

Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk

**Badania strukturalne
powtórzeń CNG w sekwencjach RNA
o znaczeniu w patogenezie TREDs**

Agnieszka Kiliszek

Praca doktorska wykonana w Zespole Struktury i Funkcji Biomolekuł
pod kierunkiem prof. dr. hab. Wojciecha Rypniewskiego

Poznań 2011

Dziękuję mojemu promotorowi prof. dr. hab. Wojciechowi Rypniewskiemu za zaangażowanie w realizację prowadzonych przeze mnie badań, owocne dyskusje oraz cenne wskazówki.

Dziękuję za współpracę prof. dr. hab. Ryszardowi Kierzkowi oraz koleżankom i kolegom z Pracowni CBB.

Dziękuję również Leszkowi za wsparcie i dyskusję „dotyczących” pracy oraz moim przyjaciołom i całej rodzinie.

Tak naprawdę, to nie rozumiesz
jakiegoś zagadnienia, dopóki nie
potrafisz go wytłumaczyć swojej
babci.

(Albert Einstein)

Dla moich bliskich

**Badania prowadzone w ramach niniejszej pracy doktorskiej
finansowane były z grantów MNiSW:**

”Struktura RNA trójnukleotydowych powtórzeń typu CNG związanych z chorobami neurologicznymi. Analiza krystalograficzna.” Numer grantu: N-N301-0171634.

”Enzymatyczna i chemiczna stabilizacja formy spinki w oligomerach RNA do badań biokrystalograficznych.” Numer grantu: N-N-301472238.

**Część prezentowanych przeze mnie wyników
została zawarta w publikacjach:**

Kiliszek A., Kierzek K., Krzyżosiak W.J. i Rypniewski W. *Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome*. Nucleic Acids Res. (2011 May 19. Epub ahead of print)

Kiliszek A., Kierzek K., Krzyżosiak W.J. i 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.

Kiliszek A., Kierzek K., Krzyżosiak W.J. i Rypniewski W. *Structural insights into CUG repeats containing the 'stretched U-U wobble': implications for myotonic dystrophy*. Nucleic Acids Res. (2009) **37**:4149-56.

Spis treści

Spis stosowanych skrótów	10
1. Cel pracy doktorskiej	12
2. Wstęp	14
2.1. Występowanie powtórzeń trójnukleotydowych w genomie	14
2.2. Patogenne powtórzenia typu CNG związane z chorobami TREDs	15
2.3. Rola powtórzeń CNG znajdujących się w mRNA niektórych genów w patomechanizmie chorób TREDs	19
2.3.1. Powtórzenia CUG	20
2.3.2. Powtórzenia CGG	23
2.3.3. Powtórzenia CCG	24
2.3.4. Powtórzenia CAG	25
2.4. Struktura drugorzędowa cząsteczek RNA zawierających powtórzenia typu CNG	26
2.5. Właściwości termodynamiczne powtórzeń trójnukleotydowych typu CNG	30
2.6. Struktura krystaliczna powtórzeń CUG	33
2.7. Struktury spinkowe RNA stabilne w warunkach krystalizacji	34
3. Wyniki	37
3.1. Struktury RNA zawierające powtórzenia CUG	37
3.1.1. Krystalizacja oligomeru GCUGCUGC	37
3.1.2. Rejestracja i przetwarzanie danych	37
3.1.3. Rozwiązanie struktury (GCUGCUGC) ₂	37
3.1.4. Udokładnianie struktury (GCUGCUGC) ₂	39
3.1.5. Struktura dupleksu [(CUG) ₆] ₂	40
3.1.6. Konformacja dupleksów RNA	42
3.1.7. Oddziaływania pomiędzy parami zasad	43
3.1.8. Hydratacja RNA i oddziaływania z ligandami	45
3.1.9. Oddziaływania warstwowe	47
3.1.10. Powierzchnia potencjału elektrostatycznego	48
3.2. Struktury RNA zawierające powtórzenia CAG	49
3.2.1. Krystalizacja dupleksu (GGCAGCAGCC) ₂	50
3.2.2. Rejestracja i przetwarzanie danych	50

3.2.3.	Rozwiązanie struktur zawierających powtórzenia CAG	51
3.2.4.	Udokładnianie struktur RNA zawierających powtórzenia CAG	52
3.2.5.	Upakowanie cząsteczek w sieci krystalicznej	53
3.2.6.	Konformacja helisy RNA oraz niekanoniczne pary A-A	54
3.2.7.	Hydratacja i oddziaływania z ligandami	56
3.2.8.	Oddziaływania warstwowe	58
3.2.9.	Powierzchnia i ukształtowanie potencjału elektrostatycznego	59
3.3.	Struktury RNA zawierające powtórzenia CGG	59
3.3.1.	Krystalizacja oligomerów zawierających powtórzenia CGG	60
3.3.2.	Pomiary dyfrakcyjne i przetwarzanie danych	60
3.3.3.	Rozwiązanie struktur zawierających powtórzenia CGG	61
3.3.4.	Udokładnianie struktur zawierających powtórzenia CGG	62
3.3.5.	Ogólny opis struktur zawierających powtórzenia CGG	63
3.3.6.	Pary G-G i ich wpływ na strukturę RNA	63
3.3.7.	Oddziaływania warstwowe	65
3.3.8.	Efekt bromowania i dystrybucja konformacji par G-G	70
3.3.9.	Oddziaływania z roztworem i hydratacja	70
3.3.10.	Powierzchnia potencjału elektrostatycznego	72
3.4.	Struktury RNA zawierające powtórzenia CCG	73
3.4.1.	Krystalizacja RNA zawierającego powtórzenia CCG	74
3.4.2.	Pomiar danych dyfrakcyjnych i ich przetwarzanie	74
3.4.3.	Rozwiązanie struktur zawierających powtórzenia CCG	74
3.4.4.	Udokładnianie struktur zawierających powtórzenia CCG	75
3.4.5.	Upakowanie cząsteczek w sieci krystalicznej	76
3.4.6.	Konformacja RNA zawierające powtórzenia CCG	76
3.4.7.	Niestandardowe pary C-C	78
3.4.8.	Uporządkowane pary 8C	81
3.4.9.	Modyfikowane reszty 4G z serii LNA	81
3.4.10.	Oddziaływania z rozpuszczalnikiem	82
3.4.11.	Oddziaływania warstwowe	83
3.4.12.	Kształt powierzchni i rozkład potencjału elektrostatycznego	85
3.5.	Opracowanie metody stabilizacji struktur spinkowych RNA w warunkach krystalizacji	86
3.5.1.	Zaprojektowanie sekwencji oligomerów RNA tworzących strukturę spinki	86
3.5.2.	Testowanie i optymalizacja reakcji cyrkularyzacji oligomeru	87
3.5.3.	Opracowanie metody oczyszczania zcyrkularyzowanego oligomeru.	90
3.5.4.	Otrzymywanie ligazy RNA z faga T4	91

3.5.5.	Procedura otrzymywania stabilnej formy spinki RNA	92
3.5.6.	Otrzymanie zcyrkularyzowanego oligomeru metodą klik	92
4.	Dyskusja	95
4.1.	Niestandardowe pary N-N	95
4.2.	Akomodacja par N-N w strukturze A-RNA	97
4.3.	Oddziaływania par N-N z cząsteczkami rozpuszczalnika	98
4.4.	Wpływ niekanonicznych par na globalną strukturę RNA	100
4.5.	Efekt wprowadzonych modyfikacji	102
4.6.	Dystrybucja konformacji par N-N w dupleksach	103
4.7.	Badania strukturalne a badania termodynamiczne	104
4.8.	Aspekt biologiczny	105
4.9.	Metodyka otrzymywania stabilnych struktur spinkowych RNA	106
5.	Materiały i Metody	108
5.1.	Materiały	108
5.1.1.	Odczynniki chemiczne i biochemiczne	108
5.1.2.	Enzymy	108
5.1.3.	Stosowane bufony, roztwory i pożywki	108
5.2.	Metody	111
5.2.1.	Elektroforeza kwasów nukleinowych w żelu poliakryloamidowym w warunkach denaturujących	111
5.2.2.	Elucja kwasów nukleinowych	111
5.2.3.	Strącanie kwasów nukleinowych	112
5.2.4.	Barwienie żeli poliakryloamidowych z mocznikiem za pomocą roztworu stains-all	112
5.2.5.	Elektroforeza białek w żelu poliakryloamidowym w obecności SDS (SDS-PAGE)	112
5.2.6.	Ekstrakcja fenolem	112
5.2.7.	Odsalanie na kolumnkach NAP-25	113
5.2.8.	Defosforylacja	113
5.2.9.	Znakowanie RNA na końcu 5' radioaktywnym izotopem ³² P	113
5.2.10.	Hydrolyzacja formamidowa	113
5.2.11.	Ograniczone trawienie RNA rybonukleazą T1 w warunkach semidenaturujących	113
5.2.12.	Reakcja cyrkularyzacji RNA z użyciem ligazy DNA	114
5.2.13.	Ligacja chemiczna	114
5.2.14.	Reakcja klik	114
5.2.15.	Trawienie enzymem restrykcyjnym HpaII	114
5.2.16.	Oczyszczanie oligomerów RNA i DNA	115

5.2.17. Otrzymywanie ligazy RNA z faga T4	116
5.2.18. Oznaczanie stężenia białka	117
5.2.19. Test aktywności T4 RNA ligazy	117
5.2.20. Otrzymywanie zcyrkularyzowanych oligomerów RNA w skali preparatywnej	117
5.2.21. Przygotowanie oligomerów RNA do krystalizacji	117
5.2.22. Krystalizacja	118
5.2.23. Pomiary dyfrakcyjne	119
5.2.24. Przetwarzanie i skalowanie danych dyfrakcyjnych	119
5.2.25. Obliczanie i analiza statystyczna amplitud czynników struktury	120
5.2.26. Rozwiązywanie struktury – podstawienie cząsteczkowe	121
5.2.27. Natywna mapa Pattersona	122
5.2.28. Udokładnianie modeli RNA	123
5.2.29. Pozostałe programy użyte w pracy	125
Literatura	127

Spis stosowanych skrótów

8-BrG – 8-bromoguanozyna

8-BrA – 8-bromoadenozyna

A – reszta adenozyny

C – reszta cytydyny

CCD — detektor półprzewodnikowy (ang. *charge coupled device*)

c.p.m. — zliczenia rozpadów promieniotwórczych na minutę

DM — dystrofia miotoniczna

DNA — kwas deoksyrybonukleinowy

DTT — ditiotreitól

DRPLA — zanik zębatoczerwienny pallidoniskowzgórzowy

FRAXE — bezsyndromowe opóźnienie umysłowe powiązane z chromosomem X (ang. *non-syndromic X-linked mental retardation*)

FXTAS — zespół drżenia i ataksji związany z łamliwym chromosomem X (ang. *fragile X-associated tremor ataxia syndrome*)

FXS — łamliwość chromosomu X

G – reszta guanozyny

HD — choroba Huntingtona

HDL2 — choroba podobna do płasawicy Huntingtona (ang. *Huntington disease like 2*)

IPTG — izopropyló β -D-tiogalaktopiranozyd

LNA — reszty nukleozydowe o zamkniętej konformacji pierścienia cukrowego (ang. *locked nucleic acid*)

MALDI-TOF — jonizacja przez resorpcję laserową w stałej matrycy – czas przelotu (ang. *matrix-assisted laser desorption ionization – time of flight*)

NMR — spektroskopia magnetycznego rezonansu jądrowego (ang. *nuclear magnetic resonance spectroscopy*)

ORF — otwarta ramka odczytu

rmsd — średnie kwadratowe odchylenie (ang. *root-mean-square deviation*)

RNA — kwas rybonukleinowy

r.p.m. – obroty na minutę

SBMA — zależny od chromosomu X rdzeniowo-opuszkowy zanik rdzeniowy (choroba Kennedy'ego)

SCA — ataksja mózdkowo-rdzeniowa

SDS — dodecylosiarczan sodu

SDS-PAGE — elektroforeza w żelu poliakryloamidowym w warunkach denaturujących w obecności SDS

TNR — powtórzenia trójnukleotydowe

TREDS — choroby związane z ekspansją trójnukleotydowych powtórzeń (ang. *trinucleotided repeat expansion diseases*)

U – reszta urydyny

UTR — region nieulegający translacji (ang. *untranslated region*)

1. Cel pracy doktorskiej

Nadrzędnym celem mojej pracy doktorskiej było określenie struktur krystalicznych oligomerów RNA o sekwencji składającej się z powtórzeń CNG (N oznacza jeden z czterech nukleotydów). Sekwencje takie są związane z patogenezą chorób neurodegeneracyjnych o wspólnej nazwie TREDs (choroby związane z ekspansją trójnukleotydowych powtórzeń) (ang. *trinucleotide repeat expansion diseases*). Wyróżnia się ponad 20 takich chorób, które są nieuleczalne i prowadzą do śmierci. Jeden z proponowanych patomechanizmów TREDs zakłada, że pre-mRNA zawierające wydłużone ciągi powtórzeń CNG posiadają właściwości toksyczne. Polegają one na nadmiernym wiązaniu czynników białkowych niezbędnych dla prawidłowego funkcjonowania komórki.

Moim zadaniem było poznanie struktury przestrzennej czterech rodzajów powtórzeń: CAG, CCG, CGG i CUG. Badane przeze mnie cząsteczki RNA zawierały 2–3 powtórzenia CNG. Podyktowane to było wydajną syntezą chemiczną krótkich cząsteczek RNA oraz większym prawdopodobieństwem uzyskania kryształów. Oligomery te tworzą dupleksy co stanowi fragment trzonu struktury typu spinki tworzonej przez ciągi powtórzeń CNG w pre-mRNA. Na podstawie otrzymanych przeze mnie struktur krystalicznych chciałam określić:

- sposób parowania się niestandardowych par typu N-N,
- wpływ par typu N-N na globalną strukturę RNA,
- oddziaływania pomiędzy RNA a cząsteczkami rozpuszczalnika,
- obliczyć powierzchnię potencjału elektrostatycznego cząsteczek RNA.

Powyższe założenia miały posłużyć do sporządzenia profilu strukturalnego powtórzeń CNG dla badań biochemicznych oraz do wyjaśnienia przyczyny patogennych właściwości tych RNA.

Moim drugim celem było opracowanie metody stabilizacji struktur spinkowych RNA w warunkach krystalizacji, czyli wysokiego stężenia oligomeru oraz soli. W roztworze istnieją dwie formy oligomeru: spinka i dupleks, które mogą przechodzić jedna w drugą. To, która forma dominuje zależy od warunków w jakich cząsteczki RNA się znajdują. Jeżeli jest to wysokie stężenie soli lub wysokie stężenie oligomeru preferencyjnie tworzy się dupleks. W efekcie powoduje to, że wykrystalizowanie struktury spinki RNA jest bardzo trudne.

Stabilne struktury spinkowe zamierzałam otrzymać z wykorzystaniem specjalnego łącznika, spinającego końce trzonu spinki za pomocą wiązania kowalencyjnego. W związku z tym opracowywanie tej metodologii polegało na:

- przetestowaniu łączników, które mogą być użyte do tego celu,
- uzyskaniu najwydajniejszego sposobu kowalencyjnego zamknięcia cząsteczki RNA,
- opracowaniu skutecznej i wydajnej metody oczyszczania uzyskanego produktu.

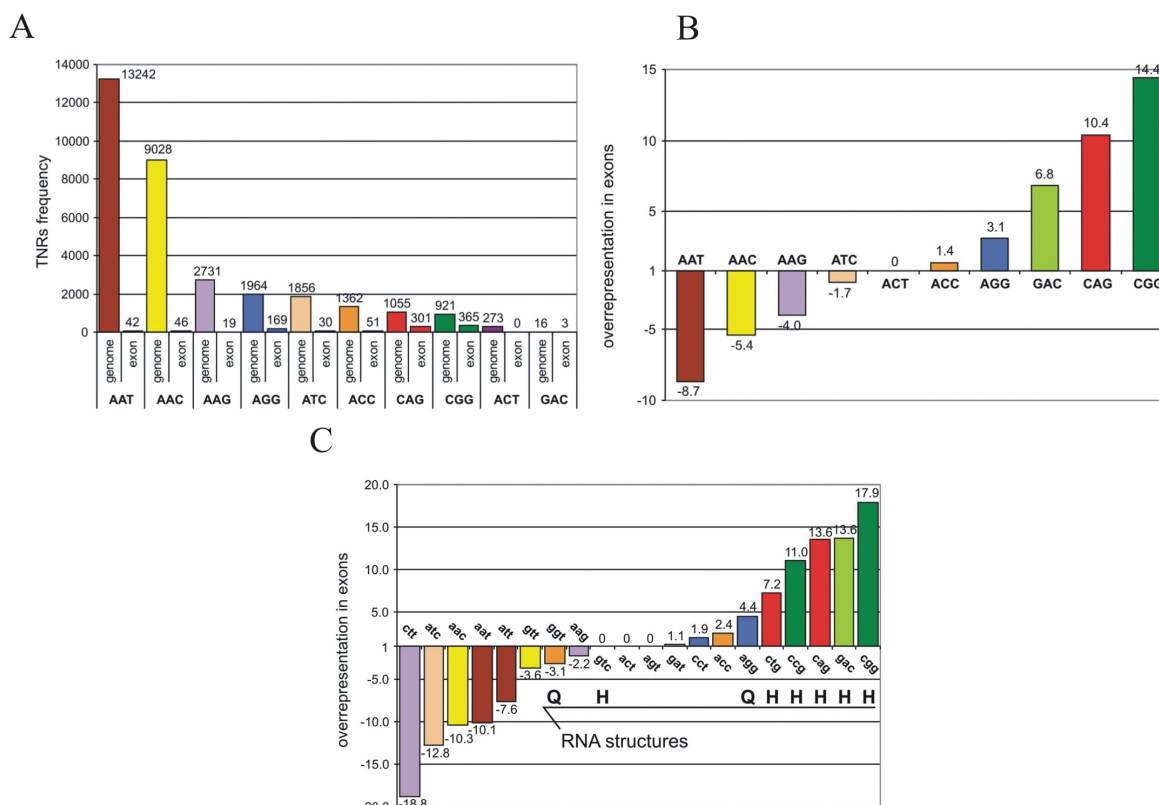
2. Wstęp

2.1. Występowanie powtórzeń trójnukleotydydowych w genomie

Sekwencje mikrosatelitarne są krótkimi powtarzającymi się motywami DNA o długości od 1 do 6 par zasad, obecnymi w genomach wszystkich organizmów (1, 2). Cechują się wysoką zmiennością osobniczą dotyczącą zarówno składu nukleotydydowego jak i długości sekwencji. To zróżnicowanie jest tak duże, że nie ma dwóch osób posiadających identyczną kombinację alleli mikrosatelitarnych. Istotnym faktem jest również to, że układ tych powtórzeń jest dziedziczony częściowo od matki a częściowo od ojca. Z tego powodu są one dobrymi markerami wykorzystywanymi m.in. w mapowaniu genetycznym czy genetyce populacji (2–4).

Specyficzną klasą sekwencji mikrosatelitarnych są powtórzenia trójnukleotydydowe (TNR) (ang. *trinucleotide repeats*). W ludzkim genomie znajduje się ich około 32000, z czego 1030 występuje w egzonach (egzon rozumiany jest jako część 5', 3' UTR (ang. *untranslated region*) lub otwartej ramki odczytu—ORF) (7). 60% TNR obecnych w egzonach zlokalizowanych jest w otwartych ramkach odczytu a pozostałe 40% w regionach niekodujących 5' i 3' (UTR). Powtórzenia trójnukleotydydowe są tolerowane w sekwencjach kodujących bardziej niż powtórzenia dwu- lub czteronukleotydydowe, ponieważ ich zmienna liczba nie zmienia ramki odczytu (8). Niemniej, nie wszystkie możliwe triplety są równocennie reprezentowane w egzonach (7). Najczęściej występują powtórzenia typu CNG (gdzie N oznacza jeden z czterech nukleotydów): CGG – 365 przypadków (ang. *occurrences*) i CAG – 301 przypadków, przy czym CGG i CAG reprezentują również drugą orientację nici DNA, czyli odpowiednio powtórzenia CCG i CTG (Rysunek 1). Przy uwzględnieniu orientacji egzonu, czyli tego jakie powtórzenia ulegają ekspresji, obliczony czynnik reprezentacji dla każdego powtórzenia CNG wynosi: CGG 17,9 ×, CCG 11,0 ×, CAG 10,4 × i CUG 7,2 × (Rysunek 1) (im wartość czynnika reprezentacji jest większa tym częściej powtórzenie występuje w sekwencji egzonów). Dodatkowo wartości oznaczają, że powtórzenia CNG są nadreprezentowane w egzonach w stosunku do ich występowania w genomie, co sugeruje, że ich obecność w sekwencjach egzonów nie jest przypadkowa. Oznacza to, że musiały być one poddane pozytywnej selekcji podczas ewolucji. Co więcej wydaje się, że lokalizacja danego typu powtórzenia w mRNA (regiony niekodujące lub otwarte ramki odczytu) jest również określona. CAG i CUG występują najczęściej w ORF (~80%) natomiast CCG i CGG w regionach 5'UTR (powyżej 50%). Sugeruje to, że być może są one funkcjonalnymi elementami genetycznymi, których polimorfizm może manifestować się

różnymi obserwowanymi fenotypami. Teza ta wymaga jednak dalszego potwierdzenia, gdyż zakres informacji na ten temat pozostaje nadal stosunkowo ubogi.



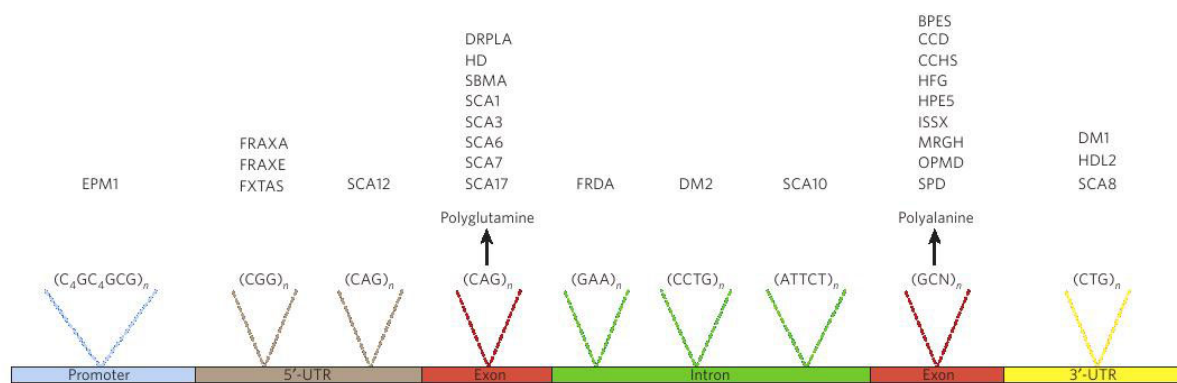
Rysunek 1. Częstość występowania powtórzeń TNR w genomie ludzkim (7). W panelu A zestawiono całkowitą liczbę danego powtórzenia, zidentyfikowaną w genomie oraz w egzonach bez uwzględnienia orientacji nici DNA. Rozkład czynników reprezentacji w egzonach: pozytywne wartości oznaczają nadreprezentację natomiast negatywne niedoreprezentowanie danego powtórzenia (B). Wskaźnik reprezentacji obliczony przy uwzględnieniu orientacji nici DNA dla poszczególnych powtórzeń TNR (C). Litery Q, H odnoszą się od tego rodzaju TNR, dla których RNA tworzą struktury spinki (H) lub kwadrupleksu (Q).

Mimo ograniczonej ilości badań nad funkcją powtórzeń CNG, są one obiektem licznych eksperymentów z innego powodu. Mianowicie, sekwencje te mogą ulegać patogennej ekspansji (nadmierne zwielokrotnienie liczby powtórzeń) prowadząc do rozwoju ludzkich nieuleczalnych chorób neurodegeneracyjnych znanych pod wspólną nazwą TREDs (ang. *trinucleotide repeat expansion diseases*) (9).

2.2. Patogenne powtórzenia typu CNG związane z chorobami TREDs

Jak sama nazwa wskazuje, większość chorób TREDs wywoływana jest ekspansją powtórzeń trójnukleotydowych CNG i GCN. Oprócz nich do tej samej grupy, ze względu na przyczynę choroby, zalicza się również powtórzenia cztero-, pięcio- a nawet dziesięcionukleotydowe (10).

Każda z tych chorób związana jest z jednym rodzajem powtórzenia znajdującym się w konkretnym genie. Obecnie znanych jest ponad 20 różnych chorób neurodegeneracyjnych. Wszystkie powtórzenia zlokalizowane są w obrębie sekwencji genów a zatem mogą występować w regionach promotorowych, 5' i 3' UTR, skończywszy na intronach oraz egzonach (Rysunek 2). W związku z tym, że moje badania dotyczą powtórzeń CNG dalszy opis literatury zawężony będzie do tego rodzaju TNR.



Rysunek 2. Lokalizacja patogennych powtórzeń w genie oraz ich powiązanie z daną chorobą TREDs (10). BPES, zespół zwężenia szpary powiekowej, opadania powieki i odwróconej zmarszczki nakątnej; CCD, dysplazja obojczykowo-czaszkowa; CCHS, pierwotna hipowentylacja pęcherzykowa — kłątwa Ondyny; DM, dystrofia miotoniczna; DRPLA, zanik zębato-czerwienno-pallidoniskowgórzkowy; EPM1, postępująca padaczka miokloniczna typu 1; FRAXA, łamliwość chromosomu X; FRAXE, bezsyndromowe opóźnienie umysłowe powiązane z chromosomem X; FXTAS, zespół drżenia i ataksji związany z łamliwym chromosomem X; HD, choroba Huntingtona; HDL2, choroba podobna do płasawicy Huntingtona; HFG, zespół dłoń-stopa-narządy płciowe; HPE5, holoprocencefalia 5; ISSX, niemowlęce drgawki sprzężone z chromosomem X; MRGH, upośledzenie umysłowe z izolowanym niedoborem hormonu wzrostu; OPMD, dystrofia oczno-gardłowa; SBMA, zależny od chromosomu X rdzeniowo-opuszkowy zanik rdzeniowy (choroba Kennedy’ego); SCA, ataksja mózdkowo-rdzeniowa; SPD, synpolidaktylia.

U zdrowych osobników trakty CNG są krótkie, stabilne a ich długość nie ulega zmianie. Dodatkowo, powtórzenia często rozdzielone są sekwencjami „przerywnikowymi” (ang. *interruptions*) (Tabela 1) (11), które prawdopodobnie pełnią funkcję stabilizującą (12–14). W rodzinach dotkniętych TREDs, powtórzenia trójnukleotydowe są dłuższe i niestabilne zarówno w komórkach somatycznych jak i rozrodczych (15). Oznacza to, że CNG mają tendencję do ulegania wydłużaniu podczas życia chorego. Długość traktów zmienia się dynamicznie, co objawia się różną liczbą powtórzeń zarówno pomiędzy chorymi osobnikami jak i w obrębie tkanek tego samego organizmu. W efekcie taka osoba z czasem może stać się swego rodzaju mozaiką komórek zawierających różną długość powtórzeń CNG. Niestabilność powtórzeń w komórkach płciowych wiąże się z dziedziczeniem wydłużonych ciągów. Zauważono, że im dłuższe trakty są przekazywane, tym symptomy chorobotwórcze stają się cięższe i ujawniają się we wcześniejszym wieku. Fenomen ten nazwano antycypacją.

Tabela 1. Zestawienie chorób związanych z powtórzeniami CNG (11).

Choroba	Gen	Rodzaj powtórzenia	Sekwencje przerywnikowe	Normalna liczba powtórzeń	Premutacja	Patogenna liczba powtórzeń
SBMA	AR	CAG	-	9–36	brak	40–55
HD	HTT	CAG	-	10–34	29–35	> 35
DRPLA	ATN1	CAG	-	7–25	brak	49–88
SCA1	ATXN1	CAG	CAT	6–39	brak	39–81
SCA2	ATXN2	CAG	CAA	13–33	brak	> 34
SCA3	ATXN3	CAG	CGG, GCC	13–44	brak	> 55
SCA6	CACNA1A	CAG	CAG,CAA	4–18	brak	20–29
SCA7	ATXN7	CAG	-	4–35	brak	37–306
SCA17	TBP	CAG	?	25–42	brak	47–63
DM1	DMPK	CUG	CGG,CTC	5–37	34–90	> 50–3000
SCA8	ATXN8	CUG	CCG, TTG, CTT, CTA, CCA, GTC	2 > 130	45–109	> 110
SCA12	PPP2R2B	CAG	-	7–45	brak	55–78
HDL2	JPH3	CUG	?	6–27	brak	51–57
DM2	ZNF9	CCUG	TCTG, GCTG	104–176	brak	75–11000
FXS	FMR1	CGG	AGG	5–54	55–200	230–2000
FXTAS	FMR1	CGG	AGG	5–54	brak	55–200
FRAXE	FMR2	CCG	?	4–39	(31–61)	200–900

Mechanizm przejścia ze stabilnej do niestabilnej liczby powtórzeń nie został jeszcze dokładnie poznany. Sugeruje się, że może mieć to związek z utrzymaniem sekwencji przerywnikowych. Z kolei, gdy są już obecne dynamiczne trakty, to ich wydłużanie może odbywać się podczas replikacji, naprawy DNA lub rekombinacji (10, 15). Co ciekawe, podczas tych procesów można zaobserwować również zjawisko kontrakcji czyli skrócenia liczby powtórzeń.

TREDs zostały odkryte około 20 lat temu (16). Wówczas po raz pierwszy powiązano wielokrotnioną liczbę powtórzeń mikrosatelitarnych z objawami chorobotwórczymi. Są to ciężkie, nieuleczalne choroby, ujawniające się zazwyczaj w wieku dojrzałym, kiedy powtórzenia osiągną odpowiedni próg liczbowy (Tabela 1). Obraz kliniczny jest szeroki i charakterystyczny dla danej choroby. Przykładowo w dystrofii miotonicznej typu 1 (DM1) występują nie tylko zaburzenia motoryczno-ruchowe lecz również m.in. zaćma, zaburzenia pracy serca czy układu pokarmowego. Początkowo objawy są niezauważalne lecz stopniowo pogłębiają się prowadząc w rezultacie do śmierci (10, 17).

Największa liczba chorób TREDs związana jest z ekspansją powtórzeń CAG (9). Należą do nich: choroba Huntingtona (HD), zależny od chromosomu X rdzeniowo-opuszkowy zanik rdzeniowy (choroba Kennedy'ego) (SBMA), zanik zębatoczerwienny pallidoniskowzgórzowy

(DRPLA) oraz kilka ataksji mózdkowo-rdzeniowych różnego typu (SCA). Powtórzenia CAG, oprócz jednego wyjątku, występują w regionach kodujących specyficznych genów (Rysunek 2). Oznacza to, że długie trakty przepisywane są na sekwencję białka. W tym przypadku jest to glutamina. Powstające białko posiada wydłużony trakt poliglutaminowy i z tego powodu choroby te nazywane są również chorobami poliglutaminowymi (ang. *poly-Q* lub *polyglutamine diseases*). Wyjątkiem, w którym powtórzenia CAG znajdują się w regionie 5'UTR, jest gen *PPP2R2B* kodujący podjednostkę regulatorową fosfatazy 2A w mózgu. Związany jest z ataksją mózdkowo-rdzeniową typu 12 (SCA12).

Powtórzenia CGG związane są z genem *FMR1* (ang. *fragile X mental retardation*) leżącym na chromosomie X. Znajdują się w regionie niekodującym 5' mRNA. Ich lokalizacja wydaje się być spodziewana ze względu na fakt, iż sekwencje bogate w CpG występują często w regionach promotorowych. Na tej podstawie sugeruje się, że powtórzenia CGG uczestniczą w regulacji ekspresji genów (7, 16). Interesujący jest fakt, że trakty te związane są z kilkoma różnymi fenotypami (18). Typowo, liczba trójek CGG w genie *FMR1* wynosi 5–54 (18, 19). Gdy liczba powtórzeń wzrasta do 45–54, wówczas zaliczają się one do podklasy nazywanej „szarą strefą” (ang. *grey zone*), gdzie istnieje większe prawdopodobieństwo dalszej ekspansji (18, 20). Trakty CGG o długości 55–200 powtórzeń są premutacjami, które u starszych mężczyzn powodują progresywną chorobę neurodegeneracyjną FXTAS (ang. *fragile X-associated tremor ataxia syndrome*) (21, 22). U kobiet będących nosicielkami premutacji może dojść do rozwoju niewydolności jajników (23). Więcej niż 200 powtórzeń CGG jest pełną mutacją, która wywołuje łamliwość chromosomu X (FXS) (24). Jest to najczęstszy dziedziczny syndrom opóźnienia umysłowego u mężczyzn.

Powtórzenia CCG występują w genie *FMR2*, który znajduje się około 6000 par zasad od genu *FMR1* i powiązane są z chorobą FRAXE czyli bezsyndromowym opóźnieniem umysłowym powiązanim z chromosomem X (ang. *non-syndromic X-linked mental retardation*) (9). Syndrom ten ujawnia się gdy osobnik posiada więcej niż 200 powtórzeń CCG (pełna mutacja) (25). Niektóre objawy kliniczne są wspólne z łamliwością chromosomu X (FXS), np. niewydolność jajników, czy problemy z mówieniem. W obu przypadkach chorzy wykazują opóźnienie umysłowe aczkolwiek dla syndromu FRAXE przyjmuje to łagodniejszą formę (9).

