

**Badania strukturalne RNA o znaczeniu
w patogenezie chorób neurodegeneracyjnych**

mgr inż. Marcin Ryczek



Poznań, 2023

Praca została wykonana w Zakładzie Struktury i Funkcji Biomolekuł
Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk

Promotor:

dr hab. Agnieszka Kiliszek, prof. ICHB

Praca przedstawiona Radzie Naukowej
Instytutu Chemii Bioorganicznej
Polskiej Akademii Nauk w Poznaniu
w celu uzyskania stopnia doktora

Praca została sfinansowana z następujących środków:
Grant NCN SONATA BIS 2017/26/E/NZ1/00950:
Opracowanie metodologii umożliwiającej stabilizację RNA w formie spinki do badań
krystalograficznych

Publikacja, która powstała w trakcie realizacji pracy doktorskiej:
Ryczek, M., Pluta, M., Błaszczuk, L., & Kiliszek, A. (2022).
Overview of Methods for Large Scale RNA Synthesis
Applied Sciences 12 (3), doi: 10.3390/app12031543

Serdecznie dziękuję:

dr hab. Agnieszce Kiliszek, prof. ICHB
za lata wspólnej pracy, przekazaną wiedzę i doświadczenie.

Wszystkim pracownikom Zakładu Struktury i Funkcji Biomolekuł
za wszystkie dyskusje naukowe i udzielone rady.

Mojej żonie Natalii
za cierpliwość i wsparcie każdego dnia.

Moim rodzicom
za wiarę we mnie.

Spis treści

Spis rycin i tabel	7
Wykaz skrótów	9
Streszczenie	12
Abstract.....	14
1. Cel pracy.....	15
2. Wprowadzenie	16
2.1. Choroby neurodegeneracyjne.....	16
2.2. Choroby neurodegeneracyjne związane z nadmierną ekspansją nukleotydów	16
2.3. Znaczenie struktury RNA w biologii	20
2.4. Struktury spinkowe RNA w badaniach krystalograficznych	22
2.5. Produkcja RNA do badań krystalograficznych.....	26
2.6. Transkrypcja <i>in vitro</i>	27
2.7. Rybozomy	29
2.8. Synteza chemiczna	33
2.9. Stwardnienie zanikowe boczne	35
2.10. Ośłupienie czołowo-skroniowe.....	38
2.11. Powtórzenia G ₂ C ₄	39
2.12. Małe cząsteczki oddziałujące z RNA.....	42
2.13. Ligand ANP77	43
3. Materiały i Metody	46
3.1. Materiały	46
3.1.1. Aparatura	46
3.1.2. Odczynniki chemiczne i biochemiczne	46
3.1.3. Stosowane bufony roztwory i pożywki.....	47
3.1.4. ANP77	49
3.1.5. Konstrukty RNA użyte w pracy	50
3.2. Metody	50
3.2.1. Produkcja polimerazy T7 i ligazy RNA T4.....	50
3.2.2. Produkcja białka U1a-RBD	51
3.2.3. Synteza chemiczna RNA	52
3.2.4. HPLC.....	53
3.2.5. Transkrypcja <i>in vitro</i>	53

3.2.6.	Oczyszczanie RNA w żelach poliakrylamidowych	53
3.2.7.	Elektroelucja RNA	54
3.2.8.	Strącanie RNA izopropanolem.....	54
3.2.9.	Cyrkularyzacja RNA	54
3.2.10.	Ekstrakcja RNA fenolem.....	54
3.2.11.	Badanie interakcji RNA – białko metodą EMSA.....	55
3.2.12.	Krystalizacja RNA.....	55
3.2.13.	Pomiary dyfrakcyjne, procesowanie danych i rozwiązanie struktury	56
3.2.14.	Udokładnianie struktury RNA.....	57
3.2.15.	Skaningowa kalorymetria różnicowa	59
3.2.16.	Kriomikroskopia elektronowa	59
4.	Wyniki	61
4.1.	Opracowanie protokołu do produkcji stabilnych struktur spinkowych RNA do badań krystalograficznych	61
4.1.1.	Produkcja RNA na drodze syntezy chemicznej	62
4.1.2.	Produkcja RNA na drodze transkrypcji <i>in vitro</i>	63
4.1.3.	Określenie struktury zaprojektowanych konstruktów	67
4.2.	Struktura krystaliczna oligomeru G ₂ C ₄	70
4.2.1.	Krystalizacja kompleksu G ₂ C ₄ z ANP77 oraz oligomeru G ₂ C ₄	71
4.2.2.	Rejestracja i przetwarzanie danych dyfrakcyjnych	71
4.2.3.	Rozwiązanie i udokładnianie struktury kompleksu oraz oligomeru G ₂ C ₄ bez liganda ANP77	72
4.2.4.	Struktura G ₂ C ₄ i G ₂ C ₄ z cząsteczką ANP77	73
4.2.5.	Ułożenie reszt cytozyny w sieci krystalicznej struktury G ₂ C ₄ bez cząsteczki ANP77	75
4.2.6.	Wiązanie G ₂ C ₄ z ANP77	76
4.2.7.	Ułożenie reszt cytozyny w sieci krystalicznej struktury G ₂ C ₄ z cząsteczką ANP77	78
4.2.8.	Skaningowa kalorymetria różnicowa oligomeru G ₂ C ₄	79
5.	Dyskusja	82
6.	Wnioski.....	92
7.	Literatura	93

Spis rycin i tabel

Rycina 1. Schemat lokalizacji powtórzeń nukleotydowych związanych z chorobami neurologicznymi.	17
Rycina 2. Mechanizmy przyczyniające się do patologii chorób wynikających z ekspansji powtórzeń nukleotydów.	20
Rycina 3. Schemat przedstawiający przykładową strukturę pierwszorzędową, drugorzędową i trzeciorzędową RNA.	21
Rycina 4. Schemat drugorzędowej struktury spinkowej RNA.	23
Rycina 5. Struktura drugorzędowa formowana przez powtórzenia CNG.	24
Rycina 6. Oddziaływanie fragmentu wiążącego RNA białka U1a z spinką II U1snRNA.	26
Rycina 7. Trifosforan 5'- guanozyny.	28
Rycina 8. Schemat hydrolizy przeprowadzanej przez rybozomy.	29
Rycina 9. Działanie rybozymów w trybie <i>cis</i> i <i>trans</i> na przykładzie rybozymu typu głowa młotka.	30
Rycina 10. Struktura drugorzędowa którą przyjmują rybozomy.	31
Rycina 11. Schemat reakcji transestryfikacji konstruktów przez rybozom.	32
Rycina 12. Amidofosforyn używany do syntezy chemicznej na podłożu stałym.	33
Rycina 13. Schemat cyklu syntezy chemicznej RNA w fazie stałej.	34
Rycina 14. Powtórzenia G ₂ C ₄ znajdują się w genie <i>C9orf72</i>	39
Rycina 15. Schemat patologicznego mechanizmu działania nadmiernych powtórzeń sześcienukleotydowych w genie <i>C9orf72</i>	40
Rycina 16. Struktura krystaliczna sekwencji zawierającej cztery powtórzenia G ₂ C ₄ z brakującymi dwiema resztami guanozyny na obu końcach konstruktów.	41
Rycina 17. Schemat przedstawiający cząsteczkę cyklofanu zawierającego dwa pierścienie aromatyczne.	43
Rycina 18. Ilustracja konformacji typu open-unstacked i typu stacked ułożonych jako cząsteczka bez ładunku, monokationu i dikationu ANP77.	44
Rycina 19. Formowanie się różnych struktur RNA w zależności od warunków i schemat łączenia końców RNA przy użyciu ligazy RNA T4.	61
Rycina 20. Chromatogramy HPLC chemicznie zsyntetyzowanych RNA.	63
Rycina 21. Ocena wydajności reakcji transestryfikacji przeprowadzonych przez rybozomy.	65
Rycina 22. Transkrypcja RNA <i>in vitro</i> przy użyciu DNA modyfikowanego C2' metoksy.	66
Rycina 23. Predykcje struktury drugorzędowej zaprojektowanych konstruktów przeprowadzone w programie RNAstructure.	68
Rycina 24. Oddziaływanie konstruktów RNA z U1a-RBD.	69
Rycina 25. Mikrografia z kriomikroskopu elektronowego.	70
Rycina 26. Drugorzędowa struktura G ₂ C ₄ z cząsteczką ANP77 i bez ANP77.	73
Rycina 27. Struktura trzeciorzędowa G ₂ C ₄ z ligandem oraz bez liganda ANP77.	74
Rycina 28. Przykład oddziaływania dwóch reszt cytydydy i jednej reszty guaniny w strukturze G ₂ C ₄	75
Rycina 29. Ułożenie reszt cytydyny w tryplecie międzycząsteczkowym bez udziału ANP77.	76

Rycina 30. Oddziaływanie liganda z 3' resztami cytydyny nici A tetrameru G ₂ C ₄	77
Rycina 31. Numeracja atomów liganda ANP77.	77
Rycina 32. Oddziaływania cząsteczki ANP77 i reszty cytydyny na pozycji 5 nici A tetrameru z wodą.	78
Rycina 33. Ułożenie reszt cytydyny w kontaktach w sieci krystalicznej z udziałem ANP77.	79
Rycina 34. Wyniki DSC dla oligomeru G ₂ C ₄ bez i w obecności liganda ANP77 w pH: 5,26, 6,0, 7,0.	80
Rycina 35. Model strukturalny antysensownego RNA genu <i>C9orf72</i>	88

Tabela 1. Tabela przedstawiająca wybrane choroby neurodegeneracyjne związane z powtórzeniami nukleotydów w DNA.	18
Tabela 2. Zestawienie struktur krystalicznych RNA zawierających powtórzenia trójnukleotydowe z ich kodami PDB.	25
Tabela 3. Zestawienie sekwencji RNA użytych w pracy.	50
Tabela 4. Zestawienie kroków oczyszczanie produktów syntezy chemicznej metodami DMT-ON i DMT-OFF.	63
Tabela 5. Spis konstruktów z rybozymami użytych w pracy.	64
Tabela 6. Zestawienie konstruktów RNA zawierających powtórzenia CUG.	68
Tabela 7. Zestawienie sesji testowych wykonywanych przy użyciu kriomikroskopu elektronowego.	69
Tabela 8. Dane krystalograficzne oraz charakterystyka zbioru danych dla kryształów struktury konstruktu G ₂ C ₄ w obecności i bez cząsteczki ANP77.	71
Tabela 9. Tabela przedstawiająca temperatury topnienia dla oligomerów G ₂ C ₄ w wariantach z i bez ANP77 w różnych warunkach pH.	81

Wykaz skrótów

A – reszta adenozyiny

ALS – stwardnienie zanikowe boczne (z ang. *Amyotrophic Lateral Sclerosis*)

ATP – trifosforan 5'- adenozyiny

C – reszta cytydyny

C9FTD/ALS - otępienie czołowo-skroniowe i stwardnienie zanikowe boczne związane z genem C9orf72 (z ang. *C9ORF72-Associated Frontotemporal Dementia and Amyotrophic Lateral Sclerosis*)

cryo-EM – kriomikroskopia elektronowa (z ang. *cryogenic electron microscopy*)

DM – dystrofia miotoniczna (z ang. *Myotonic Dystrophy*)

DMTr – 4,4-dimetoksytrytyl

DNA – kwas deoksyrybonukleinowy

DPR – powtórzenia dipeptydowe (z ang. *Dipeptide Repeats*)

DRPLA – zanik jądra zębatego, jądra czerwienego i jądra niskowzgórzowego (z ang. *Dentatorubral-Pallidolusian Atrophy*)

DSC – skaningowa kalorymetria różnicowa (z ang. *Differential Scanning Calorimetry*)

EMSA – różnicowa migracja w żelach poliakrylamidowych (z ang. *Electrophoretic Mobility Shift Assay*)

EPM1 – padaczka miokloniczna Unverrichta-Lundborga (z ang. *Progressive Myoclonus Epilepsy 1*)

fALS – rodzinne stwardnienie zanikowe boczne (z ang. *familial Amyotrophic Lateral Sclerosis*)

FRDA – ataksja Friedreicha (z ang. *Friedreich Ataxia*)

FSE – element stymulujący przesunięcie ramki odczytu (z ang. *Frameshift Stimulation Element*)

FTD – otępienie czołowo-skroniowe (z ang. *frontotemporal dementia*)

FXPOI – przedwczesne wygasanie funkcji jajników związane z zespołem łamliwego chromosomu X (z ang. *Fragile X-Associated Primary Ovarian Insufficiency*)

FXS/FRAXA – syndrom łamliwego chromosomu X (z ang. *Fragile X Syndrome*)

FXTAS – zespół drżenia i ataksji związany z zespołem łamliwego chromosomu X (z ang. *Fragile X-Associated Tremor/Ataxia Syndrome*)

G – reszta guanozyiny

GMP – monofosforan 5'- guanozyiny

HD – choroba Huntingtona (z ang. *Huntington's Disease*)

HDL2 – płasawica podobna do choroby Huntingtona 2 (z ang. *Huntington's Disease-Like 2*)

HDV – wirus zapalenia wątroby typu D (z ang. *Hepatitis Delta Virus*)

HPLC – wysokosprawną chromatografię cieczową (z ang. *High-Performance Liquid Chromatography*)

IPTG – izopropyl-β-D-tiogalaktopiranozyd

IRES – wewnętrzne miejsca wiązania rybosomu (z ang. *Internal Ribosome Entry Site*)

LB – pożywka Luria-Bertani

MBLs – ligandy wiążące niesparowane nukleotydy (z ang. *Mismatch Binding Ligands*)

MR – podstawienie cząsteczkowe (z ang. *Molecular Replacement*)

mRNA – informacyjny RNA (z ang. *messenger RNA*)

N – dowolny nukleotyd

NMR – magnetyczny rezonans jądrowy (z ang. *nuclear magnetic resonance*)

NTP – trifosforan 5' - nukleozydu

OPMD – dystrofia mięśniowa oczno-gardłowa (z ang. *Oculopharyngeal Muscular Dystrophy*)

PAGE – elektroforeza w żelu poliakrylamidowym (z ang. *Polyacrylamide Gel Electrophoresis*)

PDB – baza danych białek (z ang. *Protein Data Bank*)

R – puryna

RAN translacja – translacja bez kodonu AUG (RAN, z ang. *repeat-associated non-AUG translation*)

RBD – domena wiążąca RNA (z ang. *RNA binding domain*)

RBPs – białka wiążące RNA (z ang. *RNA binding proteins*)

RNA – kwas rybonukleinowy (od ang. *Ribonucleic Acid*)

rRNA – rybosomalny RNA (z ang. *ribosomal RNA*)

sALS – sporadyczne stwardnienie zanikowe boczne (sALS, z ang. *sporadic Amyotrophic Lateral Sclerosis*)

SBMA – opuszkowo-rdzeniowy zanik mięśni (z ang. *Spinal-Bulbar Muscular Atrophy*)

SCA – ataksja rdzeniowo-mózdkowa (z ang. *Spinocerebellar Ataxia*)

SNP – polimorfizm pojedynczych nukleotydów (z ang. *single nucleotide polymorphism*)

snRNA – mały jądrowy RNA (z ang. *small nuclear RNA*)

snRNP – mała jądrowa rybonukleoproteina (z ang. *small nuclear ribonucleoprotein*)

TBAF – fluorek tetra-n-butyloamoniowy
TBDMS – tert-butyldimetylosilil
TEA·3HF – trifluorowodorek trietyloaminy
TLC – chromatografia cienkowarstwowa (z ang. *thin layer chromatography*)
T_m – temperatura topnienia (z ang. *melting temperature*)
TOM – triizopropylodimetylosililooksymetyl
tRNA – transportujący RNA (z ang. *transfer RNA*)
U – reszta urydyny
UTR – region nieulegający translacji (z ang. *untranslated region*)
V – wolty
Y – pirymidyna

Streszczenie

Wiele chorób człowieka zwłaszcza chorób neurodegeneracyjnych i neuromuskularnych jest wywoływanych przez ekspansję powtórzeń nukleotydów zwanych mikrosatelitami w różnych genach. Transkrypty powstające na matrycy tych genów ze względu na posiadanie zwiokrotnionych powtórzeń nukleotydowych mogą formować różne struktury. Badanie struktur RNA związanych z chorobami neurodegeneracyjnymi stanowi ważny etap w zrozumieniu mechanizmów działania tych chorób oraz może przyczynić się do wyłonienia potencjalnych nowych terapeutyków celowanych.

W niniejszej rozprawie doktorskiej postawiłem przed sobą dwa cele. Moim pierwszym zadaniem było opracowanie protokołu produkcji RNA o strukturze spinki, stabilizacja tego RNA do krystalizacji oraz wyznaczenie struktury przestrzennej RNA w formie spinki stabilnej w warunkach krystalizacyjnych. Drugie zadanie dotyczyło wyznaczenia struktury przestrzennej kompleksu RNA zawierającego sekwencję G_2C_4 z syntetyczną cząsteczką ANP77. Na podstawie tej struktury chciałem określić potencjał terapeutyczny cząsteczki ANP77 oraz czy ta cząsteczka mogłaby posłużyć do stabilizacji struktur RNA.

W ramach pierwszego zadania opracowałem protokół produkcji RNA w formie spinki do badań krystalograficznych. Krótkie cząsteczki RNA (o długości do 10 nukleotydów) mogą być wydajnie produkowane za pomocą syntezy chemicznej na podłożu stałym. Natomiast RNA dłuższy niż 40 nukleotydów powinien być syntetyzowany na drodze transkrypcji *in vitro* przy użyciu rybozymów lub modyfikowanej matrycy DNA. Zaprojektowałem szereg konstruktów związanych z chorobami neurodegeneracyjnymi, dla których chciałem określić strukturę przestrzenną. Konstrukty zawierające powtórzenia trójnukleotydowe CUG (biorące udział w rozwoju dystrofii miotonicznej) stabilizowałem poprzez cyrkularyzację, a ich potencjał krystalizacyjny chciałem zwiększyć poprzez oddziaływanie RNA z domeną oddziałującą z RNA białka U1a.

W ramach drugiego zadania udało mi się uzyskać struktury przestrzenne dla konstruktu G_2C_4 z i bez cząsteczki ANP77. Opisałem sposób fałdowania się RNA, oddziaływania RNA z ligandem ANP77 oraz oddziaływania cząsteczek w sieci

krystalicznej. Przy użyciu skaningowej kalorymetrii różnicowej w zakresie pH 5,2–7,0 określiłem poziom stabilizacji struktury G₂C₄ przez cząsteczkę ANP77.

Badania przedstawione w tej pracy ukazują potencjał RNA zawierającego sekwencję G₂C₄ do formowania tripleksów złożonych z trójek nukleotydowych C-C-C, oraz C-G-C. Stanowi to ważny element w rozwoju wiedzy na temat mechanizmów związanych z chorobami neurodegeneracyjnymi opisywanymi w pracy.

Abstract

Many human diseases, especially neurodegenerative and neuromuscular diseases, are caused by the expansion of nucleotide repeats known as microsatellites in various genes. Transcripts generated from these genes, due to the presence of repeated nucleotide sequences, can form different structures. Studying RNA structures associated with neurodegenerative diseases is an important step in understanding the mechanisms behind these diseases and may contribute to the discovery of potential new therapies.

I set two goals for myself. My first task was to develop a protocol to produce RNA hairpins, stabilization of RNA in the form of hairpin for crystallization and to determine the three-dimensional structure of RNA hairpins. The second task involved determination of the three-dimensional structure of RNA complex containing the G₂C₄ sequence with a synthetic molecule ANP77. Based on this structure, I wanted to assess the therapeutic potential of the ANP77 molecule and whether it could be used to stabilize RNA structures.

For the first task, I developed a protocol to produce RNA hairpins for crystallization studies. Short RNAs (up to 10 nucleotides in length) can be efficiently produced using solid-phase chemical synthesis. However, RNAs longer than 40 nucleotides should be synthesized *via in vitro* transcription using ribozymes or modified DNA templates. I designed a series of constructs related to neurodegenerative diseases for which I wanted to determine the crystal structure. Constructs containing CUG trinucleotide repeats (associated with the development of myotonic dystrophy) were circularized and I aimed to increase their crystallization potential through the interaction of RNA structures with the U1a RNA binding domain protein.

For the second task, I successfully obtained crystal structures for the G₂C₄ construct with and without the ANP77 molecule. I described the fold of RNA, RNA-ANP77 interactions and interactions between molecules in the crystal lattice. Using differential scanning calorimetry in the pH range of 5,2–7,0, I determined the level of stabilization of the G₂C₄ structure by the ANP77 molecule.

The research presented in this work reveals the potential of RNA containing G₂C₄ sequence to form triplexes with triplets of nucleotides C-C-C and C-G-C. This is an important contribution to our understanding of the mechanisms associated with neurodegenerative diseases described in the study.

1. Cel pracy

Celem mojej pracy było poznanie struktur przestrzennych RNA mających znaczenie w patogenezie chorób neurodegeneracyjnych. Przedmiotem moich badań były cząsteczki RNA, które w swojej sekwencji posiadały powtórzenia nukleotydów charakterystyczne dla chorób neurodegeneracyjnych. Wyróżnia się ponad 40 chorób, głównie dotyczących układu nerwowego, które wywoływane są przez ekspansję powtórzeń sekwencji nukleotydowych rozmieszczonych w genach człowieka. Na pierwszym miejscu pod względem częstości występowania są choroby spowodowane ekspansjami trójnukleotydowymi. W ostatnim czasie odkryto także, że przyczyną niektórych najczęstszych zaburzeń genetycznych obserwowanych przez neurologów są ekspansje trój-, cztero-, pięcio- a nawet sześci nukleotydowe. Jeden z proponowanych patomechanizmów związanych z nadmierną ekspansją nukleotydów zakłada, że białka niezbędne do prawidłowego funkcjonowania komórki mogą być sekwestrowane przez nadmiernie wydłużone pre-mRNA.

Wiadomym jest, że RNA zawierające nadmierne wydłużenia nukleotydowe mogą przyjmować różne struktury drugo- i trzeciorzędowe, w tym struktury spinki. Jednak krystalizacja struktur spinkowych jest problematyczna ze względu na faworyzowanie formowania się dupleksów nad spinkami w warunkach obecnych w kropli krystalizacyjnej. Z tego względu moim pierwszym zadaniem było opracowanie protokołu produkcji RNA o strukturze spinki, stabilizacja tych RNA do krystalizacji oraz wyznaczenie struktury przestrzennej zaprojektowanych cząsteczek RNA w formie spinki stabilnej w warunkach krystalizacyjnych. Drugie zadanie dotyczyło wyznaczenia struktury przestrzennej kompleksu RNA zawierającego sekwencję G₂C₄ z syntetyczną cząsteczką ANP77. Na podstawie tej struktury chciałem określić potencjał terapeutyczny cząsteczki ANP77 oraz czy ta cząsteczka mogłaby posłużyć do stabilizacji struktur RNA.

2. Wprowadzenie

2.1. Choroby neurodegeneracyjne

Choroby neurodegeneracyjne są grupą wrodzonych lub nabytych chorób układu nerwowego, które polegają na utracie komórek nerwowych. Wiele z nich charakteryzuje się tworzeniem specyficznych skupisk białek w komórkach [1]. Na podstawie tego jakiego białka ulegają agregacji, choroby neurodegeneracyjne można podzielić na: tautopatie, amyloidozy, α -synukleinoz, TDP-43 proteinoz [1–4]. Wspólnymi dla powyższych chorób neurodegeneracyjnych są takie procesy biologiczne jak: stres oksydacyjny, zapalenie nerwów i zaprogramowana śmierć komórki [5–7]. Mimo trwających starań w poszukiwaniu skutecznych podejść terapeutycznych oraz leków neuroprotektyjnych, choroby neurodegeneracyjne nadal pozostają nieuleczalne. Dostępne leczenie jest jedynie objawowe, które do tego jest kosztowne. W związku z tym pacjenci borykający się z tymi chorobami często umierają, gdyż z powodu wysokich kosztów opieki medycznej są zmuszeni do rezygnacji z dostępnych terapii [8,9]. Najnowsze dane wskazują, że choroby neurodegeneracyjne dotyczą aż 50 milionów ludzi na całym świecie, głównie w krajach o niskich i średnich dochodach [10]. Prognozy sugerują, że do roku 2050, ze względu na proces starzenia się społeczeństwa, liczba osób cierpiących na te choroby sięgnie 115,4 miliona [11].

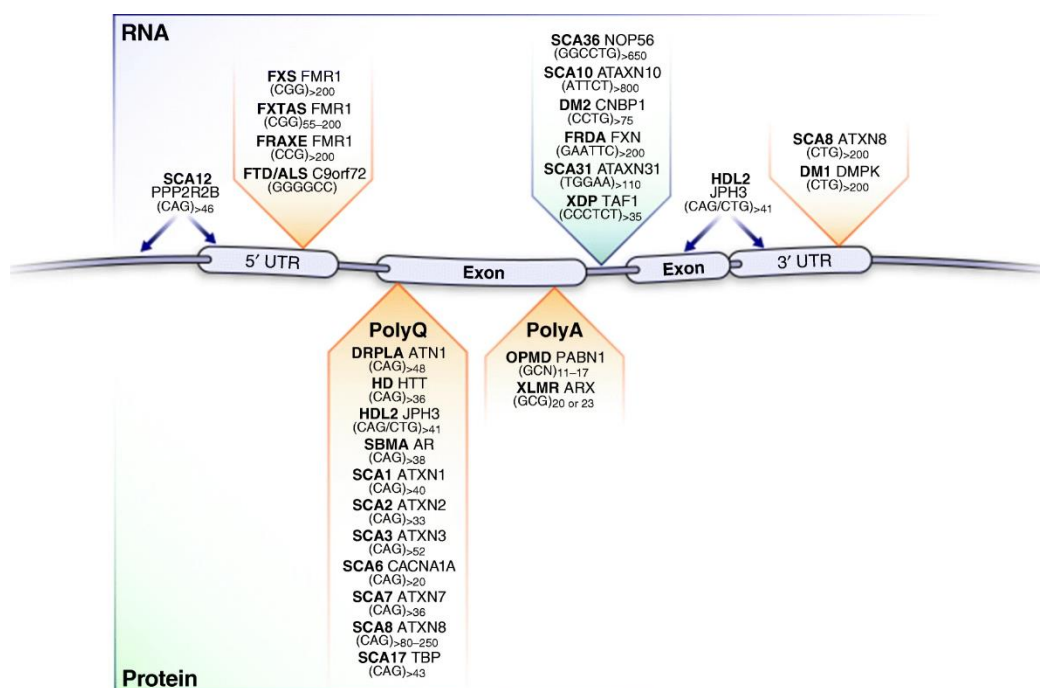
2.2. Choroby neurodegeneracyjne związane z nadmierną ekspansją nukleotydów

W genomie człowieka, jak również w genomach innych organizmów, zaobserwowano występowanie sekwencji, które powtarzają się wielokrotnie. Nie jest to jednorodny rodzaj powtarzającego się motywu, ponieważ odnotowano różnorodność w ich składzie nukleotydowym. Powtórzenia te stanowią około 50% całej sekwencji genomowej i najczęściej znajdują się w genach odpowiedzialnych za transkrypcję oraz sygnalizację wewnątrz- i międzykomórkową [12–14]. Wielokrotnie powtarzające się sekwencje podzielono na kilka kategorii, w tym:

- sekwencje rozproszone, w których powtarzające się elementy są oddzielone od siebie innymi sekwencjami,

- pseudogeny, czyli nieaktywne kopie genów,
- sekwencje tandemowe, gdzie motywy powtarzające się ułożone są obok siebie [14].

Wiele chorób człowieka, zwłaszcza chorób neurologicznych i neuromuskularnych, jest wywoływanych przez ekspansję powtórzeń nukleotydów zwanych mikrosatelitami [12,13,15]. Znajdują się one w różnych regionach genów, w tym w promotorach, regionach 5' lub 3' nieulegających translacji (UTR z ang. *untranslated region*), sekwencjach kodujących i intronach (Rycina 1) [13,16].



Rycina 1. Schemat lokalizacji powtórzeń nukleotydowych związanych z chorobami neurologicznymi [17].

W zależności od konkretnej choroby różni się próg, przy którym ekspansje powtórzeń nukleotydowych obecnych w DNA objawiają się u pacjenta jako objawy kliniczne (Tabela 1). Obecnie zidentyfikowano ponad 40 odrębnych chorób, które są spowodowane ekspansją różnych powtórzeń nukleotydowych w sekwencji DNA. W ostatnich trzech dekadach dokonano znaczącego postępu w określeniu mutacji/zmian genetycznych odpowiedzialnych za patogenezę wielu z tych chorób [13,17].

Tabela 1. Tabela przedstawiająca wybrane choroby neurodegeneracyjne związane z powtórzeniami nukleotydów w DNA.

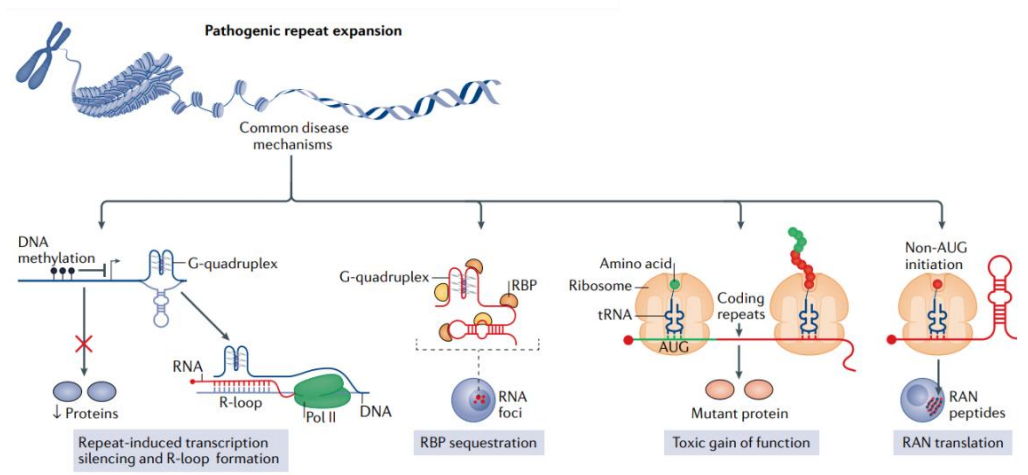
FXS/FRAXA – syndrom łamliwego chromosomu X (z ang. *Fragile X Syndrome*), DM – dystrofia miotoniczna (z ang. *Myotonic Dystrophy*), HD – choroba Huntingtona (z ang. *Huntington's Disease*), SCA – ataksja rdzeniowo-mózdkowa (z ang. *Spinocerebellar Ataxia*), DRPLA – zanik jądra zębatego, jądra czerwienego i jądra niskowzgórzowego (z ang. *Dentatorubral-Pallidoluysian Atrophy*), HDL2 – płasawica podobna do choroby Huntingtona 2 (z ang. *Huntington's Disease-Like 2*), C9FTD/ALS – ośłepienie czołowo-skroniowe i stwardnienie zanikowe boczne związane z genem *C9orf72* (z ang. *C9ORF72-Associated Frontotemporal Dementia and Amyotrophic Lateral Sclerosis*), SBMA – opuszkowo-rdzeniowy zanik mięśni (z ang. *Spinal-Bulbar Muscular Atrophy*), EPM1 – padaczka miokloniczna Unverrichta-Lundborga (z ang. *Progressive Myoclonus Epilepsy 1*), FXPOI – przedwczesne wygasanie funkcji jajników związane z zespołem łamliwego chromosomu X (z ang. *Fragile X-Associated Primary Ovarian Insufficiency*), FXTAS – zespół drżenia i ataksji związany z zespołem łamliwego chromosomu X (z ang. *Fragile X-Associated Tremor/Ataxia Syndrome*), OPMD – dystrofia mięśniowa oczno-gardłowa (z ang. *Oculopharyngeal Muscular Dystrophy*), FRDA - ataksja Friedreicha (z ang. *Friedreich Ataxia*) [18].

Choroba	Powtórzenie	Lokalizacja w genomie	Nazwa genu	Liczba powtórzeń u zdrowych osobników	Liczba powtórzeń w stanie patologicznym
FXS/ FRAXA	CGG	5' UTR	<i>FMRI</i>	6-55	200+
SBMA	CAG	Obszar kodujący	<i>AR</i>	9-36	38-62
DM1	CTG	3' UTR	<i>DMPK</i>	5-37	50-10000
HD	CAG	Obszar kodujący	<i>HTT</i>	10-35	35+
SCA1	CAG	Obszar kodujący	<i>ATXN1</i>	6-35	49-88
DRPLA	CAG	Obszar kodujący	<i>ATN1</i>	6-35	49-88
SCA3	CAG	Obszar kodujący	<i>ATXN3</i>	12-40	55-86
SCA2	CAG	Obszar kodujący	<i>ATXN2</i>	14-32	33-77

FRDA	GAA	Intron	<i>FXN</i>	8-33	90+
SCA6	CAG	Obszar kodujący	<i>CACNA1A</i>	4-18	21-30
EPM1	CCCCGC CCCCGC	Promotor	<i>CSTB</i>	2-3	30-80
SCA7	CAG	Obszar kodujący	<i>ATXN7</i>	7-17	38-120
OPMD	GCG	Obszar kodujący	<i>PABPN1</i>	6-10	12-17
SCA8	CTG	3' UTR	<i>ATXN8</i>	16-34	74+
SCA12	CAG	5' UTR	<i>PPP2R2B</i>	7-28	66-78
SCA10	ATTCT	Intron	<i>ATXN10</i>	10-20	500-4500
SCA17	CAG	Obszar kodujący	<i>TBP</i>	25-42	47-63
DM2	CCTG	Intron	<i>CNBP</i>	10-26	75-11000
FXTAS/ FXPOI	CGG	5' UTR	<i>FMR1</i>	6-55	55-200
HDL2	CTG/CAG	3' UTR/nić antysensowna	<i>JPH3</i>	<50	50+
SCA31	TGGAA	Intron	<i>TK2/BEAN</i>	0	45+
SCA36	GGCCTG	Intron	<i>Nop56</i>	3-14	650+
C9FTD/ALS	GGGGCC/ GGCCCC	5' UTR/nić antysensowna	<i>C9orf72</i>	2-25	25+

Istnieją cztery wzajemnie związane kategorie mechanizmów, które bezpośrednio przyczyniają się do patologii chorób wynikających z ekspansji powtórzeń nukleotydów [17]. Należą do nich mechanizmy oparte na DNA, które obejmują wyciszenie genów oraz tworzenie pętli R, co prowadzi do aktywacji odpowiedzi związanej z uszkodzeniami DNA lub inhibicji transkrypcji. Dodatkowo istnieją mechanizmy mające związek

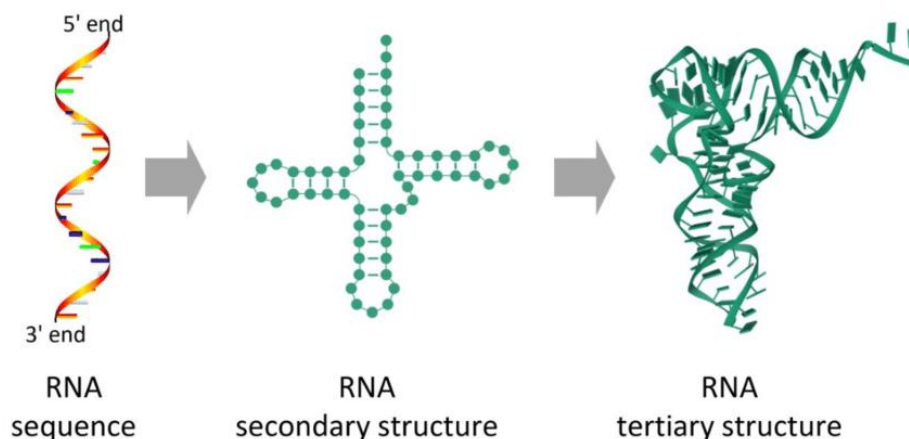
z transkrybowanym RNA. RNA formując skomplikowane struktury drugo- i trzeciorzędowe, mogą w nadmierny sposób oddziaływać z białkami wiążącymi RNA (RBPs, z ang. *RNA binding proteins*) tworząc agregaty RNA-białko. Ekspansja powtórzeń nukleotydów w sekwencjach kodujących przyczynia się do powstawania zmutowanych białek, które posiadają zmienione funkcje w porównaniu do niezmutowanego białka (zyskiwanie funkcji z ang. *gain of function* oraz strata funkcji z ang. *loss of function*). Wreszcie powtórzenia RNA, które nie zawierają kodonu inicjacji translacji AUG mogą podlegać niestandardowej translacji związanej z powtórzeniami nieopartej na kodonie AUG (RAN, z ang. *repeat-associated non-AUG translation*), co prowadzi do tworzenia toksycznych peptydów zawierających powtórzenia aminokwasów (Rycina 2) [12].



Rycina 2. Mechanizmy przyczyniające się do patologii chorób wynikających z ekspansji powtórzeń nukleotydowych [12]

2.3. Znaczenie struktury RNA w biologii

W przeciwieństwie do podwójnej helisy DNA, RNA najczęściej występuje jako jednoniciowy łańcuch, który może tworzyć struktury spinkowe, pętle wewnętrzne, wypukłości i pseudowęzły. Zarówno motywy strukturalne drugorzędowe, jak i trzeciorzędowe pełnią istotną rolę jako elementy kierujące interakcjami RNA-RNA oraz RNA-białko (Rycina 3) [19,20].



Rycina 3. Schemat przedstawiający przykładową strukturę pierwszorzędową, drugorzędową i trzeciorzędową RNA [21].

