt nukleotydowych, podczas gdy nie jest to najważniejsze oddziaływanie wpływające na strukturę RNA. Postuluje się, że oddziaływania warstwowe prowadzą do strukturyzacji cząsteczki RNA, oddziaływania elektrostatyczne i niekanoniczne wpływają na zwijanie się nici RNA, a kanoniczne parowanie się reszt nukleotydowych jest wynikiem minimalizacji energii i poprzednich oddziaływań [23, 80].

Przedstawienie struktury drugorzędowej kwasów nukleinowych w postaci nici sekwencji reszt nukleotydowych z zaznaczonymi parowaniami kanonicznymi jest powszechne i czytelne, jednak powinno być dedykowane dla cząsteczek DNA. RNA wymaga dokładniejszej reprezentacji oddziaływań wewnątrzcząsteczkowych. Próby przedstawienia struktury drugorzędowej w dokładniejszy sposób były i są podejmowane [17, 81, 82] lecz jedna konkretna metoda nie przebiła się jeszcze do powszechnego użytku.

Ujawnienie parowań kanonicznych bez znajomości struktury przestrzennej jest osiągalne z dość dobrą dokładnością dzięki eksperymentom próbkowania struktury drugorzędowej w połączeniu z programami do przewidywania struktury drugorzędowej [83, 84]. Oddziaływania niekanoniczne nie są tak łatwe w identyfikacji. Metody eksperymentalne poznawania struktury drugorzędowej dążą jednak w tym kierunku. Od momentu publikacji w 2005 roku oznaczania struktury RNA metodą SHAPE [47], metoda ta została rozwinięta poprzez włączenie dodatkowych czynników próbkujących, a także bardziej skomplikowanych metod próbkowania jak mapy mutacyjne [85], które poprzez zamianę reszt nukleotydowych pozwalają na badanie oddziaływań poprzez przestrzeń. Podobnie, użycie rodników

hydroksylowych generowanych *in situ*, które dostarcza informacji o ogólnym kształcie cząsteczki [86] zostało rozwinięte do metody mapowania rodnikami (metoda MOHCA), do stopnia w którym generator rodników jest kowalencyjnie przyłączony do jednej z reszt nukleotydowych, co dostarcza informacji o położeniu poszczególnych nukleotydów względem zmodyfikowanej reszty [49].

Nie jest właściwe myślenie o strukturze RNA jako strukturze trwałej [87]. Struktura ta zmienia się i dostosowuje do warunków w jakich się znajduje. pH, obecność soli, białek czy ligandów wpływa na strukturę RNA [88, 89]. Struktury przestrzenne RNA nawet ze zdjęć krystalograficznych obrazują pewien stan w jakim badane RNA przebywa, czasami stany te są bardziej lub mniej trwałe. Dokładniejsze zrozumienie struktury RNA jako zbioru wielu uporządkowanych oddziaływań wewnątrzcząsteczkowych, ale dostosowujących się do warunków zewnętrznych i oddziaływań międzycząsteczkowych przyczyni się do lepszego zrozumienia funkcji RNA.

6. CELE NAUKOWE PRZEDSTAWIONEGO CYKLU PRAC

Głównym celem przedstawionego cyklu prac naukowych było badanie mechanizmów ewolucji RNA i ich wpływu na strukturę i funkcję tych cząsteczek oraz wykorzystanie funkcji cząsteczek RNA zakodowanych w ich strukturze do projektowania funkcjonalnych nanocząstek. Realizacja celów opierała się na eksperymentach obliczeniowo-predykcyjnych.

Cele szczegółowe pokrywają się z tematyką poszczególnych publikacji wchodzących w skład cyklu.

1. Celem pracy Biesiada et al. *Methods Mol Biol.* 2016 było przedstawienie cech programu RNAComposer jako zaawansowanego narzędzia w przewidywaniu struktur przestrzennych RNA oraz określenie jego możliwości i ograniczeń. A także analiza wpływu dokładności/poprawności wprowadzonej struktury drugorzędowej na otrzymany model struktury przestrzennej.

RNAComposer został wykorzystany jako narzędzie obliczeniowe w dwóch pozostałych pracach.

2. Celem pracy Biesiada et al. *Methods* 2016 było opracowanie schematu modelowania nanostruktur RNA, jak również ukazanie potencjału RNAComposer w modelowaniu nanostruktur cząstek RNA, w podejściu *de novo* i w wymiarze pełnoatomowym.

3. Celem badań przedstawionych w pracy Biesiada et al. *Nucleic Acids Res.* 2022 było studiowanie procesów ewolucji na poziomie molekularnym. Praca stanowi bardzo dobry przykład wykorzystania modelowania struktur RNA do badania skomplikowanych procesów biologicznych. Naszym głównym zadaniem było zrozumienie mechanizmu ekspansji cząsteczek rybosomalnych RNA od organizmów mniej złożonych do bardziej skomplikowanych oraz jej wpływu na strukturę przestrzenną siódmego segmentu ekspansyjnego rybosomalnego RNA (ES7 rRNA). Przedstawiłem również możliwości modelowania RNA przy użyciu programu RNAComposer w wymiarze pełnoatomowym dla bardzo dużych cząsteczek RNA, na przykładzie rRNA dużej podjednostki rybosomalnej myszy.

7. SKRÓTOWY OPIS PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

Automated RNA 3D Structure Prediction with RNAComposer

Marcin Biesiada, Katarzyna J. Purzycka, Marta Szachniuk, Jacek Błażewicz and Ryszard W. Adamiak*

Methods in Molecular Biology, 2016

W przypadku kwasów rybonukleinowych określenie struktury metodami eksperymentalnymi jest ograniczone przez rozmiar tych cząsteczek, ich stabilność i możliwość przyjmowania wielu konformacji. W niniejszej pracy przedstawiam koncept zaawansowanego narzędzia bioinformatycznego do przewidywania struktur przestrzennych RNA jakim jest RNAComposer oraz jego aplikację z punktu widzenia użytkownika. Analizowałem również jak dokładność wprowadzonej struktury drugorzędowej wpływa na otrzymany model struktury przestrzennej.

7.1.1. Przewidywanie struktur RNA

Metody obliczeniowe służące do przewidywania struktur przestrzennych RNA bazują na symulacji zwijania cząsteczki RNA, modelowaniu komputerowym lub składaniu fragmentów. Większość z nich wymaga od użytkownika dużego zaangażowania oraz znajomości procesu modelowania. Istnieje jedynie kilka metod które są metodami automatycznymi, tzn. na podstawie stosunkowo prostych informacji – sekwencji i/lub struktury drugorzędowej RNA przewidują model struktury przestrzennej. Jednym z takich programów jest RNAComposer [15], który w bardzo krótkim czasie generuje struktury przestrzenne RNA w postaci pełnoatomowej, w oparciu o fragmenty obecne w jego bazie danych elementów strukturalnych RNA [34].

7.1.2. RNAComposer - webservice

RNAComposer jest dostępny dla użytkowników jako aplikacja webserwerowa w dwóch trybach – interaktywnym oraz w trybie seryjnych danych wejściowych. Aplikacja dostępna jest pod adresem: <http://rnacomposer.ibch.poznan.pl> oraz na stronie lustrzanej pod adresem: <http://rnacomposer.cs.put.poznan.pl>.

Tryb interaktywny

Tryb interaktywny jest dostępny dla każdego użytkownika w ograniczonym zakresie, w dowolnym momencie. Ten tryb dedykowany jest do przewidywania krótkołańcuchowych RNA o prostej strukturze drugorzędowej, którą można z dużym prawdopodobieństwem przewidywać z wykorzystaniem dostępnych algorytmów.

W tym trybie użytkownik może uruchamiać procesy modelowania cząsteczek RNA na podstawie sekwencji lub sekwencji i topologii struktury drugorzędowej. W przypadku podania jedynie sekwencji RNA, RNAComposer wykorzystuje jeden z trzech dostępnych algorytmów do przewidywania topologii struktury drugorzędowej: RNAstructure [54], CONTRAfold [55] lub RNAfold [56]. Generowanie pełnoatomowej struktury 3D jest bardzo szybkie, odbywa się w ciągu od kilku sekund do kilku minut, w zależności od wielkości i poziomu skomplikowania struktury. RNAComposer może przewidywać struktury przestrzenne cząsteczek RNA o wielkości do 500 reszt nukleotydowych. Wygenerowany model struktury 3D podawany jest w postaci pliku pdb, zawierającego koordynaty poszczególnych atomów budujących strukturę RNA oraz odnośnik do wizualizacji za pomocą Jmol [90].

Tryb seryjnych danych wejściowych

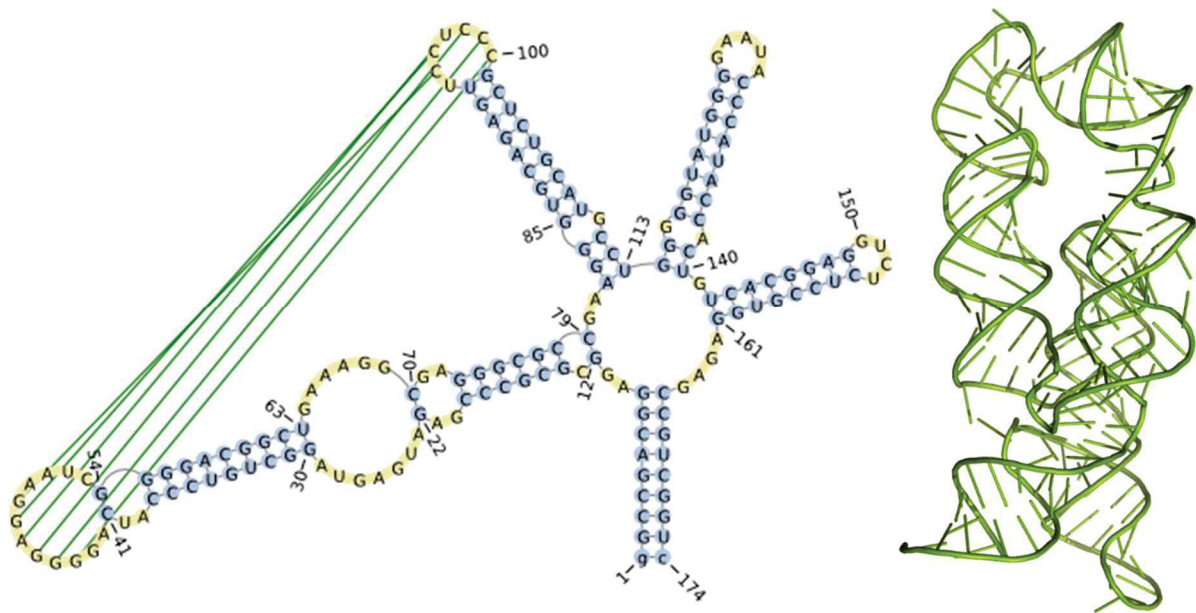
Drugi tryb, tryb seryjnych danych wejściowych, jest dostępny jedynie dla zalogowanych użytkowników. Użytkownik po zalogowaniu się do serwera RNAComposer zyskuje dostęp do dodatkowych funkcjonalności w porównaniu do trybu interaktywnego. W tym trybie dostępne jest przewidywanie do dziesięciu cząsteczek RNA jednocześnie. Dodatkowo do każdej cząsteczki można przypisać do dziesięciu różnych struktur drugorzędowych, a dla każdej struktury wygenerować do dziesięciu różnych modeli 3D. Użytkownik ma możliwość wprowadzenia więzów na odległość pomiędzy poszczególnymi atomami. Atomy te nie muszą być ze sobą związane bezpośrednio wiązaniem kowalencyjnym, co pozwala np. na modelowanie oddziaływań wyższego rzędu. Dodatkowo, wynik każdego eksperymentu modelowania jest zapisywany na serwerze i przechowywany przez określony okres czasu. Wynikiem modelowania jest struktura przestrzenna RNA oraz wszelkie dane dotyczące przetwarzania informacji przez RNAComposer w postaci zbioru plików, np. poszczególne fragmenty wykorzystane do modelowania, czy przebieg minimalizacji energii struktury.

7.1.3. Przykłady przewidywania – ryboprzełącznik lizynowy

W niniejszej pracy badaliśmy jak dokładność wprowadzonej struktury drugorzędowej wpływa na otrzymany model struktury przestrzennej RNA. Jako cząsteczkę modelową wybraliśmy ryboprzełącznik lizynowy z *Thermotoga maritima* (174 reszt nukleotydowych). Struktura krystalograficzna (PDB: 3DIL) tej cząsteczki składa się z trzonu prowadzącego do pętli pięcioramiennej, która rozgałęzia się na cztery rejony helikalne zakończone pętlami apikalnymi i helisę uformowaną przez sparowanie 5' i 3' końców cząsteczki (Rysunek 7). Pomiedzy dwoma pętlami apikalnym zachodzi oddziaływanie wyższego rzędu – pseudowęzeł typu H, który dodatkowo stabilizuje strukturę cząsteczki [91].

Początkowo struktura drugorzędowa cząsteczki ryboprzełącznika lizynowego była przewidywana *in silico* z wykorzystaniem różnych algorytmów. Do przewidywania struktury trzeciorzędowej została wytypowana struktura 2D najbardziej zbliżona do struktury krystalograficznej (PDB: 3DIL) uzyskana za pomocą programu KineFold [92]. Wykorzystanie

metod eksperymentalnego próbkowania struktury RNA poprawia przewidywanie struktury drugorzędowej, lecz wykorzystanie standardowych odczynników próbkujących nie jest wystarczające by potwierdzić występowanie struktur RNA wyższego rzędu, jak np. pseudowęzły, i ich identyfikacja wymaga dodatkowych analiz. W przypadku ryboprzełącznika lizynowego, pseudowęzeł typu pętla-pętla (ang. *kissing hairpins*) był postulowany, a następnie potwierdzony eksperymentalnie poprzez mapowanie struktury z wykorzystaniem rybonukleazy H [93].



Rysunek 7. Struktura drugorzędowa (lewa strona) i trzeciorzędowa (prawa strona) otrzymane ze struktury krystalograficznej RNA ryboprzełącznika lizynowego (PDB: 3DIL).

Do przewidywania struktury przestrzennej analizowanego RNA z wykorzystaniem programu RNAComposer przygotowałem cztery warianty danych wejściowych zawierające tę samą sekwencję RNA oraz:

- (1) strukturę przewidzianą *in silico*, najbardziej zbliżoną do struktury krystalograficznej
- (2) strukturę przewidzianą na podstawie danych eksperymentalnych (bez pseudowęzła),
- (3) strukturę przewidzianą na podstawie rozszerzonych danych eksperymentalnych z pseudowęzłem wprowadzonym jako więzy strukturalne,
- (4) strukturę przewidzianą na podstawie rozszerzonych danych eksperymentalnych z pseudowęzłem wprowadzonym w postaci zapisu kropkowo-nawiasowego.

Na potrzeby przewidywania baza elementów strukturalnych RNAComposer została pomniejszona o elementy uzyskane na podstawie struktury 3DIL [91], czyli struktury krystalograficznej ryboprzełącznika lizynowego. Dla każdego wariantu generowałem 10 modeli, lecz ze względu na najlepsze dopasowanie i najniższą energię modelu generowanego jako pierwszy, jego struktura przestrzenna była porównywana między wariantami.

Wyniki przeprowadzonych analiz wykazały, że dokładniejsza struktura drugorzędowa cząsteczki pozwalała na bardziej zbliżone do modelu rzeczywistego przewidywanie modelu przestrzennego, co przedstawia Tabela 1 poniżej. W przypadku wariantu, którego struktura drugorzędowa była generowana jedynie *in silico*, struktura pętli pięcioramiennej odbiegała znacząco od rzeczywistej i brakowało dla niej referencyjnych elementów struktury przestrzennej w bazie elementów strukturalnych RNAComposer. Brakujące elementy są modelowane w formie regularnej helisy typu A, co w przypadku elementów jednoniciowych zmniejsza jakość przewidywanej struktury przestrzennej. Dla wariantu eksperymentalnego, w którym nie wprowadziłem pseudowężła, przewidywanie struktury przestrzennej znacznie się poprawiło w porównaniu z wariantem pierwszym. Poważniejsze odchylenia od struktury rzeczywistej pojawiły się w rejonie pseudowężła. Po wprowadzeniu więzów strukturalnych na odległości między nukleotydami zaangażowanymi w tworzenie pseudowężła, struktura w tym obszarze uległa poprawieniu. Dodanie tych więzów spowodowało zbliżenie do siebie pętli apikalnych wywołując zgięcie bezpośrednio przylegających rejonów helikalnych, co z kolei miało wpływ na zmianę struktury w pozostałej części cząsteczki RNA. Pomimo lokalnej poprawy struktury, globalnie, struktura całej cząsteczki poprawiła się nieznacznie. Zamiana reprezentacji pseudowężła w danych wejściowych z więzów strukturalnych na reprezentację kropkowo-nawiasową sprawiła, że przewidziany model 3D był niemal identyczny ze strukturą krystalograficzną.

Struktura 2D	RMSD [Å]	Energia obliczona
(1) <i>In silico</i> wygenerowana przez KineFold	24.82	-3642.483
(2) Eksperymentalna bez pseudowężła	6.91	-4494.975
(3) Eksperymentalna z więzami strukturalnymi na pary tworzące pseudowężel	5.82	-4245.223
(4) Eksperymentalna z pseudowężłem oznaczonym przez reprezentację kropkowo-nawiasową	1.72	-4628.297

Tabela 1. Porównanie parametrów wygenerowanych modeli 3D rybobprzełącznika lizynowego.

7.1.4. Wnioski

RNAComposer jest dobrym narzędziem do przewidywania struktur przestrzennych RNA, bardzo przyjaznym użytkownikowi. Za jego pomocą można przewidywać strukturę cząsteczek RNA w pełnoatomowej postaci, o długości do 500 reszt nukleotydowych.

Niniejsza praca potwierdza, że przewidywanie struktur przestrzennych RNA z wykorzystaniem RNAComposer jest silnie zależne od jakości wprowadzanych danych. Im dane wejściowe będą bardziej odzwierciedlały przewidywaną strukturę, tym wyższe prawdopodobieństwo otrzymania dokładnego modelu 3D [15]. Rola użytkownika jest kluczowa podczas oceny jakości przewidywanego modelu. Użytkownik powinien zweryfikować energię przewidywanego modelu, a także obecność przeplotów łańcucha

głównego, które mogą wskazywać na błędy w przewidywaniu. Jakość przewidywanego modelu mogą obniżyć również elementy, które nie są obecne w bazie RNA FRABASE [34] przez co są generowane przez RNAComposer w oparciu o strukturę 3D helisy α .

Obecnie użytkownik posiada szereg dodatkowych funkcji zawartych w RNAComposer polepszających jakość przewidywanych modeli RNA. Do systemu RNAComposer zostały wprowadzone algorytmy CentroidFold [57], ContextFold [58] i IPknot [59], które zapewniały najlepsze wyniki przewidywania struktur drugorzędowych. W trybie interaktywnym użytkownik ma również do dyspozycji plik z danymi koordynatów atomów fragmentów budujących strukturę modelu RNA. Tryb seryjnych danych wejściowych został natomiast poszerzony o dodawanie wiązań strukturalnych na kąty torsyjne. Została dodana funkcja filtrowania bazy danych pozwalająca wykluczyć z niej struktury o określonej rozdzielczości lub struktury zdefiniowane przez identyfikator PDB. Można również wymusić przewidywanie fragmentów jednoniciowych lub dwuniciowych w postaci struktur A-RNA, jak również wprowadzać do modelowania własne fragmenty strukturalne [94].

RNAComposer and RNA 3D structure prediction for nanotechnology

Marcin Biesiada, Katarzyna Pachulska-Wieczorek, Ryszard W. Adamiak*, Katarzyna J. Purzycka*

Methods, 2016

Korelacja pomiędzy strukturą i rolą RNA może być wykorzystana podczas projektowania funkcjonalnych cząsteczek RNA. Niniejsza praca prezentuje przykład wykorzystania systemu RNAComposer do modelowania programowalnych nanocząstek RNA.

7.1.5. Nanotechnologia RNA

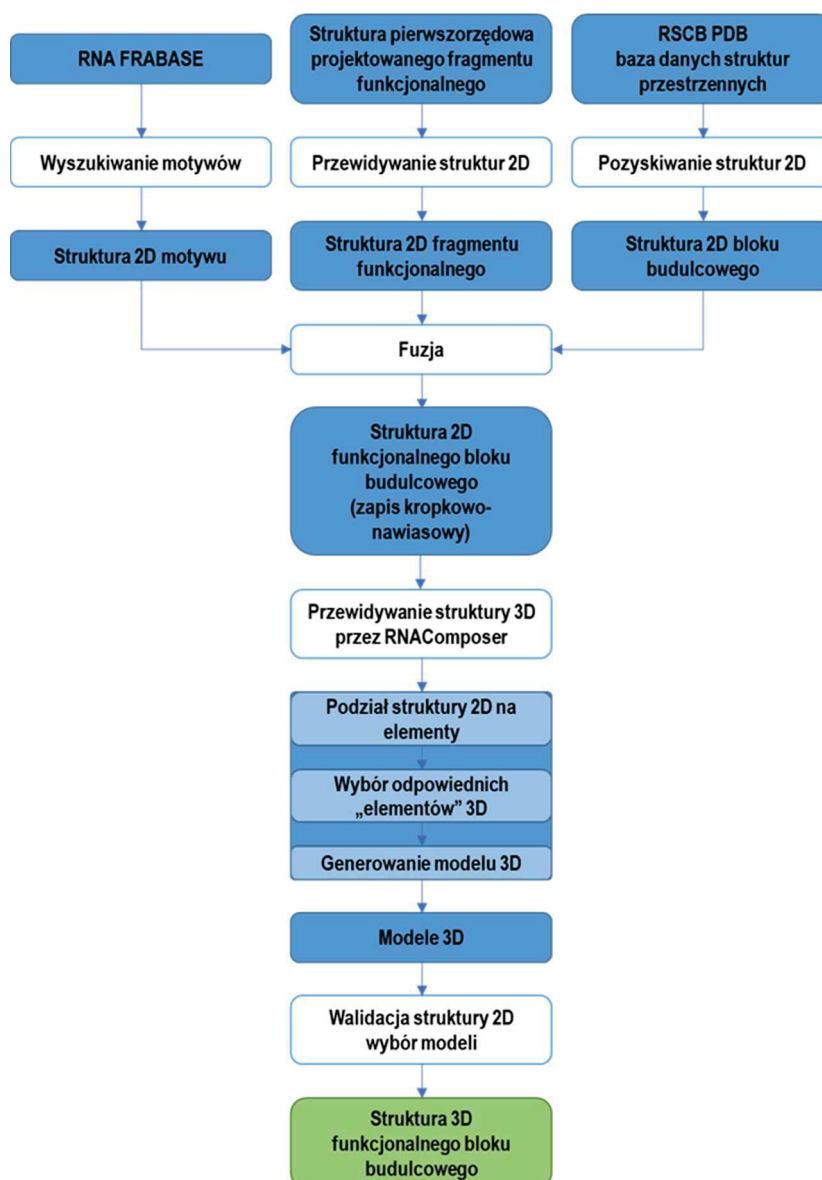
RNA jest elementem występującym naturalnie w komórkach i pełni różnorodne funkcje. Występuje w postaci katalizatora procesów komórkowych [95], jest nośnikiem informacji genetycznej [73], odpowiada za transport cząsteczek [96] czy komunikację międzykomórkową [97]. Takie funkcje RNA są interesującym celem badań nad wykorzystaniem tych wszechstronnych cząsteczek w nanotechnologii [98].

Od opublikowania naszej pracy w Methods tworzone są coraz nowsze nanocząstki RNA, jak rybobprzełączniki pozwalające wykrywać wirusa Ebola [99], czy syntetyczne RNA pozwalające regulować transkrypcję [100]. Wprowadzanie syntetycznych nanocząstek RNA do komórek będzie wiązało się z wprowadzeniem do sekwencji RNA modyfikowanych reszt nukleotydowych. Podejście takie okazało się skuteczne podczas projektowania szczepionki przeciw SARS-CoV2 [101]. Wprowadzanie pseudourydyny, 5-metylocytozyny czy 6-metyloadenozyny lub innych modyfikowanych nukleotydów powinno następować na etapie projektowania nanoczątek RNA. Wprowadzone modyfikacje wpływają na pole siłowe otaczające resztę nukleotydową, przez co wpływają na parowanie się zasad czy oddziaływania warstwowe reszt nukleotydowych [102], a co za tym idzie na ostateczny kształt nanocząstki. Modyfikowane reszty nukleotydowe mogą stabilizować strukturę nanocząstki RNA [103] jak i wpływać na jej funkcje [104]. Przykładem praktycznego użycia nanocząstki RNA, jest nanocząstka 4WJ-X wykorzystywana do transportu leku przeznaczonego do leczenia nowotworów piersi. Funkcjonalna nanocząstka RNA podnosi rozpuszczalność leku oraz jego termostabilność, przez co staje się on o wiele bardziej skuteczny w terapii [105].

7.1.6. Schemat modelowania nanostruktur RNA

Modelowanie nanostruktur RNA zaproponowaną przez nas metodą (Rysunek 8) rozpoczyna się od zaprojektowania struktury drugorzędowej nanocząstki o zaprogramowanych funkcjach. Przygotowanie takiej struktury polega na przeszukaniu odpowiedniej bazy strukturalnej w celu zidentyfikowania swoistego motywu strukturalnego, który utworzy zdefiniowane konstrukty. Poszukiwanie motywu można przeprowadzić w bazie RNA FRABASE [34]. Jednocześnie należy zaprojektować sekwencję fragmentu funkcjonalnego i przewidzieć jego strukturę drugorzędową w celu weryfikacji, czy przewidywany model jest zgodny z założeniami. W tym samym czasie należy wykorzystać bazę danych RSCB PDB [60] do wyszukiwania struktur

drugorzędowych poszczególnych bloków budulcowych. Zebrane informacje z baz RNA FRABASE, RCSB PDB oraz algorytmów do przewidywania struktur 2D RNA łączy się w celu uzyskania struktury drugorzędowej nanocząstki RNA, w postaci zapisu kropkowo-nawiasowego. Reprezentacja ta zostanie wprowadzona jako dana wejściowa do przewidywania struktury przestrzennej przez RNAComposer. Zaplanowana w ten sposób struktura 2D cząsteczki RNA ulega fragmentacji przez system RNAComposer, następuje wyszukiwanie i dobór elementów przestrzennych, złożenie w pełną strukturę i ułożenie poprzez minimalizację energii wiązań i kątów. W kolejnym etapie weryfikuje się otrzymany model nanostruktury. Należy sprawdzić, czy wygenerowany model zachowuje swoją pierwotną strukturę drugorzędową i czy w strukturze przestrzennej nie występują krytyczne odchylenia od struktury zaplanowanej. Do tego celu można wykorzystać RNApdbec [81], narzędzie, które tłumaczy strukturę przestrzenną na strukturę drugorzędową.



Rysunek 8. Schemat zaproponowanej metody modelowania nanocząstek RNA

7.1.7. RNA prokapsydu bakteriofaga phi29

RNA prokapsydu bakteriofaga $\phi 29$ ze względu na swoją budowę jest wszechstronnie wykorzystywany w dziedzinie nanotechnologii [106]. Element ten współuczestniczy w pakowaniu DNA do wnętrza prokapsydu i składa się w pięcioelementowy pierścień [107]. W prezentowanych badaniach wykorzystałem go do walidacji zaproponowanej metody modelowania programowalnych nanocząstek RNA. Strukturę pierwszorzędową podstawowego meru uzyskaliśmy ze struktury przestrzennej zdeponowanej w bazie danych RCSB PDB [60]. Struktura drugorzędowa została przewidziana za pomocą algorytmu RNAstructure [108]. Uzyskana struktura drugorzędowa została porównana ze strukturą drugorzędową RNA prokapsydu bakteriofaga $\phi 29$ [109] i wprowadzona jako dana wejściowa do systemu RNAComposer. Przed rozpoczęciem przewidywania wszystkie struktury RNA prokapsydu bakteriofaga $\phi 29$ (1FOQ [107], 2KVN [110], 3RKF [111], 4KZ2 [112]) zostały wykluczone z bazy danych elementów strukturalnych, aby zaaranżować problem braku bezpośrednich homologów strukturalnych dla zaprojektowanej struktury drugorzędowej. RNAComposer dzieli strukturę drugorzędową na fragmenty, fragmentom przypisuje odpowiednie reprezentacje przestrzenne i na tej podstawie przewiduje strukturę. Fragmenty dobierane są pod względem podobieństwa sekwencji i zgodności pirymidyno-purynowej. W przypadku przewidywanej struktury RNA prokapsydu homologia większości wybranych fragmentów była pełna, lecz kilka elementów, w tym pętla trójramienna wykazały niską homologię zarówno sekwencyjną, jak i pirymidyno-purynową. Mimo to otrzymany model 3D nakładał się ze strukturą rzeczywistą, a pozycja nukleotydów uczestniczących w oddziaływaniach międzycząsteczkowych poprzez pętle sugeruje, że takie oddziaływania mogą powstawać i tworzyć pięciocłonowy nanopierścień RNA.

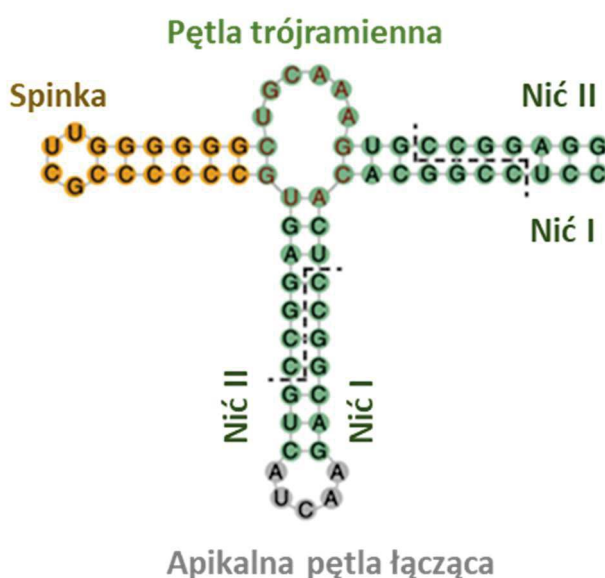
7.1.8. Struktura nanokwadratu RNA o programowalnych funkcjach

W kolejnym etapie testowaliśmy możliwości programu RNAComposer w dziedzinie modelowania nanocząstek RNA. Strukturami na których oparłem modelowanie były fragmenty RNA, które samodzielnie związały się w strukturę nanokwadratu [113]. Struktura nanokwadratu posłużyła nam za platformę do tworzenia funkcjonalnych cząsteczek RNA poprzez wprowadzenie dodatkowych elementów struktury w jego narożnikach. Wprowadzałem tam struktury w postaci spinek do włosów o programowalnych właściwościach. Przykładowo struktury tych spinek mogą zawierać sekwencje rozpoznawane przez białko Dicer czy motywy wiążące ligandy pozwalające na katalizę lub inhibicję procesów komórkowych [98, 104]. Pętle apikalne mogą być także projektowane tak, aby tworzyć połączenia typu *kissing-loop* pomiędzy poszczególnymi jednostkami nanokwadratów, co pozwala budować nanostruktury wyższego rzędu, np. biopolimerowe przestrzenne sieci [114].

W niniejsze pracy skupiłem się na wprowadzeniu spinki, która mogłaby tworzyć nieskończone biopolimerowe sieci RNA. Struktura wprowadzanej spinki zaprojektowałem tak, aby była stabilna, a pętla apikalna tej spinki miała potencjał do tworzenia oddziaływań

z inną spinką sąsiadującego nanokwadratu. Stabilność rejonu helikalnego spinki zapewniają pary kanoniczne GC, które są trwalsze niż pary AU czy GU. Sekwencja pętli apikalnej została przygotowana poprzez wyszukanie w bazie danych RNA FRABASE [34] pętli apikalnych o długości 4-10 reszt nukleotydowych tworzących oddziaływania typu pseudowęzeł. Przygotowałem 15 różnych struktur spinek, które tworzyły oddziaływania wyższego rzędu, o kątach pomiędzy rejonami helikalnymi oddziałujących cząsteczek zawierającymi się w zakresie $130^\circ - 170^\circ$. Aby wprowadzić spinkę w narożniku, pętla wewnętrzna struktury bazowej nanokwadratu została wymieniona na pętlę trójramienną. Konstruuując pętlę trójramienną ograniczyłem się do wyszukania pętli, w których pierwsza nić nie zawiera niesparowanych reszt nukleotydowych, natomiast dwie kolejne nici zawierają od 1 do 10 niesparowanych reszt nukleotydowych. Ograniczenie to miało na celu znalezienie takich pętli trójramiennych, które nie wpływałyby na strukturę pozostałych elementów nanocząstki, a z drugiej strony maksymalizowałyby wyszukanie pętli trójramiennych zachowujących kąt prosty naroża nanokwadratu. W konsekwencji znalazłem 10 różnych pętli trójramiennych spełniających przyjęte założenia.

Na każdym etapie projektowania nanocząstki (wprowadzania nowych elementów, czy zmiany sekwencji), zgodność koncepcji nowo przygotowanej struktury była sprawdzana z przewidywaniem za pomocą skryptu RNAstructure [108]. Przygotowane sekwencje oraz struktury drugorzędowe zostały wprowadzone jako dane wejściowe do przewidywania struktur przestrzennych przez RNAComposer. Ze względu na to, że RNAComposer przewiduje struktury przestrzenne pojedynczych cząsteczek zbudowanych z jednej ciągłej nici RNA, do przewidywania został stworzony konstrukt złożony z dwóch helis tworzących dwa boki nanokwadratu oraz łączącej je pętli trójramiennej i wychodzącej z niej spinki RNA. W takiej postaci uzyskałem dwie nici RNA, które na potrzeby przewidywania zostały połączone zamykającą je pętlą apikalną (Rysunek 9).



Rysunek 9. Konstrukt przewidywanej cząstki RNA, składowa nanokwadratu RNA o programowalnych funkcjach.

Po zakończeniu przewidywania powyższego konstrukt został on powielony czterokrotnie, a nici tworzące ramiona (kolor zielony, Rysunek 9) nałożone na siebie tak by tworzyły nanokwadrat. Nadmiarowe nukleotydy zostały wycięte, a struktura nowego nanokwadratu została zweryfikowana pod względem poprawności konformacji struktury przestrzennej [115].

7.1.9. Wnioski

Przeprowadzone eksperymenty potwierdziły możliwość wykorzystania programu RNAComposer do projektowania nanocząstek RNA pomimo jego pewnych ograniczeń. Wykazaliśmy, że RNAComposer jest w stanie przewidywać z dobrą dokładnością skomplikowane cząsteczki RNA, jak RNA prokapsydu bakteriofaga ϕ 29, nawet w przypadku braku w bazie danych struktur referencyjnych pochodzących od eksperymentalnie rozwiązanych struktur dla tego RNA. Jest to istotne w kontekście projektowania nanocząstek RNA, które nie bazują na cząsteczkach RNA o znanej strukturze przestrzennej.