Jak dotąd nie stwierdzono fenotypu neurodegeneracyjnego dla CCG o długości premutacyjnej (60–200), ale powtórzenia te wprowadzone do genomu muszki owocowej ujawniły właściwości toksyczne (26). Zauważono także korelację pomiędzy pośrednią (ang. *intermediate*) długością powtórzeń CCG (około 40) w allelach kilku badanych populacji a chorobą Parkinsona lub przynajmniej niektórymi jej objawami klinicznymi (27, 28). Podobnie, na podstawie badań przeprowadzonych na dużej grupie chłopców w wieku 5–18 lat, powiązano ich niewytłumaczalne problemy z nauką z długością pośrednią powtórzeń w allelach związanych z syndromami FRAXE

i FRAXA (29, 30). Na podstawie powyższych obserwacji nie wyklucza się istnienia, analogicznie do powtórzeń CGG, neurodegeneracyjnych właściwości CCG o długości premutacyjnej.

Powtórzenia CTG związane są z trzema różnymi chorobami: chorobą podobną do płasawicy Huntingtona (HDL2) (ang. *huntington disease like 2*), ataksją mózdkowo-rdzeniową typu 8 (SCA8) oraz dystrofią miotoniczną typu 1 (DM1) (17). We wszystkich przypadkach powtórzenia zlokalizowane są w regionach 3' UTR, co może wpływać na efektywność i dokładność składania pre-mRNA (31), transport mRNA (32) i regulację translacji (33, 34).

Powiązanie ekspansji CTG zlokalizowanych w genie kinazy białkowej DMPK z dystrofią miotoniczną typu 1 zostało odkryte w 1992 roku (35–37). Jest to najlepiej poznana choroba pod względem patomechanizmu spośród wszystkich TREDs. U zdrowych osobników liczba powtórzeń wynosi 5–37, natomiast zmutowana forma zawiera 50–3000 trójek (36). Obok dystrofii miotonicznej typu 1 występuje również typ 2, który wywoływany jest przez powtórzenia CCTG (38). Obie dystrofie, pomimo innego rodzaju powtórzenia oraz innej lokalizacji genowej, posiadają bardzo podobne fenotypy chorobowe (17).

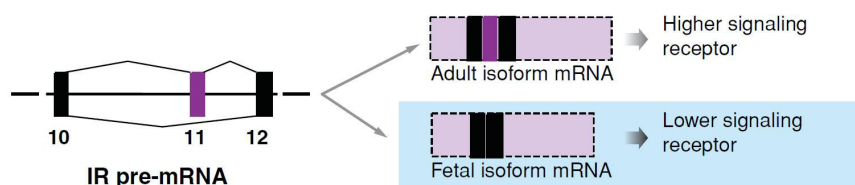
2.3. Rola powtórzeń CNG znajdujących się w mRNA niektórych genów w patomechanizmie chorób TREDs

W literaturze opisane zostały trzy mechanizmy rozwoju TREDs: utrata lub nabycie nowej funkcji białka (ang. *loss or gain of function*), utrata funkcji genu zawierającego powtórzenia (ang. *loss of function*), nabycie nowej funkcji przez mRNA zawierające długie powtórzenia CNG (ang. *mRNA gain of function*) (9). Propozycja istnienia pierwszego wymienionego patomechanizmu wydaje się być oczywista, jeżeli wydłużone trakty występują w sekwencjach kodujących białko. Jak zatem wytłumaczyć wpływ powtórzeń na rozwój choroby, gdy znajdują się one w regionach niekodujących? Odpowiedzią na to pytanie było zaproponowanie dwóch pozostałych mechanizmów. Utrata funkcji genu polega na tym, iż długie trakty zaburzają strukturę DNA uniemożliwiając ekspresję sąsiadujących genów. Z kolei nabycie nowych funkcji przez mRNA oznacza nadmierne wiązanie istotnych czynników białkowych niezbędnych do poprawnego funkcjonowania komórki. Istnienie tego mechanizmu poparte jest dowodami lecz mimo to nadal prowadzone są liczne badania w tym kierunku, ponieważ w przypadku niektórych chorób patogeneza wydaje się być bardziej złożona. Nie wyklucza się również współistnienia różnych patomechanizmów.

2.3.1. Powtórzenia CUG

Dystrofia miotoniczna (DM) (forma dystrofii mięśniowej) jest chorobą, w której pre-mRNA posiadający wydłużone trakty CUG jest bezpośrednio uwikłany w patogenezę działając *in trans* na inne specyficzne pre-mRNA, z którym nie jest powiązany (39). Początkowo uważano jednak, iż patogenezę tej choroby związana jest z haploinsuficjencją (ang. *haploinsufficiency*) (jedna kopia genu nie wystarcza do zapewnienia odpowiedniej ilości białka) ze względu na metylację histonów i wiązanie białka HP1 (białko heterochromatyny 1), które indukuje powstawanie heterochromatyny (40). W efekcie sąsiadujące geny *SIX5* i *DMPK*, które zawierają patogeniczne powtórzenia CUG, ulegają zmniejszonej ekspresji (41, 42). Obecnie wiadomo, że efekt ten odgrywa niewielką rolę w patogenezie a obniżona ilość białka DMPK jest wynikiem zaburzeń powstałych na etapach potranskrypcyjnych, w wyniku których mRNA zawierające długie powtórzenia CUG zatrzymywane są w jądrze komórkowym, gdzie tworzą złoża (ang. *foci*) (43). Najważniejszym dowodem świadczącym o patogenicznych właściwościach RNA zawierających zwiokrotnione powtórzenia CUG było odtworzenie obrazu klinicznego u myszy i muszki owocowej, w wyniku ekspresji wydłużonych traktów CUG znajdujących się w naturalnym kontekście genu *DMPK* lub wprowadzonych w regiony 3' UTR niepowiązanych transkryptów (39, 44, 45).

W toku dalszych badań odkryto, że patogeniczne RNA powodują zmiany w profilu składania transkryptów (ang. *spliceopathy*) (46). Odróżnia się to od zaburzeń składania obserwowanych w innych chorobach genetycznych tym, że proces zachodzi poprawnie. Nie następują zmiany np. poprzez wybór krytycznych miejsc składania, które powodują powstanie nienaturalnego białka (17). U chorych powstają naturalne warianty mRNA, ale nieodpowiednie dla danej tkanki i etapu rozwojowego. Innymi słowy w dojrzałych włóknach mięśniowych, w sercu czy mózgu ulegają ekspresji te izoformy białek, które są charakterystyczne dla zarodka czy płodu. Lista genów dotkniętych zmianą profilu składania zamyka się w około 20 pozycjach, ale ich całkowita pula nie została poznana i nie wyklucza się, że może dotyczyć dziesiątek czy nawet tysięcy genów. Wiadomo, że są to geny, które poddawane są alternatywnemu składaniu i uczestniczą w tym samym szlaku regulacji rozwojowej (46). Zgodnie z powyższym, obserwowane symptomy chorobowe, ujawniające się głównie w mięśniach szkieletowych, powiązane są z odkrytymi zmianami profilu procesu składania konkretnych mRNA. Przykładowo, oporność insulinowa jest wynikiem produkcji wariantu receptora insulinowego (IR) o obniżonej pojemności sygnałnej (Rysunek 3). Jest to forma płodowa białka, która powstaje w wyniku alternatywnego składania z ominięciem 11 egzonu (ang. *exon skipping*). Insulina, która rozpoznawana jest przez IR, ułatwia wnikanie glukozy do komórki. Oznacza to, że jeżeli komórka posiada receptor formy płodowej to wnika do niej mniej glukozy. Tkanka mięśniowa ma bardzo duże zapotrzebowanie na glukozę, większe niż np. komórki wątroby, dlatego komórki mięśni posiadające nieodpowiedni wariant receptora IR, najbardziej odczuwają niedobór glukozy.

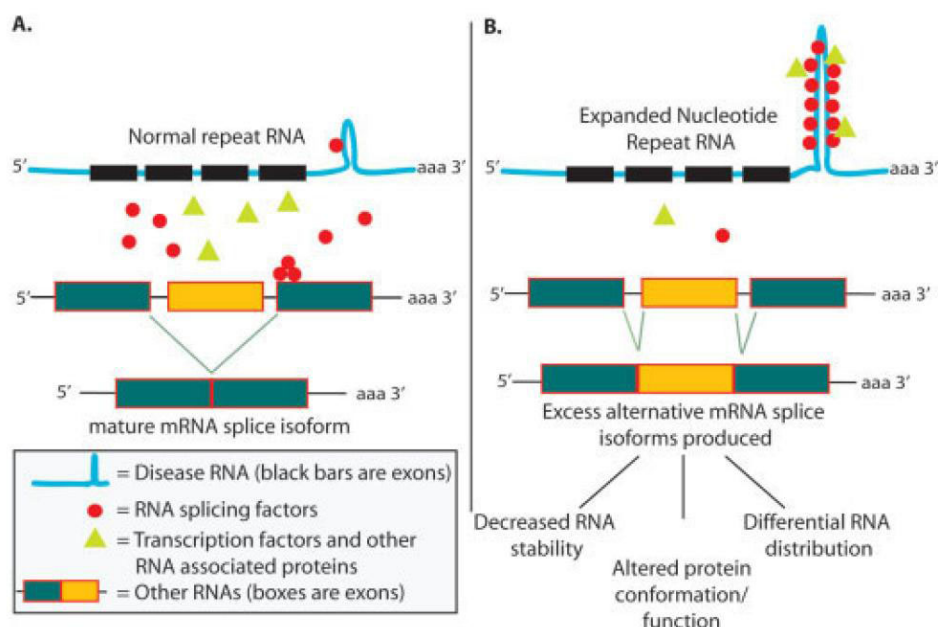


Rysunek 3. Alternatywne składanie pre-mRNA receptora insulinowego podczas rozwoju organizmu (17). mRNA osobnika dorosłego (ang. *adult isoform*) posiada zachowany egzon 11. Powoduje to, że receptor insulinowy wykazuje większą pojemność sygnałną niż forma płodowa, której mRNA (ang. *fetal isoform*) nie zawiera egzonu 11.

Proponowanym wytłumaczeniem patogenicznego działania RNA *in trans* na niepowiązane z nim transkrypty, jest nadmierne wiązanie białka MBNL1 (ang. *muscleblind-like 1*) (47). MBNL1 jest regulatorem alternatywnego składania transkryptów (48). Zatem, obniżenie jego poziomu w komórce doprowadza do zaburzeń w dojrzewaniu pre-mRNA. Białko to zlokalizowano w inkluzjach jądrowych komórek mózgu, mięśni szkieletowych i serca, zawierających wydłużone trakty CUG (49, 50). MBNL1 wraz z innym regulatorem CUG-BP1 (ang. *CUG binding protein*) zaangażowane są w kontrolę składania wielu pre-mRNA (51). Ich działanie względem siebie jest antagonistyczne. Gdy jedno z nich jest regulatorem pozytywnym to drugie jest regulatorem negatywnym. To jaką rolę pełni dane białko wydaje się być specyficzne dla danego transkryptu oraz miejsca składania. Przypuszcza się, że równowaga pomiędzy tymi dwoma białkami jest bardzo istotna, ponieważ prawdopodobnie wraz z kilkoma innymi regulatorami składania odpowiadają za tkankowo-specyficzny wzór tego procesu w komórce (52).

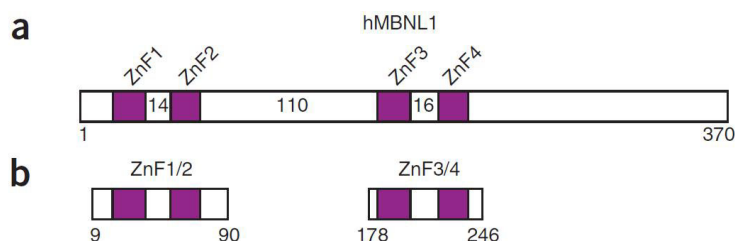
W dystrofii miotonicznej równowaga pomiędzy MBNL1 oraz CUG-BP1 jest zaburzona. Hipotetycznie, obserwowane zaburzenia składania mogą zostać wywołane przez obniżenie poziomu MBNL1 lub/i przez zwiększenie ilości CUG-BP1. Pokazano, że zarówno u myszy z usuniętym fragmentem genu *MBNL1* (ang. *knock-out*) (53) jak i u myszy z indukowaną nadekspresją białka CUG-BP1 (31, 54) odtworzony został kliniczny obraz DM. Jednak dalsze badania wskazują, że nadmierne wiązanie MBNL1 jest pierwszym i kluczowym czynnikiem powodującym rozwój dystrofii miotonicznej (Rysunek 4). W modelu komórkowym mioblastów defekty składania pre-mRNA charakterystyczne dla tej choroby nie zostały wyciszone przy obniżeniu ilości CUG-BP1 (55). Co więcej, nie jest do końca jasne czy u myszy mającej patogenne trakty CUG lub brak MBNL1 następuje zwiększenie ekspresji CUG-BP1 (54, 56). Niewykluczone jest również to, że na dalszych etapach rozwoju choroby mogą występować inne zaburzenia, np. kontroli potranskrypcyjnej.

MBNL1 jest białkiem zawierającym 386 reszt aminokwasowych (58). Posiada cztery domeny palca cynkowego typu CCCH (gdzie C oznacza resztę cysteiny, H resztę histydy), które ułożone są tandemowo parami. Pierwsza para palców cynkowych (ZnF1 oraz ZnF2) znajduje się blisko

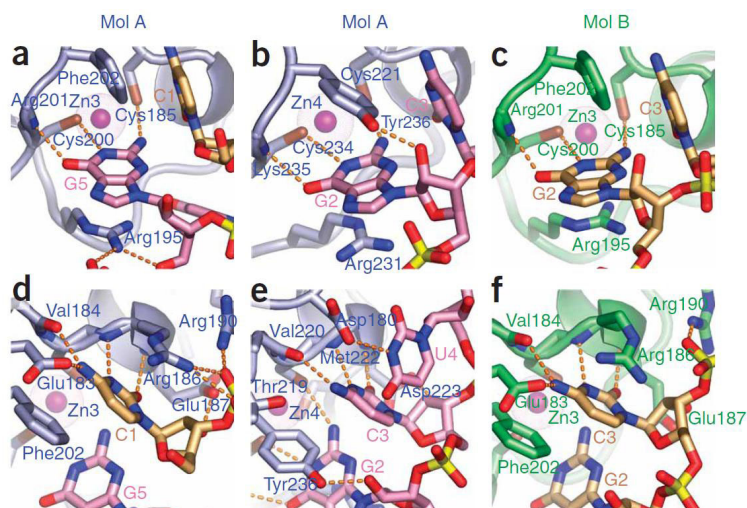


Rysunek 4. Schemat patogenezy dystrofii miotonicznej (57). Na panelu A przedstawiono fizjologiczne funkcjonowanie czynnika białkowego uczestniczącego w procesie składania pre-mRNA (czerwone kule) oraz czynnika transkrypcyjnego i innych białek (zielone trójkąty). W regionie 3' UTR pre-mRNA znajdują się powtórzenia CUG, które posiadają normalną liczbę powtórzeń. Powstający mRNA jest wariantem występującym u dojrzałych osobników. Panel B: pre-mRNA zawiera powtórzenia CUG wykazujące właściwości toksyczne, które polegają na nadmiernym wiązaniu istotnych dla funkcjonowania komórki białek. W efekcie powstaje wariant mRNA charakterystyczny dla płodu.

końca aminowego, natomiast druga (ZnF3 i ZnF4) ulokowana jest pośrodku białka (Rysunek 5). Funkcja końca karboksylowego nie została poznana, ale sugeruje się, że odpowiedzialna jest za oligomeryzację białka. Badania biochemiczne wykazały, że MBNL1 wiąże się do różnego rodzaju dwuniciowych RNA, w tym do powtórzeń CAG, CCG i CUG (59–61). Wydaje się, że preferowana dla wiązania jest obecność niestandardowych sparowań typu pirymidyna-pirymidyna oraz par G-C i C-G. Być może niestandardowe pary są istotne dla rozpoznawania RNA przez białko, lub umożliwiają częściowe rozplecenie regionu dwuniciowego i udostępnienie reszt guanozyny i cytozyny (62). Na podstawie struktury krystalicznej domeny palca cynkowego ZnF3/4 (reszty 178–246) z jednoniciowym oligomerem r(CGCGUGU) potwierdzono, że białko oddziałuje z motywem GC (63). Zagłębienie wiążące RNA utworzone jest z konserwatywnych pierścieni aromatycznych, reszt argininy oraz lizyny. Obserwowane są również oddziaływania warstwowe oraz wiązania wodorowe, które utworzone są pomiędzy krawędziami Watsona-Cricka zasad reszt nukleotydowych a atomami tlenu i azotu tworzącymi wiązanie amidowe, grupami karbonyłowymi oraz resztami cysteiny (Rysunek 6). Badania te, skupiały się jednak na poznaniu mechanizmu działania MBNL1 jako represora procesu składania. Jak dotąd nie poznano struktury kompleksu MBNL1 z dwuniciowym RNA zawierającym powtórzenia CUG. Nie zostało też wyjaśnione dlaczego białko jest nadmiernie wiązane przez trakty CUG. Nie wiadomo czy wynika to ze specyficznej struktury RNA, zwiększenia ilości miejsc, do których może się wiązać MBNL1 czy może innych niepoznanych powodów.



Rysunek 5. Schemat przedstawiający rozmieszczenie palców cynkowych w białku MBNL1 (A) (63). Do krystalizacji użyto dwa konstrukty, nazwane ZnF1/2 i ZnF3/4 (B). Każdy z konstruktyw zawiera dwa palce cynkowe.



Rysunek 6. Struktura krystaliczna kompleksu ZnF3/4 z oligomerem *r*(CGCUGU) (63). Na każdym panelu przedstawiono sposoby wiązania przez białko reszty guanazy (a-c) oraz reszty cytydyny (d-f).

2.3.2. Powtórzenia CGG

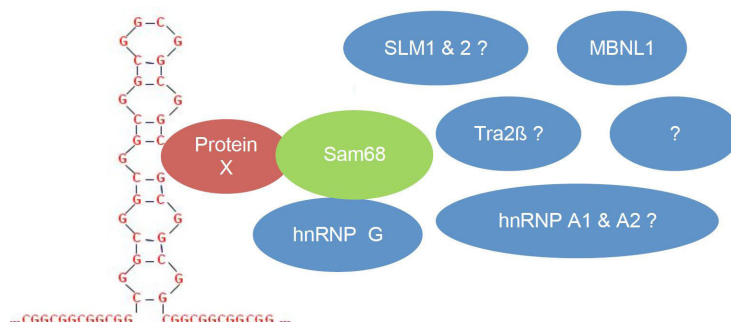
Łamliwość chromosomu X (FXS) wywoływana jest brakiem ekspresji białka FMRP, które normalnie syntetyzowane jest na wysokim poziomie w neuronach (64–66). Zaangażowane jest w eksport transkryptów z jądra komórkowego do cytoplazmy oraz w regulację transportu w synapsie (67). Wyciszenie genu *FMR1* następuje w wyniku jego metylacji, która jest indukowana przez powtórzenia CGG o pełnej mutacji (68). W przypadku choroby FXTAS (zespół drżenia i ataksji związany z łamliwym chromosomem X) poziom białka FMRP jest tylko nieznacznie obniżony, co sugeruje odmienny patomechanizm (66, 69). Pierwszym argumentem przemawiającym za tym jest występowanie szeregu objawów spotykanych tylko w syndromie FXTAS, a których nie obserwuje się w przypadku FXS (46). Jest to m.in.: ataksja, parkinsonizm, drżenie samoistne. Ponadto, charakterystyczna jest obecność inkluzji jądrowych tylko w neuronach i astrocytach dotkniętych FXTAS (70, 71). Udało się je odtworzyć u muszki owocowej wywołując ekspresję 90 powtórzeń CGG (72). Podobnie w transgenicznym mysz typu *knock-in*, ekspresja powtórzeń znajdujących się w naturalnym kontekście lub innych regionach nieulegających translacji, wywołała ten sam efekt oraz niektóre symptomy choroby (73, 74). Ponadto, poziom mRNA genu *FMR1* u chorych

osobników jest 2–8 krotnie podwyższony (66, 69). To wszystko wskazuje, że sama ekspresja wydłużonych traktów CGG jest niezbędna i wystarczająca do wywołania choroby FXTAS.

Naukowcy są zgodni co do roli RNA w patogenezie syndromu FXTAS, natomiast istnieje spór co do samego mechanizmu jej rozwoju. Początkowo uważano, że patomechanizm jest zupełnie inny niż dla dystrofii miotonicznej, ponieważ obserwowane inkluzje jądrowe są większe oraz znakowane ubikwityną (70, 71). Oprócz mRNA genu *FMR1*, składają się one z ponad 20 białek, włączając w to białka szoku cieplnego (75). Na tej podstawie zaproponowano, że dysfunkcja neuronów powstaje na skutek zaburzenia ogólnego balansu i dostępności białek regulatorowych wiążących RNA. Inną sugerowaną przyczyną FXTAS jest indukowanie przez powtórzenia CGG odpowiedzi komórki, która uruchamia procesy mające na celu ochronę przed neurodegeneracją (76, 77). Efektem wtórnym tych procesów jest tworzenie inkluzji z białek, które uległy nadekspresji. Dopiero niedawno zaproponowano mechanizm podobny do dystrofii miotonicznej (78). Przeprowadzono serię eksperymentów przy użyciu różnych linii komórkowych, do których wprowadzano na wektorze ekspresyjnym powtórzenia CGG. Na podstawie uzyskanych wyników autorzy proponują, że kluczowym momentem rozpoczęcia choroby jest związanie do ciągów CGG białka Sam68. Co ciekawe, nie wiąże się ono bezpośrednio do RNA lecz prawdopodobnie za pośrednictwem innego białka, które nie zostało jeszcze zidentyfikowane (Rysunek 7). Podobnie jak białko MBNL1, Sam68 uczestniczy, obok transportu i poliadenylacji transkryptów, w procesie składania pre-mRNA. Jego wyciszenie osłabia różnicowanie neuronów w kulturach komórkowych (79) a u myszy prowadzi do zaburzeń motorycznych (80). Inne zaburzenia, które badano, wymagają dalszej weryfikacji ze względu na nieodpowiedni model zastosowany do tego rodzaju badań. Niemniej, przedstawione wyniki sugerują, że czynnik Sam68 jest odpowiedzialny za zmianę profilu alternatywnego składania u chorych na FXTAS. Dodatkową i istotną dla patogenezy właściwością białka Sam68 jest jego zdolność do oddziaływania z różnorodnymi białkami np. hnRNP-G, hnRNP-A1, Tra2 β . Ich związanie prowadzi do dalszej masowej agregacji kolejnych białek doprowadzając do zaburzenia ogólnej równowagi w komórce.

2.3.3. Powtórzenia CCG

Syndrom FRAXE związany jest z brakiem ekspresji białka FMR2, które występuje na terenie jądra komórkowego (9). Białko FMR2 należy do rodziny białek bogatych w reszty seryny oraz proliny i prawdopodobnie jest czynnikiem transkrypcyjnym (81, 82). Podobnie jak w przypadku łamliwości chromosomu X, u chorych posiadających ponad 200 powtórzeń CCG, gen *FMR2* ulega hipermetylacji i w efekcie dezaktywacji. Brak tego białka u transgenicznych myszy upośledza percepcję zmysłową oraz odruch warunkowy na strach (83). Z kolei u muszki owocowej zaburza embriogenezę z powodu zmian, jakie zachodzą w różnych szlakach sygnałnych (kaskada sygnałna z udziałem transformującego czynnika wzrostu β czy kinazy białkowej MAPK) (81, 82).



Rysunek 7. Proponowany mechanizm patogenezy powtórzeń CGG w rozwoju FXTAS (78). Do wydłużonych ciągów CGG wiąże się niezidentyfikowane białko, które następnie przyłącza białko Sam68. Kolejne białka, które wiążą się do Sam68 to: hnRNP-G, być może również białka SLM1, SLM2, hnRNP-A oraz Tra2 β . Do takiego kompleksu wiąże się też białko MBNL1 ale prawdopodobnie na jego obrzeżach. Wiązanie to być może nie jest wystarczająco silne i dlatego białko MBNL1 nie agreguje z RNA i związanymi z nim białkami.

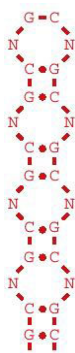
Na podstawie podobieństwa pomiędzy chorobami związanymi z genami *FMR1* i *FMR2*, Sofoła *i in.* (26) postanowili odpowiedzieć na pytanie czy powtórzenia CCG o długości premutacyjnej mogą wykazywać właściwości toksyczne. Do genomu transgenicznej muszki owocowej wprowadzili gen *CCG₉₀-EGFP*, który zawierał powtórzenia (CCG)₉₀ oraz sekwencję genu białka reporterowego EGFP (ang. *enhanced green fluorescent protein*). Ekspresja genu *CCG₉₀-EGFP* w różnych tkankach powodowała rozwój fenotypu neurodegeneracyjnego. Uzyskane wyniki sugerowały, że powtórzenia CCG o długości premutacyjnej wykazują toksyczność na poziomie mRNA i być może związane są z nieodkrytą jeszcze odmianą ataksji. Co więcej, koekspresja powtórzeń CCG z powtórzeniami CGG powodowała supresję fenotypu neurodegeneracyjnego w wyniku obniżenia poziomu patogenego transkryptu. Prawdopodobnie ciągi obu powtórzeń hybrydowały ze sobą i tworzyły długi dwuniciowy odcinek RNA rozpoznawany przez maszynę siRNA.

2.3.4. Powtórzenia CAG

Do tej pory przeprowadzono liczne badania odnośnie toksycznych właściwości białek zawierających wydłużone trakty poliglutaminowe (9, 16). Natomiast niewielu z badaczy wskazywało pre-mRNA, posiadające długie powtórzenia CAG, jako dodatkowy czynnik uczestniczący w patogenezie. Przyczyną było to, iż w różnych modelach zwierzęcych choroba rozwijała się tylko wtedy, gdy zmutowane białko ulegało translokacji do jądra komórkowego (84–86). Jednak Li *i in.* (87) pokazali na modelu ataksji mózdkowo-rdzeniowej typu 3 (SCA3) transgenicznej muszki, że pre-mRNA może brać udział w patogenezie. Wprowadzenie 100 lub 250 powtórzeń CAG w region nieulegający translacji powoduje degenerację siatkówki i dysfunkcje neuronalne. Co więcej, toksyczność jest wzmacniana przez białko MBNL1, które rozpoznaje również powtórzenia CAG, aczkolwiek dowodów na zaburzenie procesu składania pre-mRNA nie odkryto. Zastanawiające jest także, że powtórzenia CAG tworzą złogi jądrowe, które kolokalizują z MBNL1 (59, 87, 88), ale

ich rozmiar jest mniejszy niż w dystrofii miotonicznej. Być może inkluzje RNA stanowią frakcję rozpuszczalną i tylko chwilowo uczestniczą w tworzeniu złogów. W takiej sytuacji niewykluczone, że zmniejsza to właściwości toksyczne RNA, które mogą być dalej transportowane do cytoplazmy i ulegać translacji. Niemniej otrzymane wyniki sugerują, że powtórzenia CAG biorą również udział w patogenezie zwiększając intensywność zaistniałych procesów degeneracyjnych.

2.4. Struktura drugorzędowa cząsteczek RNA zawierających powtórzenia typu CNG



Rysunek 8. Struktura drugorzędowa powtórzeń CNG z pętlą czteronukleotydomą (10).

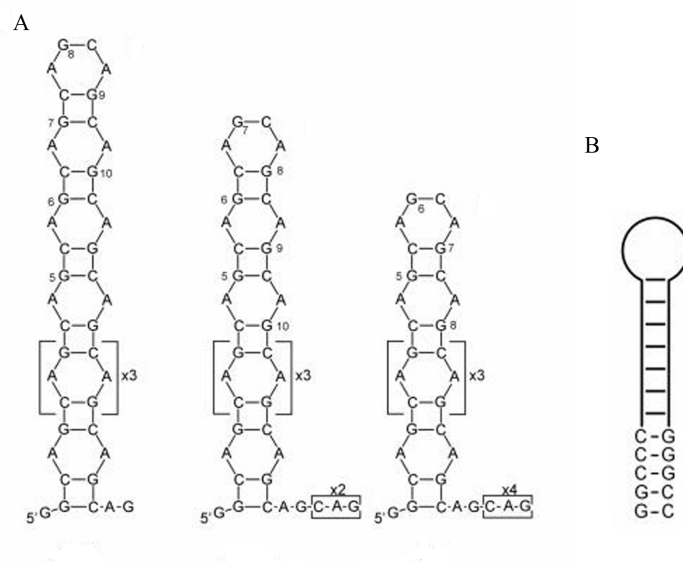
Cząsteczki RNA zawierające długie ciągi powtórzeń CNG wykazują właściwości toksyczne polegające na wiązaniu czynników niezbędnych dla poprawnego funkcjonowania komórki. Właściwości te zawdzięczają swojej strukturze. W związku z tym prof. dr hab. Włodzimierz Krzyżosiak, kierownik Pracowni Genetyki Nowotworów w Instytucie Chemii Bioorganicznej PAN w Poznaniu, podjął się zadania określenia ich struktury drugorzędowej. Głównym obiektem badań były długie ciągi powtórzeń CNG, których strukturę określano zarówno w formie izolowanej jak i w niektórych przypadkach z zachowaniem kontekstu strukturalnego i sekwencyjnego transkryptów. Sondy, jakie stosowano to jony ołowiu, które indukują cięcia w łańcuchu RNA w miejscach jednoniciowych i o zaburzonej strukturze. Do mapowania używano również różnego rodzaju rybonukleaz:

T1 (przecina po resztach guanozynowych występujących w regionach jednoniciowych), T2 (specyficzna dla regionów jednoniciowych z preferencją do cięcia po resztach adenozykowych), S1 (rozpoznaje regiony jednoniciowe bez specyficzności sekwencyjnej) oraz V1 (rozpoznaje regiony dwuniciowe).

RNA zawierające kilka powtórzeń CNG mogą tworzyć zarówno regiony jednoniciowe jak i dwuniciowe. Wykazano również, że mogą być uwikłane w oddziaływania z innym regionem mRNA (89). Natomiast długie trakty powtórzeń tworzą samodzielne struktury typu spinki zarówno jako izolowane fragmenty RNA jak i w sekwencji transkryptów (6, 90). W żelu poliakrylamidowym w warunkach niedenaturujących oligomery (CNG)₁₇ migrują jako jeden prążek a ich ruchliwość elektroforetyczna jest większa niż nieustrukturalizowanych powtórzeń np. (AAG)₁₇ (8). Ponadto, wykazują dużą stabilność, ponieważ w temperaturze 60 °C wzór cięcia nukleazami jest bardzo podobny do tego jaki obserwuje się w 37 °C.

Architektura wszystkich typów trójek CNG jest bardzo podobna (91). Trzon składa się z bloków dwóch par typu Watsona-Cricka: C-G i G-C, które rozdzielone są niestandardowymi parami N-N (A-A, C-C, G-G, U-U). Pętla apikalna składa się z czterech lub siedmiu nukleotydów. W obu przypadkach rozpoczyna się i kończy resztami, które w trzonie zaangażowane są w tworzenie pary N-N (Rysunek 8). Zauważono, że wielkość pętli zależy od liczby powtórzeń składających się na strukturę spinki. Gdy liczba trójek jest parzysta wówczas obserwuje się pętle 4 nukleotydowe, a gdy nieparzysta to jest to pętla 7 nukleotydowa.

Inną interesującą właściwością powtórzeń CNG jest to, że tworzą tzw. „poślizgnięte” struktury (ang. *slippery hairpins*) (91, 92). Oznacza to, że ten sam ciąg trójnukleotydowy może tworzyć alternatywne spinki różniące się liczbą par w trzonie. Niezaangażowane strukturalnie trójki CNG tworzą jednoniciowe, wystające końce 5' lub 3'. Dystrybucja poślizgniętych form zależy od stężenia jonów magnezu oraz od temperatury (91). Początkowo obserwowano wzór cięć, który był wypadkową alternatywnych form, ale efekt ten zlikwidowano poprzez wprowadzenie klamry czyli sekwencji tworzącej u podstawy trzonu region dwuniciowy (Rysunek 9). Spowodowało to, że liczba powtórzeń tworzących spinkę została jednoznacznie zdefiniowana a przez to i ich struktura. Naturalne klamry istnieją również w sekwencjach transkryptów niektórych genów np. *SCA6* i *SCA1* (6, 13, 89).