Odpowiednie formowanie struktury RNA jest ważnym elementem wpływającym na jego funkcję, co zostało zauważone podczas badań nad rRNA i tRNA [22]. Związek między strukturą a funkcją nie jest ograniczony do niekodujących RNA. Większość eukariotycznych RNA kodujących białka (mRNA, z ang. *messenger RNA*) poddawanych jest procesowi wycinania intronów/składania genów (z ang. *splicing*) przed translacją [23,24], a struktury drugorzędowe w rejonie pomiędzy intronem i egzonem mogą określać dokładne miejsca usunięcia intronu [25–29]. Co więcej, wybór miejsc cięcia zależy jest od wzmacniaczy lub wyciszaczy, co również jest uwarunkowane kontekstem strukturalnym [20]. Istotną cechą struktur drugorzędowych RNA jest ich dynamiczny charakter, dzięki któremu możliwe jest przyjmowanie wielu struktur o podobnej energii swobodnej, umożliwiającą szybkie zmiany konformacyjne. Klasycznymi przykładami dużych zmian strukturalnych są zmiany w strukturze rybosomu, które zachodzą podczas translacji [22] oraz zmiany konformacji ryboprzełączników [30]. Formowanie prawidłowej struktury RNA ma istotne znaczenie dla jego funkcji, a błędy w formowaniu struktury mogą prowadzić zarówno do utraty funkcji, jak i nadmiernego nabywania funkcji [12,13,18,31]. Mutacje w miejscach kluczowych dla prawidłowego formowania się struktury mogą skutkować zmianą w rozpoznawaniu RNA przez białka regulujące. Zmieniony stan równowagi formowania struktury prowadzi następnie do zaburzenia procesów komórkowych. Przykładem są mutacje w pre-mRNA białka tau, które osłabiają strukturę pętli spinki na granicy egzonu 10 i intronu 10, co zmienia jego interakcję z kompleksem U1 małej jądrowej nukleoproteiny (snRNP z ang. *small nuclear ribonucleoprotein*) i powoduje zaburzenie alternatywnego cięcia pre-mRNA [32].

W regionach dwuniciowych, RNA przyjmuje konformację helisy typu A, która charakteryzuje się mniej zwartą strukturą w porównaniu do helisy typu B DNA (11 par zasad na obrót helikalny w porównaniu do 10,5 par zasad). W tym typie helisy RNA występuje głębsza bruzda większa i szersza bruzda mniejsza. Grupa hydroksylowa przy atomie węgla 2' RNA wpływa na inną konformację cukru, stan uwodnienia i stabilność termodynamiczną w porównaniu do DNA. Tylko 60-70% par zasad stanowi klasyczne parowanie Watsona-Cricka. RNA tworzy również niekanoniczne pary, np. Hoogsteena czy pary Wobble, co przyczynia się do jego różnorodności strukturalnej i funkcjonalnej [20]. Zwijanie się RNA ma charakter hierarchiczny, gdzie sekwencja pierwotna określa elementy struktury drugorzędowej poprzez wpływ sąsiednich nukleotydów [33]. Najczęstsze struktury drugorzędowe, które tworzą łańcuchy RNA to regiony sparowanych zasad, pętle struktury spinkowej, pętle wewnętrzne, wypukłości [33], pseudowęzły, zakręty (z ang. *kink-turns*) [34] oraz G-kwadupleksy [35]. Motywy struktur drugorzędowych mogą następnie stabilizować się wzajemnie poprzez zwijanie w bardziej skomplikowane wzory trzeciorzędowe [36]. Warto zauważyć, że zarówno elementy struktury drugorzędowej, jak i trzeciorzędowej mogą być dynamiczne i przekształcać się w zależności od obecności różnych białek, małych cząsteczek i warunków środowiska [37], co wpływa następnie na funkcję takiego RNA [38,39].

Wiele z opisanych wyżej funkcjonalnych elementów struktury drugorzędowej może spontanicznie ulec zmianie na skutek powstania mutacji w sekwencji DNA. Zmiany te mogą wynikać zarówno z mutacji punktowych, delecji, insercji, jak i nadmiernej ekspansji nukleotydów w DNA. Nowo powstałe RNA, na matrycy zmutowanych fragmentów DNA, mogą inicjować patologiczne procesy w komórkach. Przykłady takiego zyskiwania funkcji obejmują sekwestrację białek wiążących RNA, aktywację innych od standardowych miejsc splicingowych, tworzenie wewnętrznych miejsc wiązania rybosomu (IRES, z ang. *internal ribosome entry site*) i następującą po tym translację niezależną od czapeczki [16], czyli translację bez kodonu AUG [40,41].

2.4. Struktury spinkowe RNA w badaniach krystalograficznych

Struktura spinkowa to powszechny motyw w cząsteczkach RNA, gdyż prawie 70% nukleotydów tworzy właśnie takie struktury w 16S rRNA [42]. Mimo, że spinka RNA charakteryzuje się prostą architekturą złożoną z trzonu (z ang. *stem*) i pętli

wierzchołkowej (z ang. *apical loop*), to jednak wykazuje znaczną złożoność strukturalną. Głównym czynnikiem różnicującym są trzony spinek, które mogą być w pełni komplementarne i dwuniciowe lub zawierać wybrzuszenia i wewnętrzne pętle [43]. Pętla wierzchołkowa prezentuje zróżnicowaną architekturę w zależności od zawartości nukleotydów (Rycina 4). To właśnie te składniki determinują różnorodność strukturalną i funkcję spinek RNA.



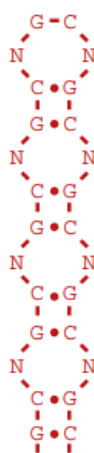
Rycina 4. Schemat drugorzędowej struktury spinek RNA [44].

Bogactwo różnorodnych struktur i funkcji, które charakteryzują struktury spinek RNA, umożliwia ich wiele zastosowań. Mogą to być syntetyzowane w laboratorium aptamery, czyli oligonukleotydy o wysokim powinowactwie i selektywności do konkretnych cząsteczek wykorzystywane w biologii molekularnej i medycynie (terapia i diagnostyka) [45–49].

Struktury spinek RNA mogą działać jako niezależne elementy uczestnicząc w inhibicji ekspresji genów w bakteriach i chroniąc transkrypty przed degradacją, a także biorąc udział w subkomórkowej lokalizacji mRNA. Spinki RNA są ważnymi składnikami ryboprzełączników, platform interakcji z czynnikami białkowymi, prekursorami mikroRNA i rybosomów. W niektórych przypadkach ich zdolność do fałdowania może jednak prowadzić do negatywnych konsekwencji, takich jak dezorganizacja procesów komórkowych, co może przyczyniać się do rozwoju chorób neurodegeneracyjnych [42,50–53].

Struktury spinek RNA stanowią istotny obiekt badań z powodu ich biologicznego znaczenia. Powtórzenia trójnukleotydowe takie jak powtórzenia sekwencji CNG (N oznacza dowolny nukleotyd), związane z wieloma chorobami neurodegeneracyjnymi mogą tworzyć struktury spinki.

Grupa prof. dr hab. Włodzimierza Krzyżosiaka z Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk podjęła się określenia struktury drugorzędowej RNA zawierających powtórzenia sekwencji CNG. RNA zawierające te powtórzenia mogą formować regiony jednoniciowe, jak również dwuniciowe. Dłuższe sekwencje powtórzeń CNG tworzą autonomiczne struktury przypominające spinki [54]. Wszystkie struktury spinek CNG układają się w podobny sposób. Ich trzon składa się z dwóch bloków utworzonych przez pary zasad Watsona-Cricka: C-G i G-C, oddzielone niestandardowymi parami nukleotydów N-N (takimi jak A-A, C-C, G-G, U-U) (Rycina 5) [16,55]



Rycina 5. Struktura drugorzędowa formowana przez powtórzenia CNG [16]

Pętla apikalna będąca końcową częścią struktury spinki CNG składa się z czterech lub siedmiu nukleotydów. Pętla ta rozpoczyna się i kończy resztami, które w trzonie struktury tworzą pary N-N. Obserwowano, że rozmiar pętli zależy od liczby powtórzeń trójnukleotydowych CNG składających się na daną strukturę typu spinki. W przypadku parzystej liczby powtórzeń obserwuje się pętle składające się z czterech nukleotydów, podczas gdy nieparzysta liczba powtórzeń skutkuje pętlą siedmionukleotydową. Inną istotną cechą powtórzeń trójnukleotydowych CNG jest zdolność do tworzenia "poślizgniętych" struktur (z ang. "*slippery hairpins*"). Oznacza to, że ta sama sekwencja trójnukleotydowa może tworzyć alternatywne struktury spinki, różniące się liczbą par w trzonie. Niezaangażowane strukturalnie sekwencje trójnukleotydowe CNG tworzą wystające końce 5' lub 3' w strukturach jednoniciowych. Dystrybucja tych poślizgniętych form zależy od stężenia jonów magnezu oraz temperatury [55,56].

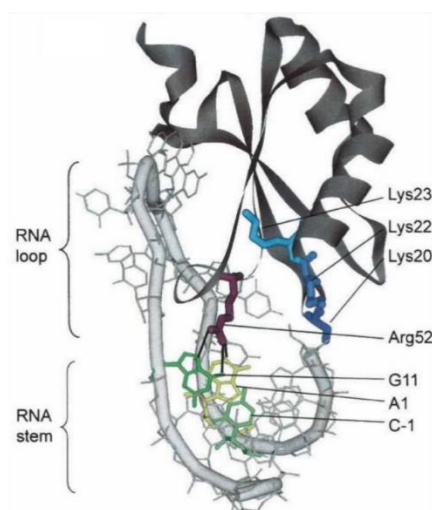
Dużą trudność sprawia otrzymanie kryształów RNA w formie spinki. Głównym powodem jest fakt, że spinki RNA są niestabilne termodynamicznie w wysokich stężeniach RNA i soli, czyli w warunkach obecnych podczas krystalizacji. W takich warunkach struktury te mają skłonność do tworzenia dupleksów nad spinkami [57,58]. Dlatego w ostatnich latach, w badaniach krystalograficznych, intensywnie badane były formowane przez powtórzenia nukleotydowe dupleksy jako modele trzonu spinki RNA (Tabela 2). Chociaż pary C-G i G-C dominują w większości struktur zawierających sekwencje powtórzone, niestandardowe pary zasad mogą także wpływać na kształt helisy RNA zaburzając jej standardową formę typu A. Jest to widoczne w przypadku powtórzeń CGG i CAG, które formują między sobą pary N-N złożone z puryn [59–64]. Pary puryna-puryna mają większe wymagania steryczne i bardziej wpływają na strukturę RNA niż pary pirymidyna-pirymidyna. Dla obu tych powtórzeń w dupleksach znajdują się reszty wykazujące rotację wokół wiązania O5'-C5'. Rotację tę można określić za pomocą kątów torsyjnych α i γ . Kąt α różni się o prawie pół skrętu od typowej wartości – 68°, podczas gdy kąt γ odbiega o około 120° od standardowej wartości 58°. Dla powtórzeń CGG ten efekt jest miejscowy i jest kompensowany w innych częściach struktury, gdzie następuje zwiększenie kąta skrętu helisy, co skutkuje średnią wartością tego parametru około 31° i jest nieco mniejszy niż dla typowej formy A-RNA (32,7°) [59,60,64]. Dla powtórzeń CAG efekt jest globalny. Helisa rozwija się, a bruzda większa poszerza się (>20 Å). Średni kąt skrętu helisy dla powtórzeń CAG wynosi tylko 28,4° [61–64]. Pary C-C w powtórzeniach CCG przede wszystkim wpływają na strukturę poprzez indukowanie przesunięcia nici [64,65]. Wartości parametrów helikalnych są podobne do formy A-RNA, co obserwuje się również w przypadku powtórzeń CUG [64,66–69]. Kąt skrętu helisy wynosi 33° dla zarówno dupleksów zawierających powtórzenia CCG, jak i CUG [64].

Tabela 2. Zestawienie struktur krystalicznych RNA zawierających powtórzenia trójnukleotydowe z ich kodami PDB.

Kod PDB struktury przestrzennej	Powtórzenie
4YN6, 3NJ6, 3NJ7, 4J50,	CAG
4E58, 4E59	CCG
3R1C, 3R1D, 3R1E, 3SJ2,	CGG
4PCJ, 1ZEV, 3GLP, 3GM7, 3SYW, 3SZX, 4E48, 4FNJ	CUG

2.5. Produkcja RNA do badań krystalograficznych

Badania strukturalne cząsteczek RNA rozpoczynają się od zaprojektowania konstruktów. Podczas projektowania konstruktów należy ustalić sekwencję pożądanego RNA, która może być dowolnie modyfikowana. Aby poprawić stabilizację struktury RNA lub ulepszyć potencjał sekwencji do krystalizacji można ją modyfikować poprzez dodawanie nukleotydów tworząc lepkie lub tępe końce. Drugorzędowa struktura cząsteczki RNA może zostać oceniona wykonując predykcję za pomocą narzędzi takich jak użyty w tej pracy program RNAStructure [70]. Aby wzmocnić potencjał krystalizacyjny, do struktury zaprojektowanego RNA można także dodać pętlę posiadającą sekwencję oddziałującą z fragmentem wiążącym się z RNA białka U1a (U1a-RBD, z ang. *U1a RNA binding domain*). Podejście to również zostało przeze mnie wykorzystane w niniejszej pracy. Białko U1a jest jednym z kluczowych białek zaangażowanych w proces *splicingu* pre-mRNA. Białko to jest częścią kompleksu snRNP, a dokładniej znajduje się w kompleksie U1 snRNP [71]. Fragment białka U1a wiążący się z RNA składa się z 98 aminokwasów i łączy się z konkretną pętlą struktury spinkowej RNA z wysokim powinowactwem (Rycina 6) [71,72]. U1a-RBD przyjmuje kształt kulisty o zwartym składzie zarówno w obecności, jak i w nieobecności RNA [72,73]. Ta właściwość sprawia, że U1a-RBD jest szczególnie odpowiednie do tworzenia wysoko uporządkowanych kryształów [74].



Rycina 6. Oddziaływanie fragmentu wiążącego RNA białka U1a z spinką II U1snRNA [71].

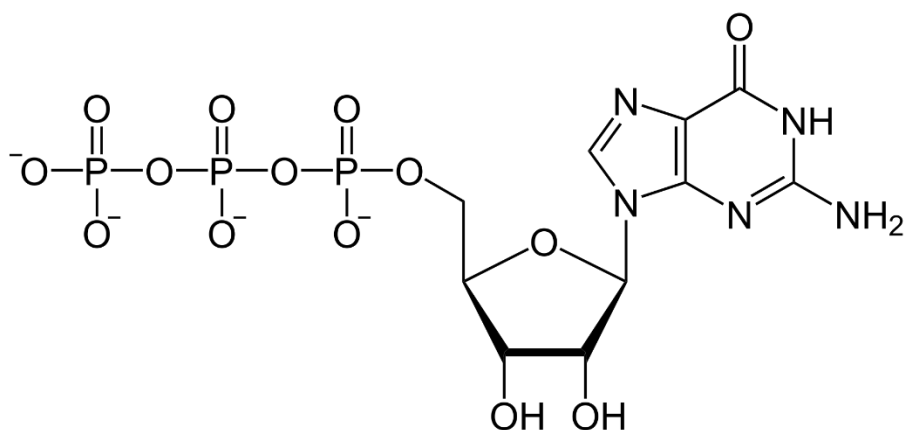
Chociaż opracowano wiele protokołów umożliwiających przygotowanie RNA do krystalizacji, uzyskanie wystarczającej ilości homogennego RNA stanowi wyzwanie [73,75–77]. RNA można pozyskać na trzy sposoby: oczyszczanie ze źródeł biologicznych, synteza chemiczna lub synteza enzymatyczna poprzez transkrypcję *in vitro*. Izolacja z komórek jest odpowiednia dla złożonych lub długich cząsteczek RNA (rybosomów lub tRNA), w wyniku czego natywne próbki mają wszystkie ważne modyfikacje potranskrypcyjne. Krótszy RNA można syntetyzować chemicznie lub enzymatycznie. Metody te są bardziej uniwersalne i preferowane do badań strukturalnych cząsteczek RNA ze względu na ilość pozyskiwanego RNA. O ile synteza chemiczna pozwala na produkcję dużych ilości RNA, to jest odpowiednia tylko dla stosunkowo krótkich oligomerów RNA. Dłuższy RNA można skutecznie syntetyzować przy użyciu transkrypcji *in vitro*, ale metoda ta charakteryzuje się uzyskaniem RNA o znacznej heterogenności na końcach 3' i ma specyficzne wymagania dotyczące sekwencji na końcu 5' [78].

Czystość i jednorodność próbki mają kluczowe znaczenie dla powodzenia eksperymentów krystalizacji oraz dyfrakcji [79]. Zarówno synteza chemiczna jak i transkrypcja *in vitro* nie są wydajne w 100%, dlatego po syntezie należy oddzielić pożądaną produkt od innych produktów tych reakcji o innej długości niż pożądana. Do tego celu standardowo używa się elektroforezy w żelu poliakrylamidowym (PAGE, z ang. *Polyacrylamide Gel Electrophoresis*) lub wysokosprawnej chromatografii cieczowej (HPLC z ang. *High-Performance Liquid Chromatography*). Obie te metody polegają na rozdzieleniu RNA nad podstawie jego długości [80,81].

2.6. Transkrypcja *in vitro*

Transkrypcja *in vitro* jest powszechną metodą stosowaną do syntezy cząsteczek RNA. Wykorzystuje matrycę DNA z sekwencją promotora polimerazy RNA T7, po której następuje fragment kodujący sekwencję docelową. Enzym wiąże matrycę w regionie promotora i rozpoczyna syntezę RNA. Wydłużenie łańcucha RNA kończy się, gdy enzym odpada z końca 3' matrycy DNA [82]. Transkrypcja *in vitro* może być wydajna, dając miligramy próbki RNA, ale wysoka wydajność zależy od kilku czynników. Najważniejszy jest rodzaj nukleotydów zlokalizowanych bezpośrednio za promotorem. Transkrypt powinien zaczynać się od co najmniej jednej reszty

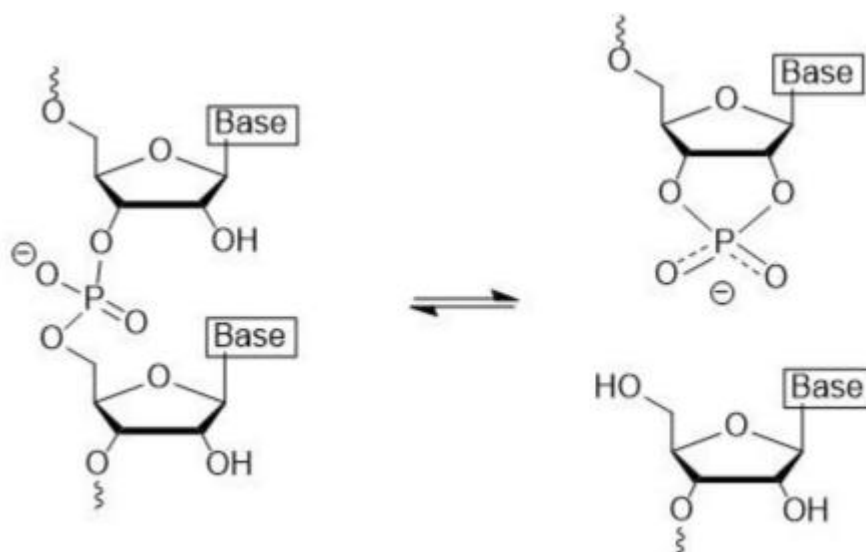
guanozyny, ale lepiej jest, gdy są obecne dwie lub trzy kolejne guanozyny [82–84]. W badaniach strukturalnych poważną wadą produkcji RNA za pomocą transkrypcji *in vitro* jest obecność transkryptów, w których powstały mutacje z powodu poślizgu polimerazy oraz transkryptów z nukleotydami dodanymi przez polimerazę RNA T7 na końcu 3' RNA, które nie są kodowane przez matrycę DNA [85]. Oddzielenie krótkich produktów od oligomerów pełnej długości jest zwykle łatwe. Wyzwanie stanowi oddzielenie transkryptów wydłużonych o dodatkowy jeden lub dwa nukleotydy. Ilości produktów N+1 i N+2, gdzie N oznacza długość pożądanego sekwencji RNA, wahają się od 10 do 50% całej syntezy, zmniejszając jednorodność, wydajność i potencjał krystalizacji próbki RNA [79,86]. Transkrypcję *in vitro* prowadzi się z użyciem mieszaniny trifosforanów nukleozydów (NTP), w wyniku czego RNA posiada trifosforan na końcu 5' (Rycina 7). W niektórych przypadkach może to negatywnie wpływać na potencjał krystalizacji RNA poprzez zapobieganie upakowaniu cząsteczek RNA w sieć krystaliczną. W związku z tym, w celu poprawy krystalizacji i jakości kryształów, konieczna może być konwersja 5'-trifosforanu w grupę hydroksylową. Tę konwersję można łatwo przeprowadzić przez transkrypcję *in vitro* w obecności reszt guanozyny lub po transkrypcji wykonać reakcję z użyciem fosfatazy alkalicznej. Jeśli do dalszej modyfikacji RNA wymagana jest grupa 5'-monofosforanowa, reakcję transkrypcji można wykonać w obecności monofosforanu 5'-guanozyny (GMP) [86,87].



Rycina 7. Trifosforan 5'- guanozyny

2.7. Rybozomy

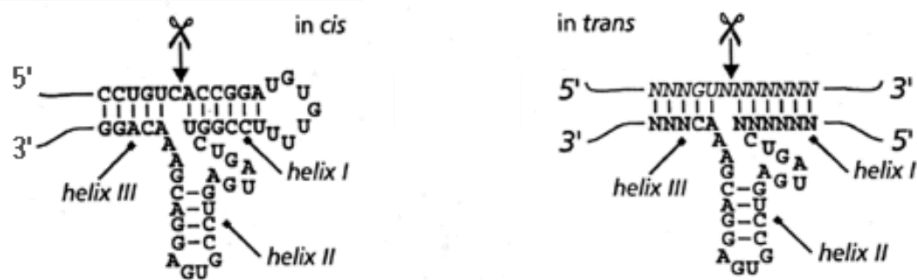
Aby rozwiązać problemy wynikające z heterogenności produktu transkrypcji *in vitro*, można użyć rybozymów. Rybozomy są klasą katalitycznych RNA [88–90]. Znajdują się głównie w wirusach atakujących rośliny i odgrywają kluczową rolę w procesie replikacji materiału genetycznego za pomocą mechanizmu toczącego się koła. Podczas procesu replikacji powstają długie, liniowe konkatamery genomu wirusa. Następnie te konkatamery są transestryfikowane przez domenę katalityczną rybozemu na jednostki o długości pojedynczych genów [89,91,92]. W tej rozprawie doktorskiej użyłem dwóch typów rybozymów: typu głowa młotka (z ang. *hammerhead*) i typu spinki (z ang. "*hairpin*"). Każdy typ rybozemu posiada swoje charakterystyczne cechy strukturalne i mechanizmy katalizy [93]. Rybozomy używane do produkcji homogennych końców RNA katalizują reakcję transestryfikacji, co prowadzi do powstania grupy hydroksylowej na końcu 5' i 2'3' cyklicznego fosforanu (Rycina 8). Do osiągnięcia optymalnej aktywności rybozomy zazwyczaj wymagają jonów metali dwuwartościowych, takich jak Mg^{2+} [90].



Rycina 8. Schemat hydrolizy przeprowadzanej przez rybozomy [78].

Rybozomy mogą działać w dwóch trybach: *cis* i *trans* (Rycina 9). W trybie *cis* substrat i rybozym występują w jednym łańcuchu RNA, a reakcja transestryfikacji prowadzi do podziału RNA na dwie różne części - domenę katalityczną rybozemu i RNA docelowe. Natomiast w trybie *trans* rybozomy rozpoznają substrat poprzez hybrydyzację z komplementarną sekwencją i następnie dokonują reakcji transestryfikacji RNA [78].

Ponadto każdy typ rybozymu ma swoje specyficzne wymagania, zwłaszcza jeśli chodzi o rozpoznawanie transestryfikowanej sekwencji RNA. Opisane zdolności katalityczne rybozymów są obiecującym obszarem badań, jednak dalsze poszerzenie wiedzy na ten temat jest niezbędne, aby w pełni zrozumieć ich potencjalne zastosowania i mechanizmy działania [78].

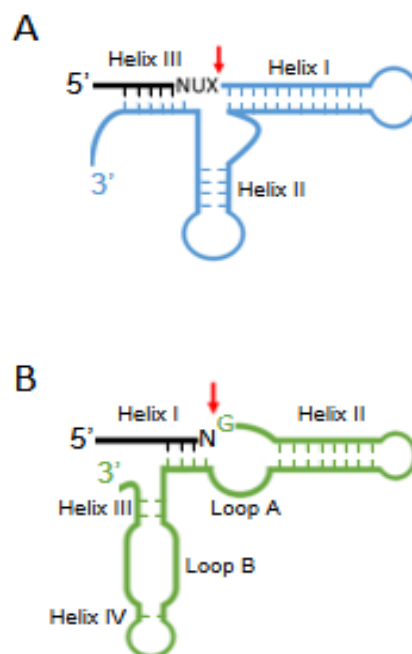


Rycina 9. Działanie rybozymów w trybie *cis* i *trans* na przykładzie rybozymu typu *głowa młotka* [94].

Rybozomy, ze względu na swoje stosunkowo niewielkie rozmiary, są atrakcyjnymi narzędziami do zastosowania w badaniach strukturalnych. Umożliwiają produkcję jednorodnych końców cząsteczek RNA oraz zwiększenie efektywności transkrypcji *in vitro* [73,76]. W pierwszym przypadku rybozomy są wbudowywane w matrycę DNA powyżej oraz/lub poniżej sekwencji docelowej. W trakcie transkrypcji *in vitro* łańcuch RNA jest transestryfikowany, co powoduje powstanie mieszanki RNA docelowego oraz domeny lub domen katalitycznych rybozymów. Następnie do oddzielenia i oczyszczenia tych cząsteczek wykorzystuje się metody takie jak elektroforeza lub HPLC [95,96]. Kiedy rybozym znajduje się na pozycji dalszej względem sekwencji docelowej, koniec 3' transkryptu RNA jest jednorodny, ale posiada grupę 2',3' cyklicznego fosforanu (Rycina 8). Jeśli grupa 3' OH jest potrzebna do dalszych eksperymentów, grupa 2',3' cyklicznego fosforanu może zostać usunięta za pomocą kinazy polinukleotydowej T4 [97]. Natomiast umieszczenie rybozymu na końcu 5' RNA docelowego zwiększa wydajność transkrypcji *in vitro* lub pozwala na produkcję RNA o jednorodnych końcach 5'. W celu osiągnięcia wyższej efektywności syntezy RNA za pomocą polimerazy RNA T7, na początku domeny katalitycznej rybozymu wprowadza się sekwencję GGGAGA, która zapewnia wysoką procesywność enzymu. [98]. Takie podejście jest szczególnie użyteczne, gdy sekwencja RNA docelowego jest kluczowa i nie może być dostosowana do wymagań polimerazy RNA T7. W przypadkach, gdy jednorodność i wysoka wydajność syntezy RNA są istotne

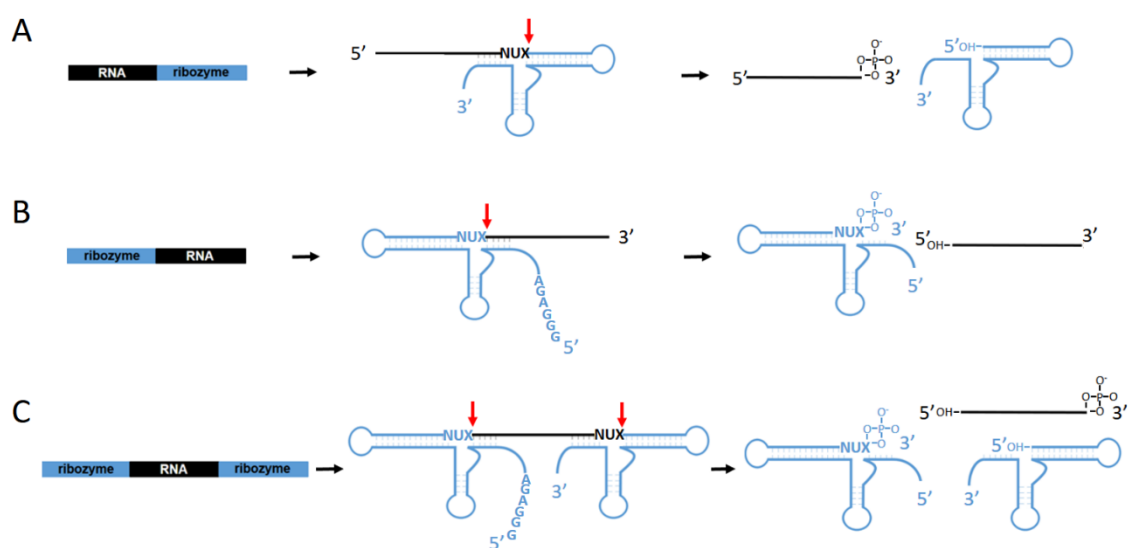
dla udanych badań krystalograficznych i innych eksperymentów, warto rozważyć konstrukcje RNA z rybozymami na obu końcach [73].

Rybozym typu głowa młotka jest jednym z najlepiej opisanych motywów katalitycznych, spotykanym głównie w wiroidach i RNA satelitarnym [99]. Jest stosunkowo niewielki, gdyż składa się z 40–50 nukleotydów. Jego struktura drugorzędowa składa się z trzech helis rozgałęziających się w regionie łączącym (Rycina 10A). W przestrzeni rybozym ten przypomina literę "Y" i jest stabilizowany przez interakcje pętla–pętla [100,101]. Miejsce transestryfikacji znajduje się na końcu 3' docelowego RNA tuż obok helisy I. Transestryfikacja RNA zachodzi po określonym motywie NUX (N oznacza dowolny nukleotyd, X może być dowolnym nukleotydem z wyjątkiem guaniny) (Rycina 10). Optymalna sekwencja transestryfikacji to AUC, GUC lub UUC, co zapewnia wysoką specyficzność aktywności rybozemu [94,102]. Dodatkowym czynnikiem wpływającym na efektywność transestryfikacji jest obecność dwuwartościowych kationów, takich jak Mg^{2+} lub Mn^{2+} , w stężeniach milimolowych. Jony metalu są zaangażowane w odpowiednie ułożenie rybozemu i pełnią rolę kofaktorów podczas reakcji transestryfikacji [103].



Rycina 10. Struktura drugorzędowa którą przyjmują rybozemy: A – 3' typu głowa młotka, B – 3' typu spinki [78]. Czerwonymi strzałkami zaznaczyłem miejsca transestryfikacji.

Rybozym typu głowa młotka znalazł szerokie zastosowanie w generowaniu RNA o jednorodnych końcach. Wykazano, że rybozym ten umieszczony zarówno na 5', jak i 3' końcu docelowego RNA skutecznie przeprowadza transestryfikację RNA o długości od 36 do 140 nukleotydów. Optymalna aktywność osiągana jest przy stężeniu jonów Mg^{2+} wynoszącym 24 mM (Rycina 11) [73]. Z powodzeniem zastosowano go w badaniach krystalograficznych, między innymi do syntez ludzkiego RNA 7SL (Protein Data Bank—PDB: 1L9A), domeny IV RNA 4.5S (PDB: 1DUH) oraz tRNA^{Ile} (PDB: 6UFH) [104–106].



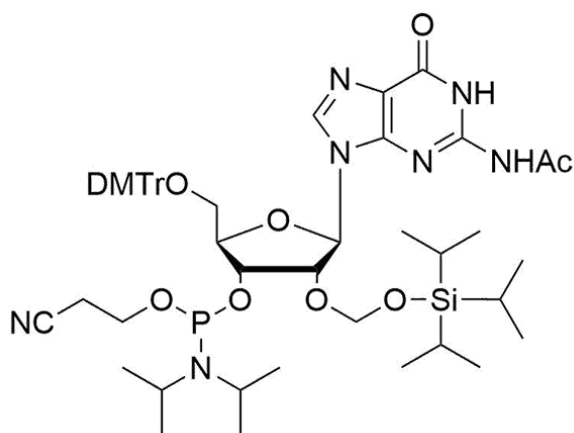
Rycina 11. Schemat reakcji transestryfikacji konstruktyw przez rybozym: A – typu głowa młotka na końcu 3' RNA docelowego, B – typu głowa młotka na końcu 5' RNA docelowego, C – typu głowa młotka jednocześnie na końcach 5' i 3' RNA docelowego [78]. Czerwonymi strzałkami oznaczyłem miejsca transestryfikacji.

Rybozym typu spinki został zidentyfikowany w łańcuchu antysensownym RNA satelitarnego wirusa przyczepkowego tytoniu. Posiada on cztery regiony helikalne oraz dwie pętle wewnętrzne (Rycina 10B) [107,108]. Podobnie jak w przypadku rybozymu typu głowa młotka, regiony helikalne rybozymu typu spinki mogą być modyfikowane pod względem długości i sekwencji bez wpływu na efektywność reakcji transestryfikacji RNA. Miejsce transestryfikacji znajduje się w jednej z pętli wewnętrznych. Optymalna sekwencja rozpoznawanego RNA przez rybozym typu spinki jest oznaczana jako RYN↓GUC (gdzie R reprezentuje purynę, Y reprezentuje pirymidynę, a N oznacza dowolny nukleotyd), a reakcja hydrolizy zachodzi między nukleotydami N i G [88]. Ze względu na ograniczenia sekwencyjne w porównaniu z rybozymem typu głowa młotka, rybozym typu spinki został zaproponowany

jako alternatywny motyw katalityczny w moich badaniach. Pomimo wysokiej efektywności reakcji transestryfikacji, wykazuje on tendencję do generowania produktów nieswoistych [73].

2.8. Synteza chemiczna

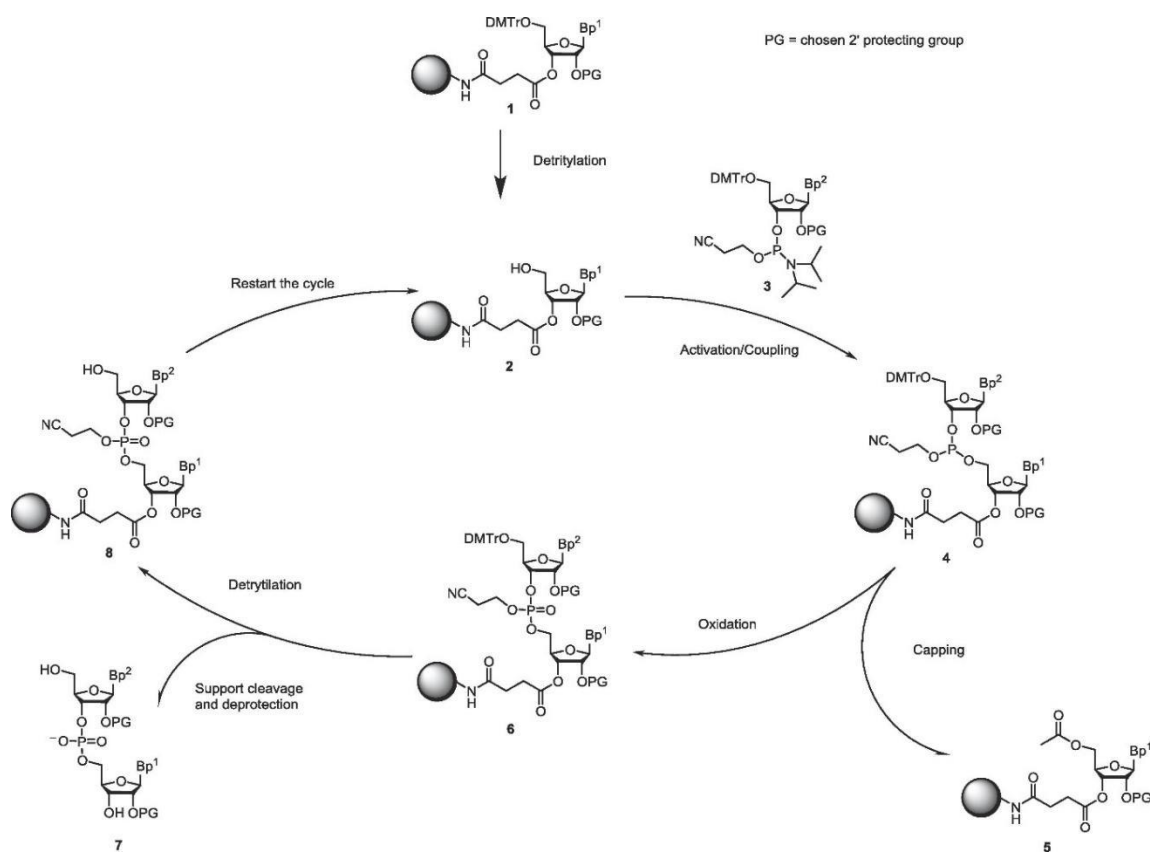
Łańcuchy RNA o długości poniżej 40 nukleotydów mogą być produkowane na drodze syntezy chemicznej. Synteza chemiczna w fazie stałej z wykorzystaniem zautomatyzowanego syntetyzera została opracowana przez Bruce'a Merrifielda i jest stosowana do syntezy kwasów deoksyrybonukleinowego i rybonukleinowego [109,110]. Wykorzystuje amidofosforyny posiadające różne grupy zabezpieczające, które blokują ich reaktywność podczas syntezy chemicznej [111]. Istnieje kilka rodzajów komercyjnie dostępnych amidofosforynów. Większość z nich posiada łagodną ochronę grupy aminowej zasady azotowej i ugrupowania fosforanowego blokowanego przez 2-cyanoetylo-N,N,N',N'-tetra-izopropyl (Rycina 12) [111].



Rycina 12. Amidofosforyn używany do syntezy chemicznej na podłożu stałym.