Projektowany nanokwadrat składa się z ośmiu niezależnych nici RNA, podczas gdy RNAComposer przewiduje struktury jednoniciowych RNA. Aby obejść to ograniczenie przewidywany konstrukt musiał być odpowiednio zaprojektowany na potrzeby modelowania, tak aby po jego zakończeniu i powrocie do wyjściowej struktury nici tworzył wiarygodną nanocząstkę. Przewidywanie nanokwadratu pokazało odmienny sposób przewidywania struktury cząsteczek RNA. W klasycznym podejściu struktura drugorzędowa (lub tylko sekwencja) jest daną wejściową do przewidywania przez RNAComposer, a wynikiem jest struktura trzeciorzędowa. W tym podejściu niejako zaczynamy od struktur trzeciorzędowych i przeszukiwania RNA FRABASE, lub innych dostępnych baz struktur przestrzennych RNA, w poszukiwaniu interesujących bloków do tworzenia nanostruktur i na ich podstawie tworzona jest struktura drugorzędowa, która poddawana jest procesowi przewidywania. Ponadto przewidywanie struktury drugorzędowej jest wykorzystane w innym celu niż w klasycznym przewidywaniu. W modelowaniu przez RNAComposer wychodzącym od sekwencji RNA pierwszym etapem jest przewidywanie struktury 2D. W projektowaniu nanostruktur konstruowanie struktur 2D spada w większym stopniu na użytkownika. Przewidywanie struktur 2D wykorzystane jest po otrzymaniu wariantu struktury przestrzennej, do weryfikacji, czy zaproponowana struktura tworzy takie pary jak w zaproponowanym modelu wyjściowym i czy zaproponowany model 3D jest wiarygodny.

rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles

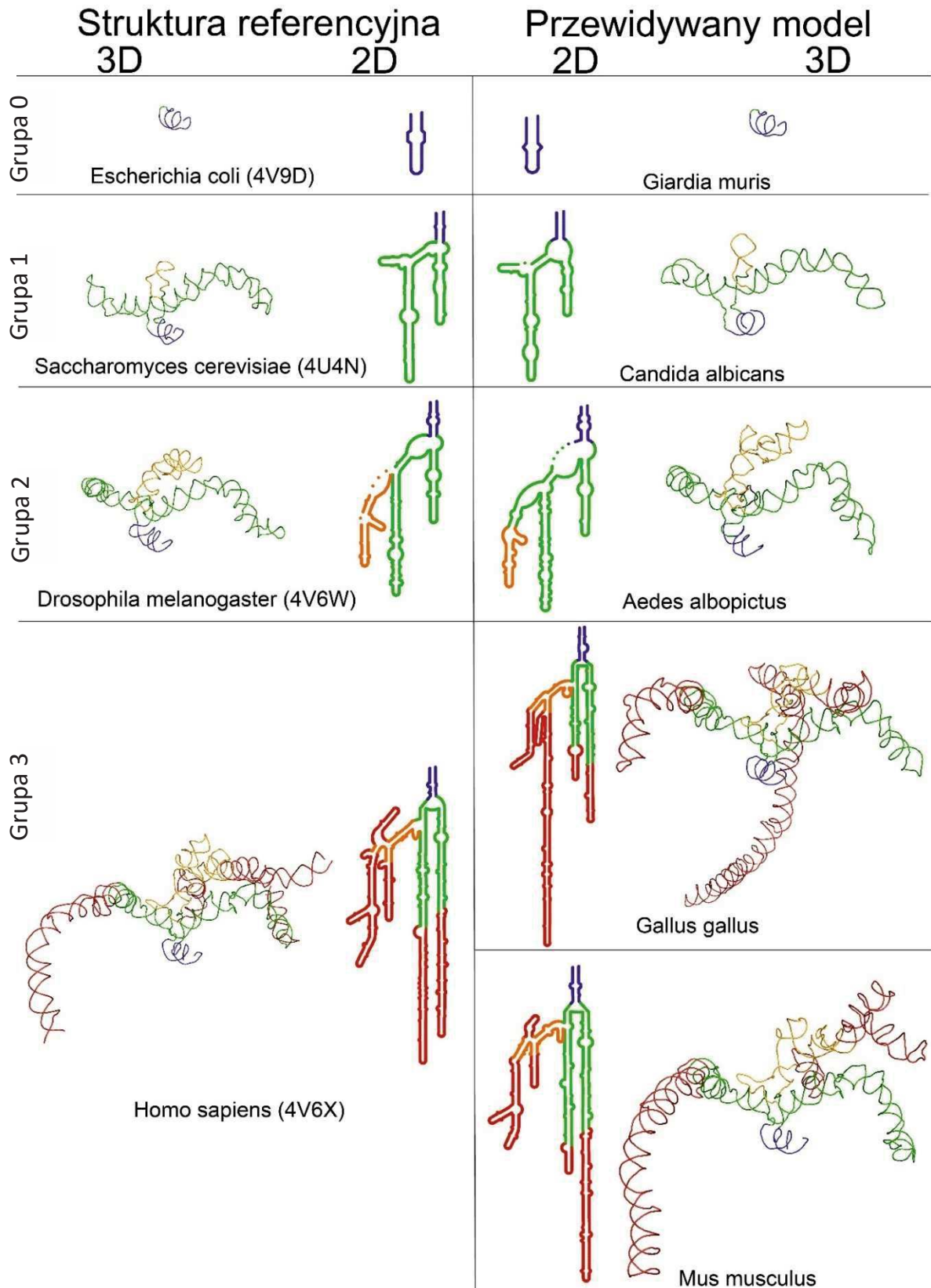
Marcin Biesiada, Michael Hu, Loren Dean Williams*, Katarzyna J. Purzycka* and Anton S. Petrov*

Nucleic Acids Research, 2022

Ewolucja organizmów przebiega nie tylko na poziomie widocznych cech zewnętrznych, ale również na poziomie molekularnym [116]. W niniejszej pracy przedstawiliśmy metodologiczny schemat postępowania prowadzący do uzyskania struktury drugo- i trzeciorzędowej rRNA w oparciu o eksperymentalnie oznaczone struktury 3D rybosomów, model akrecyjny (przyrostowy) ewolucji rybosomalnej, pokrewieństwo filogenetyczne organizmów, algorytmy zwijania RNA i modelowanie przestrzenne z wykorzystaniem programu RNAComposer [50, 117, 118].

7.1.10. Siódmy segment ekspansyjny

Porównując rRNA pochodzące od różnych gatunków obserwujemy różnorodność struktury tych cząsteczek na powierzchni rybosomu [119]. Im wyższy poziom rozwoju organizmu tym rRNA jest większe, bardziej kompleksowe, a zmiany obserwowane na powierzchni rybosomu, daleko od jego centrum aktywnego, coraz bardziej widoczne. Elementy rRNA, które ulegają rozbudowaniu nazywane są segmentami ekspansyjnymi. Najbardziej zmiennym rejonem rRNA, który jest przedmiotem niniejszej pracy, jest siódmy segment ekspansyjny (ES7) dużej podjednostki rybosomalnej. Segment ten ewoluował z kilkunastonukleotydowej struktury typu spinka u organizmów prokariotycznych, oznaczonej jako helisa 25 (H25) w rRNA dużej podjednostki rybosomalnej, do wielokrotnie rozgałęzionych struktur u ssaków, zbudowanych z ponad 750 reszt nukleotydowych. ES7 jest istotnym elementem rRNA. Udowodniono, że usunięcie tego elementu z rRNA drożdży jest śmiertelne dla tych organizmów [120-122]. ES7 jest miejscem wiązania białek rybosomalnych. Struktura rRNA człowieka, zdeponowana w bazie RCSB PDB [60], pokazuje jego silne oddziaływanie z białkiem eL6, ale ES7 oddziałuje również z innymi białkami jak eL14 [123]. Sugeruje się, że zmiany konformacyjne struktury ES7 przyczyniają się do wiązania cząsteczki rozpoznającej sygnał, kompleksu nukleoprotein odpowiedzialnego za kierowanie białek poza rybosom [124]. ES7 we wczesnej fazie formowania się rybosomu oddziałuje z czynnikami wspomagającymi ten proces [125-128]. ES7 ulega hydrolizie pod wpływem działania jonów żelaza Fe^{2+} , co prowadzi do rozpadu cząstki rRNA podczas stresu oksydacyjnego [129]. U kręgowców ES7 uczestniczy w syntezie selenoprotein [130].



Rysunek 10. Struktury drugo- i trzeciorzędowe przewidywanych cząsteczek RNA ES7 wraz z wykorzystanymi do porównania strukturami referencyjnymi.

W niniejszej pracy wyróżniliśmy cztery grupy struktur ES7. W celu usystematyzowania pogrupowaliśmy je ze względu na rozmiar i złożoność struktury. Grupa podstawowa (grupa 0) obejmuje ES7 najmniej złożonych organizmów, których RNA ES7 zbudowane jest z rejonu helikalnego zakończonego pętlą apikalną oznaczaną jako helisa 25 (H25). W grupie 1 pętla apikalna z grupy 0 zastąpiona jest pętlą trójramienną, z której wychodzą dwa dodatkowe rejonu helikalne. Jeden z nich rozrasta się na dwa kolejne rejonu helikalne poprzez kolejną pętlę trójramienną. W grupie 2 peryferyjne spinki ulegają przede wszystkim wydłużeniu, a w większości analizowanych struktur jedna z nich ulega kolejnemu rozgałęzieniu poprzez pętlę trójramienną. W grupie 3 następują kolejne ekspansje rejonów helikalnych, lecz w odróżnieniu od wcześniejszych grup są to wydłużenia wielonukleotydydowe oraz kolejne rozgałęzienia poprzez pętle trójramienne, a u najbardziej zaawansowanych ewolucyjnie kręgowców pojawiają się pętle czteroramienne (przykłady struktur ES7 - Rysunek 10).

7.1.11. Zwój sygnowy siódmego segmentu ekspansyjnego

Ważnym elementem struktury ES7 jest zwój sygnowy (ang. *signature fold*), jest to fragment struktury RNA ES7 o zakonserwowanej strukturze. Zwój sygnowy jest obecny w ES7 należących do grup 1-3. Na jego strukturę składają się trzon elementu ekspansyjnego (odpowiednik helisy 25 obecny u organizmów takich jak bakterie *E. coli*, czy wiciowce), dwie pętle trójramienne oraz wychodzące z nich trzy helisy. Wyjątek stanowią RNA ES7 z *Yarrowia lipolytica* oraz *Leishmania donovani*, zaklasyfikowanych w grupie 1, które zawierają część zwoju sygnowego ze względu na obecność tylko jednego rozgałęzienia. Zakonserwowanie struktury zwoju sygnowego pozwala na zidentyfikowanie tego elementu w analizowanych sekwencjach.

7.1.12. Modelowanie struktur RNA siódmego segmentu ekspansyjnego

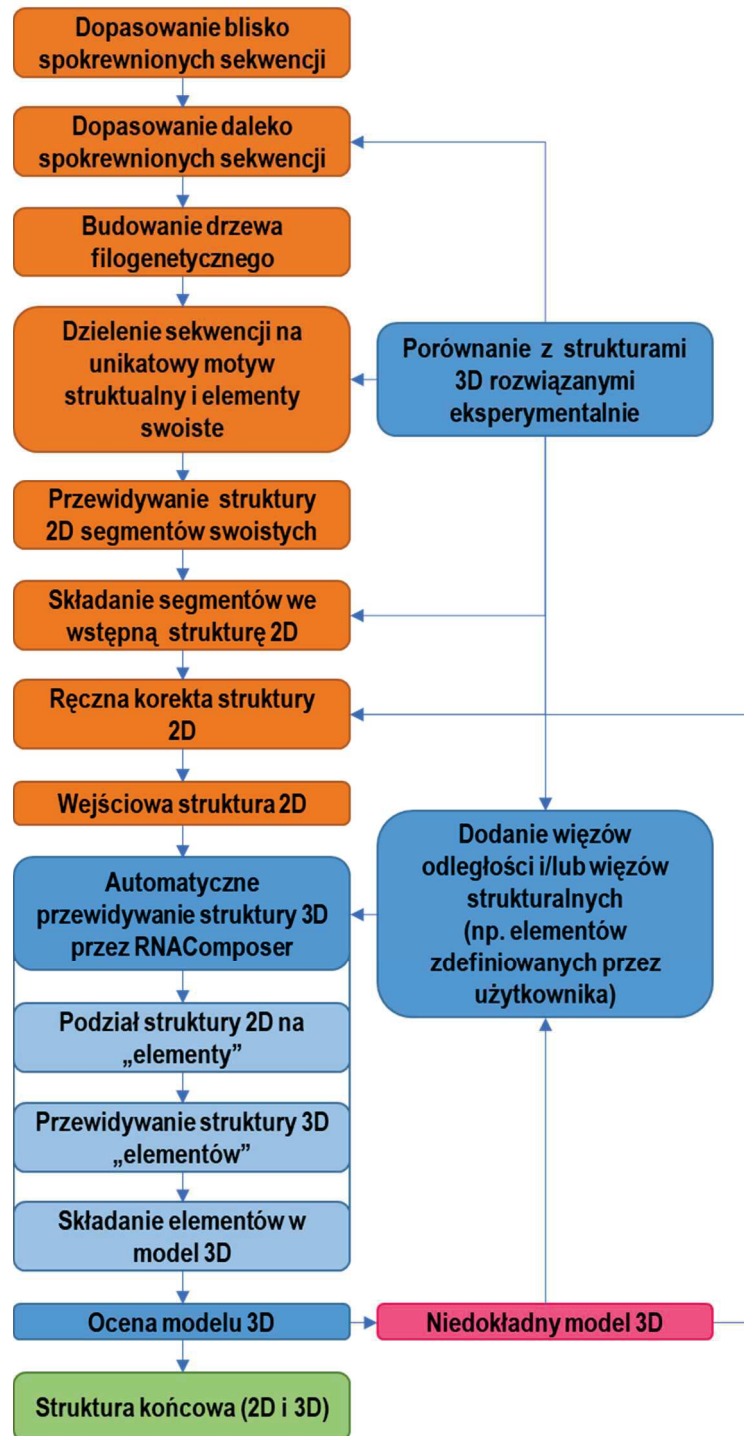
Opisana przez nas metoda przedstawia modelowanie struktury drugorzędowej i trzeciorzędowej ES7 z wykorzystaniem RNAComposer na podstawie sekwencji oraz w oparciu o model akrecyjny ewolucji rybosomalnego RNA [131]. Model ten opracowany został w oparciu o analizy porównawcze rRNA rybosomów pochodzących z różnych organizmów [132]. Model akrecyjny rRNA zakłada, że rejonu helikalne u organizmów bardziej rozwiniętych ulegają wydłużeniu, a w miejscach, gdzie występują pętle apikalne lub wewnętrzne może dochodzić do rozgałęzień. Istotnym faktem jest też, że nawet w przypadku modyfikacji sekwencji nie dochodzi do zmian w obrębie struktury przestrzennej. Oznacza to, że położenie rejonów helikalnych specyficznych dla danej struktury jest przestrzennie zakonserwowane niezależnie od organizmu lub etapu ewolucji cząsteczki RNA. Pozwoliło nam to podzielić ES7 na mniejsze części, których struktura drugorzędowa była przewidywana niezależnie. Podejście to jest koncepcyjnie podobne do wykorzystanego w systemie RNAComposer do przewidywania struktur 3D. Po uzyskaniu struktur 2D, wytypowaliśmy ES7 z różnych organizmów, każdy reprezentujący inną grupę, których struktury 3D chcieliśmy przewidzieć i opisać.

7.1.13. Metoda modelowania struktury 2D i 3D RNA

Zaproponowana przez nas metoda przewidywania struktury drugorzędowej i trzeciorzędowej RNA to metoda hybrydowa, która opiera się na unikalnym połączeniu sprawdzonych algorytmów do przewidywania struktury cząsteczek RNA z ustaleniami wynikającymi z zasad ewolucji rRNA (Rysunek 11).

Modelowanie struktury RNA ES7 dużej podjednostki rybosomalnej jest wieloetapowe. W pierwszym etapie porównuje się sekwencję przewidywanej struktury (MSA, ang. *Multiple sequence alignment*) [133] zarówno u blisko, jak i daleko spokrewnionych organizmów. Analiza MSA jest przeprowadzana z wykorzystaniem znanych algorytmów [134] i opcjonalnie korygowana manualnie. Porównanie wielu sekwencji RNA ES7, również takich, których struktura została eksperymentalnie rozwiązana, pozwala na identyfikację rejonu zwoju sygnaturowego, co z kolei prowadzi do wyznaczenia pozostałych swoistych rejonów ES7. Na podstawie analizy MSA budowane jest drzewo filogenetyczne, a dana sekwencja przypasowywana jest do grupy filogenetycznej (grupa 0 -3). Podział sekwencji na rejony swoiste wykorzystywany jest podczas przewidywania struktury drugorzędowej. Struktura 2D każdego z wyznaczonych rejonów jest przewidywana niezależnie od pozostałych, co skraca czas przewidywania przy jednoczesnym zwiększeniu dokładności. Struktury 2D RNA przewidywane są za pomocą mfold [134]. Już we wstępnej fazie przewidywania nakładane są ograniczenia na zwój sygnaturowy, aby był prawidłowo przewidywany. Uzyskana struktura 2D zwoju sygnaturowego łączona jest z przewidzianymi strukturami 2D pozostałych rejonów swoistych, budując wstępny model 2D całego ES7. Na tym etapie model 2D jest analizowany i jeżeli wprowadzenie zmian będzie zasadne, może zostać udokładniony ręcznie.

W kolejnym etapie struktura 3D rRNA przewidywana jest automatycznie poprzez platformę RNAComposer na podstawie sekwencji i struktury 2D. Przewidywanie struktury przestrzennej przebiega według schematu przedstawionego w punkcie 0. W przypadku braku elementu w bazie strukturalnej, element ten jest projektowany w oparciu o modelowanie porównawcze i wprowadzany jako zdefiniowany przez użytkownika fragment strukturalny. Otrzymany model struktury 3D podlega ocenie poprzez porównanie z dostępnymi strukturami 3D referencyjnego ES7 tej samej grupy filogenetycznej. W przypadku oceny modelu jako niedokładny analizowane są możliwe przyczyny niedokładności, w wyniku czego możliwa jest manualna modyfikacja struktury drugorzędowej lub dodanie ograniczeń na odległości lub strukturę. Proces ten jest przeprowadzany aż do momentu uzyskania pozytywnej oceny przewidzianej struktury przestrzennej w wyniku czego otrzymujemy końcowy model 2D i 3D RNA.



Rysunek 11. Schemat modelowania cząstek RNA w oparciu o znane struktury przestrzenne pokrewnych cząsteczek RNA.

7.1.14. Walidacja metody

W celu walidacji naszej metody przeprowadziliśmy dwa eksperymenty. Pierwszy eksperyment polegał porównaniu modelu ze strukturą 2D ES7 *Candida albicans* określoną doświadczalnie metodą SHAPE [117]. Eksperyment ten potwierdził zgodność struktury ES7

Candida albicans przewidzianej naszą metodą hybrydową z danymi otrzymanymi za pomocą metody SHAPE. W drugim eksperymencie na podstawie znanej struktury 3D ES7 *Eremothecium gossypii* przewidywaliśmy znaną strukturę ES7 drożdży *Saccharomyces cerevisiae*. W przewidywaniu struktury trzeciorzędowej dodatkowo wyłączyliśmy z bazy danych RNAComposer wszystkie struktury rybosomów drożdży. Otrzymany model porównaliśmy ze strukturami eksperymentalnymi (PDB: 4V6I oraz 5JUO, 4U4R) [135-137]. Poprawność przewidywania struktury drugorzędowej była bardzo wysoka, współczynnik korelacyjny Matthews'a [138] wynosił 83.8%. Otrzymany model 3D pokrywał się ze strukturą rzeczywistą ES7, odchylenie pomiędzy strukturami wyniosło 5.73 Å.

7.1.15. Przykłady przewidzianych modeli RNA ES7

W początkowej fazie badań wybraliśmy cztery sekwencje RNA ES7 dużej podjednostki rybosomalnej tak, aby każda ze struktur reprezentowała jedną grupę filogenetyczną (Rysunek 10).

Grupa 0

Do grupy 0 wytypowaliśmy sekwencję H25 pierwotniaka *Giardia muris*, której strukturę przewidywaliśmy w oparciu o motyw H25 bakterii *Escherichia coli* [139]. ES7 grupy 0 występuje w postaci spinki do włosów o długości kilkunastu reszt nukleotydowych. Przewidywanie struktury 3D tego elementu sprowadzało się do uzyskania wielu modeli na podstawie struktury drugorzędowej. Wykonałem przewidywanie 20 różnych modeli 3D, z których następnie został wybrany ten o najniższej wyznaczonej wartości energii struktury.

Grupa 1

Dla grupy 1 został wybrany ES7 bielnika białego (*Candida albicans*), którego strukturę przewidywaliśmy na podstawie ES7 drożdży *Saccharomyces cerevisiae* [136]. Cząsteczka ta posłużyła także do weryfikacji naszej metody. Struktura 2D RNA *Candida albicans* została wcześniej zaproponowana [117], jednakże nie była zgodna z naszym modelem zwoju sygnowanego (patrz rozdział 7.1.11). Struktury ES7 grupy 1 zawierają wyłącznie części zwoju sygnowanego. Wysoki stopień zakonserwowania zwoju sygnowanego przekłada się na wysokie podobieństwo sekwencji oraz struktury 2D i 3D RNA ES7 przypisanych do tej grupy. Na etapie modelowania struktury 2D rejony helikalne były łatwo identyfikowane w analizie MSA. Wszystkie fragmenty użyte do złożenia struktury trzeciorzędowej ES7 *Candida albicans* zostały pobrane przez RNAComposer z bazy danych RNA FRABASE [34], a większość wybranych fragmentów pochodziła ze struktur RNA różnych ES7. W przewidywaniu ES7 *Candida albicans* rola użytkownika została ograniczona do minimum.

Grupa 2

W grupie 2 wybraliśmy ES7 komara tygrysięgo (*Aedes albopictus*), którego struktura 2D była przewidywana na podstawie struktury ES7 muszki owocówki (*Drosophila melanogaster*) [140]. W przypadku tych dwóch struktur mamy do czynienia z przyrostem trzech rejonów helikalnych w stosunku do grupy 1, z czego jeden z dodatkowych rejonów rozgałęzia się poprzez pętlę trójramienną. Jest to zjawisko charakterystyczne dla tej grupy. Analiza pary ES7 komara i muszki owocówki była interesująca i stanowiła wyzwanie dla naszej metody ze względu na to, że sekwencje tych RNA znacznie się od siebie różnią. ES7 *Drosophila melanogaster* jest bogate w reszty AU, podczas gdy w ES7 *Aedes albopictus* dominują reszty GC. Komplikuje to znacząco porównanie sekwencji podczas analizy MSA oraz porównawcze modelowanie struktury drugorzędowej. Z drugiej strony różnica w zawartości reszt AU do GC w przewidywanym modelu nie powinna wpływać znacząco na przewidywanie struktury trzeciorzędowej. Chociaż RNAComposer promuje wybór fragmentów najbardziej kompatybilnych pod względem sekwencji, to w bazie RNA FRABASE [34] istnieje wiele fragmentów, które pomimo dużych różnic w sekwencji posiadają taką samą strukturę 2D i zbliżoną strukturę przestrzenną. Mimo tego, pierwsza pętla rozgałęziająca (pętla trójramienna), która znajduje się w rejonie zwoju sygnaturowego nie była prawidłowo przewidywana przez RNAComposer. W bazie struktur elementów, z której korzysta RNAComposer nie był dostępny element 3D odpowiedni dla tego fragmentu i był on generowany na bazie helisy A RNA. W strukturze ES7 *Aedes albopictus* pętla ta jest unikatowa, ponieważ jest większa niż w strukturach ES7 innych organizmów. Pętlę tą wygenerowałem oddzielnie na podstawie elementu ze struktury referencyjnej *Drosophila melanogaster* i wprowadziłem do obliczeń jako element zdefiniowany przez użytkownika.

Grupa 3

Do grupy 3 należą ES7 o dużych rozmiarach, bardzo złożone, których segmenty ekspansyjne są wielokrotnie rozgałęzione. W tej grupie na podstawie struktury ES7 człowieka *Homo sapiens* [140] przewidywaliśmy strukturę ES7 myszy *Mus musculus*. Ze względu na liczne rozgałęzienia struktury, odnalezienie właściwej topologii struktury w analizie MSA było bardzo istotne. Duże podobieństwo sekwencji poza zwojem sygnaturowym przewidywanej cząsteczki RNA ES7 *Mus musculus* do cząsteczki referencyjnej ES7 *Homo sapiens* ułatwiło modelowanie struktury drugorzędowej i trzeciorzędowej. Podczas przewidywania struktury przestrzennej nie odnaleziono w bazie RNA FRABASE [34] fragmentów referencyjnych struktur przestrzennych dwóch pętli rozgałęziających: drugiej pętli trójramiennej w obszarze zwoju sygnaturowego oraz pętli czteroramiennej w akrecyjnym rejonie przypominającym macki ośmiornicy. Fragmenty te przewidywałem tak jak w grupie 2 na podstawie elementów ze struktury referencyjnej, po czym wprowadziłem jako elementy strukturalne zdefiniowane przez użytkownika. Dodatkowo struktura ES7 myszy (690 reszt nukleotydowych) jest większa niż możliwości przewidywania struktur web-serwera RNAComposer (ograniczenie do 500 reszt nukleotydowych). Dlatego przed rozpoczęciem przewidywania 3D podzieliłem tą strukturę

na dwa odrębne fragmenty, tak aby posiadały wspólne, nakładające się elementy struktury. Struktury tych fragmentów przewidywałem za pomocą RNAComposer jako odrębne cząsteczki RNA. Po otrzymaniu odpowiednich struktur przestrzennych nakładałem je na siebie poprzez elementy wspólne, a powtarzające się reszty nukleotydowe usunąłem.

Generalnie długość łańcucha RNA ES7 jest zbieżna z poziomem ewolucji organizmów [132], jednak w przypadku kury *Gallus gallus* struktura ES7 odbiegała od przyjętego założenia, dlatego postanowiliśmy włączyć ten przypadek w zakres naszej pracy. W modelu akrecyjnym wzrost struktury RNA prowadzi do wydłużenia istniejących rejonów helikalnych lub powstawania nowych poprzez rozgałęzienia, dlatego struktury bardziej rozwiniętych organizmów zawierają rejony helikalne obecne w strukturach organizmów mniej rozwiniętych. W przypadku ES7 kury doszło do niebywałego wydłużenia jednego z peryferyjnych rejonów helikalnych, czego nie obserwuje się u innych organizmów bardziej rozwiniętych ewolucyjnie. Dla porównania, długość sekwencji ES7 kury jest większa niż długość sekwencji ES7 myszy. Przypadek ten był najtrudniejszym wyzwaniem podczas ustalania topologii struktury drugorzędowej.

7.1.16. Modelowanie kompleksu 28S oraz 5.8S rRNA dużej podjednostki rybosomalnej

W celu potwierdzenia użyteczności i uniwersalności naszej metody, która nie ogranicza się tylko do przewidywania struktur ES7, zastosowaliśmy ją do przedstawienia modelu całego kompleksu 28S oraz 5.8S rRNA dużej podjednostki rybosomalnej myszy (*Mus musculus*). Jako struktura referencyjna do predykcji 2D, posłużyła struktura rRNA dużej podjednostki rybosomalnej człowieka [140]. W strukturze przestrzennej dużej podjednostki rybosomalnej obecne są motywy strukturalne pseudowęzłów dalekiego zasięgu oraz liczne oddziaływania z białkami. Ze względu na ograniczenia naszej metody, która nie pozwala na przewidywanie struktury kompleksów RNA/białko, zdecydowaliśmy się nie nakładać ograniczeń na modelowanie struktury w oparciu o znane oddziaływanie z białkami. Ponadto, szczegóły tych oddziaływań jak również oddziaływań typu pseudowęzłów różnią się znacząco pomiędzy rRNA różnych organizmów. Ze względu na to również struktury pseudowęzłów nie były brane pod uwagę podczas przewidywania pomimo, że RNAComposer ma możliwość wprowadzenia informacji o pseudowęzłach do obliczeń. Analizując otrzymany model zaobserwowałem że rejony nici tworzących pseudowęzła są w bliskiej odległości od siebie (około 7 Å). W wyniku przewidywania struktury przestrzennej otrzymałem końcowy model kompleksu 28S oraz 5.8S rRNA dużej podjednostki myszy, który był strukturalnie bardzo podobny do struktury 3D rRNA dużej podjednostki człowieka (różnica RMSD: 4.020 Å). Algorytmy obliczeniowe na których bazujemy nie biorą pod uwagę oddziaływań z białkami; sekwencja, struktura drugorzędowa i 3D różni się pomiędzy organizmami; struktura RNA jest dynamiczna; wszystkie te czynniki są zapewne przyczyną obserwowanych odchyłań pomiędzy otrzymanym modelem, a strukturą eksperymentalną rRNA człowieka. Jednak otrzymany rezultat – model całego rRNA o bardzo zbliżonej strukturze do struktury referencyjnej potwierdza możliwości naszej metody.

7.1.17. Wnioski

Nasza hybrydowa metoda pozwala na modelowanie struktur 2D i 3D RNA nawet w przypadkach, które stanowią ograniczenia dla metod konwencjonalnych. Ograniczenia te wynikają z czynników takich jak konformacyjna heterogeniczność i dynamika struktury, duże rozmiary cząsteczek, zmienność sekwencji, wysoka zawartość reszt nukleotydowych guanozyny i cytozyny oraz niska kowariancja niektórych rejonów. Czynniki kluczowymi dla sukcesu naszej metody były: (i) identyfikacja zwoju sygnaturowego oraz (ii) zdolność programu RNAComposer do przewidywania struktur w przypadku, gdy fragmenty o wysokiej homologii sekwencyjnej są niedostępne. Ponieważ nasze podejście pozwala przewidzieć z dużą dokładnością największe i najbardziej kompleksowe segmenty ekspansyjne, jak i całe cząsteczki rRNA może być użyte do przewidywania struktur dowolnych segmentów ekspansyjnych rRNA różnych gatunków a także innych funkcjonalnych cząsteczek RNA lub ich swoistych części.

Wykorzystując analizę struktur otrzymanych eksperymentalnie oraz modeli RNA, które przewidywaliśmy naszą metodą hybrydową doszliśmy do wniosku, że zwój sygnaturowy jest wynikiem poważnych ograniczeń ewolucyjnych i ewolucji zbieżnej, co sprawia że nie zmienia się podczas wzrostu segmentu ekspansyjnego. Dane te potwierdzają również model ewolucyjny na poziomie molekularnym [132] w którym poprzez serię wielu powtarzających się zdarzeń dochodzi do wzrostu i rozgałęziania się struktury segmentu ekspansyjnego, nie zaburzając przy tym istniejącej struktury rybosomu. Pojawiające się w toku ewolucji nowe rejony helikalne pozostają zachowane w strukturze przestrzennej rRNA gatunków wyższego rzędu nawet pomimo znacznych różnic w sekwencji. Interesującym wyjątkiem od tej reguły jest struktura ES7 kury, w której obserwujemy rejon helikalny o nie obserwowanej u jakiegokolwiek innego gatunku wydłużonej strukturze spinki.

Szczegółowa analiza strukturalna ewolucji rRNA doprowadziła do wyprowadzenia modelu przebiegu reakcji strukturalnej, czyli koordynatów reakcji [141] dla rozgałęzień rRNA. Z zebranych danych wynika, że wzrost rozgałęzienia jest wieloetapowym procesem. Aby doszło do powstania rozgałęzienia, czyli do utworzenia nowego regionu helikalnego, musi w początkowym etapie dojść do powstania wyrzuszenia, które powiększa się i może przejść w pętlę zewnętrzną. Gdy pętla będzie już dostatecznie duża zacznie dochodzić w jej obrębie do parowań w jej obrębie i utworzenia nowego regionu helikalnego.

8. PODSUMOWANIE

Ze względu na ograniczenia eksperymentalnych metod przestrzennego określania struktury cząsteczek RNA, konieczne staje się wykorzystanie do tego celu zaawansowanych metod obliczeniowych. Przeprowadzone przeze mnie analizy wykazały, że program RNAComposer jest on dobrym narzędziem, o dużym potencjale w badaniach nad strukturą przestrzenną RNA. Jego niewątpliwą zaletą jest to, że przewiduje strukturę 3D dużych cząsteczek RNA w ujęciu pełnoatomowym, a samo przewidywane jest szybkie i dokładne. Baza danych RNA FRABASE, na której opiera się RNAComposer jest duża, bazuje na rzeczywistych cząsteczkach i ciągle się powiększa dostarczając nowych informacji. Baza danych RNA FRABASE była kluczowym składnikiem wykorzystanym przeze mnie przy projektowaniu nowych funkcjonalnych nanocząstek RNA, a w szczególności ich struktury drugorzędowej poprzez dobór odpowiednich elementów strukturalnych. Z kolei sam program RNAComposer umożliwił mi uzyskanie dużej ilości modeli przestrzennych, z których poprzez analizę geometrii rejonów helikalnych mogłem wybrać te, które mają największy potencjał do tworzenia zaprojektowanej struktury.

Przewidywanie struktur ES7 rRNA było niejako modelowaniem porównawczym, gdyż wyniki moich eksperymentów modelowania struktury 3D mogłem porównać do struktury referencyjnej o znanej strukturze przestrzennej. Mocną stroną programu RNAComposer jest przewidywanie z dobrą dokładnością fragmentów, które istotnie różnią się od struktury referencyjnej. Równocześnie na podstawie uzyskanych modeli struktury trzeciorzędowej weryfikowane były modele struktury drugorzędowej, co sprawiało, że przewidywanie było jeszcze bardziej wiarygodne.

Przewidywanie struktur przestrzennych RNA jest istotne w badaniu cząsteczek RNA, w których funkcja jest ściśle powiązana ze strukturą. Dzięki otrzymanym modelom, ale również dzięki analizie procesu modelowania zostały bez wątplenia zrealizowane wszystkie istotne założenia mojej rozprawy doktorskiej. Najważniejszymi osiągnięciami były:

- przedstawienie zależności między poprawnością przewidywanego modelu i jakością wprowadzanych danych wejściowych,
- opracowanie modelu zarządzanie funkcją cząsteczek RNA w oparciu o zaprogramowany model struktury,
- pokazanie wykorzystania programów do przewidywania struktury 2D do walidacji modeli 3D zaprojektowanych przez RNAComposer i wykorzystanie RNAComposera do walidacji modeli struktur drugorzędowych,
- ustalenie cech charakterystycznych dla ES7 rRNA dużej podjednostki rybosomalnej (zwój sygnaturowy),
- ustalenie/pokazanie gradacji ważności poziomów uporządkowania struktury rRNA i ich wpływ na funkcje cząsteczek rRNA oraz na ewolucję na poziomie molekularnym , a także wyprowadzenie modelu przebiegu ekspansji cząsteczek RNA

Przedstawione przeze mnie metody modelowania struktur przestrzennych RNA nie ograniczają się tylko do przedstawionych przykładów, mają charakter bardziej ogólny. Metoda modelowania nanocząstek RNA, może zostać wykorzystana przy projektowaniu

innych cząstek RNA o zaprogramowanych funkcjach. Z kolei metoda przewidywania struktur przestrzennych RNA zmiennych ewolucyjnie, może posłużyć do badania innych segmentów ekspansyjnych rRNA lub innych cząsteczek RNA jak np. ryboprzełączniki.

Dodatkowym efektem zasługującym na uwagę jest mój wkład w rozwój systemu RNAComposer. Przeprowadzone przeze mnie modelowania struktur przestrzennych RNA doprowadziły do usprawnienia systemu i wprowadzenia nowych funkcjonalności jak możliwości wprowadzania własnych elementów strukturalnych czy selekcja elementów z bazy danych pod względem rozdzielczości lub identyfikatora RCSB PDB.

9. WYKAZ STOSOWANYCH SKRÓTÓW I NOMENKLATURA

2D – dwuwymiarowy (struktura drugorzędowa cząsteczki RNA)

3D – trójwymiarowy (struktura trzeciorzędowa cząsteczki RNA)

A-RNA – RNA w postaci prawoskrętnej helisy typu A

ab initio – w odniesieniu do modelowania struktur przedstrzennych RNA – modelowanie wykorzystujące sekwencje jako danę wejściową, przewidywanie odbywa się na podstawie praw fizyki

AIDS – Zespół nabytego niedoboru odporności, choroba wirusowa wywołana przez ludzki wirus niedoboru odporności

COVID-19 – choroba zakaźna wywołana przez wirusa SARS-CoV2

de novo – w odniesieniu do modelowania struktur przedstrzennych RNA – modelowanie wykorzystujące sekwencje jako danę wejściową, metoda oparta o prawidłowości geometryczne

DNA – kwas deoksyrybonukleinowy

ES7 – siódmy segment ekspansyjny, fragment rRNA

H25 – helisa 25, fragment rRNA

in situ – produkt reakcji chemicznej (np. rodniki) generowany w miejscu zachodzenia reakcji i ulegający natychmiast dalszym przekształceniom

ndb – ang. *Nucleic acids data base*, portal zawierający dane struktur przestrzennych kwasów nukleinowych

NMR – ang. *Nuclear Magnetic Resonance*, spektroskopia magnetycznego rezonansu jądrowego

miRNA – ang. *micro RNA*, niewielkie, około 22 nukleotydowe dupлексы RNA, inhibujące ekspresję serii mRNA zawierające homologiczne sekwencje

MOHCA – ang. *Multiplexed hydroxyl radical Cleavage Analysis*, analiza wielopunktowego rozerwania wiązań rodnikami hydroksylowymi, jedna z metod eksperymentalnego próbkowania struktury drugorzędowej RNA

mRNA - matrycowy RNA, którego rolą jest przenoszenie informacji genetycznej między DNA a białkiem

MSA – ang. *multiple sequence alignment*, analiza porównywania wielu sekwencji (RCSB) PDB – ang. (*Research Collaboratory for Structural Bioinformatics*) *Protein Data Bank*, baza danych struktur przestrzennych dużych cząstek biologicznych zdeterminowanych eksperymentalnie

RMSD – ang. *root-mean-square deviation*, odchylenie średniej kwadratowej – miara porównawcza średniej odległości między przewidywanym modelem, a strukturą referencyjną

RNA – kwas rybonukleinowy

rRNA – rybosomalne RNA

SARS-CoV2 – ang. *severe acute respiratory syndrome coronavirus 2*, koronawirus typu 2 ciężkiego ostrego zespołu oddechowego, wirus wywołujący ostrą chorobę układu oddechowego COVID-19

SHAPE – ang. *Selective 2'-hydroxyl acylation analyzed by primer extension*, reakcja selektywnej acylacji reszty cukrowej w pozycji 2'-OH analizowana poprzez reakcję odwrotnej transkrypcji, typ reakcji eksperymentalnego próbkowania struktury drugorzędowej RNA

siRNA – ang. *small interfering RNA*, krótki interferujący kwas rybonukleionowy, dwuniciowa cząsteczka RNA o długości od kilkunastu do kilkudziesięciu reszt nukleotydowych powodująca wyciszenie ekspresji specyficznego mRNA o homologicznej sekwencji

współczynnik korelacyjny Matthews – współczynnik porównujący struktury drugorzędowe dwóch cząsteczek RNA. Współczynnik wyraża się w zakresie od -1 (lub -100%) – struktura drugorzędowa całkowicie niezgodna do 1 (lub 100%) - struktura drugorzędowa całkowicie zgodna

10. LITERATURA

1. Szczepaniak, M. and J. Moc, *Conformational studies of gas-phase ribose and 2-deoxyribose by density functional, second order PT and multi-level method calculations: the pyranoses, furanoses, and open-chain structures*. Carbohydr Res, 2014. **384**: p. 20-36.
2. Shabarova, Z. and A. Bogdanov, *Advanced organic chemistry of nucleic acids*. 1994, Weinheim ; New York: VCH. xv, 588.
3. Sychrovsky, V., et al., *Revisiting the planarity of nucleic acid bases: Pyramidalization at glycosidic nitrogen in purine bases is modulated by orientation of glycosidic torsion*. Nucleic Acids Res, 2009. **37**(21): p. 7321-31.
4. Devereaux, Z.J., et al., *Structures and Relative Glycosidic Bond Stabilities of Protonated 2'-Fluoro-Substituted Purine Nucleosides*. J Am Soc Mass Spectrom, 2019. **30**(8): p. 1521-1536.
5. Burley, S.K., et al., *RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences*. Nucleic Acids Res, 2021. **49**(D1): p. D437-D451.
6. Zhou, A.Q., *Exploration of the central dogma at the interface of chemistry and biology: 2010 Yale Chemical Biology Symposium*. Yale J Biol Med, 2010. **83**(3): p. 131-3.
7. Hughes, S.H., *Reverse Transcription of Retroviruses and LTR Retrotransposons*. Microbiol Spectr, 2015. **3**(2): p. MDNA3-0027-2014.
8. Schneider, B., Z. Moravek, and H.M. Berman, *RNA conformational classes*. Nucleic Acids Res, 2004. **32**(5): p. 1666-77.
9. Zwieb, C., *The principles of RNA structure architecture*. Methods Mol Biol, 2014. **1097**: p. 33-43.
10. Gan, H.H., S. Pasquali, and T. Schlick, *Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design*. Nucleic Acids Res, 2003. **31**(11): p. 2926-43.
11. Peselis, A. and A. Serganov, *Structure and function of pseudoknots involved in gene expression control*. Wiley Interdiscip Rev RNA, 2014. **5**(6): p. 803-22.
12. Hofacker, I.L.F., W.; Stadler P.F.; Bonhoeffer L.S.; Tacker, M.; Schuster P., *Fast folding and comparison of RNA secondary structures*. Monatshefte für Chemie/Chemical Monthly, 1994. **125**(2): p. 167-188.
13. Mattei, E., et al., *A novel approach to represent and compare RNA secondary structures*. Nucleic Acids Res, 2014. **42**(10): p. 6146-57.
14. Antczak, M., et al., *New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation*. Bioinformatics, 2018. **34**(8): p. 1304-1312.
15. Popenda, M., et al., *Automated 3D structure composition for large RNAs*. Nucleic Acids Res, 2012. **40**(14): p. e112.
16. Wang, J., et al., *3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction*. Int J Mol Sci, 2019. **20**(17).
17. Darty, K., A. Denise, and Y. Ponty, *VARNA: Interactive drawing and editing of the RNA secondary structure*. Bioinformatics, 2009. **25**(15): p. 1974-5.
18. Byun, Y. and K. Han, *PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures*. Nucleic Acids Res, 2006. **34**(Web Server issue): p. W416-22.