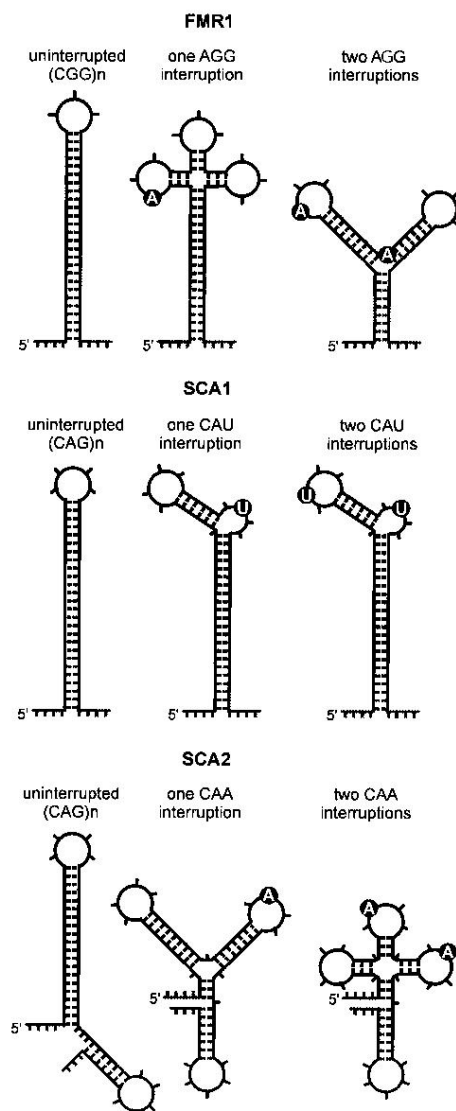
Rysunek 9. Poślizgnięte struktury tworzone przez powtórzenia CAG (91). Trzy alternatywne spinki różniące się liczbą powtórzeń CAG w trzonie oraz długością niesparowanego końca 3'(A). Spinka z klamrą na końcu trzonu stabilizuje jedną konformację struktury (B).

W mRNA genów *SCA1*, *SCA2* (powtórzenia CAG) oraz *FMR1* (powtórzenia CGG), w większość przypadków, pomiędzy ciągami powtórzeń występują sekwencje „przerwywnikowe” (zaburzenia) (6, 12–14). W allelach *SCA1* różnych populacji, w 78–95% przypadków znajdują się 1–3 zaburzenia (13). Najczęściej spotykany układ składa się z 23–43 powtórzeń z jednym

przerywnikiem CAT. Oznacza to, że w strukturach RNA o wymienionym wyżej układzie, zaburzenie spowodowało substytucję G na U i teoretycznie utworzenie pary U-C. Dla *SCA2* spotyka się głównie dwa zaburzenia CAA (zamiana G na A), które najczęściej występują w układzie 8-1-4-1-8 (gdzie cyfra oznacza liczbę powtórzeń CAG a 1 – jedno zaburzenie) (93). Natomiast w przypadku genu *FMR1* allele najczęściej zawierają dwa przerywniki AGG (zamiana C na A), przy czym trzy i cztery zaburzenia również są spotykane (12). Co ciekawe, u osobników posiadających patogeniczne powtórzenia zazwyczaj nie znajdują się sekwencje zaburzające, które mają wpływ na strukturę drugorzędową trójnukleotydowych traktów. Ich obecność może powodować, że długa spinka rozdziela się na dwie mniejsze, powiększa pętlę apikalną, bądź powoduje powstawanie jednostronnych wybrzuszeń (Rysunek 10). To, jaki efekt jest bardziej uprzywilejowany zależy od rodzaju zaburzenia. Na tej podstawie postuluje się, że funkcją sekwencji przerywnikowych jest destabilizacja struktur spinkowych powtórzeń trójnukleotydowych, co stanowi pewnego rodzaju obronę przed ich właściwościami toksycznymi. Niewykluczone, że ich rola jest również istotna na poziomie DNA, gdzie zaburzenia mogą zapobiegać ekspansji tych sekwencji podczas replikacji DNA.

Przeprowadzono szczegółowe badania strukturalne powtórzeń CUG znajdujących się w mRNA genu *DMPK* z zachowaniem jego kontekstu sekwencyjnego (92). Modele jakich użyto zawierały 5, 11, 21 oraz 49 powtórzeń oflankowanych z końca 5' 30 nukleotydami a z końca 3' 35 nukleotydami. Powtórzenia (CUG)₅ tworzyły regiony jednoniciowe, natomiast dłuższe, samodzielne struktury typu spinki nieoddziałujące z otaczającymi je resztami. Interesującą obserwacją był zanik cięć indukowanych jonami ołowiu wraz ze wzrostem liczby powtórzeń CUG. Być może efekt ten spowodowany jest usztywnieniem struktury trzonu lub wzrostem stabilności struktury w wyniku zwiększenia liczby oddziaływań warstwowych i/lub wiązań wodorowych.

Ze względu na zawartość reszt G sugeruje się, że powtórzenia CGG mogą przyjmować strukturę tetrapleksu w obecności jonów K⁺, które promują formowanie się tego typu struktur (94–96). Zauważono, że w żelach poliakrylamidowych w warunkach niedenaturujących, fragment RNA regionu 5' UTR genu *FMR1* zawierający 33 powtórzenia CGG wykazuje różną mobilność elektroforetyczną w zależności od obecności jonów K⁺ (95). Jeżeli bufor zawierał 10 mM KCl to RNA migrował szybciej, w postaci rozsmużonego prążka w stosunku do warunków gdzie nie było soli. Ponadto transkrypty poddano trawieniu RNazą T1 w warunkach niedenaturujących oraz dla porównania w denaturujących. Cięcia obserwowano w nielicznych miejscach, co wskazuje, że reszty guanozyn zaangażowane są w oddziaływania, a więc nie tworzą regionów jednoniciowych. Na tej podstawie autorzy sugerują formowanie się kompaktowej struktury, którą może być tetrapleks. Powyższe założenia nie są poparte przez wyniki eksperymentów topnienia UV (8) oraz NMR (97). Widma różnicowe UV otrzymane z pomiarów prowadzonych w obecności 100 mM NaCl lub 100 mM KCl przy dwóch różnych temperaturach 4 °C i 90 °C nie pokazały przejścia przy długości

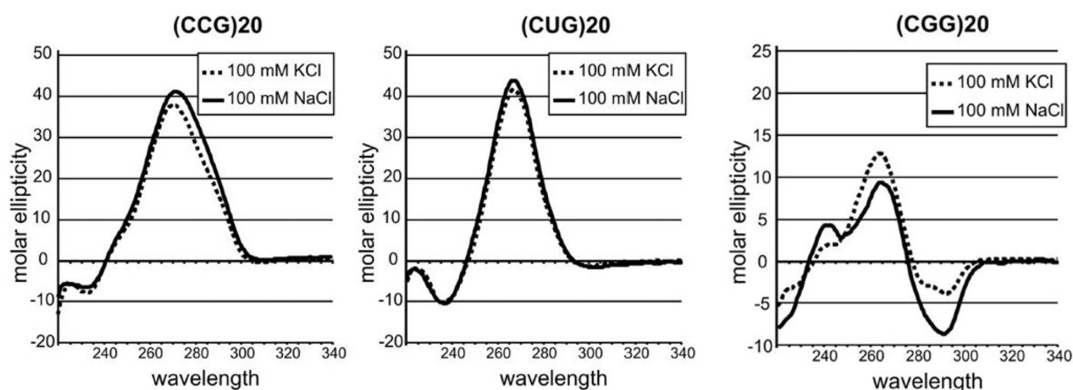


Rysunek 10. Struktury drugorzędowe fragmentów mRNA genów *FMR1*, *SCA1* i *SCA2* zawierających różną liczbę sekwencji przerywnikowych (6). Na rysunku przedstawiono wpływ jednego i dwóch przerywników (pogrubione reszty nukleotydydowe), które powodują zmianę struktury RNA.

fali 295 nm. Dla struktur tetrapleksu przy takiej długości fali obserwuje się piki negatywne. Nie nastąpił również wzrost temperatury topnienia T_m RNA w obecności jonów potasu. Badania NMR wskazały, że oligomery zawierające 6 powtórzeń CGG są strukturami dynamicznymi, co uniemożliwia jednoznaczną interpretację widm. Niemniej, nie obserwowano formowania się tetrapleksu nawet w wysokich stężeniach oligomeru używanego do pomiaru.

W przypadku wyników otrzymanych z wykorzystaniem dichroizmu kołowego sytuacja jest niejasna (8). Pomiar prowadzone były w 100 mM NaCl lub 100 mM KCl. Okazało się, że dla powtórzeń CAG, CCG i CUG widma w obu warunkach solnych są bardzo podobne i mają wysoki pozytywny pik, którego maksimum przypada przy długości fali 265–270 nm, czyli typowy dla formy A-RNA. Natomiast dla powtórzeń CGG już w buforze z NaCl widmo jest bardziej skomplikowane (Rysunek 11). Obecne są dwa maksima przy długości fali 240 nm i 265 nm oraz negatywny pik przy

290 nm. Podobny kształt widma obserwowany jest w 100 mM KCl, przy czym maksima mają inną wartość molowości eliptycznej. Sugeruje to, że struktura CGG może mieć nieco inną architekturę niż pozostałe powtórzenia ale nie jest to raczej struktura tetrapleksu.



Rysunek 11. Widma dichroizmu kołowego powtórzeń CCG, CUG i CGG otrzymane w obecności NaCl i KCl (8). Widma powtórzeń CUG i CCG są podobne do siebie. Przy długości 265 nm widoczne jest maksimum, które charakterystyczne jest dla formy A-RNA. W przypadku powtórzeń CGG otrzymane widma mają inny kształt niż dla powtórzeń CUG i CCG.

2.5. Właściwości termodynamiczne powtórzeń trójnukleotydowych typu CNG

Podobnie, jak badania nad określeniem struktury drugorzędowej powtórzeń CNG, badania termodynamiczne zostały przeprowadzone w Instytucie Chemii Bioorganicznej PAN w Poznaniu (5, 8). Obiektem były zarówno krótkie oligomery zawierające od 2 do 7 powtórzeń (5), jak i zawierające kilkanaście powtórzeń (8). Wykazano, że RNA składające się z 2–3 trójnukleotydów CNG tworzą dupleksy. Przy 4–5 powtórzeniach obserwuje się obecność obu form, dupleksu i spinki, natomiast powyżej pięciu przeważa struktura spinkowa.

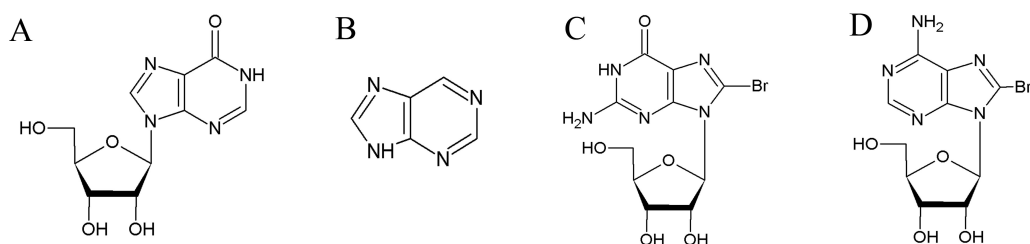
Interesującą właściwością dupleksów $G(CNG)_{2-4}C$ jest to, że ich stabilność jest bardzo podobna i w małym stopniu zależy od liczby powtórzeń oraz rodzaju niestandardowej pary N-N (5). Niemniej, najniższą energię swobodną ΔG°_{37} mają oligomery zawierające ciągi CGG (około -10 kcal/mol) potem CUG (-7,5 kcal/mol) i CAG \approx CCG (około -6,5 kcal/mol). Podobnie jest dla struktur spinkowych zawierających 5–7 powtórzeń CNG. Ich stabilność, niezależnie od rodzaju trójnukleotydu, jest bardzo podobna przy czym wydaje się, że najbardziej stabilne termodynamicznie są spinki o powtórzeniach CUG, potem CAG, CGG i CCG. Należy jednak pamiętać, że przy takiej ilości powtórzeń w roztworze istnieją obie formy – dupleksu i spinki. Równowaga przy niskich stężeniach oligomeru i NaCl jest przesunięta w kierunku spinki, ale dekonwolucja krzywych topnienia pozwala obliczyć osobno parametry termodynamiczne dla

każdej z form RNA.

Na ogólną stabilność termodynamiczną struktury spinki składa się stabilność jej trzonu oraz cztero- lub siedmionukleotydowej pętli. W przypadku spinek CNG, stabilność trzonu jest bardziej istotna niż stabilność pętli. Wyznaczone energie swobodne dla każdego typu trzonu wskazały, iż szereg stabilności jest taki sam jak w przypadku dupleksu. Najniższą wartość ΔG°_{37} mają powtórzenia CGG (-5,20 kcal/mol), następnie CUG (-3,90 kcal/mol), CAG (-3,85 kcal/mol) i CCG (-3,36 kcal/mol). Dla obu rodzajów pętli (cztero- i siedmionukleotydowej) efekt stabilizacji jest porównywalny do efektów obserwowanych dla pętli o tej samej wielkości lecz sekwencji innej niż CNG. Wydaje się, że na stabilność pętli wpływa rodzaj reszty nukleotydowej znajdującej się na końcu 5' (pierwsza reszta wchodząca w skład pętli). Taki nukleotyd można rozpatrywać jako niesparowany koniec 3'. W takim przypadku może on być źródłem stabilizacji struktury w zależności od jego zdolności do tworzenia oddziaływań warstwowych ($A \approx G > U > C$) z helikalnym regionem trzonu spinki. Zgodnie z powyższym energie swobodne pętli NGCN wynoszą: -3,30, -3,03, -2,72, -1,88 kcal/mol dla $N = A, G, U$ lub C . W przypadku pętli siedmionukleotydowych efekt ten jest bardziej złożony. Wydaje się, że tworzenie się spinek CNG z 7 nukleotydową pętlą jest mniej uprzywilejowane niż z pętlą czteronukleotydową. W konsekwencji, takie spinki mogą przyjmować alternatywne, współistniejące w roztworze struktury, np. spinka z 4 nukleotydową pętlą i trójnukleotydowym niesparowanym końcem 3'. Zatem, zbadanie efektu termodynamicznego pętli siedmionukleotydowej nie jest łatwe do określenia.

Dodatkowo, Broda *i in.* (5) określili właściwości termodynamiczne dupleksów zawierających dwa powtórzenia CAG lub CGG, w których reszty adenozyne i guanozyne, tworzące niestandardowe pary N-N, zastępowano ich modyfikowanymi odpowiednikami. Otrzymano 12 różnych wariantów takich oligomerów (Tabela 2). Na podstawie obserwacji wzrostu lub obniżenia energii swobodnej w stosunku do niemodyfikowanego dupleksu można było uzyskać informację o konformacji zasady względem pierścienia cukrowego oraz wskazać grupy funkcyjne zaangażowane w wiązania wodorowe w obrębie pary N-N. Do oligomeru GCGGCGGC wprowadzano inozynę (I) lub purynorybozyd (P) w pozycję 3 lub 6 bądź w obu miejscach jednocześnie. Ponadto, w obydwu dupleksach $(GCGGCGGC)_2$ i $(GCAGCAGC)_2$, resztę guanozyne oraz adenozyne modyfikowano atomem bromu w pozycji 8 pierścienia zasady (Rysunek 12). Podstawienie 8-bromo promuje zmianę konformacji zasady z *anti* na *syn*, co oznacza, że pierścień purynowy znajduje się nad resztą rybozy.

Otrzymane wyniki badań termodynamicznych sugerują, że pomiędzy niestandardowymi parami A-A i G-G istnieją oddziaływania wodorowe. Szczególnie silnie oddziałują ze sobą pary G-G, ponieważ wprowadzenie inozyny lub purynorybozydu, czyli zmniejszenie liczby grup funkcyjnych mogących utworzyć wiązania wodorowe, powodowało destabilizację. Efekt ten był większy dla purynorybozydu. Z kolei, oligomery zawierające 8-bromoguanozynę (8-BrG) w pozycji 3 lub 6 były bardziej trwałe. Energia swobodna była korzystniejsza o około 4 kcal/mol a temperatura topnienia



Rysunek 12. Wzór inozyzny (A), zasady purynowej (B), 8-bromoguanozyny (C) oraz 8-bromoadenozyny (D).

Tabela 2. Zestawienie parametrów termodynamicznych serii oligomerów zawierających powtórzenia CGG lub CAG (5).

RNA sequence	1/T _M vs log C _T parameters				curve fit parameters						
	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG° ₃₇ (kcal/mol)	T _m ^a (°C)	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG° ₃₇ (kcal/mol)	T _m (°C) ^b	ΔΔG° ₃₇ (kcal/mol)	ΔT _m (°C)	
GCGGCGGC	56.7 ± 2.6	161.1 ± 8.4	6.76 ± 0.04	43.0	48.9 ± 3.0	135.9 ± 9.6	6.70 ± 0.16	43.6	0	0	
GCI GCGGC	65.5 ± 7.8	192.6 ± 25.4	5.78 ± 0.19	37.5	53.9 ± 3.6	154.8 ± 11.8	5.93 ± 0.20	38.4	0.77	−5.2	
GCGGCI GC	74.8 ± 14.1	220.8 ± 45.4	6.29 ± 0.48	39.6	63.2 ± 9.1	183.4 ± 29.5	6.33 ± 0.27	40.3	0.37	−3.3	
GCI GCI GC	46.4 ± 3.1	132.9 ± 10.1	5.15 ± 0.11	33.5	48.5 ± 7.4	139.7 ± 24.0	5.14 ± 0.12	33.6	1.56	−10.0	
GCPGCGGC	57.1 ± 2.6	167.0 ± 8.4	5.29 ± 0.06	34.9	53.7 ± 5.6	155.8 ± 18.7	5.41 ± 0.18	35.5	1.29	−8.1	
GCGGCPGC	64.3 ± 6.9	186.6 ± 22.1	6.39 ± 0.18	40.5	50.9 ± 7.4	143.3 ± 24.5	6.46 ± 0.29	41.8	0.24	−1.8	
GCPGCPGC	36.5 ± 3.6	101.2 ± 11.8	5.09 ± 0.17	32.1	47.4 ± 15.2	136.8 ± 49.6	4.96 ± 0.26	32.4	1.74	−11.2	
GC ^{Br} GGCGGC	98.7 ± 5.0	280.9 ± 15.4	11.53 ± 0.24	56.6	76.4 ± 3.2	212.7 ± 9.6	10.45 ± 0.33	57.7	−3.70	14.1	
GCGGC ^{Br} GGC	74.8 ± 2.4	208.2 ± 7.5	10.20 ± 0.12	57.0	86.3 ± 8.3	243.3 ± 25.2	10.80 ± 0.53	56.6	−4.10	13.0	
GC ^{Br} GGC ^{Br} GGC	46.5 ± 3.6	129.8 ± 11.7	6.24 ± 0.09	40.8	52.7 ± 4.6	150.0 ± 15.3	6.19 ± 0.23	40.0	0.51	−3.6	
GCAGCAGC	35.0 ± 1.3	94.6 ± 4.3	5.69 ± 0.03	37.1	38.7 ± 3.7	106.4 ± 11.8	5.69 ± 0.06	37.1	0	0	
GC ^{Br} AGCAGC	49.1 ± 2.7	142.2 ± 9.0	5.03 ± 0.10	33.0	51.4 ± 4.8	149.5 ± 15.7	5.01 ± 0.11	33.1	0.68	−4.0	
GCAGC ^{Br} AGC	44.9 ± 1.7	128.2 ± 5.7	5.15 ± 0.31	33.4	52.5 ± 5.1	153.1 ± 16.6	4.96 ± 0.12	32.8	0.73	−4.3	
GC ^{Br} AGC ^{Br} AGC	33.8 ± 5.8	97.4 ± 20.0	3.63 ± 0.48	19.3	42.9 ± 30.7	127.4 ± 103.1	3.45 ± 1.33	21.7	2.24	−15.4	

^a Solutions are 1 M sodium chloride, 20 mM sodium cacodylate, and 0.5 mM Na₂EDTA, pH 7. ^b Melting temperatures are calculated for 10^{−4} M oligomer concentration.

wzrosła o 13–14°C. Gdy obie reszty były modyfikowane jednocześnie, parametry termodynamiczne były podobne do referencyjnego oligomeru. Sugeruje to, że para G-G przyjmuje konformację *syn-anti*, przy czym jest to stan dynamiczny. Rozumiany jest on w ten sposób, że następuje swobodne przechodzenie reszty guanozyny z formy *syn* do *anti*. Podobne rezultaty otrzymano w badaniach NMR (97).

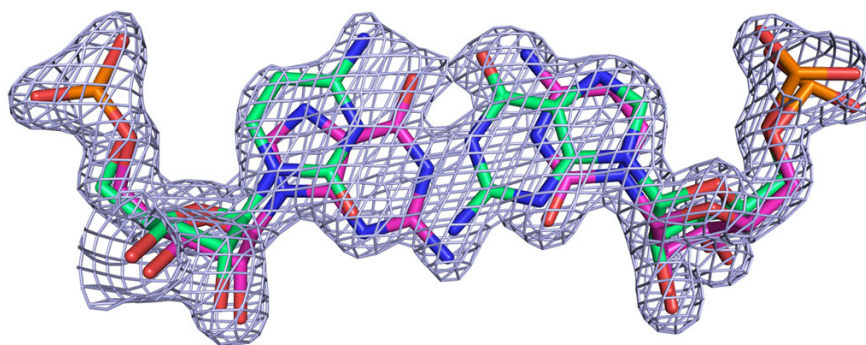
Dla dupleksów (GCAGCAGC)₂, w których w pozycję 3 lub 6 wprowadzono 8-bromoadenozynę (8-BrA), obserwowano nieznaczną destabilizację. Natomiast, gdy jednocześnie podstawiono obie reszty adenozyzny modyfikowanym nukleotydem, efekt ten był o wiele większy. Następował wzrost energii swobodnej o 2 kcal/mol i obniżenie temperatury topnienia o 15°C, co w tej sytuacji wskazywało na brak oddziaływań pomiędzy dwiema resztami 8-BrA. Na podstawie otrzymanych wyników, dotyczących dupleksu (GCPGCPGC)₂ można również wykluczyć udział grupy *exo*-aminowej adenozyzny w tworzeniu wiązań wodorowych pary A-A.

Parametry termodynamiczne wyznaczone dla dłuższych fragmentów RNA zawierających ciągi CNG są zgodne z wynikami otrzymanymi dla krótkich oligomerów (8). Wartości energii swobodnych znajdują się w zakresie od -2,38 do -6,68 kcal/mol. Zakłada się, że czynnikiem, który je różnicuje

jest rodzaj niestandardowej pary N-N, ponieważ architektura ich struktury drugorzędowej jest podobna. W 100 mM NaCl najbardziej trwałe termodynamicznie są powtórzenia CGG następnie CAG, CUG i na końcu CCG. Natomiast gdy do eksperymentu topnienia użyto buforu zawierającego 100 mM KCl szereg stabilności wyglądał nieco inaczej: CAG, CGG, CUG i CCG. W przypadku oligomerów zawierających powtórzenia CGG w roztworze KCl nastąpił największy spadek wartości energii swobodnej (o 2,36 kcal/mol). Dla pozostałych trójnukleotydów efekt destabilizacji był mniejszy i znajdował się w zakresie 0,18–0,69 kcal/mol. Taki wynik oznacza, że żaden z oligomerów nie tworzy tetrapleksu (brak wzrostu stabilności i temperatury topnienia), którego formowanie jest promowane pod wpływem KCl.

2.6. Struktura krystaliczna powtórzeń CUG

Pierwszym doniesieniem o strukturze przestrzennej powtórzeń CUG była praca opublikowana przez Mooers *i in.* (98). Autorzy określili strukturę dupleksu zawierającego 6 powtórzeń CUG o rozdzielczości 1,6 Å (kod PDB: 1zev). Zamierzeniem autorów było wykrycie spinki, ale w wysokich stężeniach oligomeru i soli nastąpiło przesunięcie równowagi w kierunku formowania się dupleksu. Opisany model składa się z dwóch nakładających się na siebie dupleksów. Ich szkielet fosforanowo-cukrowy pokrywa się i w efekcie mapa gęstości elektronowej jest niejednoznaczna (Rysunek 13). Autorzy tłumaczą, że jest to skutek statycznego nieuporządkowania. Obie helisy mają formę A-RNA i składają się z par G-C i C-G oraz U-U. Dupleksy w sieci krystalicznej układają się jedna nad drugą tworząc pseudonieskończoną helisę, naśladując w ten sposób trzon spinki składającej się z długich traktów CUG. Niekanoniczne pary U-U znajdują się naprzeciw siebie, ale mimo to nie tworzą oddziaływań wodorowych. Odległość pomiędzy atomami węgla C1' tych reszt wynosi około 10,5 Å. Dla porównania dla par G-C odległość ta wynosi około 10,7–10,8 Å.



Rysunek 13. Mapa gęstości elektronowej (kontur 1σ) struktury zdeponowanej przez Mooers *i in.* z wmodelowaną parą G-C. Na różowo zaznaczono pierwszy model a na zielono drugi.

Strukturę tą wykorzystałam w swoich badaniach, ponieważ postanowiłam zweryfikować czy niejednoznaczna mapa gęstości elektronowej jest efektem statycznego nieuporządkowania czy bliźniaczenia (*vide* Wyniki: 3.1.5).

2.7. Struktury spinkowe RNA stabilne w warunkach krystalizacji

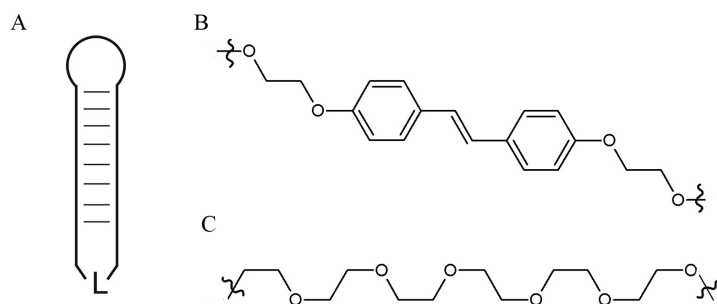
W niektórych mRNA ciągi powtórzeń CNG tworzą struktury typu spinki. Krystalizacja takich izolowanych struktur jest bardzo trudna, ponieważ są one wysoce nietrwałe w warunkach dużego stężenia soli i RNA stosowanych w badaniach biokrytalograficznych (99). Fragmenty te pozostają w równowadze pomiędzy formą spinki a dupleksem, który dominuje w roztworze o zwiększonej sile jonowej. Stąd, ze względów termodynamicznych bardziej prawdopodobne jest otrzymanie kryształów dupleksu niż spinki. Ze względu na ten fakt, realizację celu dotyczącego określenia struktur powtórzeń CNG rozpoczęłam od krystalizacji oligomerów RNA zawierających kilka powtórzeń CNG tworzących stabilny dupleks w stosowanych warunkach. Istotne było także, iż wydajność syntezy chemicznej oligomerów RNA jest bardzo duża dla krótkich oligomerów (około 10 nukleotydów) w porównaniu z dłuższymi cząsteczkami RNA.

Przedstawiona powyżej kwestia uświadomiła mi, że istnieje duża bariera w poznawaniu struktur spinkowych i stanowi to ogólny problem, który napotykają badacze, także ci zajmujący się spektroskopią NMR. W tym przypadku, utrudnieniem jest interpretacja widm otrzymanych w roztworze, w którym dynamicznie współistnieją obie formy RNA. Odzwierciedleniem tego stanu rzeczy jest mniejsza liczba poznanych struktur RNA (w bazie danych PDB znajduje się 3207 zdeponowanych struktur kompleksów kwasów nukleinowych, z czego tylko 869 to struktury RNA) w stosunku do liczby struktur białek (68924 struktur zdeponowanych w PDB). Samych struktur spinek RNA określonych metodą krystalograficzną jest zaledwie kilkanaście.

Z uwagi na szerokie spektrum działania, spinki RNA stanowią niezwykle ciekawy obiekt badań, również tych strukturalnych. Struktury spinkowe są najbardziej charakterystycznym motywem występującym w cząsteczkach RNA (100). Pomimo swojej nieskomplikowanej topologii posiadają swoistą różnorodność strukturalną. W ich obrębie mogą tworzyć się różne konformacje pętli oraz trzonu, które współtworzą globalną strukturę przestrzenną (101). Bogactwo strukturalne spinek decyduje o ich bogactwie funkcjonalnym. Struktura tego typu może być składową częścią rozbudowanych cząsteczek takich jak rybozomy, lub samodzielną jednostką funkcjonalną. Jako pojedynczy element pełni różnorodne role w komórce: moduluje stabilność mRNA, jest prekursorem mikroRNA, atenuatorem czy wreszcie miejscem specyficznego oddziaływania z białkami i innymi ligandami.

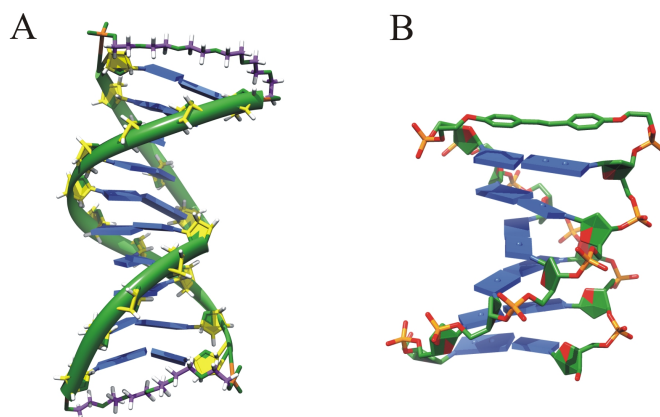
Powyższe przesłanki skłoniły mnie do opracowania uniwersalnej metody stabilizacji cząsteczek RNA w formie spinki poprzez kowalencyjne połączenie wolnych końców (cyrkularyzacja) oligomeru tak, aby zachować jego natywną strukturę. W tym celu do oligomeru musi zostać wprowadzony łącznik, który będzie znajdował się na końcu trzonu spinki, naprzeciw pętli apikalnej (Rysunek 14). Potencjalnymi łącznikami stosowanymi w biochemii są glikole polietylenowe (102) oraz pochodne eterowe stilbenu (103). Glikole polietylenowe są liniowymi związkami składającymi się z

monomerów etylenu, natomiast pochodne stilbenu zbudowane są z dwóch pierścieni fenylowych połączonych wiązaniem winylowym (Rysunek 14).



Rysunek 14. Schemat stabilnej struktury spinkowej w warunkach wysokiego stężenia soli i RNA (A). Literą L oznaczono miejsce łącznika, który spina końce trzonu spinki. Stosowanymi łącznikami mogą być, np. glikol heksaetylenu (B) lub pochodna eterowa stilbenu (C).

Zaletą tych łączników jest możliwość modulowania ich długości. W przypadku glikoli polietylenowych można swobodnie dodawać odpowiednią ilość jednostek etylenowych. Ustalono, że optymalnie łącznik dla DNA oraz RNA powinien zawierać od 5 do 7 jednostek polietylenowych (102, 104). Jak dotąd najczęściej stosowany jest glikol heksaetylenu (EG6, spacer 18). W przypadku związków stilbenu do pierścieni fenylowych mogą być symetrycznie przyłączane odpowiednie podstawiniki, np. eterowe czy amidowe, za pomocą których można dostosowywać długość łącznika spinającego dupleks (103). Ten typ łącznika jest interesujący, gdyż naśladuje sparowane zasady, z którymi może tworzyć oddziaływania warstwowe. Oba łączniki (EG6 oraz eterowa pochodna stilbenu) były wykorzystane do badań strukturalnych kwasów nukleinowych zarówno techniką NMR (105, 106) oraz krystalografii (107), przy czym badane modele nie były zcyrkularyzowanymi spinkami lecz dupleksami, których nici były połączone łącznikiem (Rysunek 15).

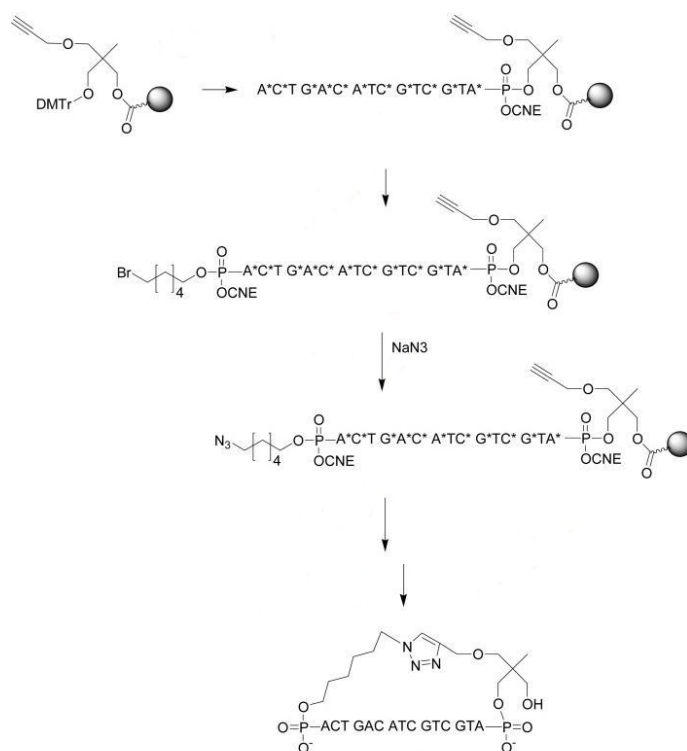


Rysunek 15. Struktury zawierające linkery EG6 (wzór patyczkowy; fioletowy) (A) i stilbenu (wzór patyczkowy; zielony) (B).