W przypadku grupy 2' OH najpowszechniej stosowanymi klasami grup blokujących są: triizopropylsililooksymetyl (TOM) i tert-butyldimetylosilil (TBDMS). Są to grupy usuwane po syntezie przy użyciu fluorku tetra-n-butyloamoniowego (TBAF) lub trifluorowodorku trietyloaminy (TEA·3HF). Grupa 5' OH jest zwykle blokowana przez 4,4-dimetoksytrytyl (DMTr) [112]. Wybór grupy zabezpieczającej grupę 2' OH rybozy jest najbardziej istotnym etapem wpływającym na wydajność i czas syntezy RNA [113]. Synteza chemiczna na podłożu stałym opiera się na cyklicznym wydłużaniu

łańcucha DNA/RNA na podłożu stałym, np. szkle o kontrolowanych rozmiarach porów lub silnie usieciowanym polistyrenie. Standardowa synteza w fazie stałej zachodzi od końca 3' do 5', a jeden cykl elongacji powoduje wydłużenie łańcucha o jeden nukleotyd. Każdy cykl obejmuje cztery etapy: detrytylację, sprzężanie, zamykanie i utlenianie (Rycina 13) [111]. Cykl rozpoczyna się od usunięcia grupy zabezpieczającej DMTr z grupy 5' OH amidofosforynu przyłączonego do stałego podłoża. Po tym następuje reakcja sprzężania, w wyniku której dochodzi do wydłużenia łańcucha RNA, który jest zablokowany na końcu 5' grupą DMTr pochodzącą z ostatniego dodanego amidofosforynu. Ponieważ wydajność sprzężania jest mniejsza niż 100%, nie wszystkie łańcuchy RNA są wydłużone. Z tego względu grupa 5' jest blokowana grupą acylową, co zapobiega ich dalszemu wydłużaniu w następnym cyklu. Ostatecznie grupa fosfinowa jest utleniana i przekształcana w stabilne ugrupowanie [77,111,114].



Rycina 13. Schemat cyklu syntezy chemicznej RNA w fazie stałej [111].

Synteza chemiczna jest odpowiednia dla stosunkowo krótkich oligomerów (do 40 nukleotydów długości). Chociaż wydajność pojedynczego cyklu wydłużania może osiągnąć 99%, ogólna wydajność systematycznie spada z każdym cyklem. W przypadku oligomerów o długości 25 nukleotydów całkowita wydajność syntezy

wynosi około 79%, przy założeniu średniej wydajności pojedynczego sprzęgania na poziomie 99%. Kiedy wydajność sprzęgania spada do 97%, całkowita wydajność wynosi tylko około 48%. W przypadku oligomerów o długości 50 nukleotydów ogólna wydajność wynosi tylko 37%, nawet jeśli średnia wydajność pojedynczego sprzęgania wynosi 98% [115]. Dlatego synteza chemiczna dłuższych RNA jest niepraktyczna. Jej uzasadnione zastosowanie występuje w momencie wprowadzania modyfikowanych nukleotydów. W przeciwieństwie do transkrypcji *in vitro*, modyfikacje chemiczne można łatwo włączyć do łańcuchów RNA podczas zautomatyzowanej syntezy. Jeśli oligomer jest zbyt długi, aby można go było skutecznie zsyntetyzować, dwa krótsze zmodyfikowane łańcuchy RNA można uzyskać oddzielnie i połączyć za pomocą ligazy RNA T4. Inną zaletą syntezy chemicznej jest mniej ograniczeń dotyczących sekwencji RNA i mniejsza heterogenność produktów w porównaniu z transkrypcją *in vitro* [44,116].

Po syntezie można zostawić grupę DMTr na końcu 5' nowo powstałego RNA (DMT-ON) lub grupę tę usunąć (DMT-OFF). Zastosowanie jednego z tych podejść determinuje zastosowanie odpowiedniego protokołu oczyszczania RNA po syntezie. W podejściu DMT-ON w celu oddzielenia cząsteczek o prawidłowej długości od wszystkich innych produktów syntezy można użyć specjalnych kolumnienek z podłożem oddziałującym z grupą DMTr. Oligomery o prawidłowej długości posiadają grupę DMTr na swoim 5' końcu i oddziałują z podłożem zawartym w kolumnie w przeciwieństwie do nieposiadających tej grupy produktów pośrednich. W metodzie DMT-OFF produkty o prawidłowej długości oraz produkty pośrednie syntezy chemicznej muszą zostać rozdzielone przy użyciu HPLC lub żelów poliakrylamidowych [78].

2.9. Stwardnienie zanikowe boczne

Stwardnienie zanikowe boczne to choroba neurodegeneracyjna układu ruchu, która prowadzi do jego stopniowego upośledzenia i ostatecznie paraliżu. W efekcie powoduje śmierć poprzez niewydolność oddechową. Objawy ALS wynikają z utraty neuronów ruchowych w rdzeniu kręgowym i pniu mózgu. Zazwyczaj początek jest ogniskowy, a następnie stopniowo rozprzestrzenia się, sugerując zaangażowanie kolejnych zestawów neuronów ruchowych w rdzeniu kręgowym [117,118]. Stwardnienie zanikowe boczne może mieć podłoże spontaniczne, co nazwane zostało sporadycznym ALS (sALS, z ang. *sporadic amyotrophic lateral sclerosis*) i podłoże dziedziczne

tw. rodzinne ALS (fALS, z ang. *familial amyotrophic lateral sclerosis*) [117–119]. Początkowo stwardnienie zanikowe boczne zostało opisane przez Jeana-Martina Charcota w 1869 roku jako choroba dotykająca wyłącznie neurony ruchowe [120]. Obecnie jednak wiadomo, że jest to wielosystemowe zaburzenie neurodegeneracyjne, które wykazuje zróżnicowanie na poziomie klinicznym, genetycznym i neuropatologicznym [118,121–123].

Ze względu na objawy wyróżnia się dwie podstawowe postaci ALS: postać kończynową (z ang. *limb onset*), która odpowiada za około 70–75% przypadków oraz postać opuszkową (z ang. *bulbar onset*) stanowiącą około 25–30% przypadków [124–126]. Niekiedy wyodrębnia się także postać oddechową (z ang. *respiratory onset*), jednak jest to bardzo rzadko występujący przebieg ALS [125]. Powyższa kategoryzacja jest uproszczeniem, a pojawiające się objawy mogą być podobne w każdej z powyższych postaci ALS. Opisuje się, że obraz kliniczny stwardnienia zanikowego bocznego obejmuje początkowo lokalne osłabienie mięśni i ich zanik, które z czasem się nasilają. W postaci kończynowej osłabienie mięśni zazwyczaj dotyczy kończyn, częściej mięśni dystalnych niż proksymalnych. Gdy choroba rozpoczyna się od objawów dotyczących mowy, połykania, głosu lub rzadziej osłabienia mięśni mówimy o ALS w formie opuszkowej [119,124–126]. Około 50% pacjentów doświadcza również objawów niezwiązanych z ruchem. Dodatkowo, u około 10% do 15% przypadków rozpoznawane jest również otępienie czołowo-skroniowe [127,128]. Wiek wystąpienia pierwszych objawów choroby, ich lokalizacja oraz tempo progresji ALS mogą się znacznie różnić. Większość pacjentów ze stwardnieniem zanikowym bocznym doświadcza nieuchronnego postępu choroby. Mediana czasu przeżycia po wystąpieniu pierwszych objawów wynosi około 3 lata, przy czym niewydolność oddechowa jest główną przyczyną śmierci [126–128].

W Europie szacuje się, że częstość występowania ALS wynosi od 10 do 12 przypadków na 100 000 osób, co może różnić się w zależności od regionu [129–131]. Mediana wieku początku objawów ALS waha się między 58 a 63 rokiem życia [132]. Szacuje się, że ryzyko zachorowania na ALS w ciągu całego życia wynosi 1 do 350 u mężczyzn i 1 do 400 u kobiet [132–134]. Mężczyźni mają większe prawdopodobieństwo rozwoju sporadycznego ALS w postaci kończynowej niż kobiety [135].

Wystąpienie stwardnienia zanikowego bocznego związane jest z mutacjami w około 20 genach. Pięć najczęstszych przyczyn genetycznych to ekspansja heksanukleotydomowa w genie *C9orf72* w chromosomie 9, mutacje w genach: *TARDBP* (kodujący białko TDP-43), *FUS* (kodującym białko FUS), *SOD1* (kodującym dysmutazę ponadtlenkową 1) i *TBKI* (kodujący kinazę wiążącą TANK 1) [121–123]. W tej grupie agregacja cytoplazmatycznego białka TDP-43, kodowanego przez gen *TARDBP*, występuje w ponad 95% przypadków i jest najbardziej charakterystyczną neuropatologiczną cechą tej choroby. Pacjenci z mutacjami w genach *SOD1* i *FUS* wykazują obecność innych agregujących się białek. Pacjenci z ekspansją *C9orf72* oraz mutacjami w genie *TBKI* również wykazują obecność inkluzji TDP-43. W niepatologicznych warunkach białko TDP-43 jest zlokalizowane głównie w jądrze komórkowym, ale w przypadku ALS nieprawidłowo przenika do cytoplazmy, gdzie tworzy agregaty i ulega fosforylacji [136,137]. *SOD1*, który jest odpowiedzialny za 20% przypadków rodzinnej postaci stwardnienia zanikowego bocznego oraz 1% do 2% przypadków sporadycznej postaci, został zidentyfikowany w 1993 roku, jako pierwszy gen związany z ALS. Mutacje tego genu zakłócają wiele kluczowych procesów komórkowych, co sprawia, że białko SOD1 jest bardziej podatne na agregację, co upośledza jego funkcję i prowadzi do rozwoju ALS [138,139]. Mutacje w genach kodujących białka TDP-43 i FUS wykryto w 2008 i 2009 roku. Odpowiadają one za 3–5% przypadków rodzinnych ALS i 1% przypadków sporadycznej postaci ALS [136,137,140–143]. W 2011 roku zidentyfikowano mutację w genie *C9orf72*, która jest odpowiedzialna za przyczynę 7–10% przypadków sporadycznego ALS i 30–50% przypadków rodzinnej postaci ALS. Nawet pomimo tego, że ekspansja heksanukleotydomowa w genie *C9orf72* znajduje się w regionie niekodującym genu, pacjenci z tą ekspansją wykazują gromadzenie się białek złożonych z powtórzeń dwóch aminokwasów. Pacjenci z ekspansją powtórzeń heksanukleotydomowych w genie *C9orf72* są bardziej narażeni na rozwój ALS o początku opuszkowym oraz na upośledzenie funkcji poznawczych i behawioralnych [144–148]. Mutacje w genie *TBKI* są prawdopodobnie piątą najczęstszą przyczyną dziedzicznej postaci ALS, stanowiąc około 1% przypadków [149–152]. Rzadko zdarza się, że konkretny pacjent posiada mutację w więcej niż jednym genie. [153].

2.10. Ołępienie czołowo-skroniowe

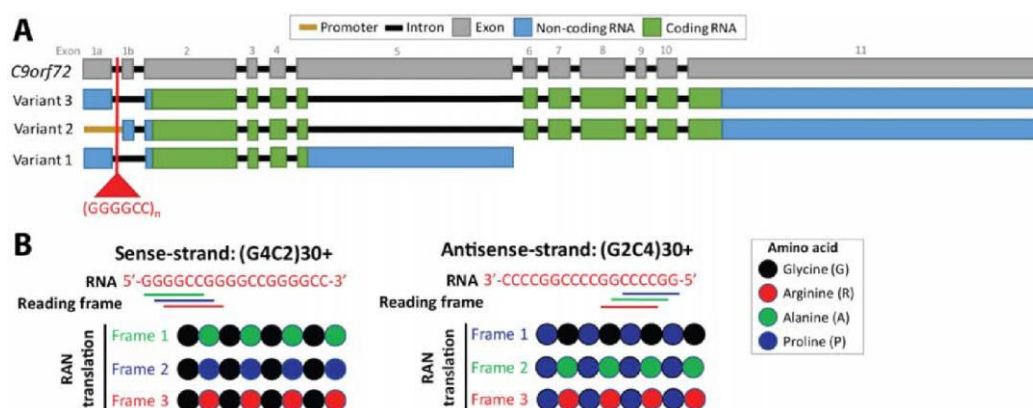
Ołępienie czołowo-skroniowe (FTD, z ang. *frontotemporal dementia*) to heterogenna grupa chorób charakteryzująca się ołępieniem niezwiązanym z chorobą Alzheimera [154,155]. Charakteryzuje się selektywną, postępującą atrofią płatów czołowych i skroniowych. Jest to grupa o fizycznie zróżnicowanym charakterze [154,156]. Pierwszy i najbardziej kompleksowy opis tej choroby został dokonany przez Arnolda Picka w XIX wieku. Nazwał ją chorobą Picka, co stanowiło historyczne określenie całego spektrum ołępienia czołowo-skroniowego. Jednak dopiero w ciągu ostatnich trzech dekad rozumiano pełną złożoność kliniczną i patologiczną tych chorób oraz ich wyjątkowe znaczenie jako przykładów selektywnej degeneracji mózgu [156–158]. Okazuje się, że główną przyczyną ołępienia czołowo-skroniowego, jak i wyżej opisanego stwardnienia zanikowego bocznego jest często ta sama mutacja w genie *C9orf72* [127,144,145].

FTD jest rzadsze niż choroba Alzheimera, a szacunki rozpowszechnienia w populacji w badaniach epidemiologicznych przeprowadzonych w Europie i Stanach Zjednoczonych wahają się od 4 do 15 na 100 000 osób w wieku od 65 lat [154,159,160]. Wartości te prawdopodobnie są niedoszacowane (głównie ze względu na błędną diagnozę), szczególnie w starszych grupach wiekowych [159,161]. FTD najczęściej diagnozuje się u pacjentów w średnim wieku, jednak zdarzają się przypadki diagnozy przed trzydziestym rokiem życia [162,163]. Ponadto ołępienie czołowo-skroniowe odpowiada za około 40% przypadków demencji o wczesnym początku [164,165].

Ołępienie czołowo-skroniowe obejmuje trzy główne zespoły kliniczne, które zostały zdefiniowane na podstawie ich charakterystycznych cech w momencie diagnozy. Około połowy przypadków FTD manifestuje się poprzez zmiany behawioralne, co nazywane jest czołowo-skroniowym wariantem behawioralnej demencji. Pozostałe przypadki charakteryzują się upośledzeniem mowy, które występuje w postaci postępującej afazji niepełnej lub upośledzeniem rozumienia słów i pamięci semantycznej, co określa się jako demencję semantyczną [163,166]. Istnieje również znaczne nakładanie się objawów klinicznych między ołępieniem czołowo-skroniowym a atypowym parkinsonizmem oraz chorobami neuronu ruchowego, co stanowi trudność w diagnostyce FTD [162,163,167–169].

2.11. Powtórzenia G₂C₄

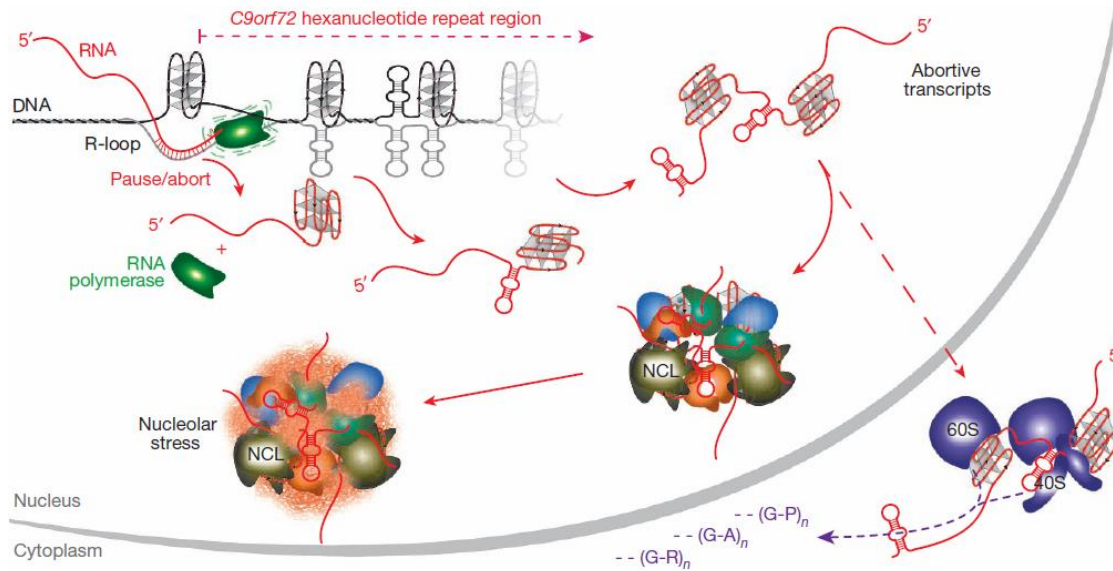
W regionie promotorowym lub pierwszym intronie genu *C9orf72* znajdują się sześci nukleotydydowe powtórzenia sekwencji G₄C₂/G₂C₄ (Rycina 14). W zdrowej komórce DNA posiada od 2 do 25 powtórzeń tej sekwencji, podczas gdy w komórkach chorych pacjentów obserwuje się wzrost liczby tych powtórzeń do nawet setek lub tysięcy. Choć pierwsze mutacje w genie *C9orf72* zostały opisane w 2011 roku [144,145], badania nad ich patogenetycznymi implikacjami szybko postępują. Owa mutacja w genie *C9orf72* jest najczęstszą przyczyną stwardnienia zanikowego bocznego oraz otępienia czołowo-skroniowego [144–146,148,170].



Rycina 14. Powtórzenia G₂C₄ znajdują się w genie *C9orf72*. A - Trzy warianty mRNA są wytwarzane z *C9orf72*, przy czym ekspansja powtórzeń występuje w intronie 1 wariantów 1 i 3 mRNA. W przypadku wariantu 2 mRNA znajduje się on w promotorze. B - Dwukierunkowa transkrypcja mutantu *C9orf72* daje dwa rodzaje RNA zawierające ekspansję powtórzeń heksanukleotydydowych: nić sensowną i antysensowną [171].

ALS i FTD należą do chorób wywołanych nadmierną ekspansją powtórzeń nukleotydydów. W związku z tym obserwuje się złożone patomechanizmy obejmujące DNA, RNA i białka (Rycina 15). Na poziomie DNA (sensownym i antysensownym) podczas transkrypcji możliwe jest tworzenie się struktur takich jak i-motywy (z ang. *i-motif*), spinka czy tetrapleks. Ponadto, możliwym jest formowanie się hybryd DNA/RNA (pętli R) hamujących proces transkrypcji. Gen *C9orf72* może być transkrybowany dwukierunkowo, co powoduje powstawanie sensownych (G₄C₂) i antysensownych (G₂C₄) powtórzeń RNA. RNA, które jest produktem transkrypcji tego genu może służyć jako matryca do translacji białek podlegających agregacji lub formować skomplikowane struktury oddziałujące z białkami, co w konsekwencji prowadzi do tworzenia się złożeń. Mimo, że powtórzenia G₄C₂/G₂C₄ znajdują się

w regionie niekodującym genu *C9orf72*, te powtarzające się sekwencje RNA mogą podlegać translacji RAN prowadząc do syntezy pięciu różnych białek z powtórzeniami dwóch aminokwasów (DPR, z ang. *dipeptide repeats*) [172].

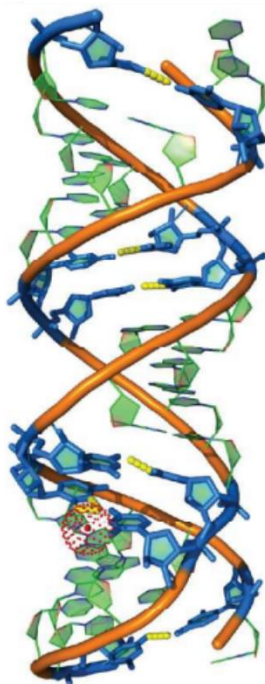


Rycina 15. Schemat patologicznego mechanizmu działania nadmiernych powtórzeń sześcienukleotydu w genie *C9orf72* [172].

Badania *in vitro* dotyczące sensownego powtórzenia DNA/RNA G_4C_2 wykazały, że krótkie (≤ 5) powtórzenia tworzą G-kwadrupeksy, które istnieją w równowadze ze strukturami spinki [173,174]. Wydaje się, że G-kwadrupeksy G_4C_2 DNA występują w topologii antyrównoległej [172,175,176], podczas gdy G-kwadrupeksy G_4C_2 RNA występują w topologii równoległej [172,177,178]. G-kwadrupeksy mogą tworzyć się zarówno w obrębie pojedynczej nici DNA/RNA, jak i między wieloma niemi [179]. Ponadto, DNA G_4C_2 i powstający RNA mogą wchodzić w interakcję tworząc hybrydowe struktury drugorzędowe takie jak pętle R, które mogą być stabilizowane przez tworzenie G-kwadrupeksów na komplementarnej nici [172,180–182]. Istnieje również dyskusja na temat tworzenia kwadrupeksów przez antysensowne DNA G_2C_4 . Sekwencje G_2C_4 mogą tworzyć i-motywy, które istnieją w równowadze ze strukturami spinki. Antysensowne DNA G_2C_4 może również tworzyć inne, stosunkowo niezbadane struktury drugorzędowe [171,178,182–184].

Wcześniejsze badania skupiały się głównie na strukturze G-kwadrupeksu dla sensownego RNA $(GGGGCC)_n$, podczas gdy struktura antysensownego RNA $(GGCCCC)_n$ nie była dobrze zbadana [178]. Sekwencje bogate w reszty cytozyny mogą

tworzyć struktury i-motywu (tetrapleksy) [185]. W badaniu przeprowadzonym przez Davida Dodda i współpracowników [186], zbadano strukturę krystaliczną sekwencji zawierającej cztery powtórzenia G_2C_4 z brakującymi dwiema guanozynami na obu końcach konstruktów (Rycina 16).



Rycina 16. Struktura krystaliczna sekwencji zawierającej cztery powtórzenia G_2C_4 z brakującymi dwiema resztami guanozyny na obu końcach konstruktów [186]

Okazało się, że zaprojektowany konstrukt RNA tworzy dwuniciową strukturę helisy typu A z powtarzającymi się jednostkami czterech par zasad Watsona-Cricka GC/CG oddzielonymi dwiema tandemowymi parami C-C. Obecność regularnie rozmieszczonych par tandemowych C-C zakłóca typową strukturę helisy typu A RNA. Zaobserwowano węższy rozmiar rowka mniejszego oraz różnicę w rozmiarze rowka większego i ogólnej średnicy helisy. Niekanoniczne pary C-C zmieniają elektrostatyczny potencjał powierzchniowy RNA w porównaniu do par zasad Watsona-Cricka GC/CG. Zmiany te obserwowano na powierzchni mniejszego rowka, co może wpływać na interakcje z białkami wiążącymi RNA. Procesy agregacji antysensownego powtórzenia RNA *C9orf72* w jądrze komórkowym i translacji RAN w cytoplazmie powiązane są z tworzeniem oddziaływań pomiędzy RNA a białkami. Unikalna struktura i elektrostatyczny potencjał powierzchniowy RNA wynikający z obecności tandemowych par C-C sugerują, że istnieje możliwość nadmiernego wiązania białek,

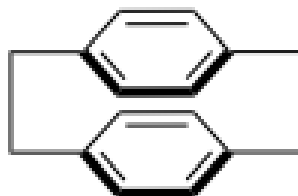
które specyficznie rozpoznają te cechy. Przykładowo, białko MBNL-1, które jest sekwestrowane przez powtarzające się sekwencje CUG naturalnie rozpoznaje regiony dwunukleotydomowe GC z różnymi odstępami między nimi. Potencjalnie specyficzne czynniki oddziałujące, takie jak białko hnRNP39, zostały zidentyfikowane dla sensownego i antysensownego powtórzenia RNA *C9orf72*. Alternatywnie, ponieważ antysensowne RNA *C9orf72* przypomina kanoniczną helisę typu A, może ono wiązać różne niespecyficzne białka oddziałujące z RNA. Interakcje białek, głównie dla sensownego RNA bogatego w reszty guanozyny, doprowadziły do identyfikacji białek o różnej specyficzności wiązania [186].

2.12. Małe cząsteczki oddziałujące z RNA

Oprócz przenoszenia informacji genetycznej, RNA regulują ekspresję genów i odgrywają kluczową rolę w wielu procesach biologicznych, w tym biorą udział w rozwoju chorób człowieka [187–189]. Mutacje DNA, a w konsekwencji w pre-mRNA, mogą prowadzić do śmierci komórki, jeśli przykładowo powodują zmiany w produkcji niezbędnych białek do prawidłowego jej funkcjonowania. Jest to aktywny obszar badań w biologii molekularnej, który polega na specyficznej detekcji uszkodzonych lub zmienionych regionów DNA i RNA poprzez małe cząsteczki, co z kolei stanowi podstawę dla potencjalnych nowych leków [190]. Atrakcyjną strategią projektowania leków jest interakcja małych cząsteczek ze specyficznymi regionami w RNA takimi jak pętle struktury spinki RNA, wybrzuszenia oraz wewnętrzne pętle obecne w rejonach dwuniciowych poprzez wiązania wodorowe, koordynację przez jony metali, hydrofobowe oddziaływanie typu „*pi-stacking*”, oddziaływanie van der Waalsa i oddziaływanie elektrostatyczne [191–195]. Wyżej wymienione regiony znajdują się w RNA zawierających nadmierne powtórzenia nukleotydomowe związane z chorobami neurodegeneracyjnymi. Interakcje małych cząsteczek z RNA potencjalnie mogłyby przeciwdziałać rozwojowi chorób neurodegeneracyjnych poprzez inhibicję formowania się patogennych struktur RNA, stabilizację prawidłowych struktur RNA lub wpływ na interakcje RNA z białkami.[192,195].

W ostatnich latach dużą popularnością cieszą się małe cząsteczki, które mogą rozpoznawać niesparowane nukleotydy w strukturach RNA i wiązać się do nich. Takimi cząsteczkami kandydującymi są cyklofany [196,197]. Cyklofany można zdefiniować

jako klasę związków cyklicznych składających się z jednostek aromatycznych połączonych grupami mostkującymi (Rycina 17) [198]. Właściwości wykazywane przez te związki można dostosować poprzez zmianę jednostki aromatycznej lub jednostki mostkującej. Takie związki cykliczne są zdolne do stabilizacji cząsteczek w obszarze hydrofobowym poprzez oddziaływania niekoordynacyjne [199–201]. Badania nad cyklofanami są intensywne ze względu na ich potencjalne zastosowania w rozpoznawaniu molekularnym, chemii polimerów, chemii supramolekularnej, nośnikach leków oraz katalizatorach w syntezie organicznej i syntezie asymetrycznej oraz ich właściwości przeciwbakteryjne i przeciwwirusowe jako inhibitorów proteazy HIV. Wiele związków cyklofanowych wykazuje specyficzne interakcje z nukleotydami i aminokwasami [202,203].



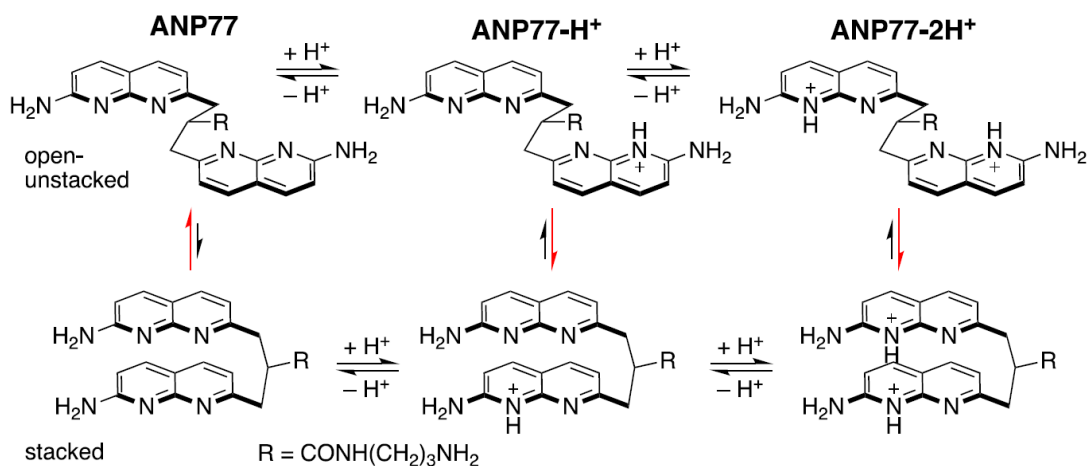
Rycina 17. Schemat przedstawiający cząsteczkę cyklofanu zawierającą dwa pierścienie aromatyczne [204]

Inną interesującą grupą małych związków chemicznych są pochodne 1,8-naftyrydyny, które wykazują różnorodne aktywności biologiczne. Wzbudzają one szczególne zainteresowanie ze względu na układ atomów azotu w cząsteczce. Układ ten jest idealny do chelatowania różnych kationów metali, w tym jonów lantanowców oraz do molekularnego rozpoznawania syntetycznych receptorów. Pochodne 1,8-naftyrydyny mogą także oddziaływać z zasadami azotowymi. Dzięki tej właściwości mogą być używane do identyfikacji niesparowanych nukleotydów zawartych w strukturach RNA z nadmiernie powtórzonymi nukleotydami [203].

2.13. Ligand ANP77

Biorąc pod uwagę biologicznie istotną rolę cyklofanów oraz pochodnych 1,8-naftyrydyny, grupa profesora Nakataniego zaprojektowała i otrzymała serię małych cząsteczek określanych jako ligandy wiążące niesparowane nukleotydy

(MBLs, z ang. *mismatch binding ligands*). Zdolność ligandów do wiązania niesparowanych zasad zależy przede wszystkim od dostępności danej zasady, interakcji z sąsiadującymi zasadami oraz komplementarności wiązań wodorowych [205]. Jedną z cząsteczek otrzymanych przez grupę prof. Nakataniego jest użyty w tej pracy ligand ANP77, który posiada trzy atomowy linker łączący dwa pierścienie 2-amino-1,8-naftyrydyny. Ze względu na swoją budowę ligand ten wykazywał potencjał do oddziaływań z resztami cytozyny tworząc pseudo-pary zasad. Pierścienie heterocykliczne ANP77 mogą być protonowane w pH zbliżonym do obojętnego, wytwarzając monokation lub dikation. Monokation ANP77 prawdopodobnie może przybierać konformację typu *stacked* przez interakcję protonowanej naftyrydyny z nieprotonowaną. Dikation może przyjmować zarówno konformację typu *unstacked* jak i typu *stacked*, jednak konformacja typu *stacked* jest mniej prawdopodobna przez odpychanie się dwóch kationów znajdujących się w bliskiej odległości (Rycina 18) [205,206].



Rycina 18. Ilustracja konformacji typu *open-unstacked* (górny rząd) i typu *stacked* (dolny rząd) ułożonych jako cząsteczka bez ładunku (lewa kolumna), monokationu (środkowa kolumna) i dikationu (prawa kolumna) ANP77 [205].

W pracy Das i współpracowników, aby zbadać stan protonowania ANP77, autorzy zmierzili zależne od pH widma absorpcyjne ANP77 i jego monomeru w przedziale pH 3–9. Zmiany w absorpcyjnych widmach zależnych od pH ujawniły, że cząsteczka ANP77 w formie monokationu przyjęła konformację typu *stacked* z pKa wynoszącym 6,6. Znacznie niższe pH wymagane do protonacji drugiej naftyrydyny w cząsteczce ANP77 sugeruje, że monokation stanowi główną frakcję wśród innych

protonowanych stanów w słabo kwasowym roztworze. Ze względu na te informacje autorzy pracy przypuszczali, że prawdopodobną formą cząsteczki wiążących się do reszt cytozyny w kwasie nukleinowym jest monokation [205].

Poznanie struktury przestrzennej RNA oddziałującego z cząsteczką ANP77 pozwoliłoby na określenie konformacji liganda oraz sposobu interakcji pomiędzy ANP77 a RNA.

3. Materiały i Metody

3.1. Materiały

3.1.1. Aparatura

- Wirówka laboratoryjna MIKRO 200 R Hettich
- Worteks Bender and Hobein
- Termoblok BioSan TS-100C
- Wyrząsarka IKA MS3 basic
- Termocykler Biometra Trio Analytik Jena
- Nanofotometr IMPLEN N60
- Różnicowy kalorymetr skaningowy Microcal PEAQ0DSC Malvern
- System do chromatografii Bio-Rad NGC
- Inkubator Innova 42
- CentriVap Concentrator Labconco
- Wirówka Avanti J-26 XPI Beckman Coulter
- Wirówka Avanti JXN-26 Beckman Coulter
- Dyfraktometr Rigaku Synergy XtaLab
- Sonikator Sinics Vibra-Cell
- Wirówka Eppendorf 5804R
- Robot do krystalizacji Gryphon Art Robbins.
- FEI Vitrobot mark IV
- Kriomikroskop elektronowy Thermo Fischer Glacios 2 Cryo-TEM

3.1.2. Odczynniki chemiczne i biochemiczne

Większość odczynników pochodziła z firm: Sigma-Aldrich, Epro, Chempur. Odczynniki były najwyższej czystości i oznaczone jako wolne od rybonukleaz. Wszystkie filtry zakupione były w firmie Millipore lub Sartorius. Amidofosforyny do syntezy chemicznej z blokadą 2'-triizopropylsililooksymetylową zakupione były

w firmie Chemgene. Screen do krystalizacji zakupiony został z firmy Hampton Research. Nukleotydy do transkrypcji *in vitro* zakupione zostały z firmy EurX.

3.1.3. Stosowane bufony roztwory i pożywki

Bufor do elektroforezy 10 x TB

890 mM Tris, 890 mM kwas borowy

Bufor do elektroforezy w warunkach denaturujących 10 x TBE

890 mM Tris, 890 mM kwas borowy, 20 mM EDTA, pH 8,3

40% roztwór żelu poliakrylamidowego 29 : 1

193,33 g akrylamidu, 6,66 g bisakrylamidu, H₂O do 500 ml

Bufor do elektroforezy białek w warunkach denaturujących

25 mM Tris-HCl pH 8,3, 190 mM glicyna, 0,1% SDS

Bufor do żelu zagęszczającego

125 mM Tris-HCl, 0,1% SDS, pH 6,8

Bufor do żelu rozdzielającego

375 mM Tris-HCl, 0,1% SDS, pH 8,8

Roztwór do denaturacji i obciążania próbek białkowych – roztwór SSB

50 mM Tris-HCl pH 6,8, 10 mM ditiotretitol, 2% SDS, 10% glicerol, 0,02% w/v błękit bromofenolowy

Pożywka LB płynna

15 g bakto-trypton, 7,5 g ekstrakt drożdżowy, 15 g NaCl, H₂O do 1500 ml, pH 7,0, ampicylina 100 µg/ml i chloramfenikol 37 µg/ml. Antybiotyki dodawane tuż przed rozpoczęciem hodowli komórek bakteryjnych.

Bufor do lizy bakterii podczas produkcji polimerazy T7 oraz ligazy RNA T4

50 mM Tris-HCl pH 8,0, 500 mM NaCl, 5% glicerol, 10 mM imidazol, 1 mM TCEP

Bufor przemywający do oczyszczania polimerazy T7

50 mM Tris-HCl pH 8,0, 500 mM NaCl, 5 % glicerol, 40 mM imidazol, 1 mM TCEP

Bufor do elucji polimerazy T7 i ligazy RNA T4

50 mM Tris-HCl pH 8,0, 500 mM NaCl, 5 % glicerol, 300 mM imidazol, 1 mM TCEP

Bufor do przechowywania polimerazy T7

40 mM Tris-HCl pH 8,0, 150 mM NaCl, 25% glicerol, 1 mM TCEP

Bufor przemywający do oczyszczania ligazy RNA T4

50 mM Tris-HCl pH 8,0, 50 mM NaCl, 5 % glicerol, 40 mM imidazol, 1 mM TCEP,
10 mM P₂O₄

Bufor do przechowywania ligazy RNA T4

50 mM HEPES pH 7,5, 150 mM KCl, 25% glicerol, 1 mM ditiotretitol

Bufor do lizy bakterii podczas produkcji białka U1a-RBD

25 mM Tris-HCl pH 8,0, 300 mM NaCl, 20 mM imidazol, 1 mM TCEP

Bufor przemywający do oczyszczania białka U1a-RBD

25 mM Tris-HCl pH 8,0, 300 mM NaCl, 40 mM imidazol, 1 mM TCEP

Bufor do elucji białka U1a-RBD ze złoża niklowego

25 mM Tris-HCl pH 8,0, 300 mM NaCl, 300 mM imidazol, 1 mM TCEP

Bufor do dializy białka U1a-RBD z proteazą TEV

25 mM Tris-HCl pH 8,0, 300 mM NaCl, 1 mM TCEP

Bufor do oczyszczania białka U1a-RBD za pomocą systemu AKTA prime i przechowywania

1M Tris-HCl pH 7,5, NaCl 100 mM

Bufor do cyrkularyzacji RNA

50 mM Tris-HCl pH 7,5, 10 mM MgCl₂, 10 mM ditiotreitol

Bufor do transkrypcji *in vitro*

40 mM Tris-HCl pH 8,0, 1 mM spermidyna, 0,01 % Triton X-100, 20 mM MgCl₂, 5 mM ditiotreitol

Bufor EMSA

10 mM kakodylan sodu pH 7,0, 5 mM MgCl₂

Bufor do krystalizacji RNA

10 mM kakodylan sodu pH 7,0, 100 mM NaCl

Roztwory do syntezy chemicznej

Aktywator: 5,7% 5-benzyltio-1H-tetrazol w acetonitrylu

Deblocking: 3% kwas trichlorooctowy w dichloroetanie

Cap A: 6% DMAP w acetonitrylu

Cap B: 15 ml 2,4,6-trimetylopirydyna, 10 ml bezwodnik octowy, 25 ml acetonitrylu

Oxidation: 250 ml jodu, 35 ml tetrahydrofuranu, 10 ml pirydyny, 5 ml H₂O

Amidofosforyny: roztwór 1 μM w acetonitrylu

3.1.4. ANP77

Cząsteczki ANP77 zostały zsyntetyzowane przez zespół prof. Kazuhiko Nakataniego (The Institute of Scientific and Industrial Research, Osaka University).

bakterii podczas produkcji polimerazy T7 oraz ligazy RNA T4. Do osadu pozyskanego z 0,6 litra hodowli dodawałem 20 ml buforu. Zawiesinę komórkową sonikowałem przez 12 sekund przy amplitudzie 70% mocy. Otrzymany ekstrakt wirowałem w temperaturze 4°C przy 20000 obr./min przez 30 minut. Uzyskany supernatant przenosiłem do kolumniek zawierających 8 ml złoża niklowego typu Ni-NTA (Qiagen), zrównoważonego wcześniej buforem do lizy bakterii podczas produkcji polimerazy T7 oraz ligazy RNA T4 i inkubowałem 15 min w 4°C. Ciecz w kolumnie usuwałem za pomocą pompy wykonanej ze strzykawki. Złoże przepłukiwałem 600 ml buforu przemywającego do oczyszczania polimerazy T7 lub ligazy RNA T4, a następnie białka eluowałem ze złoża 50 ml buforu do elucji polimerazy T7 i ligazy RNA T4. Wymyte białka dializowałem w buforze do przechowywania odpowiednim dla izolowanego białka przez 36 godzin (co 12 godzin bufor wymieniałem na świeży). Po tym czasie białka zateżałem używając probówek z filtrem AMICON do stężenia około 1,5 mg/ml. Proces oczyszczania białka monitorowałem pobierając po każdym etapie próbki rozdzielane w żelu poliakrylamidowym. Stężenie białka mierzyłem za pomocą nanofotometru IMPLEN. Aktywność polimerazy sprawdzałem wykonując reakcję transkrypcji *in vitro*. Aktywność ligazy sprawdzałem wykonując reakcję cyrkularyzacji RNA. Produkty obu reakcji rozdzielałem w 12% żelu poliakrylamidowym i obrazowałem przy użyciu światła UV.