19. Lee, J.C. and R.R. Gutell, *Helix capping in RNA structure*. PLoS One, 2014. **9**(4): p. e93664.
20. Pan, F., C. Roland, and C. Sagui, *Ion distributions around left- and right-handed DNA and RNA duplexes: a comparative study*. Nucleic Acids Res, 2014. **42**(22): p. 13981-96.
21. Popena, M., J. Milecki, and R.W. Adamiak, *High salt solution structure of a left-handed RNA double helix*. Nucleic Acids Res, 2004. **32**(13): p. 4044-54.
22. Montemayor, E.J., et al., *Structure of an RNA helix with pyrimidine mismatches and cross-strand stacking*. Acta Crystallogr F Struct Biol Commun, 2019. **75**(Pt 10): p. 652-656.
23. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs*. RNA, 2001. **7**(4): p. 499-512.
24. Zirbel, C.L., et al., *Classification and energetics of the base-phosphate interactions in RNA*. Nucleic Acids Res, 2009. **37**(15): p. 4898-918.
25. Cate, J.H., et al., *RNA tertiary structure mediation by adenosine platforms*. Science, 1996. **273**(5282): p. 1696-9.
26. Cate, J.H., et al., *Crystal structure of a group I ribozyme domain: principles of RNA packing*. Science, 1996. **273**(5282): p. 1678-85.
27. Agarwala, P., S. Pandey, and S. Maiti, *The tale of RNA G-quadruplex*. Org Biomol Chem, 2015. **13**(20): p. 5570-85.
28. Abu Almakarem, A.S., et al., *Comprehensive survey and geometric classification of base triples in RNA structures*. Nucleic Acids Res, 2012. **40**(4): p. 1407-23.
29. Coimbatore Narayanan, B., et al., *The Nucleic Acid Database: new features and capabilities*. Nucleic Acids Res, 2014. **42**(Database issue): p. D114-22.
30. Burge, S.W., et al., *Rfam 11.0: 10 years of RNA families*. Nucleic Acids Res, 2013. **41**(Database issue): p. D226-32.
31. Ponce-Salvatierra, A., et al., *Computational modeling of RNA 3D structure based on experimental data*. Biosci Rep, 2019. **39**(2).
32. Rother, M., et al., *ModeRNA: a tool for comparative modeling of RNA 3D structure*. Nucleic Acids Res, 2011. **39**(10): p. 4007-22.
33. Flores, S.C., et al., *Predicting RNA structure by multiple template homology modeling*. Pac Symp Biocomput, 2010: p. 216-27.
34. Popena, M., et al., *RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures*. BMC Bioinformatics, 2010. **11**: p. 231.
35. Zhao, Y., et al., *Automated and fast building of three-dimensional RNA structures*. Sci Rep, 2012. **2**: p. 734.
36. Xu, X., C. Zhao, and S.J. Chen, *VfoldLA: A web server for loop assembly-based prediction of putative 3D RNA structures*. J Struct Biol, 2019. **207**(3): p. 235-240.
37. Das, R. and D. Baker, *Automated de novo prediction of native-like RNA tertiary structures*. Proc Natl Acad Sci U S A, 2007. **104**(37): p. 14664-9.
38. Li, B., et al., *Advances in RNA 3D Structure Modeling Using Experimental Data*. Front Genet, 2020. **11**: p. 574485.
39. Zhang, D., J. Li, and S.J. Chen, *IsRNA1: De Novo Prediction and Blind Screening of RNA 3D Structures*. J Chem Theory Comput, 2021. **17**(3): p. 1842-1857.
40. Jonikas, M.A., et al., *Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters*. RNA, 2009. **15**(2): p. 189-99.

41. Das, R., J. Karanicolas, and D. Baker, *Atomic accuracy in predicting and designing noncanonical RNA structure*. *Nat Methods*, 2010. **7**(4): p. 291-4.
42. Sulc, P., et al., *A nucleotide-level coarse-grained model of RNA*. *J Chem Phys*, 2014. **140**(23): p. 235102.
43. Boniecki, M.J., et al., *SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction*. *Nucleic Acids Res*, 2016. **44**(7): p. e63.
44. Fang, X., et al., *An unusual topological structure of the HIV-1 Rev response element*. *Cell*, 2013. **155**(3): p. 594-605.
45. Pasquali, S. and P. Derreumaux, *HiRE-RNA: a high resolution coarse-grained energy model for RNA*. *J Phys Chem B*, 2010. **114**(37): p. 11957-66.
46. Liu, Z., et al., *Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate*. *Cell*, 2018. **173**(6): p. 1549-1550.
47. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution*. *Nat Protoc*, 2006. **1**(3): p. 1610-6.
48. Ding, F., et al., *Three-dimensional RNA structure refinement by hydroxyl radical probing*. *Nat Methods*, 2012. **9**(6): p. 603-8.
49. Cheng, C.Y., et al., *Consistent global structures of complex RNA states through multidimensional chemical mapping*. *Elife*, 2015. **4**: p. e07600.
50. Blaszczyk, L., et al., *Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression*. *Viruses*, 2017. **9**(4).
51. Yu, H., Y. Qi, and Y. Ding, *Deep Learning in RNA Structure Studies*. *Front Mol Biosci*, 2022. **9**: p. 869601.
52. Townshend, R.J.L., et al., *Geometric deep learning of RNA structure*. *Science*, 2021. **373**(6558): p. 1047-1051.
53. Shen, T., et al., *E2Efold-3D: End-to-End Deep Learning Method for accurate de novo RNA 3D Structure Prediction*. *arxiv*, 2022. **arXiv:2207.01586**.
54. Mathews, D.H., *RNA Secondary Structure Analysis Using RNAstructure*. *Curr Protoc Bioinformatics*, 2014. **46**: p. 12 6 1-25.
55. Do, C.B., D.A. Woods, and S. Batzoglou, *CONTRAFold: RNA secondary structure prediction without physics-based models*. *Bioinformatics*, 2006. **22**(14): p. e90-8.
56. Gruber, A.R., et al., *The Vienna RNA websuite*. *Nucleic Acids Res*, 2008. **36**(Web Server issue): p. W70-4.
57. Sato, K., et al., *CENTROIDFOLD: a web server for RNA secondary structure prediction*. *Nucleic Acids Res*, 2009. **37**(Web Server issue): p. W277-80.
58. Zakov, S., et al., *Rich parameterization improves RNA structure prediction*. *J Comput Biol*, 2011. **18**(11): p. 1525-42.
59. Sato, K., et al., *IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming*. *Bioinformatics*, 2011. **27**(13): p. i85-93.
60. Berman, H.M., et al., *The Protein Data Bank*. *Nucleic Acids Res*, 2000. **28**(1): p. 235-42.
61. Guntert, P., C. Mumenthaler, and K. Wuthrich, *Torsion angle dynamics for NMR structure calculation with the new program DYANA*. *J Mol Biol*, 1997. **273**(1): p. 283-98.
62. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package*. *J Magn Reson*, 2003. **160**(1): p. 65-73.

63. Korczmar, E.A.B., A.; N askręt-Barciszewska, M.Z.; Jurga, S.; Barciszewski, J.; , *100 lat RNA. Diamentowy jubileusz informacyjnego RNA*. Postępy Biochemii, 2021. **67**(3): p. 212-222.
64. Wu, L. and J.G. Belasco, *Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs*. Mol Cell, 2008. **29**(1): p. 1-7.
65. Wilson, T.J. and D.M. Lilley, *RNA catalysis--is that it?* RNA, 2015. **21**(4): p. 534-7.
66. Menza, T.W., et al., *Rapid Uptake of Home-Based HIV Self-testing During Social Distancing for SARS-CoV2 Infection in Oregon*. AIDS Behav, 2021. **25**(1): p. 167-170.
67. Noda, T., et al., *Characterization of the Ebola virus nucleoprotein-RNA complex*. J Gen Virol, 2010. **91**(Pt 6): p. 1478-83.
68. Antoine, L., et al., *RNA Modifications in Pathogenic Bacteria: Impact on Host Adaptation and Virulence*. Genes (Basel), 2021. **12**(8).
69. Murphy, E.R. and S.M. Payne, *RyhB, an iron-responsive small RNA molecule, regulates Shigella dysenteriae virulence*. Infect Immun, 2007. **75**(7): p. 3470-7.
70. Winkelsas, A.M., et al., *Targeting the 5' untranslated region of SMN2 as a therapeutic strategy for spinal muscular atrophy*. Mol Ther Nucleic Acids, 2021. **23**: p. 731-742.
71. Rossi, J.J. and D.J. Rossi, *siRNA Drugs: Here to Stay*. Mol Ther, 2021. **29**(2): p. 431-432.
72. Wu, Q., et al., *Evaluation of the safety profile of COVID-19 vaccines: a rapid review*. BMC Med, 2021. **19**(1): p. 173.
73. Manfredonia, I., et al., *Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements*. Nucleic Acids Res, 2020. **48**(22): p. 12436-12452.
74. Sreeramulu, S., et al., *Exploring the Druggability of Conserved RNA Regulatory Elements in the SARS-CoV-2 Genome*. Angew Chem Int Ed Engl, 2021. **60**(35): p. 19191-19200.
75. Dibrov, S.M., et al., *Hepatitis C virus translation inhibitors targeting the internal ribosomal entry site*. J Med Chem, 2014. **57**(5): p. 1694-707.
76. Naceri, S., et al., *Druggable Pockets at the RNA Interface Region of Influenza A Virus NS1 Protein Are Conserved across Sequence Variants from Distinct Subtypes*. Biomolecules, 2022. **13**(1).
77. Pyle, A.M., *Rediscovering RNA*. RNA, 2015. **21**(4): p. 714-5.
78. Cox, R.A. and U.Z. Littauer, *Secondary structure of ribonucleic acid in solution*. Nature, 1959. **184**(Suppl 11): p. 818-9.
79. Vicens, Q. and J.S. Kieft, *Thoughts on how to think (and talk) about RNA structure*. Proc Natl Acad Sci U S A, 2022. **119**(17): p. e2112677119.
80. Westhof, E. and V. Fritsch, *RNA folding: beyond Watson-Crick pairs*. Structure, 2000. **8**(3): p. R55-65.
81. Zok, T., et al., *RNApdbe 2.0: multifunctional tool for RNA structure annotation*. Nucleic Acids Res, 2018. **46**(W1): p. W30-W35.
82. Bernier, C.R., et al., *RiboVision suite for visualization and analysis of ribosomes*. Faraday Discuss, 2014. **169**: p. 195-207.
83. Wu, Y., et al., *Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data*. Nucleic Acids Res, 2015. **43**(15): p. 7247-59.
84. Spasic, A., et al., *Modeling RNA secondary structure folding ensembles using SHAPE mapping data*. Nucleic Acids Res, 2018. **46**(1): p. 314-323.

85. Smola, M.J., et al., *Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis*. Nat Protoc, 2015. **10**(11): p. 1643-69.
86. Costa, M. and D. Monachello, *Probing RNA folding by hydroxyl radical footprinting*. Methods Mol Biol, 2014. **1086**: p. 119-42.
87. Ganser, L.R., et al., *The roles of structural dynamics in the cellular functions of RNAs*. Nat Rev Mol Cell Biol, 2019. **20**(8): p. 474-489.
88. Al-Hashimi, H.M. and N.G. Walter, *RNA dynamics: it is about time*. Curr Opin Struct Biol, 2008. **18**(3): p. 321-9.
89. Shi, X., et al., *Probing the dynamics of the P1 helix within the Tetrahymena group I intron*. J Am Chem Soc, 2009. **131**(27): p. 9571-8.
90. Herraiez, A., *Biomolecules in the computer: Jmol to the rescue*. Biochem Mol Biol Educ, 2006. **34**(4): p. 255-61.
91. Serganov, A., L. Huang, and D.J. Patel, *Structural insights into amino acid binding and gene control by a lysine riboswitch*. Nature, 2008. **455**(7217): p. 1263-7.
92. Xayaphoummine, A., T. Bucher, and H. Isambert, *Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W605-10.
93. Grundy, F.J., S.C. Lehman, and T.M. Henkin, *The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes*. Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12057-62.
94. Antczak, M., et al., *New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure*. Acta Biochim Pol, 2016. **63**(4): p. 737-744.
95. Lilley, D.M., *Mechanisms of RNA catalysis*. Philos Trans R Soc Lond B Biol Sci, 2011. **366**(1580): p. 2910-7.
96. Green, R. and H.F. Noller, *Ribosomes and translation*. Annu Rev Biochem, 1997. **66**: p. 679-716.
97. Dinger, M.E., T.R. Mercer, and J.S. Mattick, *RNAs as extracellular signaling molecules*. J Mol Endocrinol, 2008. **40**(4): p. 151-9.
98. Afonin, K.A., et al., *Multifunctional RNA nanoparticles*. Nano Lett, 2014. **14**(10): p. 5662-71.
99. Chau, T.H.T., et al., *Developments of Riboswitches and Toehold Switches for Molecular Detection-Biosensing and Molecular Diagnostics*. Int J Mol Sci, 2020. **21**(9).
100. Chappell, J., M.K. Takahashi, and J.B. Lucks, *Creating small transcription activating RNAs*. Nat Chem Biol, 2015. **11**(3): p. 214-20.
101. Roncati, L. and L. Corsi, *Nucleoside-modified messenger RNA COVID-19 vaccine platform*. J Med Virol, 2021. **93**(7): p. 4054-4057.
102. D'Esposito, R.J., et al., *Challenges with Simulating Modified RNA: Insights into Role and Reciprocity of Experimental and Computational Approaches*. Genes (Basel), 2022. **13**(3).
103. Liu, D., et al., *Structures of artificially designed discrete RNA nanoarchitectures at near-atomic resolution*. Sci Adv, 2021. **7**(39): p. eabf4459.
104. Shu, Y., et al., *Stable RNA nanoparticles as potential new generation drugs for cancer therapy*. Adv Drug Deliv Rev, 2014. **66**: p. 74-89.
105. Guo, S., et al., *Ultra-thermostable RNA nanoparticles for solubilizing and high-yield loading of paclitaxel for breast cancer therapy*. Nat Commun, 2020. **11**(1): p. 972.

106. Shu, Y., B. Seremi, and P. Guo, *Fabrication Methods for RNA Nanoparticle Assembly Based on Bacteriophage phi29 Packaging RNA Structural Features*, in *RNA Nanotechnology and Therapeutics*. 2013, CRC Press. p. 285-302.
107. Simpson, A.A., et al., *Structure of the bacteriophage phi29 DNA packaging motor*. *Nature*, 2000. **408**(6813): p. 745-50.
108. Reuter, J.S. and D.H. Mathews, *RNAstructure: software for RNA secondary structure prediction and analysis*. *BMC Bioinformatics*, 2010. **11**: p. 129.
109. Bailey, S., et al., *Phylogenetic analysis and secondary structure of the Bacillus subtilis bacteriophage RNA required for DNA packaging*. *J Biol Chem*, 1990. **265**(36): p. 22365-70.
110. Harris, S. and S.J. Schroeder, *Nuclear magnetic resonance structure of the prohead RNA E-loop hairpin*. *Biochemistry*, 2010. **49**(29): p. 5989-97.
111. Ding, F., et al., *Structure and assembly of the essential RNA ring component of a viral DNA packaging motor*. *Proc Natl Acad Sci U S A*, 2011. **108**(18): p. 7357-62.
112. Zhang, H., et al., *Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA*. *RNA*, 2013. **19**(9): p. 1226-37.
113. Dibrov, S.M., et al., *Self-assembling RNA square*. *Proc Natl Acad Sci U S A*, 2011. **108**(16): p. 6405-8.
114. Jasinski, D., et al., *Advancement of the Emerging Field of RNA Nanotechnology*. *ACS Nano*, 2017. **11**(2): p. 1142-1164.
115. Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation*. *Protein Sci*, 2018. **27**(1): p. 293-315.
116. Lehman, N., *RNA in evolution*. *Wiley Interdiscip Rev RNA*, 2010. **1**(2): p. 202-13.
117. Gomez Ramos, L.M., et al., *Eukaryotic Ribosomal Expansion Segments as Antimicrobial Targets*. *Biochemistry*, 2017. **56**(40): p. 5288-5299.
118. Miao, Z., et al., *RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme*. *RNA*, 2017. **23**(5): p. 655-672.
119. Fox, G.E., *Origin and evolution of the ribosome*. *Cold Spring Harb Perspect Biol*, 2010. **2**(9): p. a003483.
120. Shankar, V., et al., *rRNA expansion segment 27Lb modulates the factor recruitment capacity of the yeast ribosome and shapes the proteome*. *Nucleic Acids Res*, 2020. **48**(6): p. 3244-3256.
121. Fujii, K., et al., *Decoding the Function of Expansion Segments in Ribosomes*. *Mol Cell*, 2018. **72**(6): p. 1013-1020 e6.
122. Ramesh, M. and J.L. Woolford, Jr., *Eukaryote-specific rRNA expansion segments function in ribosome biogenesis*. *RNA*, 2016. **22**(8): p. 1153-62.
123. Ben-Shem, A., et al., *The structure of the eukaryotic ribosome at 3.0 Å resolution*. *Science*, 2011. **334**(6062): p. 1524-9.
124. Nilsson, J., et al., *Comparison of fungal 80 S ribosomes by cryo-EM reveals diversity in structure and conformation of rRNA expansion segments*. *J Mol Biol*, 2007. **369**(2): p. 429-38.
125. Greber, B.J., et al., *Cryo-EM structures of Arx1 and maturation factors Rei1 and Jjj1 bound to the 60S ribosomal subunit*. *Nat Struct Mol Biol*, 2012. **19**(12): p. 1228-33.
126. Wu, S., et al., *Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes*. *Nature*, 2016. **534**(7605): p. 133-7.

127. Kater, L., et al., *Visualizing the Assembly Pathway of Nucleolar Pre-60S Ribosomes*. Cell, 2017. **171**(7): p. 1599-1610 e14.
128. Bradatsch, B., et al., *Structure of the pre-60S ribosomal subunit with nuclear export factor Arx1 bound at the exit tunnel*. Nat Struct Mol Biol, 2012. **19**(12): p. 1234-41.
129. Smethurst, D.G.J., et al., *Iron-mediated degradation of ribosomes under oxidative stress is attenuated by manganese*. J Biol Chem, 2020. **295**(50): p. 17200-17214.
130. Kossinova, O., et al., *The SBP2 protein central to selenoprotein synthesis contacts the human ribosome at expansion segment 7L of the 28S rRNA*. RNA, 2014. **20**(7): p. 1046-56.
131. Petrov, A.S., et al., *History of the ribosome and the origin of translation*. Proc Natl Acad Sci U S A, 2015. **112**(50): p. 15396-401.
132. Petrov, A.S., et al., *Evolution of the ribosome at atomic resolution*. Proc Natl Acad Sci U S A, 2014. **111**(28): p. 10251-6.
133. Notredame, C., *Recent evolutions of multiple sequence alignment algorithms*. PLoS Comput Biol, 2007. **3**(8): p. e123.
134. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Res, 2003. **31**(13): p. 3406-15.
135. Armache, J.P., et al., *Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution*. Proc Natl Acad Sci U S A, 2010. **107**(46): p. 19748-53.
136. Garreau de Loubresse, N., et al., *Structural basis for the inhibition of the eukaryotic ribosome*. Nature, 2014. **513**(7519): p. 517-22.
137. Abeyrathne, P.D., et al., *Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome*. Elife, 2016. **5**.
138. Matthews, B.W., *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*. Biochim Biophys Acta, 1975. **405**(2): p. 442-51.
139. Dunkle, J.A., et al., *Structures of the bacterial ribosome in classical and hybrid states of tRNA binding*. Science, 2011. **332**(6032): p. 981-4.
140. Anger, A.M., et al., *Structures of the human and Drosophila 80S ribosome*. Nature, 2013. **497**(7447): p. 80-5.
141. Buerger, H.B., *Chemical reaction coordinates from crystal structure data*. Inorganic Chemistry, 1973. **12**(10): p. 2321–2325.

11. ZAŁĄCZNIKI**ZAŁĄCZNIK NR 1. Życiorys naukowy**

Wykształcenie i doświadczenie zawodowe:

Od 01.09.2018	Laboratorium Fizykochemii i Analityki, DPL; Laboratorium Badawcze SORBOLAB; Poznań (<i>Specjalista chemik analityk</i>)
2016-2018	Zakład Struktury i Funkcji Retrotranspozonów; Instytut Chemii Bioorganicznej, Polska Akademia Nauk; Poznań (<i>studia III stopnia, stacjonarne</i>)
2014-2016	Zespół Struktury i Funkcji Retroelementów; Instytut Chemii Bioorganicznej, Polska Akademia Nauk; Poznań (<i>studia III stopnia, stacjonarne</i>)
01.09.2013-8.04.2016	Europejskie Centrum Bioinformatyki i Genomiki; Instytut Informatyki; Politechnika Poznańska, Poznań (<i>Asystent naukowy</i>)
2013-2014	Zespół Struktury i Funkcji RNA; Instytut Chemii Bioorganicznej, Polska Akademia Nauk; Poznań (<i>studia III stopnia, stacjonarne</i>)
2011-2013	Wydział Chemii; Uniwersytet im. Adama Mickiewicza; Poznań (<i>studia II stopnia, stacjonarne</i>)
2008-2011	Wydział Chemii; Uniwersytet im. Adama Mickiewicza; Poznań (<i>studia I stopnia, stacjonarne</i>)
2005-2008	I LO im. Marii Skłodowskiej-Curie; Piła (<i>szkoła ponadgimnazjalna</i>)

Staż:

05-06.2017	The Center for Ribosomal Origins and Evolution; School of Chemistry & Biochemistry; Georgia Institute of Technology, Atlanta, USA
09.2012-02.2013	Laboratory of Inorganic Chemistry; Department of Chemistry; University of Ioannina; Ioannina, Grecja

Szkolenia:

03.2023	Technika LC-MS/MS we współczesnej analizie próbek biologicznych i żywności
01.2023	Optymalizacja qPCR oraz RT-qPCR
11.2022	Chromatografia gazowa – omówienie techniki, dobór kolumny chromatograficznej, wskazówki praktyczne
11.2022	Western Blot – praktyczne aspekty związane z elektroforezą próbek
11.2022	Wykorzystanie chromatografii cienkowarstwowej do szybkiej analizy jakościowej
10.2022	Sekrety techniki Elisa
05.2022	Filtracja membranowa - rozwiązania EZ-Fit Family dedykowane do badania wody i napojów
04.2022	Chromatografia cieczowa: przygotowanie układu pomiarowego
04.2022	Ekstrakcja do fazy stałej w praktyce

07.2021	Badania pozostałości w środkach ochrony roślin w żywności i środowisku. Praktyczne wykorzystanie technik GC-MS/MS i LC-MS/MS
06.2021	Duolink PLA – detekcja i lokalizacja oddziaływań lub modyfikacji białkowych w komórce
05.2021	Western blotting – wskazówki i porady dotyczące optymalizacji procesu
05.2021	Zastosowanie metody ultrafiltracji do odsalania, wymiany buforu, zatężania oraz separacji białka
03.2021	Dobór właściwych membran i rozwiązań w różnych aplikacjach filtracji analitycznej i przygotowania prób
12.2020	Spektroskopia ICP-MS: eleganckie narzędzie w analizie wielopierwiastkowej
10.2020	Badania pozostałości w środkach ochrony roślin w żywności i środowisku. Praktyczne wykorzystanie technik GC-MS/MS i LC-MS/MS
01.2019	Obsługa chromatografu ciekłego Shimadzu z elementami projektowania metod HPLC zgodnie z QbD; Poznań
05.2016	Zaawansowane użytkowanie systemu GeXP połączone z warsztatami z sekwencjonowania oraz analizy fragmentów; Darmstadt; Niemcy
05.2013	Auditor wewnętrzny systemów zarządzania w laboratorium badawczym według PN-EN ISO/IEC 17025:2005; Poznań
05.2013	Zasady dobrej praktyki produkcyjnej – GMP; Poznań

Wyróżnienia:

2016-2018	Specjalne stypendium naukowe dla studentów studiów III stopnia finansowane z dotacji podmiotowej w ramach Krajowego Naukowego Ośrodka Wiodącego (KNOW) w projekcie “Poznańskie Konsorcjum RNA”; Poznań
2016-2017	Stypendium dla najlepszych doktorantów ICHB PAN, Poznań
2013	Wyróżnienie pracy magisterskiej przyznane przez Radę Naukową Wydziału Chemii Uniwersytetu im. Adama Mickiewicza; Poznań
2011-2013	Stypendium motywacyjne dla studentów studiów II stopnia współfinansowane ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego w projekcie “Chemia Warta Poznania”; Poznań

Projekty badawcze:

2016	Opracowanie nowej metody hybrydowej określania struktury drugorzędowej RNA; realizacja zadania badawczego służącego rozwojowi młodych naukowców lub uczestników studiów doktoranckich w roku 2016, działalność statutowa ICHB finansowana przez MNiSW (<i>kierownik</i>);
2013-2016	Automatyczne, wysokoprzepustowe modelowanie struktur przestrzennych RNA; projekt MAESTRO finansowany przez Narodowe Centrum Nauki [2012/06/A/ST6/00384] (<i>wykonawca</i>)

Referaty:

- 2017 RNAComposer. Recent advances and applications.
RNA Run; Obrzycko
- 2016 RNAComposer and RNA 3D structure prediction for nanotechnology.
RNA Rush; Obrzycko
- 2016 Projektowanie nowych nanocząsteczek RNA o zdefiniowanych funkcjach.
III Poznańskie Sympozjum Młodych Naukowców; Poznań
- 2016 The specificity of GagΔp6, NC and MA interactions with 5'-UTR of HIV-2.
3rd Innovative Approaches for Novel Antiviral Agents Summer School;
Pula; Włochy
- 2016 RNA 3D structure prediction for RNA nanodesign
Kongres BIO; Wrocław
- 2014 Rola danych eksperymentalnych w przewidywaniu struktur 3D RNA.
I Poznańskie Sympozjum Młodych Naukowców. Nowe Oblicze Nauk
Przyrodniczych; Poznań

Postery:

- 2015 Próbkowanie chemiczne i przewidywanie struktur przestrzennych RNA.
M. Biesiada, K. Pachulska-Wieczorek, R.W. Adamiak, K.J. Purzycka
I. Wielkopolskie Sympozjum Chemii Bioorganicznej, Organicznej i
Biomateriałów, BioOrg 2015; Poznań
- 2015 Insight into the biological functions of retroelement's RNA through the
predictions and analyses of their RNA 3D structures.
M. Biesiada, K. Pachulska-Wieczorek, L. Błaszczyk, J. Gumna, R.W.
Adamiak, K.J. Purzycka
RNA Society Meeting 2015; Madison; WI; USA
- 2014 Nucleic acid chaperone activity of retroviral proteins.
K. Pachulska-Wieczorek, A.K. Stefaniak, K.J. Purzycka, L. Błaszczyk, M.
Biesiada, R.W. Adamiak
Kongres BIO; Warszawa
- 2014 The importance of the experimental data in the 3D RNA structure
predictions.
M. Biesiada, K. Pachulska-Wieczorek, L. Błaszczyk, R.W. Adamiak, K.J.
Purzycka
XXI Round Table on Nucleosides, Nucleotides and Nucleic Acids.
Chemical Biology of Nucleic Acids; Poznań
- 2013 Badanie wpływu warunków reakcji 4-fenylimidazolo-2-tiolu z jodem na
otrzymany produkt.
M. Biesiada, A.M. Owczarzak, M. Kubicki
55 Konwersatorium Krystalograficzne, VI Sesja Naukowa i Warsztaty
PTK; Wrocław

Publikacje:

Biesiada M, Hu M, Williams LD, Purzycka KJ, Petrov AS. *rRNA Expansion in Eukaryotes: from the Signature Fold to Tentacles*. **Nucleic Acids Research**. **2022**; 50(18): p. 10717-10732

Gómez Ramos LM, Degtyareva N, Kovacs NA, Holguin SY, Jiang L, Biesiada M, Purzycka KJ, Arya DY, Williams LD. *Eukaryotic Ribosomal Expansion Segments as Antimicrobial Targets*. **Biochemistry**. **2017**; 56(40): p. 5288-5299

Błaszczak L, Biesiada M, Saha A, Garfinkel DJ, Purzycka KJ. *Structure of Ty1 internally initiated RNA influences restriction factor expression*. **Viruses**. **2017**; 9(4). pii: E74.

Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, Boniecki MJ, Bujnicki J, Chen SJ, Cheng CY, Chou FC, Ferré-D'Amaré AR, Das R, Dawson WK, Feng D, Dokholyan NV, Dunin-Horkawicz S, Geniesse C, Kappel K, Kladwang W, Krokhotin A, Łach GE, Major F, Mann TH, Magnus M, Pachulska-Wieczorek K, Patel DJ, Piccirilli JA, Popena M, Purzycka KJ, Ren A, Rice GM, Santalucia J Jr, Sarzynska J, Szachniuk M, Tandon A, Trausch JJ, Tian S, Wang J, Weeks KM, Williams B 2nd, Xiao Y, Xu X, Zhang D, Zok T, Westhof E. *RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme*. **RNA**. **2017**; 23(5): p. 655-672.

Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW. *Automated RNA 3D structure prediction with RNAComposer*. **Methods Mol Biol**. **2016**; 1490:199-215.

Biesiada M; Pachulska-Wieczorek K; Adamiak RW; Purzycka KJ. *RNAComposer and RNA 3D structure prediction for nanotechnology*. **Methods**. **2016**; 103:120-7.

Pachulska-Wieczorek K, Błaszczak L, Biesiada M, Adamiak RW, Purzycka KJ. *The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag*. **Retrovirology**. **2016**; 13:18.

Pachulska-Wieczorek K, Błaszczak L, Gumna J, Nishida Y, Saha A, Biesiada M, Garfinkel D, Purzycka KJ. *Characterizing the functions of Ty1 Gag and the Gag-derived restriction factor p22/p18*. **Mob Genet Elements**. **2016**; 6(2):e1154637.

Biesiada M, Kourkoumelis N, Kubicki M, Owczarzak AM, Balas V, Hadjikakou SK. *Fundamental chemistry of iodine. The reaction of di-iodine towards thiourea and its methyl-derivative: formation of aminothiazoles and aminothiadiazoles through dicationic disulfides*. **Dalton Trans**. **2014**; 43(12):4790-806.

ZAŁĄCZNIK NR 2. Prace naukowe wchodzące w skład rozprawy doktorskiej

Automated RNA 3D Structure Prediction with RNAComposer

Marcin Biesiada*, Katarzyna J. Purzycka*, Marta Szachniuk, Jacek Blazewicz, and Ryszard W. Adamiak

Abstract

RNAs adopt specific structures to perform their activities and these are critical to virtually all RNA-mediated processes. Because of difficulties in experimentally assessing structures of large RNAs using NMR, X-ray crystallography, or cryo-microscopy, there is currently great demand for new high-resolution 3D structure prediction methods. Recently we reported on RNAComposer, a knowledge-based method for the fully automated RNA 3D structure prediction from a user-defined secondary structure. RNAComposer method is especially suited for structural biology users. Since our initial report in 2012, both servers, freely available at <http://rnacomposer.ibch.poznan.pl> and <http://rnacomposer.cs.put.poznan.pl> have been often visited. Therefore this chapter provides guidance for using RNAComposer and discusses points that should be considered when predicting 3D RNA structure. An application example presents current scope and limitations of RNAComposer.

Key words RNA tertiary structure, RNA three-dimensional structure, RNA modeling

1 Introduction

RNAs adopt specific structures to perform their activities, beginning with transcription and ending with turnover. RNA structure is critical to virtually all RNA-mediated processes ranging from splicing to viral replication in eukaryotes. Studying RNA structure helps understanding how it guides RNA function. The ability of RNA strands to fold back on themselves and to form stable tertiary architectures is fundamental to RNA function. In many cases, the RNA tertiary structure is crucial for recognition by cellular factors. RNA structure is influenced by primary sequence, cellular environment, trans-acting factors, or ion homeostasis. These factors

*These authors contributed equally to this work.

Dedication: This work is dedicated to Professor David Shugar, one of the pioneers in the field of molecular biophysics, on the occasion of his 100th birthday anniversary.

contribute to the difficulty of obtaining RNA 3D structures using X-ray crystallography or NMR.

Recent advancements in RNA secondary structure probing, including new reagents or structure mapping integrated with systematic mutagenesis, allowed to determine secondary structures of large RNAs with reasonable accuracy [1, 2]. The bottleneck is an RNA 3D structure determination. One possibility to overcome this limitation is RNA 3D structure prediction. Computational methods for RNA tertiary structure prediction are based on simulation of RNA folding (DMD [3], NAST [4]), comparative modeling (ModeRNA [5]) or fragment assembly (FARNA [6], MC-Fold/MC-Sym [7]). Most of them are time-consuming and require high level of expertise to accomplish prediction of 3D model. Few methods are automated (Vfold [8], iFoldRNA [9], 3dRNA [10]) and among them is our method called RNAComposer [11]. All available approaches have substantial limitations, but the field of computational methods for RNA tertiary structure prediction is currently coming into bloom.

1.1 RNA 3D Structure Prediction with RNAComposer

The RCSB PDB [12] database contains over 2500 spatial structures of RNA, and this number still increases. Every spatial structure has corresponding secondary structure. Every RNA secondary structure contains single and double stranded regions, and these can be classified as loops, stems, etc. If we divide RNA secondary structures into specific elements, like stems or loops, many of these elements will be identical or similar. The idea of RNAComposer [11] is based on the observation that many of these elements repeat. Therefore unknown 3D structures can be predicted and built based on the known elements derived from the solved 3D structures. RNAComposer predicts 3D structures of RNA molecules based on their sequence and secondary structure topology. It uses a dedicated database (a dictionary of structure elements) that contains 3D RNA fragments derived from RNA FRABASE [13, 14]. The dictionary relates RNA secondary and tertiary structure elements. In the first step, secondary structure provided by the user is divided into fragments according to its graph representation [15]. Next, RNAComposer algorithm searches through the dictionary of elements to find best matching 3D structure fragments with the RNA secondary structure of interest. Initial 3D model is built by assembling selected 3D structural elements. Next, it is refined to the final 3D structure by minimization in torsion angle space and in Cartesian atom coordinate space using CHARMM force field (both steps use incorporated X-Plor suite [16]).

RNAComposer is based on the concept of machine translation. It is a knowledge-based method that employs fully automated fragment assembly based on the user-specified secondary RNA structure. Structure prediction is very fast and accomplished using web-servers: <http://rnacomposer.ibch.poznan.pl> and mirror <http://rnacomposer.cs.put.poznan.pl>.