Odpowiednio przygotowany oligomer RNA zawierający łącznik musi zostać zcyrkularyzowany (połączenie końców 5' i 3' wiązaniem kowalencyjnym). Reakcję tą można przeprowadzić na drodze

enzymatycznej, za pomocą enzymu ligazy RNA bądź DNA z faga T4, które są standardowo stosowane w badaniach biochemicznych. Alternatywnie można również wykorzystać metodę ligacji chemicznej, która polega na aktywowaniu grupy fosforanowej czynnikiem kondensującym (najczęściej stosowany jest bromocyjan). Głównym czynnikiem decydującym o efektywności ligacji chemicznej jest obecność monofosforan na końcu 3' (108), wówczas wydajność reakcji jest większa oraz unika się ryzyka powstania produktu ubocznego zawierającego wiązanie 2'-5' fosfodiesterowe. Miejsce ligacji powinno znajdować się w regionie dwuniciowym i wydaje się, że poprawność jego utworzenia decyduje o wydajności reakcji, która może wynieść do 85% (109). Zaletą tej metody jest niska cena używanych odczynników. Wadą jest toksyczność czynnika kondensującego.

Modyfikacją zaproponowanego podejścia jest użycie tzw. „chemii klik” (ang. *click chemistry*) (110). Polega ona na szybkiej i powtarzalnej syntezie związku poprzez łączenie małych jednostek ze sobą. Najczęściej stosowaną reakcją jest utworzenie wiązania pomiędzy grupą azydkową a alkinową w obecności związków miedzi. Zastosowanie tej reakcji jest możliwe również w celu syntezy kolistych oligomerów (Rysunek 16). W tym podejściu stosuje się zmodyfikowane podłoże i grupę fosforoamidową. Syntetyzowany oligomer nie posiada łącznika, ponieważ powstaje on w wyniku reakcji klik.



Rysunek 16. Schemat syntezy zcykularyzowanego oligomeru metodą klik (110). Wydłużanie łańcucha kwasu nukleinowego odbywa się na modyfikowanym podłożu, które zawiera grupę acylową. Syntezę kończy przyłączenie bromowej pochodnej fosforamidu, która modyfikowana jest do azydku. Ligowanie oligomeru następuje w obecności $Cu(I)$.

3. Wyniki

3.1. Struktury RNA zawierające powtórzenia CUG

Określiłam dwie struktury RNA zawierające powtórzenia CUG: dupleksu $(GCUGCUGC)_2$ oraz $[(CUG)_6]_2$. Pierwszy z nich został zaprojektowany, oczyszczony i wykryty przez mnie. Drugą strukturę określiłam w oparciu o czynniki struktury zdeponowane wcześniej przez Mooers *i in.* (98) w PDB (*Protein Data Bank*) (kod: 1zev). Model ten udokładniałam z uwzględnieniem bliźniaczenia.

3.1.1. Krystalizacja oligomeru GCUGCUGC

Liofilizowany oligomer GCUGCUGC został rozpuszczony w 5 mM $MgCl_2$ (stężenie końcowe RNA wynosiło 2 mM), podgrzany do temperatury 65 °C przez 5 minut a następnie schładzany stopniowo do temperatury pokojowej przez noc. Kryształy otrzymane były techniką wiszącej kropli z dyfuzją par w temperaturze 19 °C. Początkowo kropla zawierała 2 μl RNA oraz 2 μl roztworu rezerwuarowego (50 mM octan sodu pH 5,5, 100 mM $MgCl_2$, 1,5 M Li_2SO_4). Kryształy powstawały po upływie 2–3 dni.

3.1.2. Rejestracja i przetwarzanie danych

Pomiary dyfrakcyjne kryształu $(GCUGCUGC)_2$ zostały zarejestrowane na linii pomiarowej X13 znajdującej się w ośrodku synchrotronowym DESY w Hamburgu. Pomiaru dokonano w temperaturze 100 K a jako krioprotektantu użyto 25% glicerolu. Zarejestrowano 900 obrazów dyfrakcyjnych z oscylacją 1° o maksymalnej rozdzielczości 1,23 Å. Ostatecznie po uważnym przetworzeniu danych, pomimo wartości kąta $\beta = 90^\circ$, grupę przestrzenną wyznaczono jako C2. Charakterystyka danych dyfrakcyjnych znajduje się w Tabeli 3.

3.1.3. Rozwiązanie struktury $(GCUGCUGC)_2$

Analiza zawartości niezależnej części komórki elementarnej przy pomocy współczynnika Matthews (111) wskazywała na obecność trzech cząsteczek dupleksu przy zawartości roz-

Tabela 3. Dane krystalograficzne oraz charakterystyka zbioru danych dla kryształu struktury dupleksu (GCUGCUGC)₂. W nawiasie podano wartości dotyczące ostatniego zakresu rozdzielczości.

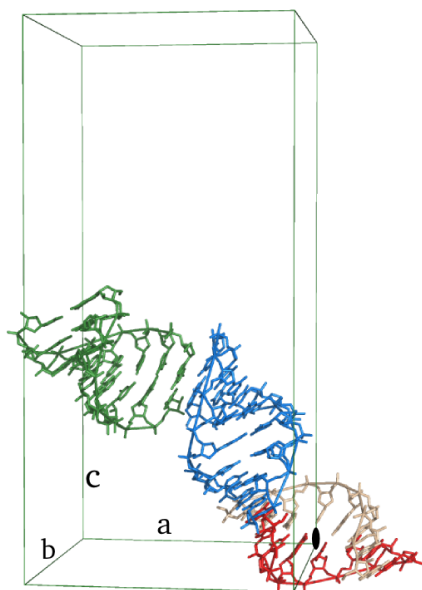
Dupleks	(GCUGCUGC) ₂
Źródło promieniowania	EMBL-X13
Temperatura (K)	100
Długość fali (Å)	0,808
Grupa przestrzenna	C2
Parametry komórki (Å)	a = 37,0 b = 38,9 c = 77,7 β = 90,0°
Zakres rozdzielczości (Å)	20,0–1,23 (1,25–1,23)
R _{merge}	0,058 (0,30)
Liczba unikalnych refleksów	32055
Kompletność (%)	99,1 (98,1)
Mozaikowość (°)	1
Krotność obserwacji	7,1(6,2)
<I/σ(I)>	21,3 (5,7)
Refleksy > 2σ (%)	86 (72)
Czynnik B z wykresu Wilsona (Å ²)	14,2

puszczalnika wynoszącej 47%. Do podstawienia cząsteczkowego użyto struktury dupleksu [GC(8-BrG)GCGGC]₂, w którym odpowiednie reszty guanozynowe zmutowano do reszt urydynowych. Próby rozwiązania problemu fazowego w grupie przestrzennej C2 zakończyły się niepowodzeniem. Jedną z przyczyn mógł być nieprawidłowy model zastosowany do obliczeń bądź nieodpowiednia grupa przestrzenna. W celu wyeliminowania drugiej z przyczyn dane przetworzono w grupie P1 (Tabela 4). Było to możliwe dzięki odpowiednio dużej liczbie zarejestrowanych obrazów dyfrakcyjnych. Zapewniało to dobrą krotność zbioru refleksów. Fazowanie z użyciem programu PHASER (112) w tym przypadku przebiegło prawidłowo.

Tabela 4. Statystyka danych dyfrakcyjnych kryształu oligomeru GCUGCUGC przetworzonych w grupie przestrzennej P1. W nawiasie podano wartości dotyczące ostatniego zakresu rozdzielczości.

Grupa przestrzenna	P1
Parametry komórki (Å)	a = 26,8 b = 26,8 c = 77,7 α = 90° β = 90,0° γ = 87,15°
Zakres rozdzielczości (Å)	20,0–1,23 (1,25–1,23)
R _{merge}	0,049 (0,30)
Liczba unikalnych refleksów	63053
Kompletność (%)	96 (90)
Mozaikowość (°)	1
Krotność danych	4 (3,7)
<I/σ(I)>	17 (4,3)
Refleksy > 2σ (%)	83 (77,3)

Dużym zaskoczeniem był wynik analizy upakowania cząsteczek w sieci krystalicznej. Okazało się, że odpowiednią grupą przestrzenną jest wcześniej określona grupa C2. Ostatecznie, w części asymetrycznej znajdują się trzy duplekisy: A+B, C+D i E+E', gdzie E' jest nicią symetrycznie ekwiwalentną do łańcucha E, generowaną przez oś dwukrotną. Sąsiednie cząsteczki RNA w sieci krystalicznej oddziałują pomiędzy sobą warstwowo, tworząc pseudonieskończone kolumny helis, równoległe do płaszczyzny a-c i nachylone pod kątem około 45° do osi a i c (Rysunek 17).



Rysunek 17. Upakowanie cząsteczek w komórce elementarnej struktury dupleksu $(GCUGCUGC)_2$. Kolorem zielonym zaznaczono dupleks składający się z nici C+D, niebieskim A+B, czerwonym nić E a jej symetryczny odpowiednik E' kolorem różowym.

3.1.4. Udokładnianie struktury $(GCUGCUGC)_2$

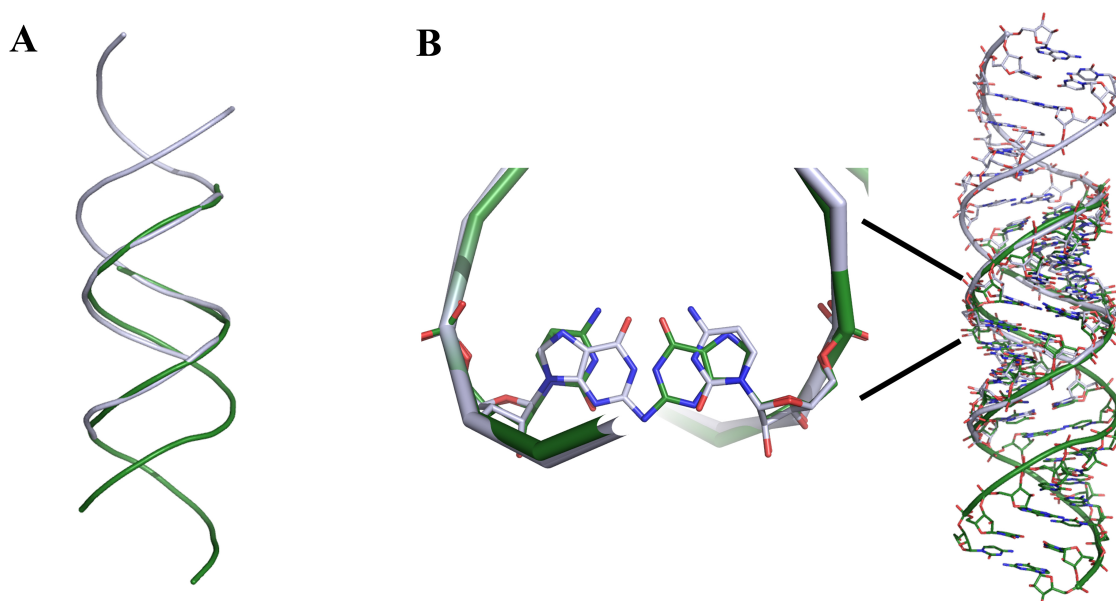
Proces udokładniania prowadzony był głównie za pomocą programu REFMAC5 (113), natomiast do inspekcji map i przebudowy modelu zastosowałam program COOT (114). Cząsteczki wody dodawane były przy użyciu programu ARP/wARP (115), który pracował w trybie domyślnym. Jakość i ilość danych pozwoliła na przypisanie anizotropowych czynników temperatury wszystkim atomom modelu. W mapie gęstości elektronowej zidentyfikowałam dwa jony siarczanowe oraz jedną cząsteczkę glicerolu. Ostatnie cykle udokładniania prowadziłam za pomocą programu SHELXL (116), który udokładniał obsadzenia cząsteczek ligandów oraz części łańcuchów RNA posiadające alternatywne konformacje. Dodatkowo zwiększyłam ilość danych poprzez włączenie wszystkich refleksów, również tych stosowanych wcześniej do obliczenia czynnika R_{free} . Statystyki zebrane są w Tabeli 5.

Tabela 5. Zestawienie danych dotyczących procesu udokładniania struktur zawierających powtórzenia CUG.

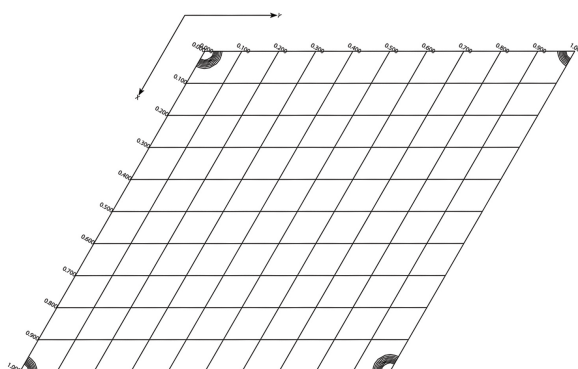
Dupleks	(GCUGCUGC) ₂	[(CUG) ₆] ₂ wg Mooers <i>i in.</i>	[(CUG) ₆] ₂ po rozdzieleniu
Średnia wartość czynnika B (Å ²)	22,8	28	33,8
Liczba refleksów R/R _{free}	30062/1602	9965/753	9925/1084
R (%)	14,8	21,8	21,9
R _{free} (%)	18,4	27,9	26,2
Liczba atomów cząsteczek RNA	934	1 500 (obsadzenie połówkowe)	750
Liczba cząsteczek wód	194	81 (obsadzenie połówkowe)	53
Liczba cząsteczek jonu siarczanowego	2	-	-
Liczba cząsteczek glicerolu	1	-	-
Rmsd na wiązaniach/wartość docelowa (Å)	0,018/0,21	0,011/0,21	0,01/0,02
Rmsd na kątach /wartość docelowa	2,75/3,0°	2,06/3,0°	0,028/0,04 Å
Kod PDB	3GLP	1ZEV	3GM7

3.1.5. Struktura dupleksu [(CUG)₆]₂

Pierwotnie struktura dupleksu [(CUG)₆]₂ (struktura romboedryczna) została zdeponowana w PDB (kod 1zev) przez Mooers *i in.* (98). Statystyki danych krystalograficznych przedstawione są w Tabeli 5. Model ten składa się z dwóch cząsteczek, które względem siebie są przesunięte i obrócone (Rysunek 18). Autorzy tłumaczyli ten fakt istnieniem statycznego nieuporządkowania (ang. *static disorder*). Jednakże grupa przestrzenna R3, sposób nakładania się cząsteczek a także wnikliwa inspekcja mapy gęstości elektronowej sugerowała, że przyczyna może być inna – obecność bliźniaczenia. W celu weryfikacji tej hipotezy przeanalizowałam czynniki struktury za pomocą programu PHENIX.XTRIAGE (117). Detekcja bliźniaczenia opiera się na fakcie, iż w jego wyniku rozkład intensywności refleksów jest zbliżony do wartości średniej. Oznacza to, że czynniki struktury przyjmują mniej ekstremalne wartości. Istnieje kilka dostępnych testów na wykrycie bliźniaczenia takich jak: analiza rozkładu statystycznego intensywności, L-test, H-test (118). Na ich podstawie nie stwierdzono bliźniaczenia, ale na mapie Pattersona (*vide* Materiały i Metody: 5.2.27) zlokalizowano dodatkowy pik, który stanowił 68% wysokości pików w początku układu współrzędnych (ang. *origin peak*) (Rysunek 19). Wskazywało to na obecność pseudotranslacji, która była rezultatem okresowego charakteru helisy kwasu nukleinowego. Pseudotranslacja ma również wpływ na czynniki struktury. Powoduje zwiększenie liczby refleksów o dużej i małej intensywności, co może maskować efekt bliźniaczenia. Dodatkowym istotnym faktem było to, że analiza wykazała wyższą symetrię sieci krystalicznej R32 niż ta, która została nadana.



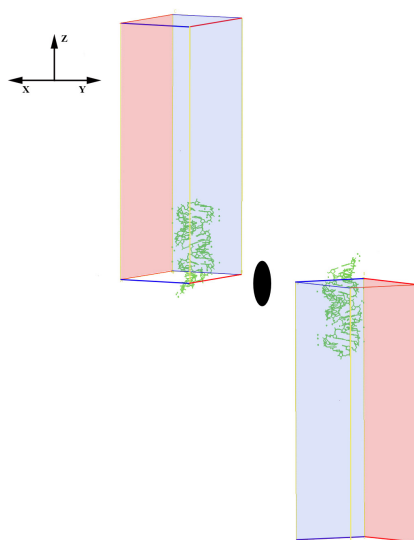
Rysunek 18. Zdeponowany model dupleksu $[(CUG)_6]_2$ (kod 1zev) w PDB przez Mooers i in. Dwa dupleksy (szary i zielony) nakładają się na siebie tak, że ich szkielety fosforanowo-cukrowe (A) pokrywają się niemal idealnie przy czym rodzaj nakładających się reszt jest inny. Przykładowo na parę G-C (szary) nakłada się para C-G (zielony) (B).



Rysunek 19. Przekrój przez mapę Pattersona, na której znajduje się silny pik o współrzędnych $(0,00, 0,00, 0,020)$. Pik zlokalizowany jest 28 \AA od pików w początku układu współrzędnych w kierunku osi z, co odpowiada długości jednego skreću helisy ($10,8$ par zasad).

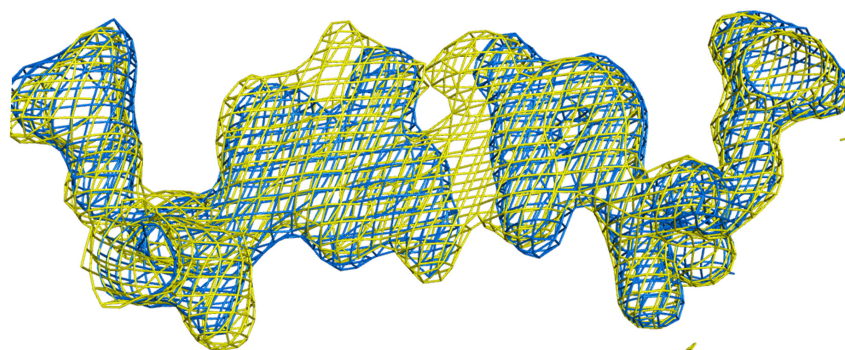
Powyższe wyniki sugerowały, że możliwe jest bliźniaczenie, w którym dwie domeny obrócone są względem siebie o 180° wzdłuż osi prostopadłej do osi c (Rysunek 20). Początkowy stopień bliźniaczenia $0,42$ oraz operator bliźniaczenia $(k \ h \ -l)$ wyznaczyłam za pomocą serwera Yeates and Fam UCLA (<http://www.doe-mbi.ucla.edu/Services/Twinning>) (119). W wyniku procesu udokładniania struktury stopień bliźniaczenia osiągnął wartość $0,5$. W takim przypadku niemożliwe jest algebraiczne wydzielenie czynników struktury i w związku z tym udokładnianie musi być prowadzone względem zblizniczonych danych. Programy SHELXL, PHENIX i REFMAC5 mają opcje udokładniania względem takich danych.

Do udokładniania struktury $[(CUG)_6]_2$ wykorzystałam program SHELXL. Jako model początkowy wybrałam jedną z dwóch helis (nazwaną przeze mnie G+H). Mapy gęstości



Rysunek 20. Schemat przedstawiający sposób bliźniaczenia struktury $[(CUG)_6]_2$. Domeny powiązane są ze sobą za pomocą operatora bliźniaczenia (czarna elipsa), który znajduje się prostopadle do kierunku osi z . Za jego pomocą (obrót o 180°) można przekształcić jedną domenę w drugą.

elektronowej obliczane były przez program COOT na podstawie pliku *.fcf generowanego przez SHELXL. Plik ten zawiera „rozblźniaczone” na podstawie udokładnionego modelu czynniki struktury z przypisanymi odpowiednimi fazami. Ostatecznie, mapa $2F_o - F_c$ jest jednoznaczna (Rysunek 21) a wartości czynników R zadowalające (Tabeli 5).



Rysunek 21. Dwie mapy gęstości elektronowej $2F_o - F_c$ obliczone dla zblźniaczonej (żółty) oraz rozblźniaczonej struktury (niebieski).

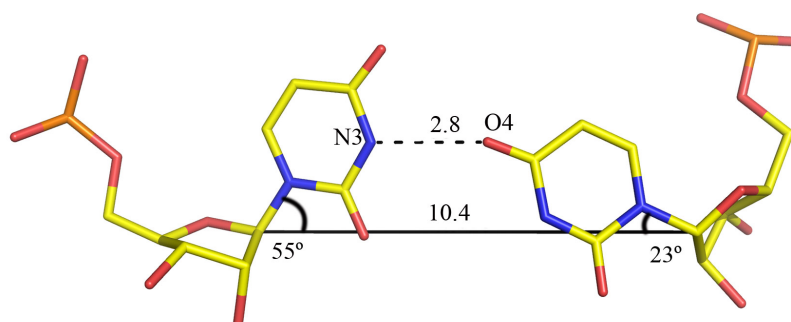
3.1.6. Konformacja dupleksów RNA

Zarówno w strukturze $(GCUGCUGC)_2$ (jednoskośnej) jak i $[(CUG)_6]_2$ romboedrycznej dupleksy przyjmują formę A-RNA, w których reszty cukrowe przyjmują konformację C3'-endo lub w niektórych przypadkach C2'-endo. Parametry helikalne zostały obliczone na podstawie wektorów niezależnych od sekwencji wyznaczonych przy użyciu pozycji atomów C1' sparowanych reszt nukleotydowych (szczegóły w rozdziale Materiały i Metody: 5.2.29.). Przemieszczenie (ang. *displacement*), kąt nachylenia (ang. *angle*) oraz odległość między płaszczyznami zasad (ang. *rise*)

wyznaczone pomiędzy wektorem atomów C1' a osią helisy, pomimo obecności niestandardowych par, przyjmują wartości nieodbiegające od typowych (120) (Tabela 6). Jedyne kąt skrętu helisy (ang. *twist*) dupleksu A+B struktury jednoskośnej przyjmuje rozrzucone wartości (odchylenie standardowe = $8,1^\circ$). Największe wartości, powyżej 40° , wyznaczone są pomiędzy parami G-C a U-U. W przypadku pozostałych dupleksów wartości kąta skręcenia nie wykazują takiej zmienności. Dla dupleksu C+D odchylenie standardowe wynosi $3,6^\circ$, dla E+E' $2,6^\circ$ a dla G+H $3,2^\circ$. Z kolei wartości średnie dla wszystkich helis są bardzo zbliżone, $32\text{--}34^\circ$, a zatem typowe dla formy A-RNA. W efekcie wszystkie dupлексы są bardzo podobne do siebie. Wartości rmsd wynoszą $0,9\text{--}1,4 \text{ \AA}$ dla cząsteczek struktury jednoskośnej oraz $1,0\text{--}1,7 \text{ \AA}$ dla struktury rozblźnianej porównywanej z modelem (GCUGCUGC)₂.

3.1.7. Oddziaływania pomiędzy parami zasad

Wszystkie pary G-C tworzą typowe pary Watsona-Cricka. Z kolei pary U-U oddziałują ze sobą jednym wiązaniem wodorowym utworzonym pomiędzy grupą karbonylową O4 a grupą aminową N3. Ta reszta urydyny, która jest akceptorem wiązania wodorowego, nachylona jest w kierunku bruzdy małej. Ma to odzwierciedlenie w wartości kąta λ , który wyznaczany jest pomiędzy wiązaniem N-glikozydowym a prostą łączącą atomy C1' sparowanych nukleotydów. Dla „nachylonej reszty urydyny” wynosi on jedynie 30° . Dla porównania, typowa wartość to około 55° (Rysunek 22). Dystans pomiędzy atomami C1' ($10,4 \text{ \AA}$) par U-U jest mniejszy tylko o $0,1 \text{ \AA}$ od średniej wartości ($10,5 \text{ \AA}$) dla par G-C. Kąt otwarcia pary (ang. *opening*) U-U jest równy $23,6^\circ$ (wartość średnia) (Tabela 6). Powyższe właściwości są obserwowane w obu strukturach, dla wszystkich niestandardowych par. Na podstawie nomenklatury przedstawionej przez Leontisa i Westhofa (121) takie parowanie można opisać jako U *cis* (*wobble*) W+C+/W+C+, z zaznaczeniem, iż pomiędzy resztami jest jedno wiązanie wodorowe zamiast dwóch możliwych (*vide* Dyskusja: 4.1). Parę tą można opisać również jako „rozciągnięta para U-U” (ang. „*stretched U-U wobble*”).

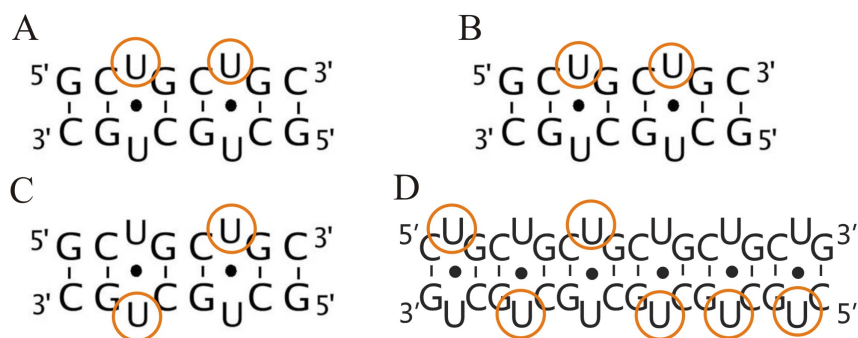


Rysunek 22. „Rozciągnięta para U-U”. Pomiedzy parą utworzone jest jedno wiązanie wodorowe (przerwana linia). Jedna z reszt urydynowych nachylona jest w kierunku bruzdy dużej, co odzwierciedla kąt λ , który wynosi 23° . Odległość pomiędzy atomami C1' (czarna linia) jest nieznacznie mniejsza od średniej wartości równej $10,5 \text{ \AA}$.

Tabela 6. Parametry helikalne wyznaczone dla struktur dupleksów zawierających powtórzenia CUG. Szczegółowy opis parametrów przesunięcie (ang. displacement), kąt (ang. angle), kąt skrętu helisy (ang. twist) oraz odległość między płaszczynami zasad (ang. rise) zostały zdefiniowane w Materiałach i Metodach: 5.2.29. Odchylenie standardowe oznaczono skrótem s.d.

Para	Przemieszczenie (Å)			Kąt (°)			Kąt skrętu helisy (°)			Odległość między płaszczynami zasad (Å)			Opening (Å)							
	A+B	C+D	E+E'	G+H	A+B	C+D	E+E'	G+H	A+B	C+D	E+E'	G+H	A+B	C+D	E+F	G+H				
G-C	6,8	5,9	6,8		14,4	15,7	12,5		27,1	33,5	34,4		3,3	2,5	2,7		-1,1	2,1	-2,3	
C-G	6,3	6,5	7,2	6,8	14,6	14,2	10,2	20,5	44,8	34,9	33,7	39,0	2,8	2,3	2,4	2,0	0,1	-1,2	-1,0	-13,2
U-U	6,5	6,6	7,1	6,8	14	15,3	11,3	18,2	33,8	34,6	35	31,9	2,2	2,9	2,7	2,4	-32	-34	-31	-21,2
G-C	5,2	6,2	6,4	6,9	15	15,7	14,1	13,8	22,3	33,2	27,7	33,3	2,7	2,9	2,9	2,7	-2,1	-0,5	0,5	-6,8
C-G	5,8	6,4	6,4	6,8	11,4	15	14,1	13,1	40,9	30,2	35	31,7	3,2	2,4	2,7	2,8	0,1	-4,4	0,5	0,8
U-U	7,1	6,4	7,1	6,4	7,4	15,3	11,3	12,5	26,5	41,4	33,7	35	2,9	2,3	2,4	2,6	-16	-21	-31	-22,1
G-C	7,1	6,3	7,2	5,7	9,9	14,7	10,2	14,9	31,5	31,8	34,4	35,1	2,4	2,8	2,7	2,4	1,3	1,9	-1,0	-3,5
C-G	6,4	7,0	6,8	5,6	14,1	12,1	12,5	17,8				36,3				2,4	0,6	-1,5	-2,3	-1,5
U-U				6,1				14,2				34,9				2,7				-21
G-C				6,8				12,5				31,1				2,7				-0,7
C-G				7,5				7,7				32,9				3,3				1,7
U-U				7,8				8,0				33,6				2,8				-11,9
G-C				6,6				13,3				27,7				2,4				-5,1
C-G				5,9				18,3				29,3				2,3				1,9
U-U				7,0				14,9				34,5				2,7				-28
G-C				7,9				13				32,4				2,7				-2,2
C-G				8,1				12,9				32,3				2,6				2,0
U-U				7,7				16,1				40,5				2,3				-14,8
G-C				7,2				18,1												1,3
średnia	6,4	6,4	6,9	6,9	12,6	14,7	12,0	14,4	32,4	34,2	33,4	33,6	2,8	2,6	2,7	2,6	-6,1	-7,4	-8,4	-8,0
s.d.	0,6	0,3	0,4	0,8	2,8	1,2	1,6	3,4	8,1	3,6	2,6	3,2	0,4	0,3	0,2	0,3	11,8	13,2	13,7	9,8

W jednym powtórzeniu CUG można rozpatrywać dwie różne konformacje pary U-U w zależności od tego, która z reszt urydynowych jest nachylona w kierunku bruzdy małej (Rysunek 23). W duplesie A+B oraz C+D w dwóch parach U-U „nachylone urydyny” znajdują się na tej samej nici, odpowiednio A i D. W przypadku duplesu E+E' tylko druga U od końca 5' jest nachylona w każdej nici. W strukturze romboedrycznej w nici G są tylko dwie nachylone reszty urydynowe: pierwsza i trzecia, pozostałe cztery znajdują się na łańcuchu H.

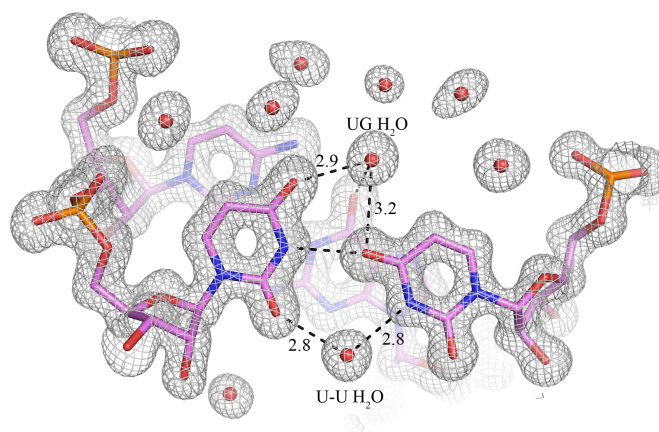


Rysunek 23. Dystrybucja „nachylonych” reszt urydynowych (pomarańczowe koła) w strukturze $(GCUGCUGC)_2$ (dupleks A+B—panel A, C+D—panel B i C+D—panel C) oraz $[(CUG)_6]_2$ (D).

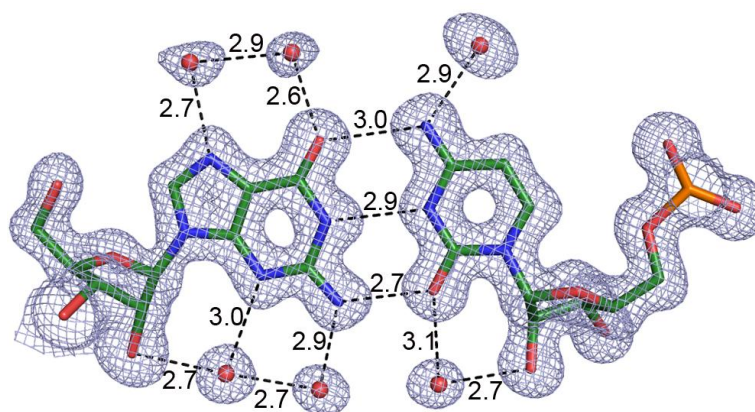
3.1.8. Hydratacja RNA i oddziaływania z ligandami

Pary U-U oddziałują z dwiema charakterystycznymi cząsteczkami wody (Rysunek 24). Jedną z nich, nazwaną „wodą U-U”, zlokalizowaną została w bruzdzie małej. Oddziałuje z grupą aminową N3 nachylonej urydyny oraz z atomem O2 drugiej reszty U. To charakterystyczne oddziaływanie obserwowane jest dla wszystkich sześciu par U-U struktury jednoskośnej i dla czterech par U-U w strukturze romboedrycznej. Druga cząsteczka wody („woda UG”) znajduje się w bruzdzie dużej i tworzy wiązanie wodorowe z grupą karbonylową O4 nienachylonej reszty U oraz z tlenem karbonylowym O6 sąsiadującej reszty guanozyny znajdującej się na przeciwnej nici. W strukturze jednoskośnej cząsteczka „wody UG” jest przy każdej parze U-U, natomiast w romboedrycznej przy trzech z sześciu par U-U.

Hydratacja par C-G jest również charakterystyczna (Rysunek 25). W wysokorozdzielczej strukturze jednoskośnej, większość reszt guanozynowych oddziałuje z czterema cząsteczkami wody. Dwie z nich znajdują się w bruzdzie dużej, przy czym jedna z nich tworzy wiązanie wodorowe z grupą N7 a druga z O6. W bruzdzie małej cząsteczki wody oddziałują z grupami: pierwsza z *exo*-aminową, druga z iminową. Z kolei reszty cytydynowe są typowo hydratowane przez dwie cząsteczki wody, po jednej w każdej bruzdzie. W trzech przypadkach grupa *exo*-aminowa reszt cytydynowych zamiast z wodą oddziałuje z jonem siarczanowym lub z cząsteczką glicerolu.