3.2.2. Produkcja białka U1a-RBD

Produkcję białka U1a-RBD rozpocząłem od przeniesienia małej ilości zamrożonego stoku bakteryjnego do pożywki LB o objętości 25 ml z antybiotykiem. Taką hodowlę prowadziłem przez noc w 37°C z wytrząsaniem przy 230 obr./min. Uzyskaną zawiesinę zaszczepiłem 1,6 litra pożywki LB z antybiotykiem (4 kolby o pojemności 1 litra, w każdej 0,4 litra hodowli) i dalej inkubowałem do uzyskania gęstości optycznej $OD_{600} = 0,8$, po czym hodowlę schłodziłem do 18°C. Następnie zaindukowałem ekspresję używając IPTG do końcowego stężenia 1 mM i inkubowałem dalej przez noc w 18°C. Następnego dnia komórki oddzieliłem od pożywki wirując je w 4°C przy 4000 obr./min przez 15 minut. Osad komórkowy rozpuszczałem w buforze do lizy bakterii podczas produkcji białka U1a-RBD. Na osad pozyskany z 0,4 litra hodowli dodawałem 15 ml buforu. Zawiesinę komórkową sonikowałem przez 12 sekund przy amplitudzie

70% mocy. Ekstrakt wirowałem w temperaturze 4°C przy 20000 obr./min przez 30 minut. Uzyskany supernatant przenosiłem do kolumnienek zawierających 4 ml złoża niklowego typu Ni-NTA (Qiagen) zrównoważonego wcześniej buforem do lizy bakterii podczas produkcji białka U1a-RBD i inkubowałem 15 min w 4°C. Ciecz w kolumnie usuwałem za pomocą pompy wykonanej ze strzykawki. Złoże przepłukiwałem 200 ml buforu przemywającego do oczyszczania białka U1a, a następnie białko eluowałem ze złoża 20 ml buforu do elucji białka U1a-RBD ze złoża niklowego. Wyeluowane białko dializowałem w buforze do dializy białka U1a-RBD z proteazą TEV przez 36 godzin (co 12 godzin bufor wymieniałem na świeży) i jednocześnie poddawałem działaniu proteazy TEV (0,1 mg proteazy na 1 mg białka U1a-RBD), aby odciąć od białka U1a-RBD łańcuch aminokwasowy niezbędny do oddziaływania białka ze złożem niklowym. Po tym czasie białko oczyszczałem używając AKTA prime FPLC system z użyciem buforu do oczyszczania białka U1a-RBD za pomocą systemu AKTA prime i przechowywania. Następnie białko U1a-RBD zatężałem używając probówek z filtrem AMICON do stężenia około 9 mg/ml. Proces oczyszczania białka monitorowałem pobierając po każdym etapie próbki, które rozdzielałem w żelu poliakrylamidowym. Stężenie białka mierzyłem za pomocą nanofotometru IMPLN. Aktywność U1a-RBD sprawdzałem wykonując reakcję wiązania białka do pętli RNA wiążącej białko U1a.

3.2.3. Synteza chemiczna RNA

Syntezę oligonukleotydów przeprowadzałem z użyciem syntetyzera DNA/RNA Applied Biosystem 392 metodą amidofosforynową na podłożu stałym. Wszystkie oligomery oddzieliłem od podłoża zgodnie z protokołem AMA – wodorotlenek amonu/metyloamina w wodzie (40%) (1:1 v/v) przez 4 godziny. Oligomery DMT-ON dalej odbezpieczałem i oczyszczałem za pomocą kolumnienek do oczyszczania DNA Glen-Pak lub Glen-Pak RNA (Glen Research) zgodnie z protokołem producenta. Dla oligomerów RNA DMT-OFF blokującą grupę 2' usuwałem 1 M fluorkiem tetra-n-butyloamoniowym w tetrahydrofuranie w 37°C przez 20 godzin. Oligomery odsalałem przy użyciu kolumn Illustra NAP-25 (GE health) lub Bio-Scale Mini Bio-Gel P-6 Desalting Cartridge (Bio-Rad). Czystość oligomerów oceniałem za pomocą wysokosprawnej chromatografii ciekowej lub 10-15% denaturujących żeli poliakryloamidowych.

3.2.4. HPLC

Około 5 µg próbki RNA wstrzykiwałem na kolumnę Thermo Fisher Scientific DNAPac PA 100 4 × 250 mm. Rozdzielanie prowadziłem przy szybkości przepływu 1 ml/min w temperaturze pokojowej w gradiencie 0-60% 0,5 M buforu nadchloranu sodu w 25 mM Tris pH 8,0 przez 45 minut.

3.2.5. Transkrypcja *in vitro*

Reakcja transkrypcji *in vitro* zawierała 1 µg zlinearyzowanego plazmidu lub chemicznie zsyntetyzowanego DNA, 15 µg polimerazy RNA T7 własnej produkcji, 4 mM mieszaninę NTP i bufor transkrypcyjny. Stężenie jonów magnezu dostosowałem do każdego konstruktu RNA i typu rybozomu (zakres stężeń: 20-32 mM). W przypadku chemicznie syntetyzowanej matrycy DNA, przed reakcją transkrypcji *in vitro*, 500 ng każdej nici hybrydowałem poprzez denaturację w 95°C przez 5 min i renaturację przez 10 min w temperaturze pokojowej. Reakcję transkrypcji prowadziłem przez 4 godziny lub przez noc w temperaturze 37°C. Następnie matrycowy DNA trawiłem przy użyciu 10 U DNAzy I inkubując mieszaninę w 37°C przez 20 min. Produkty reakcji oczyszczałem przy użyciu 10–15% denaturujących żeli poliakrylamidowych, eluowałem z żelu za pomocą elektrolecucji oraz strącałem przy użyciu izopropanolu.

3.2.6. Oczyszczanie RNA w żelach poliakrylamidowych

Elektroforezę w skali preparatywnej prowadziłem żelach 8–12% w zależności od długości RNA. Każdą elektroforezę prowadziłem w buforze 1 x TBE, którą poprzedzałem preelektroforezą przy natężeniu prądu 10 lub 15 mA. Przed nałożeniem na żel, próbki mieszałem w stosunku 1:1 z roztworem 8 M mocznika. Elektroforezę prowadziłem przy stopniowo zwiększonym natężeniu prądu do uzyskania wartości 50 mA.

3.2.7. Elektroelucja RNA

Po oczyszczeniu w żelu wycięte prążki RNA przenosiłem do worka dializacyjnego ThermoFisher Snake Skin, który zalewałem buforem 1 x TBE. Następnie woreczek dializacyjny zamykałem z obu stron. Tak przygotowany wycinek żelu w woreczku dializacyjnym umieszczałem w zlewce zawierającej 1 x TBE i przykładałem prąd o napięciu 120 V przez 1 h. Po tym czasie przenosiłem roztwór z wyeluowanym RNA z worka dializacyjnego do próbki typu Eppendorf. Następnie strącałem tak otrzymany RNA za pomocą izopropanolu.

3.2.8. Strącanie RNA izopropanolem

Oligomery strącałem 1 objętością izopropanolu (w stosunku do objętości RNA) w obecności 100 mM octanu sodu o pH 5,26. RNA strącałem w temperaturze -20°C przez godzinę, a następnie wirowałem przez 30 minut przy 14000 obr./min. w temperaturze 4°C . Po usunięciu supernatantu, osad suszyłem i rozpuszczałem w dejonizowanej wodzie. Stężenie kwasów nukleinowych oznaczałem na podstawie pomiarów absorpcji przy długości fali $\lambda = 260 \text{ nm}$ z wykorzystaniem nanofotometru.

3.2.9. Cyrkularyzacja RNA

Celem cyrkularyzacji, RNA o stężeniu 10 pmol/ μl inkubowałem w buforze zawierającym 80 mM KCl kolejno: 5 min w 95°C , 10 min w 4°C i 10 min w temperaturze pokojowej. Następnie tak przygotowany RNA cyrkularyzowałem przy użyciu ligazy RNA T4 w obecności 10 mM ATP w 1x buforze do cyrkularyzacji przez 5 godzin. Powodzenie cyrkularyzacji sprawdzałem rozdzielając w żelu poliakrylamidowym RNA po reakcji cyrkularyzacji z RNA, którego nie poddawałem działaniu ligazy RNA T4.

3.2.10. Ekstrakcja RNA fenolem

Do próbki zawierającej RNA dodawałem 1 objętość fenolu o pH 4,5 (w stosunku do objętości próbki) i wytrząsałem przez 1 minutę w temperaturze pokojowej. Mieszaninę

wirowałem przez 1 minutę przy 10000 obr./min. Warstwę wodną przenosiłem do nowej probówki typu Eppendorf. Aby pozbyć się resztek fenolu z próbki, do fazy wodnej dodawałem 1 objętość mieszaniny chloroform:alkohol izoamylowy w stosunku 24:1 (v/v) i wytrząsałem przez 1 minutę w temperaturze pokojowej, po czym wirowałem przez 1 minutę przy 10000 obr./min. Fazę wodną następnie strącałem izopropanolem.

3.2.11. Badanie interakcji RNA–białko metodą EMSA

Do badania oddziaływań RNA z białkiem U1a-RBD wykorzystałem metodę różnicowej migracji w żelach poliakrylamidowych (EMSA, z ang. *Electrophoretic Mobility Shift Assay*). Reakcję prowadziłem w objętości 10 μ l. Mieszanina reakcyjna zawierała 30 pmoli RNA zawieszonego w buforze zawierającym 10 mM kakodylan sodu pH 7,0. Tak przygotowaną reakcję inkubowałem 5 min w 95°C oraz 10 min w 4°C. Następnie do reakcji dodałem chlorek magnezu do stężenia 5 mM i dalej inkubowałem mieszaninę przez 10 min w temperaturze pokojowej. Później do RNA dodawałem białko U1a-RBD i inkubowałem mieszaninę 20 min w temperaturze pokojowej. Próbkę obciążałem 2 μ l 12% fikolem i prowadziłem elektroforezę w żelu poliakrylamidowym w warunkach natywnych w 4°C przy natężeniu 10 mA. Żel wizualizowałem używając toluidyny.

3.2.12. Krystalizacja RNA

Liofilizowane oligomery RNA rozpuszczałem w wodzie dejonizowanej. Następnie dodawałem roztwór soli NaCl do stężenia 100 mM oraz kakodylanu sodu do stężenia 10 mM. Tak przygotowaną mieszaninę denaturowałem przez 5 minut w temperaturze 95°C. Dalej mieszaninę schładzałem do temperatury pokojowej przez co najmniej 30 minut. Następnie, w przypadku krystalizacji w obecności liganda lub białka U1a-RBD, do mieszaniny dodawałem ligand ANP77/białko U1a-RBD i inkubowałem ją w temperaturze pokojowej przez 20 min. Warunki krystalizacji ustaliłem przy użyciu komercyjnie dostępnych zestawów roztworów krystalizacyjnych Natrix 1, Natrix 2 lub Crystal Screen (Hampton Research) w temperaturach 4°C, 19°C i 30°C. RNA mieszałem z roztworem krystalizacyjnym w stosunku 1:1, 1:2 i 2:1.

Krystalizacje oligomerów nastawiałem ręcznie lub za pomocą robota do krystalizacji Gryphon Art Robbins.

3.2.13. Pomiary dyfrakcyjne, procesowanie danych i rozwiązanie struktury

Pomiary dyfrakcyjne prowadziłem przy użyciu sprzętu XtaLAB Synergy-R. Kryształy przed pomiarem zamrażałem w ciekłym azocie w obecności krioprotektanta - glicerolu. Zarejestrowane dane dyfrakcyjne procesowałem za pomocą programu CrysAllis Pro.

Celem badania struktur krystalicznych za pomocą metod dyfrakcji rentgenowskiej jest określenie rozkładu gęstości elektronowej w symetrycznie niezależnej części komórki elementarnej, ponieważ maksima gęstości elektronowej wskazują miejsca, gdzie znajdują się atomy. Informacje o rozkładzie gęstości elektronowej są ukryte we wzorcu dyfrakcji uzyskanym z badanego kryształu. Funkcja matematycznie opisująca wzorec dyfrakcji nazywana jest czynnikiem struktury i oznaczana symbolem $F(hkl)$. Rozkład gęstości elektronowej można matematycznie wyrazić jako odwrotną transformację Fouriera czynnika struktury na podstawie analizowanego wzorca dyfrakcji. Proces ten pozwala uzyskać cenne informacje na temat układu atomów w sieci krystalicznej i zrozumieć jej trójwymiarową strukturę.

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| e^{-2\pi i(hx+ky+lz) - i\alpha(hkl)}$$

ρ – wartość gęstości elektronowej w punkcie xyz

$|F(hkl)|$ – amplituda czynnika struktury dla refleksu (hkl)

$\alpha(hkl)$ – kąt fazowy dla refleksu (hkl)

x, y, z – współrzędne w komórce elementarnej w przestrzeni rzeczywistej

h, k, l – współrzędne w przestrzeni odwrotnej

Wartości kąta fazowego nie mogą być zmierzone za pomocą detektora, podczas gdy amplituda może być obliczona na podstawie intensywności odbicia $I(hkl)$ na podstawie zależności:

$$I(hkl) = |F(hkl)|^2$$

Problem fazowy jest zatem fundamentalnym problemem do rozwiązania w analizie strukturalnej. W podanej pracy do rozwiązania tego problemu wykorzystałem metodę podstawienia cząsteczkowego (MR, z ang. *molecular replacement*). Może być ona używana, gdy badany RNA ma strukturalne podobieństwo do już istniejącego modelu. Rozwiązanie problemu fazowego opiera się na nakładaniu funkcji Pattersona, jednej pochodzącej z eksperymentalnych amplitud czynnika struktury, a drugiej z amplitud obliczonych z "modelu poszukiwania", w którym poszukuje się największej korelacji między wektorami eksperymentalnymi a tymi pochodzącymi ze znanego modelu. Aby metoda podstawienia cząsteczkowego była skuteczna, "model szukający" musi spełniać pewne kryteria. Powinien być podobny z badaną cząsteczką RNA na poziomie większym niż 30%. Modele używane w tej metodzie są pozyskiwane z PDB. Wstępna ocena poprawności rozwiązania opiera się na dwóch parametrach: LLG oraz z-score. LLG informuje o tym, jak lepiej można przewidzieć dane przy użyciu aktualnego modelu w porównaniu z modelem, w którym atomy są rozmieszczone losowo. Wartość tego parametru powinna być dodatnia i rosnąć na przestrzeni kolejnych etapów dopasowania pełnego modelu. Z kolei z-score jest związany z przyrostem logarytmu wiarygodności i mierzy, ile standardowych odchyień od średniej wynosi wartość parametru LLG. Ważne jest, aby z-score był jak najwyższy. Przyjmuje się, że z-score powyżej 8 wskazuje na osiągnięcie poprawnego rozwiązania. Kluczowym jest dokładna inspekcja mapy gęstości elektronowej obliczonej na podstawie uzyskanych wstępnych faz. Czasem, mimo niskiej wartości z-score model może być częściowo poprawny i może być wykorzystany do skonstruowania dokładniejszego modelu.

We wszystkich opisywanych w tej pracy strukturach zastosowałem metodę podstawienia cząsteczkowego przy użyciu programu PHASER z pakietu programów CCP4 [207].

3.2.14. Udokładnianie struktury RNA

Rozwiązanie struktury to tylko pierwszy etap, który pozwala zdefiniować współrzędne poszczególnych atomów – uzyskujemy jedynie przybliżony model. W przypadku metody MR, zdefiniowane są jedynie atomy modelu początkowego, dlatego konieczne jest udokładnienie współrzędnych wszystkich atomów, a w niektórych przypadkach także odbudowanie części modelu. Udokładnienie odbywa się w kolejnych

cyklach. Każdy krok obejmuje ręczną korektę modelu oraz rundę obliczeń w celu zminimalizowania różnic między eksperymentalnymi a obliczonymi wartościami czynników struktury [208]. Ten proces jest realizowany poprzez dostosowanie czynników temperaturowych (izotropowo lub anizotropowo), położenia oraz obsadzenia każdego pojedynczego atomu. W trakcie udokładniania struktur RNA często korzysta się z danych stereochemicznych, które zostały wywnioskowane z wysokorozdzielczych struktur małowzrostkowych [209]. Gdy te informacje są wykorzystywane jako więzy miękkie, dodatkowe równania są włączane do procesu obliczeniowego. Te równania uwzględniają pewne obserwowane wartości, takie jak długości wiązań i kąty pomiędzy poszczególnymi atomami, które dążą do osiągnięcia wartości "idealnych", przy uwzględnieniu pewnego poziomu zmienności. Z kolei, gdy takie informacje są traktowane jako więzy sztywne, liczba parametrów jest zmniejszana poprzez ustalenie ich stałych wartości. Dodatkowo, wykorzystuje się parametry TLS (z ang. *translation, libration, screw motion*), które częściowo oddają anizotropię drgań grup atomów. Do konstrukcji modelu użyłem programu COOT [210].

Każdy etap udokładniania jest monitorowany przez sprawdzenie wskaźników rozbieżności R i R_{free} . Są to wartości rozbieżności między eksperymentalnymi amplitudami współczynnika struktury a tymi obliczonymi na podstawie modelu. Wskaźnik R , obliczany dla całego zbioru refleksów, ma wadę polegającą na tym, że może się poprawiać (zmniejszać wartość) wskutek nadmiernego dopasowania modelu i danych przy każdym rozszerzeniu modelu o dodatkowe parametry (nawet jeśli nie są one potwierdzone przez dane eksperymentalne). Aby temu zapobiec, wprowadzono wskaźnik R_{free} [211], oparty na losowo wybranych (około 1000) refleksach, które są wyłączone z procesu udokładniania od samego początku [212]. Dla dobrze udokładnionych struktur, wskaźniki R i R_{free} nie mogą przekraczać 30%, a różnica między nimi nie powinna przekraczać około 3–7 punktów procentowych [213]. Jednak parametry te zależą od rozdzielczości i jakości danych.

Do wizualizacji map gęstości elektronowych i budowy modeli struktur opisanych w tej pracy użyłem programu COOT. Struktury udokładniłem za pomocą programu REFMAC5, który jest częścią pakietu CCP4. Ocena jakości struktur była przeze mnie wykonywana poprzez kontrolę czynników rozbieżności R i R_{free} .

3.2.15. Skaningowa kalorymetria różnicowa

W tej pracy użyłem skaningowej kalorymetrii różnicowej (DSC, z ang. *differential scanning calorimetry*) do pomiaru temperatury topnienia konstruktu G_2C_4 oraz kompleksu G_2C_4 z ANP77 w różnych warunkach pH. Podstawą metody DSC jest zasada, że jeśli w trakcie podgrzewania zachodzą przemiany fizyczne w badanej próbce, tempo dostarczania ciepła będzie różnić się od materiału odniesienia (buforu, w którym próbka jest rozpuszczona). Jeśli w próbce zachodzą procesy endotermiczne to tempo dostarczania ciepła będzie wyższe, w przeciwnym razie zachodzi reakcja egzotermiczna.

Denaturację cieplną badałem za pomocą systemu MICROCAL PEAQ-DSC (Malvern Instruments Ltd.). Przeprowadzane przeze mnie eksperymenty DSC składały się z dwóch pomiarów: a) pięć skanów buforu, w którym rozpuszczone były próbki, aby ustalić historię termiczną aparatu i uzyskać powtarzalność linii bazowej; b) pięć skanów próbki RNA w odniesieniu do buforu w którym RNA został rozpuszczony.

Próbki do pomiarów za pomocą skaningowej kalorymetrii różnicowej przygotowałem za pomocą dializy w buforze zawierającym 10 mM kakodylan sodu oraz 100 mM NaCl. Wykonywałem trzy pomiary w buforze o pH 5,2, 6,0 i 7,0. Stężenie RNA w próbce było równe 300 pmol/ μ l. RNA denaturowałem przez 5 min w 95°C, następnie schładzałem w lodzie przez 10 min i inkubowałem w temperaturze pokojowej przez 20 minut. W przypadku próbki z ligandem po inkubacji w temperaturze pokojowej do RNA dodawałem ligand w stężeniu 300 pmol/ μ l. Tak przygotowaną próbkę inkubowałem przez 20 minut w temperaturze pokojowej. Następnie próbki mierzyłem przy pomocy skaningowej kalorymetrii różnicowej w pięciu cyklach w zakresie temperatur (0-110°C), następnie otrzymane wykresy analizowałem przy użyciu programu producenta.

3.2.16. Kriomikroskopia elektronowa

Kriomikroskopia elektronowa (cryo-EM, z ang. *cryogenic electron microscopy*) to technika obrazowania wykorzystywana do wizualizacji struktur przestrzennych biomolekuł. Rewolucjonizuje dziedzinę biologii strukturalnej dostarczając

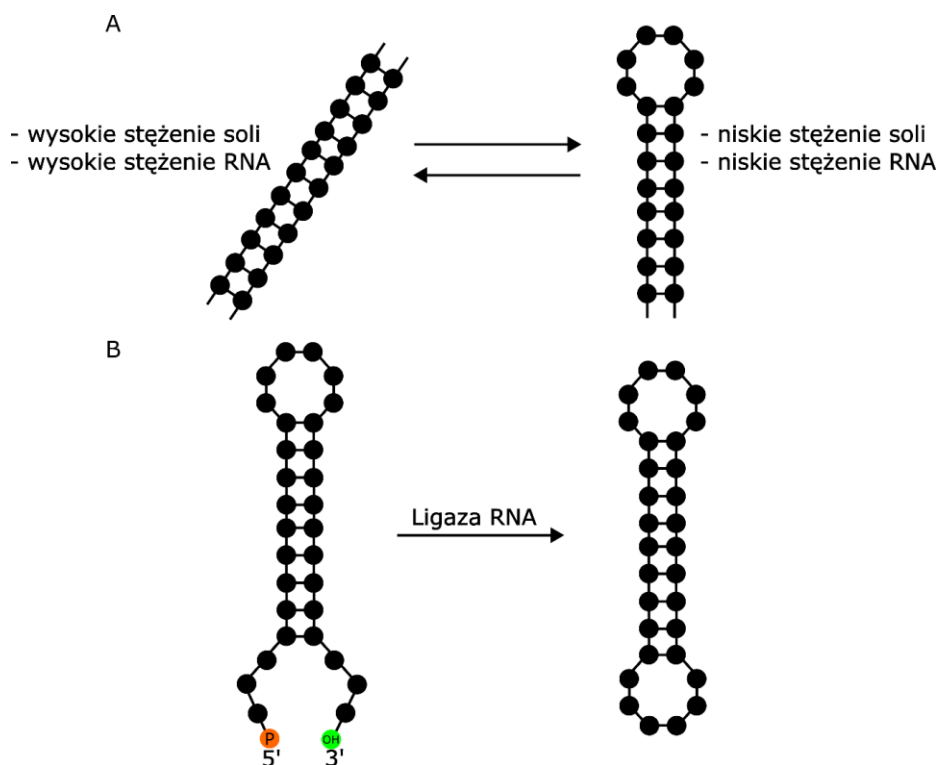
trójwymiarowe obrazy biomolekuł bez konieczności uzyskiwania kryształów. Przygotowanie próbki do cryo-EM polega na szybkim zamrożeniu oczyszczonej próbki badanej biomolekuły w cienkiej warstwie szklanego lodu. Szybkie zamrażanie pomaga zachować naturalną strukturę próbki w możliwie zbliżonym stanie do jej naturalnego stanu. Następnie zamrożoną próbkę umieszcza się w specjalnym mikroskopie krioelektronowym. Mikroskop działa w ekstremalnie niskich temperaturach, aby zapobiec odparowywaniu lub sublimacji lodu w trakcie obrazowania. Wiązka elektronów jest kierowana na próbkę. Elektrony te są rozpraszane na podstawie kształtu cząsteczki. Powstały wzór rozproszonych elektronów jest rejestrowany przez detektor. Detektor rejestruje serię dwuwymiarowych mikrografii elektronowych. Te obrazy 2D reprezentują różne ułożenie cząsteczki w próbce, dzięki czemu można określić wygląd struktury pod różnym kątem. Poprzez połączenie informacji z wielu obrazów wykonanych pod różnymi kątami można zrekonstruować trójwymiarową strukturę biomolekuły. Proces ten obejmuje dopasowanie i połączenie obrazów w celu redukcji szumów [214].

Konstrukty RNA (CUG11-U1a, CUG12-U1a, CUG21-U1a) do badań przy użyciu kriomikroskopu elektronowego przygotowałem w czterech wariantach: o stężeniu 50, 100, 200 i 300 μM . RNA wiązałem z białkiem U1a-RBD w stosunku 1:1 (ilość RNA:ilość białka) w buforze zawierającym 10 mM Tris pH 8,0, 100 mM NaCl, 5 mM MgCl_2 w temperaturze pokojowej. Próbki były mrożone przy użyciu FEI Vitrobot mark IV (parametry mrożenia: *blot time* – 3s, *wait time* – 5s, *drain time* – 0s, *blot force* – 5). Testowe mikrografie były wykonywane przy użyciu kriomikroskopu elektronowego Thermo Fischer Glacios 2 Cryo-TEM. Mrożenie i testowanie było wykonywane w Narodowym Centrum Promieniowania Synchrotronowego SOLARIS w Krakowie.

4. Wyniki

4.1. Opracowanie protokołu do produkcji stabilnych struktur spinkowych RNA do badań krystalograficznych

Pierwszym celem niniejszej pracy było przetestowanie dostępnych protokołów pozyskiwania RNA do badań krystalograficznych i optymalizacja tych protokołów do otrzymywania stabilnych struktur spinkowych. Problemem w krystalografii struktur spinkowych RNA jest promowanie formowania się dupleksu RNA nad strukturą spinkową RNA w warunkach wysokiego stężenia soli oraz wysokiego stężenia RNA, czyli w warunkach panujących w kropli krystalizacyjnej (Rycina 19A). Jednym ze sposobów rozwiązania tego problemu jest cyrkularyzacja spinki RNA z wykorzystaniem ligazy RNA T4, którą zastosowałem w niniejszej pracy (Rycina 19B). Ligaza RNA T4 do przeprowadzenia reakcji cyrkularyzacji wymaga, aby cyrkularyzowany RNA posiadał monofosforan na 5' końcu oraz grupę hydroksylową na 3' końcu. Enzym ten powoduje kowalencyjne połączenie końców 5' i 3', co uniemożliwia rozfałdowanie struktury.



Rycina 19. A – Formowanie się różnych struktur RNA w zależności od warunków. B – Schemat łączenia końców RNA przy użyciu ligazy RNA T4.

Do struktury spinki można dodać sekwencję oddziałującą z fragmentem białka U1a wiążącym się do RNA, aby zwiększyć jej potencjał krystalizacyjny. Podejście to zostało przeze mnie wykorzystane w niniejszej pracy. W przedstawionych badaniach używane białko U1a-RBD, ligaza RNA T4 oraz polimeraza T7 wyprodukowane były przeze mnie poprzez nadekspresję w bakteriach *E. coli*, co pozwoliło na redukcję kosztów. Otrzymane enzymy charakteryzowały się porównywalną aktywnością do komercyjnych odpowiedników.

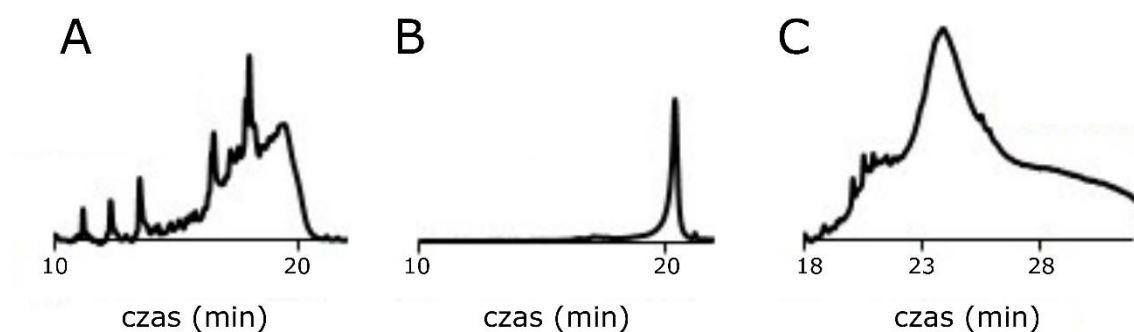
4.1.1. Produkcja RNA na drodze syntezy chemicznej

W związku z pierwszym zadaniem chciałem przetestować dwa protokoły otrzymywania oligomerów: DMT-ON i DMT-OFF. Skupiłem się na homogenności i czasie oczyszczania próbki. Ocenę rozpocząłem od produkcji modelowych sześci nukleotydowych oligomerów. Wybrałem tak krótką sekwencję ze względu na wysoką wydajność jej syntezy. Jednocześnie sprawiło to, że ocena protokołu była łatwiejsza. Oligomery otrzymałem stosując dwa podejścia: DMT-OFF, w którym grupa 5' DMT jest usuwana z ostatniego syntetyzowanego nukleotydu oraz DMT-ON, w którym grupa 5' DMT jest zachowywana i wykorzystywana do dodatkowego etapu oczyszczania. Oba oligomery odciąłem od podłoża, a grupy blokujące usunąłem zgodnie ze standardowymi procedurami. Oligomer DMT-ON oczyściłem przy użyciu kolumniek posiadających powinowactwo do grupy DMTr, podczas gdy oligomer DMT-OFF tylko odsoliłem. Porównanie chromatogramów próbek z wysokosprawnej chromatografii ciekłowej wykazało wysoką czystość i jednorodność RNA oczyszczonego za pomocą kolumniek (Rycina 20B). Oligomer DMT-OFF zawierał wiele krótkich produktów (Rycina 20A), co wskazało na potrzebę dodatkowego etapu oczyszczania z zastosowaniem chromatografii cienkowarstwowej (TLC, z ang. *thin layer chromatography*), elektroforezy w żelu poliakrylamidowym lub HPLC. Powyższa analiza pokazała, że synteza DMT-ON jest szybsza oraz łatwiejsza do produkcji RNA wysokiej czystości (Tabela 4).

Tabela 4. Zestawienie kroków oczyszczania produktów syntezy chemicznej metodami DMT-ON i DMT-OFF.

DMT-ON	DMT-OFF
Odcięcie oligomeru od podłoża	Odcięcie oligomeru od podłoża
Odblokowanie pozycji 2' OH oligomeru za pomocą fluorowodoru trietyloamoniowego	Odblokowanie pozycji 2' OH oligomeru za pomocą fluorku tetrabutylamoniowego
Oczyszczenie RNA za pomocą kolumnienek o powinowactwie do grupy DMTr	Odsolenie RNA
Analityczne HPLC	Oczyszczenie RNA za pomocą żelu poliakrylamidowego, HPLC lub TLC
	Analityczne HPLC

Ze względu na powyższe wyniki postanowiłem, że dłuższy 43-nukleotydowy RNA otrzymam za pomocą metody DMT-ON (Rycina 20C). Jednak heterogenność otrzymanych produktów była bardzo wysoka, co skutkowało niższą wydajnością metody oczyszczania. W związku z tym, do produkcji RNA powyżej 40 nukleotydów zdecydowałem się zastosować inną metodę syntezy.



Rycina 20. Chromatogramy HPLC chemicznie zsyntetyzowanych RNA: A i B – oligomer RNA o długości 6 nt zsyntetyzowany odpowiednio w trybie DMT-OFF i DMT-ON; C – RNA o długości 43 nt zsyntetyzowany w trybie DMT-ON.

4.1.2. Produkcja RNA na drodze transkrypcji *in vitro*

Drugim podejściem, które zastosowałem do uzyskania RNA powyżej 40 nukleotydów do krystalizacji jest jego produkcja na drodze transkrypcji *in vitro*.

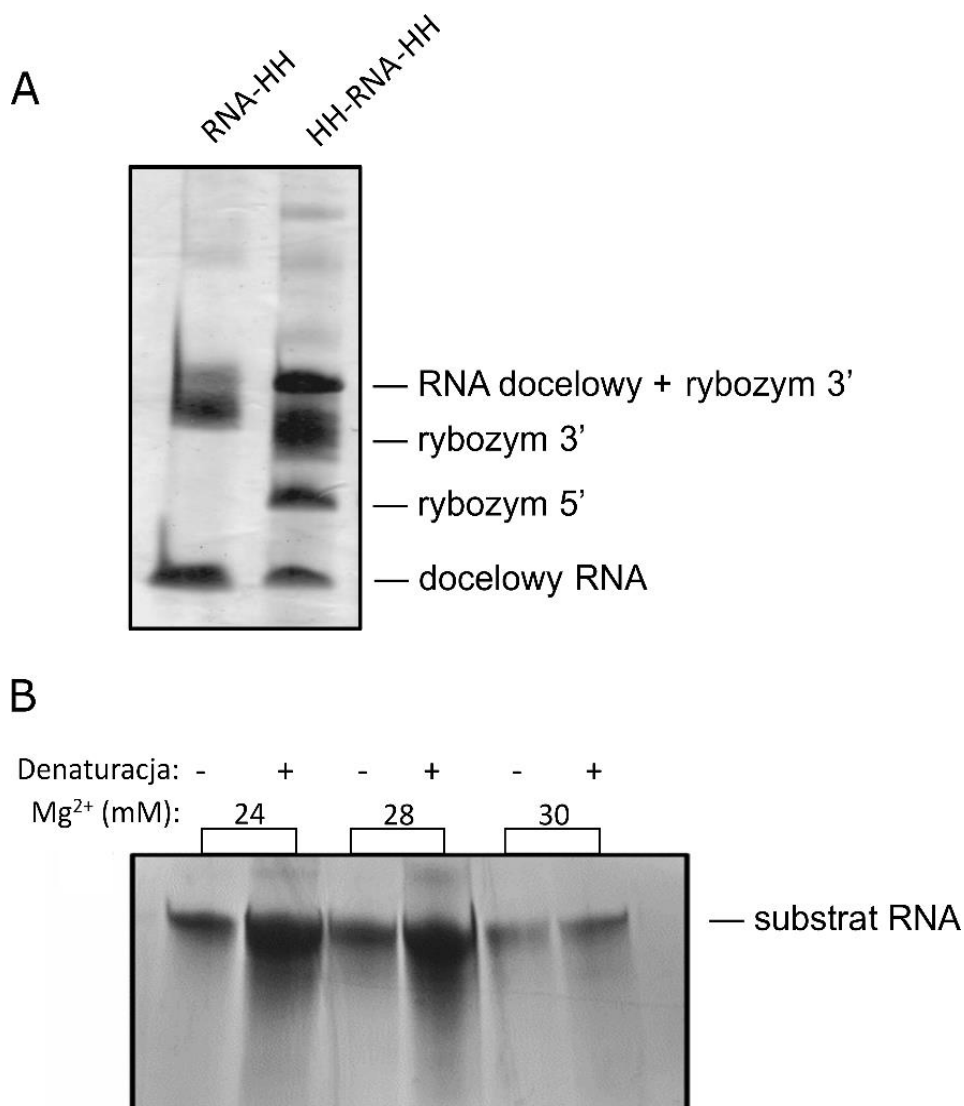
Podczas transkrypcji *in vitro* z wykorzystaniem polimerazy T7 napotyka się najczęściej dwa problemy, które mogą być niepożądane w badaniach krystalograficznych. Pierwszy dotyczy 5' końca RNA. Podczas reakcji polimeraza T7 jako pierwszą transkrybuje resztę guanozyny. Drugim problemem jest dodawanie nukleotydów na 3' końcu przez polimerazę T7, które nie występują w matrycy. Problemy te mogą być rozwiązane przez użycie rybozymów. Aby przetestować skuteczność transkrypcji *in vitro* na matrycy zawierającej rybozomy, zaprojektowałem i zsyntetyzowałem trzy konstrukty różniące się lokalizacją, typem zastosowanego rybozomu oraz długością i sekwencją docelowego RNA (Tabela 5).

Tabela 5. Spis konstruktów z rybozymami użytych w pracy.