Two procedures are implemented in RNAComposer web-server to obtain RNA 3D structures. The interactive mode, for fast inspection of RNA molecule of interest is dedicated to all visitors, while the batch mode allows modeling larger RNA structures in large-scale. Batch mode is available after registration.

2 Materials

RNAComposer should be run via an Internet web browser. It has been implemented in the client–server model and requires from the user (i.e., the client side) to have an Internet access and a web browser installed locally. The system works with most of the available browsers. Windows users can execute it in Microsoft Internet Explorer (from version 8.0), Mozilla Firefox (from version 3.6), Opera (from version 10.53), or Google Chrome (from version 5.0), whereas for those working under Linux or Mac OS, Mozilla Firefox (from version 3.6) and Opera (from version 10.53) are suggested. In every case, the latest versions of web browsers are strongly recommended. In the interactive mode, RNAComposer provides a possibility to display the output 3D model. The visualization is performed by the incorporated Jmol [17] applet that runs on a web browser's Java Virtual Machine. Therefore, an installation of Java (www.java.com), preferably its latest version, is required to execute this option.

3 Methods

3.1 *RNAComposer Website at a Glance*

When opening the RNAComposer website, one can see Homepage with interactive mode activated, vertical menu panel displayed on the left side, and bottom bar with quick links (Fig. 1).

Menu panel contains Main Menu, login area, visitor counter, and links to supporting institutions. Main menu allows to navigate between the system pages: “Home”, “Tools”, “Help”, “About”, “References”, “Links”, and “Contact us”. “Home” is the starting page. Its main content, in both modes, is the task entry box, where the input sequence and secondary structure topology should be typed in. Three examples are available at hand for quick upload and processing. Clicking on “Compose” button runs the process of 3D structure modeling. “Tools” page provides support for users having their secondary structures encoded in CT or BPSEQ format. Two converters from these notations to dot–bracket are available there. “Help” gives a detailed description of the system, including usage scenarios, input and output data formats with examples, and user account. RNAComposer authors, developers, supporting team, and funders are acknowledged on the “About” page. “References” enumerate RNAComposer-related publications for



Fig. 1 RNAComposer interactive mode homepage (<http://rnacomposer.cs.put.poznan.pl>). (1) Main menu. (2) Examples. (3) Reset button. (4) Entry box with input example. (5) Compose button. (6) E-mail address box

citation. On the “Links” page we have collected links to selected tools for RNA structure storage, analysis and processing. Finally, “Contact us” provides the message form to be used in case of questions, problems, or suggestions concerning RNAComposer. Messages typed here are e-mailed to the RNAComposer team members.

Login area differs between modes. In the interactive mode it allows to register within the system (“Create an account” option), reset forgotten password (“Forgot your password?” option), and enter the batch mode by logging into the system. In the batch mode it provides links to access user workspace and account settings page.

Quick links located on the bottom bar direct the user to “Terms and conditions”, “System requirements”, and “Report problems” pages. The first page collects information about policy of using RNAComposer system. Software requirements that should be met to make the use of RNAComposer possible are reported on “System requirements” page. Finally, the “Report problems” link directs to the contact page, similarly as “Contact us” option in the Main Menu.

3.2 Getting Started: How to Prepare the Input Data

RNAComposer requires sequence and defined secondary structure topology as an input. RNA length is limited to 500 nucleotides. Input consists of three clearly defined lines. The first line serves to identify the molecule while following two lines are critical for the three-dimensional structural model building. The first line begins with the right angle bracket “>” and identifies the RNA strand of interest. After “>” only basic letters “A–Z”, “a–z”, numbers 0–9, underscore “_” and colon “:” signs are accepted. The next line

contains RNA sequence annotated in four-letter ribonucleotide code: “A, C, G, U”. Modified bases cannot be marked. The third line represents RNA secondary structure topology encoded in dot–bracket notation. This notation represents secondary structure of RNA in a simple way, but few principles must be followed to prepare correct input. In this notation, dots are assigned to unpaired nucleotides or noncanonical pairing. Parentheses are assigned to canonical-paired nucleotides so that left “(” represents first nucleotide from the base-pair (closer to the RNA 5′-end) and the right “)” is attributed to the nucleotide closing this pair (Fig. 2a). If the secondary structure is more complicated, e.g., contains structural elements of higher order, it is necessary to use other brackets for the structure representation. The square brackets “[]”, braces “{ }”, or angle brackets “< >” are allowed. For example first order pseudoknot within RNA structure should be annotated using square brackets.

Lines 2 and 3 must have equal number of characters and each dot or bracket sign from the third line should correspond with the letter identifying the base and located above this sign. Only canonical A-U, C-G and a wobble G-U base pairs are permitted in the input data. Any additional comments can be added in any new line starting with hash “#” sign. An example of the correct input data (structure of RNA function modulator, PDB id: 4K27 [18]) is shown on Fig. 2a. Its secondary structure is visualized on panel b.

Three examples are available in the web-server interactive mode to display correct input data. First example is a model hairpin, the second shows RNA structure with the pseudoknot and the third one exemplifies the use of the secondary structure prediction program (described in detail in Subheading 3.3). Button “Reset” above the entry box allows to clear the entry box contents.

In the batch mode the user can provide a batch of up to ten RNA sequences with 1–10 secondary structure topologies for

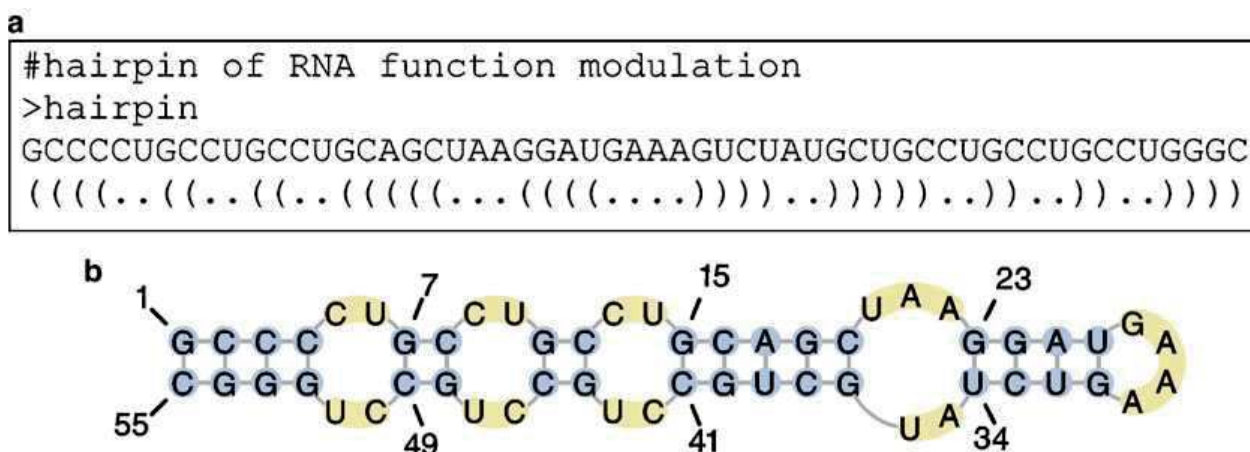


Fig. 2 Example of correct input data. (a) Secondary structure of functional modulator hairpin in dot–bracket notation and (b) its PseudoViewer visualization

every sequence. Secondary structure topology is represented in dot–bracket notation. Each sequence must be introduced with a new line containing sequence identifier following the angle bracket. Every additional dot–bracket notation of the secondary structure must be written in separated line directly under the sequence line.

To present correct input data in the batch mode, three examples are provided on the RNAComposer website. The first example is a 274 nts RNA containing pseudoknot, based on the crystal structure of tetrahymena ribozyme (PDB id 1X8W). The second example is a 120 nts RNA, based on 5S rRNA *E. coli* mutant (12C>12A; PDB id 2AWB). For this RNA, three different secondary structures are predefined. Example 3 is oriented towards high-throughput approach available in RNAComposer and exemplifies input of multiple sequences of pre-miRNA with several defined secondary structures for each sequence.

If the above described rules are not obeyed, the computation is stopped and the error messages are displayed under the entry box. Non-critical errors (displayed in yellow) like presence of non-canonical base pairs in secondary structure topology or lowercase letter in the sequence can be automatically corrected or ignored. If ignored, the RNAComposer treats noncanonical base pair as the canonical one and the lowercase letter is changed to the capital. Critical errors (written in red) obstruct launching of the computation and this type of error must be repaired.

3.3 Structure Modeling in the Interactive Mode

Interactive mode (Fig. 1), available without registration is dedicated mostly to the inexperienced users who are interested in the fast inspection of simple structures or want to test RNAComposer performance. Interactive mode is useful for smaller and not complex RNAs that are readily predicted *in silico* (*see* **Notes 1** and **2**).

Before launching structure modeling, correct input data are required (*see* Subheading 3.2). Secondary structure topology of RNA can be entered by the user in dot–bracket notation. However, in this mode, the secondary structure can be also predicted making use of one of three programs: RNAstructure [19], CONTRAfold [20], or RNAfold [21] that have been incorporated into RNAComposer system. For this purpose the name of selected program should be provided in the third line instead of dot–bracket representation (example 3 in the RNAComposer interactive mode).

Pushing “Compose” button results in opening the “Task progress information” page (Fig. 3). In this page, the user obtains output information containing: task identifier and input data, information about the task processing phases released in real-time with partial and completed computing times, and two links. First link allows to download “pdb” file with coordinates of the predicted 3D model (Fig. 4). The second link allows for instant 3D visualization of the model using Jmol [17] applet. Additionally, upon user request, the pdb and log files may be sent to the supplied e-mail

Task progress information

1 Task identifier: #b6dd0d-6df3-4523-b02c-a7538ff6f67e
 Task description: hairpin
 Sequence: GCGCCGCGCCGCGAGCGAAGGAGUAAAGGCGAGCGCGCCGCGCCGCGGCG
 Secondary structure: ((((((((.....(((.....)))))).....)))))).....))))))

2

2014-12-19 13:50:56.849 Task launched.
 2014-12-19 13:56:57.144 STEP 1: Input data validation.
 2014-12-19 13:56:57.209 Secondary structure provided by the user.
 2014-12-19 13:56:57.209 Input data validation completed (0:0:0:132).
 2014-12-19 13:57:09.470 STEP 2: Secondary structure fragmentation.
 2014-12-19 13:57:09.496 Secondary structure fragmentation completed (0:0:0:17).
 2014-12-19 13:57:09.528 STEP 3: 3D structure elements preparation.
 2014-12-19 13:57:09.571 3D structure elements preparation completed (0:0:11:718).
 2014-12-19 13:57:09.579 3D structure elements ready for composition.
 2014-12-19 13:57:09.780 STEP 4: 3D structure elements rigid body transformation.
 2014-12-19 13:57:09.786 3D structure elements rigid body transformation completed (0:0:0:0).
 2014-12-19 13:57:09.802 MODEL 1
 2014-12-19 13:57:09.812 STEP 5: 3D structure elements merging.
 2014-12-19 13:57:09.821 3D structure elements merging completed (0:0:0:179).
 2014-12-19 13:57:09.829 STEP 6: Initial 3D structure minimization in torsion angle space.
 2014-12-19 13:57:09.836 Initial 3D structure minimization in torsion angle space completed (0:0:0:57).
 2014-12-19 13:57:09.846 STEP 7: Composed 3D structure minimization in atomic Cartesian coordinates space.
 2014-12-19 13:57:09.850 Composed 3D structure minimization in atomic Cartesian coordinates space completed (0:0:10:418).
 2014-12-19 13:57:09.862 Composed 3D structure refinement completed(0:0:33:493).
 2014-12-19 13:57:09.871 MODEL 1 successfully composed.
 2014-12-19 13:57:09.880 MODEL 1 composition time: 0:0:11:04.
 2014-12-19 13:57:09.888 Task completed.

3 [hairpin.pdb](#) [3D visualization](#)

Fig. 3 The task progress information page (<http://rnacomposer.cs.put.poznan.pl>). (1) Task identifier and input data. (2) Information about task processing phases. (3) Links to output data: pdb file and Jmol visualization

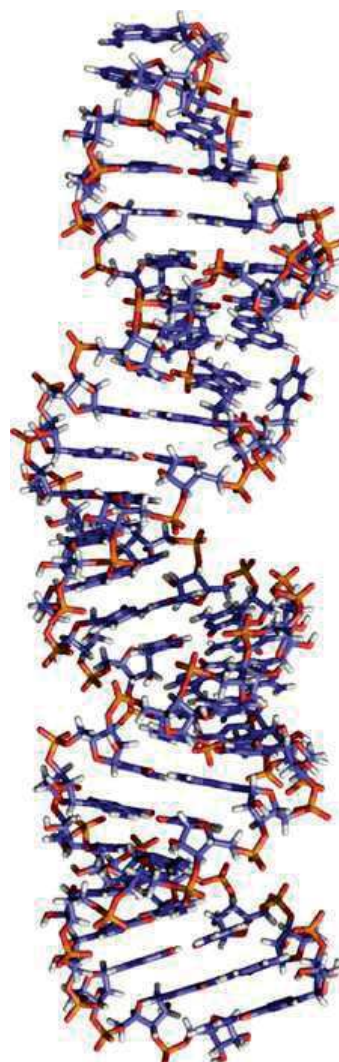


Fig. 4 3D representation of hairpin model

address. This is possible when the appropriate option is set in the user workspace.

One query in RNAComposer's interactive mode allows to build only one 3D RNA model. It is usually accomplished within seconds to minutes, depending on the RNA molecule size and complexity [11].

3.4 Creating User Account

For the purpose of using the RNAComposer batch mode, a user should register in the system by creating a personal account. Upon selecting "Create an account" option, located in the menu panel login area (Fig. 1), an account creation page is launched. The user is required to provide an e-mail address, unique username (user ID), and password (all of them are case sensitive). After quick data validation, an e-mail with activation link is sent to the user. Clicking onto the link ends the registration. Immediately after that, the new account is available and ready to use. Logged user can change the account settings, like e-mail address, password, and notification about completed batches. Three options are defined for e-mail notifications: no notifications, notifications with or without prediction results attached.

3.5 Structure Modeling in the Batch Mode

Batch mode (Fig. 5) is designed for RNA 3D structure prediction in large scale and allows to use additional RNAComposer functionality:

- Single input with multiple sequences defined and multiple secondary structures for each sequence
- Prediction of multiple models for one secondary structure
- Addition of distance restraints
- User workspace for short term storage of obtained data
- Detailed information about prediction steps and results provided in the log file (*see Note 3*)

Process of RNA structure prediction in the batch mode is similar to the interactive mode. However, in this mode the user can provide a batch of up to ten sequences with 1–10 secondary structure topologies for every sequence. No limit has been defined for a number of batches that can be run by the user. All batches are queued in the system and served according to the queuing protocol. Similarly to the interactive mode, the secondary structure topology is presented in dot-bracket notation (*see Notes 1 and 2*). In the single batch prediction, the user can launch as many as 100 tasks (one task consists of one sequence and secondary structure) that would have to be introduced separately in the interactive mode. Additionally, for each of these tasks RNAComposer can compute up to ten structural models, resulting in 1000 RNA 3D structural models that can be generated at a time, upon a single batch. The user can choose the required number of

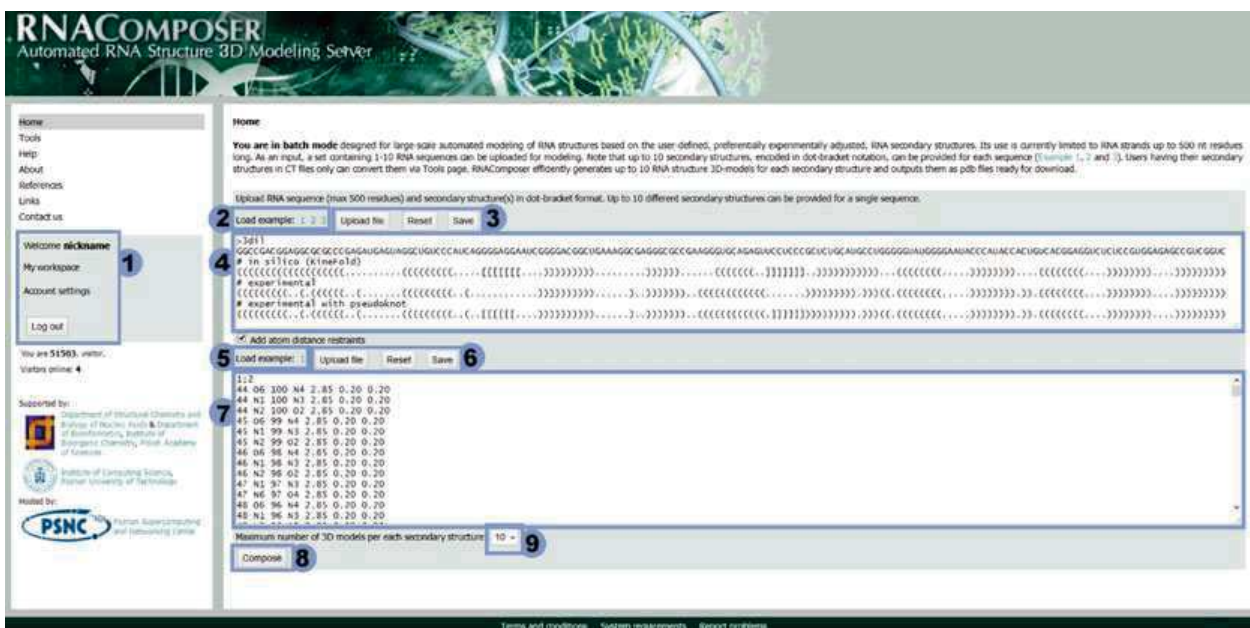


Fig. 5 RNAComposer batch mode homepage (<http://rnacomposer.ibch.poznan.pl>). (1) Batch mode's login area. (2) Examples. (3, 5, 6) Action buttons. (4) Entry box with input example of lysine riboswitch. (7) Example of correct atom distance restraints. (8) Compose button. (9) Drop-down list of maximum number of 3D models

structures by selecting option “maximum number of 3D models per each secondary structure” (Fig. 5). The predicted models are ranked based on the following criteria:

- Secondary structure topology compliance
- Similarity of the input and output sequences
- Purine–pyrimidine compatibility
- Source structure resolution
- Energy

The first output model is the one with best values of all the above criteria. In case of other models, the resolution and energy criteria are omitted.

3.5.1 Atom Distance Restraints

In addition to the sequence and secondary structure input, atom distance restraints can be optionally provided in the batch mode. Information about interatomic distances in RNA can be obtained experimentally or from analysis of particular databases, e.g., RNA FRABASE [13, 14]. Long range contacts or information about base-pairs can be revealed using UV or chemically induced cross-linking [22], nuclear magnetic resonance (NMR) [23], fluorescence resonance energy transfer (FRET) [24], or electron paramagnetic resonance (EPR) [25] studies. Such information can greatly improve 3D structure prediction. Atom distance restraints can be entered into additional entry box (Fig. 5). Format of atom distance restraints for RNAComposer is strictly defined and divided into two parts. The first part is a headline that defines the sequence

number and the number of the secondary structure topology for which restraints will be applied. Numbers (from 1 to 10) referring to the sequences should be separated from those referring to the topologies by semicolons (example 3 in the RNAComposer batch mode). If the restraints are applied to more than one sequence (or secondary structure topology), then the sequences (topologies) should be separated by commas. As shown in Fig. 5, headline: 1;2 indicates that restraints will be applied to the first sequence and second dot–bracket notation. Every new description of interatomic distances must be specified under new headline.

In the second part, the user provides restraints as rigid atom distances. One line corresponds to the single restraint and consists of seven fields separated by spaces. The following fields represent:

- Serial number of first residue
- Name of the atom from the first residue
- Serial number of second residue
- Name of the atom from the second residue
- Interatomic distance value [\AA]
- Allowable lower deviation from distance value [\AA]
- Allowable upper deviation from distance value [\AA]

Number of nucleotide residue and name of the atom can be described by at most four characters. Numbers of nucleotide residues correspond to nucleotides in chain from 5' to 3' end. Names of the atoms consist of the atom symbol and its number ascribed in accordance with IUPAC rules. Values defining distances must be described in %5.2f format, which means that the total number of digits is five, and number of decimal points is 2. No limit is set to the number of distance restraints that can be defined for single headline. Input data in specified format can be introduced into the entry boxes by pasting, writing or uploading text file. Additionally, “Save” button allows writing input data on hard disk (Fig. 5).

3.5.2 RNA 3D Structure Prediction Results

When the computation of query finishes, the user can download the results. Results are provided in pdb and log files. Pdb file contains atom coordinates of RNA 3D structural model(s). Log is a text file with data and parameters describing obtained model(s) and all steps of the 3D structure prediction. It contains:

- Input data
- Processing phases with partial and completed computing times
- Secondary structure elements resulting from the fragmentation of the input secondary structure
- 3D structure elements selected for the model building with their PDB IDs and similarity (%) to the structure provided by the user (*see* **Notes 4** and **5**)

Table 1
Approximate energy values of the final 3D structure expected for different RNA strand lengths

RNA length (nt)	Energy (kcal/mol)
30–100	–300 to –2000
100–200	–2000 to –4000
200–300	–4000 to –6000
300–400	–6000 to –8000
400–500	–8000 to –10,000

- 3D structure energy after every 100 steps of minimization in the Cartesian coordinates space
- Final energy of the 3D structure

Included information allows evaluation of model accuracy. The log file should be thoroughly analyzed by the user and several points should be taken under consideration. A good indicator of the 3D structural model quality is the final structure energy. Expected energy values were estimated in the original report [11] and are summarized in Table 1.

Log file contains also information about structural elements that due to the absence in the RNAComposer dictionary (although containing nearly half million of 3D structure elements) were generated de novo by the system. It should be noted that 3D structures containing such elements are usually of lower accuracy. Additional experimental data should be provided in such cases to help refine the structure and make the prediction more reliable (*see Note 6*). RNAComposer does not require homologs of RNA in question to predict the 3D structure. However, structure prediction accuracy tends to be better if the homology of the fragments used for the structure assembly is higher. Especially more complex elements, such as multi-helical junctions, should be inspected for the level of similarity. This information is provided as a percentage for every fragment in the log file.

3.6 Application Example

Usage of the RNAComposer batch mode is presented on the example of the lysine riboswitch. This RNA regulates biosynthesis and transport of amino acids, mainly lysine [26, 27]. Active domain of *Thermotoga maritima* lysine riboswitch RNA contains 174 nts and its secondary and three-dimensional structures have been solved [28] (PDB id 3DIL). The core of this RNA consists of a five-way junction leading to four hairpins and helix formed by pairing of the 5' and 3' ends. This RNA is also stabilized by the

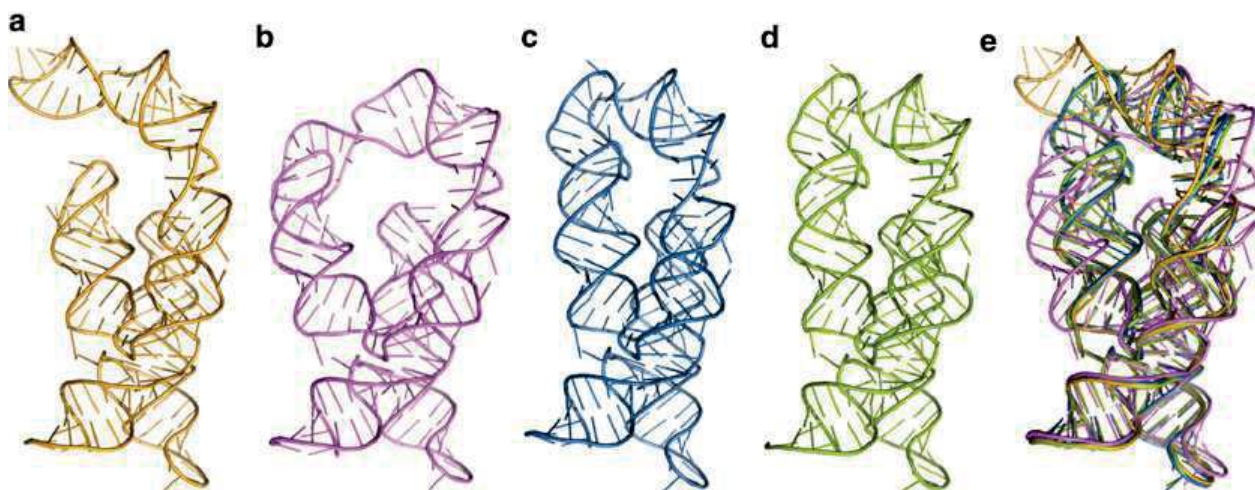


Fig. 6 3D structure representation of lysine riboswitch models obtained from (a) experimental data without pseudoknot, (b) experimental data with pseudoknot determined as atom distance restraints, (c) experimental data with pseudoknot determined as brackets. (d) Reference crystal structure of lysine riboswitch (PDB id. 3DIL). (e) Superposition of obtained models and crystal structure

pseudoknot between two hairpin apical loops. (Fig. 6). This RNA is an interesting example of complex structure and was therefore chosen to present scope and limitations of RNAComposer. To simulate a situation in which 3D structure of unknown RNA is predicted we removed all structural elements derived from 3DIL from the RNAComposer dictionary.

Correct secondary structure information is critical for the prediction of 3D structure. To present RNAComposer dependence on the input secondary structure we have undertaken two approaches: (1) secondary structure of lysine riboswitch RNA was predicted *in silico* or (2) was based on the experimental data from probing experiments [27, 29].

Different web-accessible tools were used for the *in silico* experiment. Secondary structure predicted using KineFold [30] was most similar to that observed in the crystal structure with the Matthews correlation coefficient (MCC) [31] of 0.877. This structure was used for further analyses (Fig. 5 entry box).

Secondary structure probing experiments provide valuable information that help to develop RNA secondary structure models. Recent advancement in probing methods, including new reagents [32] and experiments prepared in high-throughput fashion, made it possible to obtain good quality secondary structure models of large RNAs [33–35]. However, these methods still have limitations. In most cases complex elements, like pseudoknots, require additional analysis to confirm their existence [36]. In the case of lysine riboswitch RNA loop–loop interaction was postulated [27] and further confirmed [29]. To reflect this situation in our tests, experimentally supported structure was used in two options: without and with pseudoknot (Fig. 5 entry box). RNAComposer allows to annotate pseudoknots using square brackets (Fig. 5) or

by providing appropriate distance restraints (Fig. 5). In summary, the following input options for 3D structure prediction of lysine riboswitch were used:

Case 1. The *in silico* predicted secondary structure.

Case 2. Experimentally supported secondary structure without pseudoknot.

Case 3. Experimentally supported secondary structure with the pseudoknot annotated using distance restraints.

Case 4. Experimentally supported secondary structure with the pseudoknot annotated using square brackets notation.

For each RNA secondary structure 10 three-dimensional structural models were predicted. Entire prediction took 26 min, thus single 3D structure was predicted in about 40 s. For clarity, we analyze here in detail only model no. 1 generated for each case.

According to the previously reported estimation [11] total energy calculated by RNAComposer for the 174 nts RNA should not exceed -3474 kcal/mol. Examination of the log file showed that all of the models meet this criterion suggesting that they can be further analyzed.

Case 1.

During the modeling of 3D structure using *in silico* predicted RNA secondary structure input, one 3D element ought to be generated by the RNAComposer engine. The 3D element representing required secondary structure topology of the five-way junction predicted *in silico* was not present in the RNAComposer dictionary. The predicted 3D model showed departure from the RNA crystal structure, as demonstrated by RMSD of 24.82 Å (Table 2). This could be anticipated since an input secondary structure was inaccurate. In such case total energy criterion might not be informative. Moreover, this case confirms that the quality of the predicted 3D models is strongly dependent on the RNAComposer dictionary content.

Case 2.

When the correct secondary structure was used as input but the pseudoknot was omitted, calculated RMSD of 6.91 Å (Table 2)

Table 2
Energy and global RMSD of models obtained for lysine riboswitch compared with crystal structure of 3DIL

RNA secondary structure model	RMSD (Å)	Energy (kcal/mol)
In silico predicted	24.82	-3642.483
Experimental without pseudoknot	6.91	-4494.975
Experimental with atom distance restraints	5.82	-4245.223
Experimental with square brackets	1.72	-4628.297

showed that predicted 3D model closely resembled RNA crystal structure (Fig. 6a, e). Local departure from the correct structure was observed within the region corresponding to the apical loops forming pseudoknot in lysine riboswitch RNA, underlying the importance of the fully correct input secondary structure. It is important to note that five-way junction was well predicted and therefore correct orientation of all helices was observed.

Case 3.

To correct prediction in Case 2 model, distances between residues involved in pseudoknot formation were fixed (Fig. 5). For this purpose RNA helix with the sequence identical to the loop-loop interaction was extracted from RNA FRABASE [13, 14] and distances between atoms characteristic for this helix were used to restrain the prediction. Use of this additional RNAComposer functionality allowed to obtain model with slightly better global RMSD (5.82 Å) (Table 2). However, kink introduced by the restraints within one of the hairpins affected conformation of adjoining structural elements (Fig. 6b, e). Sequence similarity between input structure and element used by RNAComposer for loop 1 (nts 41–54) was 42.86%, while for the loop 2 (nts 93–101) it was 66.67%. Still, predicted and crystal structure showed good overall agreement.

Case 4.

When correct input secondary structure was used and loop-loop interaction was annotated with square brackets predicted structure showed global RMSD of 1.72 Å (Table 2, Fig. 6c, e) demonstrating predictive power of RNAComposer.

3.7 User Workspace

Registered users have the opportunity to view their latest batches in the personal workspace. It is accessible after login and selecting “My workspace” option in the login area of menu panel. Also after running a new batch, the user is immediately taken to the workspace. There, the list of all uploaded batches that are unfinished or were finished less than 2 weeks ago, is displayed. For every batch in the list, the user can see the time of its uploading to the system (i.e., the time of pressing “Compose” button), its launching (i.e., actual time in which it left the queue and started to be processed), and its completion. The details concerning batch input data can be recalled after clicking on its identifier which takes the user to batch details page. Batch processing results (pdb and log files) can be saved to a local disk upon selecting the “Download” option. Every completed batch, together with its results, can be removed from the system after selecting it from the list and pressing the “Delete selected” button. Otherwise, it is stored in the system for 2 weeks and next removed after notifying the user with a warning e-mail.

4 Notes

RNAComposer fidelity was validated on 40 RNA 3D structures of different complexity and ranging in length from 33 to 161 nts [11]. The method is fully applicable to the prediction of 3D structures of large RNA, up to 500 nucleotides [36], including folding intermediates [37]. RNAComposer is being instantly developed and its dictionary volume increases, thus predictions are more accurate. However several points should be considered.

Note 1. As demonstrated in the application example, correct secondary structure is critical to obtain accurate 3D structure.

Note 2. The in silico RNA structure prediction of larger RNA should be reinforced by incorporation of the restraints obtained in RNA structure probing experiments.

Note 3. Log file should be thoroughly examined by the user. Total energy value is an important indicator of prediction quality. High total energy value disqualifies the model.

Note 4. Models containing 3D elements generated by RNAComposer usually show lower accuracy.

Note 5. In general, RNAComposer prediction does not depend on fragments homology. However if low homology fragments are used by RNAComposer for critical structural elements like multihelical junctions, quality of the predicted structure might be lower.

Note 6. Users might want to test predicted models experimentally. Possibilities include hydroxyl radical probing or introduction of mutations that would compromise predicted long-range contacts in RNA 3D models [36]. In addition, RNAComposer can be used to derive models based on NMR experimental data [38].

Acknowledgments

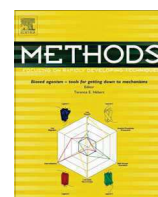
This work was supported by the National Science Center Poland [MAESTRO 2012/06/A/ST6/00384 (to R.W.A)] and Ministry of Science and Higher Education [0492/IP1/2013/72 (to K.J.P.)].

References

1. Spitale RC, Flynn RA, Torre EA, Kool ET, Chang HY (2014) RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip Rev RNA* 5(6):867–881. doi:10.1002/wrna.1253
2. Tian S, Cordero P, Kladwang W, Das R (2014) High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* 20(11):1815–1826. doi:10.1261/rna.044321.114
3. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms.

- RNA 14(6):1164–1173. doi:10.1261/rna.894608
4. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15(2):189–199. doi:10.1261/rna.1270809
 5. Rother M, Rother K, Puton T, Bujnicki JM (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* 39(10):4007–4022. doi:10.1093/nar/gkq1320
 6. Das R, Baker D (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci USA* 104(37):14664–14669. doi:10.1073/pnas.0703836104
 7. Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452(7183):51–55. doi:10.1038/nature06684
 8. Cao S, Chen SJ (2011) Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* 115(14):4216–4226. doi:10.1021/jp112059y
 9. Sharma S, Ding F, Dokholyan NV (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* 24(17):1951–1952. doi:10.1093/bioinformatics/btn328
 10. Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y (2012) Automated and fast building of three-dimensional RNA structures. *Sci Rep* 2:734. doi:10.1038/srep00734
 11. Popena M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40(14):e112. doi:10.1093/nar/gks339
 12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242
 13. Popena M, Blazewicz M, Szachniuk M, Adamiak RW (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res* 36(Database issue):D386–D391. doi:10.1093/nar/gkm786
 14. Popena M, Szachniuk M, Blazewicz M, Wasik S, Burke EK, Blazewicz J, Adamiak RW (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinformatics* 11:231. doi:10.1186/1471-2105-11-231
 15. Gan HH, Pasquali S, Schlick T (2003) Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res* 31(11):2926–2943
 16. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73
 17. Herraes A (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem Mol Biol Educ* 34(4):255–261. doi:10.1002/bmb.2006.494034042644
 18. Childs-Disney JL, Yildirim I, Park H, Lohman JR, Guan L, Tran T, Sarkar P, Schatz GC, Disney MD (2014) Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. *ACS Chem Biol* 9(2):538–550. doi:10.1021/cb4007387
 19. Mathews DH (2014) RNA secondary structure analysis using RNAstructure. *Curr Protoc Bioinformatics* 46:12.16.11–12.16.25. doi:10.1002/0471250953.bi1206s46
 20. Do CB, Woods DA, Batzoglou S (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* 22(14):e90–e98. doi:10.1093/bioinformatics/btl246
 21. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte Fur Chem* 125(2):167–188. doi:10.1007/Bf00818163
 22. Sergiev PV, Dontsova OA, Bogdanov AA (2001) Chemical methods for the structural study of the ribosome: judgment day. *Mol Biol* 35(4):472–495. doi:10.1023/A:1010506522897
 23. Furtig B, Richter C, Wohnert J, Schwalbe H (2003) NMR spectroscopy of RNA. *ChemBiochem* 4(10):936–962. doi:10.1002/cbic.200300700
 24. Wozniak AK, Nottrott S, Kuhn-Holsken E, Schroder GF, Grubmuller H, Luhrmann R, Seidel CA, Oesterhelt F (2005) Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA* 11(10):1545–1554. doi:10.1261/rna.2950605
 25. Frolow O, Endeward B, Schiemann O, Prisner TF, Engels JW (2008) Nitroxide spin labeled RNA for long range distance measurements by EPR-PELDOR. *Nucleic Acids Symp Ser (Oxf)* 52:153–154. doi:10.1093/nass/nrn078
 26. Huang LL, Serganov A, Patel DJ (2010) Structural insights into ligand recognition by a

- sensing domain of the cooperative glycine riboswitch. *Mol Cell* 40(5):774–786. doi:10.1016/j.molcel.2010.11.026
27. Grundy FJ, Lehman SC, Henkin TM (2003) The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci U S A* 100(21):12057–12062. doi:10.1073/pnas.2133705100
 28. Serganov A, Huang L, Patel DJ (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* 455(7217):1263–1267. doi:10.1038/nature07326
 29. Blouin S, Lafontaine DA (2007) A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* 13(8):1256–1267. doi:10.1261/Rna.560307
 30. Xayaphoummine A, Bucher T, Isambert H (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res* 33(Web Server issue):W605–W610. doi:10.1093/nar/gki447
 31. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 405(2):442–451
 32. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY (2013) RNA SHAPE analysis in living cells. *Nat Chem Biol* 9(1):18–20. doi:10.1038/nchembio.1131
 33. Purzycka KJ, Pachulska-Wieczorek K, Adamiak RW (2011) The in vitro loose dimer structure and rearrangements of the HIV-2 leader RNA. *Nucleic Acids Res* 39(16):7234–7248. doi:10.1093/nar/gkr385
 34. Legiewicz M, Zolotukhin AS, Pilkington GR, Purzycka KJ, Mitchell M, Uranishi H, Bear J, Pavlakis GN, Le Grice SF, Felber BK (2010) The RNA transport element of the murine musD retrotransposon requires long-range intramolecular interactions for function. *J Biol Chem* 285(53):42097–42104. doi:10.1074/jbc.M110.182840
 35. Purzycka KJ, Legiewicz M, Matsuda E, Eizentstat LD, Lusvarghi S, Saha A, Le Grice SF, Garfinkel DJ (2013) Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res* 41(1):463–473. doi:10.1093/nar/gks983
 36. Huang Q, Purzycka KJ, Lusvarghi S, Li D, Legrice SF, Boeke JD (2013) Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA* 19(3):320–332. doi:10.1261/rna.035535.112
 37. Lusvarghi S, Sztuba-Solinska J, Purzycka KJ, Pauly GT, Rausch JW, Grice SF (2013) The HIV-2 Rev-response element: determining secondary structure and defining folding intermediates. *Nucleic Acids Res* 41(13):6637–6649. doi:10.1093/nar/gkt353
 38. Krahenbuhl B, Lukavsky P, Wider G (2014) Strategy for automated NMR resonance assignment of RNA: application to 48-nucleotide K10. *J Biomol NMR* 59(4):231–240. doi:10.1007/s10858-014-9841-3



RNAComposer and RNA 3D structure prediction for nanotechnology



Marcin Biesiada^{a,b}, Katarzyna Pachulska-Wieczorek^a, Ryszard W. Adamiak^{a,b}, Katarzyna J. Purzycka^{a,*}

^a Department of Structural Chemistry and Biology of Nucleic Acids, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

^b Institute of Computing Science, Poznan University of Technology, Poznan, Poland

ARTICLE INFO

Article history:

Received 4 January 2016

Received in revised form 4 March 2016

Accepted 21 March 2016

Available online 22 March 2016

Keywords:

RNA
RNA 3D structure
RNA structure prediction
Structural bioinformatics
RNA nanotechnology

ABSTRACT

RNAs adopt specific, stable tertiary architectures to perform their activities. Knowledge of RNA tertiary structure is fundamental to understand RNA functions beginning with transcription and ending with turnover. Contrary to advanced RNA secondary structure prediction algorithms, which allow good accuracy when experimental data are integrated into the prediction, tertiary structure prediction of large RNAs still remains a significant challenge. However, the field of RNA tertiary structure prediction is rapidly developing and new computational methods based on different strategies are emerging.