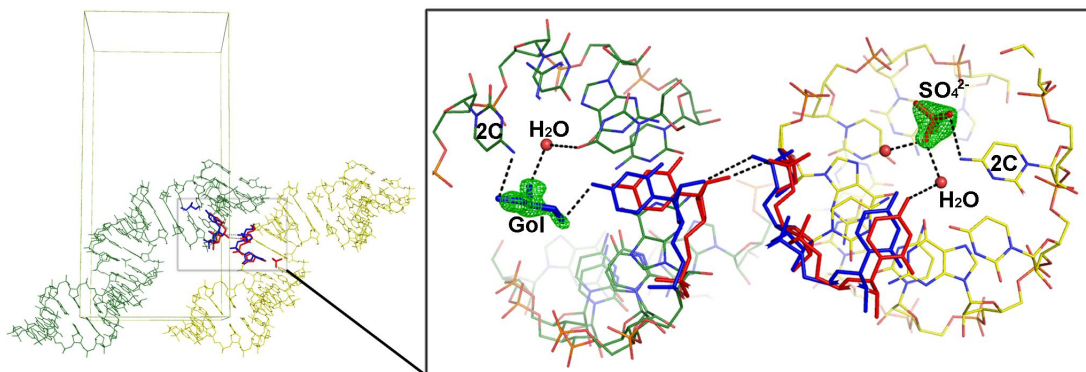
Rysunek 24. Hydratacja pary U-U. W bruzdce dużej znajduje się „woda UG”, która oddziałuje z resztami urydynowymi oraz sąsiadującą resztą G. W bruzdce małej „woda U-U” tworzy wiązania wodorowe z obiema resztami U. Mapa gęstości elektronowej $2F_o-F_c$ ma kontur 1σ .



Rysunek 25. Hydratacja pary G-C. W bruzdce dużej i małej znajdują się po trzy cząsteczki wody, z czego dwie oddziałują z resztą guanozyny a jedna z resztą cytydyny. Mapa gęstości elektronowej $2F_o-F_c$ ma kontur 1σ .

Oba ligandy zlokalizowane w strukturze $(GCUGCUGC)_2$ znajdują się w bruzdce dużej. Anion siarczanowy oddziałuje z nicią C+D natomiast glicerol z A+B. Każdy z nich tworzy dwa takie same wiązania wodorowe: z grupą aminową reszty 2C (nić A dla glicerolu lub nieć D dla jonu siarczanowego) oraz z „wodą UG” pary 3U-6U. Oba ligandy mają obsadzenie połówkowe i oddziałują z jedną z alternatywnych konformacji nieuporządkowanego łańcucha RNA (glicerol z nicią B, anion siarczanowy z nicią C). W sieci krystalicznej obie nici oddziałują ze sobą a nieuporządkowanie konformacyjne jest skorelowane z rodzajem związanego ligandu. W rezultacie, albo cząsteczka jonu siarczanowego wiąże się do RNA, albo glicerol (Rysunek 26). Dodatkowo trzecia grupa OH glicerolu oddziałuje z grupą *exo*-aminową reszty 5C łańcucha B, spinając obie reszty znajdujące się na przeciwnych niciach. Odległość pomiędzy atomami azotu tych reszt (2C nici A i 4D nici B) w konformacji preferowanej dla wiązania glicerolu wynosi 5,8 Å. W chwili, gdy

wiązany jest siarczan i następuje przesunięcie nici odległość ta wzrasta do 6,9 Å. Dla porównania średni dystans dla innych ekwiwalentnych par wynosi 6,6 Å.

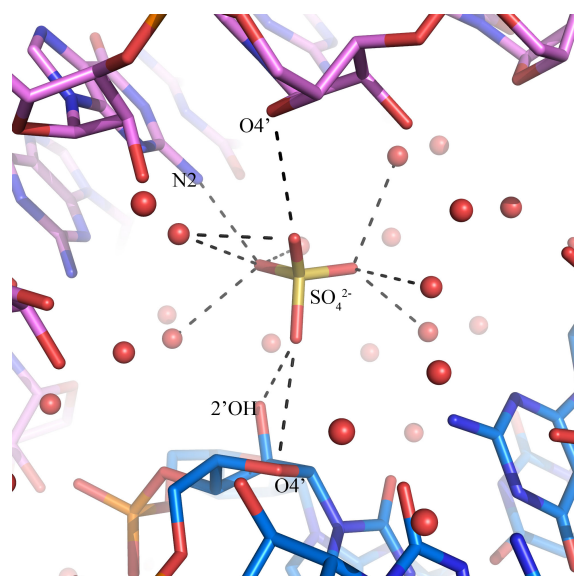


Rysunek 26. Oddziaływania ligandów z strukturze jednoskośnej. Dwa łańcuchy RNA tworzą kontakty w sieci krystalicznej (przerywana linia). Wiązanie glicerolu (Gol) lub jonu siarczanowego powiązane jest z lokalnym nieuporządkowaniem dwóch sąsiadujących łańcuchów RNA. Oba ligandy mają obsadzenie połówkowe, więc albo cząsteczka glicerolu jest wiązana (konformacja zaznaczona na niebiesko) albo jon siarczanowy (konformacja zaznaczona na czerwono). Na rysunku pokazana jest na zielono mapa typu „omit” (kontur 3σ).

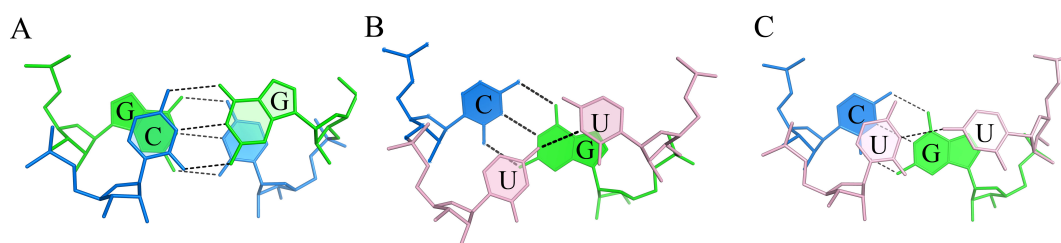
W strukturze jednoskośnej zlokalizowano także drugą cząsteczkę anionu siarczanowego. Znajduje się ona pomiędzy dupleksem A+B i jego symetrycznym dupleksem A'+B', w przestrzeni utworzonej przez pierścienie cukrowe (Rysunek 27). Dwa atomy tlenu jonu siarczanowego oddziałują z każdym atomem O2' rybozy: reszt 3U nici B i 2C nici A. Trzeci atom tlenu jonu SO_4^{2-} tworzy wiązanie wodorowe z grupą N2 reszty 1G nici A.

3.1.9. Oddziaływania warstwowe

Wyróżniono trzy rodzaje oddziaływań warstwowych: dwa pomiędzy parami CU/UG i jedno dla par GC/GC (Rysunek 28). W przypadku tych ostatnich powierzchnie pierścieni zasad w znacznym stopniu nachodzą na siebie. Jest to typowe dla sąsiadujących par typu Watsona-Cricka zawartych w helisie A-RNA. Pomiedzy parą U-U a sąsiednią parą typu Watsona-Cricka obserwuje się ograniczone oddziaływania warstwowe. We wszystkich przypadkach uracyl znajduje się nad pięciocząłowym pierścieniem sąsiedniej guanozyny, natomiast położenie zasad reszt urydynowych względem cytydynowych zależy od jej konformacji. W przypadku, gdy U jest nachylona w kierunku bruzdy małej, nie obserwuje się oddziaływań z C, lecz z niewielką częścią sześciocząłowego pierścienia G z przeciwnej nici. W przypadku, gdy urydyna nie jest nachylona, jej zasada znajduje się pomiędzy pierścieniami sąsiedniej pary G-C, z którymi tworzy słabe oddziaływania warstwowe.



Rysunek 27. Jon siarczanowy wiąże się w przestrzeni utworzonej przez bruzdy małe dwóch symetrycznych cząsteczek RNA struktury jednoskośnej i oddziałuje bezpośrednio z pierścieniami cukrowymi reszt nukleotydowych oraz grupą *exo*-aminową guanozyny. Tworzy też liczne wiązania wodorowe (przerwane linie) z cząsteczkami wody (czerwone kule).

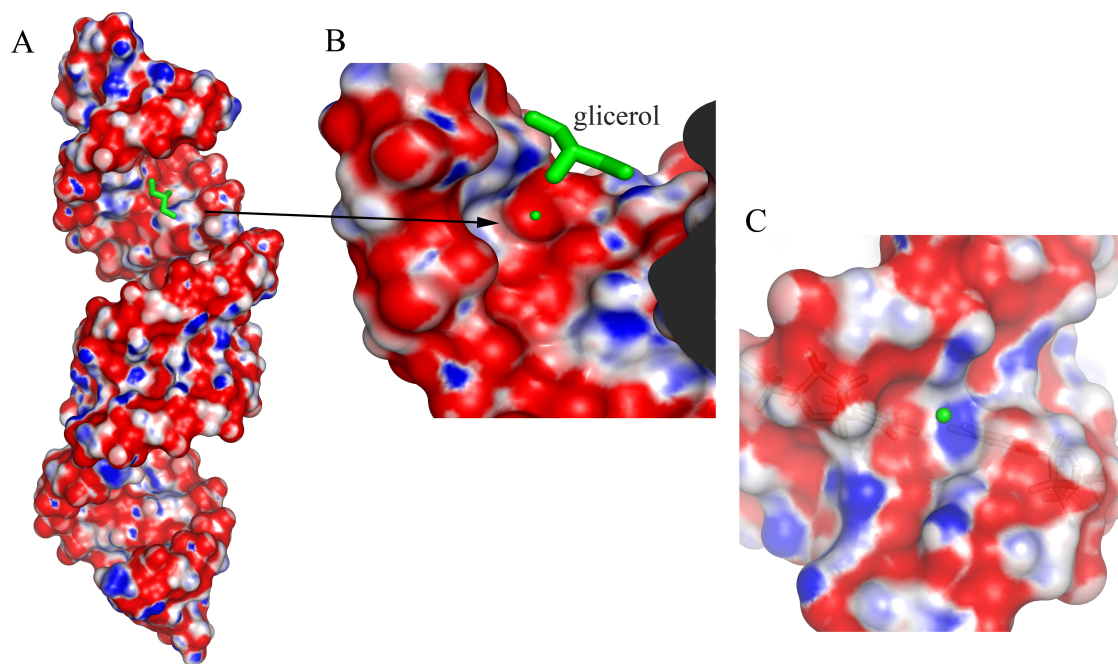


Rysunek 28. Trzy typy oddziaływań warstwowych obserwowane pomiędzy parami GC/GC (A) oraz pomiędzy parami CU/UG (B, C) w zależności od konformacji pary U-U.

3.1.10. Powierzchnia potencjału elektrostatycznego

Rozkład potencjału elektrostatycznego dla wszystkich dupleksów jest podobny. Bruzda duża wykazuje niejednorodną dystrybucję obu potencjałów z przewagą ujemnego. Dodatni potencjał jest generowany przez grupy *exo*-aminowe reszt cytydynowych oraz nieznacznie przez krawędzie Hoogsteena reszt urydynowych. Ponadto dodatni potencjał posiadają odsłonięte po stronie 3' pierścienie cytozyn z powodu braku oddziaływań warstwowych z sąsiadującą resztą urydyny. Są to m.in. reszty 2C dla nici A i D, które są zaangażowane w oddziaływania odpowiednio z cząsteczką glicerolu i jonom siarczanowym. „Woda UG”, z którą oddziałują ligandy znajduje się we wgłębieniu o potencjale ujemnym. Zostało ono utworzone poprzez wypchnięcie nachylonej reszty urydyny w kierunku bruzdy mniejszej i odsłonięcie reszty cytydyny, która w konsekwencji tworzy jedną ze ścian wgłębienia (Rysunek 29). Rozkład potencjału elektrostatycznego w bruzdzie małej tworzy charakterystyczny wzór. Wzdłuż osi helisy przeplatają się pasy o potencjale dodatnim i ujemnym. Prążki o potencjale dodatnim są utworzone przez elektrododatnie atomy warstwowo ułożonych reszt C, G i U. Z kolei negatywne pasma generowane są przez grupy karbonylowe reszt

cytydynowych i nachylonych reszt urydynowych. Architektura pary U-U sprawia, że w bruzdzie małej jest również wgłębienie ale znacznie płytsze i obejmujące obie reszty. Jedną ze ścian stanowi eksponowany na zewnątrz karbonylowy atom tlenu O2. Wnęka ma dwa obszary: jeden o potencjale ujemnym (atom O2 nienachylonej U) a drugi o potencjale dodatnim (N3 nachylonej U). W jej obszarze znajduje się „woda U-U”, która swoją sferą van der Waalsa obejmuje większość zagłębienia o potencjale dodatnim.



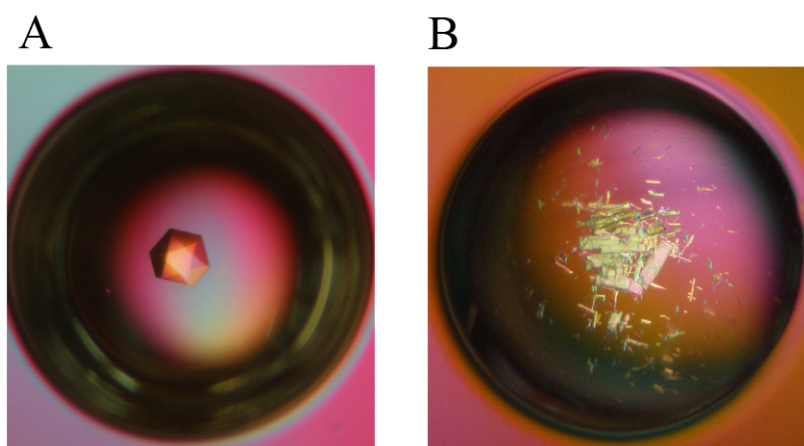
Rysunek 29. Powierzchnia potencjału elektrostatycznego trzech kolejnych dupleksów struktury jednoskośnej. Helisa ma głównie potencjał ujemny (czerwony) z obszarami o potencjale dodatnim (niebieski) (A). Glicerol (zielony) wiąże się w bruzdzie dużej i oddziałuje z grupami funkcyjnymi RNA wykazującymi potencjał dodatni. „Woda UG” (zielona kula) znajduje się w zagłębieniu o potencjale ujemnym (B). W bruzdzie małej utworzone jest wgłębienie, w którym zlokalizowana jest „woda U-U” (zielona kula) (C).

3.2. Struktury RNA zawierające powtórzenia CAG

Określiłam strukturę dwóch form krystalicznych tego samego oligomeru o sekwencji GGCAG-CAGCC, który zaprojektowałam, oczyściłam i wykryzowałam. Pierwsza z form to wysokorozdzielcza (0,95 Å) struktura romboedryczna, natomiast druga to struktura trygonalna określona z rozdzielczością 1,9 Å. Dupleks posiada na końcach dodatkową parę G-C. Została ona wprowadzona ze względu na to, iż próby krystalizacji oligomeru GCAGCAGC zakończyły się niepowodzeniem. Dodatkowe pary miały za zadanie zwiększyć stabilność dupleksu lub/i jednoznacznie określić jego strukturę.

3.2.1. Krystalizacja dupleksu (GGCAGCAGCC)₂

Liofilizowany oligomer został rozpuszczony w 100 mM KCl (stężenie końcowe RNA wynosiło 1 mM), następnie inkubowany przez 10 minut w temperaturze 65 °C i schładzany powoli do temperatury pokojowej przez 2–3 godziny. Forma romboedryczna wykryła się po upływie roku w 25 mM MgSO₄, 50 mM Tris-HCl o pH 8,5 i 1,8 M (NH₄)₂SO₄ w temperaturze 19 °C. Otrzymałam tylko jeden kryształ o wymiarach 0,4 x 0,4 x 0,5 mm (Rysunek 30). Kryształ formy trygonalnej otrzymano w 10 mM MgSO₄, 50 mM kakodylan-NaOH pH 6,5 i 2M (NH₄)₂SO₄. Kryształy wykryły się w temperaturze 30 °C w przeciągu 2–3 tygodni po czym przeniesiono je do temperatury 19 °C. Stosowano metodę dyfuzji par w układzie wiszącej kropli, która początkowo zawierała 2 μl RNA oraz 2 μl roztworu rezerwuarowego.



Rysunek 30. Kryształy dupleksu (GGCAGCAGCC)₂ formy romboedrycznej (A) oraz formy trygonalnej (B).

3.2.2. Rejestracja i przetwarzanie danych

Dane struktury wysokorozdzielczej zostały zarejestrowane na linii BL14.1 w ośrodku synchrotronowym BESSY w Berlinie, natomiast struktury trygonalnej na linii EMBL-X13 w ośrodku DESY w Hamburgu. W obu przypadkach kryształ podczas pomiaru znajdował się w oparach azotu. Krioprotektantem był 20% glicerol. Kryształ formy romboedrycznej miał zdolność do rozpraszania z wyższą rozdzielczością niż zarejestrowano dane (stosunek wartości I/σ na ostatniej powłoce wynosi 7,7) (Tabela 7). Niestety była to możliwie najbliższa odległość detektora do kryształu (95 mm). Zarejestrowano dwa zbiory danych: nisko- i wysokorozdzielczych. W przypadku struktury o rozdzielczości 1,9 Å, symetria sieci została wyznaczona jako P3. Wygaszenia systematyczne wskazywały, że oś trójrotna była osią śrubową, w związku z tym podczas rozwiązywania struktury brano pod uwagę dwie możliwe grupy przestrzenne P3₁ oraz P3₂. Dane zostały przetworzone pakietem programów DENZO i SCALEPACK (Tabela 7) (122).

Tabela 7. Zestawienie statystyki danych dyfrakcyjnych kryształów formy romboedrycznej i trygonalnej. W nawiasie podano wartości dotyczące ostatniego zakresu rozdzielczości.

Forma krystaliczna	romboedryczna	trygonalna
Dupleks	(GGCAGCAGCC) ₂	(GGCAGCAGCC) ₂
Źródło promieniowania	BESSY BL 14.1	EMBL-X13
Temperatura (K)	100	100
Długość fali (Å)	0,8	0,8123
Grupa przestrzenna	R32:H	P3 ₂
Parametry komórki (Å)	a = b = 46,4 c = 82,7	a = b = 47,4 c = 83,2
Zakres rozdzielczości (Å)	23,2 – 0,95 (0,97–0,95)	18,6–1,9 (1,93–1,9)
R _{merge}	0,036 (0,291)	0,105 (0,518)
Liczba unikalnych refleksów	21756	16346
Kompletność (%)	99,9 (98,0)	99,8 (100)
Mozaikowość (°)	0,34	0,7
Krotność obserwacji	15,3 (11,0)	6,4 (6,2)
$\langle I/\sigma(I) \rangle$	74,1 (7,7)	18,2 (4,4)
Refleksy > 2σ (%)	94,4 (72,9)	81,6 (81,9)
Czynnik B z wykresu Wilsona (Å ²)	6,8	25,7

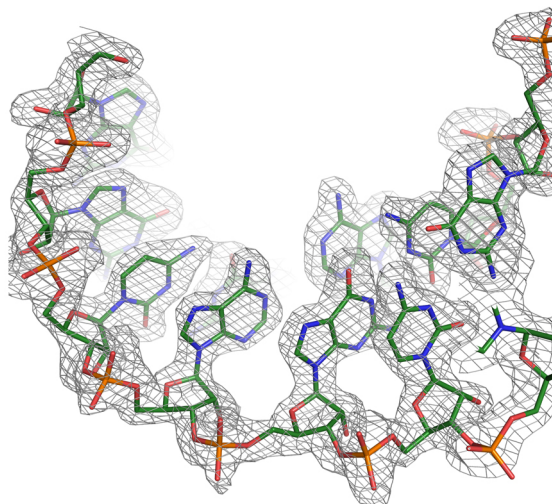
3.2.3. Rozwiązanie struktur zawierających powtórzenia CAG

Początkowo posiadałam zbiór danych formy trygonalnej, dla której wówczas nie rozwiązałam problemu fazowego z wykorzystaniem podstawienia cząsteczkowego. Stosowałam różne modele dupleksów oraz pojedyncze łańcuchy RNA, które posiadały odpowiednio zmutowane reszty nukleotydowe. Stosowałam różne programy: PHASER, MolRep (123), EPMP (124). Dokonałam również analizy czynników struktury z użyciem programu PHENIX.XTRIAGE. Dane były dobrej jakości, nie wykryto obecności bliźniaczenia, grupa punktowa była poprawna. Wykryto jedynie obecność pseudotranslacji.

W międzyczasie otrzymałam dane dyfrakcyjne formy romboedrycznej. Strukturę rozwiązałam za pomocą programu PHASER. Jako modelu użyłam odpowiednio zmutowany dupleks [GC(8-BrG)GCGGC]₂. W części niezależnej znajdowała się jedna nić.

Porównując parametry komórek formy romboedrycznej i trygonalnej okazało się, że są do siebie bardzo podobne, co sugerowało, że są to te same formy krystaliczne (Tabela 7). W celu potwierdzenia tego przypuszczenia przetworzyłam dane w nowej grupie R32 a następnie wykorzystałam model struktury wysokorozdzielczej do rozwiązania problemu fazowego. Próba zakończyła się niepowodzeniem, co jednoznacznie potwierdziło, że są to jednak dwie różne formy krystaliczne. Skłoniło mnie to do poszukiwania innej, „poprawnej” grupy przestrzennej. Dane

przeskalowałam w grupie $P3_212$ a następnie otrzymałam rozwiązanie – trzy nici w części niezależnej. Oznaczało to, że jeden z dupleksów znajduje się na krystalograficznej osi dwukrotnej i jego druga nić jest nicią symetryczną. Mimo dobrej mapy gęstości elektronowej (Rysunek 31) czynniki R osiągnęły wartość 33,5/43,7 i nie ulegały zmniejszeniu. Dodatkowo niemożliwe okazało się zbudowanie na podstawie mapy gęstości elektronowej „symetrycznego dupleksu” o poprawnej sekwencji. Wynikało to z faktu, że oś dwukrotna biegła przez środek pary a powinna być pomiędzy dwiema parami (Rysunek 32). Wróciłam do danych przetworzonych w grupie $P3_2$. Do rozwiązania struktury użyłam modelu z grupy $P3_212$, przy czym usunęłam nić symetrycznego dupleksu, ponieważ był on niekompletny. Program PHASER znalazł trzy dupleksy, dla których czynniki z-score i LLG wyniosły: pierwszy zorientowany dupleks—z-score = 5,9, LLG = 114; drugi zorientowany dupleks—z-score = 22,2, LLG = 661; trzeci zorientowany dupleks—z-score = 25,6, LLG = 1846 i było to poprawne rozwiązanie.



Rysunek 31. Fragment mapy gęstości elektronowej $2F_o-F_c$ (kontur 1σ) otrzymanej po rozwiązaniu struktury w grupie przestrzennej $P3_212$.



Rysunek 32. Struktura drugorzędowa dupleksu $(GGCAGCAGCC)_2$, który jest symetryczny. Oś dwukrotna (czerwony) znajduje się pomiędzy parami GC/GC.

3.2.4. Udokładnianie struktur RNA zawierających powtórzenia CAG

Początkowo udokładnianie prowadziłam przy użyciu programu REFMAC5, a następnie PHENIX. Do obliczania czynnika R_{free} wybrałam dla obu struktur około 1000 refleksów (5%). Czynniki temperaturowe struktury romboedrycznej udokładniane były anizotropowo. Do modelu dodałam również atomy wodoru z wiązaniami geometrycznymi (ang. *in riding position*). Ostatnie

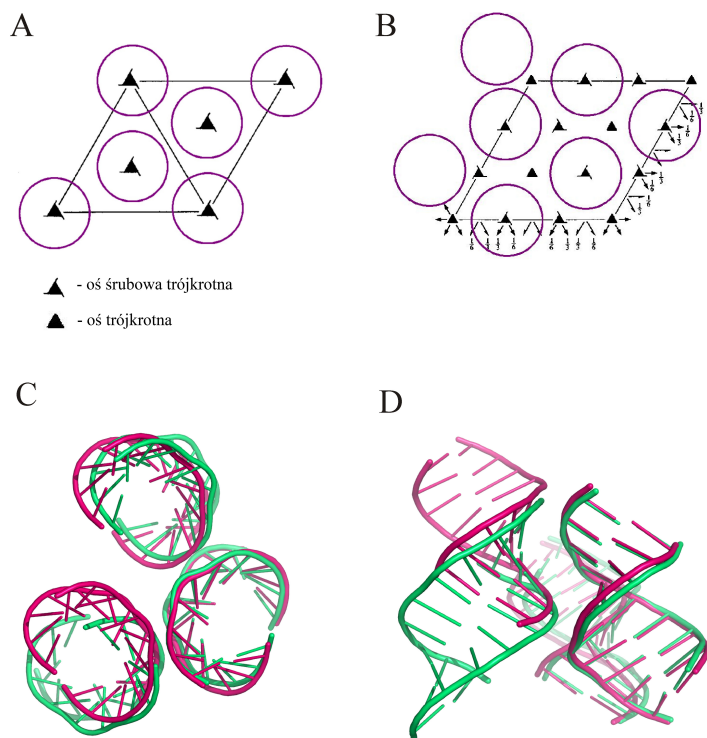
Tabela 8. Zestawienie danych dotyczących procesu udokładniania struktur zawierających powtórzenia CAG.

Forma krystaliczna	romboedryczna	trygonalna
Średnia wartość czynnika B (Å ²)	11,4	24,4
Liczba refleksów: work/test	21748	15413 /821
R (%)	10,6	21,17
R _{free} (%)	-	24,82
Liczba atomów cząsteczek RNA	213	1278
Liczba atomów cząsteczek wody	86	170
Liczba jonów siarczanowych	1	6
Rmsd na wiązaniach (Å)	0,016	0,006
Rmsd na kątach (°)	2	1,4

cykle prowadziłam z pełnym zbiorem refleksów bez stosowania jakichkolwiek więzów (ang. *unrestrained refinement*). Stosunek liczby danych do udokładnianych parametrów wyniósł 8. Struktura trygonalna była udokładniana z uwzględnieniem parametrów TLS zdefiniowanych dla poszczególnych łańcuchów RNA przy pomocy programu TLSMD (125). Szczegółowe zestawienie statystyk otrzymanych po procesie udokładniania znajduje się w Tabeli 8.

3.2.5. Upakowanie cząsteczek w sieci krystalicznej

Parametry komórki obu form są bardzo zbliżone jednakże ich grupy przestrzenne inne. Oznacza to, że dupleks (GGCAGCAGCC)₂ może upakować się w sieć krystaliczną trygonalną lub romboedryczną. W części asymetrycznej struktury romboedrycznej znajduje się jedna nić. Druga nić dupleksu jest generowana przez oś dwukrotną prostopadłą do helisy RNA. Oś helisy biegnie w tym samym miejscu co krystalograficzna oś śrubowa 3₂, natomiast pomiędzy dupleksami są osie trójrotne (Rysunek 33). W strukturze trygonalnej znajdują się trzy dupleksy: A+B, C+D, E+F. Trzy helisy są również owinięte wokół osi 3₂: A+B i E+F wokół osi biegnących wewnątrz komórki, C+D wokół osi znajdującej się na krawędzi komórki. W obu formach dupleksy leżą jeden nad drugim tworząc pseudonieskończoną helisę wzdłuż krawędzi c komórki. Po nałożeniu sieci krystalicznych obu form, kolumny RNA znajdują się w tym samym miejscu. Jednakże dupleks C+D formy trygonalnej jest przesunięty i obrócony o pół skrętu helisy (o 5 par zasad) w stosunku do odpowiadającego mu dupleksu formy romboedrycznej (Rysunek 33). Ta subtelna różnica powoduje, że dupleks (GGCAGCAGCC)₂ może krystalizować w dwóch różnych grupach przestrzennych. Wartości odchyleń rmsd pomiędzy atomami helis struktury wysoko i niskorozdzielczej wynoszą 0,9–1,0 Å, natomiast dla porównanych między sobą dupleksów formy trygonalnej wynoszą 0,1–0,5 Å.

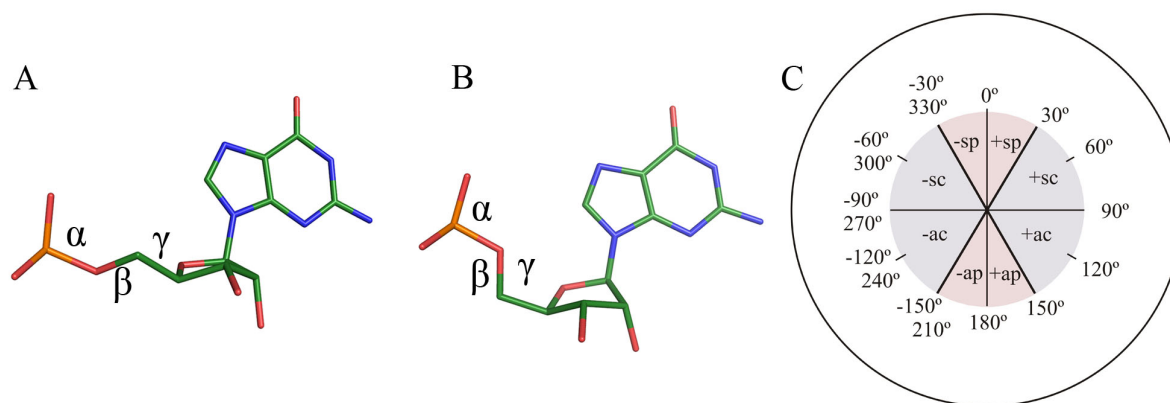


Rysunek 33. Schemat upakowania dupleksów $(GGCAGCAGCC)_2$ (fioletowe koła) wzdłuż osi helisy w komórce trygonalnej (A) i romboedrycznej (B). Po nałożeniu dwóch sieci na siebie wzdłuż kierunku z (C) helisy znajdują się w tym samym miejscu. W kierunku prostopadłym do osi z (D) jedna z helis jest przesunięta względem drugiej o połowę skreću helisy. Na różowo zaznaczono formę trygonalną a na zielono romboedryczną.

3.2.6. Konformacja helisy RNA oraz niekanoniczne pary A-A

Wszystkie dupleksy obu form krystalicznych przyjmują formę A-RNA. Pierścienie cukrowe są w konformacji C3'-endo oraz w dwóch przypadkach (reszty 5G nici A oraz 5G nici E) w C2'-exo. Parametr Z_p , który opisuje przesunięcie atomów fosforu sąsiednich par zasad od płaszczyzny xy lokalnego układu odniesienia wyznaczonego dla tych par wynosi 2,3–3,1 Å (126) (dla formy A-RNA powinien być większy od 1,5 Å). Wartości kątów torsyjnych są typowe dla A-RNA za wyjątkiem kątów α (określa rotację wokół wiązania P-O5') oraz γ (określa rotację wokół wiązania C5'-C4') nukleotydu 5G we wszystkich łańcuchach obu struktur (Tabela 9). Typowa konformacja kąta α opisywana jest jako *-gauche*-sc, natomiast dla reszty guanozyny piątej kąt znajduje się w następujących regionach konformacyjnych: *+ap*—łańcuch H, *+ac*—łańcuch A, C i E, *-ac* dla B, D i F. Kąt γ ma standardową konformację *+sc* jedynie dla łańcucha B. W pozostałych przypadkach jest: *+ap* w nici H, *-ap* nie A, C i E oraz *+ac* dla D i F. Może być to wizualizowane jako obrócenie atomu tlenu O5' w wyniku rotacji wiązania O5'-C5' (Rysunek 34).

W wyniku zmiany konformacji łańcucha fosforanowo-cukrowego, reszty rybozy 5G oraz 4A znajdują się niemal w tej samej płaszczyźnie (Rysunek 35). Odpowiadający im kąt skreću helisy, wyznaczony pomiędzy parami AG/CA, wynosi 18–22°. Jest to mała wartość w porównaniu do



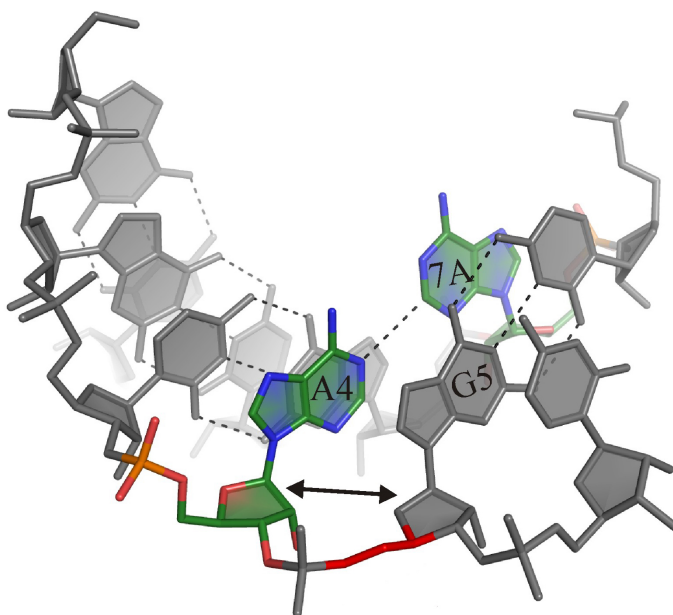
Rysunek 34. Dwie różne konformacje szkieletu fosforanowo-cukrowego reszty guanozyny: z rotacją wokół wiązania O5'-C5' (A) oraz najczęściej obserwowana w kwasach nukleinowych (B). Na panelu C znajduje się koło zakresu kątów torsyjnych i odpowiadających im nazw konformacji.