Nazwa konstruktu	Użyty rybozym i jego lokalizacja	Długość RNA bez rybozymów (nt)
RNA-HH	Rybozym typu głowa młotka (HH) na 3' końcu	43
HH-RNA-HH	Rybozym typu głowa młotka na 5' oraz 3' końcu	43
HH-RNA-HP	Rybozym typu głowa młotka na 5' oraz rybozym typu spinki (HP) na 3' końcu	43

Konstrukt RNA-HH wykazywał wysoką wydajność transestryfikacji. Oszacowałem, że około 90% transkryptów przereagowało do rybozomu i docelowego RNA (Rycina 21A, ścieżka 1). Projektując następny konstrukt, dodałem rybozym typu głowa młotka na końcu 5' docelowego RNA w matrycy DNA (HH-RNA-HH), aby zwiększyć wydajność transkrypcji *in vitro*. Co zaskakujące, wydajność transestryfikacji spadła do około 50% (Rycina 21A, ścieżka 2). Zaobserwowałem prążek odpowiadający pośredniemu produktowi RNA (RNA docelowe z przyłączonym rybozymem na końcu 3'). Sugeruje to, że obecność rybozomu na końcu 5' docelowego RNA wpływa na nieprawidłowe fałdowanie rybozomu znajdującego się na końcu 3', zmniejszając jego aktywność. Dlatego w kolejnym konstrukcie zamieniłem rybozym typu głowa młotka na końcu 3' docelowego RNA na rybozym typu spinki, aby sprawdzić, czy inny typ rybozomu obecny na końcu 3' może zwiększyć wydajność powstawania produktu RNA. Jednak ta modyfikacja nie spowodowała separacji docelowego RNA od rybozymów w konstrukcie HH-RNA-HP. Optymalizacja warunków reakcji poprzez

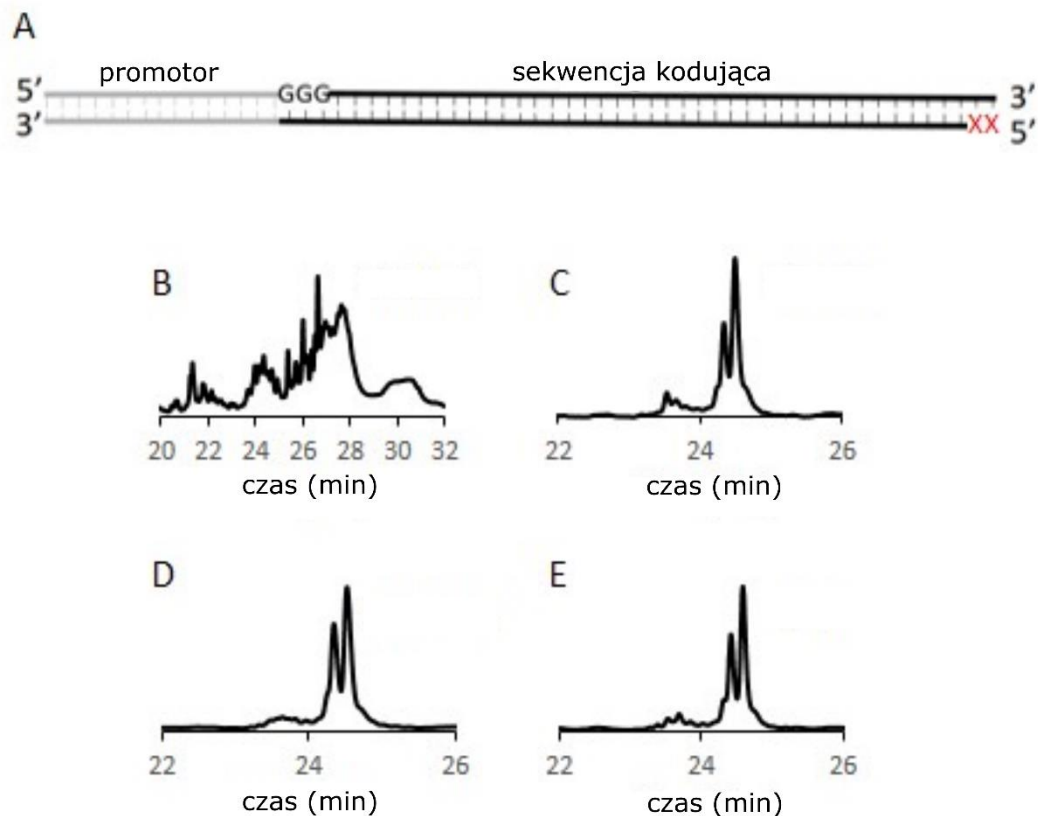
zwiększenie stężenia jonów Mg^{2+} i wykonanie cykli denaturacji-renaturacji nie poprawiły wydajności reakcji (Rycina 21B).



Rycina 21. Ocena wydajności reakcji transestryfikacji przeprowadzonych przez rybozomy. A – konstrukty z rybozymem typu głowa młotka, B – konstrukt z rybozymem typu spinki. Rozdziały przeprowadziłem w 12% żelu poliakrylamidowym.

W związku z powyższymi problemami zastosowałem metodę zmniejszania heterogenności końca 3' RNA za pomocą modyfikacji matrycy DNA do reakcji transkrypcji *in vitro*. Wymagało to wprowadzenia grupy metoksylovej w pozycję drugą atomu węgla C2' jednej lub dwóch reszt nukleotydowych znajdujących się na końcu 5' nici antysensownej matrycy DNA (Rycina 22A). Zaprojektowałem i zsyntetyzowałem matrycę DNA z dwoma resztami nukleotydowymi zawierającymi C2'-metoksył na 5' końcu nici antysensownej kodującej docelowe RNA. Następnie przeprowadziłem

transkrypcję *in vitro* na zmodyfikowanych i niezmodyfikowanych matrycach DNA celem porównania jednorodności produktów RNA. Analiza produktów przy użyciu HPLC wykazała znaczące zmniejszenie heterogenności RNA w przypadku zmodyfikowanej matrycy DNA, gdzie zaobserwowałem tylko dwa piki odpowiadające dwóm RNA o różnej długości (Rycina 22C). W przypadku niezmodyfikowanej matrycy, zaobserwowałem wysoką heterogenność produktów transkrypcji (Rycina 22B). Stosując tę samą metodę, uzyskałem RNA z grupą 5' OH lub grupą monofosforanową na końcu 5', przeprowadzając transkrypcję *in vitro* w obecności odpowiednio guanozyny lub GMP. Analiza produktów transkrypcji przy użyciu HPLC nie wykazała różnic pomiędzy reakcjami transkrypcji *in vitro* z wykorzystaniem guanozyny i GMP (Rycina 22D oraz E).



Rycina 22. Transkrypcja RNA *in vitro* przy użyciu modyfikowanego DNA. A – schematyczne przedstawienie matrycy DNA. Zmodyfikowane nukleotydy są oznaczone czerwonymi krzyżykami. Chromatogramy przedstawiające produkty transkrypcji *in vitro* przy użyciu: B – niezmodyfikowanej matrycy DNA, C – matrycy DNA z dwiema resztami nukleotydowymi zawierającymi C2'-metoksyl na 5' końcu nici antysensownej, D – matrycy DNA z dwiema resztami nukleotydowymi zawierającymi C2'-metoksyl na 5' końcu nici antysensownej w obecności guanozyny, E – matrycy DNA z dwiema resztami nukleotydowymi zawierającymi C2'-metoksyl na 5' końcu nici antysensownej w obecności GMP.

Biorąc pod uwagę długość, sekwencję oraz grupy, które powinny znajdować się na końcach zaprojektowanych przez mnie konstruktów RNA o strukturze spinki zdecydowałem się użyć poniższego protokołu:

- Transkrypcja *in vitro* z wykorzystaniem modyfikowanej matrycy w obecności GMP.
- Oczyszczanie RNA po transkrypcji w 10% żelu poliakrylamidowym w warunkach denaturujących.
- Elektroelucja RNA z żelu.
- Strącanie wyeluowanych z żelu próbek RNA przy użyciu izopropanolu.
- Cyrkularyzacja RNA przy użyciu ligazy RNA T4.
- Oczyszczanie cyrkularyzowanej spinki RNA fenolem i chloroformem.
- Strącanie RNA przy użyciu izopropanolu.
- Wiązanie RNA z białkiem U1a-RBD.

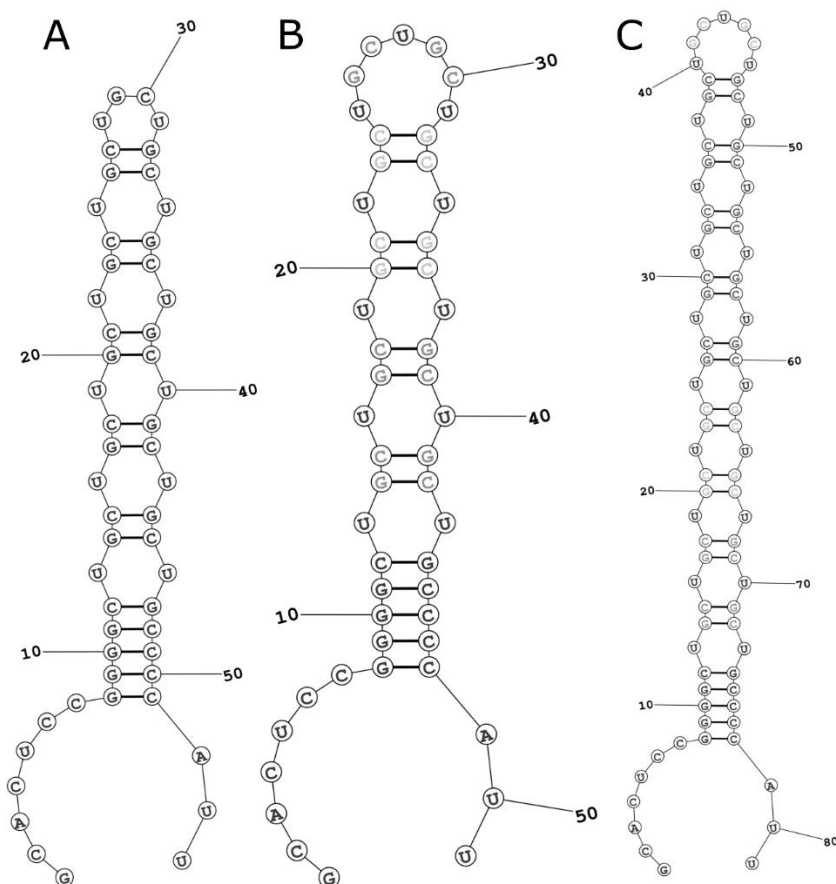
4.1.3. Określenie struktury zaprojektowanych konstruktów

Końcowym etapem zadania pierwszego było podjęcie próby określenia struktury przestrzennej RNA w formie spinki. Zaprojektowałem kilka konstruktów zawierających powtórzenia sekwencji CUG związanej z dystrofią miotoniczną (Tabela 6). Wybrałem powtórzenia CUG ze względu na wcześniejsze udane krystalizacje tych powtórzeń w Zakładzie Struktury i Funkcji Biomolekuł oraz dużą liczbę struktur zawierających te powtórzenia w bazie PDB. Sugerowało to wysokie prawdopodobieństwo uzyskania kryształów tych powtórzeń. Ze względu na możliwość tworzenia tzw. „poślizgniętych” struktur przez powtórzenia CUG, dodałem kilka par GC mających temu przeciwdziałać. Zaprojektowałem trzy konstrukty z 11 (CUG11-U1a), 12 (CUG12-U1a) i 21 powtórzeniami (CUG21-U1a), ponieważ chciałem zaobserwować formowanie struktur spinkowych o pętli cztero- i siedmionukleotydoj. Po przeciwnej stronie od pętli apikalnej złożonej z powtórzeń CUG dodałem sekwencję wiążącą białko U1a-RBD, w której znajdowało się miejsce ligacji. Miejsce ligacji wybrałem ze względu na rozmiary tego regionu jednoniciowego, co znacznie ułatwiało dostępność tego miejsca dla ligazy RNA T4.

Przed przystąpieniem do prób określenia struktur konstruktów RNA zawartych w Tabeli 6 wykonałem predykcje ich struktury drugorzędowej przy użyciu programu RNAStructure (Rycina 23) [70]. Predykcje potwierdziły, że zaprojektowane przeze mnie sekwencje formują struktury spinki.

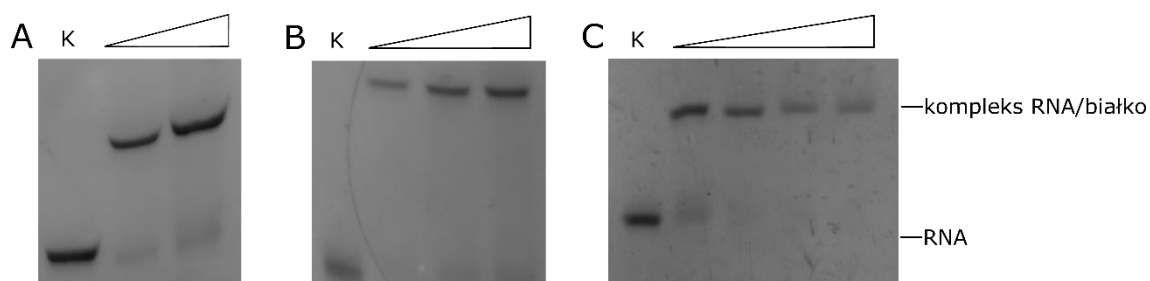
Tabela 6. Zestawienie konstruktów RNA zawierających powtórzenia CUG.

Nazwa konstrukt	Długość (nt)	Masa cząsteczkowa (g/mol)
CUG11-U1a	51	16425
CUG12-U1a	54	17592
CUG21-U1a	81	25991



Rycina 23. Predykcje struktury drugorzędowej zaprojektowanych konstruktów przeprowadzone w programie RNAStructure. A – CUG11-U1a, B – CUG12-U1a, C – CUG21-U1a.

Kolejnym krokiem była produkcja zaprojektowanych konstruktów. Do tego celu użyłem wcześniej ustalonego przeze mnie protokołu. Następnie sprawdzałem, czy wyprodukowane przeze mnie białko U1a-RBD wiąże zaprojektowane konstrukty przy pomocy metody EMSA (Rycina 24).



Rycina 24. Oddziaływanie konstruktów RNA z białkiem U1a-RBD (A – CUG21-U1a, B – CUG11-U1a, C – CUG12-U1a). RNA (30 pmoli) inkubowałem w samym buforze reakcyjnym (K) lub z dodatkiem białka U1a-RBD (w kolejnych reakcjach ilość białka U1a-RBD wzrastała, odpowiednio: A: 30, 60 pmoli; B: 30, 45 i 60 pmoli; C: 30, 45, 60, 90 pmoli). Trójkątem oznaczyłem wzrastające stężenia białka U1a-RBD. Rozdział prowadziłem w 10% żelu poliakrylamidowym.

Następnie otrzymane konstrukty zostały przeze mnie wykorzystane do krystalizacji. Krystalizację prowadziłem w kilku różnych stężeniach RNA (0,5; 0,75 i 1 mM) dobierając również różne stosunki stężenia RNA:białko (1:1, 1:2, 1:3) oraz różne temperatury krystalizacji (4°C, 19 °C, 30 °C). W każdym wariancie testowałem 192 różne warunki krystalizacji. W żadnym z przetestowanych warunków nie otrzymałem kryształów.

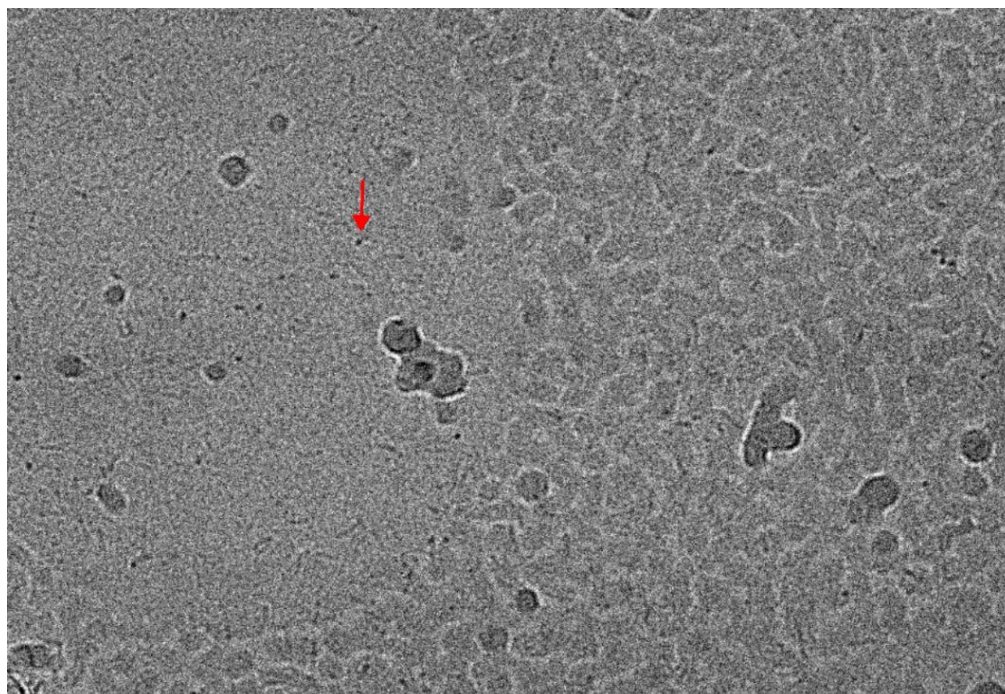
Po nieudanych próbach krystalizacji postanowiłem podjąć próbę określenia struktury tych konstruktów za pomocą kriomikroskopii elektronowej. W tej metodzie, podobnie jak w krystalografii, jakość i homogenność próbki mają kluczowe znaczenie. Przed przystąpieniem do właściwego pomiaru testuje się przygotowane próbki. Weryfikację próbek wykonuje się poprzez wizualną inspekcję testowych mikrografii. Związku z tym miały miejsce trzy sesje testowe (Tabela 7).

Tabela 7. Zestawienie sesji testowych wykonywanych przy użyciu kriomikroskopu elektronowego.

Sesja	Testowane próbki	Stężenia RNA (pmol/μl)
1	CUG11-U1a	300, 200
	CUG12-U1a	300, 200

2	CUG12-U1a	100, 50
	CUG21-U1a	100, 50
3	CUG-21-U1a	100

Podczas tych prób przygotowałem szereg rozcieńczeń próbek tak, aby cząsteczki RNA były jak najlepiej widoczne na mikrografiach (każdy konstrukt był przeze mnie wykonywany w stężeniach 50, 100, 200 i 300 pmol/μl, kompleksy RNA/białko w stosunku ilościowym 1:1). Niestety dla otrzymanych kompleksów RNA/białko napotkałem trudności w identyfikacji cząsteczek RNA. (Rycina 25).



Rycina 25. Mikrografia z kriomikroskopu elektronowego. Strzałką zaznaczyłem prawdopodobną cząsteczkę RNA CUG21-U1a.

4.2. Struktura krystaliczna oligomeru G_2C_4

Drugim zadaniem wykonanym w tej pracy doktorskiej było uzyskanie struktury krystalicznej RNA zawierającego sekwencję 5' GGCCCC 3' (G_2C_4). Uzyskałem dwie struktury: strukturę oligomeru G_2C_4 oraz strukturę tego oligomeru w kompleksie z cząsteczką ANP77.

4.2.1. Krystalizacja kompleksu G₂C₄ z ANP77 oraz oligomeru G₂C₄

Oligomer RNA G₂C₄ zsyntetyzowałem chemicznie i oczyściłem metodą DMT-ON. Ligand ANP77 użyty w pracy otrzymałem w ramach współpracy z grupą prof. Nakataniego. Kryształy otrzymałem techniką siedzącej kropli z dyfuzją par w temperaturze 19°C. Kropla zawierała 0,5 µl RNA oraz 0,5 µl roztworu rezerwuarowego (0,08 M chlorek sodu, 0,04 M kakodylan sodu pH 6,0, 45% 2-metylo-2,4-pentanodiol, 0,012 M tetrachlorowodorek sperminy).

W przypadku oligomeru G₂C₄ bez ANP77 kryształy otrzymałem techniką siedzącej kropli z dyfuzją par w temperaturze 19°C podobnie, jak w przypadku konstruktu stanowiącego kompleks G₂C₄ z syntetyczną cząsteczką ANP77. Kropla zawierała 0,5 µl RNA oraz 0,5 µl roztworu rezerwuarowego (0,08 M chlorek sodu, 0,02 M chlorek magnezu, 0,04 M kakodylan sodu pH 6,0, 35% 2-metylo-2,4-pentanodiol, 0,012 M tetrachlorowodorek sperminy).

4.2.2. Rejestracja i przetwarzanie danych dyfrakcyjnych

Pomiary dyfrakcyjne kryształów G₂C₄ z cząsteczką ANP77 i bez tej cząsteczki przeprowadziłem przy użyciu dyfraktometru XtalLab-R. Zarejestrowałem obrazy dyfrakcyjne o maksymalnej rozdzielczości 1,1 Å (wariant z ANP77) i 1,5 Å (wariant bez ANP77). Wyzaczyłem grupę przestrzenną C222₁ dla obu wariantów. Szczegółowa charakterystyka danych dyfrakcyjnych znajduje się w Tabeli 8.

Tabela 8. Dane krystalograficzne oraz charakterystyka zbioru danych dla kryształów struktury konstrukt G₂C₄ w obecności i bez cząsteczki ANP77. W nawiasie znajdują się wartości dotyczące ostatniego zakresu rozdzielczości.

	G ₂ C ₄ -ANP77	G ₂ C ₄
Statystyki danych dyfrakcyjnych		
Grupa przestrzenna	C222 ₁	C222 ₁
Parametry komórki (Å, °)	a = 29,34 Å, b = 45,18 Å, c = 86,28 Å α = 90°, β = 90°, γ = 90°	a = 41,76 Å, b = 46,64 Å, c = 51,49 Å α = 90°, β = 90°, γ = 90°
Rozdzielczość (Å)	24,62-1,1 (1,16-1,1)	21,24-1,5 (1,58-1,5)

R_{merge}	0,076 (0,560)	0,116 (0,553)
I/σ	26,9 (4,4)	10 (2,4)
Kompletność (%)	99,9 (99,4)	99,9 (100)
Redundancja	16,3 (11)	7,3 (4,8)
Liczba unikalnych refleksów	23771	8334
Statystyki udokładniania struktur		
Liczba refleksów: work/test	22774/949	7905/404
Średnia wartość czynnika B (\AA^2)	12,508	11,155
$R_{\text{work}}/R_{\text{free}}$	0,1287/0,1634	0,1742/0,2224
Liczba atomów cząsteczek RNA	492	492
Liczba atomów cząsteczek wody	111	116
RMSD na wiązaniach (\AA)	0,016	0,010
RMSD na kątach ($^\circ$)	1,930	1,721

4.2.3. Rozwiązanie i udokładnianie struktury kompleksu oraz oligomeru G_2C_4 bez liganda ANP77

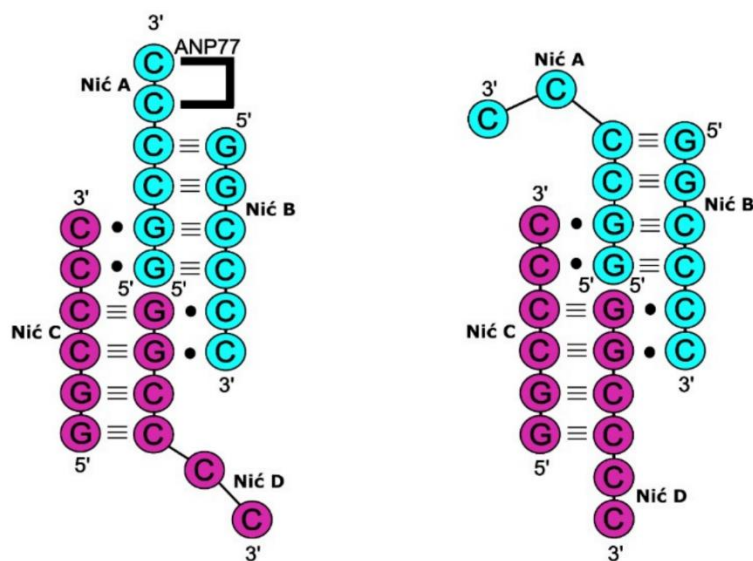
W celu rozwiązania i udokładnienia struktury kompleksu oligomeru RNA G_2C_4 z cząsteczką ANP77 wykonałem następujące kroki. Najpierw przeprowadziłem analizę zawartości komórki elementarnej przy pomocy współczynnika Matthews [215], która wykazała obecność czterech nici RNA przy obecności rozpuszczalnika w 59,27 % (prawdopodobieństwo na poziomie 51%, współczynnik Matthews = 2,04). Następnie strukturę rozwiązałem używając programu PHASER_MR [216]. Do podstawienia molekularnego użyłem struktury dupleksu (CCCCGG)₃CCCC (kod PDB: 5ew7), z którego wyodrębniłem sekwencję 5' GGCCCC 3'. Do udokładniania struktury użyłem programu REFMAC5, natomiast przebudowę modelu prowadziłem przy użyciu programu COOT. Po udokładnieniu struktura posiadała następujące współczynniki: $R = 0,1287$ i $R_{\text{free}} = 0,1634$. Wzór strukturalny liganda ANP77 wykonałem przy użyciu programu Marvin JS, który pozwolił mi uzyskać wzór SMILES tego związku.

Następnie wzór ten posłużył do wygenerowania plików .pdb i .cif, niezbędnych do wprowadzenia liganda do struktury G₂C₄ i dalej do udokładniania. Pliki .pdb i .cif wygenerowałem używając Grade Web Server (https://grade.globalphasing.org/cgi-bin/grade2_server.cgi).

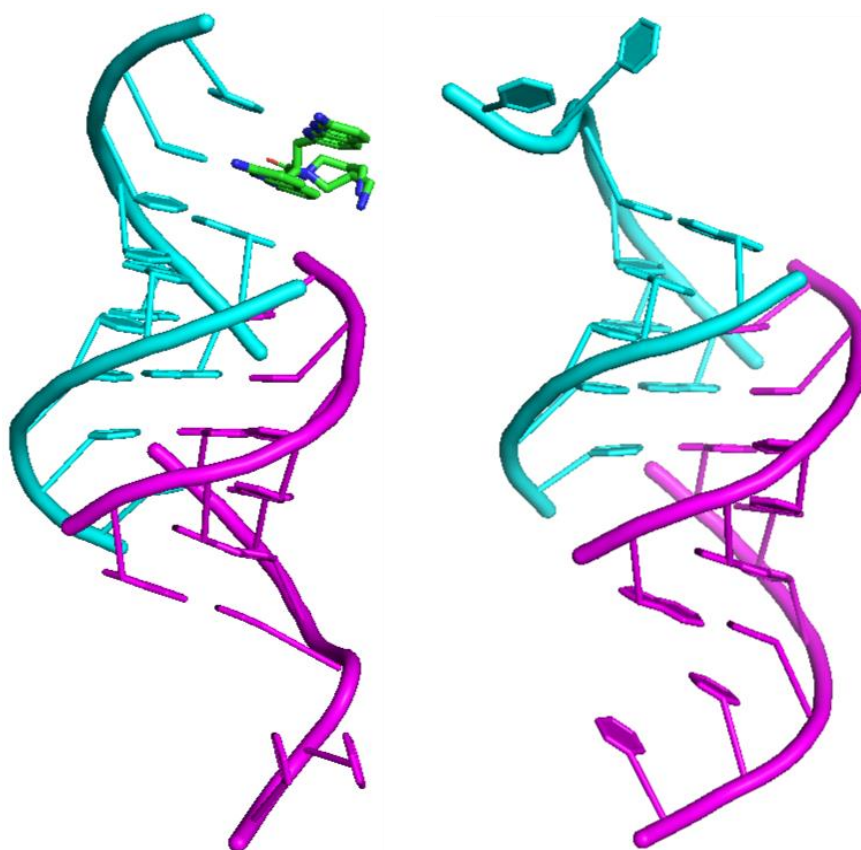
Rozwiązanie i udokładnianie struktury G₂C₄ bez ANP77 przebiegało podobnie do rozwiązania i udokładniania struktury G₂C₄ z ANP77. Analiza zawartości komórki elementarnej przy pomocy współczynnika Matthews'a wykazała obecność czterech nici RNA przy obecności rozpuszczalnika w 65,17 % (prawdopodobieństwo na poziomie 56%, współczynnik Matthews'a = 2,38). Strukturę rozwiązałem używając programu PHASER_MR. Do podstawienia molekularnego użyłem struktury G₂C₄ z ligandem ANP77, którą wcześniej udokładniłem. Do udokładniania struktury użyłem programu REFMAC5, natomiast przebudowę modelu prowadziłem przy użyciu programu COOT. Po udokładnieniu struktura posiadała następujące współczynniki: R = 0,1742 i R_{free} = 0,2224.

4.2.4. Struktura G₂C₄ i G₂C₄ z cząsteczką ANP77

Struktury RNA zawierające sekwencję G₂C₄ z cząsteczką ANP77 oraz bez tej cząsteczki są do siebie bardzo podobne. Składają się z dwóch oddziałujących ze sobą dupleksów, które tworzą jeden tetramer (Rycina 26 – model 2D, Rycina 27 – model 3D).



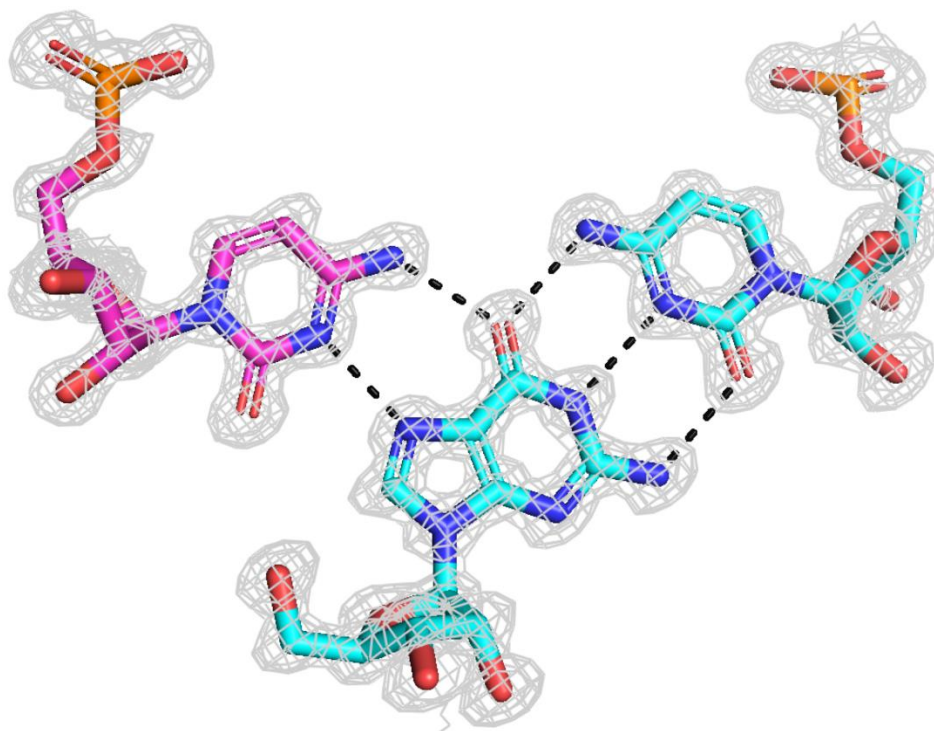
Rycina 26. Drugorzędowa struktura G₂C₄ z cząsteczką ANP77 i bez ANP77. Kolorami cyjan i fioletowym zazaczyłem dupleksy tworzące tetramer.



Rycina 27. Struktura trzeciorzędowa G_2C_4 z ligandem oraz bez liganda ANP77. Kolorami cyjan i fioletowym oznaczyłem dwa dupleksy w strukturze tetramery. Kolorem zielonym oznaczyłem ligand ANP77.

Dwie nici tetramery (A i B) tworzą jeden dupleks, natomiast nici C i D tworzą drugi dupleks. W każdym dupleksie tworzą się cztery kanoniczne pary GC, które oflankowane są dwiema wystającymi resztami cytydyny. Dwie reszty cytydyny znajdujące się na końcu 5' nici A oddziałują z cząsteczką ANP77. W strukturze krystalicznej, dupleksy AB i CD oddziałują między sobą. Dwie reszty cytozyny z nici C oddziałują z dupleksem AB, a dwie reszty cytozyny z nici B oddziałują z dupleksem CD. Jednocześnie warstwowe ułożenie reszt guanozyny w niciach A i D na końcach 5' sprawia wrażenie, że stworzona jest jedna długa nić RNA. Struktura ta w pewnym fragmencie przyjmuje formę tripleksu, który flankowany jest kanonicznymi parami Watsona-Cricka CG tworzącymi w tym fragmencie dupleks, po których następuje jednoniciowy fragment dwóch wystających reszt cytydyny. Motyw tripleksu jest złożony z czterech warstwowo ułożonych trójek reszt nukleotydowych $C^+ \cdot G \cdot C$ (C^+ oznacza protonowaną resztę cytozyny). Reszta guaniny tworzy parę Watsona-Cricka z resztą cytozyny w ramach dupleksu. Ta sama reszta guaniny tworzy parę Hogsteena z protonowaną resztą cytozyny pochodzącą z sąsiedniego dupleksu.

Oddziaływanie pomiędzy protonowaną resztą cytozyny a resztą guaniny odbywa się poprzez dwa wiązania wodorowe: jedno wiązanie wodorowe pomiędzy grupą egzoaminową reszty cytozyny i karbonylową reszty guaniny, drugie natomiast pomiędzy atomem N3 reszty cytozyny i N7 reszty guaniny (Rycina 28). Wiązanie to jest możliwe dzięki protonacji reszty cytydyny poprzez konwersję grupy iminowej na aminową.

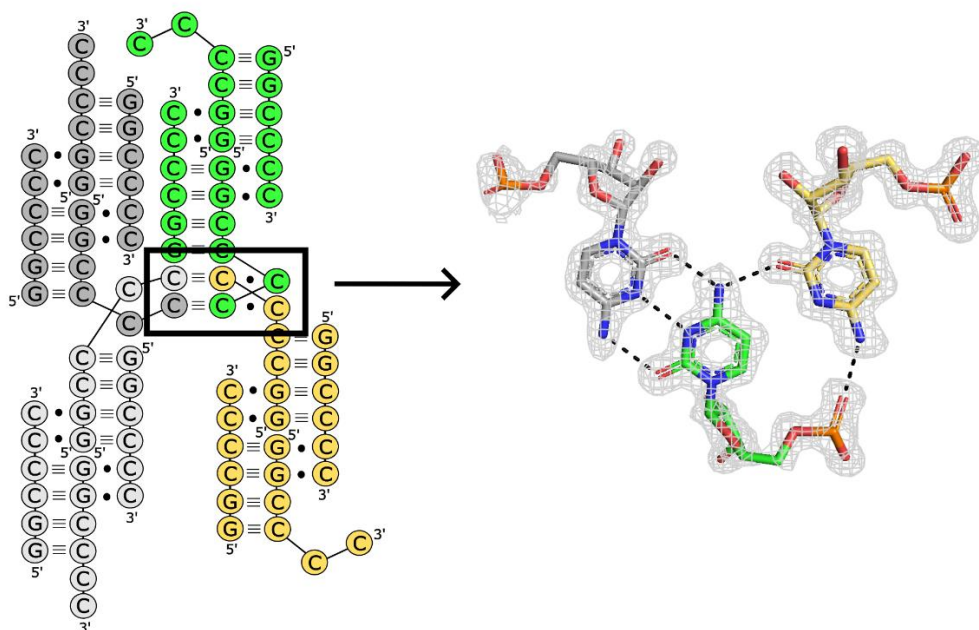


Rycina 28. Przykład oddziaływania dwóch reszt cytydyny i jednej reszty guaniny w strukturze G_2C_4 . Kolorem cyjan oznaczone są reszty nukleotydowe z dupletu AB, natomiast kolorem fioletowym reszta cytydyny z nici C. Przerywaną linią oznaczyłem wiązania wodorowe.

4.2.5. Ułożenie reszt cytozyny w sieci krystalicznej struktury G_2C_4 bez cząsteczki ANP77

W strukturze sekwencji G_2C_4 bez liganda ANP77 wystające z tetrameru reszty cytozyny biorą udział w oddziaływaniach cząsteczek w sieci krystalicznej. Dwa tryplety formowane są przez reszty cytozyny pochodzące z czterech tetramerów. Pierwsza trójka powstaje poprzez oddziaływania piątej reszty cytozyny z nici D (Rycina 29, kolor zielony) z resztą cytozyny z pozycji szóstej nici A' z symetrycznej cząsteczki (Rycina 29, kolor jasnoszary) i resztą cytozyny z pozycji szóstej nici D' z symetrycznej cząsteczki (Rycina 29, kolory żółty). Natomiast druga trójka powstaje poprzez oddziaływania reszty

cytozyny z pozycji szóstej nici D (Rycina 29, kolor zielony) z resztą cytozyny z pozycji szóstej nici A' (Rycina 29, kolor ciemnoszary) oraz z resztą cytozyny z pozycji piątej nici D' (Rycina 29, kolor żółty) z symetrycznych cząsteczek. Na drodze tych kontaktów powstaje oddziaływanie C-C⁺•C, gdzie jedna z reszt cytozyny (atom N3) jest protonowana (Rycina 29) i tworzy trzy wiązania wodorowe z drugą resztą cytozyny w konformacji *trans*. Trzecia cytozyna z trypletu oddziałuje poprzez dwa wiązania wodorowe z jedną z reszt cytozyny z pary C•C⁺.

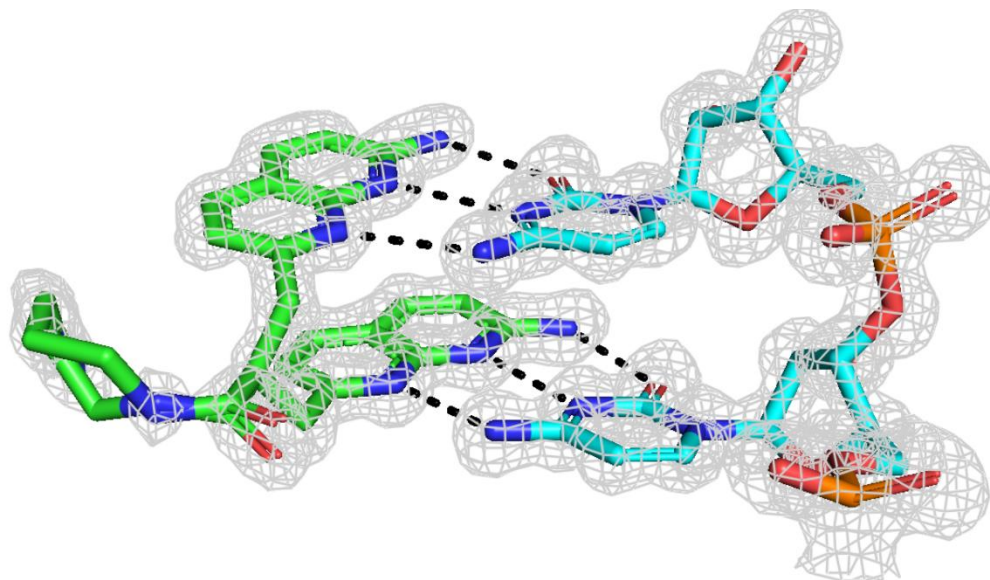


Rycina 29. Ułożenie reszt cytydyny w tryplecie międzycząsteczkowym bez udziału ANP77. Każdy kolor odpowiada innej cząsteczce tetrameru z sieci krystalicznej.