RNAComposer is a user-friendly and freely available server for 3D structure prediction of RNA up to 500 nucleotide residues. RNAComposer employs fully automated fragment assembly based on RNA secondary structure specified by the user. Importantly, this method allows incorporation of distance restraints derived from the experimental data to strengthen the 3D predictions. The potential and limitations of RNAComposer are discussed and an application to RNA design for nanotechnology is presented.

© 2016 Elsevier Inc. All rights reserved.

1. Background

RNA molecules are engaged in almost every aspect of cell life and new RNA functions are still to be discovered. The function of an RNA is dictated by its sequence and structure. RNA secondary structure contains single and double stranded regions. It is estimated that more than 50% of RNA secondary structure consists of A-form helices [1]. However, the elements joining the helical regions are important for the relative positioning of helical regions and therefore RNA 3D structure. Longer RNAs can acquire complex structures through the coalescence of independently folded subdomains. RNA structures are critical for the function of both mRNAs and non-coding RNAs, including microRNAs (miRNAs), riboswitches, ribozymes and long non-coding RNAs (lncRNAs). The fact that specific RNA tertiary architectures are fundamental to RNA function makes RNA molecules promising drug targets as well as therapeutic agents. Knowledge about RNA 3D structure is not only instructive to understand RNA function but is also required for structure-based drug design and useful in the emerging field of RNA nanotechnology [2].

RNA secondary structure prediction algorithms are advanced and allow reasonable accuracy especially when experimental data are integrated into the prediction [3]. In addition to the gold standards of RNA secondary structure determination, such as DMS probing or more recent techniques like SHAPE, structure mapping

integrated with systematic mutagenesis makes it possible to obtain accurate secondary structure models even for large RNAs [3,4]. RNA atomic structures can be obtained by crystallography, NMR, and cryo-electron microscopy, however, RNA size, conformational heterogeneity, and multiple functional states can complicate structural determinations. In 2015, only 73 new RNA structures were deposited into the RCSB PDB, with the total number of RNAs only at 1165. In contrast, over 100 thousand protein structures have been deposited into the RCSB PDB. The largest RNAs for which 3D structures were solved are components of the ribosome, and those are stabilized by interactions with proteins. Although RNA 3D structure prediction still remains a significant challenge [5], new computational strategies are emerging for RNA 3D structure predictions that are based on alternative strategies [6–8]. Physics-based automated methods use fragment assembly [9,10], coarse-grained and atomic-level molecular dynamics [11–15] and internal coordinate space dynamics [16,17]. Knowledge-based comparative modeling [18] depends on 3D structural templates and unequivocal sequence alignment. Several methods are automated such as Vfold [11], iFoldRNA [14], 3dRNA [19], and RNAComposer [20]. All have limitations but the field of computational RNA tertiary structure prediction is rapidly developing. New experimental approaches like MOHCA-seq (Multiplexed .OH Cleavage Analysis with paired-end sequencing) [21] may substantially strengthen the 3D predictions, since MOHCA-seq provides tertiary proximities. However, among automated methods, only iFoldRNA [14] and RNAComposer [20] allow incorporation of data during 3D structure modeling.

* Corresponding author.

E-mail address: purzycka@ibch.poznan.pl (K.J. Purzycka).

2. The outline of RNA 3D structure prediction using RNAComposer

RNAComposer [20] employs fully automated fragment assembly based on a secondary RNA structure specified by the user, and is freely available through web-servers: <http://rnacomposer.ibch.poznan.pl> and mirror <http://rnacomposer.cs.put.poznan.pl>.

The user provides sequence and secondary structure in dot-bracket notation. Based on the graph representation [22], secondary structure is divided into fragments containing overlapping canonical base pairs. Those smaller secondary structure elements are related to the corresponding 3D structure elements. The appropriate 3D element is chosen using a dedicated dictionary tailored from the RNA FRABASE database [23,24]. This dictionary contains 3D structure elements derived from nearly all RNA structures deposited into the RCSB PDB. The dictionary is automatically updated and contains almost 500,000 3D elements. 3D structure elements are related to the secondary structure elements based on the secondary structure and sequence homology, and purine/pyrimidine compatibility. If more than one appropriate 3D element is found, additional criteria, such as the resolution of the original structure deposited into the RCSB PDB and its energy, are used to help define the most suitable element. If the appropriate 3D structure element is missing, which can happen for complex RNA structures, RNAComposer will generate the element. Once all the secondary structure elements have matching 3D elements, RNAComposer will automatically assemble the 3D elements using overlapping canonical base pairs. Resultant structures are subjected to the energy minimization in the torsion angle space [25] and subsequently in the Cartesian atom coordinate space [26]. Final, high-quality RNA 3D models are returned to the user.

RNAComposer was initially evaluated using a representative benchmark set of 40 RNAs of different complexity, and the secondary structures derived from the X-ray structures [20]. We demonstrated that RNAComposer was characterized by high efficiency and superior performance. Importantly, even the most complex structures from this set showed the correct conformation of *n*-way junctions, orientation of helices or close proximity of the loops that in the reference structures were involved in loop-loop interactions. More recently, we have tested RNAComposer using 10 riboswitches with complex 3D structures, containing pseudoknots and extensive tertiary interactions [27]. Even within this set, nine out of ten examples were characterized by high accuracy and very good recovery of canonical and noncanonical base pairing and stacking.

3. RNAComposer input and output data

RNAComposer is available in two modes: interactive mode (for unregistered users) and batch mode (for registered users). Interactive mode is tailored to the inexperienced user and can operate on RNA sequence only; secondary structure can be predicted using one of the methods incorporated within RNAComposer: RNAfold [28], RNAstructure [29], or CONTRAfold [30]. The interactive mode is useful for modeling simple RNA structures. RNAComposer allows for very short computing times. Therefore in batch mode, 1000 3D structures can be obtained within a single RNAComposer batch. The user can provide an up to 10 RNA sequences, with 1–10 secondary structure topologies (in dot-bracket notation) for every sequence, and 10 RNA 3D structure models can be predicted for each sequence and secondary structure. RNAComposer prediction strongly depends on the quality of the input data since secondary structure topology will not be altered during the prediction [20,27]. Therefore, it is important to provide the correct secondary structure. For larger RNAs, we recommend working on experimentally supported secondary structures. Additional information from

RNA secondary structure probing experiments can substantially improve *in silico* RNA secondary structure prediction [3].

Batch mode also allows for incorporation of distance restraints that can considerably limit the diversity of possible 3D structure models [4], especially for larger RNAs. Distance restraints can be obtained from chemical probing experiments using hydroxyl radicals [31], UV or chemically induced RNA crosslinking [32], mutational analyses [33,34], MOHCA-seq [21], NMR [35], EPR [36] or FRET [37]. Those can be introduced into RNAComposer as interatomic distances specified for chosen atoms and can be input as a range of values.

The user is provided with a PDB file containing coordinates of the predicted 3D models and a log file. The log file contains information about the prediction process and is an important source of data that can help to evaluate the accuracy of the obtained model. This file includes lists of the structure elements resulting from secondary structure fragmentation, and the tertiary elements (sequence and topology) selected from the dictionary for assembly, together with their origin (PDB ID code). These lists also provide information about sequence similarity and the final energy of the 3D model.

RNAComposer does not require high homology between the secondary structure element and the 3D element chosen from the dictionary; reasonable models are obtained even with elements containing less than 5% homology [27]. However if one uses fragments with low homology for critical structural elements like multihelical junctions, the quality of the predicted structure may be lower. In such case, additional data, like distance restraints, should be considered by the user. If RNAComposer must generate a specific element due to its absence in the dictionary, the resultant 3D RNA structure is usually less accurate. This information is provided within the log file.

The user should also inspect the energy of the final structure. We have shown that the energy values change linearly with RNA strand length [20], and this value is a good indicator of the quality of the 3D structure model. However, a predicted model can have the appropriate energy and still show extensive departure from the correct 3D structure when the input secondary structure is extensively inaccurate. This phenomenon is demonstrated by comparing the *c*-di-GMP-II riboswitch generated computationally with its known 3D structure [27]. For highly accurate prediction, it is important to ensure incorporation of the correct secondary structure topology and distance restraints, especially for large RNAs.

4. Applications

We thoroughly tested the accuracy and quality of RNAComposer predictions using RNAs of different complexity with known X-ray structures [20,27]. We have also recently participated in collective and blind experiments in 3D RNA structure prediction as part of RNA Puzzles [4]. The experimental material contained sequences from large RNA structures with limited or no homology with previously solved RNA molecules. RNA 3D structure predicted by RNAComposer with the aid of experimental data [4] ranked as one of the best models, suggesting that RNAComposer can already provide useful structural information for biological problems. RNAComposer was also used to derived models based on NMR [38], SAXS [39] and cryo-EM [40]. When data from structure probing experiments was used, models of RNA larger than 300 nt [33], or folding intermediates [31] were deciphered, answering functional questions. When supported by additional software, even models of RNAs larger than 500 nt were predicted [41]. RNAComposer constitutes a central element of new approaches for RNA structure analysis and quality control [42,43]. Its strengths and speed make it ideal for *in silico* prediction of RNA secondary structures containing non-canonical base-pairs [44]. Since 2012, both

RNAComposer servers have been visited over 150 thousand times. In our first report describing RNAComposer, we included predictions of RNA structures characteristic of randomized sequence, demonstrating the potential of our method to build artificial RNA 3D structures and its suitability for the design of artificial structures for potential therapeutic applications.

4.1. Nanotechnology

RNA is a powerful biomaterial for nanotechnology applications due to its structural versatility, property of self-assembly, low free energy of annealing and the possibilities for controlling its structure [2]. The feasibility of RNA nanotechnology in disease therapy was recently outlined [40,45] and new roles for RNA in nanomedicine applications are expanding [46]. Here, we present the workflow for utilizing RNAComposer in RNA design for potential applications in nanotechnology (Fig. 1). First, we tested RNAComposer accuracy in RNA 3D structure prediction using the known structure of the bacteriophage phi29 prohead RNA that assembles into a ring [47]. Subsequently, we applied RNAComposer to design a novel RNA motif and functionalize a known nanosquare structure [48].

4.1.1. Predicting the tertiary structure of the bacteriophage phi29 prohead RNA

High accuracy RNA 3D structure prediction is critical if the model obtained will be used for programmed self-assembly. Aberration from the desired orientation of the helices may result in an inability to construct higher order structures. Only a few 3D structures of RNA nanoparticles are available from RCSB database. The bacteriophage phi29 prohead RNA (pRNA) has been extensively applied in the nanotechnology field [46]. The 3D structure of the pentameric ring assembled from pRNAs has been determined using electron microscopy [47]. In the first step of our analyses, RNAstructure [29] was used for the RNA secondary structure prediction. The resultant structure was compared with the available experimental data [49]. Sequence and secondary structure in dot-bracket notation were introduced to RNAComposer server. Importantly, for this test, all the 3D structure elements comprised by the model of the bacteriophage phi29 prohead RNA PDB structure (1FOQ) [47] were excluded from the RNAComposer dictionary before prediction. Moreover, we also excluded elements derived from homologous structures: Phi29 RNA E-loop hairpin (2KVN [50]), essential RNA ring component (3R4F [51]); and phi29 RNA 3-way junction (4KZ2 [52]). This allowed our experiment to mimic the intricate problem of the lack of obvious homologs that users may face. RNAComposer predicted the 3D structure of pRNA within seconds. The total energy value for this 3D model was -2552.13 kcal/mol, which is within expected range. The accuracy, as indicated by network fidelity parameters as well as stereochemical quality of the predicted model, was very high (Table 1). Several

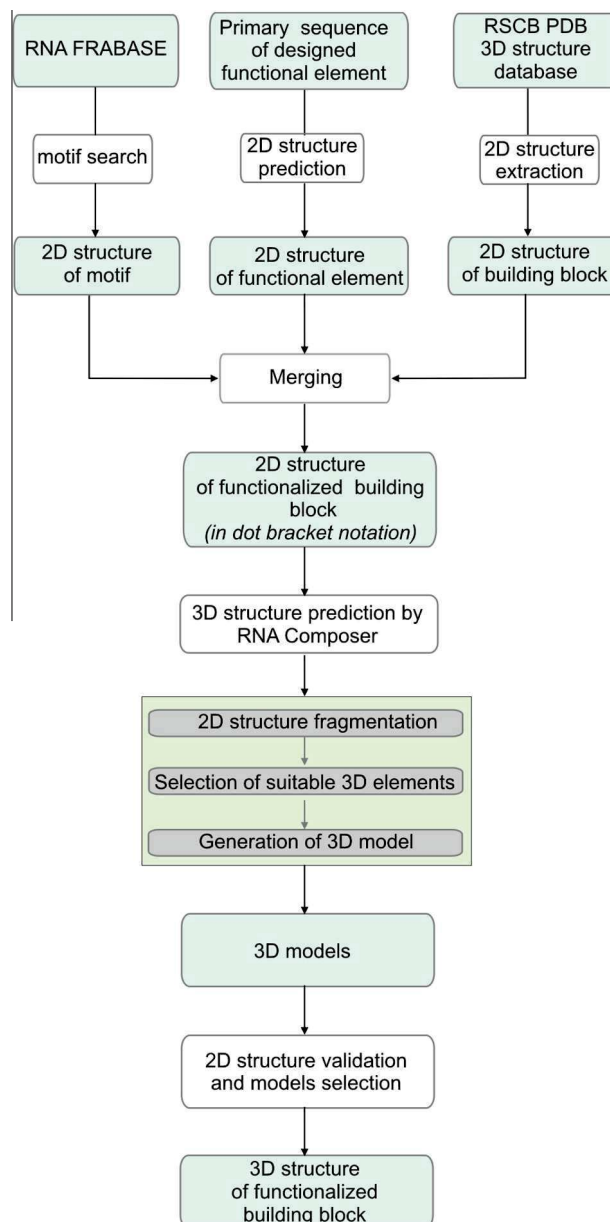


Fig. 1. The scheme of the workflow for utilizing RNAComposer in RNA design.

secondary structure elements of pRNA did not have highly similar corresponding 3D elements in RNAComposer dictionary, including the 3-way junction. Therefore elements with low sequence similarity and pyrimidine/purine compatibility were used for pRNA 3D model assembly (Table 2). Despite this, the orientation of the

Table 1
Accuracy and quality of the RNAComposer predicted pRNA 3D model and 3D structures of bacteriophage phi29 prohead RNA (1FOQ), Phi29 RNA E-loop hairpin (2KVN), essential RNA ring component (3R4F) and phi29 RNA 3-way junction (4KZ2).

Reference PDB id.	Accuracy*			Nucleic acid geometry					
	INF ^{cbp} **	INF ^{all} **	RMSD [Å]	Clash-score, all atoms	Outlier bonds [%]	Outlier angles [%]	Potentially incorrect		
							Sugar puckers[%]	Backbone[%]	
RNAComposer model	–	–	–	0.00	0.00	0.00	5.00	32.50	
1FOQ	0.78	0.77	9.4	75.65	3.03	6.42	5.50	0.00	
2KVN	0.83	0.65	2.3	42.28	0.00	0.00	0.00	0.00	
3R4F	0.85	0.81	11.4	37.81	0.00	0.53	7.58	24.24	
4KZ2	0.88	0.85	9.2	6.40	0.00	0.00	0.00	7.41	

* Calculated for the predicted model with the structure of indicated PDB id.

** INF – interaction network fidelity, INF score range from 0 (worst) to 1.0 (best).

Table 2
3D structure prediction details for the pRNA.

Structural element	Target secondary structure element				Source tertiary structure element					Sequence homology	
	Localization		Secondary structure	Topology	Localization		Secondary structure			Sequence similarity [%]	Pyrimidines/purines compatibility [%]
	Residues				PDB ID	Residues		Sequence	Topology		
	First	Last	Sequence	First		Last	Sequence		Topology		
stem 1	1	4	GGAA	(((3OI2	133	136	GGAA	(((100	100
	114	117	UUCC)))		218	221	UUCC)))		
loop1	4	6	AUG	(.(4A1D	1207	1209	AUG	(.(100	100
	48	52	UUGAG)...)		1259	1263	UGGAG)...)		
stem2	6	9	GGUA	(((3ZEX	1918	1921	GGUA	(((100	100
	110	113	UACU)))		2374	2377	UACU)))		
loop2	9	10	AC	((1NBR	12	13	AC	((100	100
	108	110	GCU	.)		17	19	GCU	.)		
stem3	10	11	CG	((1K01	1930	1931	CG	((100	100
	107	108	CG)		1941	1942	CG)		
loop3	11	12	GG	((1YL4	931	932	GG	((100	100
	108	110	CAC	.)		1208	1210	CAC	.)		
stem4	12	17	GUACUU	((((3KNH	925	930	GCAUGU	((((66.7	83.3
	100	105	AAGUGC))))		1211	1216	ACGUGC))))		
loop4	17	21	UGCAU	(...(3ZEX	2564	2568	GGUUG	(...(14.3	28.6
	99	100	AA)		2598	2599	CU)		
stem5	21	28	UUGUCAUG	(((((((2IHX	1	8	CUGCCUC	(((((((56.2	75
	92	99	CAUGGCAA))))))		68	75	GAGGGCAG))))))		
loop5	28	30	GUG	(.(3J44	637	639	UGG	(.(20	40
	71	75	CUUUG)...)		827	831	UAAGA)...)		
	91	92	UC)		931	932	UA)		
stem6	30	34	GUAUG	(((1GAX	48	52	GUAGG	(((80	80
	67	71	CAUAC))))		60	64	CCUAC))))		
loop6	34	36	GUU	(.(1HR2	186	188	GUU	(.(100	100
	66	67	AC)		249	250	AC)		
stem7	36	39	UGGG	(((3IZV	125	128	UGGG	(((100	100
	63	66	CCCA)))		233	236	CCCA)))		
loop7	39	48	GGAUUAAACCC	(.....(3J44	2233	2243	GCGCAGCGUAG	(.....(15.4	53.8
	62	63	GC)		2301	2301	CC)		
stem8	49	52	CUGA	(((2QBI	31	34	CUGA	(((100	100
	59	62	UCAG)))		48	51	UCAG)))		
loop8	52	59	AUUGAGUU	(.....)	3BBN	1082	1089	GUUGAGUU	(.....)	87.5	100
stem9	75	80	GUUGAU	((((4A19	479	484	GUCGAU	((((83.3	100
	86	91	GUCAAU))))		494	499	GUCGAU))))		
loop9	80	86	UUGGUUG	(.....)	3CC7	2586	2592	UUGUUCG	(.....)	71.4	85.7
single strand	117	120	CUAA)...)	3J45	283	286	CUAA)...)	100	100

helices in the predicted pRNA 3D structure was correct and superposition with the pentameric ring structure (1FOQ) demonstrated the potential for correct assembly (Fig. 2A). Loop-loop interactions are not present in the model, however this is expected since we predicted the 3D structure of monomeric pRNA. The apical loops of monomers are in close proximity (Fig. 2B). The rmsd calculated with the structures of pRNA and its fragments were within the expected range, taking into account the resolution of the source structures (Table 1).

4.1.2. Functionalizing nanosquare structure

We applied RNAComposer to design novel RNA motif and functionalize a previously tested nanosquare building block (Fig. 3A,E) for which Dibrov et al. [48] demonstrated the programmed self-assembly of RNA squares from complex mixtures of corner units (building blocks). This RNA square can be potentially used as a combinatorial nanoscale platform. We wanted to introduce a new hairpin (HP1) into the nanosquare building block. Such a hairpin can be designed to produce multifunctional RNA nanoparticles that contain Dicer substrate RNA or a specific motif such as a ligand

binding site, as presented by Afonin et al. [40]. Moreover, functionalized corner units can potentially serve for the controlled construction of RNA nano-objects. Using functionalized corner units with different topologies and orientations of HP1, which contains a sequence able to form a kissing-loop interaction, it should be possible to modulate the architecture of RNA higher-order nanostructures, i.e. two-dimensional arrays. The sequence of HP1 is G-C rich and was designed to form a stable structure. The stability of this hairpin was estimated using RNAstructure [29]. The calculated lowest free-energy secondary structure corresponded to the desired RNA conformation (Fig. 3C) and no other secondary structure was predicted within 20% of the lowest energy structure. By introducing different sequences in the HP1 apical loop one could choose an angle between the z-axes of the stems of respective hairpins interacting through the kissing-loop. Therefore after introducing HP1 into the nanosquare building block, control of the architecture of higher-order nanostructures could be achieved by modulating intermolecular interactions. FRABASE [23,24] was used to select desirable kissing-loops, but RNAJunction Database [53] can be also used as an alternative tool. We searched FRABASE for all available structures of kissing-loop complexes with apical loops smaller than 10 nt and at least 4 nt involved in loop-loop interactions. We selected 15 examples with the angles between hairpins z-axes ranging from ~ 130 to 170° (Table 3). To functionalize a nanosquare building block with the engineered HP1, its internal loop must be replaced with a 3-way junction. Such a junction is another element that allows regulation of the architecture of two-dimensional arrays. This can be achieved by controlling the positioning of HP1 relative to the corner unit helices. FRABASE [23,24] was searched for 3-way junctions of different topologies. RNAJunction Database [53] is another excellent tool that can also be used. For our search the first strand was left intact to limit the number of elements while the second and third varied the number of nucleotides from 1 to 10 (Fig. 3D). Applying “Strand shift operation” in FRABASE resulted in 2514 different 3-way junction hits. Junctions with repeated sequences and similar 3D structures were removed to obtain 158 diverse 3D elements. Taking into account the predictive power of RNAComposer, all of the elements were converted into RNAComposer dot-bracket annotated input, and 16 different batches were run. Each batch consisted of 10 tasks with different sequences and secondary structures (8 for the last batch) and three 3D structures were predicted for each task. Since RNAComposer can only predict 3D structures for one RNA strand, the nanosquare building block was closed on one side by an apical loop (Fig. 3B,F). The entire procedure took less than three hours. Total energy values of the resultant functionalized building blocks were analyzed. Energy values change linearly with the RNA strand length and for this example they should not be higher than -1557 to -1278 kcal/mol. All of the predicted 3D structures were within the expected range. 3D structures for which desired orientation of the corner unit helices was preserved were selected for further analyses (Fig. 3G,H). Secondary structures of functionalized building blocks can be different than assumed at the stage of merging dot-bracket representations of corner unit, 3-way junction and HP1 (Fig. 1). Therefore for the chosen subset of predicted 3D models, RNAstructure [29] was used to verify if the introduction of additional sequences constituting the 3-way junction do not destabilize HP1 or do not base pair with the corner unit sequence. After validating their stereochemical quality, ten different 3D models were chosen that represent the desired secondary structure and different topologies of the 3D structures (Fig. 3H,I). This experiment demonstrates the possibility to utilize RNAComposer for RNA design in RNA nanotechnology.

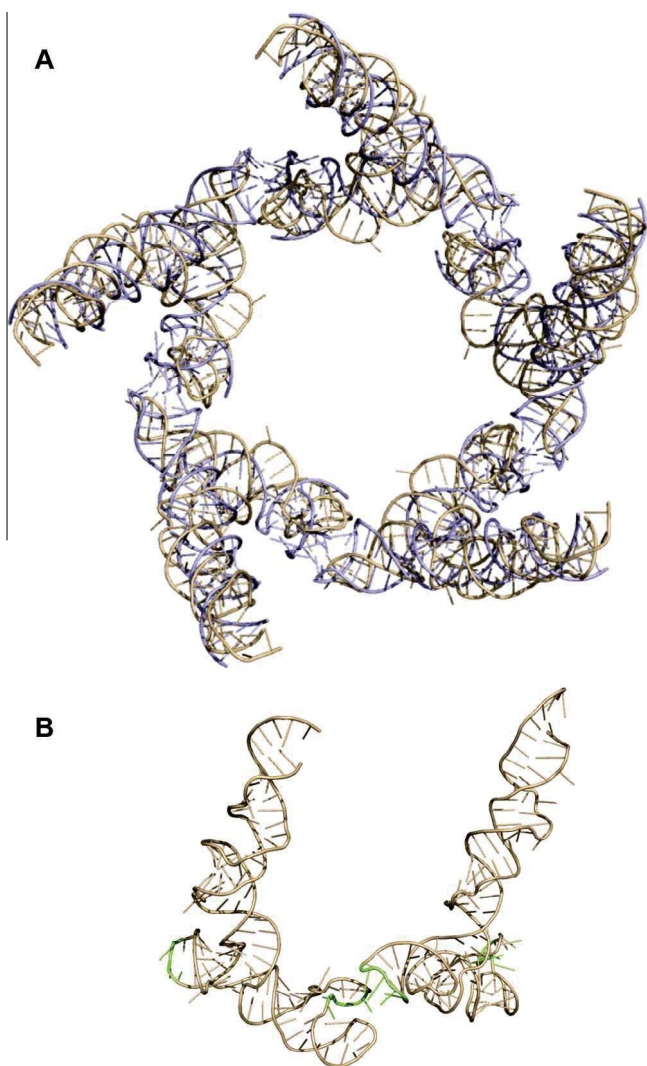


Fig. 2. (A) Superposition of the predicted 3D model of pRNA (gold) and structure of the pentameric ring assembled from pRNAs (1FOQ; lilac). (B) Kissing-loop interfaces are marked in green on the model of dimeric unit of pRNAs.

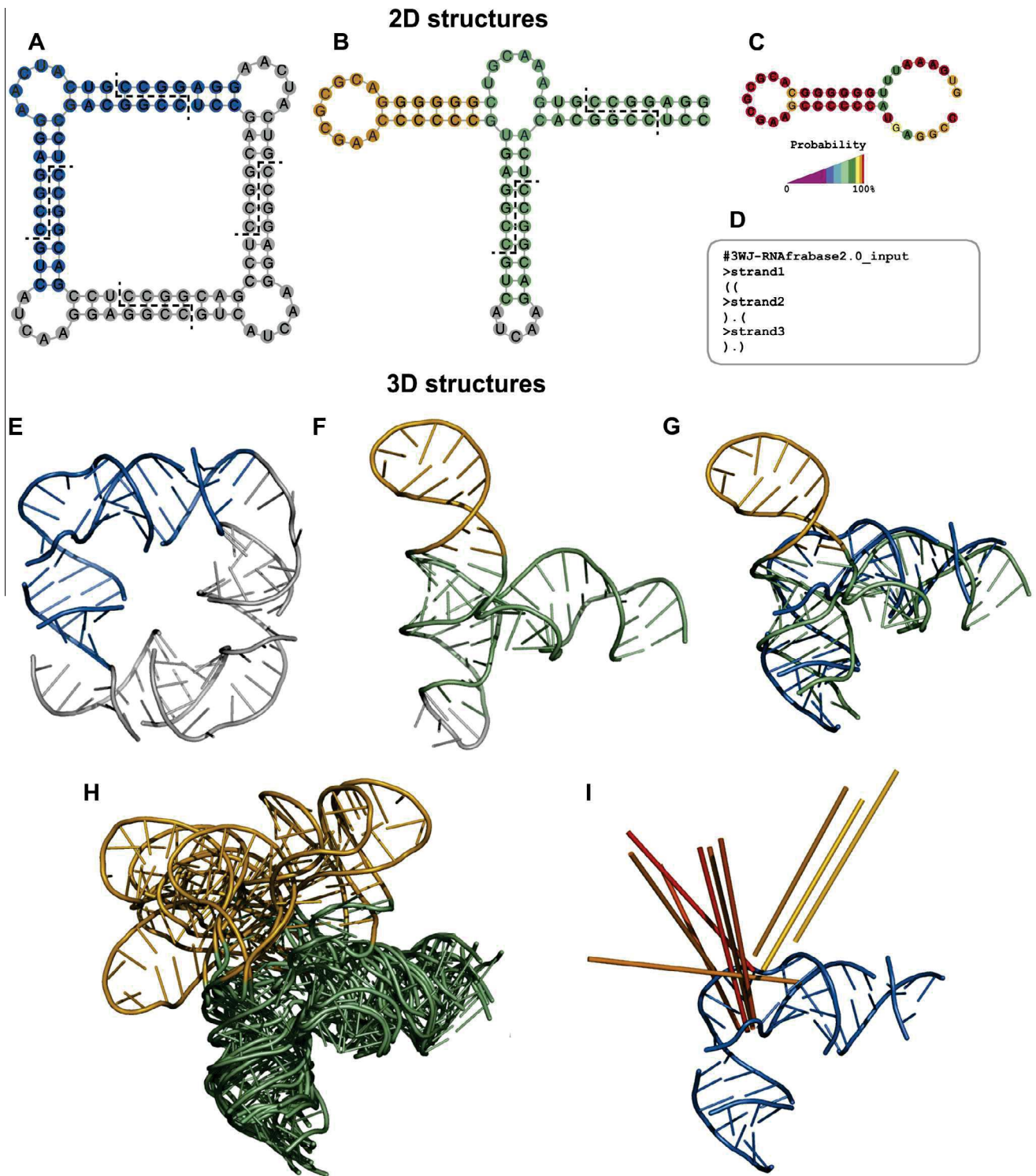


Fig. 3. Representation of nanosquare secondary (A) and 3D (E) structure with the building block used for RNAComposer prediction shown in blue. Secondary (B) and 3D (F) structure of the element containing 3-way junction are shown in green, with the additional hairpin (HP1) in orange. (C) RNAstructure prediction of HP1 with the base-pairing probability annotated. (D) RNA FRABASE 3-way junction search input. (G) 3D structure of nanosquare building block superimposed with the predicted structure of the functionalized element. (H) Ten superimposed representative models are shown and in (I) HP1 is represented by sticks.

5. Concluding remarks

Similar to understanding protein structure, studies of RNA structure have a history decades long. However, contrary to proteins, methods for RNA 3D structure prediction have just recently

emerged. Although these methods are still in the early phases of development and are not free from problems and limitations, several have been used successfully to test biological hypotheses [4]. RNAComposer has successfully predicted the 3D structures of diverse RNAs [20,27] based on the secondary structures, was

Table 3
Examples of the kissing-loops complexes.

PDB id.	stems z-axes angle [°]
4GMA	133.5
2BJ2	142.7
2JLT	143.2
1VW3	149.4
3J3F	154.3
3U5H	157.3
4A1D	158.8
3J3E	160.4
4KJB	160.5
3V2F	161.0
1VQO	166.5
3ZEX	167.5
1XPF	168.6
1XPE	170.8
1ZCI	173.3

already used in nanotechnology field [40,54,55] and allows incorporation of distance restraints derived from experimental data to guide 3D structure predictions. RNAComposer continues to be developed as its dictionary volume increases, resulting in more accurate predictions of 3D structure that can be functionally validated.

Acknowledgements

This work was supported by the National Science Center Poland [2012/06/A/ST6/00384], the Ministry of Science and Higher Education Poland [0492/IP1/2013/72] and [01/KNOW2/2014]. Poznan Supercomputing and Networking Centre is acknowledged for hosting the RNAComposer server at <http://rnacomposer.ibch.poznan.pl>. KJP also acknowledges support from the Ministry of Science and Higher Education Poland (MNiSW, fellowship for outstanding young scientists).

References

- [1] N.B. Leontis, E. Westhof, Geometric nomenclature and classification of RNA base pairs, *RNA* 7 (4) (2001) 499–512.
- [2] P. Guo, F. Haque, B. Hallahan, R. Reif, H. Li, Uniqueness, advantages, challenges, solutions, and perspectives in therapeutics applying RNA nanotechnology, *Nucleic Acid Ther.* 22 (4) (2012) 226–245, <http://dx.doi.org/10.1089/nat.2012.0350>.
- [3] K.E. Deigan, T.W. Li, D.H. Mathews, K.M. Weeks, Accurate SHAPE-directed RNA structure determination, *Proc. Natl. Acad. Sci. U.S.A.* 106 (1) (2009) 97–102, <http://dx.doi.org/10.1073/pnas.0806929106>.
- [4] Z. Miao, R.W. Adamiak, M.F. Blanchet, M. Boniecki, J.M. Bujnicki, S.J. Chen, C. Cheng, G. Chojnowski, F.C. Chou, P. Cordero, J.A. Cruz, A.R. Ferre-D'Amare, R. Das, F. Ding, N.V. Dokholyan, S. Dunin-Horkawicz, W. Kladwang, A. Krokhotin, G. Lach, M. Magnus, F. Major, T.H. Mann, B. Masquida, D. Matelska, M. Meyer, A. Peselis, M. Popena, K.J. Purzycka, A. Serganov, J. Stasiewicz, M. Szachniuk, A. Tandon, S. Tian, J. Wang, Y. Xiao, X. Xu, J. Zhang, P. Zhao, T. Zok, E. Westhof, RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures, *RNA* 21 (6) (2015) 1066–1084, <http://dx.doi.org/10.1261/rna.049502.114>.
- [5] M.G. Seetin, D.H. Mathews, Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints, *J. Comput. Chem.* (2011), <http://dx.doi.org/10.1002/jcc.21806>.
- [6] F. Jossinet, T.E. Ludwig, E. Westhof, Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels, *Bioinformatics* 26 (16) (2010) 2057–2059, <http://dx.doi.org/10.1093/bioinformatics/btq321>.
- [7] H.M. Martinez, J.V. Maizel Jr., B.A. Shapiro, RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA, *J. Biomol. Struct. Dyn.* 25 (6) (2008) 669–683, <http://dx.doi.org/10.1080/07391102.2008.10531240>.
- [8] N. Leontis, Westhof, 2011, zielona.
- [9] R. Das, J. Karanickolas, D. Baker, Atomic accuracy in predicting and designing noncanonical RNA structure, *Nat. Methods* 7 (4) (2010) 291–294, <http://dx.doi.org/10.1038/nmeth.1433>.
- [10] M. Parisien, F. Major, The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature* 452 (7183) (2008) 51–55, <http://dx.doi.org/10.1038/nature06684>.
- [11] S. Cao, S.J. Chen, Physics-based de novo prediction of RNA 3D structures, *J. Phys. Chem. B* 115 (14) (2011) 4216–4226, <http://dx.doi.org/10.1021/jp112059y>.
- [12] M.A. Jonikas, R.J. Radmer, R.B. Altman, Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models, *Bioinformatics* 25 (24) (2009) 3259–3266, <http://dx.doi.org/10.1093/bioinformatics/btp576>.
- [13] M.A. Jonikas, R.J. Radmer, A. Laederach, R. Das, S. Pearlman, D. Herschlag, R.B. Altman, Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters, *RNA* 15 (2) (2009) 189–199, <http://dx.doi.org/10.1261/rna.1270809>.
- [14] S. Sharma, F. Ding, N.V. Dokholyan, IFoldRNA: three-dimensional RNA structure prediction and folding, *Bioinformatics* 24 (17) (2008) 1951–1952, <http://dx.doi.org/10.1093/bioinformatics/btn328>.
- [15] X. Xu, P. Zhao, S.J. Chen, Vfold: a web server for RNA structure and folding thermodynamics prediction, *PLoS One* 9 (9) (2014), <http://dx.doi.org/10.1371/journal.pone.0107504> e107504.
- [16] S.C. Flores, R.B. Altman, Turning limited experimental information into 3D models of RNA, *RNA* 16 (9) (2010) 1769–1778, <http://dx.doi.org/10.1261/rna.2112110>.
- [17] S.C. Flores, M.A. Sherman, C.M. Bruns, P. Eastman, R.B. Altman, Fast flexible modeling of RNA structure using internal coordinates, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8 (5) (2011) 1247–1257, <http://dx.doi.org/10.1109/TCBB.2010.104>.
- [18] M. Rother, K. Rother, T. Puton, J.M. Bujnicki, ModeRNA: a tool for comparative modeling of RNA 3D structure, *Nucleic Acids Res.* 39 (10) (2011) 4007–4022, <http://dx.doi.org/10.1093/nar/gkq1320>.
- [19] Y. Zhao, Y. Huang, Z. Gong, Y. Wang, J. Man, Y. Xiao, Automated and fast building of three-dimensional RNA structures, *Sci. Rep.* 2 (2012) 734, <http://dx.doi.org/10.1038/srep00734>.
- [20] M. Popena, M. Szachniuk, M. Antczak, K.J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R.W. Adamiak, Automated 3D structure composition for large RNAs, *Nucleic Acids Res.* 40 (14) (2012) e112, <http://dx.doi.org/10.1093/nar/gks339>.
- [21] C.Y. Cheng, F.C. Chou, W. Kladwang, S. Tian, P. Cordero, R. Das, Consistent global structures of complex RNA states through multidimensional chemical mapping, *Elife* 4 (2015), <http://dx.doi.org/10.7554/eLife.07600> e07600.
- [22] H.H. Gan, S. Pasquali, T. Schlick, Exploring the repertoire of RNA secondary motifs using graph theory: implications for RNA design, *Nucleic Acids Res.* 31 (11) (2003) 2926–2943.
- [23] M. Popena, M. Blazewicz, M. Szachniuk, R.W. Adamiak, RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures (Database issue), *Nucleic Acids Res.* 36 (2008) D386–D391, <http://dx.doi.org/10.1093/nar/gkm786>.
- [24] M. Popena, M. Szachniuk, M. Blazewicz, S. Wasik, E.K. Burke, J. Blazewicz, R. W. Adamiak, RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures, *BMC Bioinf.* 11 (2010) 231, <http://dx.doi.org/10.1186/1471-2105-11-231>.
- [25] P. Guntert, C. Mumenthaler, K. Wuthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol.* 273 (1) (1997) 283–298, <http://dx.doi.org/10.1006/jmbi.1997.1284>.
- [26] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson.* 160 (1) (2003) 65–73.
- [27] K.J. Purzycka, M. Popena, M. Szachniuk, M. Antczak, P. Lukasiak, J. Blazewicz, R.W. Adamiak, Automated 3D RNA structure prediction using the RNAComposer method for riboswitches, *Methods Enzymol.* 553 (2015) 3–34, <http://dx.doi.org/10.1016/bs.mie.2014.10.050>.
- [28] P. Schuster, W. Fontana, P.F. Stadler, I.L. Hofacker, From sequences to shapes and back: a case study in RNA secondary structures, *Proc. Biol. Sci.* 255 (1344) (1994) 279–284, <http://dx.doi.org/10.1098/rspb.1994.0040>.
- [29] J.S. Reuter, D.H. Mathews, RNAstructure: software for RNA secondary structure prediction and analysis, *BMC Bioinf.* 11 (2010) 129, <http://dx.doi.org/10.1186/1471-2105-11-129>.
- [30] C.B. Do, D.A. Woods, S. Batzoglou, CONTRAfold: RNA secondary structure prediction without physics-based models, *Bioinformatics* 22 (14) (2006) e90–e98, <http://dx.doi.org/10.1093/bioinformatics/btl246>.
- [31] S. Lusvardi, J. Sztuba-Solinska, K.J. Purzycka, G.T. Pauly, J.W. Rausch, S.F. Grice, The HIV-2 Rev-response element: determining secondary structure and defining folding intermediates, *Nucleic Acids Res.* 41 (13) (2013) 6637–6649, <http://dx.doi.org/10.1093/nar/gkt353>.
- [32] P.V. Sergiev, O.A. Dontsova, A.A. Bogdanov, Chemical methods for the structural study of the ribosome: Judgment day, *Mol. Biol.* 35 (4) (2001) 472–495, <http://dx.doi.org/10.1023/A:1010506522897>.
- [33] Q. Huang, K.J. Purzycka, S. Lusvardi, D. Li, S.F. Legrice, J.D. Boeke, Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription, *RNA* 19 (3) (2013) 320–332, <http://dx.doi.org/10.1261/rna.035535.112>.
- [34] S. Tian, P. Cordero, W. Kladwang, R. Das, High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states, *RNA* 20 (11) (2014) 1815–1826, <http://dx.doi.org/10.1261/rna.044321.114>.
- [35] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, NMR spectroscopy of RNA, *ChemBioChem* 4 (10) (2003) 936–962, <http://dx.doi.org/10.1002/cbic.200300700>.
- [36] O. Frolow, B. Endeward, O. Schiemann, T.F. Prisner, J.W. Engels, Nitroxide spin labeled RNA for long range distance measurements by EPR-PELDOR, *Nucleic Acids Symp. Ser. (Oxford)* 52 (2008) 153–154, <http://dx.doi.org/10.1093/nass/nnr078>.