Tabela 9. Wartości kątów torsyjnych reszt 5G w dupleksach obu struktur krystalicznych. W nawiasie podano wartości odchylenia standardowego.

	α	β	γ	δ	ϵ	ζ	χ
5G	153,2	-174,4	179,5	81,2	-151,7	-71,1	-175,5
	135,2	-178,1	-160,7	92,9	-145,8	-78,3	-174,9
	-96,9	175,3	84	68,9	-159,2	-74,8	-174,6
	149,3	178,1	-171,7	90,9	-147,8	-74,5	-177
	-114	-176,4	97,4	67,8	-155,8	-79,1	-178,8
	-114,8	178	-163,9	97	-148,2	-70,2	-171,5
	138,6	-173,2	93,9	66,9	-160,7	-78,9	-179,3
	-169,9 (58,0)	181,3 (4,3)	148,4 (53,5)	80,8 (13,0)	-152,7 (5,9)	-75,3 (3,7)	-175,9 (2,7)

średniej wartości 31° pozostałych par dupleksu (Tabela 10). Wypadkowy rezultat jest taki, iż helisa jest „niedokręcona” (ang. *underwound*) co oznacza, że przypada 12,5–12,9 par zasad na jeden skręt. Dodatkowo bruzda duża rozszerza się do ponad 20 Å w środkowej części dupleksu (Tabela 11). Dystans pomiędzy atomami C1' sparowanych nukleotydów wynosi $10,7 \pm 0,2$ Å i jest typowy dla formy A-RNA. Jest on dłuższy tylko o 0,3 Å dla par A-A.

Pary C-G tworzą typowe oddziaływania typu Watsona-Cricka. Natomiast pary A-A są nietypowe. Wszystkie reszty adenozynowe są w konformacji *anti* w stosunku do pierścienia cukrowego. Pomiedzy nimi jest jedno wiązanie wodorowe utworzone przez atom wodoru węgla C2 oraz atom azotu N1 (C2-H2...N1) (Rysunek 36). W wysokorozdzielczej strukturze, długość wiązania C2...N1 wynosi 3,4 Å, a gdy włączony jest atom wodoru z wiązaniami geometrycznymi, wynosi 2,4 Å, zatem o 0,3 Å bliżej niż wynosi suma ich promieni van der Waalsa. Wszystkie trzy atomy C, H i N leżą w jednej płaszczyźnie a kąt pomiędzy nimi ma 176° . Podobnie jest w strukturze trygonalnej gdzie atomy C2 oraz N1 oddalone są od siebie o 3,1–3,4 Å. We wszystkich parach A-A reszta adenozyny znajdująca się bliżej końca 3' jest przesunięta w kierunku bruzdy dużej, co skutkuje wartością kąta λ równą 87° (wartość średnia). Wystaje ona ponad inne reszty i wygląda jakby pokazywała znak „kciuk do góry” (ang. *thumbs up*) (Rysunek 35). Druga adenozyna z pary



Rysunek 35. Para A-A (zielony) i jej otoczenie strukturalne w strukturze romboedrycznej. Kąty torsyjne α i γ (czerwony) reszty 5G przyjmują niestandardowe wartości, co skutkuje rozkręceniem helisy. Taka konformacja łańcucha fosforanowo-cukrowego powoduje, że reszty 4A i 5G znajdują się niemal w tej samej płaszczyźnie (strzałka). Efekt ten jest obserwowany w każdej nici obu struktur i powiązany jest z resztą adenozyńską mniej nachyloną w kierunku bruzdy dużej.

Tabela 10. Parametry helikalne dupleksów $(GGCAGCAGCC)_2$. Odchylenie standardowe oznaczono skrótem s.d.

Para	Przemieszczenie (Å)				Kąt (°)				Kąt skrętu helisy (°)				Odległość między płaszczyznami zasad (Å)			
	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F
G-C	7,8	8,8	8,8	8,3	13,1	10,8	11,8	13,1	30,0	31,3	32,9	33,2	2,7	2,8	2,7	2,7
G-C	8,7	10,0	10,3	9,8	8,3	7,4	8,4	8,7	30,9	30,6	31,5	32,1	3,0	3,1	3,1	3,1
C-G	9,0	10,2	10,4	10,0	8,0	6,1	7,0	7,4	33,7	30,5	29,6	28,3	2,9	3,0	2,9	3,0
A-A	7,5	8,3	8,7	8,6	10,7	9,0	9,5	9,5	18,8	17,8	17,8	22,4	3,0	2,7	2,8	2,9
G-C	7,4	8,1	8,5	8,4	10,4	10,7	10,2	9,9	33,4	26,0	29,2	25,8	2,4	2,5	2,4	2,4
C-G	7,4	8,4	8,5	8,1	10,4	10,0	10,1	10,7	18,8	22,1	17,9	17,6	3,0	2,9	2,9	2,7
A-A	7,5	8,7	8,7	8,3	10,7	9,4	9,8	9,1	33,7	28,2	28,9	30,4	2,9	3,0	2,9	2,9
G-C	9,0	9,9	10,4	10,1	8,0	7,5	7,2	6,1	30,9	32,6	31,9	30,8	3,0	3,0	3,1	3,1
C-G	8,7	9,8	10,3	10,0	8,3	8,6	8,2	7,2	30,0	33,0	33,0	31,6	2,7	2,7	2,7	2,8
C-G	7,8	8,3	8,8	8,7	13,1	12,9	11,6	10,3	-	-	-	-	-	-	-	-
średnia	8,1	9,0	9,3	9,0	10,1	9,2	9,4	9,2	28,9	28,0	28,1	28,0	2,8	2,9	2,8	2,9
s.d.	0,7	0,8	0,9	0,8	2,0	2,0	1,7	2,0	5,9	5,2	6,0	5,2	0,2	0,2	0,2	0,2

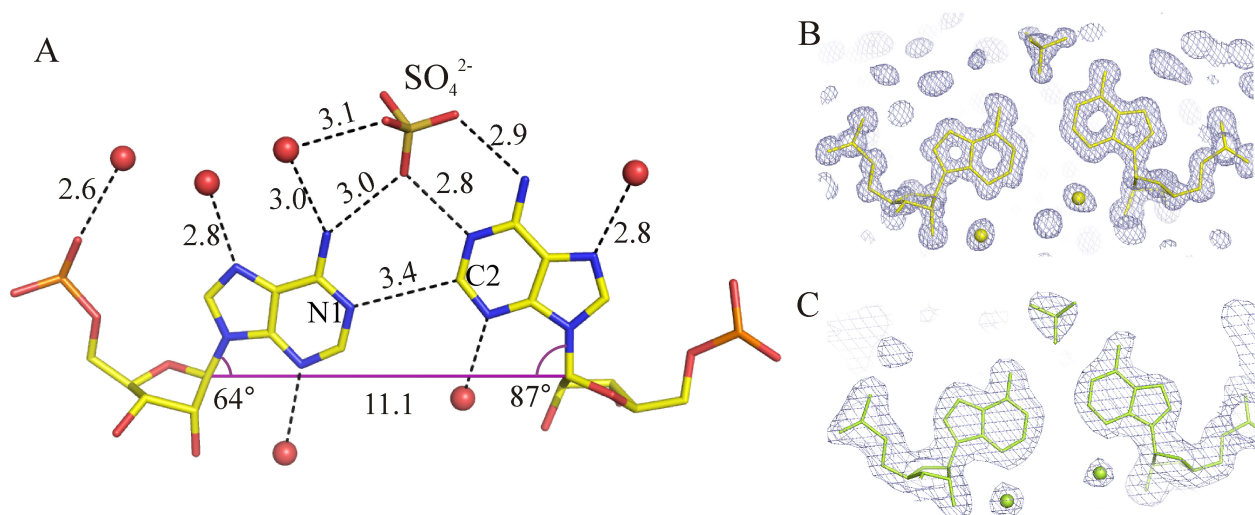
jest również wysunięta ($\lambda = 64^\circ$). Średnia wartość kąta dla pozostałych reszt w obu strukturach wynosi 55° .

3.2.7. Hydratacja i oddziaływania z ligandami

Solvatacja par A-A jest podobna we wszystkich dupleksach obu form krystalicznych. W bruzdzie małej każda z reszt adenozyńskich oddziałuje z cząsteczką wody za pomocą dwóch oddziaływań utworzonych z atomami N3 oraz O2 grupy 2' hydroksylowej (Rysunek 35). Podobnie

Tabela 11. Wyznaczone szerokości bruzd dupleksów (GGCAGCAGCC)₂ na podstawie odległości pomiędzy atomami P-P / udokładniony dystans z uwzględnieniem kierunku szkieletu fosforanowo-cukrowego.

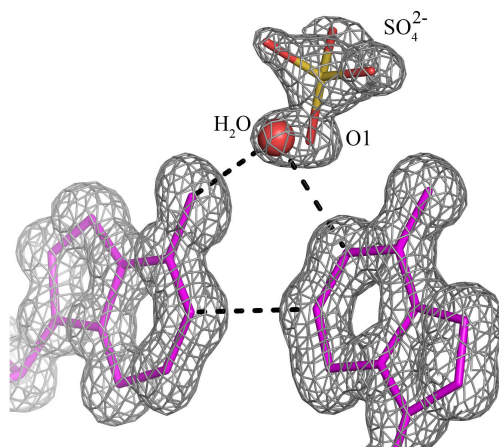
skok	Szerokość bruzdy małej (Å)				Szerokość bruzdy dużej (Å)			
	H+H'	A+B	C+D	E+F	H+H'	A+B	C+D	E+F
GG/CC	---	---	---	---	---	---	---	---
GC/GC	---	---	---	---	---	---	---	---
CA/AG	17,3/--	17/--	17,3/--	17,3/--	18,5/--	20,7/--	20,2/--	19,5/--
AG/CA	16,9/15,4	16,5/14,9	16,7/15,2	16,8/15,4	22,2/21,4	24,1/24,0	24,4/24,3	24/23,9
GC/GC	16,6/15,1	16,3/14,9	16,4/14,9	16,3/14,9	22,8/22,2	25,2/25,2	26/26,0	25,3/25,3
CA/AG	16,9/15,4	16,8/15,4	16,7/15,1	16,4/14,9	22,2/21,4	24,0/23,9	24,5/24,4	24,2/24,2
AG/CA	17,3/--	17,3/--	17,2/--	17/--	18,5/--	19,5/--	20,2/--	20,7/--
GC/GC	---	---	---	---	---	---	---	---
CC/GG	---	---	---	---	---	---	---	---



Rysunek 36. Niestandardowa para A-A oraz jej oddziaływania z cząsteczkami rozpuszczalnika (A). Na panelach B i C przedstawiono mapę gęstości elektronowej (kontur 1σ) struktury wysokorozdzielczej (B) i niskorozdzielczej (C).

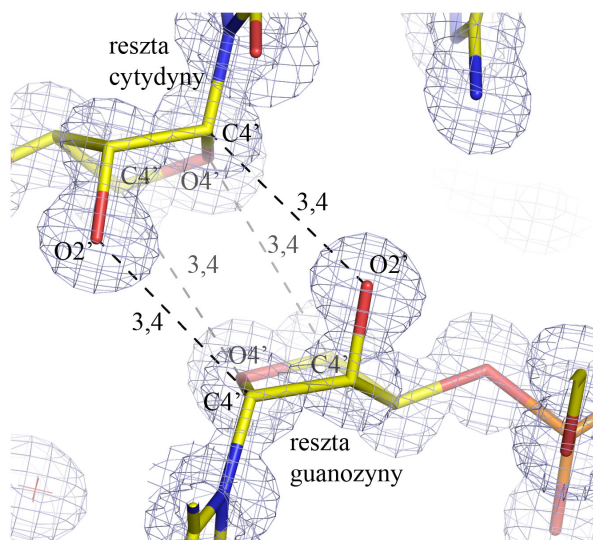
jest w bruzdzie dużej gdzie zazwyczaj znajduje się cząsteczka wody tworząca wiązanie wodorowe z atomem N7. Dodatkowo zlokalizowałam anion siarczanowy, który „zaklinowany” jest pomiędzy sparowanymi pierścieniami adenin. Jeden z atomów tlenu (O1) jonu siarczanowego oddziałuje jednocześnie z atomem N6 mniej wysuniętej reszty A oraz z N1 drugiej reszty A. Z atomem N3 drugiej reszty A kolejny atom tlenu (O2) jonu siarczanowego tworzy wiązanie wodorowe. Sposób oddziaływania jonu siarczanowego można opisać jako scalenie miejsc wiązania anionu dwóch reszt, które w literaturze określane są jako ADE_WC_H oraz ADE_WC (127). Obsadzenie siarczanu jest połówkowe i w jego nieobecności, w pozycji tlenu (O1), znajduje się woda (Rysunek 37).

W sieci krystalicznej formy trygonalnej znajduje się sześć oddziaływań międzycząsteczkowych typu ryboza-ryboza. Każde składa się z czterech oddziaływań utworzonych pomiędzy atomami C,



Rysunek 37. Oddziaływanie ligand-RNA w obrębie pary A-A. Jon siarczanowy ma obsadzenie połówkowe i w jego nieobecności, w pozycji tlenu O1, znajduje się cząsteczka wody (czerwona kula).

H i O. Dwa z nich stanowią wiązanie C1'-H1'...O2' a pozostałe dwa to C4'-H1'...O4'. Resztami zaangażowanymi w to oddziaływanie są wszystkie reszty 3C oraz 8G na poszczególnych niciach. Parowanie pierścieni cukrowych zachodzi pomiędzy resztami dwóch cytozyn (dwa przypadki), dwóch guanozyn (dwa przypadki) lub cytozyną a guanozyną (dwa przypadki). Podobnie jest w strukturze romboedrycznej, gdzie są dwie symetryczne pary: 3C-8G (Rysunek 38). Średnia odległość donor-akceptor wynosi 3,4 Å z odchyleniem standardowym równym 0,1 Å.

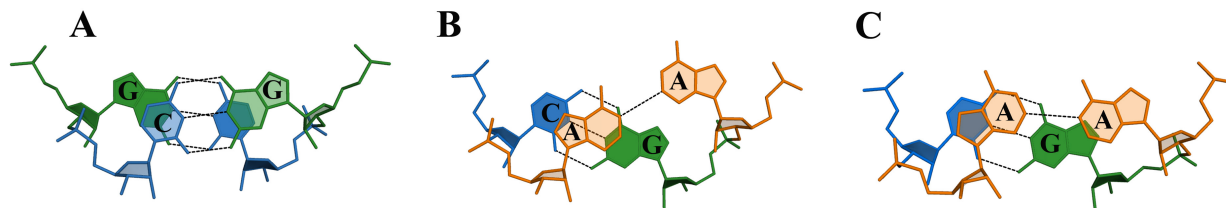


Rysunek 38. Oddziaływania w sieci krystalicznej pomiędzy pierścieniami ryboz dwóch symetrycznie powiązanych dupleksów (GGCAGCAGCACC)₂. Utworzone są cztery oddziaływania typu C-H...O, ich długości podane są w Å. Mapa gęstości elektronowej 2F_o-F_c ma kontur 1σ.

3.2.8. Oddziaływania warstwowe

W obu strukturach można wyróżnić trzy rodzaje oddziaływań warstwowych: jedno pomiędzy parami GC/GC i dwa pomiędzy CA/AG, zależne od konformacji reszt adenykowych (Rysu-

nek 39). Zasady par G-C i C-G nakładają się w dużym stopniu, co jest typowe dla par kanoicznych (Rysunek 39 panel A). Natomiast w przypadku niekanonicznych par A-A oddziaływania warstwowe są ograniczone. Zasada bardziej nachylonej w kierunku bruzdy dużej reszty A asocjuje jedynie z pierścieniem cytozyny sąsiadującej po stronie 5' (średnia powierzchnia nałożenia wynosi $2,2 \pm 1,1 \text{ \AA}^2$) (Rysunek 39 panel B i C). Mniej nachylona reszta adenozyiny tworzy oddziaływania warstwowe z obiema sąsiadującymi resztami (Rysunek 39 panel B i C) przy czym tylko w niewielkim zakresie (powierzchnia nałożenia wynosi $0,8 \pm 0,1 \text{ \AA}^2$ z resztą cytozyny i $0,3 \pm 0,3 \text{ \AA}^2$ z resztą guanozyny).



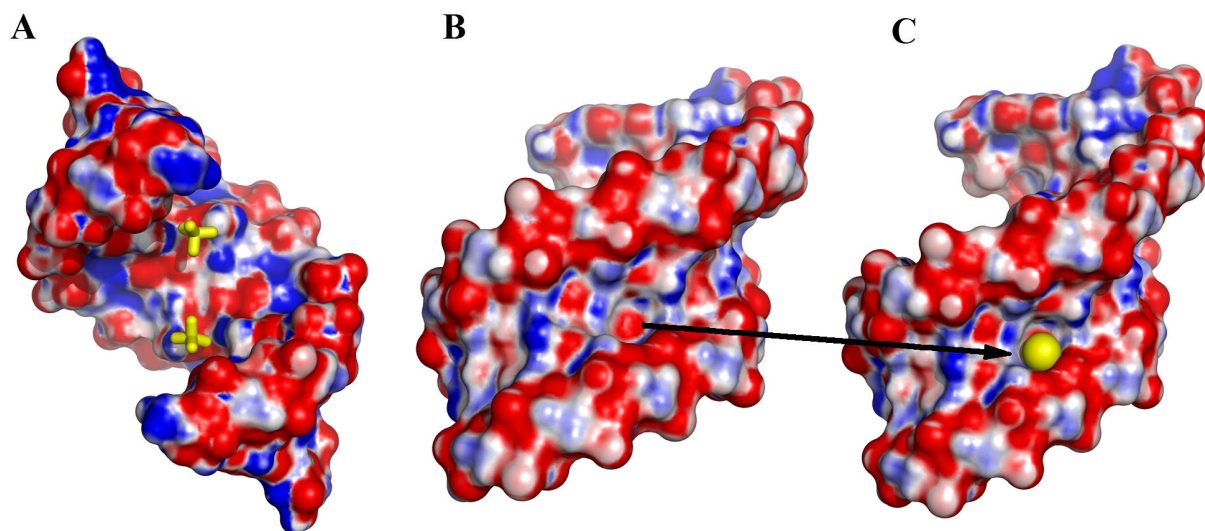
Rysunek 39. Trzy rodzaje oddziaływań warstwowych: pomiędzy parami typu Watson-Crick (A) oraz pomiędzy parami CA/AG (B, C) w zależności od konformacji pary A-A.

3.2.9. Powierzchnia i ukształtowanie potencjału elektrostatycznego

Potencjał elektrostatyczny posiada zbliżoną dystrybucję dla wszystkich dupleksów (Rysunek 40). W bruzdzie małej pasy o pozytywnym i negatywnym ładunku ułożone są równoległe do kierunku osi helisy i przeplatają się na przemian tworząc charakterystyczny wzór. Pozytywne pasma utworzone są głównie przez grupy *exo*-aminowe reszt guanozyn oraz przez pierścienie cukrowe. Bruzda duża wykazuje większe zróżnicowanie w dystrybucji potencjałów. Obszary o potencjale dodatnim generują grupy *exo*-aminowe reszt cytydyn oraz adenozyn. Do tych ostatnich wiążą się jony siarczanowe, które pasują zarówno geometrią jak i ładunkiem. Pary A-A znajdują się w centralnej części bruzdy dużej i tworzą, na skutek ich wypchnięcia, wybrzuszenia. Natomiast za resztami adenozyn w konformacji „kciuk do góry” tworzy się wgłębienie. Reszta guanozyny po stronie 3' jest znacznie przesunięta w dół w stosunku do adenozyiny w konformacji „kciuk do góry” i nie tworzy z nią oddziaływań warstwowych.

3.3. Struktury RNA zawierające powtórzenia CGG

Określiłam strukturę trzech różnych dupleksów zawierających powtórzenia CGG. Dwa z nich, $[\text{GC}(8\text{-BrG})\text{GCGGC}]_2$ i $[\text{GC}(8\text{-BrG})\text{GCGGCGGC}]_2$, zostały zaprojektowane i oczyszczone przez prof. dr. hab. Ryszarda Kierzka. Zawierają one modyfikowaną resztę 8-bromoguanozynę (8-BrG) w pozycji 3 łańcucha RNA. Prof. dr hab. Wojciech Rypniewski wykrystalizował i zarejestrował dane z kryształów obu oligomerów. Rozwiązał również strukturę dupleksu $[\text{GC}(8\text{-BrG})\text{GCGGC}]_2$. Moja praca rozpoczęła się od udokładnienia struktury tego dupleksu oraz od rozwiązania



Rysunek 40. Potencjał elektrostatyczny dupleksu $(GGCAGCAGCC)_2$ struktury romboedrycznej. Elektronegatywny potencjał oznaczono na czerwono a elektropozytywny na niebiesko. Jony siarczano-we (żółty) wiążą się do bruzdy dużej i oddziałują z parami A-A (A). W bruzdzie małej (B i C) znajduje się wgłębienie, które wypełnione jest cząsteczką wody (żółta sfera) (C).

struktury dupleksu $[GC(8-BrG)GCGGCGGC]_2$. Dla trzeciego oligomeru o sekwencji GCGGCGGC zrealizowałam wszystkie etapy na drodze do określenia jego struktury krystalicznej.

3.3.1. Krystalizacja oligomerów zawierających powtórzenia CGG

Roztwór rezerwurowy używany do krystalizacji oligomeru GCGGCGGC zawierał 10 mM $MgCl_2$, 50 mM kakodylan sodu pH 6 i 1 M Li_2SO_4 . Początkowo stężenie RNA wynosiło 2,4 mg/ml. Kryształy rosły w przeciągu kilku dni w postaci małych zrosniętych igieł, które użyto jako zarodki krystaliczne. Szczepione kryształy rosły w takich samych warunkach krystalizacyjnych ale przy mniejszym stężeniu RNA, 1,2 mg/ml. Taki sam roztwór rezerwurowy stosowany był do krystalizacji oligomeru $GC(8-BrG)GCGGCGGC$ o stężeniu 10 mg/ml. Kryształy pojawiły się po dziesięciu miesiącach. Natomiast kryształy $[GC(8-BrG)GCGGC]_2$ rosły przez dwa miesiące w 10 mM $CaCl_2$, 0,2 M NH_4Cl , 50 mM Tris-HCl pH 8,5 i 30% w/v PEG 4000. RNA miało stężenie 10 mg/ml. Krystalizacja odbywała się w temperaturze 19 °C. Objętość roztworu rezerwurowego wynosiła 500 μl , natomiast RNA mieszano z roztworem rezerwurowym w stosunku 1:1.

3.3.2. Pomiary dyfrakcyjne i przetwarzanie danych

Dane dyfrakcyjne rejestrowane były w temperaturze 100 K: kryształy $(GCGGCGGC)_2$ na linii BL 14.2 synchrotron BESSY (Berlin), kryształy $[GC(8-BrG)GCGGC]_2$ i $[GC(8-BrG)GCGGCGGC]_2$ na linii EMBL-X11, DESY (Hamburg). Jako krioprotektantu użyto 20% glicerol. Grupa przestrzenna kryształów dupleksów modyfikowanych została w

obu przypadkach wyznaczona jako C2 (jednoskośna), natomiast niemodyfikowanej jako P1 (trójskośna). Takie niskie grupy przestrzenne wymagają zarejestrowania dużej liczby obrazów dyfrakcyjnych obejmujących sumaryczny kąt oscylacji przynajmniej 180° . Dla kryształu $[\text{GC}(8\text{-BrG})\text{GCGGC}]_2$ wynosił on $353,5^\circ$, dla $[\text{GC}(8\text{-BrG})\text{GCGGCGGC}]_2$ $360^\circ + 30^\circ$ a dla $(\text{GCGGCGGC})_2$ $360^\circ + 56^\circ$. Kryształy rozpraszały odpowiednio do rozdzielczości 0,97, 1,45 oraz 2,05 Å. Wszystkie zbiory danych były integrowane i skalowane przy użyciu programów DENZO/SCALEPACK. Charakterystyka danych dyfrakcyjnych znajduje się w Tabeli 12.

Tabela 12. Zestawienie statystyk przetworzonych danych dyfrakcyjnych struktur zawierających powtórzenia CGG. W nawiasie podano wartości dotyczące ostatniego zakresu rozdzielczości.

Oligomer	GCGGCGGC	GC(8-BrG)GCGGC	GC(8-BrG)GCGGCGGC
Źródło promieniowania	BESSY BL 14.2	EMBL-X11	EMBL-X11
Długość fali (Å)	0,92	0,8126	0,8126
Grupa przestrzenna	P1	C2	C2
Parametry komórki	a = 39,7 b = 76,9 c = 85,4 Å; $\alpha = 90,0 \beta = 88,6 \gamma = 77,3^\circ$	a = 50,7 b = 22,5 c = 44,2 Å; $\beta = 117,8^\circ$	a = 118,6 b = 28,6 c = 61,8 Å; $\beta = 118,0^\circ$
Zakres Rozdzielczości (Å)	20,0– 2,05 (2,09– 2,05)	20,0– 0,97 (0,99– 0,97)	20,0– 1,45 (1,47– 1,45)
Mozaikowość (°)	0,3	1	0,3
R _{merge}	0,093 (0,488)	0,132 (0,934)	0,112 (0,620)
$\langle I/\sigma(I) \rangle$	17 (2,8)	9 (2,5)	10 (3,3)
Kompletność (%)	98,4 (97,5)	95,7 (93,4)	99,8 (100,0)
Liczba unikalnych refleksów	60 328	24 974	32 941
Krotność obserwacji	4,4 (3,5)	5,4 (4,7)	7,5 (6,3)
Refleksy > 3 σ (%)	71 (40)	77 (29)	77 (50)
Czynnik B z wykresu Wilsona (Å ²)	33	7,8	21,2

3.3.3. Rozwiązanie struktur zawierających powtórzenia CGG

Struktura $[\text{GC}(8\text{-BrG})\text{GCGGC}]_2$ została rozwiązana metodą SAD (ang. *single-wavelength anomalous dispersion*) jako pierwsza ze wszystkich oligomerów zawierających powtórzenia CNG. Z powyższego względu stosowana była przede mną jako model do podstawienia cząsteczkowego przy rozwiązywaniu innych struktur RNA. Pozycje atomów bromu zostały określone za pomocą programu SHELXD (128). Następnie za pomocą programu SHELXE wybrano odpowiedni enancjomer oraz obliczono wstępne fazy. Obliczona mapa gęstości elektronowej nie była na tyle dokładna, aby móc na jej podstawie zbudować model, natomiast tworzyła równoległe kolumny wskazując na dupleksy RNA ułożone warstwowo jeden nad drugim. Do modyfikacji gęstości elektronowej zastosowano program DM (129), ale nie poprawiło to jakości mapy. Kolejnym krokiem było udokładnianie faz z użyciem wolnych atomów z wytrząsaniem za pomocą programu ARP/wARP. Po 25 cyklach dodawania i usuwania atomów, czynniki R/R_{free} zaczęły maleć osiągając po 50 cyklach wartość 0,188/0,256. Otrzymana mapa gęstości elektronowej pokazywała

wolne atomy stanowiące rozwiązana strukturę RNA.

Obie struktury $[GC(8-BrG)GCGGCGGC]_2$ i $(GCGGCGGC)_2$ rozwiązałam podstawieniem cząsteczkowym przy użyciu programu PHASER. Modelem był dupleks $[GC(8-BrG)GCGGCGGC]_2$, który dla 11 nukleotydowego oligomeru GC(8-BrG)GCGGCGGC został przedłużony o odpowiednie trzy pary zasad. Dla struktury modyfikowanej obliczyłam przy użyciu współczynnika Matthews'a, iż w części asymetrycznej znajdują się dwie lub trzy cząsteczki. Programowi udało się obliczyć fazy dla dwóch dupleksów. Wskaźniki z-score były stosunkowo niskie, wynosiły odpowiednio 4,3 oraz 6,0 ale inspekcja mapy gęstości elektronowej pokazała, że początkowe fazy są względnie dobrze obliczone tylko dla części modelu. Cząsteczki były nie do końca poprawnie usytuowane w komórce, co generowało nakładanie się fragmentów symetrycznych cząsteczek na siebie. Wymagało to przebudowy modelu.

W podobny sposób rozwiązałam strukturę natywnego dupleksu. Obliczyłam, że w części asymetrycznej znajduje się 16–20 cząsteczek RNA. Niestety moc obliczeniowa komputera była za słaba, aby znaleźć fazę dla wszystkich 16 dupleksów. Następowало przedwczesne zakończenie działania programu. Z powyższego powodu zadałam programowi PHASER znalezienie 8 dupleksów. Wartości z-score dla większości cząsteczek były większe od 7 a jakość mapy gęstości elektronowej zadowalająca. Na podstawie mapy różnicowej mogłam dobudować stopniowo kolejne dupleksy.

3.3.4. Udokładnianie struktur zawierających powtórzenia CGG

Wszystkie trzy modele udokładnianie były za pomocą programu REFMAC5 i PHE-NIX (Tabela 13). Do obliczania statystyk R_{free} wydzielono 2% refleksów (modyfikowane RNA) oraz 5% w przypadku natywnej struktury. Pierwsze cykle udokładniania struktury $[GC(8-BrG)GCGGCGGC]_2$ i $(GCGGCGGC)_2$ polegały na budowaniu modelu. Monitorowanie wartości czynników R pozwoliło na ocenę poprawności wykonanej pracy. Gdy modele zostały całkowicie określone, dodałam cząsteczki wody za pomocą programu ARP/wARP. Czynniki temperaturowe obu struktur udokładniałam izotropowo. W przypadku 11 nukleotydowego dupleksu stosunek liczby refleksów (31971) do liczby udokładnianych parametrów ($1864 \times 9 = 16776$) teoretycznie pozwalał na zastosowanie anizotropowych czynników temperaturowych. Niestety po ich uwzględnieniu czynniki R zmniejszyły tylko nieznacznie swoją wartość. Wysokorozdzielcza struktura $[GC(8-BrG)GCGGCGGC]_2$ była udokładniania z anizotropowymi czynnikami temperatury. Ostatnie cykle prowadzone były z pełnym zbiorem refleksów oraz bez stosowania więzów geometrycznych (ang. *unrestrained refinement*). Słownik zawierający stereochemiczny opis nukleotydu 8-bromoguanozyny wygenerowałam za pomocą serwera PRODRG2 (130) oraz standardowego słownika używanego dla guanozyny przez programy do udokładniania.

Tabela 13. Zestawienie statystyk otrzymanych po udokładnieniu struktur zawierających powtórzenia CGG.

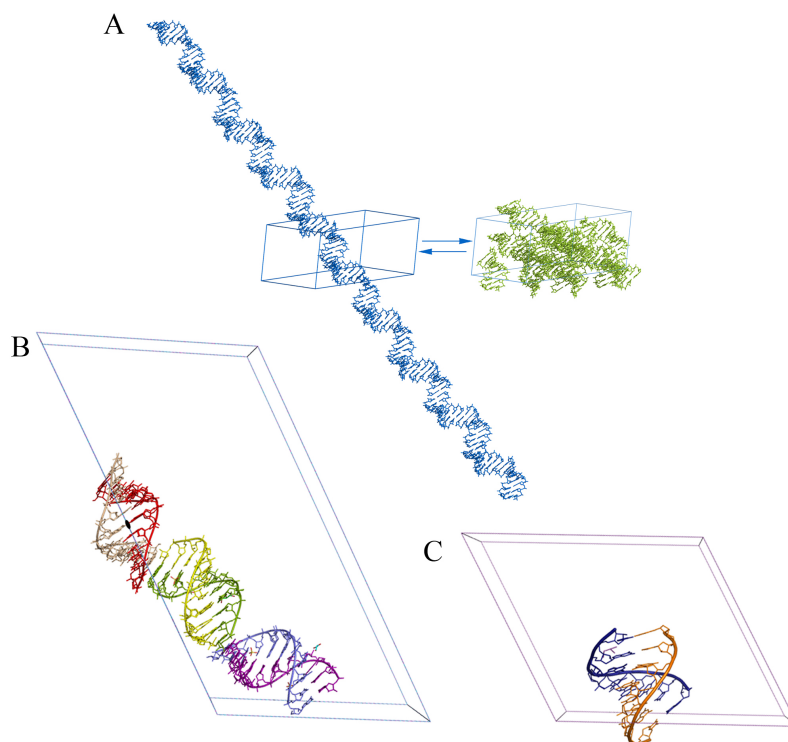
Oligomer	GCGGCGGC	GC(8-BrG)GCGGC	GC(8-BrG)GCGGCGGC
Liczba refleksów: work/free	57051/3038	24947/0	32646/671
R (%)	21,56	13,66	23,21
R _{free} (%)	25,71	-	27,02
Liczba atomów cząsteczek RNA	6304	370	1617
Liczba cząsteczek jonów	7 SO ₄ ²⁻	2 Ca ²⁺	8 SO ₄ ²⁻
Liczba cząsteczek wody	524	103	207
Inne cząsteczki	-	1 glicerol	-
Odchylenie rmsd od wartości idealnej			
długość wiązania (Å)	0,005	0,011	0,006
kąty (°)	1,145	1,746	1,282
Kod PDB	3R1C	3R1E	3R1D

3.3.5. Ogólny opis struktur zawierających powtórzenia CGG

We wszystkich trzech strukturach cząsteczki RNA ułożone są w komórce jedna nad drugą i tworzą długie równoległe kolumny. W części asymetrycznej struktury trójskośnej (niemodyfikowanej) znajduje się 18 dupleksów. Są one uporządkowane w przestrzeni w taki sposób, że mogą utworzyć jedną pseudohelisę składającą się ze wszystkich 18 symetrycznie niezależnych cząsteczek (Rysunek 41). W części asymetrycznej wysokorozdzielczej struktury [GC(8-BrG)GCGGC]₂ jest jeden dupleks, natomiast w przypadku [GC(8-BrG)GCGGCGGC]₂ znajduje się 5 nici (Rysunek 41). Tworzą one trzy helisy, z których jedna składa się z dwóch symetrycznych łańcuchów.