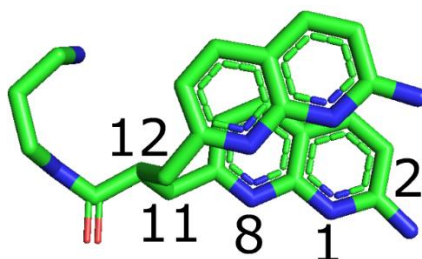
4.2.6. Wiązanie G₂C₄ z ANP77

Dwie wystające reszty cytozyny z tetrameru G₂C₄ są idealnym miejscem wiązania ANP77. Ligand w konformacji typu *stacked* (Rycina 30) oddziałuje z dwiema wystającymi resztami cytozyny w nici A poprzez trzy wiązania wodorowe tworząc pseudokanoniczne pary zasad (Rycina 30). Pierwsze wiązanie obecne jest pomiędzy atomem N4 reszty cytozyny z atomem N8 cząsteczki ANP77. Drugie wiązanie zachodzi pomiędzy atomem N3 reszty cytozyny i atomem N1 ANP77, natomiast trzecie wiązanie obecne jest pomiędzy atomem O2 reszty cytozyny z grupą 2-aminową liganda (Rycina 30 i 31). Grupa karbonylowa cząsteczki ANP77 oddziałuje z cząsteczką wody. Z tą samą

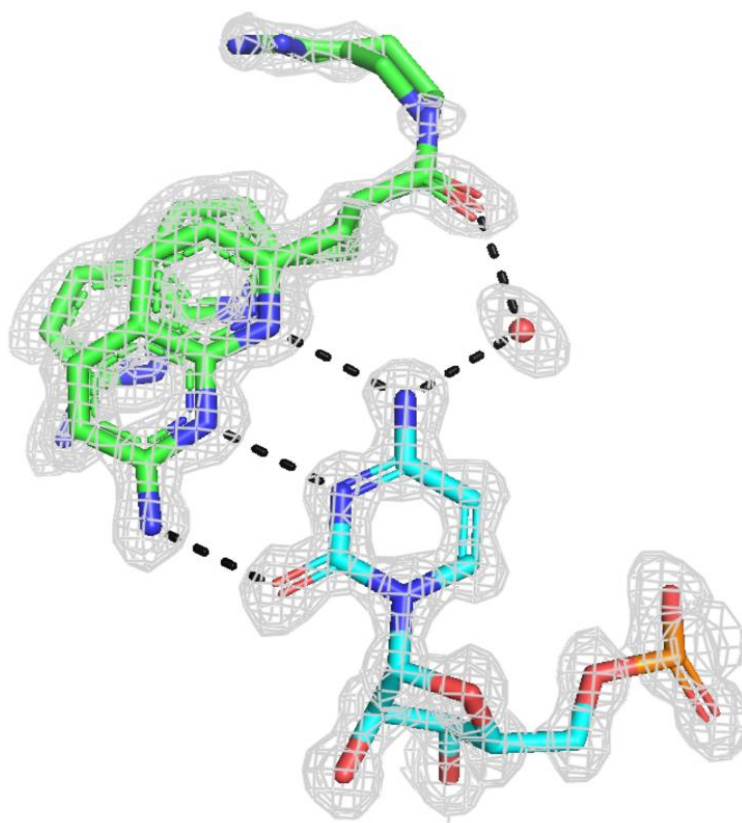
cząsteczką wody oddziałuje grupa egzoaminowa reszty cytozyny znajdującej się w pozycji 5 nici A tetrameru (Rycina 32).



Rycina 30. Oddziaływanie liganda (kolor zielony) z resztami cytydyny nici A tetrameru G_2C_4 (kolor cyjan). Przerywaną linią oznaczyłem wiązania wodorowe.



Rycina 31. Numeracja atomów liganda ANP77.

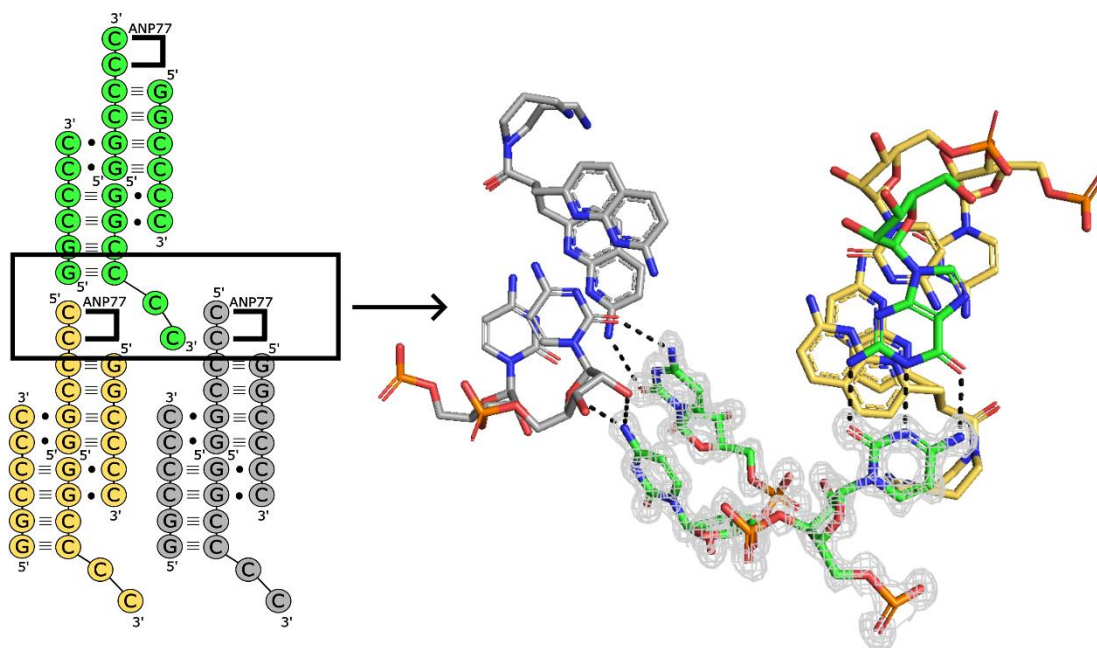


Rycina 32. Oddziaływania cząsteczki ANP77 (kolor zielony) i reszty cytydyny na pozycji 5 nici A tetrameru (kolor cyjan) z wodą. Wiązania wodorowe oznaczyłem przerywaną linią.

4.2.7. Ułożenie reszt cytozyny w sieci krystalicznej struktury G₂C₄ z cząsteczką ANP77

Ligand ANP77 bierze udział w oddziaływaniach pomiędzy cząsteczkami tetrameru w sieci krystalicznej (Rycina 33). W strukturze z ligandem, wystające reszty cytydyny z nici D, które nie oddziałują z cząsteczką ANP77, biorą udział w tworzeniu kontaktów między cząsteczkami tetrameru w sieci krystalicznej. Reszta cytozyny z pozycji szóstej nici D (Rycina 33, kolor zielony) oddziałuje zarówno z ligandem, jak i resztą cytozyny z pozycji szóstej nici A' cząsteczki symetrycznej (Rycina 33, kolor szary). Reszta cytozyny z pozycji piątej nici D (Rycina 33, kolor zielony) oddziałuje z pierścieniem rybozy reszty cytydyny z pozycji szóstej nici A' cząsteczki symetrycznej (Rycina 33, kolor szary). Ostatnia para zasad, powstająca pomiędzy resztą cytozyny z pozycji czwartej nici D a resztą guaniny z pozycji pierwszej nici C (Rycina 33, kolor

zielony), jest ułożona warstwowo nad oddziaływaniem dwóch reszt cytozyny z nici A' z ligandem ANP77 z cząsteczki symetrycznej (Rycina 33, kolor żółty).

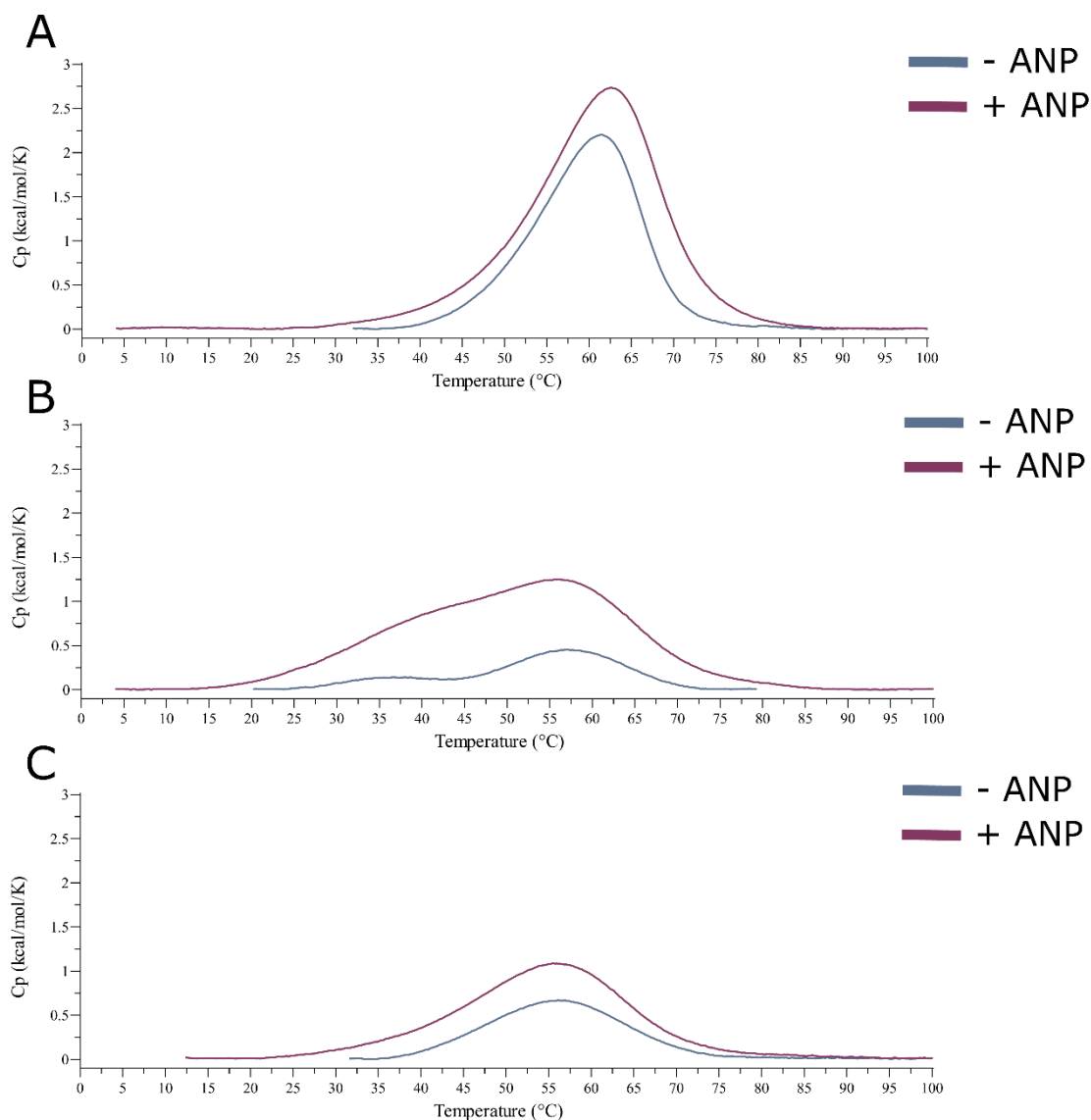


Rycina 33. Ułożenie reszt cytydyny w kontaktach w sieci krystalicznej z udziałem ANP77. Kolorami zielonym, żółtym i szarym oznaczyłem trzy tetrametry G_2C_4 .

4.2.8. Skaningowa kalorymetria różnicowa oligomeru G_2C_4

Biofizyczną ewaluację kompleksu G_2C_4 z ligandem ANP77, jak i samego oligomeru G_2C_4 wykonałem przy użyciu skaningowej kalorymetrii różnicowej w trzech warunkach pH: 5,26, 6,0 oraz 7,0 (Rycina 34).

Na samym początku przeprowadziłem pomiary w pH 6,0, czyli w warunku, w którym RNA podlegał krystalizacji (Rycina 34B). Dla próbki bez liganda zaobserwowałem dwa piki, które oznaczały temperatury topnienia struktur formowanych przez G_2C_4 : T_{m1} była równa $37,42^\circ\text{C}$ (T_m – temperatura topnienia), natomiast T_{m2} wyniosła $57,24^\circ\text{C}$. W przypadku kompleksu G_2C_4 z ligandem ANP77 zaobserwowałem przesunięcie obu pików: T_{m1} wzrosła o $4,53^\circ\text{C}$ ($T_{m1} = 41,95^\circ\text{C}$), natomiast T_{m2} wzrosła tylko o $0,66^\circ\text{C}$ ($T_{m2} = 57,90^\circ\text{C}$). Mając na uwadze, jak ważna jest wartość pH dla protonacji reszt cytozyny zdecydowałem się przeprowadzić podobne eksperymenty w pH 7,0 i 5,26.



Rycina 34. Wyniki DSC dla oligomeru G_2C_4 bez i w obecności liganda ANP77 w pH: 5,26 (A), 6,0 (B), 7,0 (C).

Dla próbki bez liganda w pH 7,0 obserwowałem tylko jeden pik (Rycina 34C). Dodanie liganda powodowało obniżenie temperatury topnienia o 1°C w pH 7,0. Temperatura topnienia oligomeru bez liganda była równa $55,92^\circ\text{C}$, natomiast z ligandem wyniosła $54,92^\circ\text{C}$. W pH 5,26, tak samo jak w przypadku pH 7,0, obserwowałem tylko jeden pik dla próbki bez, jak i z ligandem (Rycina 34A). Dla próbki z ligandem obserwowałem przesunięcie się pik temperatury topnienia o około $1,25^\circ\text{C}$ względem próbki bez liganda. W pH 5,26 temperatura topnienia próbki bez liganda była równa $60,04^\circ\text{C}$, natomiast z ligandem wyniosła $61,29^\circ\text{C}$. Wartości temperatur topnienia obu próbek w pH 5,26 były wyższe od wartości T_{m2} uzyskanych w przypadku eksperymentów w pH 6,0 o odpowiednio $2,8^\circ\text{C}$ i $3,39^\circ\text{C}$ dla próbki bez liganda

i z ligandem. Podsumowanie otrzymanych przeze mnie wyników przy użyciu skaningowej kalorymetrii różnicowej dla oligomeru G₂C₄ z i bez liganda ANP77 w trzech różnych pH przedstawiłem w Tabeli 9.

Tabela 9. Tabela przedstawiająca temperatury topnienia dla oligomerów G₂C₄ w wariantach z i bez ANP77 w różnych warunkach pH.

Testowany wariant	Tm₁ (°C)	Tm₂ (°C)
pH 5,26 - ANP77	60,04	-
pH 5,26 + ANP77	61,29	-
pH 6,0 - ANP77	37,42	57,24
pH 6,0 + ANP77	41,95	57,90
pH 7,0 - ANP77	55,92	-
pH 7,0 + ANP77	54,92	-

5. Dyskusja

Głównym problemem w krystalizowaniu RNA o strukturze spinki jest to, że formowanie się tej struktury faworyzowane jest w warunkach niskiego stężenia soli oraz niskiego stężenia RNA. Wysokie stężenie RNA i wysokie stężenie soli są nieodłącznymi elementami towarzyszącymi krystalizacji i są to warunki, które faworyzują formowanie się dupleksów. Aby rozwiązać ten problem, w niniejszej rozprawie zaproponowałem kowalencyjne zamknięcie spinki RNA używając ligazy RNA T4. Uformowany w ten sposób cyrkularyzowany łańcuch RNA byłby bardziej stabilny i mniej podatny na rozfałdowanie od struktury spinki. W tym podejściu jest kilka aspektów, które trzeba wziąć pod uwagę projektując eksperyment. Ligaza RNA T4 wymaga do optymalnej cyrkularyzacji RNA monofosforanu na końcu 5' RNA oraz grupy hydroksylowej na końcu 3' [217]. Stanowi to duże ograniczenie w swobodnym pozyskiwaniu materiału do badań krystalograficznych. W niniejszej pracy przetestowałem dostępne protokoły służące do pozyskiwania RNA oraz stworzyłem swojego rodzaju przewodnik o tym, jak skutecznie pozyskiwać RNA do badań strukturalnych w zależności od potrzeb [78]. Istnieje wiele czynników, które muszą zostać wzięte pod uwagę podczas projektowania RNA do badań strukturalnych takie jak: długość oligomeru, czas i koszt produkcji, dostępny sprzęt czy obecność modyfikacji chemicznych w kwasie nukleinowym.

Poza wymaganiami dotyczącymi grup funkcyjnych obecnych na końcach RNA, istotnym aspektem jest produkcja kwasów nukleinowych do badań krystalograficznych. W badaniach strukturalnych niezbędne są nanomolowe ilości homogenego materiału o ściśle zdefiniowanych końcach. Istnieją trzy metody pozyskiwania RNA: izolacja z materiału biologicznego, synteza chemiczna i transkrypcja *in vitro*. Do badań strukturalnych nie używa się izolacji z materiału biologicznego przez małe ilości pozyskiwanego pożądanego RNA. Dlatego też do badań krystalograficznych używa się syntezy chemicznej lub transkrypcji *in vitro* [78].

Badania przeprowadzone w tej pracy oraz moje doświadczenie pokazują, że kilkunukleotydowe RNA mogą być wydajnie produkowane za pomocą syntezy chemicznej. Dłuższe RNA (powyżej 40 nukleotydów) mogą być pozyskiwane tą metodą ze znacznie mniejszą wydajnością. Dla wysokiej wydajności syntezy ważnym jest,

aby wszystkie reagenty zostały rozpuszczone w acetonitrylu o jak najniższej zawartości wody (do 20 ppm). Wysokie uwodnienie acetonitrylu może skutkować rozpadaniem się reagentów i konsekwencji prowadzić do niskich wydajności syntezy. Podczas pracy zauważyłem, że najlepszą praktyką jest przygotowywanie reagentów na świeżo, gdyż z upływem czasu reagenty mogą się uwadniać [78]. Drugim ważnym czynnikiem jest używanie wysokiej jakości amidofosforynów. Rekomendowane jest, aby używać amidofosforynów z blokadą TOM nad TBDMS. Jest to związane z migracją blokady TBDMS z węgla 2' na węgiel 3' rybozy w warunkach zasadowych, co prowadzi do odsłonięcia grupy 2'. W konsekwencji mogą powstać nieaktywne biologicznie wiązania 2'-5' pomiędzy kolejnymi resztami w nowo zsyntetyzowanym RNA [218]. Koszty komercyjnego zakupu amidofosforynów TOM i TBDMS są porównywalne. Kolejnym ważnym elementem w syntezie chemicznej RNA jest jego oczyszczanie po syntezie. RNA, który nie zawiera modyfikacji lub innych grup niż grupa hydroksylowa na końcu 5' powinien być oczyszczany metodą DMT-ON. Ta metoda oczyszczania jest szybsza i łatwiejsza, niż oczyszczanie metodą DMT-OFF. Dla długich RNA, niezbędne jest zastosowanie kolejnych etapów oczyszczania materiału, włączając oczyszczanie RNA w żelach poliakrylamidowych, oczyszczanie przy użyciu HPLC w warunkach denaturujących lub TLC.

Dłuższe RNA (powyżej 40 nukleotydów) mogą być wydajniej produkowane przy użyciu transkrypcji *in vitro*. Główną wadą tej metody jest produkcja heterogennego końca 3' RNA przez polimerazę T7. Problem ten może zostać rozwiązany poprzez użycie modyfikowanej matrycy DNA. W takiej matrycy dwa ostatnie nukleotydy (lub tylko ostatni nukleotyd) są modyfikowane i zawierają grupę 2'-metoksyłową [85,219]. Okazuje się, że takie działanie istotnie obniża heterogenność próbki RNA, jednak dla badanych przeze mnie oligonukleotydów RNA nie wyeliminowało to tego problemu. Obecność dwóch odrębnych pików na chromatogramie pokazuje, że RNA o takim obrazie rozdziału może zostać skutecznie rozdzielone używając HPLC. Zaletą metody wykorzystującej modyfikowaną matrycę DNA do transkrypcji *in vitro* jest łatwe uzyskanie monofosforanu na końcu 5' pożądanego RNA. Monofosforan na 5' końcu RNA jest niezbędny do reakcji cyrkularyzacji RNA przy udziale ligazy RNA T4. W takiej sytuacji do reakcji transkrypcji *in vitro* należy dodać 5' monofosforan guanozyny w stężeniu około 3 razy wyższym niż stężenie pojedynczego 5' trifosforanu guanozyny. Metoda ta nie pozwala jednak na wyeliminowanie drugiego poważnego problemu dotyczącego produkcji RNA

przy użyciu polimerazy RNA T7, czyli wymagania przez to białko przynajmniej jednej reszty guanozyny na początku sekwencji do rozpoczęcia transkrypcji. Sposobem pozwalającym rozwiązać tę przeszkodę jest między innymi użycie rybozymów [73]. W przedstawionych badaniach użyłem rybozymów typu głowa młotka oraz typu spinki. Rybozomy te przeprowadzają reakcję transestryfikacji łańcucha RNA w wybranym miejscu. Użycie rybozymów nie jest jednak bezproblemowe. Po pierwsze, rybozomy użyte w pracy charakteryzują się wymaganiami sekwencyjnymi do rozpoznawania miejsca transestryfikacji łańcucha RNA [88,94,102]. Opisany problem użycia rybozymów ma szczególne znaczenie podczas generowania końców 3' w pożądanym RNA do krystalizacji. W tym przypadku sekwencja rozpoznawana przez rybozym musi znajdować się w sekwencji pożądanego RNA. Kolejnym poważnym ograniczeniem zastosowania rybozymów jest sama wydajność transestryfikacji rybozymów oraz oddzielenie rybozymów od pożądanego RNA. Należy pamiętać, aby rybozomy nie były tej samej długości co pożądanym RNA. W przeciwnym wypadku identyfikacja oraz odróżnienie pożądanego sekwencji od rybozymu może być niemożliwa. Trzecim poważnym ograniczeniem zastosowania rybozymów są grupy chemiczne, które generowane są podczas transestryfikacji łańcucha RNA. Przy użyciu rybozymów typu głowa młotka i typu spinki na 5' końcu RNA generowana jest grupa hydroksylowa, natomiast na końcu 3' RNA generowany jest 2'3' cykliczny fosforan. Jeżeli tak przygotowany RNA musi być następnie cyrkularyzowany, należy usunąć 2'3' cykliczny fosforan oraz dokonać fosforylacji końca 5'. Można to wykonać przy użyciu kinazy polinukleotydowej T4 z aktywnością fosfatazy [220]. Takie podejście skutkuje jednak zwiększeniem kosztów oraz obniżeniem wydajności produkcji RNA. Oprócz użytych w pracy rybozymów typu głowa młotka i typu spinki, do generowania homogenego końca 3' RNA można użyć także innych rybozymów. Popularnym rybozymem jest rybozym HDV pozyskany z wirusa zapalenia wątroby typu D (HDV, z ang. *hepatitis delta virus*) [221,222]. Sekwencja tuż przed miejscem transestryfikacji nie jest ograniczona, dlatego może być stosowany do generowania jednorodnych 3' końców dla różnych cząsteczek RNA. Ponieważ rybozym HDV nie ma ścisłych wymagań co do sekwencji, znalazł szerokie zastosowanie wśród badaczy strukturalnych do przygotowywania RNA do krystalizacji. Przykłady jego zastosowania obejmują przygotowanie domeny IV rybosomalnego RNA 4,5S z *E. coli* (PDB: 1DUH), lub ryboprzełącznika nadA (PDB: 7D82) [105,223]. Jednak w przypadku tego rybozymu opisano, że formowanie się złożonych struktur w docelowym RNA utrudnia właściwe

uformowanie się struktury rybozomu, co prowadzi do obniżonej wydajności transestryfikacji [224–226]. Jest to sytuacja podobna do tej, którą zaobserwowałem w przypadku zastosowania rybozomu typu spinki na końcu 3' zaprojektowanego przeze mnie RNA, dlatego rybozom HDV nie został wykorzystany w moich doświadczeniach.

W niniejszej pracy zaprojektowałem kilka konstruktów zawierających opisane powyżej rybozomy. Rybozom typu głowa młotka znajdujący się w konstrukcie na końcu 3' docelowego RNA ulegał wydajnej transestryfikacji. W przypadku konstruktów, w którym zastosowałem dwa rybozomy typu głowa młotka na obu końcach docelowego RNA nie obserwowałem tak wydajnej transestryfikacji. Prawdopodobnie rybozom typu głowa młotka znajdujący się na końcu 5' formując swoją strukturę hamował formowanie struktury drugiego rybozomu typu głowa młotka znajdującego się na końcu 3' docelowego RNA, jednocześnie inhibując jego zdolność do transestryfikacji łańcucha RNA. W ten sposób zmniejszył wydajność produkcji oczekiwanego RNA. Podobna sytuacja miała miejsce z konstruktem zawierającym rybozom typu głowa młotka na końcu 5' docelowego RNA i rybozom typu spinki na końcu 3'. Różnica polegała na tym, że transestryfikacja w przypadku tych rybozomów nie zachodziła wcale. Prawdopodobnie było to spowodowane brakiem poprawnego formowania się struktury całego konstruktów RNA [227].

Do produkcji RNA użytych w tej pracy wykorzystałem także transkrypcję *in vitro* prowadzoną z wykorzystaniem modyfikowanej matrycy DNA. Po wykonaniu reakcji konstrukty RNA wymagały oczyszczenia, do czego zastosowałem żele poliakrylamidowe. Jest to szybka metoda pozwalająca na otrzymanie homogennego RNA w jeden dzień. Procedura opiera się na wycięciu odpowiednich prążków z żelu, które odpowiadają oczekiwanym łańcuchom RNA. W procedurze tej należy być precyzyjnym, aby nie wyciąć przez przypadek prążka z niepożądaną sekwencją, jak również należy wykonywać operację wycięcia prążka szybko, ponieważ naświetlany światłem ultrafioletowym RNA może ulegać procesowi sieciowania (z ang. *cross-linking*) [228]. Wycięty prążek z pożądanym RNA poddałem elektrolecji. Jest to metoda zdecydowanie szybsza od elucji fizycznej i skraca całą procedurę pozyskiwania RNA. Następnie wyeluowany RNA był przeze mnie odsolony poprzez strącanie izopropanolem, które ma zdecydowaną przewagę nad procedurą strącania RNA przy użyciu etanolu przez objętości użyte w procedurach. Podczas strącania RNA potrzebna jest taka sama objętość izopropanolu co objętość mieszaniny, w której rozpuszczony jest RNA. W przypadku

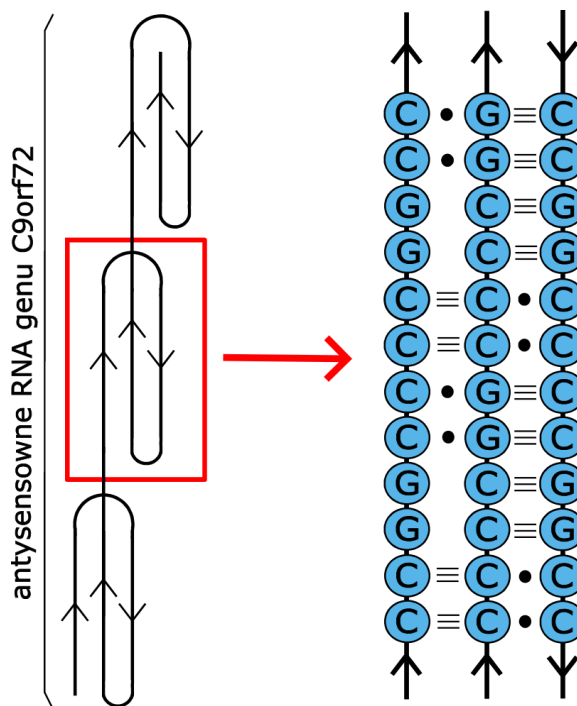
użycia etanolu niezbędna jest aż trzy razy większa objętość. Tak przygotowany RNA był przez mnie poddawany reakcji ligacji przy pomocy ligazy RNA T4 do cyrkularyzacji. Konstrukty, które w ten sposób otrzymałem posiadały pętlę wiązania białka U1a-RBD na przeciwnym biegunie do pętli badanej spinki. Oddziaływanie domeny wiążącej RNA białka U1a miało na celu zwiększenie potencjału krystalizacyjnego konstruktów. Podejście to było z sukcesem stosowane w krystalizacji spinki II U1snRNA [71]. W przypadku otrzymanych przez mnie konstruktów (CUG21-U1a, CUG12-U1a, CUG11-U1a) podejście to niestety nie pozwoliło na uzyskanie kryształów. Pomimo podjęcia prób optymalizacji procesu krystalizacji konstruktów, użycia robota do nastawiania krystalizacji, jak również nastawienia krystalizacji ręcznie, nie otrzymałem kryształów zaprojektowanych konstruktów RNA. Trudno jest wyjaśnić powód braku krystalizacji zaprojektowanych konstruktów przez bogactwo czynników mających znaczenie w tym procesie. Czynniki takie jak: temperatura, wilgotność, pH, stężenie, struktura i sekwencja krystalizowanej cząsteczki, stężenie soli czy obecność małych cząsteczek wspomagających proces krystalizacji mogą mieć wpływ na pojawienie się kryształów RNA [86,229].

Po nieudanych próbach krystalizacji podjąłem próbę poznania struktur RNA zaprojektowanych przez mnie konstruktów oddziałujących z domeną wiążącą RNA białka U1a przy użyciu kriomikroskopii elektronowej. Metoda ma tę przewagę nad standardową krystalizacją, że nie wymaga do determinacji struktury biomolekuł uzyskania kryształów [230]. Niestety i to podejście okazało się nieskuteczne w przypadku moich konstruktów. Zbyt małe rozmiary zaprojektowanych RNA nie pozwoliły na ich obserwację pod kriomikroskopem elektronowym, co jest zgodne z danymi literaturowymi. Najmniejszą cząsteczką RNA, dla której do tej pory udało się poznać strukturę przestrzenną był liczący 88 nukleotydów element stymulujący przesunięcie ramki odczytu (FSE z ang. *frameshift stimulation element*), niezbędny do zrównoważonej ekspresji białek wirusowych (28 kDa, rozdzielczość 6.9 Å) [231]. Ta rozdzielczość nie jest zadowalająca. Z wyższą rozdzielczością (3.7 Å) zdeterminowano strukturę 40 kDa SAM-IV ryboprzełącznika RNA [232].

W ramach drugiego zadania podjąłem się określenia struktury sekwencji G₂C₄ związanej ze stwardnieniem zanikowym bocznym oraz otępieniem czołowo-skroniowym. Chociaż przewiduje się, że sensowny RNA (GGGGCC)_n może formować strukturę G-kwadrupeksu i spinki, to struktura antysensownego RNA (GGCCCC)_n

nie była intensywnie badana [233,234]. Sekwencje bogate w reszty cytozyny często przyjmują struktury i-motywu (tetrapleksu) [185]. Sposób tworzenia się struktury sekwencji G₂C₄ antysensownego RNA genu *C9orf72* i potencjalny wpływ tej struktury na patologię chorób neurodegeneracyjnych oraz możliwość jej interakcji z małymi cząsteczkami pozostają niejasne.

Powtórzenia G₂C₄ wykazują wysoki polimorfizm strukturalny. Podczas badań strukturalnych określiłem, że powtórzenia te mają potencjał do formowania trypletów nukleotydowych. W zaprezentowanych w tej pracy doktorskiej strukturach istnieją dwa motywy zawierające trójki nukleotydowe. Pierwsza trójka jest zawarta w strukturze tetrameru, gdzie para Watsona-Cricka G-C oddziałuje z resztą cytozyny znajdującą się w sekwencji z nici sąsiedniego dupleksu (C⁺•G-C). Tryplet C⁺•G-C jest wrażliwy na zmiany pH, ponieważ pK_a na pozycji N3 reszty cytozyny wynosi około 4,6 [235]. W przypadku pary zasad Hoogsteena C⁺•G, protonacja na pozycji N3 reszty cytozyny umożliwia tworzenie wiązania wodorowego z atomem N7 reszty guaniny, co stabilizuje oddziaływanie tych dwóch reszt. Dlatego tryplety zawierające C⁺•G-C są bardziej stabilne w niższym pH w porównaniu do pH obojętnego. Ważne jest również, że pK_a reszty cytozyny Hoogsteena zależy od jej położenia w tripleksie oraz obecności sąsiednich trójek nukleotydów. Druga trójka nukleotydów formowana jest przez trzy reszty cytozyny tworzące kontakty w sieci krystalicznej. Badania zaprezentowane w tej rozprawie doktorskiej ukazują potencjał reszty cytozyny do wspomaganie tworzenia struktur wyższego rzędu. Na bazie moich badań można więc wnioskować, że wydłużone powtórzenia G₂C₄ obecne w antysensownej nici RNA genu *C9orf72* mogą tworzyć skomplikowane struktury. Ponadto, obserwowana przeze mnie para C⁺•C występuje w strukturach typu i-motywu. Bazując na uzyskanych strukturach krystalicznych zaproponowałem model, w którym długie ciągi powtórzeń G₂C₄ mogą formować potrójną helisę (Rycina 35). Helisa ta składa się z ciągu trzech różnych trypletów: C-C⁺•C, C⁺•G-C i G•C-G.



Rycina 35. Model strukturalny antysensownego RNA genu *C9orf72*.

W konstrukcie RNA G_2C_4 , ligand ANP77 oddziaływał z wystającymi resztami cytozyny. Pomiar DSC wykazał, że konstrukt G_2C_4 w roztworze występował jako tetramer z regionem trójniciowym. Najlepsze warunki do tworzenia tetramery występowały przy pH 6,0, gdzie w skanie DSC obserwowaliśmy dwa piki. Pierwszy pik prawdopodobnie odpowiadał rozpadowi struktury tetramery, podczas gdy drugi pik reprezentował rozpad regionu dwuniciowego. Przy pH 7,0 wymagana do tworzenia trypletów protonacja cytozyn zachodzi mniej wydajnie, co wyjaśnia obecność jednego pik w skanie DSC odpowiadającemu dupleksowi. Widma DSC przy pH 5,26 również sugerowały obecność jednej stabilnej struktury (pojedynczy pik), przypuszczalnie reprezentowanej przez strukturę typu i-motywu, która łatwo tworzyła się w warunkach kwasowych [185,236].

Zależne od pH stabilizacje trypletów C-C-C zostały zademonstrowane w badaniach termodynamicznych i badaniach magnetycznego rezonansu jądrowego (NMR, z ang. *nuclear magnetic resonance*). W warunkach kwasowych, wyższe tempo protonacji reszt cytozyny prowadziło do wzmocnienia interakcji Hoogsteena w trypletach [237,238]. Wykazano, że konieczność protonacji reszt cytozyn do utworzenia kontaktów trzeciorzędowych nie wskazywała na niską stabilność tripleksów w środowisku komórkowym [239]. Chociaż obliczone pKa reszty cytozyny było poniżej pH 5,0,

kontekst strukturalny i zawartość rozpuszczalnika mogą prowadzić do protonacji reszty cytozyny nawet w warunkach zasadowych [240,241]. Przesunięcie pKa reszt cytozyny zaobserwowano w kilku strukturach krystalicznych pseudowęzłów otrzymanych z RNA wirusa żółtej krzyżówki buraków, wirusa mozaiki bobu i rybozemu HDV [242–247]. Te badania wykazały, że tworzenie motywu C⁺GCA było kluczowe dla odpowiedniego złożenia i aktywności rybozemu. Formowanie potrójnej helisy również wymagało protonacji cytozyn. Wyniki krystalograficzne i biofizyczne wskazują, że można zaobserwować protonowane reszty cytozyny w warunkach zbliżonych do fizjologicznych, co sugeruje, że mogą one być również stabilne w macierzy wewnątrzkomórkowej. Podobne wnioski przedstawiono w badaniach NMR zawierających powtórzenia G₂C₄ w sekwencji DNA [248].

Do konstruktu G₂C₄ dodałem ligand ANP77. W strukturze zaobserwowałem jedną cząsteczkę liganda. Pierścienie naftyrydyny ANP77 zachowują się jak zasady azotowe nukleotydów i tworzą trzy wiązania wodorowe z grupami funkcyjnymi dwóch wystających reszt cytozyn z łańcucha A tetrameru G₂C₄. Dokładniej, obserwowałem jedno wiązanie wodorowe między grupą 2-aminową ANP77 a atomem karbonylu O2 reszty cytozyny, natomiast drugie między grupą iminową N8 liganda a grupą 2-egzoaminową reszty cytozyny. Trzecia interakcja występuje między atomem N1 naftyrydyny a atomem N3 reszty cytozyny. Pomimo, że mapa gęstości elektronowej została dobrze zdefiniowana i obliczona przy rozdzielczości atomowej 1.1Å, lokalizacja protonów w obrębie dwóch pseudo-par jest niejednoznaczna. Najprawdopodobniej jeden proton znajduje się w jednej z jednostek naftyrydyny cząsteczki ANP77, ponieważ przy neutralnym pH równym 7, ligand obecny jest w formie monokationu. Drugi proton można przypisać albo reszcie cytozyny, albo drugiej jednostce naftyrydyny cząsteczki ANP77. Druga para wystających reszt cytozyny wchodzących w skład łańcucha D jest uporządkowana i oddziałuje z sąsiednią cząsteczką RNA oraz grupą 2-egzoaminową sąsiedniego liganda z sieci krystalicznej.