- [37] A.K. Wozniak, S. Nottrott, E. Kuhn-Holsken, G.F. Schroder, H. Grubmuller, R. Luhrmann, C.A. Seidel, F. Oesterhelt, Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements, *RNA* 11 (10) (2005) 1545–1554, <http://dx.doi.org/10.1261/rna.2950605>.
- [38] B. Krahenbuhl, P. Lukavsky, G. Wider, Strategy for automated NMR resonance assignment of RNA: application to 48-nucleotide K10, *J. Biomol. NMR* (2014), <http://dx.doi.org/10.1007/s10858-014-9841-3>.
- [39] C.P. Jones, W.A. Cantara, E.D. Olson, K. Musier-Forsyth, Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry, *Proc. Natl. Acad. Sci. U.S.A.* 111 (9) (2014) 3395–3400, <http://dx.doi.org/10.1073/pnas.1319658111>.
- [40] K.A. Afonin, M. Viard, A.Y. Koyfman, A.N. Martins, W.K. Kasprzak, M. Panigaj, R. Desai, A. Santhanam, W.W. Grabow, L. Jaeger, E. Heldman, J. Reiser, W. Chiu, E. O. Freed, B.A. Shapiro, Multifunctional RNA nanoparticles, *Nano Lett.* 14 (10) (2014) 5662–5671, <http://dx.doi.org/10.1021/nl502385k>.
- [41] J. Sztuba-Solinska, T. Teramoto, J.W. Rausch, B.A. Shapiro, R. Padmanabhan, S.F. Le Grice, Structural complexity of Dengue virus untranslated regions: cis-acting RNA motifs and pseudoknot interactions modulating functionality of the viral genome, *Nucleic Acids Res.* 41 (9) (2013) 5075–5089, <http://dx.doi.org/10.1093/nar/gkt203>.
- [42] M. Antczak, T. Zok, M. Popenda, P. Lukasiak, R.W. Adamiak, J. Blazewicz, M. Szachniuk, RNApdbee—a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs, *Nucleic Acids Res.* 42 (Web Server issue) (2014) W368–W372, <http://dx.doi.org/10.1093/nar/gku330>.
- [43] P. Lukasiak, M. Antczak, T. Ratajczak, M. Szachniuk, M. Popenda, R.W. Adamiak, J. Blazewicz, RNAssess – A web server for quality assessment of RNA 3D structures, *Nucleic Acids Res.* 43 (W1) (2015) W502–W506, <http://dx.doi.org/10.1093/nar/gkv557>.
- [44] A. Rybarczyk, N. Szostak, M. Antczak, T. Zok, M. Popenda, R. Adamiak, J. Blazewicz, M. Szachniuk, New in silico approach to assessing RNA secondary structures with non-canonical base pairs, *BMC Bioinf.* 16 (2015) 276, <http://dx.doi.org/10.1186/s12859-015-0718-6>.
- [45] H.M. Zhang, Y. Su, S. Guo, J. Yuan, T. Lim, J. Liu, P. Guo, D. Yang, Targeted delivery of anti-coxsackievirus siRNAs using ligand-conjugated packaging RNAs, *Antiviral Res.* 83 (3) (2009) 307–316, <http://dx.doi.org/10.1016/j.antiviral.2009.07.005>.
- [46] Y. Shu, B. Seremi, P. Guo, Fabrication methods for RNA nanoparticle assembly based on bacteriophage phi29 packaging RNA structural features, in: P. Guo, F. Haque (Eds.), *RNA Nanotechnology and Therapeutics*, CRC Press, 2014. pp. 273–262.
- [47] A.A. Simpson, Y. Tao, P.G. Leiman, M.O. Badasso, Y. He, P.J. Jardine, N.H. Olson, M.C. Morais, S. Grimes, D.L. Anderson, T.S. Baker, M.G. Rossmann, Structure of the bacteriophage phi29 DNA packaging motor, *Nature* 408 (6813) (2000) 745–750, <http://dx.doi.org/10.1038/35047129>.
- [48] S.M. Dibrov, J. McLean, J. Parsons, T. Hermann, Self-assembling RNA square, *Proc. Natl. Acad. Sci. U.S.A.* 108 (16) (2011) 6405–6408, <http://dx.doi.org/10.1073/pnas.1017999108>.
- [49] S. Bailey, J. Wichtwechkarn, D. Johnson, B.E. Reilly, D.L. Anderson, J.W. Bodley, Phylogenetic analysis and secondary structure of the *Bacillus subtilis* bacteriophage RNA required for DNA packaging, *J. Biol. Chem.* 265 (36) (1990) 22365–22370.
- [50] S. Harris, S.J. Schroeder, Nuclear magnetic resonance structure of the prohead RNA E-loop hairpin, *Biochemistry* 49 (29) (2010) 5989–5997, <http://dx.doi.org/10.1021/bi100393r>.
- [51] F. Ding, C. Lu, W. Zhao, K.R. Rajashankar, D.L. Anderson, P.J. Jardine, S. Grimes, A. Ke, Structure and assembly of the essential RNA ring component of a viral DNA packaging motor, *Proc. Natl. Acad. Sci. U.S.A.* 108 (18) (2011) 7357–7362, <http://dx.doi.org/10.1073/pnas.1016690108>.
- [52] H. Zhang, J.A. Endrizzi, Y. Shu, F. Haque, C. Sauter, L.S. Shlyakhtenko, Y. Lyubchenko, P. Guo, Y.I. Chi, Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA, *RNA* 19 (9) (2013) 1226–1237, <http://dx.doi.org/10.1261/rna.037077.112>.
- [53] E. Bindewald, R. Hayes, Y.G. Yingling, W. Kasprzak, B.A. Shapiro, RNAjunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign (Database issue), *Nucleic Acids Res.* 36 (2008) D392–D397, <http://dx.doi.org/10.1093/nar/gkm842>.
- [54] K.A. Afonin, M. Viard, I. Kagiampakis, C.L. Case, M.A. Dobrovolskaia, J. Hofmann, A. Vrzak, M. Kireeva, W.K. Kasprzak, V.N. KewalRamani, B.A. Shapiro, Triggering of RNA interference with RNA-RNA, RNA-DNA, and DNA-RNA nanoparticles, *ACS Nano* 9 (1) (2015) 251–259, <http://dx.doi.org/10.1021/nn504508s>.
- [55] D. Jedrzejczyk, A. Chworos, Structural identification of the novel 3 way-junction motif, *DNA RNA Nanotechnol.* 2 (1) (2015), <http://dx.doi.org/10.1515/rnan-2015-0004>.

rRNA expansion segment 7 in eukaryotes: from Signature Fold to tentacles

Marcin Biesiada¹, Michael Y. Hu^{2,3}, Loren Dean Williams^{2,3,*}, Katarzyna J. Purzycka^{1,*} and Anton S. Petrov^{2,3,*}

¹Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan 61-704, Poland, ²Center for the Origins of Life, Georgia Institute of Technology, Atlanta, GA 30332, USA and ³School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received November 04, 2020; Revised September 13, 2022; Editorial Decision September 16, 2022

ABSTRACT

The ribosomal core is universally conserved across the tree of life. However, eukaryotic ribosomes contain diverse rRNA expansion segments (ESs) on their surfaces. Sites of ES insertions are predicted from sites of insertion of micro-ESs in archaea. Expansion segment 7 (ES7) is one of the most diverse regions of the ribosome, emanating from a short stem loop and ranging to over 750 nucleotides in mammals. We present secondary and full-atom 3D structures of ES7 from species spanning eukaryotic diversity. Our results are based on experimental 3D structures, the accretion model of ribosomal evolution, phylogenetic relationships, multiple sequence alignments, RNA folding algorithms and 3D modeling by RNAComposer. ES7 contains a distinct motif, the ‘ES7 Signature Fold’, which is generally invariant in 2D topology and 3D structure in all eukaryotic ribosomes. We establish a model in which ES7 developed over evolution through a series of elementary and recursive growth events. The data are sufficient to support an atomic-level accretion path for rRNA growth. The non-monophyletic distribution of some ES7 features across the phylogeny suggests acquisition via convergent processes. And finally, illustrating the power of our approach, we constructed the 2D and 3D structure of the entire LSU rRNA of *Mus musculus*.

INTRODUCTION

The ribosome, a massive assembly of rRNAs and ribosomal proteins (rProteins), lies at the heart of translation, synthesizing all coded protein. The cytosolic ribosome contains a ‘common core’ with universal structure in all organisms

(1–3). Common core rRNA is approximated by bacterial rRNA (4); around 90% of *Escherichia coli* rRNA is universal in secondary (2D) and three-dimensional (3D) structure in cytosolic ribosomes of extant organisms (2).

The ribosomal common core grew by accretion during six phases of evolution, culminating with the prokaryotic ribosome (5). The expansion of ribosomes then paused for several billion years, between the last universal common ancestor and the rise of eukarya.

Eukaryotic rRNAs are larger than prokaryotic rRNAs, and contain a diverse array of rRNA expansion segments [ESs, (6–9)] (Figure 1, <http://apollo.chemistry.gatech.edu/RibosomeGallery/index.html>) that emerge from a small number of sites on the surface of the common core, and are remote from functional centers, as shown by chemical probing and confirmed by structure determination (10–14). The acquisition of the eukaryotic shell, on the surface of the common core, is a seventh phase of ribosomal evolution (2,4). Protist and fungal ESs combine with eukaryotic-specific rProteins to form a secondary shell around the common core (2). The growth of ‘rRNA tentacles’, which are long double-helical elements observed in endothermic eukaryotes, marks an eighth phase of ribosomal evolution. In birds and mammals, tentacles extend for hundreds of Ångströms from the ribosomal surface (2,4) (Figure 1B). Variation of ES size and structure over evolution is evident by comparison of ribosomes of eukarya (15–18) and prokaryotes (19–22). An accretion mechanism of ESs growth over evolution was initially proposed by Michot (23).

Expansion segment 7 (ES7) is a nexus of ribosomal growth and diversity (1,4,6–9,24–26); it arose in the seventh phase of evolution and continued growth in the eighth phase. ES7 is the largest eukaryotic ES, with the most complex structure (Figure 2). ES7 tends to increase in size in concert with species complexity, and with overall size of the large ribosomal subunit (LSU), accounting for ~20% of LSU rRNA in mammals (2,5). Obligate parasites show

*To whom correspondence should be addressed. Tel: +1 404 894 8338; Email: anton.petrov@biology.gatech.edu
Correspondence may also be addressed to Loren Dean Williams. Tel: +1 404 385 6258; Email: loren.williams@chemistry.gatech.edu
Correspondence may also be addressed to Katarzyna J. Purzycka. Tel: +48 61 8 528503; Email: purzycka@ibch.poznan.pl

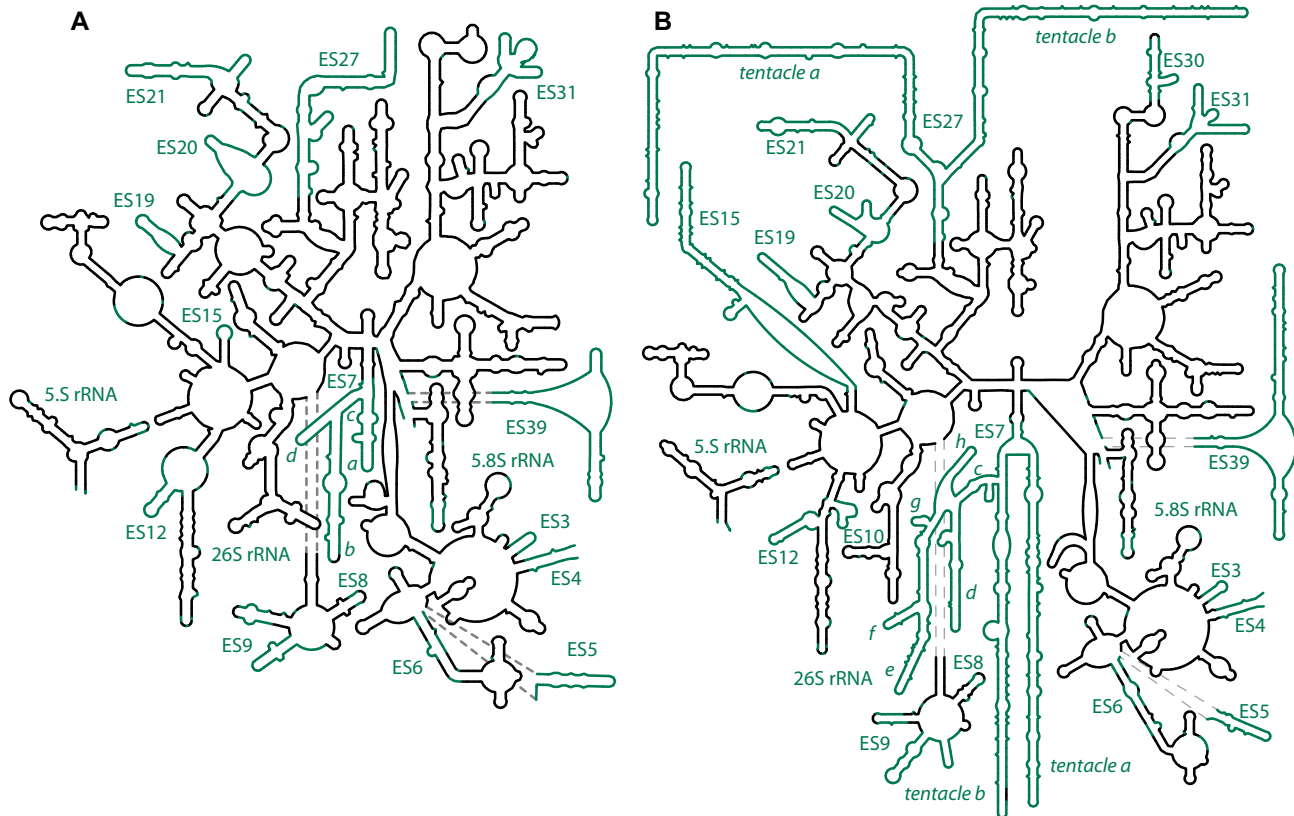


Figure 1. Secondary structures of two eukaryotic LSU rRNAs. (A) *S. cerevisiae*, and (B) *H. sapiens*. rRNA of ribosomal common core as defined in (2) is black. ESs are highlighted in teal and labeled according to Gerbi (8).

reduced rRNA expansion segments and genomes (27), and are not discussed in detail here.

In this report, we determine 2D and 3D structures of ES7 across phylogeny using a hybrid method that combines evolutionary mechanisms with RNA folding methods (Figure 3). We introduce the ES7 Signature Fold, which is conserved in 3D structure in all eukaryotes and is the foundation for broad diversity over phylogeny. We also improve previous ES7 structural and phylogenetic predictions (4,23,28–32). The structures here have been integrated into R2DT web-server (33) and can serve as templates for rRNA secondary structure for a broad variety of species.

MATERIALS AND METHODS

Modeling ES7: the approach

We have established a hybrid method to construct 2D and 3D models of ES7s of species broadly distributed across eukarya. The method introduces a novel combination of conventional 2D and 3D RNA folding algorithms with the accretion model of rRNA growth over evolution.

The accretion model stipulates that rRNA accumulates new elements onto pre-existing structure in the absence of remodeling (5), allowing identification of ancestral (structurally invariant) elements of ESs across all species.

We filtered the evolutionary information from a multiple sequence alignment (MSA, Supplementary Data Set 1)

through folding algorithms (35), and predicted an initial set of secondary structures (Figure 3). Addition of structural constraints from the accretion model substantially reduced the number of plausible structures. We used the resulting secondary structures as input for 3D structure prediction by RNAComposer (36) and applied the 3D models to discriminate between several secondary structural using the plausibility of the 3D models and lack of the steric clashes as selection criteria. The results allow us to describe ES7 in 2D and 3D for a variety of eukaryotic species (Supplementary Figure S1). The resulting secondary and 3D structures of ES7 are given in Supplementary Data Sets 2 and 3.

ES7 Signature Fold

We have identified the ES7 Signature Fold (Figure 4), a structurally conserved portion of ES7 seen in essentially all eukaryotic ribosomes. The ES7 Signature Fold contains H25 in addition to the archaeal expansion (ES7b, numbered here and after according to *Homo sapiens* schema, Figure 2, inset), plus two additional helices (ES7a and ES7c), and junctions that link them. The ES7 Signature Fold is conserved in 2D and 3D structure. Exceptions are ES7 from *Yarrowia lipolytica*, which lacks ES7c and ES7 of kinetoplastic parasites, which lack ES7a (Figure 2).

ES7 in protists generally consists of the ES7 Signature Fold and little else. Variability of ES7 across phylogeny is focused on animals, and arises from elongation of helices of

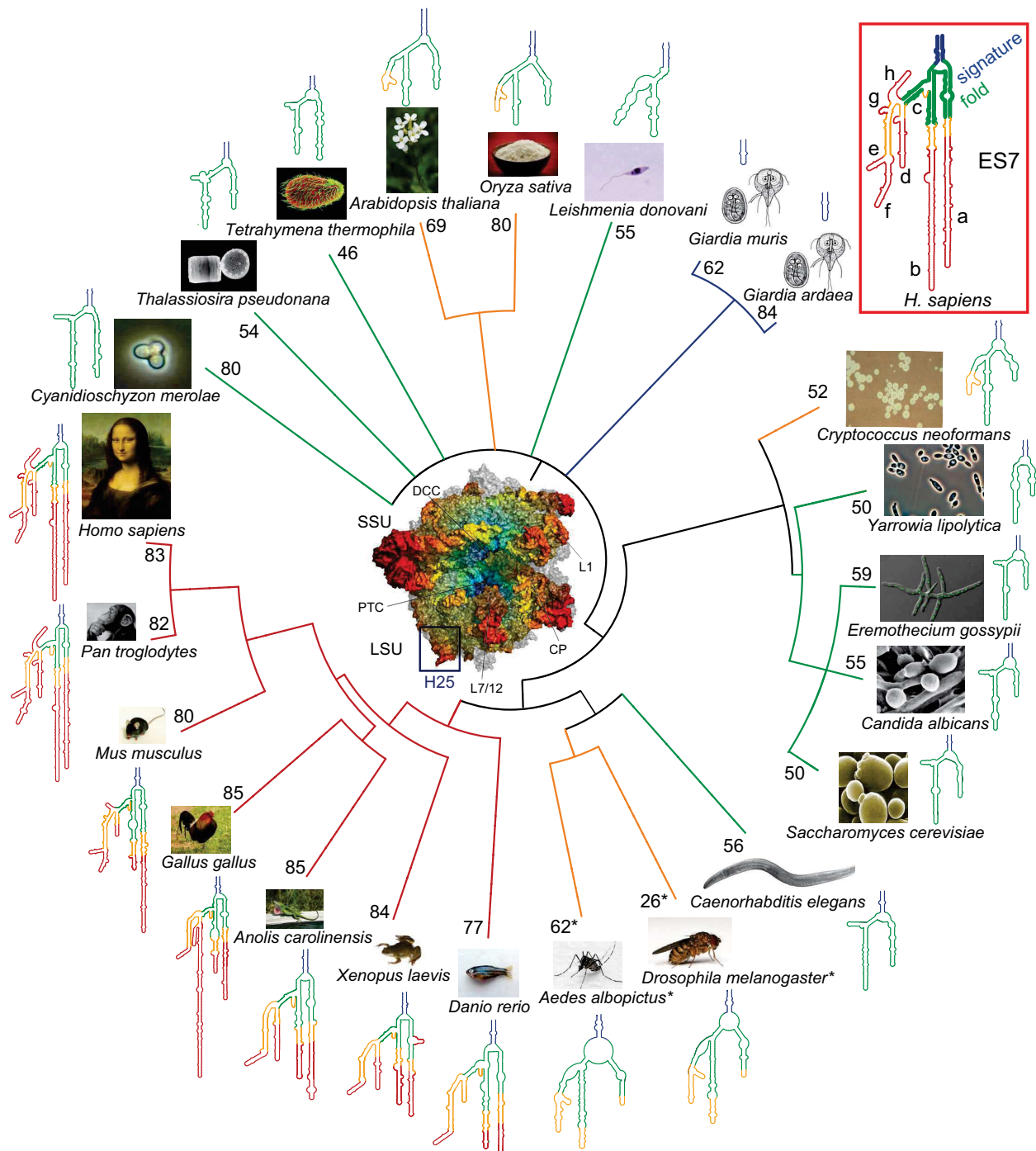


Figure 2. Secondary structures of ES7 mapped onto the canonical eukaryotic tree of life. Basal Helix 25 (boxed in the center) is part of the universal core; it is shown in blue in secondary structures as well as in the inset of *H. sapiens* ES7. Green indicates the ES7 Signature Fold, which is universal to rRNA of all eukaryotes. Yellow is universal to metazoans. Red is metazoan specific rRNA extensions, some of which reach extreme lengths in birds and mammals (tentacles). Colors of branches and internal nodes represent allocation of species (leaves) into Groups 0–3. The GC content (shown as percentage) for each ES7 is displayed next to each leaf of the tree. The asterisks indicate related species (*D. melanogaster* and *A. albopictus*) with a highly divergent GC content. Inset: the structure of ES7 of *H. sapiens* with the Signature Fold in bold green and blue. Helices labeled according to *H. sapiens* schema (4). All secondary structures in this document were visualized with the web portal RiboVision (34). The 3D structure of a prokaryotic ribosome with H25 highlighted in the blue box. In this image, colored by distance from the functional centers (peptidyl transferase center in the LSU and decoding center in the SSU) the ribosomal proteins are in gray.

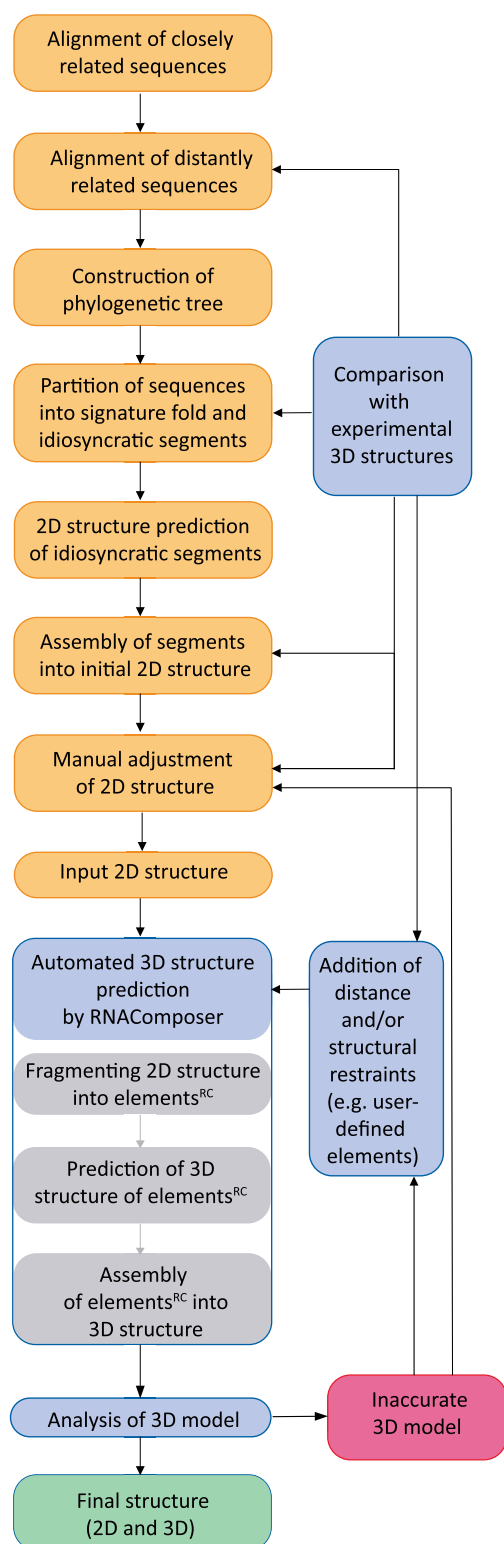


Figure 3. A schematic of the work-flow used to model 2D and 3D structures of eukaryotic rRNAs. RC indicates automated steps performed by RNAComposer.

the Signature Fold (ES7a–ES7c) and insertion of new helices (ES7d–ES7h, inset of Figure 2). The largest ES7s, in birds and mammals, contain the ES7 Signature Fold elaborated by linked arrays of junctions, helices, bulges and loops (Figure 2). The ES7 Signature Fold is absent only from a few obligate pathogenic eukaryotes (Figure 2: *Giardia*) that are also characterized by reduced genomes and ribosomes (37,38).

Modeling ES7: methodology

The ES7 Signature Fold is maintained over evolution as helices accrete and elongate. This critical finding enabled us to partition sections of ES7s of various species into sub-fragments and to independently predict sub-fragment secondary structures. We have integrated evolutionary and structural constraints into a workflow of RNA 2D and 3D structure prediction, surmounting limitations of the size, complexity and heterogeneity of ES7s. Our modeling is enabled in part by a) the process of accretion, as opposed to remodeling, as ES7 increased in size over evolution, and b) structural conservation of the ES7 Signature Fold. Specific modeling information is provided in the Supplementary Text.

The MSA was used to identify the Signature Fold in each ES7. The starting seed MSA was derived from the structural superimposition and contained the regions within the Signature Fold and enriched by additional sequences using mafft (39). The Signature Fold-annotated MSA (Supplementary DataSet S4) allowed us to pinpoint variable helical stem loops. By combining the information from the MSA with known 2D structures, 2D ES7 models of closely related species were predicted. Finally, secondary structures of ES7s of more distantly related species were predicted by bootstrapping the available information from the MSA. Predicted 2D and 3D structures were constrained by the conserved Signature Fold and variable helical stems. Secondary structures of individual helical elements were predicted by mfold (35) and merged with the ES7 Signature Fold to form the complete ES7 secondary structure. Variable stem loop regions of the MSA were manually adjusted when necessary, based on the 2D predictions.

Three-dimensional models were predicted by inputting sequences and secondary structures into RNAComposer (36) (Figure 3). Experimental 3D structures of ES7 are known for several eukaryotic ribosomes (4,16,17,40). We leveraged these experimental ribosomal structures and mined a library of RNA structural elements (36,41,42). In instances, when RNAComposer's dictionary lacked required elements (within the Signature Fold or helical stems, Supplementary Figure S2), the missing segments were modeled using standard protocols of comparative RNA 3D structure prediction based on 3D RNA templates (43,44), subsequently introduced into the prediction protocol as user-defined elements (45), and further refined by RNAComposer.

RNAComposer uses secondary structural topology over the sequence homology to assemble small 3D elements into a 3D model (see Supplementary Text and Supplementary Figure S2). Application of the hybrid methodology for the prediction of 2D structures limited the number of plausi-

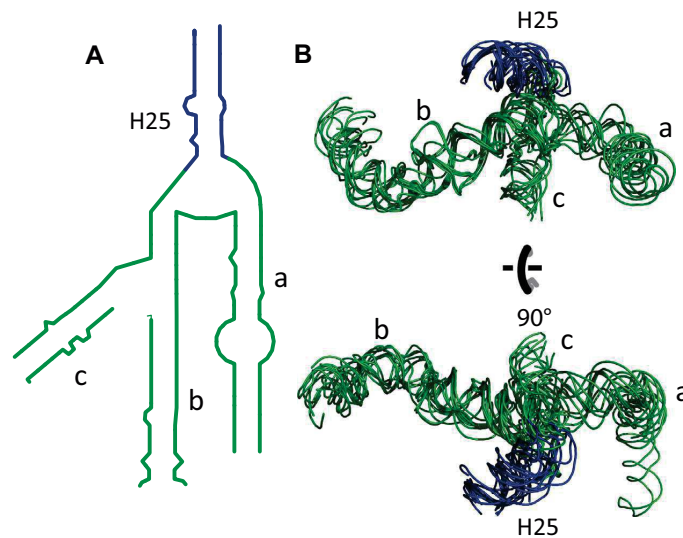


Figure 4. The ES7 Signature Fold. (A) Secondary structure. (B) Three-dimensional structure. The coloring scheme is the same as in Figure 2. In panel (B), all experimental and modeled structures from Table 1 are superimposed.

ble topologies and eliminated many 3D structures of unfavorable steric interactions between the structural domains. Multiple models were predicted for each structure studied.

Final models were selected based on (i) lowest energy, (ii) preservation of relative orientations of helices at junctions (46) and (iii) consistency with the evolutionary model (the underlying core tends to be preserved in all ES7 structures). The use of RNAComposer for 3D structure prediction allowed to overcome the problem of computational cost for larger RNA molecules, demanding addition of atomic details to coarse-grain models or prediction dependent on structural templates. Furthermore, the application of the hybrid approach eliminates the problem of homologs for RNAs that differ vastly in sequence or contain unique expansions which are not found within the closest homologs. Therefore, our approach allows us to overcome challenges to structure prediction seen in other methods (43,44,47–52).

Validation of the approach

To validate the hybrid approach, we performed two tests. First, using ES7 of *Saccharomyces cerevisiae* as a template, we predicted the structure of ES7 of *Candida albicans* and experimentally validated it by a chemical probing using Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) (53). The structure predicted by the hybrid approach is in good agreement with the experimental SHAPE data (54) and significantly improved the model of the *C. albicans* ES7 structure compared to a previous effort (55) (Supplementary Figure S3). Second, we performed a reciprocal prediction, by modeling the known structure of ES7 of *S. cerevisiae* based on ES7 of another fungus (*Eremothecium gossypii*) and comparing the reverse prediction with experimentally determined structures (PDB entry 4V6I, 5JUO, 4U4R) (15,56,57) (Supplementary Figure S4). The accuracy of the *S. cerevisiae* ES7 model was high as determined by comparison with ES7 of the experimentally determined structure (PDB entry 4V6I), result-

ing in Matthews correlation coefficient = 83.8%, sensitivity = 98.5%, specificity = 90.3% (58). Additional details of this validation of the hybrid approach are reported in the Supplementary Text.

Allocation of ES7s to Groups based on topology and length of variable expansions

To systematize ES7 modeling, ES7s from 23 species were allocated into groups based on size and complexity (Figure 2, Table 1). Group 0 represents ES7s in species with H25 alone. Groups 1–3 (Figure 2) represent ES7s with various extents of recursive accretion onto the ES7 Signature Fold, with Group 1 ES7 < Group 2 ES7 < Group 3 ES7. Representatives from each group are compiled in Table 1. Auxiliary MSAs for ES7s within each group are compiled in Supplementary Data Set 4.

Group 0 ES7s are simple stem loops, seen in obligate pathogens with reduced genomes and ribosomes similar to those of bacteria.

Group 1 ES7s are up to 250 nts in length and are composed essentially of the Signature Fold. *Y. lipolytica* is an exception that lacks ES7c (Figure 2). Parasitic kinetoplasts (*Leishmania donovani*, *Tripanosoma brucei* and *Tripanosoma cruzi*) lack ES7a and contain elongated ES7c (18,59).

Group 2 ES7s, seen in invertebrate metazoans, plants and some fungi, are larger and more complex than Group 1 ES7s (Figure 2). ES7s of Group 2 contain the ES7 Signature Fold, with extensions of helices ES7a–c and the addition of ES7d–e, which branch off of ES7c.

Group 3 ES7s are found in chordates (fish, amphibians, reptiles, mammals). Each Group 3 ES7 contains a subset of five core helices (ES7a–e, Figure 2, inset). Additional growth (ES7a–b; d–e) and further branching (ES7f–h) of ES7 is observed in endothermic vertebrates (birds and mammals), which contain GC rich tentacles. In these species, helices of ES7 break through the rRNA-protein

Table 1. ES7 groups

Group 0	nts	GC%	Group 1	nts	GC%	Group 2	nts	CG%	Group 3	nts	GC%
<i>Giardia ardeae</i>	31	82	<i>Caenorhabditis elegans</i>	232	56	<i>Aedes albopictus</i> ^a	352	62	<i>Anolis carolinensis</i>	521	85
<i>Giardia intestinalis</i>	8	100	<i>Candida albicans</i> ^a	209	55	<i>Arabidopsis thaliana</i>	224	69	<i>Danio rerio</i>	473	77
<i>Giardia muris</i> ^a	29	62	<i>Cyanidioschyzon merolae</i>	255	80	<i>Cryptococcus neoformans</i>	235	52	<i>Gallus gallus</i> ^a	799	85
			<i>Eremothecium gossypii</i>	212	59	<i>Drosophila melanogaster</i> ^b	340	26	<i>Homo sapiens</i> ^b	876	83
			<i>Saccharomyces cerevisiae</i> ^b	210	50	<i>Oryza sativa</i>	214	80	<i>Mus musculus</i> ^a	690	80
			<i>Tetrahymena thermophila</i> ^b	220	46				<i>Pan troglodytes</i>	866	82
			<i>Thalassiosira pseudonana</i>	212	54				<i>Xenopus laevis</i>	510	84
			<i>Yarrowia lipolytica</i>	161	50						
			<i>Leishmania donovani</i>	200	55						

^aSpecies whose ES7 3D structures were predicted here by RNAComposer (bold).

^bSpecies whose ES7 3D structures were determined experimentally by others (4,16,17).

core and extend out from the characteristic protist shell of eukaryotic ribosomes (1,4).