3.3.6. Pary G-G i ich wpływ na strukturę RNA

Wszystkie helisy mają formę A-RNA, w których kąt skręcenia znajduje się w zakresie 30-32°. Pierścienie cukrowe mają upakowanie C3'-endo lub w kilku przypadkach C2'-exo a parametr Zp wynosi $2,63 \pm 0,22$ Å (dla formy A-RNA powinien być większy od 1,5 Å). Pomędzy kanonicznymi parami G-C znajdują się pary G-G, w których zawsze jedna guanozyna jest w konformacji *syn* natomiast druga w *anti* (Rysunek 42). Według nomenklatury zaproponowanej przez Leontisa i Westhofa (121) można tę parę opisać jako G/G *cis* Watson-Crick/Hoogsteen. Pomędzy resztami guanozynowymi, wchodzących w skład niestandardowej pary, utworzone są dwa wiązania wodorowe pomędzy tlenem karbonylowym a N1 oraz atomem N7 a grupą *exo*-aminową. Długości wiązań mieszczą się w zakresie od 2,6 do 3,3 Å. Dodatkowo konformacja G(*syn*) stabilizowana jest przez wewnętrzne oddziaływanie grupy *exo*-aminowej z atomem tlenu grupy fosforanowej (3,0 Å). Dystans pomędzy atomami C1' wynosi $11,3 \pm 0,1$ Å, dla porównania pomędzy parami G-C wynosi $10,7 \pm 0,2$ Å. Kąt λ między wiązaniem N-glikozydowym a linią łączącą atomy C1' jest

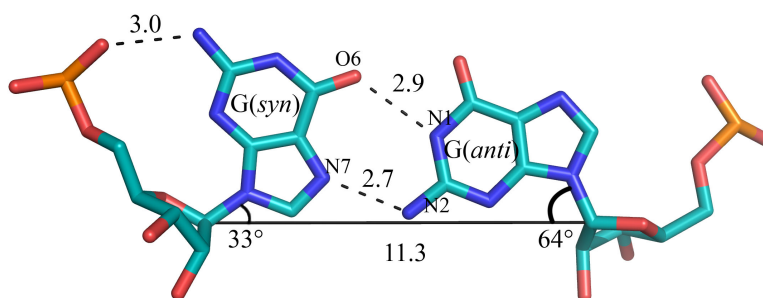


Rysunek 41. Upakowanie dupleksów $(GCGGCGGC)_2$ (A), $[GC(8-BrG)GCGGCGGC]_2$ (B) i $[GC(8-BrG)GCGGCGGC]_2$ (C) w komórce elementarnej. W strukturze trójskośnej (A) 18 dupleksów może być upakowane w sieci na dwa sposoby: w postaci długiej pseudonieskończonej helisy (niebieski) bądź równomiernie w przestrzeni komórki elementarnej (żółty). W strukturze jednoskośnej (B) w części asymetrycznej znajduje się pięć łańcuchów $GC(8-BrG)GCGGCGGC$, które tworzą trzy dupleksy. Jeden z nich (czerwony i różowy) leży na osi dwukrotnej. Wysokorozdzielcza struktura jednoskośna (C) w części asymetrycznej ma jeden dupleks.

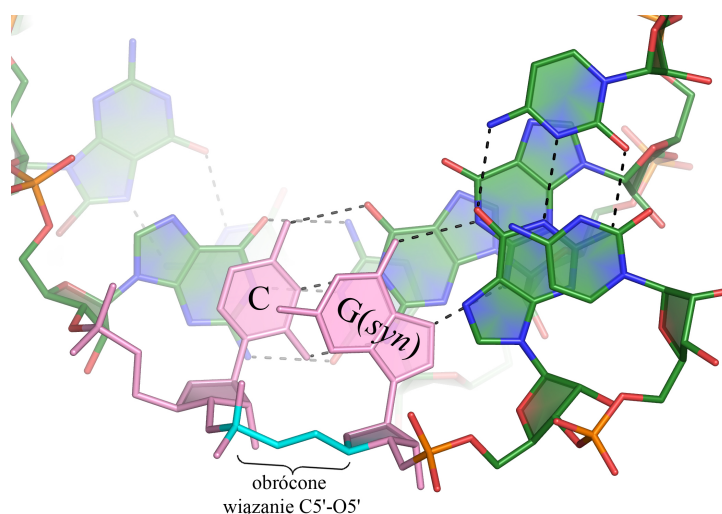
równy $33 \pm 4^\circ$ dla $G(\text{syn})$, natomiast dla $G(\text{anti})$ $54 \pm 3^\circ$. Oznacza to, że reszta w konformacji *syn* jest przesunięta w kierunku bruzdy małej natomiast druga reszta w kierunku bruzdy dużej.

Kąty torsyjne α (określa rotację wokół wiązania P-O5') i γ (określa rotację wokół wiązania C5'-C4') reszty $G(\text{syn})$ są nietypowe. Przyjmują one odmienną konformację. Typowa dla kąta α jest konformacja *-sc*, natomiast w strukturach obserwujemy *+ac*, *+ap* lub w jednym przypadku *-ap* (Rysunek 43). Oznacza to, że jego wartość waha się od 107 – 182° , natomiast średnia wynosi 142° i jest to prawie pół obrotu więcej od standardowej wartości -60° dla *-sc*. Z kolei kąt γ jest albo *+ap* lub *-ap* co odpowiada wartościom 152° lub -152° . Jest to około 120° więcej od typowej konformacji *-sc*. Innymi słowy oznacza to, że wiązanie O5'-C5' jest „przekręcone” a szkielet fosforanowo-cukrowy „wyprostowany” (Rysunek 43).

Skutkiem lokalnego wyprostowania wiązania fosforanowo-cukrowego jest to, iż pierścienie ryboz: $G(\text{syn})$ oraz sąsiadującej po jej 5' stronie reszty cytozyny leżą niemal w tej samej płaszczyźnie. Ponadto następuje lokalne „rozwiniecie” helisy, które jest kompensowane w innym miejscu dupleksu. W konsekwencji, średnia wartość skrętu helisy nie różni się od typowej. Uśredniona szerokości bruzdy dużej wynosi dla natywnego dupleksu $17,9 \pm 0,9 \text{ \AA}$, dla dłuższego modyfikowanego $17,8 \pm 2,5 \text{ \AA}$ i dla krótszego modyfikowanego $14,3 \text{ \AA}$. Z kolei bruzda mała ma



Rysunek 42. Para G-G. Jedna z reszt jest w konformacji *syn* i jest nachylona w kierunku bruzdy małej, natomiast druga jest w konformacji *anti* i jest wypchnięta w kierunku bruzdy dużej. Reszty oddziałują ze sobą dwoma wiązaniami wodorowymi. Dodatkowo istnieje jedno wiązanie wewnątrzcząsteczkowe pomiędzy grupą *exo*-aminową G(*syn*) a atomem tlenu jej grupy fosforanowej.



Rysunek 43. Reszta guanozyny w konformacji *syn* (różowy) w otoczeniu strukturalnym. Wiązanie pomiędzy atomami C5' i O5' (niebieski) jest nietypowe. W rezultacie szkielet fosforanowo-cukrowy znajduje się w tej samej płaszczyźnie co sąsiednia reszta cytydyny a kąt skrętu helisy ulega zmniejszeniu.

szerokość odpowiednio: $15,8 \pm 0,5 \text{ \AA}$, $15,4 \pm 0,5 \text{ \AA}$ i $16,1 \text{ \AA}$. Są to wartości typowe dla formy A-RNA. Parametry helikalne są zebrane w Tabelach 14, 15, 16 i 17.

3.3.7. Oddziaływania warstwowe

W dupleksach zawierających powtórzenia CGG występują trzy rodzaje oddziaływań warstwowych (Rysunek 44). Pierwsze z nich tworzy się tylko między sąsiadującymi ze sobą kanonicznymi parami CG/GC, których zasady w dużym stopniu asocjują ze sobą. W pozostałych dwóch rodzajach oddziaływań zaangażowane są pary G-G. Reszta G(*syn*) oddziałuje jedynie w małym zakresie z resztami G i C znajdującymi po jej stronie 5'. Podobnie jest w przypadku drugiej reszty G(*anti*), której pierścień zasady znajduje się nad obiema sąsiadującymi resztami po stronie 5' podczas, gdy po swojej stronie 3' oddziałuje z pięciocłonowym pierścieniem guaniny.

Tabela 14. Zestawienie parametru helikalnego przesunięcie (ang. displacement) (\AA) dla dupleksów (GGGGGGC)₂ (A), [GC(8-BrG)GGGGGGC]₂ (B) i [GC(8-BrG)GGGGC]₂ (C). W panelach B i C podane są wartości dla różnych konformacji dupleksów RNA obserwowanych w danej strukturze.

łańcuch	AB	CD	EF	GH	IJ	KL	MN	OP	QR	SY	TU	VZ	WX	ab	cd	ef	gh	ij
G-C	6,4	7,0	7,2	6,5	6,6	6,7	7,4	6,8	7,3	7,1	7,1	6,9	7,0	6,7	6,5	6,6	6,5	6,7
C-G	6,7	6,9	7,0	6,3	6,8	6,7	7,2	6,7	7,0	6,9	6,8	6,7	7,2	7,1	6,7	7,0	6,7	7,0
G-G	6,3	6,8	6,3	6,2	6,2	6,2	6,4	6,3	6,3	6,4	6,4	6,1	6,6	6,6	6,3	6,8	6,5	6,5
G-C	6,5	6,5	6,7	6,8	6,5	6,6	6,6	6,7	6,6	6,6	6,6	6,7	6,9	6,7	6,8	7,1	6,9	6,4
C-G	6,0	7,1	6,9	6,5	6,8	6,8	6,4	6,5	6,4	6,7	6,9	6,8	7,1	6,7	7,0	6,7	6,4	6,6
G-G	6,1	6,8	6,3	6,1	6,4	6,3	6,3	6,4	6,2	6,5	6,4	6,5	6,9	6,4	6,4	6,6	6,2	6,2
G-C	6,9	6,9	6,7	6,8	6,7	6,4	6,8	6,8	6,7	7,0	6,9	6,7	7,2	6,9	6,8	7,1	6,8	6,6
C-G	6,6	6,5	6,4	6,7	6,2	6,1	6,2	6,8	6,2	6,5	7,0	6,7	6,7	7,0	6,6	6,9	6,8	6,4
średnia	6,4	6,8	6,7	6,5	6,5	6,5	6,7	6,6	6,6	6,7	6,8	6,6	7,0	6,8	6,6	6,8	6,6	6,5
s.d.	0,3	0,2	0,3	0,3	0,3	0,3	0,4	0,2	0,4	0,3	0,3	0,2	0,2	0,2	0,2	0,2	0,2	0,2

A

łańcuch	AB
G-C	5,5
C-G	5,8
G-Br-G	5,7
G-C	6,2
C-G	6,6
G-GBr	6,0 / 5,7
G-C	6,2
C-G	6,7
średnia	6,0
s.d.	0,4

C

łańcuch	AB	CD	EE'
G-C	7,9 / 7,5	7,7 / 7,6 / 6,8 / 6,7	8,1 / 8,0 / 8,4 / 8,2
C-G	8,2 / 7,7	8,0 / 7,9 / 7,4 / 7,2	7,8 / 7,7 / 8,3 / 8,2
G-Br-G	7,8 / 7,9 / 7,3 / 7,3	7,3 / 7,1 / 6,8 / 6,6	7,2 / 7,0 / 7,7 / 7,5
G-C	7,6 / 7,7 / 7,3 / 7,4	7,0 / 6,9 / 6,6 / 6,5	8,3 / 8,2 / 8,6 / 8,5
C-G	6,4 / 6,5 / 6,4 / 6,5	6,0 / 5,4 / 6,0 / 5,4	8,0 / 8,3 / 8,3 / 8,5
G-G	7,0 / 6,9	5,9 / 5,4 / 5,9 / 5,3	8,3 / 8,6
G-C	7,8 / 7,7 / 7,7 / 7,5	6,5 / 6,6 / 6,4 / 6,5	8,3 / 8,0 / 8,5 / 8,3
C-G	8,1 / 7,9 / 8,0 / 7,8	6,9 / 7,0 / 6,9 / 7,0	8,2 / 8,3 / 8,5 / 8,6
G-GBr	6,5 / 6,3 / 6,5 / 6,2	5,6 / 5,7 / 5,7 / 5,8	7,0 / 7,2 / 7,5 / 7,7
G-C	6,2 / 6,3 / 6,4 / 6,4	5,1 / 5,4	7,7 / 7,8 / 8,2 / 8,3
C-G	6,1 / 5,5 / 6,5 / 5,9	5,0 / 4,9 / 5,3 / 5,3	8,0 / 8,1 / 8,2 / 8,4
średnia	7,1	6,3	8,1
s.d.	0,7	0,9	0,4

B

Tabela 15. Zestawienie parametru helikalnego kąt (ang. angle) ($^{\circ}$) dla dupleksów (GGGGGGG)₂ (A), [GC(8-BrG)GGGGGGC]₂ (B) i [GC(8-BrG)GGGGC]₂ (C). W panelach B i C podane są wartości dla różnych konformacji dupleksów RNA obserwowanych w danej strukturze.

łańcuch	AB	CD	EF	GH	IJ	KL	MN	OP	QR	SY	TU	VZ	WX	ab	cd	ef	gh	ij
G-C	6,5	8,8	7,3	6,2	9,6	10,6	5	3,6	5,5	4,2	5,7	4,2	6,8	7,5	6	5,2	4,2	4,4
C-G	3,8	7	5,9	4,3	10,1	10,6	4,7	3,1	5,5	3,8	4,4	4	6,2	4,9	5,8	3,8	3,5	3,1
G-G	5,3	7,1	6,8	5,9	8,6	9	7,8	4,8	8,2	8,3	6,1	6,2	6	5	5,9	3,4	4,4	5,7
G-C	4,6	4,3	7,2	5,7	4,3	5,4	8,5	3,7	8,9	5,6	5,3	6,4	3,2	5,9	2,6	1,8	2,7	6,2
C-G	5,8	4,6	6,6	6,5	3,6	4,3	6,9	3,8	7,7	7,3	5,1	5,3	3	3,9	1,8	2,6	5,5	4,6
G-G	6,2	7,7	8	5,8	6	7,2	7	5,1	8	7,4	5,7	6,8	5,2	6,5	5,6	4,6	5,1	5,2
G-C	2,3	7,9	7,6	3,6	7,2	9,5	4,6	2,7	5,2	5,4	3,1	4,4	5,2	3,3	6,3	5,1	3,4	5,9
C-G	5,2	7,3	10,2	5,3	9,3	10,6	7,9	3,5	8,4	8,8	4,4	7,3	8,1	3,2	7,3	5,4	3,7	6,9
średnia	5	6,8	7,4	5,4	7,3	8,4	6,5	3,8	7,2	6,4	5	5,6	5,5	5	5,2	4	4	5,2
s.d.	1,4	1,6	1,3	1	2,5	2,5	1,5	0,8	1,5	1,9	1	1,3	1,7	1,6	1,9	1,3	0,9	1,2

B

łańcuch	AB	CD	EE'
G-C	2,2 / 2,6 / 5,0 / 5,5	0,9 / 1,0 / 3,1 / 3,2	9,1 / 8,9 / 5,6 / 5,5
C-G	2,3 / 2,3 / 7,4 / 7,6	0,9 / 0,9 / 3,8 / 3,8	14,1 / 13,9 / 8,3 / 8,1
G-Br-G	6,2 / 6,0 / 10,2 / 10,1	4,0 / 4,0 / 6,5 / 6,4	18,9 / 18,6 / 12,9 / 12,7
G-C	7,4 / 6,9 / 6,7 / 6,2	4,6 / 4,5 / 4,5 / 4,4	10,6 / 10,4 / 8,9 / 8,7
C-G	10,0 / 9,5 / 9,5 / 8,8	6,2 / 7,8 / 5,8 / 7,5	6,9 / 6,1 / 6,1 / 5,2
G-G	7,2 / 6,7 / 7,3 / 6,7	6,3 / 6,8 / 6,4 / 7,0	7,0 / 6,0
G-C	1,3 / 0,9 / 2,1 / 1,6	2,5 / 2,4 / 3,1 / 3,1	6,1 / 6,9 / 5,2 / 6,1
C-G	2,3 / 2,2 / 3,6 / 3,5	2,0 / 2,0 / 2,9 / 3,0	10,4 / 10,6 / 8,7 / 8,9
G-GBr	8,9 / 9,0 / 10,3 / 10,4	6,7 / 7,7	18,6 / 18,9 / 12,7 / 12,9
G-C	10,9 / 10,4 / 12,1 / 11,7	9,5 / 9,6 / 10,2 / 10,3	13,9 / 14,1 / 8,1 / 8,3
C-G	12,5 / 11,5 / 13,2 / 12,1	10,3 / 10,4 / 10,6 / 10,7	8,9 / 9,1 / 5,5 / 5,6
średnia	7,1	5,5	9,7
s.d.	3,6	3	4

C

łańcuch	AB
G-C	13,2 / 13,3
C-G	12,9 / 12,8
G-Br-G	13,5 / 13,2
G-C	9,0 / 8,7
C-G	6,0 / 5,8
G-GBr	10,8 / 10,4
G-C	10
C-G	9,5 / 9,7
średnia	10,5
s.d.	2,5

Tabela 16. Zestawienie parametru helikalnego kąta skrętu helisy (*ang. helical twist*) ($^{\circ}$) dla dupleksów (*GGGGGGC*)₂ (A), [*GC(8-BrG)GGGGGGC*]₂ (B) i [*GC(8-BrG)GGGGC*]₂ (C). W panelach B i C podane są wartości dla różnych konformacji dupleksów RNA obserwowanych w danej strukturze.

łańcuch	AB	CD	EF	GH	IJ	KL	MN	OP	QR	SY	TU	VZ	WX	ab	cd	ef	gh	ij
G-C	31,0	34,8	29,6	32,3	29,7	28,8	30,5	32,1	30,7	31,8	31,3	33,6	31,8	29,3	30,3	30,9	32,0	31,3
C-G	29,4	31,1	27,6	33,5	32,4	33,6	27,2	25,3	26,6	27,8	25,6	27,2	33,8	29,3	33,2	33,5	26,4	27,4
G-G	33,5	26,2	32,7	26,8	27,3	28,2	32,6	33,8	33,5	32,8	33,1	32,9	25,8	30,9	28,4	27,0	35,4	32,7
G-C	30,7	34,1	31,1	33,4	31,8	32,7	29,5	31,9	30,8	30,6	32,1	31,7	27,2	32,0	29,5	27,7	32,1	31,9
C-G	32,7	25,6	29,5	32,7	28,1	29,6	34,8	32,4	34,6	32,3	34,5	32,6	27,9	32,8	27,8	25,3	30,1	29,6
G-G	27,4	32,2	32,8	26,9	34,0	33,7	28,2	26,1	28,8	30,3	24,8	26,7	33,3	27,5	34,1	33,5	28,3	32,6
G-C	30,1	31,1	28,0	31,6	29,1	28,2	32,6	31,3	32,4	29,9	30,7	31,1	30,6	30,8	28,3	31,7	31,2	32,4
C-G	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
średnia	30,7	30,7	30,2	31,0	30,3	30,7	30,8	30,4	31,1	30,8	30,3	30,8	30,1	30,4	30,2	29,9	30,8	31,1
s.d.	2,1	3,6	2,1	2,9	2,4	2,5	2,7	3,3	2,8	1,7	3,7	2,8	3,1	1,8	2,5	3,3	2,9	2,0

C

łańcuch	AB
G-C	32,6
C-G	36,7
GBr-G	27,6
G-C	34,0 / 33,9
C-G	28,0 / 25,5
G-GBr	33,1 / 35,6
G-C	33,8
C-G	---
średnia	32,4
s.d.	3,8

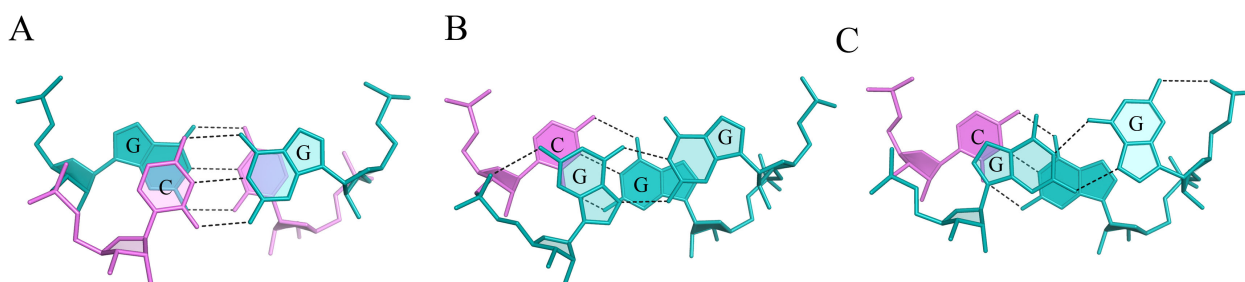
B

łańcuch	AB	CD	EE'
G-C	35,1	27,8 / 29,9	32,0 / 34,7
C-G	32,4 / 30,8	31,8 / 33,2	36,8 / 33,7
GBr-G	25,3 / 27,8	29,1 / 29,3	20,4 / 16,7
G-C	32,1 / 35,8	26,9 / 33,5 / 31,0 / 37,6	35,7 / 32,2 / 33,4 / 30,0
C-G	27,1	28,03 / 28,6	17,7 / 31,9
G-G	27,0	34,5 / 27,2	31,9 / 17,7
G-C	33,9	33,6	32,2 / 35,7 / 30,0 / 33,4
C-G	21,8	28,5	16,7 / 20,4
G-GBr	36,1 / 33,1	35,1	33,7 / 36,8
G-C	30,5 / 39,2	28,6	34,7 / 32,0
C-G	---	---	---
średnia	30,2	30,8	29,0
s.d.	4,8	2,9	7,3

Tabela 17. Zestawienie parametru helikalnego odległość między płaszczyznami sąsiadujących par zasad (ang. rise) (\AA) dla dupleksów (GGGGGGC)₂ (A), [GC(8-BrG)GGGGGGC]₂ (B) i [GC(8-BrG)GGGC]₂ (C). W panelach B i C podane są wartości dla różnych konformacji dupleksów RNA obserwowanych w danej strukturze.

łańcuch	AB	CD	EF	GH	IJ	KL	MN	OP	QR	SY	TU	VZ	WX	ab	cd	ef	gh	ij
G-C	3,1	3,1	3,0	3,2	3,4	3,3	3,1	3,1	3,0	3,1	3,2	3,2	3,4	3,0	3,3	3,4	3,0	3,1
C-G	3,0	3,0	2,8	2,7	3,1	3,0	2,8	3,0	2,7	2,9	2,9	2,7	3,1	3,0	3,1	3,2	3,1	2,9
G-G	3,1	2,8	2,9	3,1	2,7	2,7	3,1	3,2	3,1	2,8	3,0	2,9	2,9	3,0	2,9	3,0	3,1	3,1
G-C	3,4	3,0	3,4	3,5	3,1	3,0	3,5	3,5	3,5	3,3	3,5	3,6	3,2	3,3	3,2	3,2	3,6	3,2
C-G	2,9	2,8	2,8	3,0	2,9	2,8	3,0	3,1	2,9	3,1	3,1	3,1	2,9	3,0	2,9	3,0	2,9	3,0
G-G	3,0	3,1	2,8	2,9	3,0	3,0	3,0	3,0	2,9	2,9	3,0	2,8	3,1	3,0	3,1	3,2	2,9	3,0
G-C	3,2	3,1	3,2	3,1	3,2	3,2	2,9	3,2	2,9	2,9	3,0	3,0	3,3	3,1	3,3	3,4	3,2	3,1
C-G	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
średnia	3,1	3,0	3,0	3,1	3,0	3,0	3,0	3,1	3,0	3,0	3,1	3,1	3,1	3,1	3,1	3,2	3,1	3,1
s.d.	0,2	0,1	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,3	0,2	0,1	0,2	0,2	0,2	0,1

łańcuch	AB	CD	EE'	łańcuch	AB
G-C	3,1 / 3,0 / 3,4 / 3,4	3,3 / 3,3 / 3,4 / 3,4	3,0 / 3,1 / 2,8 / 2,8	G-C	2,9
C-G	3,2	3,3	2,5 / 2,6	C-G	3
G-Br-G	3,0 / 2,6	3,1 / 3,1 / 2,8 / 2,8	2,3 / 2,3 / 2,7 / 2,8	G-Br-G	2,6 / 2,7
G-C	3,0	3,0 / 2,8 / 3,0 / 2,8	2,8 / 2,9 / 3,0 / 3,1	G-C	2,9
C-G	2,8 / 2,8 / 2,8 / 2,9	2,7	3,1 / 2,7 / 3,2 / 2,8	C-G	2,9
G-G	2,8	2,9 / 3,1 / 2,9 / 3,1	2,7 / 3,1 / 2,8 / 3,2	G-GBr	2,5
G-C	3,3 / 3,2	3,5	2,9 / 2,8 / 3,1 / 3,0	G-C	2,9
C-G	3,0 / 2,9	3,1	2,3 / 2,3 / 2,8 / 2,7	C-G	---
G-GBr	2,6 / 2,8 / 2,6 / 2,7	2,9 / 2,8 / 2,8 / 2,8	2,5 / 2,5 / 2,6 / 2,6	średnia	2,8
G-C	3,0 / 2,9	2,7	3,1 / 3,0 / 2,8 / 2,8	s.d.	0,2
C-G	---	---	---		
średnia	3,0	3,0	2,8		
s.d.	0,2	0,3	0,3		



Rysunek 44. Oddziaływania warstwowe pomiędzy parami typu Watsona-Cricka (A) oraz pomiędzy parą G-C a parą G-G (B, C).

3.3.8. Efekt bromowania i dystrybucja konformacji par G-G

Wszystkie bromowane reszty guanozyn są w konformacji *syn*. Oznacza to, że utworzona jest zawsze para 8-BrG(*syn*)-G(*anti*). W przypadku dupleksów niemodyfikowanej struktury para G-G występuje w trzech możliwych ułożeniach: G(*syn*)-G(*anti*), G(*anti*)-G(*syn*) lub jako statycznie nieuporządkowana (ang. *statically disordered*) mieszanina dwóch powyższych konformacji (w dwóch z 36 par) (Tabela 18). W jednym dupleksie są dwie pary G-G, z których każda przyjmuje jedną z trzech powyższych aranżacji. Oznacza to, że w dupleksach obserwowane są różne kombinacje konformacji obu par G-G. Najbardziej preferowana jest kombinacja symetryczna (naprzemienna), tzn. za parą *anti-syn* jest para *syn-anti* lub na odwrót, za parą *syn-anti* jest para *anti-syn* (w 14 z 18 przypadków) (Tabela 18), przy czym pierwszy wymieniony układ występuje nieznacznie częściej (8 przypadków w stosunku do 6). W pozostałych czterech dupleksach, w dwóch jest ułożenie asymetryczne, a w kolejnych dwóch jedna z par G-G jest nieuporządkowana. W strukturze [GC(8-BrG)GCGGCGGC]₂ środkowe pary G-G są niemodyfikowane. W dupleksie C+D i E+E' para ta jest nieuporządkowana (mieszanina dwóch konformacji), natomiast w dupleksie A+B jest uporządkowana (Tabela 18).

Parametry helikalne struktur modyfikowanych i niemodyfikowanych są zbliżone do siebie (Tabele 14, 15, 16 i 17). Odzwierciedleniem tego jest wartość odchylenia rmsd otrzymana po nałożeniu na siebie dupleksów, która wynosi około 1 Å.

3.3.9. Oddziaływania z roztworem i hydratacja

Eksponowana w kierunku bruzdy dużej krawędź Watsona-Cricka reszty G(*syn*) oddziałuje z jonem siarczanowym. W niektórych przypadkach natywnego modelu oraz dłuższego modyfikowanego dupleksu anion jest uporządkowany, ale najczęściej jego orientacja jest nieokreślona. Można go jednak odróżnić od cząsteczki wody na podstawie rozmiaru i kształtu mapy gęstości elektro- nowej oraz na podstawie jego odległości od RNA (Rysunek 45). W wysokorozdzielczej strukturze, gdzie w medium krystalizacyjnym nie ma anionów siarczanowych, zlokalizowany jest dwuwarto-

Tabela 18. Dystrybucja konformacji par G-G w dupleksach w $(GGGGGGC)_2$ (A), $[GC(8-BrG)GGGGGGC]_2$ (B) i $[GC(8-BrG)CGGGC]_2$ (C).

A

łańcuch	AB	CD	EF	GH	IJ	KL	MN	OP	QR	SY	TU	VZ	WX	ab	cd	ef	gh	ij
G-C																		
C-G																		
G-G	anti-syn	syn-anti	anti-syn	nieuporządkowana	syn-anti	anti-syn	anti-syn	anti-syn	anti-syn	anti-syn	anti-syn	anti-syn	syn-anti	anti-syn	syn-anti	syn-anti	anti-syn	anti-syn
G-C																		
C-G																		
G-G	syn-anti	anti-syn	anti-syn	syn-anti	anti-syn	anti-syn	syn-anti	syn-anti	syn-anti	syn-anti	syn-anti	syn-anti	anti-syn	syn-anti	anti-syn	anti-syn	nieuporządkowana	anti-syn
G-C																		
C-G																		

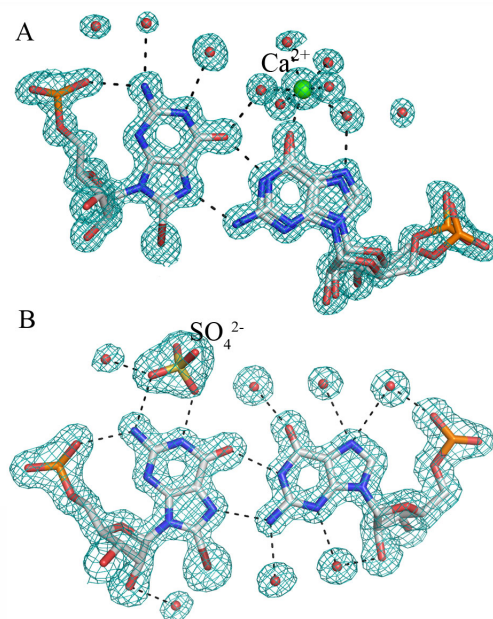
B

łańcuch	AB	CD	EE'
G-C			
C-G			
G-Br-G	syn-anti	syn-anti	syn-anti
G-C			
C-G			
G-G	anti-syn	nieuporządkowana	nieuporządkowana
G-C			
C-G			
G-GBr	anti-syn	anti-syn	anti-syn
G-C			
C-G			

C

łańcuch	AB
G-C	
C-G	
G-Br-G	syn-anti
G-C	
C-G	
G-GBr	anti-syn
C-G	

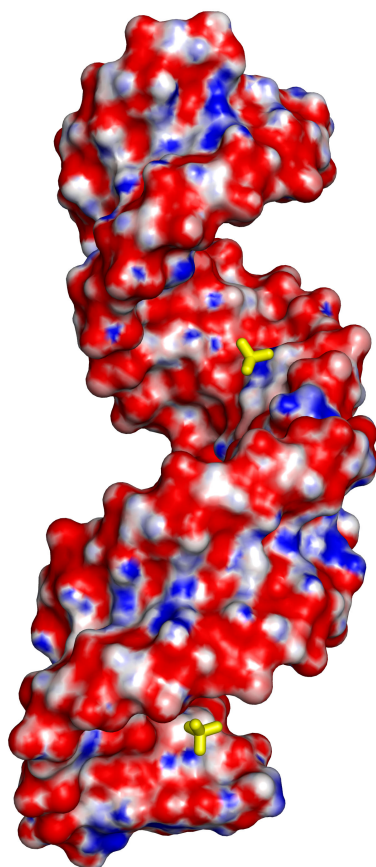
ściowy kation wapnia. Sferę koordynacyjną kationu stanowi cztery lub pięć cząsteczek wody oraz atom karbonylowy G(*anti*) (Rysunek 45). Długości wiązań wodorowych wynoszą 2,2–2,6 Å. Gdy jonu siarczanowego nie ma, krawędź Watsona-Cricka reszty G(*syn*) jest hydratowana przez trzy cząsteczki wody, które razem z guanozyną tworzą w jednej płaszczyźnie „grzebień”. Analogiczna sytuacja występuje wtedy, gdy nie ma jonu wapnia. Z atomami tworzącymi krawędź Hoogsteena oddziałują dwie lub trzy uporządkowane cząsteczki wody. Interesujący jest fakt, że we wszystkich strukturach, w których obecny jest jon siarczanowy, bruzda duża ma szerokość 18 Å, natomiast dla posiadających Ca²⁺ jest węższa prawie o 4 Å.