Trudno jest wyjaśnić, dlaczego (pomimo obecności dwóch miejsc wiązania) tylko jedna cząsteczka ANP77 wiąże się z sekwencją konstruktu RNA G₂C₄. Wystające reszty cytozyny na obu końcach dupleksu wydają się równe pod względem zdolności do tworzenia wiązań wodorowych i środowiska strukturalnego. Ponadto, z punktu widzenia termodynamicznego, oddziaływanie z dwoma cząsteczkami ANP77 powinno być korzystne dla układu, dzięki dodatkowym wiązaniom wodorowym i rozszerzonym

interakcjom między pierścieniami naftyrydynowymi a resztami guanozyny znajdującymi się na końcu 5' nici C. Jednym z powodów takiej sytuacji może być różny potencjał krystalizacyjny kompleksów RNA-ligand w zależności od liczby związanych cząsteczek liganda. W sieci krystalicznej niezwiązane reszty cytozyny łańcucha D uczestniczą w kontaktach krystalicznych między związanymi symetrycznie cząsteczkami. Ten efekt stabilizacji może ułatwić wzrost kryształów.

W przedstawionym modelu, ligand ANP77 przyjmuje konformację typu *stacked* z jednostkami naftyrydyny przesuniętymi względem siebie. Skutkuje to ograniczonymi oddziaływaniami warstwowymi między aromatycznymi fragmentami liganda. Cząsteczka ANP77 była projektowana z myślą o oddziaływaniu z dwiema sąsiednimi resztami cytozyn. Z tego powodu, podczas projektowania liganda trzeba było wziąć pod uwagę wiele cech strukturalnych, aby zapewnić specyficzne oddziaływanie z sekwencją docelową. Po pierwsze, podczas projektowania cząsteczki ANP77 autorzy wybrali grupy funkcyjne jednostek naftyrydyny jako matrycę dla bezpośrednich oddziaływań z krawędziami Watsona-Cricka reszt cytozyny. Po drugie, połączyli pierścienie aromatyczne krótkim łącznikiem składającym się z trzech atomów węgla, aby ograniczyć swobodę konformacyjną cząsteczki. Po trzecie, do cząsteczki ANP77 dołączyli boczny łańcuch aminopropanokarboksyamidowy, który łatwo podlegał protonacji, zwiększając rozpuszczalność w wodzie i oddziaływania z negatywnie naładowanymi kwasami nukleinowymi [205,206]. Inspekcja modelu krystalograficznego kompleksu G₂C₄-ANP77 potwierdza, że ligand zdolny jest do bezpośredniego oddziaływania z dwiema wystającymi resztami cytozyny, tworząc pseudo-pary. Ligand ANP77 wpasowuje się w helisę RNA rozszerzając długość regionu podwójnej nici. Wskazuje to na dopasowanie liganda ANP77 do struktury nie tylko pod względem wiązań wodorowych, ale także kształtu helisy RNA. Geometria łącznika pozwala na ułożenie jego pierścieni aromatycznych w sposób podobny do układu zasad azotowych w helisie kwasu nukleinowego. Jednostki naftyrydyny cząsteczki ANP77 ustawione są jeden nad drugim, oddalone od siebie o około 3Å, odzwierciedlając wartość parametru wzrostu w formie A-RNA. Ponadto, są one skręcone względem siebie o około 25°, a stopień skrętu może ulegać regulacji przez obrót wokół wiązania C11-C12 liganda ANP77. Długość łącznika w cząsteczce ANP77 wydaje się być odpowiednio dobrana, ponieważ pozwala na wiązanie tylko dwóch kolejnych reszt cytozyny. Dodanie większej liczby atomów węgla zwiększyłoby elastyczność i swobodę konformacyjną liganda, co mogłoby

negatywnie wpłynąć na selektywność ANP77 (np. rozpoznawanie reszt cytozyn oddzielonych dodatkowymi resztami nukleotydowymi).

6. Wnioski

Przedstawione w niniejszej rozprawie doktorskiej wyniki pozwalają na sformułowanie następujących wniosków:

1. RNA do badań krystalograficznych o długości do 10 nukleotydów może z powodzeniem być produkowany przy użyciu syntezy chemicznej na podłożu stałym, a następnie oczyszczany przy użyciu żeli poliakrylamidowych, TLC lub HPLC.
2. Oligomery RNA o długości powyżej 40 nukleotydów powinny być produkowane przy użyciu transkrypcji *in vitro*. Problemy wynikające z heterogenności 3' końca RNA mogą zostać rozwiązane przy użyciu rybozymów lub modyfikowanej matrycy DNA w połączeniu z oczyszczaniem w żelu poliakrylamidowym lub HPLC.
3. Współczesna technologia ustalania struktury przestrzennej RNA za pomocą kriomikroskopii elektronowej nie pozwala na rozwiązywanie struktur RNA mniejszych niż 20 kDa.
4. Konstrukt zawierający sekwencję G₂C₄ ma tendencję do formowania tetramerów z rejonami trójniciowymi. W tej rozprawie doktorskiej zaobserwowałem formowanie się tripleksów C-G-C oraz C-C-C.
5. Cząsteczka ANP77 oddziałuje z wystającymi resztami cytozyn struktury G₂C₄ w formie typu *stacked* poprzez trzy wiązania wodorowe.
6. ANP77 stabilizuje formowanie się struktury G₂C₄ w pH 5,26 i 6,0. Dodatkowo formowanie tetramerów jest faworyzowane w pH 6,0.

7. Literatura

1. Dugger, B.N.; Dickson, D.W. Pathology of Neurodegenerative Diseases. *Cold Spring Harb Perspect Biol* **2017**, *9*, a028035, doi:10.1101/CSHPERSPECT.A028035.
2. Soto, C.; Pritzkow, S. Protein Misfolding, Aggregation, and Conformational Strains in Neurodegenerative Diseases. *Nat Neurosci* **2018**, *21*, 1332–1340, doi:10.1038/S41593-018-0235-9.
3. Wells, C.; Brennan, S.E.; Keon, M.; Saksena, N.K. Prionoid Proteins in the Pathogenesis of Neurodegenerative Diseases. *Front Mol Neurosci* **2019**, *12*, doi:10.3389/FNMOL.2019.00271.
4. Bruzova, M.; Rusina, R.; Stejskalova, Z.; Matej, R. Autopsy-Diagnosed Neurodegenerative Dementia Cases Support the Use of Cerebrospinal Fluid Protein Biomarkers in the Diagnostic Work-Up. *Scientific Reports* **2021**, *11*, 1–12, doi:10.1038/s41598-021-90366-5.
5. Cenini, G.; Lloret, A.; Cascella, R. Oxidative Stress and Mitochondrial Damage in Neurodegenerative Diseases: From Molecular Mechanisms to Targeted Therapies. *Oxid Med Cell Longev* **2020**, *2020*, doi:10.1155/2020/1270256.
6. Guzman-Martinez, L.; Maccioni, R.B.; Andrade, V.; Navarrete, L.P.; Pastor, M.G.; Ramos-Escobar, N. Neuroinflammation as a Common Feature of Neurodegenerative Disorders. *Front Pharmacol* **2019**, *10*, doi:10.3389/FPHAR.2019.01008.
7. Erekat, N.S. Apoptosis and Its Therapeutic Implications in Neurodegenerative Diseases. *Clin Anat* **2022**, *35*, 65–78, doi:10.1002/CA.23792.
8. Salman, M.M.; Al-Obaidi, Z.; Kitchen, P.; Loreto, A.; Bill, R.M.; Wade-Martins, R. Advances in Applying Computer-Aided Drug Design for Neurodegenerative Diseases. *Int J Mol Sci* **2021**, *22*, doi:10.3390/IJMS22094688.
9. Aldewachi, H.; Al-Zidan, R.N.; Conner, M.T.; Salman, M.M. High-Throughput Screening Platforms in the Discovery of Novel Drugs for Neurodegenerative Diseases. *Bioengineering (Basel)* **2021**, *8*, 1–20, doi:10.3390/BIOENGINEERING8020030.

10. Livingston, G.; Huntley, J.; Sommerlad, A.; Ames, D.; Ballard, C.; Banerjee, S.; Brayne, C.; Burns, A.; Cohen-Mansfield, J.; Cooper, C.; et al. Dementia Prevention, Intervention, and Care: 2020 Report of the Lancet Commission. *Lancet* **2020**, *396*, 413–446, doi:10.1016/S0140-6736(20)30367-6.
11. Pereira, T.M.C.; Côco, L.Z.; Ton, A.M.M.; Meyrelles, S.S.; Campos-Toimil, M.; Campagnaro, B.P.; Vasquez, E.C. The Emerging Scenario of the Gut-Brain Axis: The Therapeutic Actions of the New Actor Kefir against Neurodegenerative Diseases. *Antioxidants (Basel)* **2021**, *10*, doi:10.3390/ANTIOX10111845.
12. Malik, I.; Kelley, C.P.; Wang, E.T.; Todd, P.K. Molecular Mechanisms Underlying Nucleotide Repeat Expansion Disorders. *Nat Rev Mol Cell Biol* **2021**, *22*, 589–607, doi:10.1038/S41580-021-00382-6.
13. Figueiredo, A.S.; Loureiro, J.R.; Macedo-Ribeiro, S.; Silveira, I. Advances in Nucleotide Repeat Expansion Diseases: Transcription Gets in Phase. *Cells* **2023**, *12*, doi:10.3390/CELLS12060826.
14. Rohilla, K.J.; Gagnon, K.T. RNA Biology of Disease-Associated Microsatellite Repeat Expansions. *Acta Neuropathol Commun* **2017**, *5*, 63, doi:10.1186/S40478-017-0468-Y.
15. Xu, Z.; Poidevin, M.; Li, X.; Li, Y.; Shu, L.; Nelson, D.L.; Li, H.; Hales, C.M.; Gearing, M.; Wingo, T.S.; et al. Expanded GGGGCC Repeat RNA Associated with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Causes Neurodegeneration. *Proc Natl Acad Sci U S A* **2013**, *110*, 7778–7783, doi:10.1073/PNAS.1219643110.
16. Mirkin, S.M. Expandable DNA Repeats and Human Disease. *Nature* **2007**, *447*, 932–940, doi:10.1038/NATURE05977.
17. Ellerby, L.M. Repeat Expansion Disorders: Mechanisms and Therapeutics. *Neurotherapeutics* **2019**, *16*, 924–927, doi:10.1007/S13311-019-00823-3/FIGURES/1.
18. Rohilla, K.J.; Gagnon, K.T. RNA Biology of Disease-Associated Microsatellite Repeat Expansions. *Acta Neuropathologica Communications 2017 5:1* **2017**, *5*, 1–22, doi:10.1186/S40478-017-0468-Y.

19. Schwartz, J.L.; Jones, K.L.; Yeo, G.W. Repeat RNA Expansion Disorders of the Nervous System: Post-Transcriptional Mechanisms and Therapeutic Strategies. *Crit Rev Biochem Mol Biol* **2021**, *56*, 31–53, doi:10.1080/10409238.2020.1841726.
20. Bernat, V.; Disney, M.D. RNA Structures as Mediators of Neurological Diseases and as Drug Targets. *Neuron* **2015**, *87*, 28–46, doi:10.1016/J.NEURON.2015.06.012.
21. Zhao, Q.; Zhao, Z.; Fan, X.; Yuan, Z.; Mao, Q.; Yao, Y. Review of Machine Learning Methods for RNA Secondary Structure Prediction. *PLoS Comput Biol* **2021**, *17*, doi:10.1371/JOURNAL.PCBI.1009291.
22. Korostelev, A.; Trakhanov, S.; Laurberg, M.; Noller, H.F. Crystal Structure of a 70S Ribosome-TRNA Complex Reveals Functional Interactions and Rearrangements. *Cell* **2006**, *126*, 1065–1077, doi:10.1016/j.cell.2006.08.032.
23. Li, Q.; Lee, J.A.; Black, D.L. Neuronal Regulation of Alternative Pre-mRNA Splicing. *Nat Rev Neurosci* **2007**, *8*, 819–831, doi:10.1038/NRN2237.
24. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **2008**, *456*, 470–476, doi:10.1038/NATURE07509.
25. Buratti, E.; Muro, A.F.; Giombi, M.; Gherbassi, D.; Iaconcig, A.; Baralle, F.E. RNA Folding Affects the Recruitment of SR Proteins by Mouse and Human Polypurinic Enhancer Elements in the Fibronectin EDA Exon. *Mol Cell Biol* **2004**, *24*, 1387–1400, doi:10.1128/MCB.24.3.1387-1400.2004.
26. Warf, M.B.; Berglund, J.A. Role of RNA Structure in Regulating Pre-mRNA Splicing. *Trends Biochem Sci* **2010**, *35*, 169–178, doi:10.1016/J.TIBS.2009.10.004.
27. De Conti, L.; Baralle, M.; Buratti, E. Exon and Intron Definition in Pre-mRNA Splicing. *Wiley Interdiscip Rev RNA* **2013**, *4*, 49–60, doi:10.1002/WRNA.1140.

28. Buratti, E.; Baralle, F.E. Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol Cell Biol* **2004**, *24*, 10505–10514, doi:10.1128/MCB.24.24.10505-10514.2004.
29. Black, D.L.; Grabowski, P.J. Alternative Pre-mRNA Splicing and Neuronal Function. *Prog Mol Subcell Biol* **2003**, *31*, 187–216, doi:10.1007/978-3-662-09728-1_7.
30. Tucker, B.J.; Breaker, R.R. Riboswitches as Versatile Gene Control Elements. *Curr Opin Struct Biol* **2005**, *15*, 342–348, doi:10.1016/J.SBI.2005.05.003.
31. La Spada, A.R.; Taylor, J.P. Repeat Expansion Disease: Progress and Puzzles in Disease Pathogenesis. *Nat Rev Genet* **2010**, *11*, 247–258, doi:10.1038/NRG2748.
32. Jiang, Z.; Cote, J.; Kwon, J.M.; Goate, A.M.; Wu, J.Y. Aberrant Splicing of Tau Pre-mRNA Caused by Intronic Mutations Associated with the Inherited Dementia Frontotemporal Dementia with Parkinsonism Linked to Chromosome 17. <https://doi.org/10.1128/MCB.20.11.4036-4048.2000> **2023**, *20*, 4036–4048, doi:10.1128/MCB.20.11.4036-4048.2000.
33. Mathews, D.H.; Disney, M.D.; Childs, J.L.; Schroeder, S.J.; Zuker, M.; Turner, D.H. Incorporating Chemical Modification Constraints into a Dynamic Programming Algorithm for Prediction of RNA Secondary Structure. *Proc Natl Acad Sci U S A* **2004**, *101*, 7287–7292, doi:10.1073/PNAS.0401799101.
34. Klein, D.J.; Schmeing, T.M.; Moore, P.B.; Steitz, T.A. The Kink-Turn: A New RNA Secondary Structure Motif. *EMBO J* **2001**, *20*, 4214–4221, doi:10.1093/EMBOJ/20.15.4214.
35. Kim, J.; Cheong, C.; Moore, P.B. Tetramerization of an RNA Oligonucleotide Containing a GGGG Sequence. *Nature* **1991**, *351*, 331–332, doi:10.1038/351331A0.
36. Brion, P.; Westhof, E. Hierarchy and Dynamics of RNA Folding. *Annu Rev Biophys Biomol Struct* **1997**, *26*, 113–137, doi:10.1146/ANNUREV.BIOPHYS.26.1.113.

37. Schroeder, R.; Barta, A.; Semrad, K. Strategies for RNA Folding and Assembly. *Nat Rev Mol Cell Biol* **2004**, *5*, 908–919, doi:10.1038/NRM1497.
38. Bugaut, A.; Murat, P.; Balasubramanian, S. An RNA Hairpin to G-Quadruplex Conformational Transition. *J Am Chem Soc* **2012**, *134*, 19953–19956, doi:10.1021/JA308665G.
39. Serganov, A.; Yuan, Y.R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A.T.; Hobartner, C.; Micura, R.; Breaker, R.R.; Patel, D.J. Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing MRNAs. *Chem Biol* **2004**, *11*, 1729–1741, doi:10.1016/J.CHEMBIOL.2004.11.018.
40. Zu, T.; Gibbens, B.; Doty, N.S.; Gomes-Pereira, M.; Huguet, A.; Stone, M.D.; Margolis, J.; Peterson, M.; Markowski, T.W.; Ingram, M.A.C.; et al. Non-ATG-Initiated Translation Directed by Microsatellite Expansions. *Proc Natl Acad Sci U S A* **2011**, *108*, 260–265, doi:10.1073/PNAS.1013343108.
41. Castelli, L.M.; Huang, W.P.; Lin, Y.H.; Chang, K.Y.; Hautbergue, G.M. Mechanisms of Repeat-Associated Non-AUG Translation in Neurological Microsatellite Expansion Disorders. *Biochem Soc Trans* **2021**, *49*, 775–792, doi:10.1042/BST20200690.
42. Bevilacqua, P.C.; Blose, J.M. Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins. *Annu Rev Phys Chem* **2008**, *59*, 79–103, doi:10.1146/ANNUREV.PHYSCHEM.59.032607.093743.
43. Warner, K.D.; Chen, M.C.; Song, W.; Strack, R.L.; Thorn, A.; Jaffrey, S.R.; Ferré-D'Amaré, A.R. Structural Basis for Activity of Highly Efficient RNA Mimics of Green Fluorescent Protein. *Nat Struct Mol Biol* **2014**, *21*, 658–663, doi:10.1038/NSMB.2865.
44. Kiliszek, A.; Błaszczyk, L.; Kierzek, R.; Rypniewski, W. Stabilization of RNA Hairpins Using Non-Nucleotide Linkers and Circularization. *Nucleic Acids Res* **2017**, *45*, 1–9, doi:10.1093/nar/gkx122.
45. Ouellet, J. RNA Fluorescence with Light-Up Aptamers. *Front Chem* **2016**, *4*, doi:10.3389/FCHEM.2016.00029.

46. Pfeiffer, F.; Mayer, G. Selection and Biosensor Application of Aptamers for Small Molecules. *Front Chem* **2016**, *4*, 188250, doi:10.3389/FCHEM.2016.00025/BIBTEX.
47. Wei, Z.; Zhou, Y.; Wang, R.; Wang, J.; Chen, Z. Aptamers as Smart Ligands for Targeted Drug Delivery in Cancer Therapy. *Pharmaceutics* **2022**, *14*, doi:10.3390/PHARMACEUTICS14122561.
48. Mollasalehi, N.; Francois-Moutal, L.; Porciani, D.; Burke, D.H.; Khanna, M. Aptamers Targeting Hallmark Proteins of Neurodegeneration. *Nucleic Acid Ther* **2022**, *32*, 235–250, doi:10.1089/NAT.2021.0091.
49. Guo, P. The Emerging Field of RNA Nanotechnology. *Nat Nanotechnol* **2010**, *5*, 833–842, doi:10.1038/nnano.2010.231.
50. Serganov, A.; Nudler, E. A Decade of Riboswitches. *Cell* **2013**, *152*, 17–24, doi:10.1016/J.CELL.2012.12.024.
51. Svoboda, P.; Di Cara, A. Hairpin RNA: A Secondary Structure of Primary Importance. *Cell Mol Life Sci* **2006**, *63*, 901–918, doi:10.1007/S00018-005-5558-5.
52. Georgakopoulos-Soares, I.; Parada, G.E.; Hemberg, M. Secondary Structures in RNA Synthesis, Splicing and Translation. *Comput Struct Biotechnol J* **2022**, *20*, 2871–2884, doi:10.1016/J.CSBJ.2022.05.041.
53. Xu, J.; Hou, J.; Ding, M.; Wang, Z.; Chen, T. Riboswitches, from Cognition to Transformation. *Synth Syst Biotechnol* **2023**, *8*, 357–370, doi:10.1016/J.SYNBIO.2023.05.008.
54. Jasinska, A.; Michlewski, G.; de Mezer, M.; Sobczak, K.; Kozlowski, P.; Napierala, M.; Krzyzosiak, W.J. Structures of Trinucleotide Repeats in Human Transcripts and Their Functional Implications. *Nucleic Acids Res* **2003**, *31*, 5463–5468, doi:10.1093/NAR/GKG767.
55. Sobczak, K.; de Mezer, M.; Michlewski, G.; Krol, J.; Krzyzosiak, W.J. RNA Structure of Trinucleotide Repeats Associated with Human Neurological Diseases. *Nucleic Acids Res* **2003**, *31*, 5469–5482, doi:10.1093/NAR/GKG766.

56. Napierała, M.; Krzyzosiak, W.J. CUG Repeats Present in Myotonin Kinase RNA Form Metastable “Slippery” Hairpins. *J Biol Chem* **1997**, *272*, 31079–31085, doi:10.1074/JBC.272.49.31079.
57. Rypniewski, W.; Adamiak, D.A.; Milecki, J.; Adamiak, R.W. Noncanonical G(Syn)-G(Anti) Base Pairs Stabilized by Sulphate Anions in Two X-Ray Structures of the (GUGGUCUGAUGAGGCC) RNA Duplex. *RNA* **2008**, *14*, 1845–1851, doi:10.1261/RNA.1164308.
58. Holbrook, S.R.; Cheong, C.; Tinoco, I.; Kim, S.H. Crystal Structure of an RNA Double Helix Incorporating a Track of Non-Watson-Crick Base Pairs. *Nature* **1991**, *353*, 579–581, doi:10.1038/353579A0.
59. Kumar, A.; Fang, P.; Park, H.; Guo, M.; Nettles, K.W.; Disney, M.D. A Crystal Structure of a Model of the Repeating r(CG) Transcript Found in Fragile X Syndrome. *Chembiochem* **2011**, *12*, 2140–2142, doi:10.1002/CBIC.201100337.
60. Kiliszek, A.; Kierzek, R.; Krzyzosiak, W.J.; Rypniewski, W. Crystal Structures of CGG RNA Repeats with Implications for Fragile X-Associated Tremor Ataxia Syndrome. *Nucleic Acids Res* **2011**, *39*, 7308–7315, doi:10.1093/NAR/GKR368.
61. Yildirim, I.; Park, H.; Disney, M.D.; Schatz, G.C. A Dynamic Structural Model of Expanded RNA CAG Repeats: A Refined X-Ray Structure and Computational Investigations Using Molecular Dynamics and Umbrella Sampling Simulations. *J Am Chem Soc* **2013**, *135*, 3528–3538, doi:10.1021/JA3108627.
62. Kiliszek, A.; Kierzek, R.; Krzyzosiak, W.J.; Rypniewski, W. Atomic Resolution Structure of CAG RNA Repeats: Structural Insights and Implications for the Trinucleotide Repeat Expansion Diseases. *Nucleic Acids Res* **2010**, *38*, 8370–8376, doi:10.1093/NAR/GKQ700.
63. Tawani, A.; Kumar, A. Structural Insights Reveal the Dynamics of the Repeating r(CAG) Transcript Found in Huntington’s Disease (HD) and Spinocerebellar Ataxias (SCAs). *PLoS One* **2015**, *10*, doi:10.1371/JOURNAL.PONE.0131788.
64. Błaszczyk, L.; Rypniewski, W.; Kiliszek, A. Structures of RNA Repeats Associated with Neurological Diseases. *Wiley Interdiscip Rev RNA* **2017**, *8*, 1–25, doi:10.1002/wrna.1412.

65. Kiliszek, A.; Kierzek, R.; Krzyzosiak, W.J.; Rypniewski, W. Crystallographic Characterization of CCG Repeats. *Nucleic Acids Res* **2012**, *40*, 8155–8162, doi:10.1093/NAR/GKS557.
66. Coonrod, L.A.; Lohman, J.R.; Berglund, J.A. Utilizing the GAAA Tetraloop/Receptor to Facilitate Crystal Packing and Determination of the Structure of a CUG RNA Helix. *Biochemistry* **2012**, *51*, 8330–8337, doi:10.1021/BI300829W.
67. Kiliszek, A.; Kierzek, R.; Krzyzosiak, W.J.; Rypniewski, W. Structural Insights into CUG Repeats Containing the “Stretched U-U Wobble”: Implications for Myotonic Dystrophy. *Nucleic Acids Res* **2009**, *37*, 4149–4156, doi:10.1093/NAR/GKP350.
68. Mooers, B.H.M.; Logue, J.S.; Berglund, J.A. The Structural Basis of Myotonic Dystrophy from the Crystal Structure of CUG Repeats. *Proc Natl Acad Sci U S A* **2005**, *102*, 16626–16631, doi:10.1073/PNAS.0505873102.
69. deLorimier, E.; Coonrod, L.A.; Copperman, J.; Taber, A.; Reister, E.E.; Sharma, K.; Todd, P.K.; Guenza, M.G.; Berglund, J.A. Modifications to Toxic CUG RNAs Induce Structural Stability, Rescue Mis-Splicing in a Myotonic Dystrophy Cell Model and Reduce Toxicity in a Myotonic Dystrophy Zebrafish Model. *Nucleic Acids Res* **2014**, *42*, 12768–12778, doi:10.1093/NAR/GKU941.
70. Reuter, J.S.; Mathews, D.H. RNAstructure: Software for RNA Secondary Structure Prediction and Analysis. *BMC Bioinformatics* **2010**, *11*, doi:10.1186/1471-2105-11-129.
71. Law, M.J.; Rice, A.J.; Lin, P.; Laird-Offringa, I.A. The Role of RNA Structure in the Interaction of U1A Protein with U1 Hairpin II RNA. *RNA* **2006**, *12*, 1168–1178, doi:10.1261/RNA.75206.
72. Oubridge, C.; Ito, N.; Evans, P.R.; Teo, C.H.; Nagai, K. Crystal Structure at 1.92 Å Resolution of the RNA-Binding Domain of the U1A Spliceosomal Protein Complexed with an RNA Hairpin. *Nature* **1994**, *372*, 432–438.
73. Price, S.R.; Ito, N.; Oubridge, C.; Avis, J.M.; Nagai, K. Crystallization of RNA-Protein Complexes I. Methods for the Large-Scale Preparation of RNA Suitable

- for Crystallographic Studies. *J Mol Biol* **1995**, *249*, 398–408, doi:10.1006/jmbi.1995.0305.
74. Ferré-D'Amaré, A.R. Use of the Spliceosomal Protein U1A to Facilitate Crystallization and Structure Determination of Complex RNAs. *Methods* **2010**, *52*, 159–167, doi:10.1016/J.YMETH.2010.06.008.
75. Ferré-D'Amaré, A.R.; Doudna, J.A. Methods to Crystallize RNA. *Curr Protoc Nucleic Acid Chem* **2000**, *1*, 7.6.1-7.6.13, doi:10.1002/0471142700.NC0706S00.
76. Ahmed, Y.L.; Ficner, R. RNA Synthesis and Purification for Structural Studies. *RNA Biol* **2014**, *11*, 427, doi:10.4161/RNA.28076.
77. Baronti, L.; Karlsson, H.; Marušič, M.; Petzold, K. A Guide to Large-Scale RNA Sample Preparation. *Anal Bioanal Chem* **2018**, *410*, 3239–3252, doi:10.1007/s00216-018-0943-8.
78. Ryczek, M.; Pluta, M.; Błaszczuk, L.; Kiliszek, A. Overview of Methods for Large-Scale RNA Synthesis. *Applied Sciences (Switzerland)* **2022**, *12*, 1543, doi:10.3390/APP12031543/S1.
79. Holbrook*, S.R.; Holbrook, E.L.; Walukiewicz, H.E. Crystallization of RNA. *Cellular and Molecular Life Sciences* **2001**, *58*, 234–243, doi:10.1007/PL00000851.
80. Petrov, A.; Wu, T.; Puglisi, E.V.; Puglisi, J.D. RNA Purification by Preparative Polyacrylamide Gel Electrophoresis. *Methods Enzymol* **2013**, *530*, 315–330, doi:10.1016/B978-0-12-420037-1.00017-8.
81. HPLC Purification of RNA for Crystallography and NMR. Available online: <https://rnajournal.cshlp.org/content/2/2/110.short> (accessed on 27 August 2023).
82. Beckert, B.; Masquida, B. Synthesis of RNA by in Vitro Transcription. *Methods Mol Biol* **2011**, *703*, 29–41, doi:10.1007/978-1-59745-248-9_3.
83. Coleman, T.M.; Wang, G.; Huang, F. Superior 5' Homogeneity of RNA from ATP-Initiated Transcription under the T7 Phi 2.5 Promoter. *Nucleic Acids Res* **2004**, *32*, 2–5, doi:10.1093/nar/gnh007.

84. Dunn, J.J.; Studier, F.W. Complete Nucleotide Sequence of Bacteriophage T7 DNA and the Locations of T7 Genetic Elements. *J Mol Biol* **1983**, *166*, 477–535, doi:10.1016/S0022-2836(83)80282-4.
85. Kao, C.; Zheng, M.; Rüdiger, S. A Simple and Efficient Method to Reduce Nontemplated Nucleotide Addition at the 3' Terminus of RNAs Transcribed by T7 RNA Polymerase. *RNA* **1999**, *5*, 1268–1272, doi:10.1017/S1355838299991033.
86. Golden, B.L.; Kundrot, C.E. RNA Crystallization. *J Struct Biol* **2003**, *142*, 98–107, doi:10.1016/S1047-8477(03)00042-X.
87. Pujari, N.; Saundh, S.L.; Acquah, F.A.; Mooers, B.H.M.; Ferré-D'amaré, A.R.; Kwun-Wai Leung, A. Engineering Crystal Packing in RNA Structures I: Past and Future Strategies for Engineering RNA Packing in Crystals. *Crystals (Basel)* **2021**, *11*, doi:10.3390/CRYST11080952.
88. Tanner, N.K. Ribozymes: The Characteristics and Properties of Catalytic RNAs. *FEMS Microbiol Rev* **1999**, *23*, 257–275, doi:10.1111/j.1574-6976.1999.tb00399.x.
89. Weinberg, C.E.; Weinberg, Z.; Hammann, C. Novel Ribozymes: Discovery, Catalytic Mechanisms, and the Quest to Understand Biological Function. *Nucleic Acids Res* **2019**, *47*, 9480–9494, doi:10.1093/NAR/GKZ737.
90. Doherty, E.A.; Doudna, J.A. Ribozyme Structures and Mechanisms. *Annu Rev Biochem* **2000**, *69*, 597–615, doi:10.1146/ANNUREV.BIOCHEM.69.1.597.
91. Müller, S.; Appel, B.; Balke, D.; Hieronymus, R.; Nübel, C. Thirty-Five Years of Research into Ribozymes and Nucleic Acid Catalysis: Where Do We Stand Today? [Version 1; Referees: 2 Approved]. *F1000Res* **2016**, *5*, doi:10.12688/F1000RESEARCH.8601.1/DOI.
92. Serganov, A.; Patel, D.J. Ribozymes, Riboswitches and beyond: Regulation of Gene Expression without Proteins. *Nat Rev Genet* **2007**, *8*, 776–790, doi:10.1038/nrg2172.
93. Maurel, M.C.; Leclerc, F.; Hervé, G. Ribozyme Chemistry: To Be or Not To Be under High Pressure. *Chem Rev* **2020**, *120*, 4898–4918, doi:10.1021/ACS.CHEMREV.9B00457.

94. Ferré-D'Amaré, A.R.; Doudna, J.A. Use of Cis- and Trans-Ribozymes to Remove 5' and 3' Heterogeneities from Milligrams of in Vitro Transcribed RNA. *Nucleic Acids Res* **1996**, *24*, 977–978, doi:10.1093/nar/24.5.977.
95. Kieft, J.S.; Batey, R.T. A General Method for Rapid and Nondenaturing Purification of RNAs. *RNA* **2004**, *10*, 988–995, doi:10.1261/rna.7040604.
96. Salvail-Lacoste, A.; Di Tomasso, G.; Piette, B.L.; Legault, P. Affinity Purification of T7 RNA Transcripts with Homogeneous Ends Using ARiBo and CRISPR Tags. *RNA* **2013**, *19*, 1003–1014, doi:10.1261/rna.037432.112.
97. Schürer, H.; Lang, K.; Schuster, J.; Mörl, M. A Universal Method to Produce in Vitro Transcripts with Homogeneous 3' Ends. *Nucleic Acids Res* **2002**, *30*, doi:10.1093/nar/gnf055.
98. Hartmann, R.K.; Bindereif, A.; Schön, A.; Westhof, E. Handbook of RNA Biochemistry. *Handbook of RNA Biochemistry* **2008**, 1–931, doi:10.1002/9783527619504.
99. De La Peña, M.; García-Robles, I.; Cervera, A. The Hammerhead Ribozyme: A Long History for a Short RNA. *Molecules* *2017*, Vol. 22, Page 78 **2017**, *22*, 78, doi:10.3390/MOLECULES22010078.
100. Pena, M. d. I.; Dufour, D.; Gallego, J. Three-Way RNA Junctions with Remote Tertiary Contacts: A Recurrent and Highly Versatile Fold. *RNA* **2009**, *15*, 1949–1964, doi:10.1261/rna.1889509.
101. Martick, M.; Scott, W.G. Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* **2006**, *126*, 309–320, doi:10.1016/j.cell.2006.06.036.Tertiary.
102. Amarguioui, M.; Prydz, H. Hammerhead Ribozyme Design and Application. *Cell Mol Life Sci* **1998**, *54*, 1175–1202, doi:10.1007/s000180050247.
103. Walter, F.; Westhof, E. *Catalytic RNA* ; Academic Press, 2001; ISBN 9780123812865.
104. Oubridge, C.; Kuglstatter, A.; Jovine, L.; Nagai, K. Crystal Structure of SRP19 in Complex with the S Domain of SRP RNA and Its Implication for the Assembly

- of the Signal Recognition Particle. *Mol Cell* **2002**, *9*, 1251–1261, doi:10.1016/S1097-2765(02)00530-0.
105. Jovine, L.; Hainzl, T.; Oubridge, C.; Scott, W.G.; Li, J.; Sixma, T.K.; Wonacott, A.; Skarzynski, T.; Nagai, K. Crystal Structure of the Ffh and EF-G Binding Sites in the Conserved Domain IV of Escherichia Coli 4.5S RNA. *Structure* **2000**, *8*, 527–540, doi:10.1016/S0969-2126(00)00137-4.
106. Battaglia, R.A.; Grigg, J.C.; Ke, A. Structural Basis for tRNA Decoding and Aminoacylation Sensing by T-Box Riboregulators. *Nat Struct Mol Biol* **2019**, *26*, 1106–1113, doi:10.1038/s41594-019-0327-6.
107. Walter, N.G.; Burke, J.M. The Hairpin Ribozyme: Structure, Assembly and Catalysis. *Curr Opin Chem Biol* **1998**, *2*, 24–30, doi:10.1016/S1367-5931(98)80032-X.
108. Rupert, P.B.; Ferré-D'Amaré, A.R. Crystal Structure of a Hairpin Ribozyme-Inhibitor Complex with Implications for Catalysis. *Nature* **2001**, *410*, 780–786, doi:10.1038/35071009.
109. Merrifield, R.B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem Soc* **1963**, *85*, 2149–2154, doi:10.1021/ja00897a025.
110. Josephson, S.; Lagerholm, E.; Palm, G.; Mörch, L.; Norin, T. Automatic Synthesis of Oligodeoxynucleotides and Mixed Oligodeoxynucleotides Using the Phosphoamidite Method. *Acta Chem Scand* **1984**, *38b*, 539–545, doi:10.3891/acta.chem.scand.38b-0539.
111. Flamme, M.; McKenzie, L.K.; Sarac, I.; Hollenstein, M. Chemical Methods for the Modification of RNA. *Methods* **2019**, *161*, 64–82, doi:10.1016/j.ymeth.2019.03.018.
112. Bartosik, K.; Debiec, K.; Czarnecka, A.; Sochacka, E.; Leszczynska, G. Synthesis of Nucleobase-Modified RNA Oligonucleotides by Post-Synthetic Approach. *Molecules* **2020**, *25*, 3344, doi:10.3390/molecules25153344.
113. Somoza, Á. Protecting Groups for RNA Synthesis: An Increasing Need for Selective Preparative Methods. *Chem Soc Rev* **2008**, *37*, 2668, doi:10.1039/b809851d.

114. Sekine, M. Recent Development of Chemical Synthesis of RNA. *Synthesis of Therapeutic Oligonucleotides*; 2018; pp. 41–65 ISBN 9789811319129.
115. Reese, C.B. Oligo- and Poly-Nucleotides: 50 Years of Chemical Synthesis. *Org Biomol Chem* **2005**, *3*, 3851–3868, doi:10.1039/b510458k.
116. Höbartner, C.; Micura, R. Chemical Synthesis of Selenium-Modified Oligoribonucleotides and Their Enzymatic Ligation Leading to an U6 SnRNA Stem-Loop Segment. *J Am Chem Soc* **2004**, *126*, 1141–1149, doi:10.1021/ja038481k.
117. Mead, R.J.; Shan, N.; Reiser, H.J.; Marshall, F.; Shaw, P.J. Amyotrophic Lateral Sclerosis: A Neurodegenerative Disorder Poised for Successful Therapeutic Translation. *Nat Rev Drug Discov* **2023**, *22*, 185–212, doi:10.1038/S41573-022-00612-2.
118. Chua, J.P.; De Calbiac, H.; Kabashi, E.; Barmada, S.J. Autophagy and ALS: Mechanistic Insights and Therapeutic Implications. *Autophagy* **2022**, *18*, doi:10.1080/15548627.2021.1926656.
119. Peters, O.M.; Brown, R.H. Amyotrophic Lateral Sclerosis. *Neurobiology of Brain Disorders: Biological Basis of Neurological and Psychiatric Disorders* **2015**, 262–280, doi:10.1016/B978-0-12-398270-4.00018-5.
120. Ferrari, R.; Kapogiannis, D.; Huey, E.; Momeni, P. FTD and ALS: A Tale of Two Diseases. *Curr Alzheimer Res* **2011**, *8*, 273–294, doi:10.2174/156720511795563700.
121. Woolley, S.; Goetz, R.; Factor-Litvak, P.; Murphy, J.; Hupf, J.; Lomen-Hoerth, C.; Andrews, H.; Heitzman, D.; Bedlack, R.; Katz, J.; et al. Longitudinal Screening Detects Cognitive Stability and Behavioral Deterioration in ALS Patients. *Behavioural neurology* **2018**, *2018*, doi:10.1155/2018/5969137.
122. Beeldman, E.; Raaphorst, J.; Twennaar, M.K.; De Visser, M.; Schmand, B.A.; De Haan, R.J. The Cognitive Profile of ALS: A Systematic Review and Meta-Analysis Update. *J Neurol Neurosurg Psychiatry* **2016**, *87*, 611–619, doi:10.1136/JNNP-2015-310734.