RESULTS

ESs are highly diverse across phylogeny and can be dynamic and disordered *in vivo*. Modeling of ESs presents challenges at the levels of both 2D (30) and 3D structure (4). Significant portions of ESs are omitted from most experimental structures. We have predicted 2D structures of ES7s of 23 species in Groups 0–3 from across the phylogenetic tree, and modelled 3D structures for representative species from each group (Movie 1, Supplementary Table S1, Supplementary Data Set 3). Group 0 contains ES7s of obligate parasite species with reduced rRNAs. Mapping of ES7 structures onto the eukaryotic phylogenetic tree (Figure 2) reveals non-monophyletic distribution for Groups 1–2, while ES7 of Group 3 represents a monophyletic group of vertebrate species.

ES7 predictions

ES7 structures modeled in the current study differ in size, branching topologies and sequence compositions. These differences posed a number of modeling challenges that have been independently addressed in the process of predicting the ES7 structures of species from each group.

Sequences from Group 1 were very similar to each other in length and in 2D and 3D structure. ES7 in Group 1 is essentially composed of the Signature Fold. Due to the similarity of size and secondary structure, identification of the individual helical regions within the MSA and overall 2D and 3D modelling was straightforward.

The ES7s of Group 2 have accreted beyond the Signature Fold. Most of the sequences from Group 2 have high GC content, ranging from 60% to 80% (Figure 2) and are represented within multiple phylogenetic groups, revealing a substantial sequence variability. *Drosophila melanogaster*, the conventional biological model whose 3D structure has

been determined (4), is anomalous in ES7 GC composition, and is the only metazoan we have discovered thus far with low GC content of ES7 (26%, Supplementary Figure S5). In the current study, we used the Signature Fold of *D. melanogaster* as a guide to model the 2D structure of ES7 for another insect, *A. albopictus*, whose ES7 contains a representative GC content for Group 2 (Supplementary Figure S6, Supplementary Data Set 5). This modeled ES7 structure was further used as a template to model 2D ES7s of other species from Group 2 that are similar in sequence and its GC content (Supplementary Data Sets 2 and 3).

Group 3 contains ES7s of endothermic vertebrates, which are characterized by large size and high complexity, with further diversity in helical branching. The tentacle-like regions are hypervariable in length and branching topology (Figure 2; compare *Xenopus laevis*, *Gallus gallus*, *Mus musculus* and *Homo sapiens*) among the members of the group, representing the main modeling challenge. The conservation of the ES7 Signature Fold in the 3D structure of the *H. sapiens* ribosome and the high sequence similarity among Chordates in this core region (Supplementary Figure S7) allowed us to model these ES7s using our hybrid approach in spite of the complexity. Here, we illustrate the general features of the modeling approach by presenting the specific method used to determine the structure ES7 of *G. gallus* (Supplementary Text). This ES7 contains a unique extension segment ES7d1 located in between ES7c and ES7d. We further describe the 2D and 3D models of the entire LSU rRNA for *M. musculus* and compare it with the previous predictions in the context of the experimental structures available for *H. sapiens*.

Comparative analysis of ES7s from Groups 1–3

Validation of the 2D models of ES7 (Figure 2) was accomplished by computing sequence conservation and covariation as well as by quantitative comparison of predicted and experimental 3D structures. Conservation and co-variation analyses were performed independently for ES7s of each

Group 1–3 from MSAs in Supplementary Data Set 4. Structural comparisons were performed for species whose ES7 structures were modeled here or were previously determined, as specified in Table 1.

Conservation and co-variation scores. To estimate the accuracy of 2D structures we computed conservation and co-variation scores from the ES7 MSA (see Supplementary Materials and Methods). The conservation score used here is a gap adjusted Shannon entropy (GASE), which is a measure of conservation of nucleotide identity (2). The co-variation score PASE (pair adjusted Shannon entropy) characterizes paired nucleotides by their adherence to co-variation. Thus, a nucleotide position is considered conserved if it is always canonically paired, even if its nucleotide identity is not conserved. The computed scores (ranging from 0 to 2, where 0 represents universally conserved) were mapped onto ES7 representatives for Group 1 (*S. cerevisiae*), Group 2 (*D. melanogaster*) and Group 3 (*H. sapiens*), (Figure 5; Supplementary Figures S8–S10).

The structure of the Signature Fold is highly conserved. Base pairing within the Signature Fold is reasonably predicted in all groups despite variability of sequence. Conservation/co-variation metrics of ES7s are highest within the Signature Fold, and the covariation score is substantially higher than the conservation score (Figure 5). The base pair Shannon Entropy (BPSE, Figures S8B, S9B and S10B) and combined PASE score (Figures 5B, D and F) reveal a moderate to high degree of the base pairing conservation (regardless of the specific identity of the base pairs) within the Signature Fold, thus providing the relationship between the structural data and the sequence information within the MSAs for this nearly universal region of the ES7 Helical junctions in Groups 1 and 2 tend to be less conserved than the base paired regions. ES7s of group 1 exhibit moderate conservation and covariation despite their non-monophyletic origins.

Helical stems (ES7a–d) from Group 2, outside the Signature Fold, exhibit little conservation or co-variation, reflective of the polyphyletic nature of this group (Figure 5C and D). An abundance of irregularities of helical stems by bulges or internal loops pose challenges for predicting these elements by a conventional co-variational approach for remotely related species, as the irregularities vary between the species and complementary strands cannot be easily identified from the MSAs. We note that the co-variational method has been successfully applied to a subset of closely related species from Group 2 (e.g. within Hymenoptera) (60) due to high structural similarity of the extended regions.

Conservation/co-variation metrics for the respective helical stems of Group 3 (ES7a–d, Figure 5E and F) are higher for than those for Group 2. The Group 3 specific stems (ES7e–h) also reveal moderate to high conservation/covariation (if present). Overall, spatial alignment of the 3D elements of Group 3 is outstanding (Supplementary Data Set 3), especially in a subgroup of species that represents endothermal vertebrates. A majority of tentacles exhibit substantial sequence conservation (especially in mammalian species). Tentacle ES7b is highly variable in structure and ES7d is the least conserved in sequence. Due to high sequence conservation, co-variational analysis has

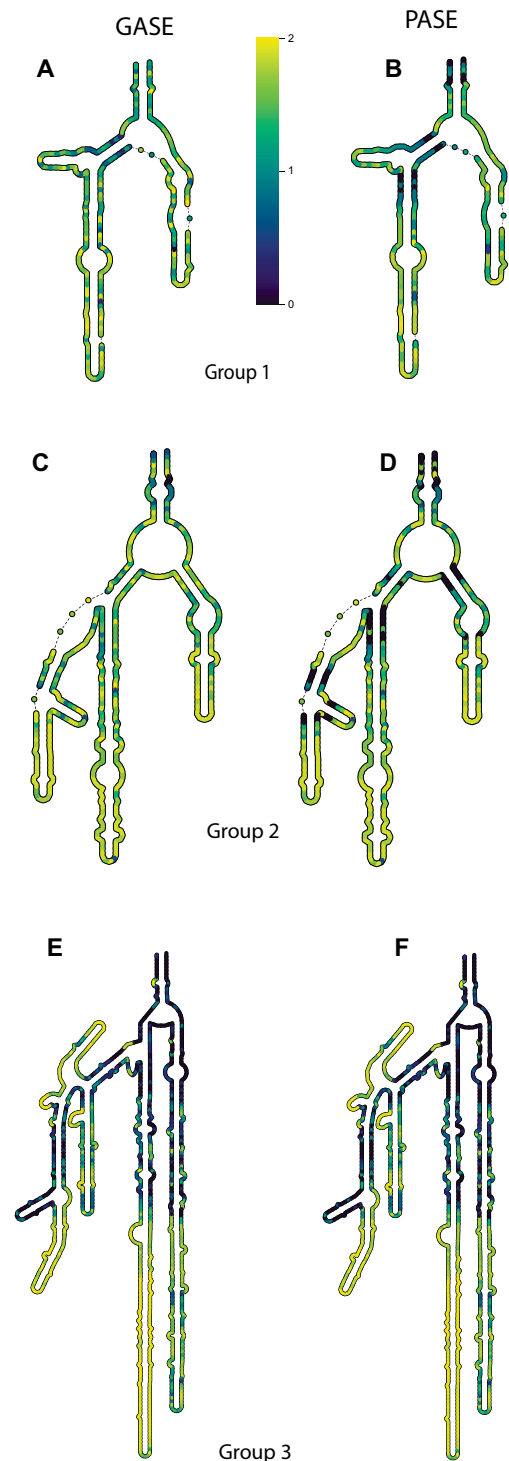


Figure 5. Conservation (GASE) and co-variation (PASE) scores mapped onto the secondary structures of ES7 from Groups 1–3. (A) GASE and (B) PASE computed from ES7 MSA of Group 1 and mapped onto ES7 of *S. cerevisiae*; (C) GASE and (D) PASE computed from ES7 MSA of Group 2 and mapped onto ES7 of *D. melanogaster*; (E) GASE and (F) PASE computed from ES7 MSA of Group 3 and mapped onto ES7 of *H. sapiens*. Both scores range from 0 (dark blue, absolute conservation or co-variation) to 2 (yellow, random signal, no conservation or covariation). Intermediate values are indicated by the color bar. The ES7 MSAs of Groups 1–3 are given in Supplementary Data Set 4.

Table 2. Structural analysis of experimental and modeled structures by RMSD metrics

Modeled/experimental structures	RMSD, Å	Experimental/experimental structures	RMSD, Å
<i>C. albicans</i> (1) ^a / <i>T. thermophila</i> (1)	6.713	<i>T. thermophila</i> (1)/ <i>D. melanogaster</i> (2)	5.794
<i>A. albopictus</i> (2)/ <i>D. melanogaster</i> (2)	4.934	<i>T. thermophila</i> (1)/ <i>H. sapiens</i> (3)	5.700
<i>M. musculus</i> (3)/ <i>H. sapiens</i> (3)	4.224	<i>D. melanogaster</i> (2)/ <i>H. sapiens</i> (3)	6.831

^aGroup numbers as defined in Table 1 are shown in parentheses.

limited power for species in Group 3, as co-variation signals cannot be detected.

Structural analysis by RMSD metrics and long-range interactions. Comparison of the root-mean square deviations (RMSDs) of atomic positions confirms that the structure of the ES7 Signature Fold is conserved among all species; structural differences between predicted and experimental structures (RMSD = 4–7 Å) are comparable to differences between experimental structures (RMSD = 5–7 Å), Table 2, Supplementary Table 2 and Supplementary Figure S11.

The RMSDs between experimental and modeled structures within each group reveal important patterns:

- The Signature Fold is structurally conserved (Figure 4).
- Junctions within the Signature Fold for Groups 1 and 2 vary in sequence and reveal subtle variations in 2D structure; they may substantially differ within a group (e.g. *T. thermophila* versus *C. albicans*, Group 1; *D. melanogaster* versus *A. albopictus*; Group 2); 3D junctions of species within Group 3 tend to be highly conserved (*H. sapiens* versus *M. musculus*).
- In Group 2, helical stems of ES7 that link to the Signature Fold vary substantially in their structures because of irregular local elements. Each RNA helix (7a, 7b or 7c) is characterized by an idiosyncratic set of irregularities, making them quite different from the helical elements within the common core, where local motifs within the helices are generally well-defined (within a given taxonomic domain).
- Structural deviation between helical stems of Group 2 is greater than between those of Group 3. Species allocated to Group 2 are not monophyletic, and different local patterns likely reflect convergent evolution. Conservation within the helical elements of the Signature Fold varies between the groups, with Group 3 being most conserved.

More similar structures tend to preserve long range interactions in 3D. Base-phosphate and base-sugar contacts have been extracted from the modeled and experimental 3D structures of ES7 for Groups 1–3 and mapped onto the 2D structure (Supplementary Figures S12 and S13). These long-range interactions tend to be somewhat similar for the structures that belong to the same group and vary between the groups. Thus, despite the overall high structural conservation of the ES7 Signature Fold, the specific features (base

pairing pattern, junction architecture, specific long-range interactions) are not fully preserved.

Patterns and exceptions

We observed informative patterns within the ES7 Groups, which show group-specific personalities. Comparison of sequence and structure revealed that across Groups 1–3 (Figure 5, Table 2), the size and complexity of ES7 is interdependent with sequence conservation, co-variation, heterogeneity of the local motifs within helical elements, and structural variability of the junctions and extended regions.

Group 1 ES7s are moderately conserved in sequence and highly conserved in structure. Group 2 ES7s are diverse in sequence with modest diversity in structure. Group 3 ES7s are conserved in both sequence and structure, in common regions. Group 3 ES7s show extensive variation in morphology and length of tentacles, especially in mammals, suggesting that the rate of accretion is high relative to the rate of change of sequence. It was instructive to model ES7 of each group independently, accounting for their unique sets of features.

rRNA growth mechanisms

The acquisition of the eukaryotic shell, on the surface of the common core, is a seventh phase of ribosomal evolution (2,4). This phase of rRNA evolution produced ES7 Groups 1 and 2. The growth of tentacles, which are long double-helical protrusions seen in endothermic eukaryotes, marks an eighth phase of ribosomal evolution. This phase of rRNA evolution produced ES7 Group 3.

The accretion model is consistent with the following constraints on ES7: (i) the core of ES7, the Signature Fold, is conserved in 2D and 3D structure and is universal to eukaryotes, (ii) new rRNA is added onto pre-existing ES7 rRNA by accretion processes such that addition of new rRNA fragments does not remodel pre-existing ES7 rRNA, (iii) 2D and 3D structures of ES7 in common between daughter species are ancestral and (iv) 3D structure of ES7 is more conserved than sequence (5).

The diversity of ribosomal structures and their known phylogenetic relationships have allowed construction of ES7 accretion pathways. We assume that rRNA ancestral states can be reasonably modeled by elements that are shared among all daughter species. This assumption suggests that ES7s of simple extant species reasonably represent those of extinct ancestors. We have ranked extant ES7s by size and complexity.

The *H. sapiens* accretion pathway (*hs*-accretion pathway) of ES7 describes a process that initiates with H25 of the Last Universal Common Ancestor (LUCA) (61) and terminates with ES7 of extant *H. sapiens* (Figure 6). Our best estimate of ES7 at LUCA is H25 of bacterial ribosomes (*E. coli* is used here as a prototype). Helix 25 is universal to all cytosolic ribosomes. The first step from LUCA in the *hs*-accretion pathway is inferred from archaeal ribosomes where H25 contains a micro-ES (μ -ES). μ -ES's of 5–30 nucleotides are commonly observed in archaeal ribosomes at sites of much larger eukaryotic ESs (Figure 6). An insertion fingerprint in H25 of archaeal ribosomes at the base of the μ -ES marks

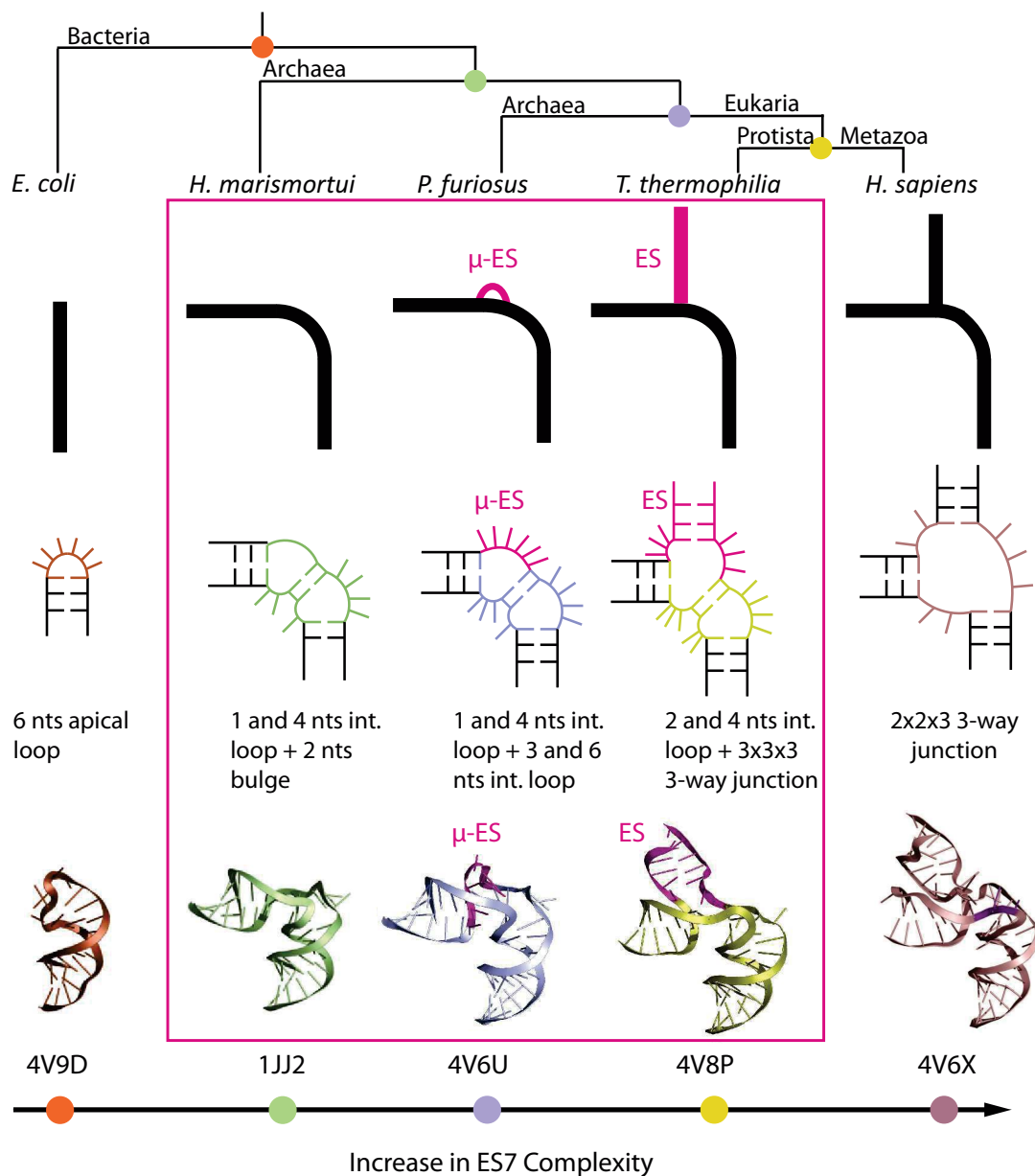


Figure 6. RNA accretion in ES7, based on comparison of experimental 3D structures and their evolutionary relationships. The lineage leading to *H. sapiens* (Metazoan) is highlighted. The rRNA at each ancestral node, highlighted by colored circles, is approximated by rRNA that is common to daughters (2). rRNA is depicted schematically, in secondary structures, and in three dimensional structures. A stem loop of H25 (*E. coli*) establishes the base of ES7. An expansion of the apical loop extends H25 in *H. marismortui*; a bulge expands into an internal loop in *P. furiosus*, which extends into a new expansion ES7a in *T. thermophila*; this element undergoes further adjustments in *H. sapiens*. PDB IDs of the source structures are indicated below the 3D structures. Chain IDs and nucleotide numbers are given in Supplementary Table S3. The 3D structures are compiled in Movie 2.

the site of archaeal expansion of the H25 stem. Additional structures of ES7 over the eukaryotic tree of life support an *hs*-accretion pathway in which further extension of this internal loop leads to a three-way junction (ES7a/ES7b, *T. thermophila*). This three-way junction is further elaborated by tertiary interactions between the branched stem and the underlying trunk (*H. sapiens*; Figure 6; Movie 2).

The *hs*-accretion pathway can be expanded to a fine-grained progression (Figure 7), of elementary steps, by interleaving structures of the phylogenetic accretion process

with transitional structures extracted from the RNA structural database (41,42). Elementary steps of RNA growth include the accretion of one (or several) nucleotides, yielding a bulge. The bulge is expanded by insertion of additional nucleotides. Repetition of nucleotide accretion into the same bulge region results in more complex bulges (if insertion occurs in the same RNA strand) or internal loops (if the insertion appears in the opposing strand). Finally, accretion of a critical number of nucleotides (7–9) results in extrusion of a new helix and formation of a three-way junction.

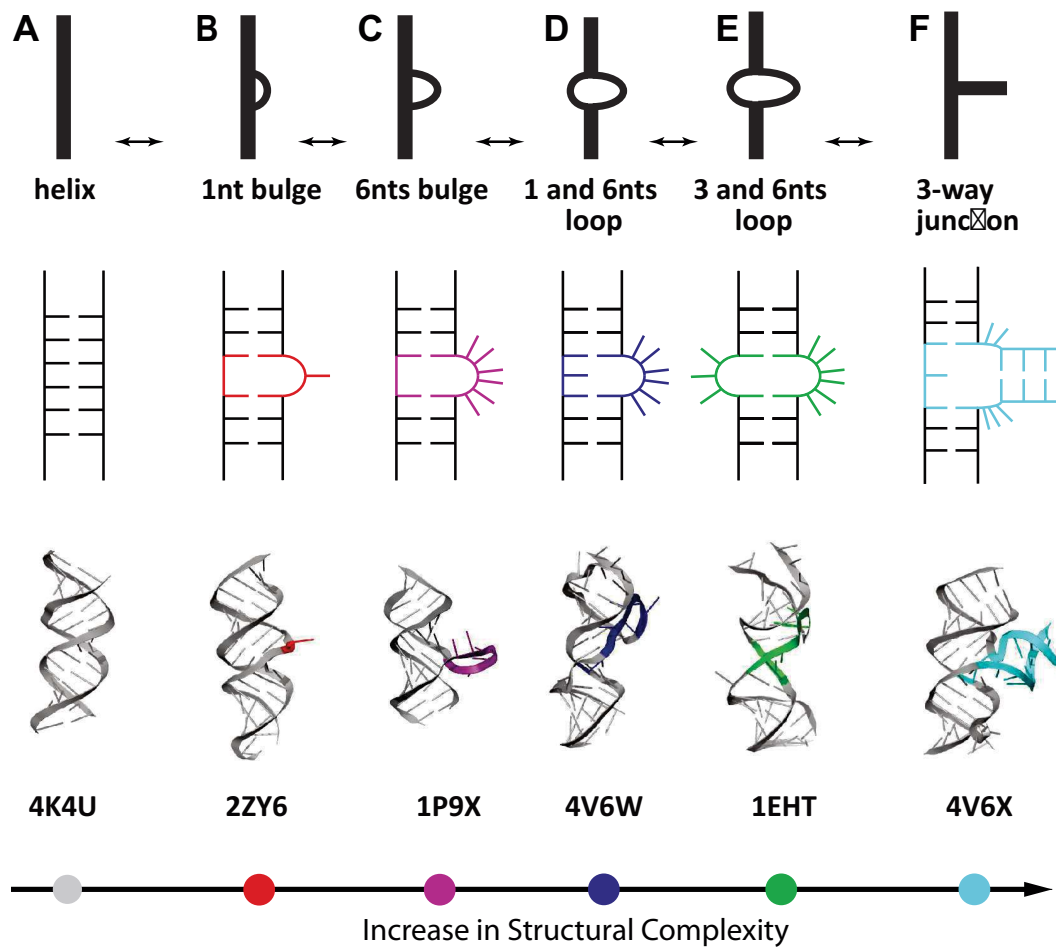


Figure 7. A three-way junction emerges from a helix through a series of elementary steps of rRNA accretion. Existing and emerging elements are depicted by their secondary and 3D structures and by schematic representations. (A) Double helical region; (B) 1 nt bulge; (C) 6 nt bulge; (D) 1 and 6 nt loop; (E) 3 and 6 nt loop; (F) three-way junction. This trajectory is based on experimental 3D structures that are ordered by size. The evolutionary relationship is inferred. PDB IDs of the source structures are indicated below the structures. Chain IDs and nucleotide numbers are given in Supplementary Table S4.

This process converts one RNA topology to another in a series of elementary steps that maintain the underlying structure. The construction of this fine-grained molecular progression during rRNA evolution is analogous to structure-based descriptions of evolutionary processes (62,63) and macromolecular dynamics (64–66) that are inspired by the concept of reaction coordinates for chemical transformation (67).

LSU of *M. musculus*

To illustrate the utility and generality of our approach, we modeled the 2D and 3D structures of the LSU rRNA of *M. musculus*. This rRNA contains large eukaryotic ESs that contribute to a total LSU rRNA length of 4887 nts. 2D structures of all ESs were modeled by methods analogous to those described above for ES7, using the LSU rRNA of *H. sapiens* (68), which has been experimentally determined by Cryo-EM (4), as a guide.

The 2D model of *M. musculus* LSU rRNA presented here differs from the previous models (23,28). The differences are discussed in the SI text and Supplementary Figure

S14. Comparison of the 2D structures of *M. musculus* LSU rRNA (Figure 8) and *H. sapiens* (Figure 1B) demonstrates their similarity. Yet, some helices of *M. musculus* rRNA are shorter than in *H. sapiens* (e.g. ES7a,b,d,h, ES15, ES27a, ES27b), while some are absent (ES7g, ES15a, ES30a). The result lends support to the accretion model and is consistent with a rapid growth rate of rRNA in mammalian lineages (2). We anticipate that the *M. musculus* LSU rRNA 2D and 3D structures will be useful for explaining and predicting a variety of data.

ES7 of *M. musculus*. *M. musculus* LSU rRNA contains one of the largest known ES7s (691 nts). 3D modeling produced a *M. musculus* ES7 structure nearly superimposable on that of *H. sapiens* (Movie 1, Supplementary Data Set 3). A comparison of the tentative 2D structure with the final 2D structure extracted from a 3D model reveals minor peculiar discrepancies discussed in the Supplementary Text and shown in Supplementary Figure S15. The tips of expansion segments ES7a,b in both *H. sapiens* and *M. musculus* are enriched in G and C and polarized in their distribution, so that one strand is G rich, while the other is C rich. Such polar-

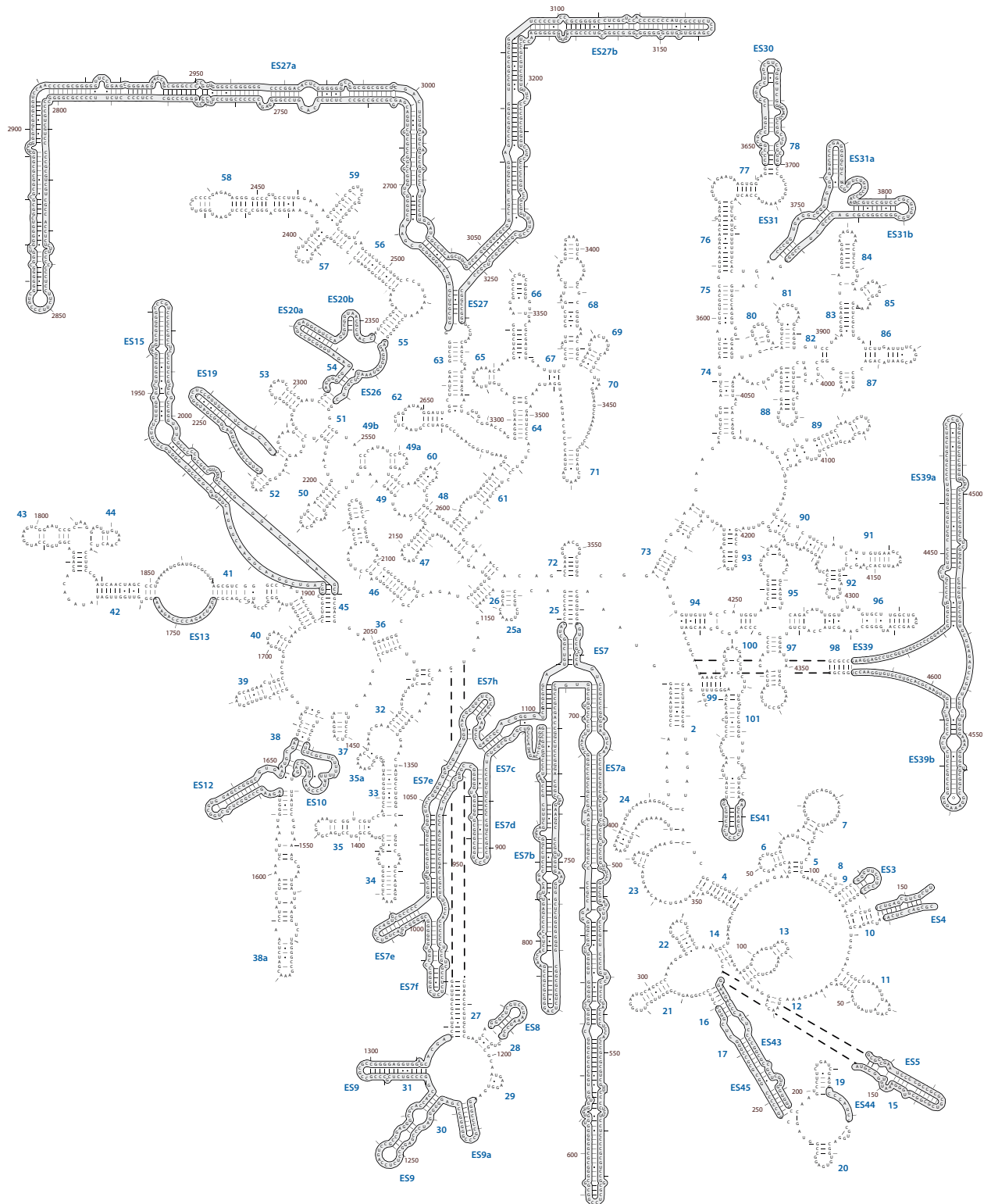


Figure 8. Secondary structure of the LSU rRNA of *M. musculus*. Expansion segments are outlined.

ization may promote the folding of the long expansions and may accelerate the formation of G-quadruplexes (25,26).

Complete 3D model of the LSU rRNA of *M. musculus*. To further illustrate the utility and universality of our approach, we modeled the 3D structure of the *M. musculus* LSU rRNA, using the 2D structure as a basis. The resulting *M. musculus* rRNA model is nearly superimposable on that of *H. sapiens* (PDB ID 4V6X), revealing an overall good agreement (RMSD = 4.020 Å) of the common core regions and several tentacle-like expansion segments (ES7a, ES27a,b, ES39a). Yet, some of the tentacles were found to have different directionalities in the two structures. Examples of such tentacles include ES7b,e,f. These differences could arise from the lack of interactions with proteins, variability in the local base pairing, high mobilities and polymorphism of the tentacles themselves, or limitations of our modelling approach.

ES7 forms pseudoknots with ES15 and ES39. In rRNA from *H. sapiens* (PDB ID 4V6X), the first pseudoknot is formed by C1289 from ES7b and G4941 and C4942 from ES39 (Supplementary Figure S16). In the modeled structure of the LSU rRNA from *M. musculus*, the nucleotides equivalent to those in the ES7/ES39 pseudoknot are found to be in proximity (~7 Å) (Supplementary Figure S16). The structural analysis of ribosomal complexes available in Ribovision reveals that the ES7/ES39 pseudoknot exists in eukaryotic species that represent major phylogenetic groups (mammalians, insects, fungi, plants, SAR).

The second pseudoknot is formed by base pairing between G963 from ES7 and C2250 from ES15. This pseudoknot has been observed in the ribosomes of mammalian species (4) and some kinetoplastids (18).

The corresponding nucleotide region in ES15 of *M. musculus* is significantly shorter than that of *H. sapiens* (Supplementary Figure S17). However, our prediction locates corresponding regions of ES15 and ES7 in proximity to each other (~3 Å).

Both ES7/ES39 and ES7/ES15 pseudoknots are also stabilized by eukaryotic specific protein eL6 (Supplementary Figure S18), whose globular domain binds to ES39 and the intrinsically disordered tail wraps around ES7 and ES15 (in mammalian species) providing the tight structural integration of the eukaryotic specific motif ES7/ES15/ES39/eL6. Previous analyses have suggested that the components of this region of eukaryotic LSU have co-evolved (4,59). While formation of intact pseudoknots in the absence of rProteins is not expected in our models without introduction of specific constraints (69) (see Discussion below), the relative proximity of the corresponding fragments illustrates the power, universality, and robustness of the presented hybrid approach for ribosomal systems of any size.

DISCUSSION

The ES7 Signature Fold

ES7 is largest and the most complex ES and contains the most diverse rRNA over phylogeny. Yet, with the exception of ribosomes from some parasites, the ES7 Signature Fold (Figure 4) is highly conserved in 2D and 3D structures in all eukaryotes. The ES7 Signature Fold provides a foundation

for rRNA elaboration and is a platform for helical elongation and branching. ES7 of species in Group 1 (Table 1) is essentially comprised of the Signature Fold.

Despite overall conservation, the ES7 Signature Fold varies over phylogeny by local differences in secondary structure (single nucleotide insertions, base pair disruptions) and long range interactions (base-sugar and base-phosphate). ES7 Signature Fold rRNA is engaged in the interactions with other components of the LSU, such as rRNA of ES39 and rProtein eL6. We used invariant Base Pair Shannon Entropy (2) to validate the ES7 Signature Fold, which was initially identified at the level of 3D structure. Our data confirms that covariation can predict 2D structures of some closely related species that contain divergent sequences. ES7 size correlates with size of other ESs, especially ES15, ES27 and ES39.

Beyond the ES7 Signature Fold

Elaboration of ES7 over phylogeny is achieved primarily by elongation of Signature Fold helices and insertion of stem loops into the elongated helices. Stem loops that insert outside the Signature Fold (as in Groups 2 and 3, Table 1) exhibit moderate conservation of position. These stem loops demonstrate substantial variation in length, and sequence composition over phylogeny, and are characterized by small insertions as well as variations in base pairing (even for closely related species such as *D. melanogaster* and *A. albopictus*).

2D models of ES7 were described previously by Gutell (30,70) for a small subset of phyla. These models are accurate for Hymenoptera (29,60), Acari (71) and some protists (72) even though sequence divergence, size and branching complexity presented challenges. However, structures of ES7s from many mammals and birds remained unresolved until now. We note that the 2D structure of ES7 of *M. musculus* presented here (ES7^{Petrov}) is substantially different from an early model by Bachellerie (ES7^{Bachellerie}) (9). ES7^{Bachellerie} contains two four-way junctions, while ES7^{Petrov} contains several three-way junctions (Supplementary Figure S14), and resembles a morphology of ES7 of *H. sapiens*.

Convergent evolution of ES7

Phylogenetic comparison of ES7s reveals evidence of convergence. Polyphyletic distribution of ES7 features from Groups 2 and 3 across phylogeny suggests that some insertions took place in parallel. Analogous elaborations, that project from the Signature Fold and result in extension of ES7abc and branching of ES7d, are broadly dispersed over phylogeny and are found in insects (*D. melanogaster* and *A. albopictus*), plants (*A. thaliana* and *O. sativa*) and fungi (*C. neoformans*). These elaborations are not observed in rRNAs of the deepest ancestral nodes. Indicators of convergent evolution are also seen for tentacles of Group 3.

Long tentacles are attached at different sites in rRNAs of different species. ES7 of *G. gallus* contains one long tentacle (ES7d1) attached at the position of a small stem-loop (that branches form ES7d) in mammalian rRNAs. ES7 of *H. sapiens* contains two long tentacles (ES7a and ES7b),

which attach at sites of shorter tentacles in the rRNA of *G. gallus* (Supplementary Figure S7). The comparison suggests similar function of highly elongated extensions that transcends specific sites of attachment.

Biological roles of ES7

Ribosomal expansion in eukaryotes might arise from conventional adaptive pressures, but might also be explained in part by Constructive Neutral Evolution (CNE) (73–75). CNE is a multi-step process in which neutral, non-adaptive change opens capacities for complementation and co-dependency. In a first CNE step, an intrinsic or environmental change would alter ribosomal structure, expanding the rRNA, with little or no gain in function. This rRNA expansion might not initially confer advantage. Subsequent steps would stumble upon functions for the expansion segment, ratcheting the complexity.

Experimental manipulation of rRNA of eukaryotic ribosomes *in vivo* is problematic, complicating the determination of functions (76). Nevertheless, expansion segment ES7 is associated with multiple functions, many of which have been recently highlighted for fungal and protist ribosomes (Group 1). Thus, removal of ES7 has been found to be lethal for fungi (77–79). ES7 interacts with ribosomal protein eL14 via ES7b (16). The tip of ES7b can be in proximity to a flexible L7/12-stalk and may transiently interact with ribosomal protein uL10 (80). It has been suggested that conformational changes of fungal ES7 (together with ES39) affect binding of the Signal Recognition Particle, facilitating a conformational switch of ES7/ES39 (81). ES7 has been found to interact with ribosomal assembly factors Nsa1p, Mak16p and Rpf1p in fungi, highlighting its functional importance in early ribosomal assembly (82–85). Recently, it has been shown that ES7 in fungi interacts with the assembly factor Noc2p (86) in functionally coordinated modules in the early 66S pre-ribosome, suggesting that ESs may have coevolved with ribosomal assembly factors to drive differentiation of the pathways of ribosomal biogenesis. ES7 is cleaved by Fe²⁺, causing a rapid hydrolysis of ribosomal particles under oxidative stress (87). ES7a with ES27a help to localize N-terminal acetylases (88) and influence translation fidelity (78).

ES7 of higher eukaryotes gained additional functions. Ribosomes of Group 3 (Table 1), but not those of Groups 1 and 2 contain ES7e. This stem-loop plays a role in selenoprotein synthesis by interacting with SECIS-binding protein SBP2 via a K-rich motif found to be highly conserved in vertebrates (89). Sequences of the K-rich motif of SBP2 vary between Group 2 and Group 3 (90). Based on these observations we suggest a co-evolution of ES7e and SBP2(90).