123. Hardiman, O.; Al-Chalabi, A.; Chio, A.; Corr, E.M.; Logroscino, G.; Robberecht, W.; Shaw, P.J.; Simmons, Z.; Van Den Berg, L.H. Amyotrophic Lateral Sclerosis. *Nat Rev Dis Primers* **2017**, *3*, doi:10.1038/NRDP.2017.71.
124. Goutman, S.A.; Hardiman, O.; Al-Chalabi, A.; Chió, A.; Savelieff, M.G.; Kiernan, M.C.; Feldman, E.L. Recent Advances in the Diagnosis and Prognosis of Amyotrophic Lateral Sclerosis. *Lancet Neurol* **2022**, *21*, 480–493, doi:10.1016/S1474-4422(21)00465-8.
125. van Es, M.A.; Hardiman, O.; Chio, A.; Al-Chalabi, A.; Pasterkamp, R.J.; Veldink, J.H.; van den Berg, L.H. Amyotrophic Lateral Sclerosis. *Lancet* **2017**, *390*, 2084–2098, doi:10.1016/S0140-6736(17)31287-4.
126. Grad, L.I.; Rouleau, G.A.; Ravits, J.; Cashman, N.R. Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). *Cold Spring Harb Perspect Med* **2017**, *7*, doi:10.1101/CSHPERSPECT.A024117.
127. Masrori, P.; Van Damme, P. Amyotrophic Lateral Sclerosis: A Clinical Review. *Eur J Neurol* **2020**, *27*, 1918–1929, doi:10.1111/ENE.14393.
128. Phukan, J.; Pender, N.P.; Hardiman, O. Cognitive Impairment in Amyotrophic Lateral Sclerosis. *Lancet Neurol* **2007**, *6*, 994–1003, doi:10.1016/S1474-4422(07)70265-X.
129. Logroscino, G.; Piccininni, M.; Marin, B.; Nichols, E.; Abd-Allah, F.; Abdelalim, A.; Alahdab, F.; Asgedom, S.W.; Awasthi, A.; Chaiah, Y.; et al. Global, Regional, and National Burden of Motor Neuron Diseases 1990-2016: A Systematic Analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* **2018**, *17*, 1083–1097, doi:10.1016/S1474-4422(18)30404-6.
130. Marin, B.; Boumé diene, F.; Logroscino, G.; Couratier, P.; Babron, M.C.; Leutenegger, A.L.; Copetti, M.; Preux, P.M.; Beghi, E. Variation in Worldwide Incidence of Amyotrophic Lateral Sclerosis: A Meta-Analysis. *Int J Epidemiol* **2017**, *46*, 57–74, doi:10.1093/IJE/DYW061.
131. Logroscino, G.; Traynor, B.J.; Hardiman, O.; Chió, A.; Mitchell, D.; Swingler, R.J.; Millul, A.; Benn, E.; Beghi, E. Incidence of Amyotrophic Lateral Sclerosis in Europe. *J Neurol Neurosurg Psychiatry* **2010**, *81*, 385–390, doi:10.1136/JNNP.2009.183525.

132. Longinetti, E.; Fang, F. Epidemiology of Amyotrophic Lateral Sclerosis: An Update of Recent Literature. *Curr Opin Neurol* **2019**, *32*, 771–776, doi:10.1097/WCO.0000000000000730.
133. Johnston, C.A.; Stanton, B.R.; Turner, M.R.; Gray, R.; Blunt, A.H.M.; Butt, D.; Ampong, M.A.; Shaw, C.E.; Leigh, P.N.; Al-Chalabi, A. Amyotrophic Lateral Sclerosis in an Urban Setting: A Population Based Study of Inner City London. *J Neurol* **2006**, *253*, 1642–1643, doi:10.1007/S00415-006-0195-Y.
134. Ryan, M.; Heverin, M.; McLaughlin, R.L.; Hardiman, O. Lifetime Risk and Heritability of Amyotrophic Lateral Sclerosis. *JAMA Neurol* **2019**, *76*, 1367–1374, doi:10.1001/JAMANEUROL.2019.2044.
135. Manjaly, Z.R.; Scott, K.M.; Abhinav, K.; Wijesekera, L.; Ganesalingam, J.; Goldstein, L.H.; Janssen, A.; Dougherty, A.; Willey, E.; Stanton, B.R.; et al. The Sex Ratio in Amyotrophic Lateral Sclerosis: A Population Based Study. <https://doi.org/10.3109/17482961003610853> **2010**, *11*, 439–442, doi:10.3109/17482961003610853.
136. Piol, D.; Robberechts, T.; Da Cruz, S. Lost in Local Translation: TDP-43 and FUS in Axonal/Neuromuscular Junction Maintenance and Dysregulation in Amyotrophic Lateral Sclerosis. *Neuron* **2023**, *111*, 1355–1380, doi:10.1016/J.NEURON.2023.02.028.
137. Tamaki, Y.; Urushitani, M. Molecular Dissection of TDP-43 as a Leading Cause of ALS/FTLD. *Int J Mol Sci* **2022**, *23*, doi:10.3390/IJMS232012508.
138. Peggion, C.; Scalcon, V.; Massimino, M.L.; Nies, K.; Lopreiato, R.; Rigobello, M.P.; Bertoli, A. SOD1 in ALS: Taking Stock in Pathogenic Mechanisms and the Role of Glial and Muscle Cells. *Antioxidants (Basel)* **2022**, *11*, doi:10.3390/ANTIOX11040614.
139. Rosen, D.R.; Siddique, T.; Patterson, D.; Figlewicz, D.A.; Sapp, P.; Hentati, A.; Donaldson, D.; Goto, J.; O'Regan, J.P.; Deng, H.X.; et al. Mutations in Cu/Zn Superoxide Dismutase Gene Are Associated with Familial Amyotrophic Lateral Sclerosis. *Nature* **1993**, *362*, 59–62, doi:10.1038/362059a0.
140. Kwiatkowski, T.J.; Bosco, D.A.; LeClerc, A.L.; Tamrazian, E.; Vanderburg, C.R.; Russ, C.; Davis, A.; Gilchrist, J.; Kasarskis, E.J.; Munsat, T.; et al. Mutations in the

- FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *Science* **2009**, *323*, 1205–1208, doi:10.1126/SCIENCE.1166066.
141. Vance, C.; Rogelj, B.; Hortobágyi, T.; De Vos, K.J.; Nishimura, A.L.; Sreedharan, J.; Hu, X.; Smith, B.; Ruddy, D.; Wright, P.; et al. Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. *Science* **2009**, *323*, 1208–1211, doi:10.1126/SCIENCE.1165942.
142. Sreedharan, J.; Blair, I.P.; Tripathi, V.B.; Hu, X.; Vance, C.; Rogelj, B.; Ackerley, S.; Durnall, J.C.; Williams, K.L.; Buratti, E.; et al. TDP-43 Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis. *Science* **2008**, *319*, 1668–1672, doi:10.1126/SCIENCE.1154584.
143. Kabashi, E.; Valdmanis, P.N.; Dion, P.; Spiegelman, D.; McConkey, B.J.; Velde, C. Vande; Bouchard, J.P.; Lacomblez, L.; Pochigaeva, K.; Salachas, F.; et al. TARDBP Mutations in Individuals with Sporadic and Familial Amyotrophic Lateral Sclerosis. *Nature Genetics* **2008**, *40*:5, 572–574, doi:10.1038/ng.132.
144. DeJesus-Hernandez, M.; Mackenzie, I.R.; Boeve, B.F.; Boxer, A.L.; Baker, M.; Rutherford, N.J.; Nicholson, A.M.; Finch, N.C.A.; Flynn, H.; Adamson, J.; et al. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **2011**, *72*, 245–256, doi:10.1016/J.NEURON.2011.09.011.
145. Renton, A.E.; Majounie, E.; Waite, A.; Simón-Sánchez, J.; Rollinson, S.; Gibbs, J.R.; Schymick, J.C.; Laaksovirta, H.; van Swieten, J.C.; Myllykangas, L.; et al. A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **2011**, *72*, 257–268, doi:10.1016/J.NEURON.2011.09.010.
146. Masrori, P.; Beckers, J.; Gossye, H.; Van Damme, P. The Role of Inflammation in Neurodegeneration: Novel Insights into the Role of the Immune System in C9orf72 HRE-Mediated ALS/FTD. *Mol Neurodegener* **2022**, *17*, doi:10.1186/S13024-022-00525-Z.
147. Diab, R.; Pilotto, F.; Saxena, S. Autophagy and Neurodegeneration: Unraveling the Role of C9ORF72 in the Regulation of Autophagy and Its Relationship to ALS-

- FTD Pathology. *Front Cell Neurosci* **2023**, *17*, doi:10.3389/FNCEL.2023.1086895.
148. Ghaffari, L.T.; Trotti, D.; Haeusler, A.R.; Jensen, B.K. Breakdown of the Central Synapses in C9orf72-Linked ALS/FTD. *Front Mol Neurosci* **2022**, *15*, doi:10.3389/FNMOL.2022.1005112.
149. Cirulli, E.T.; Lasseigne, B.N.; Petrovski, S.; Sapp, P.C.; Dion, P.A.; Leblond, C.S.; Couthouis, J.; Lu, Y.F.; Wang, Q.; Krueger, B.J.; et al. Exome Sequencing in Amyotrophic Lateral Sclerosis Identifies Risk Genes and Pathways. *Science* **2015**, *347*, 1436–1441, doi:10.1126/SCIENCE.AAA3650.
150. Freischmidt, A.; Wieland, T.; Richter, B.; Ruf, W.; Schaeffer, V.; Müller, K.; Marroquin, N.; Nordin, F.; Hübers, A.; Weydt, P.; et al. Haploinsufficiency of TBK1 Causes Familial ALS and Fronto-Temporal Dementia. *Nat Neurosci* **2015**, *18*, 631–636, doi:10.1038/NN.4000.
151. Oakes, J.A.; Davies, M.C.; Collins, M.O. TBK1: A New Player in ALS Linking Autophagy and Neuroinflammation. *Mol Brain* **2017**, *10*, 1–10, doi:10.1186/S13041-017-0287-X.
152. Gurfinkel, Y.; Polain, N.; Sonar, K.; Nice, P.; Mancera, R.L.; Rea, S.L. Functional and Structural Consequences of TBK1 Missense Variants in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Neurobiol Dis* **2022**, *174*, doi:10.1016/J.NBD.2022.105859.
153. Van Blitterswijk, M.; Van Es, M.A.; Hennekam, E.A.M.; Dooijes, D.; Van Rheenen, W.; Medic, J.; Bourque, P.R.; Schelhaas, H.J.; Van der Kooi, A.J.; De Visser, M.; et al. Evidence for an Oligogenic Basis of Amyotrophic Lateral Sclerosis. *Hum Mol Genet* **2012**, *21*, 3776–3784, doi:10.1093/HMG/DDS199.
154. Puppala, G.K.; Gorthi, S.P.; Chandran, V.; Gundabolu, G. Frontotemporal Dementia - Current Concepts. *Neurol India* **2021**, *69*, 1144–1152, doi:10.4103/0028-3886.329593.
155. Kelley, R.E.; El-Khoury, R. Frontotemporal Dementia. *Neurol Clin* **2016**, *34*, 171–181, doi:10.1016/j.ncl.2015.08.007.

156. Younes, K.; Miller, B.L. Neuropsychiatric Aspects of Frontotemporal Dementia. *Psychiatric Clinics of North America* **2020**, *43*, 345–360, doi:10.1016/j.psc.2020.02.005.
157. Graff-Radford, N.R.; Woodruff, B.K. Frontotemporal Dementia. *Semin Neurol* **2007**, *27*, 48–57, doi:10.1055/S-2006-956755.
158. Solomon, D.A.; Mitchell, J.C.; Salcher-Konrad, M.T.; Vance, C.A.; Mizielinska, S. Review: Modelling the Pathology and Behaviour of Frontotemporal Dementia. *Neuropathol Appl Neurobiol* **2019**, *45*, 58–80, doi:10.1111/nan.12536.
159. Warren, J.D.; Rohrer, J.D.; Rossor, M.N. Clinical Review. Frontotemporal Dementia. *BMJ* **2013**, *347*, doi:10.1136/BMJ.F4827.
160. Chen, J.A.; Wang, Q.; Davis-Turak, J.; Li, Y.; Karydas, A.M.; Hsu, S.C.; Sears, R.L.; Chatzopoulou, D.; Huang, A.Y.; Wojta, K.J.; et al. A Multiancestral Genome-Wide Exome Array Study of Alzheimer Disease, Frontotemporal Dementia, and Progressive Supranuclear Palsy. *JAMA Neurol* **2015**, *72*, 414–422, doi:10.1001/jamaneurol.2014.4040.
161. Sivasathiaselan, H.; Marshall, C.R.; Agustus, J.L.; Benhamou, E.; Bond, R.L.; Van Leeuwen, J.E.P.; Hardy, C.J.D.; Rohrer, J.D.; Warren, J.D. Frontotemporal Dementia: A Clinical Review. *Semin Neurol* **2019**, *39*, 251–263, doi:10.1055/S-0039-1683379.
162. Young, J.J.; Lavakumar, M.; Tampi, D.; Balachandran, S.; Tampi, R.R. Frontotemporal Dementia: Latest Evidence and Clinical Implications. *Ther Adv Psychopharmacol* **2018**, *8*, 33–48, doi:10.1177/2045125317739818.
163. Barczak, A.; Hintze, B. Early Diagnosis of Behavioral Variant of Frontotemporal Dementia – a Clinical Perspective. *Neuropsychiatria I Neuropsychologia/Neuropsychiatry and Neuropsychology* **2022**, *17*, 16–23, doi:10.5114/NAN.2022.117952.
164. Livingston, G.; Huntley, J.; Sommerlad, A.; Ames, D.; Ballard, C.; Banerjee, S.; Brayne, C.; Burns, A.; Cohen-Mansfield, J.; Cooper, C.; et al. Dementia Prevention, Intervention, and Care: 2020 Report of the Lancet Commission. *Lancet* **2020**, *396*, 413–446, doi:10.1016/S0140-6736(20)30367-6.

165. Chokesuwattanaskul, A.; Jiang, H.; Bond, R.L.; Jimenez, D.A.; Russell, L.L.; Sivasathiaseelan, H.; Johnson, J.C.S.; Benhamou, E.; Agustus, J.L.; van Leeuwen, J.E.P.; et al. The Architecture of Abnormal Reward Behaviour in Dementia: Multimodal Hedonic Phenotypes and Brain Substrate. *Brain Commun* **2023**, *5*, fcad027, doi:10.1093/braincomms/fcad027.
166. Harciarek, M.; Cosentino, S. Language, Executive Function and Social Cognition in the Diagnosis of Frontotemporal Dementia Syndromes. *International Review of Psychiatry* **2013**, *25*, 178–196, doi:10.3109/09540261.2013.763340.
167. Devenney, E.M.; Ahmed, R.M.; Hodges, J.R. Frontotemporal Dementia. *Handb Clin Neurol* **2019**, *167*, 279–299, doi:10.1016/B978-0-12-804766-8.00015-7.
168. Boeve, B.F.; Boxer, A.L.; Kumfor, F.; Pijnenburg, Y.; Rohrer, J.D. Advances and Controversies in Frontotemporal Dementia: Diagnosis, Biomarkers, and Therapeutic Considerations. *Lancet Neurol* **2022**, *21*, 258–272, doi:10.1016/S1474-4422(21)00341-0.
169. Wang, J.; Wang, B.; Zhou, T. The Advance on Frontotemporal Dementia (FTD)’s Neuropathology and Molecular Genetics. *Mediators Inflamm* **2022**, *2022*, doi:10.1155/2022/5003902.
170. Hautbergue, G.M.; Cleary, J.D.; Guo, S.; Ranum, L.P.W. Therapeutic Strategies for C9orf72 Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Curr Opin Neurol* **2021**, *34*, 748–755, doi:10.1097/WCO.0000000000000984.
171. Goodman, L.D.; Bonini, N.M. New Roles for Canonical Transcription Factors in Repeat Expansion Diseases. *Trends Genet* **2020**, *36*, 81–92, doi:10.1016/J.TIG.2019.11.003.
172. Haeusler, A.R.; Donnelly, C.J.; Periz, G.; Simko, E.A.J.; Shaw, P.G.; Kim, M.S.; Maragakis, N.J.; Troncoso, J.C.; Pandey, A.; Sattler, R.; et al. C9orf72 Nucleotide Repeat Structures Initiate Molecular Cascades of Disease. *Nature* **2014**, *507*, 195–200, doi:10.1038/nature13124.
173. Ash, P.E.A.; Bieniek, K.F.; Gendron, T.F.; Caulfield, T.; Lin, W.L.; DeJesus-Hernandez, M.; Van Blitterswijk, M.M.; Jansen-West, K.; Paul, J.W.; Rademakers, R.; et al. Unconventional Translation of C9ORF72 GGGGCC

- Expansion Generates Insoluble Polypeptides Specific to C9FTD/ALS. *Neuron* **2013**, *77*, 639–646, doi:10.1016/j.neuron.2013.02.004.
174. Wang, Z.F.; Ursu, A.; Childs-Disney, J.L.; Guertler, R.; Yang, W.Y.; Bernat, V.; Rzuczek, S.G.; Fuerst, R.; Zhang, Y.J.; Gendron, T.F.; et al. The Hairpin Form of r(G4C2)Exp in C9ALS/FTD Is Repeat-Associated Non-ATG Translated and a Target for Bioactive Small Molecules. *Cell Chem Biol* **2019**, *26*, 179-190.e12, doi:10.1016/J.CHEMBIOL.2018.10.018.
175. Šket, P.; Pohleven, J.; Kovanda, A.; Štalekar, M.; Župunski, V.; Zalar, M.; Plavec, J.; Rogelj, B. Characterization of DNA G-Quadruplex Species Forming from C9ORF72 G4C2-Expanded Repeats Associated with Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. *Neurobiol Aging* **2015**, *36*, 1091–1096, doi:10.1016/J.NEUROBIOLAGING.2014.09.012.
176. Brčić, J.; Plavec, J. ALS and FTD Linked GGGGCC-Repeat Containing DNA Oligonucleotide Folds into Two Distinct G-Quadruplexes. *Biochim Biophys Acta Gen Subj* **2017**, *1861*, 1237–1245, doi:10.1016/J.BBAGEN.2016.11.018.
177. Su, Z.; Zhang, Y.; Gendron, T.F.; Bauer, P.O.; Chew, J.; Yang, W.Y.; Fostvedt, E.; Jansen-West, K.; Belzil, V. V.; Desaro, P.; et al. Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in C9FTD/ALS. *Neuron* **2014**, *83*, 1043–1050, doi:10.1016/J.NEURON.2014.07.041.
178. Reddy, K.; Zamiri, B.; Stanley, S.Y.R.; Macgregor, R.B.; Pearson, C.E. The Disease-Associated r(GGGGCC)_n Repeat from the C9orf72 Gene Forms Tract Length-Dependent Uni- and Multimolecular RNA G-Quadruplex Structures. *J Biol Chem* **2013**, *288*, 9860–9866, doi:10.1074/JBC.C113.452532.
179. Li, F.; Zhou, J. G-Quadruplexes from Non-Coding RNAs. *J Mol Med (Berl)* **2023**, *101*, 621–635, doi:10.1007/S00109-023-02314-7.
180. Reddy, K.; Schmidt, M.H.M.; Geist, J.M.; Thakkar, N.P.; Panigrahi, G.B.; Wang, Y.H.; Pearson, C.E. Processing of Double-R-Loops in (CAG)_n(CTG)_n and C9orf72 (GGGGCC)_n(GGCCCC)_n Repeats Causes Instability. *Nucleic Acids Res* **2014**, *42*, 10473–10487, doi:10.1093/NAR/GKU658.
181. Diab, M.A.; Mor-Shaked, H.; Cohen, E.; Cohen-Hadad, Y.; Ram, O.; Epsztejn-Litman, S.; Eiges, R. The G-Rich Repeats in FMR1 and C9orf72 Loci Are

- Hotspots for Local Unpairing of DNA. *Genetics* **2018**, *210*, 1239–1252, doi:10.1534/GENETICS.118.301672.
182. Zamiri, B.; Mirceta, M.; Bomsztyk, K.; Macgregor, R.B.; Pearson, C.E. Quadruplex Formation by Both G-Rich and C-Rich DNA Strands of the C9orf72 (GGGGCC)₈•(GGCCCC)₈ Repeat: Effect of CpG Methylation. *Nucleic Acids Res* **2015**, *43*, 10055–10064, doi:10.1093/NAR/GKV1008.
183. Zhang, Y.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices Obtained from the GGGGCC and CCCC GG Hexanucleotide Repeats That Are the Hallmark of C9FTD/ALS Diseases. *ACS Chem Neurosci* **2017**, *8*, 578–591, doi:10.1021/ACSCHEMNEURO.6B00348.
184. Zamiri, B.; Mirceta, M.; Abu-Ghazalah, R.; Wold, M.S.; Pearson, C.E.; Macgregor, R.B. Stress-Induced Acidification May Contribute to Formation of Unusual Structures in C9orf72-Repeats. *Biochim Biophys Acta Gen Subj* **2018**, *1862*, 1482–1491, doi:10.1016/J.BBAGEN.2018.03.001.
185. Snoussi, K.; Nonin-Lecomte, S.; Leroy, J.L. The RNA I-Motif. *J Mol Biol* **2001**, *309*, 139–153, doi:10.1006/JMBI.2001.4618.
186. Dodd, D.W.; Tomchick, D.R.; Corey, D.R.; Gagnon, K.T. Pathogenic C9ORF72 Antisense Repeat RNA Forms a Double Helix with Tandem C:C Mismatches. *Biochemistry* **2016**, *55*, 1283–1286, doi:10.1021/ACS.BIOCHEM.6B00136/SUPPL_FILE/BI6B00136_SI_001.PDF.
187. Esteller, M. Non-Coding RNAs in Human Disease. *Nat Rev Genet* **2011**, *12*, 861–874, doi:10.1038/nrg3074.
188. Cooper, T.A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777–793, doi:10.1016/j.cell.2009.02.011.
189. Nossent, A.Y. The Epitranscriptome: RNA Modifications in Vascular Remodelling. *Atherosclerosis* **2023**, *374*, 24–33, doi:10.1016/J.ATHEROSCLEROSIS.2022.11.004.
190. Mukherjee, S.; Dohno, C.; Asano, K.; Nakatani, K. Cyclic Mismatch Binding Ligand CMBL4 Binds to the 5'-T-3'/5'-GG-3' Site by Inducing the Flipping out

- of Thymine Base. *Nucleic Acids Res* **2016**, *44*, 7090–7099, doi:10.1093/NAR/GKW672.
191. Connelly, C.M.; Moon, M.H.; Schneekloth, J.S. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016**, *23*, 1077–1090, doi:10.1016/J.CHEMBIOL.2016.05.021.
192. Crooke, S.T.; Witztum, J.L.; Bennett, C.F.; Baker, B.F. RNA-Targeted Therapeutics. *Cell Metab* **2018**, *27*, 714–739, doi:10.1016/J.CMET.2018.03.004.
193. Yu, A.M.; Choi, Y.H.; Tu, M.J. RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. *Pharmacol Rev* **2020**, *72*, 862–898, doi:10.1124/PR.120.019554.
194. Hof, F.; Craig, S.L.; Nuckolls, C.; Rebek Jr., J. Molecular Encapsulation. *Angewandte Chemie International Edition* **2002**, *41*, 1488–1508, doi:10.1002/1521-3773(20020503)41:9<1488::AID-ANIE1488>3.0.CO;2-G.
195. Childs-Disney, J.L.; Yang, X.; Gibaut, Q.M.R.; Tong, Y.; Batey, R.T.; Disney, M.D. Targeting RNA Structures with Small Molecules. *Nat Rev Drug Discov* **2022**, *21*, 736–762, doi:10.1038/S41573-022-00521-4.
196. Ramaiah, D.; Neelakandan, P.P.; Nair, A.K.; Avirah, R.R. Functional Cyclophanes: Promising Hosts for Optical Biomolecular Recognition. *Chem Soc Rev* **2010**, *39*, 4158–4168, doi:10.1039/B920032K.
197. Murata, A.; Nakamori, M.; Nakatani, K. Modulating RNA Secondary and Tertiary Structures by Mismatch Binding Ligands. *Methods* **2019**, *167*, 78–91, doi:10.1016/J.YMETH.2019.05.006.
198. Yoon, J.; Kim, S.K.; Singh, N.J.; Kim, K.S. Imidazolium Receptors for the Recognition of Anions. *Chem Soc Rev* **2006**, *35*, 355–360, doi:10.1039/B513733K.
199. Choquesillo-Lazarte, D.; Brandi-Blanco, M. del P.; García-Santos, I.; González-Pérez, J.M.; Castiñeiras, A.; Niclós-Gutiérrez, J. Interligand Interactions Involved in the Molecular Recognition between Copper(II) Complexes and Adenine or Related Purines. *Coord Chem Rev* **2008**, *252*, 1241–1256, doi:10.1016/J.CCR.2007.09.018.

200. Wright, A.T.; Anslyn, E. V. Differential Receptor Arrays and Assays for Solution-Based Molecular Recognition. *Chem Soc Rev* **2006**, *35*, 14–28, doi:10.1039/B505518K.
201. Nonaka, H.; Tsukiji, S.; Ojida, A.; Hamachi, I. Non-Enzymatic Covalent Protein Labeling Using a Reactive Tag. *J Am Chem Soc* **2007**, *129*, 15777–15779, doi:10.1021/JA074176D.
202. Ramaiah, D.; Neelakandan, P.P.; Nair, A.K.; Avirah, R.R. Functional Cyclophanes: Promising Hosts for Optical Biomolecular Recognition. *Chem Soc Rev* **2010**, *39*, 4158–4168, doi:10.1039/B920032K.
203. Mukherjee, S.; Dohno, C.; Nakatani, K. Design and Synthesis of Cyclic Mismatch-Binding Ligands (CMBLs) with Variable Linkers by Ring-Closing Metathesis and Their Photophysical and DNA Repeat Binding Properties. *Chemistry – A European Journal* **2017**, *23*, 11385–11396, doi:10.1002/CHEM.201702064.
204. Nozawa, R.; Kim, J.; Oh, J.; Lamping, A.; Wang, Y.; Shimizu, S.; Hisaki, I.; Kowalczyk, T.; Fliegl, H.; Kim, D.; et al. Three-Dimensional Aromaticity in an Antiaromatic Cyclophane. *Nature Communications 2019 10:1* **2019**, *10*, 1–7, doi:10.1038/s41467-019-11467-4.
205. Das, B.; Nagano, K.; Kawai, G.; Murata, A.; Nakatani, K. 2-Amino-1,8-Naphthyridine Dimer (ANP77), a High-Affinity Binder to the Internal Loops of C/CC and T/CC Sites in Double-Stranded DNA. *Journal of Organic Chemistry* **2022**, *87*, 340–350, doi:10.1021/ACS.JOC.1C02383/SUPPL_FILE/JO1C02383_SI_001.PDF.
206. Das, B.; Murata, A.; Nakatani, K. A Small-Molecule Fluorescence Probe ANP77 for Sensing RNA Internal Loop of C, U and A/CC Motifs and Their Binding Molecules. *Nucleic Acids Res* **2021**, *49*, 8462–8470, doi:10.1093/NAR/GKAB650.
207. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser Crystallographic Software. *J Appl Crystallogr* **2007**, *40*, 658–674, doi:10.1107/S0021889807021206.
208. Tronrud, D.E. Introduction to Macromolecular Refinement. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2156–2168, doi:10.1107/S090744490402356X.

209. Parkinson, G.; Vojtechovsky, J.; Clowney, L.; Brünger, A.T.; Berman, H.M. New Parameters for the Refinement of Nucleic Acid-Containing Structures. *Acta Crystallogr D Biol Crystallogr* **1996**, *52*, 57–64, doi:10.1107/S0907444995011115.
210. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486–501, doi:10.1107/S0907444910007493.
211. Kleywegt, G.J.; Jones, T.A. Model Building and Refinement Practice. *Methods Enzymol* **1997**, *277*, 208–230, doi:10.1016/S0076-6879(97)77013-7.
212. Brünger, A.T. Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *Nature* **1992**, *355*, 472–475, doi:10.1038/355472A0.
213. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures. *FEBS J* **2008**, *275*, 1–21, doi:10.1111/J.1742-4658.2007.06178.X.
214. Thompson, R.F.; Walker, M.; Siebert, C.A.; Muench, S.P.; Ranson, N.A. An Introduction to Sample Preparation and Imaging by Cryo-Electron Microscopy for Structural Biology. *Methods* **2016**, *100*, 3–15, doi:10.1016/J.YMETH.2016.02.017.
215. Matthews, B.W. Solvent Content of Protein Crystals. *J Mol Biol* **1968**, *33*, 491–497, doi:10.1016/0022-2836(68)90205-2.
216. Storoni, L.C.; McCoy, A.J.; Read, R.J. Likelihood-Enhanced Fast Rotation Functions. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 432–438, doi:10.1107/S0907444903028956.
217. Kaufmann, G.; Kallenbach, N.R. Determination of Recognition Sites of T4 RNA Ligase on the 3'-OH and 5'-P Termini of Polyribonucleotide Chains. *Nature* **1975**, *254*, 452–454, doi:10.1038/254452A0.

218. Teranishi, K.; Ueno, F. Mechanism of 2-O→3-O Silyl Migration in Cyclomaltohexaose (α -Cyclodextrin). *Tetrahedron Lett* **2003**, *44*, 4843–4848, doi:10.1016/S0040-4039(03)01149-3.
219. Kao, C.; Rüdiger, S.; Zheng, M. A Simple and Efficient Method to Transcribe RNAs with Reduced 3' Heterogeneity. *Methods* **2001**, *23*, 201–205, doi:10.1006/meth.2000.1131.
220. Das, U.; Shuman, S. Mechanism of RNA 2',3'-Cyclic Phosphate End Healing by T4 Polynucleotide Kinase-Phosphatase. *Nucleic Acids Res* **2013**, *41*, 355–365, doi:10.1093/NAR/GKS977.
221. Tanner, N.K.; Schaff, S.; Thill, G.; Petit-Koskas, E.; Crain-Denoyelle, A.M.; Westhof, E. A Three-Dimensional Model of Hepatitis Delta Virus Ribozyme Based on Biochemical and Mutational Analyses. *Current Biology* **1994**, *4*, 488–498, doi:10.1016/S0960-9822(00)00109-3.
222. Webb, C.-H.T.; Lupták, A. HDV-like Self-Cleaving Ribozymes. *RNA Biol* **2011**, *8*, 719–727, doi:10.4161/rna.8.5.16226.
223. Chen, H.; Egger, M.; Xu, X.; Flemmich, L.; Krasheninina, O.; Sun, A.; Micura, R.; Ren, A. Structural Distinctions between NAD⁺-riboswitch Domains 1 and 2 Determine Differential Folding and Ligand Binding. *Nucleic Acids Res* **2020**, *48*, 12394–12406, doi:10.1093/nar/gkaa1029.
224. Doudna, J.A. Preparation of Homogeneous Ribozyme RNA for Crystallization. *Methods in molecular biology* **1997**, *74*, 365–370, doi:10.1385/0-89603-389-9:365.
225. Peselis, A.; Gao, A.; Serganov, A. Preparation and Crystallization of Riboswitches. *Methods in Molecular Biology* **2016**, *1320*, 21–36, doi:10.1007/978-1-4939-2763-0_3.
226. Steckelberg, A.L.; Akiyama, B.M.; Costantino, D.A.; Sit, T.L.; Nix, J.C.; Kieft, J.S. A Folded Viral Noncoding RNA Blocks Host Cell Exoribonucleases through a Conformationally Dynamic RNA Structure. *Proc Natl Acad Sci U S A* **2018**, *115*, 6404–6409, doi:10.1073/pnas.1802429115.

227. Fedor, M.J. Structure and Function of the Hairpin Ribozyme. *J Mol Biol* **2000**, 297, 269–291, doi:10.1006/JMBI.2000.3560.
228. Harris, M.E.; Christian, E.L. RNA Crosslinking Methods. *Methods Enzymol* **2009**, 468, 127–146, doi:10.1016/S0076-6879(09)68007-1.
229. McPherson, A.; Cudney, B. Optimization of Crystallization Conditions for Biological Macromolecules. *Acta Crystallogr F Struct Biol Commun* **2014**, 70, 1445–1467, doi:10.1107/S2053230X14019670.
230. Wang, H.W.; Wang, J.W. How Cryo-Electron Microscopy and X-Ray Crystallography Complement Each Other. *Protein Sci* **2017**, 26, 32–39, doi:10.1002/PRO.3022.
231. Zhang, K.; Zheludev, I.N.; Hagey, R.J.; Haslecker, R.; Hou, Y.J.; Kretsch, R.; Pintilie, G.D.; Rangan, R.; Kladwang, W.; Li, S.; et al. Cryo-EM and Antisense Targeting of the 28-KDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. *Nat Struct Mol Biol* **2021**, 28, 747–754, doi:10.1038/S41594-021-00653-Y.
232. Zhang, K.; Li, S.; Kappel, K.; Pintilie, G.; Su, Z.; Mou, T.C.; Schmid, M.F.; Das, R.; Chiu, W. Cryo-EM Structure of a 40 KDa SAM-IV Riboswitch RNA at 3.7 Å Resolution. *Nature Communications* 2019 10:1 **2019**, 10, 1–6, doi:10.1038/s41467-019-13494-7.
233. Douglas, A.G.L. Non-Coding RNA in C9orf72-Related Amyotrophic Lateral Sclerosis and Frontotemporal Dementia: A Perfect Storm of Dysfunction. *Noncoding RNA Res* **2018**, 3, 178–187, doi:10.1016/J.NCRNA.2018.09.001.
234. Fratta, P.; Mizielińska, S.; Nicoll, A.J.; Zloh, M.; Fisher, E.M.C.; Parkinson, G.; Isaacs, A.M. C9orf72 Hexanucleotide Repeat Associated with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Forms RNA G-Quadruplexes. *Sci Rep* **2012**, 2, doi:10.1038/SREP01016.
235. Levene, P.A.; Bass, L.W.; Simms, H.S. THE IONIZATION OF PYRIMIDINES IN RELATION TO THE STRUCTURE OF PYRIMIDINE NUCLEOSIDES. *Journal of Biological Chemistry* **1926**, 70, 229–241, doi:10.1016/S0021-9258(18)84540-9.

236. Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid I-Motif Structures. *RSC Adv* **2014**, *4*, 26956–26980, doi:10.1039/C4RA02129K.
237. Sun, P.D.; Davies, D.R. The Cystine-Knot Growth-Factor Superfamily. *Annu Rev Biophys Biomol Struct* **1995**, *24*, 269–292, doi:10.1146/ANNUREV.BB.24.060195.001413.
238. Holland, J.A.; Hoffman, D.W. Structural Features and Stability of an RNA Triple Helix in Solution. *Nucleic Acids Res* **1996**, *24*, 2841–2848, doi:10.1093/NAR/24.14.2841.
239. Moody, E.M.; Lecomte, J.T.J.; Bevilacqua, P.C. Linkage between Proton Binding and Folding in RNA: A Thermodynamic Framework and Its Experimental Application for Investigating PKa Shifting. *RNA* **2005**, *11*, 157–172, doi:10.1261/RNA.7177505.
240. Jones, E.L.; Mlotkowski, A.J.; Hebert, S.P.; Schlegel, H.B.; Chow, C.S. Calculations of p Ka Values for a Series of Naturally Occurring Modified Nucleobases. *J Phys Chem A* **2022**, *126*, 1518–1529, doi:10.1021/ACS.JPCA.1C10905.
241. Tang, C.L.; Alexov, E.; Pyle, A.M.; Honig, B. Calculation of PKas in RNA: On the Structural Origins and Functional Roles of Protonated Nucleotides. *J Mol Biol* **2007**, *366*, 1475–1496, doi:10.1016/J.JMB.2006.12.001.
242. Ferré-D'Amaré, A.R.; Zhou, K.; Doudna, J.A. Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature* **1998**, *395*, 567–574, doi:10.1038/26912.
243. Ke, A.; Zhou, K.; Ding, F.; Cate, J.H.D.; Doudna, J.A. A Conformational Switch Controls Hepatitis Delta Virus Ribozyme Catalysis. *Nature* **2004**, *429*, 201–205, doi:10.1038/NATURE02522.
244. Nixon, P.L.; Cornish, P. V.; Suram, S. V.; Giedroc, D.P. Thermodynamic Analysis of Conserved Loop-Stem Interactions in P1-P2 Frameshifting RNA Pseudoknots from Plant Luteoviridae. *Biochemistry* **2002**, *41*, 10665–10674, doi:10.1021/BI025843C.

245. Nixon, P.L.; Giedroc, D.P. Energetics of a Strongly PH Dependent RNA Tertiary Structure in a Frameshifting Pseudoknot. *J Mol Biol* **2000**, *296*, 659–671, doi:10.1006/JMBI.1999.3464.
246. Su, L.; Chen, L.; Egli, M.; Berger, J.M.; Rich, A. Minor Groove RNA Triplex in the Crystal Structure of a Ribosomal Frameshifting Viral Pseudoknot. *Nat Struct Biol* **1999**, *6*, 285–292, doi:10.1038/6722.
247. Nixon, P.L.; Rangan, A.; Kim, Y.G.; Rich, A.; Hoffman, D.W.; Hennig, M.; Giedroc, D.P. Solution Structure of a Luteoviral P1-P2 Frameshifting MRNA Pseudoknot. *J Mol Biol* **2002**, *322*, 621–633, doi:10.1016/S0022-2836(02)00779-9.
248. Kovanda, A.; Zalar, M.; Šket, P.; Plavec, J.; Rogelj, B. Anti-Sense DNA d(GGCCCC)_n Expansions in C9ORF72 Form i-Motifs and Protonated Hairpins. *Sci Rep* **2015**, *5*, doi:10.1038/SREP17944.