A recent bioinformatics study suggests that the substantial growth of ESs (including ES7) in mammals may reflect a gain of function related to interactions with non-ribosomal RNAs (91). ES7 contains a multitude of antisense matches (longer than 10nts) to many 5'-UTRs of human mRNAs as well as G-rich tracts which can facilitate triple helix RNA formation.

The structural models of ES7 provided here could be useful for designing, testing, confirming, or rejecting biological or evolutionary hypotheses as well as for modeling the in-

teractions between the expansion segments and other constituents of the translational machinery at the atomic level of details (e.g. rationalizing the data from the pull down experiments).

Evaluation of the hybrid modeling approach

Our hybrid approach predicts 2D and 3D structure despite impediments by dynamics and conformational heterogeneity, large size, sequence hypervariability, high GC content, and absence of covariation signal, and thus is a powerful pipeline for modeling both 2D and 3D structures of rRNAs. The ES7 models presented here appear to correctly capture the vast majority of 2D and 3D elements of ES7, and are useful in interpreting and predicting a variety of data. The hybrid method can be applied to other eukaryotic rRNA expansions as exemplified by modeling of the rRNA LSU from *M. musculus*. Inaccuracies are most pronounced in tertiary interactions between helical motifs, helical and unpaired motifs, or non-paired motifs (such as A-platforms, ribose zippers or pseudoknots). If such interactions are experimentally detected, they could be constrained by incorporation of additional data into RNAComposer modeling (69). The methodology proposed in the current work cannot predict G-quadruplexes. Our 3D modelling may be biased towards high levels of topological similarity between related species (*H. sapiens* vs *S. cerevisiae* or *H. sapiens* vs *C. albicans*); stem *c* (*D. melanogaster* and *A. albopictus*). 2D models are based on conservation of junctions, and RNAComposer predictions rank topology over sequence. Models are predicted in isolation from proteins. The absence of proteins can result in errors when compared with experimental structures, since proteins can stabilize RNA structure or alternatively prohibit formation of canonical base pairs.

CONCLUSIONS

The ES7 is the most diverse expansion segment within eukaryotic ribosomes. It is the product of intense restraints and convergent evolution, and is stringently maintained as helices accrete and elongate. ES7 structures have been produced here by efforts that integrated techniques and constraints from 1D, 2D, and 3D-based algorithms and datasets. The data support an evolutionary model in which ES7 of metazoan lineages emerged through a series of elementary and recursive growth events. As a general pattern, a given stem-loop, once introduced into an ancestral rRNA, appears to freeze, and remain conserved in rRNAs of daughter species. Similar 2D topologies and 3D morphologies can be achieved by a variety of sequences. Common ES7 morphology is observed among species with significant differences in sequence.

Expansions that perturb underlying ribosomal structure or functional regions of the ribosome are not observed. The number of distinct topological and morphological solutions exploited by nature appears to be limited. Comparison of ES7 structures across phylogeny reveals evidence of convergent evolution. The existence of a broad range intermediate structures in non-related RNAs suggest that nucleotide-level growth of the rRNA expansion segments was an incremental and reversible process in which many closely related states were sampled.

Our work is inspired by pioneering efforts of Gutell (29,30,72,92). The identification of the ES7 Signature Fold of eukaryotes and the inference of rRNA accretion during evolution (5,93) enabled our prediction of 2D and 3D structures for diverse eukaryotes. Our hybrid method is facilitated by the availability of 3D structures of ES7 for diverse eukaryotic species and the ability of RNAComposer to predict RNA 3D structures when structural templates are missing. Our results provide a framework for further development of high-throughput predictions using statistical approaches (Infernal, R-scape) (94–96).

The structures modelled here reveal principles of rRNA accretion observed in the experimentally derived structures, and support the Accretion Model of ribosomal evolution (5,93,97). A recent experimental Cryo-EM structure of a plant *Solanum lycopersicum* (98) confirmed our predictions for ES7 of Group 3. Since the hybrid approach proposed here appears to successfully predict the largest eukaryotic expansion segments, it should be capable of resolving structures of ES's of essentially any rRNA, as demonstrated by the model of the LSU rRNA from *M. musculus*. The models predicted in the current study can now be used as templates in R2DT webserver (33) to derive accurate secondary structures of the other species.

DATA AVAILABILITY

The 2D and 3D models produced in the current study are available as Supplementary Datasets.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We would like to acknowledge Dr Elena Rivas for helpful discussions.

FUNDING

National Aeronautics and Space Administration [80NSSC18K1139 to L.D.W., A.S.P.]; National Science Center Poland [2012/06/A/ST6/00384]; Polish Ministry of Science and Higher Education, under the KNOW programme; Foundation for Polish Science [HOMING PLUS/2012–6/12 to K.J.P.]. Funding for open access charge: NASA [80NSSC18K1139].

Conflict of interest statement. None declared.

References

- Melnikov,S., Ben-Shem,A., Garreau de Loubresse,N., Jenner,L., Yusupova,G. and Yusupov,M. (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*, **19**, 560–567.
- Bernier,C.R., Petrov,A.S., Kovacs,N.A., Penev,P.I. and Williams,L.D. (2018) Translation: the universal structural core of life. *Mol. Biol. Evol.*, **35**, 2065–2076.
- Bowman,J.C., Petrov,A.S., Frenkel-Pinter,M., Penev,P.I. and Williams,L.D. (2020) Root of the tree: the significance, evolution, and origins of the ribosome. *Chem. Rev.*, **120**, 4848–4878.
- Anger,A.M., Armache,J.P., Berninghausen,O., Habeck,M., Subklewe,M., Wilson,D.N. and Beckmann,R. (2013) Structures of the human and drosophila 80S ribosome. *Nature*, **497**, 80–85.
- Petrov,A.S., Bernier,C.R., Hsiao,C., Norris,A.M., Kovacs,N.A., Waterbury,C.C., Stepanov,V.G., Harvey,S.C., Fox,G.E., Wartell,R.M. *et al.* (2014) Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 10251–10256.
- Ware,V.C., Tague,B.W., Clark,C.G., Gourse,R.L., Brand,R.C. and Gerbi,S.A. (1983) Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*. *Nucleic Acids Res.*, **11**, 7795–7817.
- Clark,C.G., Tague,B.W., Ware,V.C. and Gerbi,S.A. (1984) *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. *Nucleic Acids Res.*, **12**, 6197–6220.
- Gerbi,S.A. (1996) In Zimmermann,R. A. and Dahlberg,A. E. (eds.), *Ribosomal RNA—Structure, Evolution, Processing, and Function in Protein Synthesis*. CRC Press, Boca Raton, FL, pp. 71–87.
- Hassouna,N., Michot,B. and Bachelierie,J.P. (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Res.*, **12**, 3563–3583.
- Han,H., Schepartz,A., Pellegrini,M. and Dervan,P.B. (1994) Mapping RNA regions in eukaryotic ribosomes that are accessible to methidiumpropyl-edta-fe(ii) and EDTA-Fe(II). *Biochemistry*, **33**, 9831–9844.
- Larsson,S.L. and Nygård,O. (2001) Proposed secondary structure of eukaryote specific expansion segment 15 in 28S rRNA from mice, rats, and rabbits. *Biochemistry*, **40**, 3222–3231.
- Alkemar,G. and Nygård,O. (2004) Secondary structure of two regions in expansion segments ES3 and ES6 with the potential of forming a tertiary interaction in eukaryotic 40S ribosomal subunits. *RNA (New York, NY)*, **10**, 403–411.
- Alkemar,G. and Nygård,O. (2006) Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry*, **45**, 8067–8078.
- Spahn,C.M., Beckmann,R., Eswar,N., Penczek,P.A., Sali,A., Blobel,G. and Frank,J. (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*: trna-Ribosome and subunit-subunit interactions. *Cell*, **107**, 373–386.
- Armache,J.P., Jarasch,A., Anger,A.M., Villa,E., Becker,T., Bhushan,S., Jossinet,F., Habeck,M., Dindar,G., Franckenberg,S. *et al.* (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Ångstrom resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 19748–19753.
- Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
- Klinge,S., Voigts-Hoffmann,F., Leibundgut,M., Arpagaus,S. and Ban,N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science*, **334**, 941–948.
- Vicens,Q., Bochler,A., Jobe,A., Frank,J. and Hashem,Y. (2021) In Harris,J. R. and Marles-Wright,J. (eds.), *Macromolecular Protein Complexes III: Structure and Function*. Springer International Publishing, Cham, pp. 433–450.
- Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
- Watson,Z.L., Ward,F.R., Méheust,R., Ad,O., Schepartz,A., Banfield,J.F. and Cate,J.H. (2020) Structure of the bacterial ribosome at 2 Å resolution. *Elife*, **9**, e60482.
- Harms,J., Schluenzen,F., Zarivach,R., Bashan,A., Gat,S., Agmon,I., Bartels,H., Franceschi,F. and Yonath,A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679–688.
- Selmer,M., Dunham,C.M., Murphy,F.V., Weixlbaumer,A., Petry,S., Kelley,A.C., Weir,J.R. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
- Michot,B., Hassouna,N. and Bachelierie,J.P. (1984) Secondary structure of mouse 28S rRNA and general model for the folding of the large rRNA in eukaryotes. *Nucleic Acids Res.*, **12**, 4259–4279.
- Parker,M.S., Sallee,F.R., Park,E.A. and Parker,S.L. (2015) Homoiterons and expansion in ribosomal RNAs. *FEBS Open Biol*, **5**, 864–876.
- Mestre-Fos,S., Ito,C., Moore,C.M., Reddi,A.R. and Williams,L.D. (2020) Human ribosomal G-quadruplexes regulate heme bioavailability. *J. Biol. Chem.*, **295**, 14855–14865.

26. Mestre-Fos, S., Penev, P.I., Suttapitugsakul, S., Hu, M., Ito, C., Petrov, A.S., Wartell, R.M., Wu, R. and Williams, L.D. (2019) G-quadruplexes in human ribosomal RNA. *J. Mol. Biol.*, **431**, 1940–1955.
27. Melnikov, S.V., Manakongtreecheep, K., Rivera, K.D., Makarenko, A., Pappin, D.J. and Söll, D. (2018) Muller's ratchet and ribosome degeneration in the obligate intracellular parasites microsporidia. *Int. J. Mol. Sci.*, **19**, 4125.
28. Holmberg, L., Melander, Y. and Nygård, O. (1994) Probing the structure of mouse ehrlich ascites cell 5.8S, 18S and 28S ribosomal RNA in situ. *Nucleic Acids Res.*, **22**, 1374–1382.
29. Gillespie, J.J., Johnston, J.S., Cannone, J.J. and Gutell, R.R. (2006) Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *apismellifera* (Insecta: hymenoptera): structure, organization, and retrotransposable elements. *Insect Mol. Biol.*, **15**, 657–686.
30. Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M. et al. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinf.*, **3**, 2.
31. Hashem, Y., des Georges, A., Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C. et al. (2013) High-Resolution cryo-electron microscopy structure of the *trypanosomabrucei* ribosome. *Nature*, **494**, 385–389.
32. Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F. and Backofen, R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA (New York, NY)*, **18**, 900–914.
33. Sweeney, B.A., Hoksza, D., Nawrocki, E.P., Ribas, C.E., Madeira, F., Cannone, J.J., Gutell, R., Maddala, A., Meade, C.D., Williams, L.D. et al. (2021) R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat. Commun.*, **12**, 3494.
34. Bernier, C.R., Petrov, A.S., Waterbury, C.C., Jett, J., Li, F., Freil, L.E., Xiong, X., Wang, L., Migliozi, B.L., Hershkovits, E. et al. (2014) Ribovision suite for visualization and analysis of ribosomes. *Farad. Discuss.*, **169**, 195–207.
35. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
36. Popena, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J. and Adamiak, R.W. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.
37. Moran, N.A. (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, **108**, 583–586.
38. McCutcheon, J.P. and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.*, **10**, 13–26.
39. Katoh, K. and Standley, D.M. (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*, **32**, 1933–1942.
40. Jenner, L., Melnikov, S., Garreau de Loubresse, N., Ben-Shem, A., Iskakova, M., Urzhumtsev, A., Meskauskas, A., Dinman, J., Yusupova, G. and Yusupov, M. (2012) Crystal structure of the 80S yeast ribosome. *Curr. Opin. Struct. Biol.*, **22**, 759–767.
41. Popena, M., Blazewicz, M., Szachniuk, M. and Adamiak, R.W. (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res.*, **36**, D386–D391.
42. Popena, M., Szachniuk, M., Blazewicz, M., Wasik, S., Burke, E.K., Blazewicz, J. and Adamiak, R.W. (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinf.*, **11**, 231.
43. Flores, S.C., Wan, Y.Q., Russell, R. and Altman, R.B. (2010) Predicting RNA structure by multiple template homology modeling. Pacific Symposium on Biocomputing 2010, 216–227.
44. Rother, M., Rother, K., Puton, T. and Bujnicki, J.M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
45. Antczak, M., Popena, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., Adamiak, R.W. and Szachniuk, M. (2016) New functionality of RNAComposer: an application to shape the axis of mir160 precursor structure. *Acta Biochim. Pol.*, **63**, 737–744.
46. Laing, C. and Schlick, T. (2009) Analysis of four-way junctions in RNA structures. *J. Mol. Biol.*, **390**, 547–559.
47. Touzet, H. and Perriquet, O. (2004) CARNAC: folding families of related RNAs. *Nucleic Acids Res.*, **32**, W142–W145.
48. Yao, Z.Z., Weinberg, Z. and Ruzzo, W.L. (2006) CMfinder - a covariance model based RNA motif finding algorithm. *Bioinformatics*, **22**, 445–452.
49. Lindgreen, S., Gardner, P.P. and Krogh, A. (2007) MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing. *Bioinformatics*, **23**, 3304–3311.
50. Kiryu, H., Tabei, Y., Kin, T. and Asai, K. (2007) Mulet: a practical multiple alignment tool for structural RNA sequences. *Bioinformatics*, **23**, 1588–1598.
51. Xu, Z.J. and Mathews, D.H. (2011) Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics*, **27**, 626–632.
52. Wu, Y., Shi, B.B., Ding, X.Q., Liu, T., Hu, X.H., Yip, K.Y., Yang, Z.R., Mathews, D.H. and Lu, Z.J. (2015) Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.*, **43**, 7247–7259.
53. Low, J.T. and Weeks, K.M. (2010) Shape-Directed RNA secondary structure prediction. *Methods*, **52**, 150–158.
54. Ramos, L.M.G., Degtyareva, N.N., Kovacs, N.A., Holguin, S.Y., Jiang, L.W., Petrov, A.S., Biesiada, M., Hu, M.Y., Purzycka, K.J., Arya, D.P. et al. (2017) Eukaryotic ribosomal expansion segments as antimicrobial targets. *Biochemistry*, **56**, 5288–5299.
55. Srikantha, T., Gutell, R.R., Morrow, B. and Soll, D.R. (1994) Partial nucleotide-sequence of a single Ribosomal-RNA coding region and secondary structure of the large subunit 25S-Ribosomal RNA of *candida-albicans*. *Curr. Genet.*, **26**, 321–328.
56. Abeyrathne, P.D., Koh, C.S., Grant, T., Grigorieff, N. and Korostelev, A.A. (2016) Ensemble cryo-em uncovers inchworm-like translocation of a viral ires through the ribosome. *Elife*, **5**, e14874.
57. Garreau de Loubresse, N., Prokhorova, I., Holtkamp, W., Rodnina, M.V., Yusupova, G. and Yusupov, M. (2014) Structural basis for the inhibition of the eukaryotic ribosome. *Nature*, **513**, 517–522.
58. Matthews, B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta*, **405**, 442–451.
59. Li, Z., Guo, Q., Zheng, L., Ji, Y., Xie, Y.-T., Lai, D.-H., Lun, Z.-R., Suo, X. and Gao, N. (2017) Cryo-EM structures of the 80S ribosomes from human parasites *trichomonas vaginalis* and *toxoplasma gondii*. *Cell Res.*, **27**, 1275–1288.
60. Gillespie, J.J., Munro, J.B., Heraty, J.M., Yoder, M.J., Owen, A.K. and Carmichael, A.E. (2005) A secondary structural model of the 28S rRNA expansion segments D2 and D3 for chalcidoid wasps (Hymenoptera: chalcidoidea). *Mol. Biol. Evol.*, **22**, 1593–1608.
61. Kyriades, N., Overbeek, R. and Ouzounis, C. (1999) Universal protein families and the functional content of the last universal common ancestor. *J. Mol. Evol.*, **49**, 413–423.
62. Grishin, N.V. (2001) Fold change in evolution of protein structures. *J. Struct. Biol.*, **134**, 167–185.
63. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A. and Williams, L.D. (2006) Single nucleotide RNA choreography. *Nucleic Acids Res.*, **34**, 1481–1491.
64. Vargason, J.M., Henderson, K. and Ho, P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 7265–7270.
65. Mohan, S., Hsiao, C., VanDeusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, R.M. and Williams, L.D. (2009) Mechanism of RNA double helix-propagation at atomic resolution. *J. Phys. Chem. B*, **113**, 2614–2623.
66. Sundaralingam, M. and Sekharudu, Y.C. (1989) Water-Inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333–1337.
67. Burgi, H.B. (1973) Chemical reaction coordinates from crystal-structure data 1. *Inorg. Chem.*, **12**, 2321–2325.
68. Petrov, A.S., Bernier, C.R., Gulen, B., Waterbury, C.C., Hershkovits, E., Hsiao, C., Harvey, S.C., Hud, N.V., Fox, G.E., Wartell, R.M. et al. (2014) Secondary structures of rRNAs from all three domains of life. *PLoS One*, **9**, e88222.
69. Purzycka, K.J., Popena, M., Szachniuk, M., Antczak, M., Lukasiak, P., Blazewicz, J. and Adamiak, R.W. (2015) Automated 3D RNA structure prediction using the RNAComposer method for riboswitches. *Methods Enzymol.*, **553**, 3–34.

70. Gutell,R.R. and Fox,G.E. (1988) A compilation of large subunit RNA sequences presented in a structural format. *Nucleic Acids Res.*, **16Suppl**, r175–r269.
71. Zhao,Y.-E., Wang,Z.-H., Xu,Y., Wu,L.-P. and Hu,L. (2013) Secondary structure prediction for complete rDNA sequences (18S, 5.8S, and 28S rDNA) of *Demodex folliculorum*, and comparison of divergent domains structures across acari. *Exp. Parasitol.*, **135**, 370–381.
72. Schnare,M.N., Damberger,S.H., Gray,M.W. and Gutell,R.R. (1996) Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23S-Like) ribosomal RNA. *J. Mol. Biol.*, **256**, 701–719.
73. Muñoz-Gómez,S.A., Bilollikar,G., Wideman,J.G. and Geiler-Samerotte,K. (2021) Constructive neutral evolution 20 years later. *J. Mol. Evol.*, **89**, 172–182.
74. Stoltzfus,A. (1999) On the possibility of constructive neutral evolution. *J. Mol. Evol.*, **49**, 169–181.
75. Lynch,M. and Walsh,B. (2007) *The Origins of Genome Architecture*. Sinauer Associates Sunderland, MA.
76. Hariharan,N., Ghosh,S. and Palakodeti,D. (2022) The story of rRNA expansion segments: finding functionality amidst diversity. *Wiley Interdiscip. Rev. RNA*, e1732.
77. Shankar,V., Rauscher,R., Reuther,J., Gharib,W.H., Koch,M. and Polacek,N. (2020) rRNA expansion segment 27Lb modulates the factor recruitment capacity of the yeast ribosome and shapes the proteome. *Nucleic Acids Res.*, **48**, 3244–3256.
78. Fujii,K., Susanto,T.T., Saurabh,S. and Barna,M. (2018) Decoding the function of expansion segments in ribosomes. *Mol. Cell*, **72**, 1013–1020.e1016.
79. Ramesh,M. and Woolford,J.L. Jr. (2016) Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. *RNA*, **22**, 1153–1162.
80. Diaconu,M., Kothe,U., Schlünzen,F., Fischer,N., Harms,J.M., Tonevitsky,A.G., Stark,H., Rodnina,M.V. and Wahl,M.C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*, **121**, 991–1004.
81. Nilsson,J., Sengupta,J., Gursky,R., Nissen,P. and Frank,J. (2007) Comparison of fungal 80S ribosomes by cryo-EM reveals diversity in structure and conformation of rRNA expansion segments. *J. Mol. Biol.*, **369**, 429–438.
82. Greber,B.J., Boehringer,D., Montellese,C. and Ban,N. (2012) Cryo-EM structures of arx1 and maturation factors reil1 and jji1 bound to the 60S ribosomal subunit. *Nat. Struct. Mol. Biol.*, **19**, 1228–1233.
83. Wu,S., Tutuncuoglu,B., Yan,K., Brown,H., Zhang,Y., Tan,D., Gamalinda,M., Yuan,Y., Li,Z., Jakovljevic,J. *et al.* (2016) Diverse roles of assembly factors revealed by structures of late nuclear Pre-60S ribosomes. *Nature*, **534**, 133–137.
84. Kater,L., Thoms,M., Barrio-Garcia,C., Cheng,J., Ismail,S., Ahmed,Y.L., Bange,G., Kressler,D., Berninghausen,O., Sinning,I. *et al.* (2017) Visualizing the assembly pathway of nucleolar Pre-60S ribosomes. *Cell*, **171**, 1599–1610.e1514.
85. Bradatsch,B., Leidig,C., Granneman,S., Gnädig,M., Tollervey,D., Böttcher,B., Beckmann,R. and Hurt,E. (2012) Structure of the Pre-60S ribosomal subunit with nuclear export factor arx1 bound at the exit tunnel. *Nat. Struct. Mol. Biol.*, **19**, 1234–1241.
86. Wang,X., Yue,Z., Xu,F., Wang,S., Hu,X., Dai,J. and Zhao,G. (2021) Coevolution of ribosomal RNA expansion segment 7L and assembly factor Noc2p specializes the ribosome biogenesis pathway between *Saccharomyces cerevisiae* and *Candida albicans*. *Nucleic Acids Res.*, **49**, 4655–4667.
87. Smethurst,D.G.J., Kovalev,N., McKenzie,E.R., Pestov,D.G. and Shcherbik,N. (2020) Iron-Mediated degradation of ribosomes under oxidative stress is attenuated by manganese. *J. Biol. Chem.*, **295**, 17200–17214.
88. Knorr,A.G., Schmidt,C., Tesina,P., Berninghausen,O., Becker,T., Beatrix,B. and Beckmann,R. (2019) Ribosome–NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol.*, **26**, 35–39.
89. Kossinova,O., Malygin,A., Krol,A. and Karpova,G. (2014) The SBP2 protein central to selenoprotein synthesis contacts the human ribosome at expansion segment 7L of the 28S rRNA. *RNA*, **20**, 1046–1056.
90. Takeuchi,A., Schmitt,D., Chapple,C., Babaylova,E., Karpova,G., Guigo,R., Krol,A. and Allmang,C. (2009) A short motif in drosophila secis binding protein 2 provides differential binding affinity to secis RNA hairpins. *Nucleic Acids Res.*, **37**, 2126–2141.
91. Parker,M.S., Balasubramanian,A. and Parker,S.L. (2018) The expansion segments of human 28S rRNA match microRNAs much above 18S rRNA or core segments. *MicroRNA*, **7**, 128–137.
92. Gutell,R.R. (1992) Evolutionary characteristics of 16S and 23S rRNA structures. In Hartman,H. and Matsuno,K. (eds.), *The Origin and Evolution of the Cell: Proceedings of the Conference on the Origin and Evolution of Prokaryotic and Eukaryotic Cells*, Shimoda, Japan, 22–25 April 1992. World Scientific, Singapore, pp. 243–309.
93. Petrov,A.S., Gulen,B., Norris,A.M., Kovacs,N.A., Bernier,C.R., Lanier,K.A., Fox,G.E., Harvey,S.C., Wartell,R.M., Hud,N.V. *et al.* (2015) History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 15396–15401.
94. Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.
95. Rivas,E. (2020) RNA structure prediction using positive and negative evolutionary information. *PLoS Comp. Biol.*, **16**, e1008387.
96. Rivas,E., Clements,J. and Eddy,S.R. (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods*, **14**, 45–48.
97. Kovacs,N.A., Petrov,A.S., Lanier,K.A. and Williams,L.D. (2017) Frozen in time: the history of proteins. *Mol. Biol. Evol.*, **34**, 1252–1260.
98. Cottilli,P., Itoh,Y., Nobe,Y., Petrov,A.S., Lisón,P., Taoka,M. and Amunts,A. (2022) *Cryo-EM Structure and rRNA Modification Sites of a Plant Ribosome*. *Plant Communications*, **3**, 100342.

ZAŁĄCZNIK NR 3. Oświadczenia autorów określające wkład w powstanie prac naukowych wchodzących w skład rozprawy doktorskiej

Poznań, dnia 01.03.2023

mgr Marcin Biesiada

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE KANDYDATA DO NADANIA
STOPNIA NAUKOWEGO DOKTORA

Tytuł pracy naukowej: Automated RNA 3D Structure Prediction with RNAComposer

Autorzy: Marcin Biesiada, Katarzyna J. Purzycka, Marta Szachniuk, Jacek Błażewicz i Ryszard W. Adamiak

Czasopismo: Methods in Molecular Biology

Rok wydania: 2016

Oświadczam, że mój wkład w powstanie wyżej wymienionej pracy polegał na:

- opracowanie planu pracy pod kierunkiem prof. dr hab. Ryszarda W. Adamiaka
- wykonaniu eksperymentów przewidywania struktur 3D RNA
- przygotowaniu wszystkich ilustracji
- opracowaniu danych ujętych w tabeli 2
- przygotowaniu manuskryptu, a w szczególności opisaniu punktu *Metody*

Marcin Biesiada

Poznań, dnia 01.03.2023

mgr Marcin Biesiada

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE KANDYDATA DO NADANIA
STOPNIA NAUKOWEGO DOKTORA

Tytuł pracy naukowej: RNAComposer and RNA 3D structure prediction for nanotechnology

Autorzy: Marcin Biesiada, Katarzyna Pachulska-Wieczorek, Ryszard W. Adamiak i Katarzyna J. Purzycka

Czasopismo: Methods

Rok wydania: 2016

Oświadczam, że mój wkład w powstanie wyżej wymienionej pracy polegał na:

- opracowaniu planu pracy pod kierunkiem dr Katarzyny J. Purzyckiej i prof. dr hab. Ryszarda W. Adamiaka
- wykonaniu eksperymentów przewidywania struktur 3D RNA
- przygotowaniu ilustracji i tabel
- pomoc w przygotowaniu manuskryptu, a w szczególności wprowadzenia

Marcin Biesiada

Poznań, dnia 01.03.2023

mgr Marcin Biesiada

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE KANDYDATA DO NADANIA
STOPNIA NAUKOWEGO DOKTORA

Tytuł pracy naukowej: rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles

Autorzy: Marcin Biesiada, Michael Y. Hu, Loren Dean Williams, Katarzyna J. Purzycka i Anton S. Petrov

Czasopismo: Nucleic Acids Research

Rok wydania: 2022

Oświadczam, że mój wkład w powstanie wyżej wymienionej pracy polegał na:

- opracowanie planu pracy pod kierunkiem dr Katarzyny J. Purzyckiej oraz dr. Anton S. Petrov
- wykonaniu eksperymentów przewidywania struktur 3D RNA
- przygotowaniu ilustracji 2 (wspólnie z Michael Y. Hu i Anton S. Petrov), ilustracji 3 (wspólnie z Katarzyna J. Purzycka i Anton S. Petrov), ilustracji 4 (wspólnie z Anton S. Petrov), oraz ilustracji 6, 7, S1, S2, S4, S5, S6, S14, S15
- przygotowaniu tabeli S1, S2, S3, oraz tabeli S4 (wspólnie z Michael Y. Hu i Anton S. Petrov)
- przygotowaniu filmu 1 i 2
- dyskusowaniu wyników i obieraniu nowych strategii
- pomoc w przygotowaniu manuskryptu

Marcin Biesiada

Poznań, 15 marca 2023 r.

Prof. dr hab. Ryszard W. Adamiak

OŚWIADCZENIE

Poniżej przedstawiam informacje dotyczące mojego wkładu merytorycznego w przygotowanie dwóch poniższych publikacji doktoranta mgr Marcina Biesiady.

Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW§. *Automated RNA 3D structure prediction with RNAComposer*. *Methods Mol Biol.* 2016; 1490:199-215

Publikacja to rozdział na zaproszenie edytorów monografii. Nakreślenie ram i zakresu szczegółowości opisu metody RNAComposer, koordynacja prac, współudział w przygotowaniu manuskryptu i jego finalnej edycji.

Biesiada M; Pachulska-Wieczorek K; Adamiak RW; Purzycka KJ§. *RNAComposer and RNA 3D structure prediction for nanotechnology*. *Methods.* 2016, 103:120-7

Przedyskutowanie podstawowych celów badań na obszarze nowego pola badawczego jakim było wykorzystanie metod komputerowego przewidywania struktur 3D RNA w nanotechnologii RNA, dyskusja wyników, udział w przygotowaniu finalnej wersji manuskryptu.



Prof. dr hab. Ryszard W. Adamiak

Poznań, 26 marca 2023 r.

Prof. dr hab. inż. Jacek Błażewicz

OŚWIADCZENIE

W odniesieniu do pracy naukowej

Biesiada M, Purzycka KJ, Szachniuk M, Błażewicz J, Adamiak RW (2016). *Automated RNA 3D structure prediction with RNAComposer*, in Doug H. Turner, David H. Mathews (eds.) *RNA Structure Determination: Methods and Protocols (Methods Molecular Biology 1490)*, Springer, Humana Press, 199-215

oświadczam, że mój wkład w jej powstanie polegał na:

- dyskusji nad koncepcją przygotowywanego manuskryptu
- współudziale w przygotowaniu manuskryptu



Prof. dr hab. inż. Jacek Błażewicz

Subject:

Biesiada M; Hu MY; Williams LD; Purzycka KJ; Petrov AS. rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles. **Nucleic Acids Res.** 2022, 50(18): p. 10717-10732

To Whom It May Concern:

I performed experiments that resulted in predictions of the 2D structures presented in Figure 2 and Table 1, and I wrote the first draft describing these results. Prediction of 2D structures entailed breaking the 1D rRNA sequences into meaningful chunks and applying folding algorithms to the chunks.

Sincerely,

A handwritten signature in black ink that reads "Michael Y. Hu". The signature is written in a cursive style with a large initial 'M' and 'H'.

Michael Hu

Poznań, 29.05.2023

dr hab. Katarzyna Pachulska-Wieczorek, prof. IChB PAN
Kierownik Zakładu Struktury i Funkcji Retrotranspozonów
E-mail: kasiapw@ibch.poznan.pl

OŚWIADCZENIE WSPÓŁATORA PUBLIKACJI NAUKOWEJ

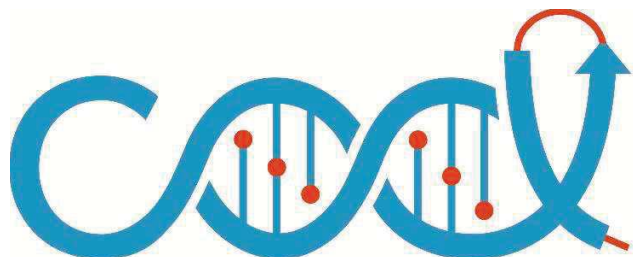
Oświadczam, że w pracy:

Biesiada M; Pachulska-Wieczorek K; Adamiak RW; Purzycka KJ[§]. *RNAComposer and RNA 3D structure prediction for nanotechnology*. **Methods**. 2016, 103:120-7,

brałam udział w tworzeniu koncepcji pracy, analizie i dyskusji uzyskanych wyników oraz przygotowaniu manuskryptu.



dr hab. Katarzyna Pachulska-Wieczorek



Dr. Anton S. Petrov
School of Chemistry and Biochemistry

CENTER FOR THE ORIGIN OF LIFE

Georgia Institute of Technology, Atlanta, GA 30332
anton.petrov@biology.gatech.edu

Subject:

Biesiada M; Hu MY; Williams LD; Purzycka KJ; Petrov AS. rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles. **Nucleic Acids Res.** 2022, 50(18): p. 10717-10732

To Whom It May Concern:

I designed the research project, I supervised analyses and experiments performed by Michael Hu that resulted in prediction of 2D structures presented in Figures 1 and 2 and Table 1, and co-supervised the modelling performed by Mr. Biesiada (together with Dr. Purzycka). Together with Prof. Purzycka, we designed the overall pipeline of the study (Figure 3).

I performed experiments shown in Figures 5 and 8 and provided their graphical representation. I participated in the discussion of the results, general and detailed concepts of this work. Together with Prof. Williams and Prof. Purzycka, I participated in preparation of the initial drafts and the final version of the manuscript. I also participated in revision of the manuscript and responding to reviewer comments.

Sincerely,

Anton S. Petrov

Athens, 30 maja 2023 r.

dr Katarzyna J. Purzycka

OŚWIADCZENIE

Oświadczam, że w pracy:

Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW^s. *Automated RNA 3D structure prediction with RNAComposer. Methods Mol Biol. 2016; 1490:199-215*

mój udział polegał na współtworzeniu ogólnej i szczegółowej koncepcji pracy. Byłam odpowiedzialna za część pracy ukazującej wpływ danych eksperymentalnych na wyniki predykcji. Pod moim kierunkiem mgr Marcin Biesiada wykonywał eksperymenty przewidywania struktury. Byłam współodpowiedzialna za ostateczną wersję manuskryptu.



dr Katarzyna J. Purzycka

Athens, 30 maja 2023 r.

dr Katarzyna J. Purzycka

OŚWIADCZENIE

Oświadczam, że w pracy:

Biesiada M; Pachulska-Wieczorek K; Adamiak RW; Purzycka KJ[§]. *RNAComposer and RNA 3D structure prediction for nanotechnology*. **Methods**. 2016, 103:120-7

byłam pomysłodawcą ogólnej koncepcji pracy oraz współtwórcą planu szczegółowego. Brałam udział w analizach i dyskusji otrzymanych wyników. Byłam współodpowiedzialna za przygotowanie ostatecznej wersji publikacji, korespondencję z recenzentami oraz korekty i uzupełnia pracy na etapie recenzji.



dr Katarzyna J. Purzycka

Athens, 30 maja 2023 r.

dr Katarzyna J. Purzycka

OŚWIADCZENIE

Oświadczam, że w pracy:

Biesiada M; Hu MY; Williams LD; Purzycka KJ[§]; Petrov AS[§]. *rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles*. **Nucleic Acids Res.** **2022**, 50(18): p. 10717-10732

pod moim kierunkiem mgr Marcin Biesiada prowadził eksperymenty predykcji struktury 3D. Byłam współautorem koncepcji badań. Uczestniczyłam w analizie, interpretacji i dyskusji wyników zawartych w pracy. Byłam współodpowiedzialna za przygotowanie ostatecznej wersji publikacji, korespondencję z recenzentami oraz korekty i uzupełnia pracy na etapie recenzji.



dr Katarzyna J. Purzycka

Poznań, dnia 13.03.2023r.

Prof. dr hab. inż. Marta Szachniuk
Zakład Bioinformatyki Strukturalnej
Instytut Chemii Bioorganicznej PAN
ul. Noskowskiego 12/14
61-704 Poznań

Oświadczenie współautora publikacji naukowej
zawartej w rozprawie doktorskiej mgr Marcina Biesiady

W odniesieniu do pracy naukowej

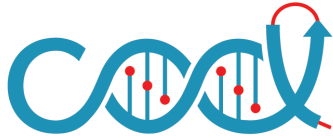
Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW (2016) *Automated RNA 3D structure prediction with RNAComposer*, in Doug H. Turner, David H. Mathews (eds.) *RNA Structure Determination: Methods and Protocols (Methods in Molecular Biology 1490)*, Springer, Humana Press, 199-215,

oświadczam, że mój wkład w jej powstanie polegał na:

- zaprojektowaniu aplikacji sieciowej systemu RNAComposer, nadzorowaniu jego implementacji oraz przygotowaniu i przeprowadzeniu testów oprogramowania,
- opracowaniu podręcznika użytkownika systemu RNAComposer, na podstawie którego została napisana w/w publikacja,
- udziale w przygotowaniu pierwszej wersji manuskryptu.



Prof. dr hab. inż. Marta Szachniuk



**CENTER
FOR THE
ORIGIN OF LIFE**

Loren Williams, Professor and Director
ldw@gatech.edu



Tuesday, 18 April 2023

Re: Biesiada M; Hu MY; Williams LD; Purzycka KJ; Petrov AS. rRNA Expansion ES7 in Eukaryotes: from Signature Fold to Tentacles. **Nucleic Acids Res.** 2022, 50(18): p. 10717-10732

To Whom It May Concern:

I certify that along with Dr. Anton Petrov, I conceived the idea of breaking rRNA into pieces and using folding algorithms and multiple sequence alignments to determine secondary structures of rRNA expansion segments. Michael Hu performed the bulk of the fragment folding. Marcin Biesiada and Kasia Purzycka joined the ongoing effort and extended the approach to include three-dimensional modelling and videos, resulting in a more powerful and accurate hybrid approach. Additionally, Marcin Biesiada made important contributions to the discovery of reaction pathways. Anton Petrov, Kasia Purzycka and I wrote the manuscript. The three of us each made important editorial and technical contributions to the manuscript.

Sincerely,

Loren Williams
Professor,
Center Director