Rysunek 45. Oddziaływania par G-G z cząsteczkami rozpuszczalnika. W strukturze dupleksu [GC(8-BrG)GCGGC]₂ związał się jon wapnia (zielony) (A), który oddziałuje z pięcioma cząsteczkami wody oraz z grupą karbonylową reszty G(*anti*). W pozostałych dwóch strukturach eksponowana krawędź Watsona-Cricka reszty G(*syn*) oddziałuje z jonem siarczanowym (B).

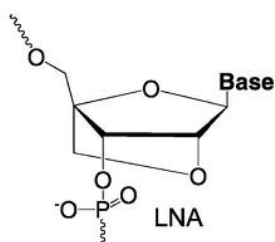
3.3.10. Powierzchnia potencjału elektrostatycznego

Rozkład potencjału elektrostatycznego wygląda podobnie jak w RNA zawierających powtórzenia CUG i CAG. Pionowe pasy potencjału dodatniego i ujemnego przeplatają się wzdłuż bruzdy małej (Rysunek 46). Bruzda duża jest elektroujemna, ale znajdują się w niej obszary o potencjale dodatnim generowane przez eksponowaną krawędź Watsona-Cricka G(*syn*) oraz poprzedzającą ją grupę *exo-aminową* reszty cytydyny. Miejsce wiązania jonu siarczanowego odpowiada elektrodo-datnym właściwościom G(*syn*), które wydają się być większe niż dla sąsiedniej cytozyny nieprzy-ciągającej żadnego anionu.



Rysunek 46. Przykład powierzchni potencjału elektrostatycznego obliczonej dla trzech dupleksów $[GC(8-BrG)GCGGCGGC]_2$. Jony siarczanowe (żółty) wiążą się do miejsc o potencjale dodatnim generowanych przez eksponowaną krawędź Watsona-Cricka reszty G(syn).

3.4. Struktury RNA zawierające powtórzenia CCG



Rysunek 47. Struktura reszty nukleotydowej z serii LNA.

Określenie struktury powtórzeń CCG okazało się zadaniem najtrudniejszym począwszy od krystalizacji po proces udokładniania. Ostatecznie rozwiązałam dwie struktury dupletu $(GCCGCCGC)_2$ (rozdzielczość 1,54 Å) oraz $(GCCG^LCCGC)_2$ (rozdzielczość 1,95 Å) zawierającego modyfikowaną resztę 4G z serii LNA (G^L). LNA jest syntetycznym analogiem reszty nukleotydowej, który zawiera mostek metylenowy łączący atom tlenu 2' pierścienia rybozy z atomem węgla 4' (Rysunek 47). „Zamknięcie” pierścienia cukrowego powoduje, że jego konformacja, typowa dla formy A kwasów nukleinowych (C3'-endo), zostaje usztywniona. Zauważono, że wzmacnia to oddziaływania warstwowe pomiędzy zasadami azotowymi nukleotydów oraz ułatwia rearanżację szkieletu fosforanowo-cukrowego. Innymi słowy zwiększa to zdolność do hy-

brydyzacji oligomeru. Z powyższych względów wprowadziłam nukleotyd z serii LNA, aby ułatwić tworzenie się dupleksu i w efekcie jego krystalizację.

3.4.1. Krystalizacja RNA zawierającego powtórzenia CCG

Oba liofilizowane oligomery rozpuszczane były w 200 mM lub 400 mM octanie amonu. Stężenie końcowe RNA wynosiło 1 mM. Następnie przeprowadzałam inkubację w temperaturze 95 °C przez 5–10 minut. Oligomer schładzany był do temperatury pokojowej przez przynajmniej 20 minut. Do niemodyfikowanego RNA przed krystalizacją dodawałam sperminę do stężenia końcowego 25 mM. Krystalizował on w temperaturze 30 °C z roztworu zawierającego 10 mM octan magnezu, MES pH 5,6 i 2,5 M $(\text{NH}_4)_2\text{SO}_4$. Oligomer GCCG^LCCGC wykazywał bardzo duży potencjał do krystalizacji. Udało się otrzymać monokryształy w różnych warunkach w temperaturze 19 °C, ale największe i najlepiej rozprasające krystalizowały w 10 mM MgCl_2 , MES pH 5,6 oraz 1,8 M Li_2SO_4 . RNA mieszany był z roztworem rezerwuarowym w stosunku 1:1.

3.4.2. Pomiar danych dyfrakcyjnych i ich przetwarzanie

Pomiary dyfrakcyjne prowadzone były w ośrodku synchrotronowym BESSY w Berlinie na linii BL14.1 (modyfikowany oligomer) i BL14.2 (niemodyfikowany oligomer) w temperaturze 100 K. Jako krioprotektant użyto 20% glicerol. Dla kryształu dupleksu $(\text{GCCGCCGC})_2$ zarejestrowano dane o sumarycznym kącie oscylacji 360 ° do rozdzielczości 1,54 Å. Obrazy charakteryzują się małą liczbą refleksów, co w efekcie, po przetworzeniu danych w grupie przestrzennej R3 daje tylko 5119 unikalnych obserwacji. Dla kryształu oligomeru modyfikowanego zarejestrowano również dane z sumarycznym kątem oscylacji 360 ° (Tabela 19). Grupę przestrzenną wyznaczono jako P4₁22 a maksymalna rozdzielczość wyniosła 1,95 Å (tabela). Podczas rejestracji danych zauważyłam, że refleksy są podłużne, niekiedy o kształcie przecinków. Taka obserwacja może wskazywać na obecność efektu bliźniaczenia lub może świadczyć o lokalnym nieuporządkowaniu w sieci krystalicznej. Po przetworzeniu danych i przeanalizowaniu ich za pomocą programu PHENIX.XTRIAGE nie stwierdziłam bliźniaczenia.

3.4.3. Rozwiązanie struktur zawierających powtórzenia CCG

Obie struktury zostały rozwiązane za pomocą programu PHASER. W obu przypadkach użytym modelem był dupleks $(\text{GCUGCUGC})_2$ lub jego fragment składający się z czterech par. Rozwiązaniem struktury niemodyfikowanej był jeden dupleks w części asymetrycznej, którego z-score wynosił 8,3 a LLG = 231. Zaskakujące było, że w obliczoną gęstość elektronową można było wbudować dwie nici ale, składające się z siedmiu zamiast ośmiu nukleotydów. Ponadto trudno było znaleźć początek i koniec dupleksu. Mapa gęstości tworzyła ciągłe kolumny helis, których najdłuższe osie były ułożone równoległe do osi trójкратно. W związku z tym, że oligomer na

Tabela 19. Zestawienie statystyk otrzymanych po przetworzeniu danych dyfrakcyjnych oraz z procesu udokładniania struktur zawierających powtórzenia CCG.

Oligomer	GCCGCCGC	GCCG ^L CCGC
Źródło promieniowania	BESSY BL 14.2	BESSY BL 14.1
Długość fali (Å)	0,91841	0,918
Grupa przestrzenna	R3	P4 ₁ 22
Parametry komórki (Å)	a = 40,6 c = 56,1	a = 42,0 c = 81,7
Rozdzielczość (Å)	30,0–1,54 (1,57–1,54)	20,0–1,95 (1,98–1,95)
Mozaikowość (°)	0,3	0,85
R _{merge}	0,051 (0,928)	0,067 (0,673)
<I/σ(I)>	40,5 (3,2)	37 (2,2)
Kompletność (%)	99,9 (100)	99,8 (97,5)
Liczba unikalnych refleksów	5119	5762
Krotność danych	16,4 (12,3)	28 (15,6)
Refleksy > 3σ (%)	82,6 (30,9)	77,8 (27,5)
Czynnik B z wykresu Wilsona (Å ²)	29,7	50,7
R (%)	24,35	24,9
R _{free} (%)	26,79	26,35
Liczba atomów cząsteczek RNA	303	514
Liczba cząsteczek jonu siarczanowego	–	2
Liczba cząsteczek wody	19	19
Odchylenie rmsd od wartości idealnej		
długości wiązań (Å)	0,023	0,018
kąty (°)	2,05	2,609

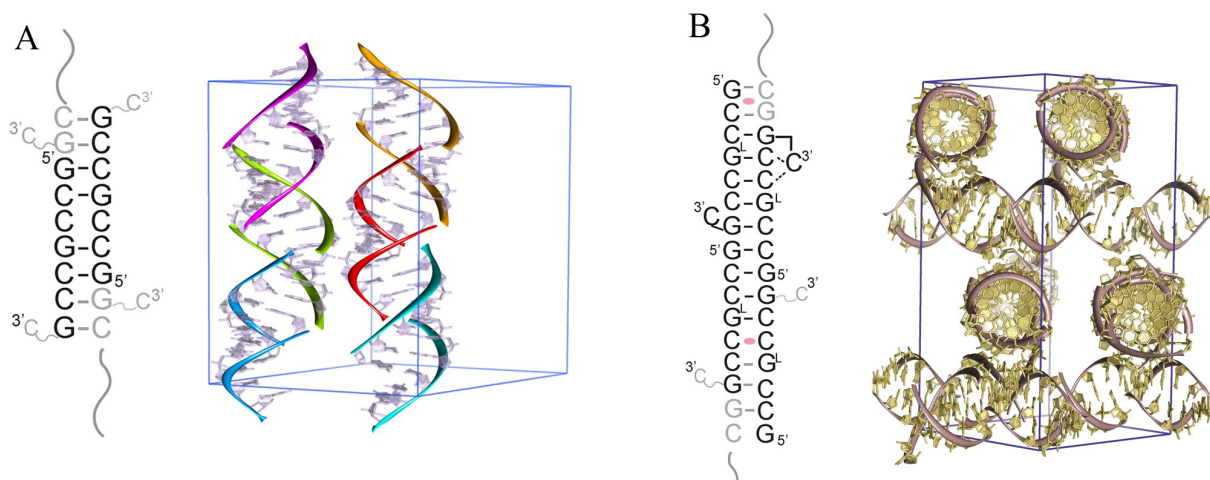
końcu 5' nie zawierał grupy fosforanowej w mapie gęstości elektronowej powinna być „przerwa”. Rozwiązanie tej zagadki nastąpiło po rozwiązaniu struktury dupleksu (GCCG^LCCGC)₂. W tym przypadku przy obliczeniach przeprowadzanych za pomocą programu PHASER, kluczowe okazało się wprowadzenie ograniczenia zakresu rozdzielczości do 3,2 Å.

3.4.4. Udokładnianie struktur zawierających powtórzenia CCG

Proces udokładniania obu struktur był trudnym etapem badań. Podstawowym problemem były zbyt duże rozbieżności pomiędzy czynnikami R. Przyczyną był zbyt mały zbiór refleksów wybrany do obliczenia statystyk R_{free}, który powinien zawierać około 1000 refleksów. Początkowo nie wyznaczyłam zbioru R_{free} o zalecanej ilości ponieważ wykluczało to w tym przypadku aż jedną piątą (20%) zarejestrowanych refleksów z procesu udokładniania. Jednak powyższe względy spowodowały, że zbiór ten odpowiednio powiększyłam do obliczania czynnika R_{free}. Proces udokładniania prowadziłam za pomocą programu PHENIX. Słownik zawierający opis stereochemiczny reszty G^L wygenerowałam przy pomocy serwera PRODRG2 (<http://davapc1.bioch.dundee.ac.uk/prodrng/submit.html>). Statystyki zebrane są w Tabeli 19.

3.4.5. Upakowanie cząsteczek w sieci krystalicznej

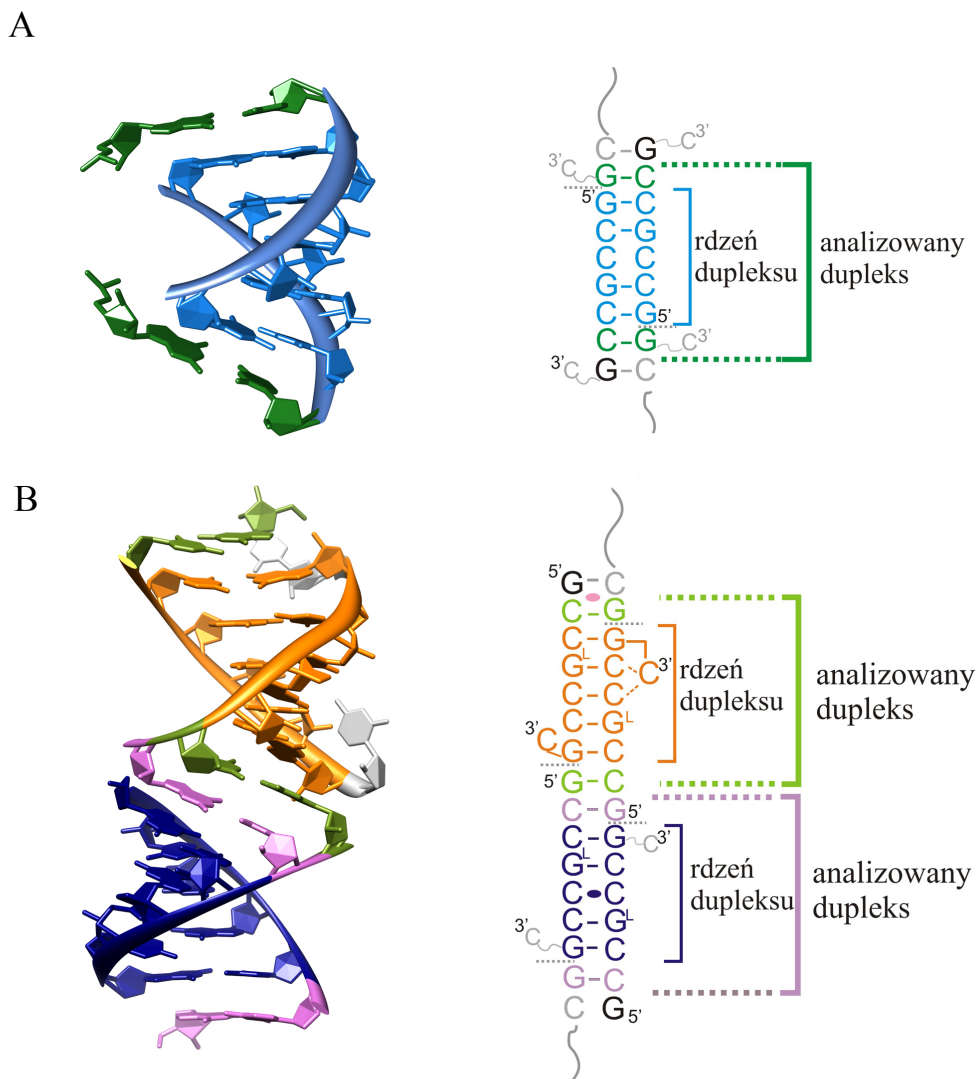
W części asymetrycznej struktury romboedrycznej (niemodyfikowanej) znajdują się dwie nici: A i B, które względem siebie są przesunięte o trzy reszty nukleotydowe. Dodatkowo ostatnia reszta 8C w obu łańcuchach jest nieuporządkowana i jej lokalizacja nie została ustalona. W związku z powyższym dupleks składa się z 5 par zasad. Taką strukturę można opisać jako „poślizgnięty” dupleks z niesparowanymi końcami 3'. W sieci krystalicznej niesparowane reszty 6C i 7G parują się ze swoimi symetrycznymi odpowiednikami tworząc 4 dodatkowe pary (Rysunek 48). Powstające „na zakładkę” pseudonieskończone helisy układają się wzdłuż osi trójkrotnych. Podobnie jest w strukturze tetragonalnej, w której oligomer tworzy identyczne, „poślizgnięte” dupleksy przy czym w części asymetrycznej znajdują się trzy nici: C, D i E. Jedna z helis (E+E') utworzona jest przez dwa symetryczne względem siebie łańcuchy. Interesujące jest, że w dupleksie C+D reszty 8C są uporządkowane i zaangażowane w oddziaływania wewnątrz i międzycząsteczkowe. Natomiast w nici E, podobnie jak w strukturze romboedrycznej, reszta 8C jest nieuporządkowana. W tetragonalnej sieci krystalicznej są dwa pasma pseudonieskończonych helis. Oba biegną równolegle do płaszczyzny a-b ale względem siebie są obrócone o 90° (Rysunek 48). Symetryczne helisy oddziałują ze sobą bruzdami małymi tworząc wiązania wodorowe oraz warstwowe (utworzone pomiędzy resztami 8C łańcuchów C i D).



Rysunek 48. Struktura drugorzędowa oraz upakowanie cząsteczek w sieci krystalicznej struktury romboedrycznej (niemodyfikowanej) (A) oraz tetragonalnej (modyfikowanej) (B). Szarym zaznaczono nici symetryczne oraz nieuporządkowane reszty 8C. Krystalograficzna oś dwukrotna zaznaczona jest na różowo.

3.4.6. Konformacja RNA zawierające powtórzenia CCG

W celu ułatwienia jednoznacznego opisu i dalszych porównań wyznaczona przeze mnie helisa składa się z 7 par zasad (analizowany dupleks) (Rysunek 49), na które przypada 6 par G-C oraz 1 para C-C. W takim dupleksie wyróżnia się dodatkowo jego rdzeń (zawiera 5 par zasad), w którym zachowana jest ciągłość łańcucha fosforanowo-cukrowego.



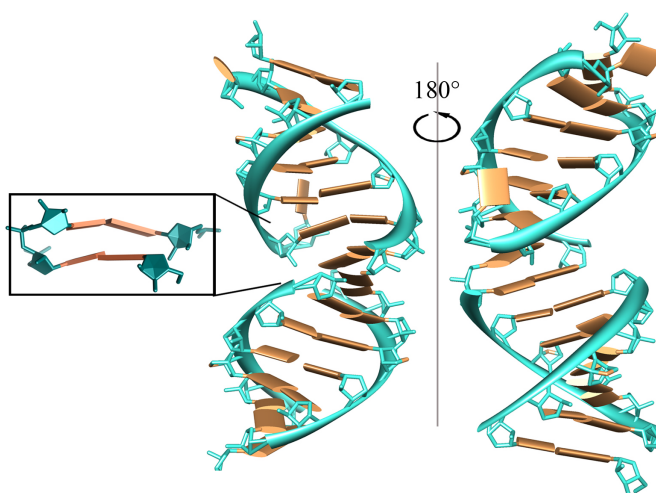
Rysunek 49. Struktura przestrzenna oraz drugorzędowa analizowanych dupleksów zawierających powtórzenia CCG: niemodyfikowanego (A) i modyfikowanych (B). Każdy z dupleksów składa się z 7 par zasad z czego 5 par stanowi rdzeń dupleksu, który ma zachowaną ciągłość łańcucha fosforanowo-cukrowego. Oś dwukrotna oznaczona jest elipsą.

W niemodyfikowanej strukturze helisa ma formę A-RNA, gdzie wszystkie reszty nukleotydowe mają konformację C3'-endo a średnia wartość parametru Z_p jest równa $2,5 \pm 0,06$ Å. Średnia wartość kąta skrętu helisy jest równa $34,7 \pm 3,3^\circ$, natomiast dla rdzenia dupleksu wynosi ona $33,3 \pm 2,4^\circ$ (10,8 par zasad na jeden skręt helisy) (Tabela 20). Podobnie w modyfikowanej strukturze, dupleks E+E' przyjmuje formę A (Z_p wynosi $2,5 \pm 0,03$ Å), w której wartość kąta skrętu helisy wynosi $33,7 \pm 0,4^\circ$ (dla rdzenia helisy $33,8 \pm 0,5^\circ$). W dupleksie C+D 11 z 14 reszt nukleotydowych ma pierścień cukrowy w konformacji C3'-endo, jedna C2'-exo oraz dwie C2'-endo (konformacja typowa dla formy B-DNA), przy czym wartości parametru Z_p są większe od 1,6 Å. Ze względu na ten fakt drugi dupleks, a przynajmniej jego rdzeń, przyjmuje również formę A-RNA, jednakże pewne zniekształcenia helisy są widoczne (Rysunek 50). Odzwierciedlone jest to w takich

parametrach jak: *propeller* i *buckle*, które przyjmują szeroki zakres wartości odpowiednio: od $-19,5^\circ$ do $2,6^\circ$ oraz -16° do $22,4^\circ$ (Tabela 21). Dla porównania dla wysokorozdzielczych struktur A-DNA wynoszą one $-8,3^\circ$ oraz $-2,4^\circ$ (131). Równie dużym zakresem wartości charakteryzuje się skręt helisy ($33,6 \pm 11,8^\circ$) oraz w mniejszym stopniu odległość pomiędzy płaszczyznami sąsiadujących zasad (ang. *rise*) – $3,0 \pm 0,7 \text{ \AA}$. W duplesie E+E' parametry te przyjmują bardziej zbliżone wartości aczkolwiek *propeller* wynosi $-14,4 \pm 4,9^\circ$ i jest wyższy od typowej wartości.

Tabela 20. Parametry helikalne struktur krystalicznych dupleksów $(GCCGCCGC)_2$ i $(GCCG^L CCGC)_2$. Odchylenie standardowe oznaczono jako s.d.

Para	Przemieszczenie (Å)			Kąt (°)			Kąt skrętu helisy (°)			Odległość między płaszczyznami zasad (Å)		
	A+B	C+D	E+E'	A+B	C+D	E+E'	A+B	C+D	E+E'	A+B	C+D	E+E'
G-C	5,5	7,9	5,2	15,6	12,4	9,1	40,0	39,0	33,5	2,5	2,1	2,7
G-C	6,5	7,2	5,3	13,8	16,1	8,7	32,2	30,2	33,8	2,7	3,6	3,2
C-G	6,9	6,3	4,9	12,8	10,0	8,1	31,6	29,3	33,4	2,7	2,6	2,9
C-C	6,7	5,2	5,6	13,3	8,5	7,3	32,6	40,4	33,4	2,7	2,6	2,8
G-C	6,4	4,2	5,4	14,2	12,7	9,9	36,9	14,3	34,4	2,4	4,1	3,1
C-G	6,5	4,9	6,2	12,8	15,7	12,1	34,9	48,4	33,4	2,5	2,8	2,5
C-G	6,6	6,1	6,4	10,8	9,0	12,4	---	---	---	---	---	---
średnia	6,4	6,0	5,6	13,3	12,1	9,7	34,7	33,6	33,7	2,6	3,0	2,9
s.d.	0,4	1,3	0,5	1,5	3,1	1,9	3,3	11,8	0,4	0,1	0,7	0,3
rdzeń	6,6	5,6	5,5	13,4	12,6	9,2	33,3	28,6	33,8	2,6	3,2	3,0
s.d.	0,2	1,2	0,5	0,6	3,4	1,9	2,4	10,8	0,5	0,1	0,8	0,2



Rysunek 50. Struktura krystaliczna duplesu modyfikowanego. Helisa ma formę A-RNA przy czym niektóre parametry helikalne przyjmują nietypowe wartości. W ramce przedstawiono przykład, w którym parametr *buckle* przyjmuje wartości powyżej 12° .

3.4.7. Niestandardowe pary C-C

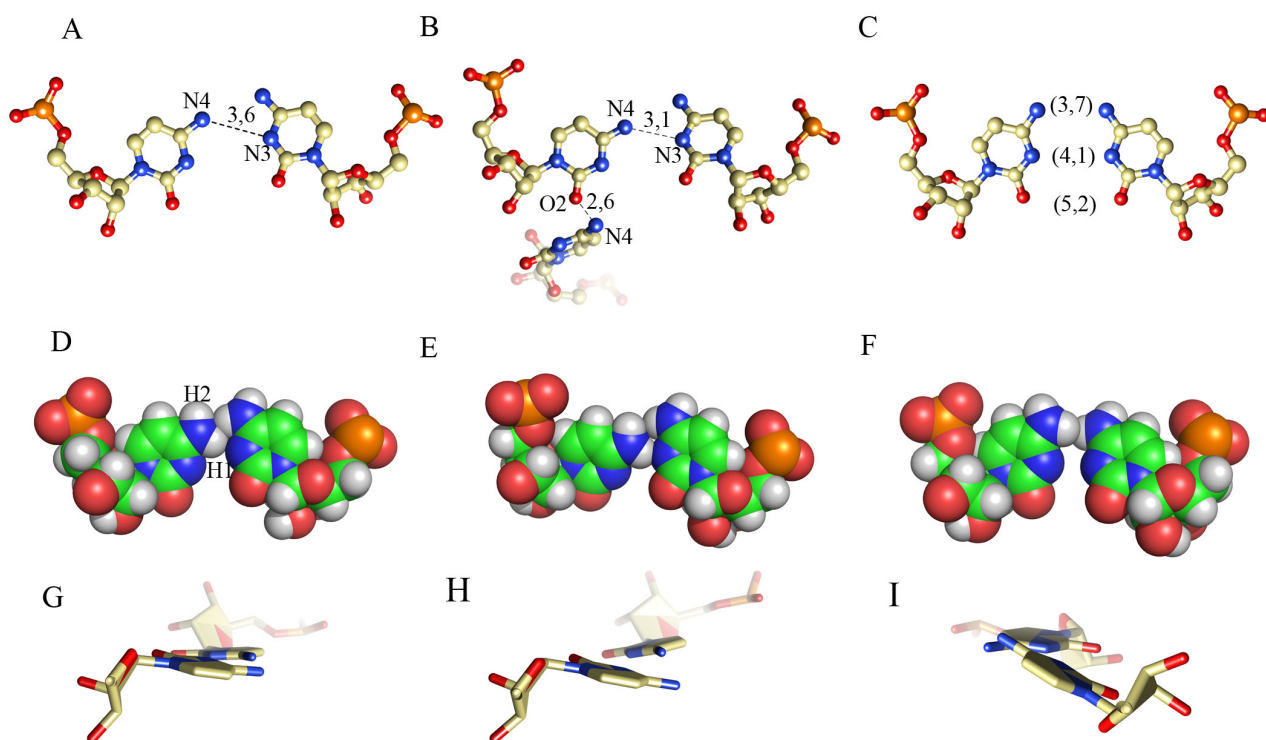
W każdym z trzech dupleksów znajduje się jedna para C-C i każda z nich paruje się w inny sposób. W strukturze niemodyfikowanej jedna z reszt cytydyn nachyla się w kierunku bruzdy

Tabela 21. Parametry helikalne struktur krystalicznych dupleksów $(GCCGCCGC)_2$ $(GCCG^L CCGC)_2$. Odchylenie standardowe oznaczono jako s.d.

Para	Propeller (°)			Buckle (°)			Opening (°)		
	A+B	C+D	E+E'	A+B	C+D	E +E'	A+B	C+D	E+E'
G-C	-13,0	-7,7	-12,7	-1,7	-4,4	0,6	2,6	-3,2	-2,2
G-C	-11,6	0,6	-9,7	-1,7	-16,0	-5,4	0,1	-5,2	-1,2
C-G	-9,1	-14,3	-18,0	-1,4	-0,7	-5,3	-0,4	-0,4	-2,7
C-C	-11,9	-19,5	-21,6	-1,2	-13,0	0,0	-12,5	-22,9	-6,2
G-C	-7,8	-14,3	-18,0	4,5	-12,3	5,3	0,3	-3,1	-2,7
C-G	-14,4	2,6	-8,2	4,9	22,4	6,4	-0,1	-0,6	-1,5
C-G	-14,3	-9,3	-12,7	2,4	12,1	-0,6	5,3	-0,8	-2,2
średnia	-11,7	-8,8	-14,4	0,8	-1,7	0,1	-0,7	-5,2	-2,7
s.d.	2,5	8,1	4,9	3,0	14,3	4,6	5,6	8,0	1,6
rdzeń	-11,0	-9,0	-15,1	1,0	-3,9	0,2	-2,5	-6,4	-2,9
s.d.	2,6	9,9	5,8	3,4	15,8	5,6	5,6	9,4	2,0

małej. Jej kąt λ równy jest $38,9^\circ$ a parametr *opening* wynosi $-12,5^\circ$ i jest to najmniejsza wartość wyznaczona dla wszystkich par w dupleksie (Tabela 21). Reszty oddziałują słabym wiązaniem wodorowym ($3,6 \text{ \AA}$), które utworzone jest pomiędzy grupą *exo*-aminową N4 nachylonej cytydyny a atomem N3 drugiej reszty. W takiej konformacji grupy *exo*-aminowe obu reszt znajdują się w bliskim sąsiedztwie w wyniku czego dochodzi między nimi do oddziaływania. Odległość pomiędzy atomami N4 wynosi $3,3 \text{ \AA}$, a pomiędzy atomami wodoru H1 (znajduje się po stronie krawędzi Watsona-Cricka) nachylonej reszty i H2 (znajduje się po stronie krawędzi Hoogsteena) drugiej $2,2 \text{ \AA}$. Grupy te dopasowały się zatem swoim kształtem do siebie. Wspomniany atom H2 znajduje się w przestrzeni utworzonej pomiędzy atomami wodoru drugiej grupy i oddziałuje z atomem azotu. Można to opisać jako „zakleszczenie” grup *exo*-aminowych (Rysunek 51).

W dupleksie C+D struktury $(GCCG^L CCGC)$ również jedna z reszt cytydyn (z nici D) jest nachylona w kierunku bruzdy małej ($\lambda = 30,8^\circ$; *opening* = $-22,9^\circ$). Konformacja ta stabilizowana jest przez oddziaływanie grupy karbonylowej nachylonej cytydyny z grupą *exo*-aminową reszty 8C (nić C), która ulokowana jest w bruzdzie małej tego dupleksu. Pomiedzy parą C-C jest jedno wiązanie wodorowe ($3,1 \text{ \AA}$) utworzone pomiędzy atomami N4 nachylonej reszty a N3 drugiej cytydyny (Rysunek 51). Atomy azotu obu grup *exo*-aminowych znajdują się dalej od siebie ($3,5 \text{ \AA}$) niż w wyżej opisanej parze C-C, natomiast dystans pomiędzy atomami H1 nachylonej cytydyny a H2 drugiej reszty C nieznacznie się zwiększył (o $0,1 \text{ \AA}$). Wzajemne ułożenie grup *exo*-aminowych jest odmienne. Atomy azotu oraz atomy H1 i H2 (Rysunek 51 panel D) nie znajdują się w jednej płaszczyźnie (Rysunek 51 panel H). Mijają się w przestrzeni, co odzwierciedla parametr *propeller*, który wynosi $-19,5^\circ$ (Tabela 21). Oznacza to, że płaszczyzny pierścieni zasad są w znacznym stopniu skręcone względem siebie (jak śmigła samolotu).



Rysunek 51. Pary C-C dupleksu A+B (A, D, G), C+D (B, E, H) oraz E+E' (C, F, I). Każda z par C-C oddziałuje w inny sposób. W strukturze romboedrycznej pomiędzy resztami jest prawdopodobnie jedno słabe wiązanie wodorowe (A). Atomy wodorów reszt *exo-aminowych* są w bliskim sąsiedztwie. Atom H1 nienachylonej cytydyny jest zakleszczony pomiędzy atomami wodorów drugiej cytydyny (D). W strukturze tetragonalnej jedna z par tworzy również jedno wiązanie wodorowe (B), przy czym atomy wodorów mijają się w przestrzeni na skutek skręcenia płaszczyzn zasad względem siebie (E, H). Druga para struktury tetragonalnej prawdopodobnie nie tworzy wiązań wodorowych (C). Jedynie atomy wodorów H1 są w bliskim sąsiedztwie (F) ale znajdują się w innych płaszczyznach (I).

Trzecia para C-C znajdująca się w dupleksie E+E' jest symetryczna nie tylko chemicznie, ale również w sposobie parowania się oraz krystalograficznie. Obie reszty są lekko nachylone w kierunku bruzdy małej ($\lambda = 46,5^\circ$; $\text{opening} = -6,2^\circ$) (Rysunek 51). Te same grupy funkcyjne obu reszt, które należą do krawędzi Watsona-Cricka, są naprzeciw siebie. Dystans pomiędzy poszczególnymi grupami wynosi dla $\text{N4} \cdots \text{N4}$ 3,7 Å, dla $\text{N3} \cdots \text{N3}$ 4,1 Å i dla $\text{O2} \cdots \text{O2}$ 5,2 Å. W przypadku atomów N4 jest jeszcze możliwe istnienie słabego oddziaływania pomiędzy nimi, natomiast pomiędzy pozostałym atomami jest to mało prawdopodobne. Płaszczyzny zasad są skręcone względem siebie o $-21,6^\circ$ (Tabela 21), co powoduje, że atomy wodorowe H1 w pewnym stopniu mijają się w przestrzeni pomimo bliskiej odległości wynoszącej 1,9 Å.

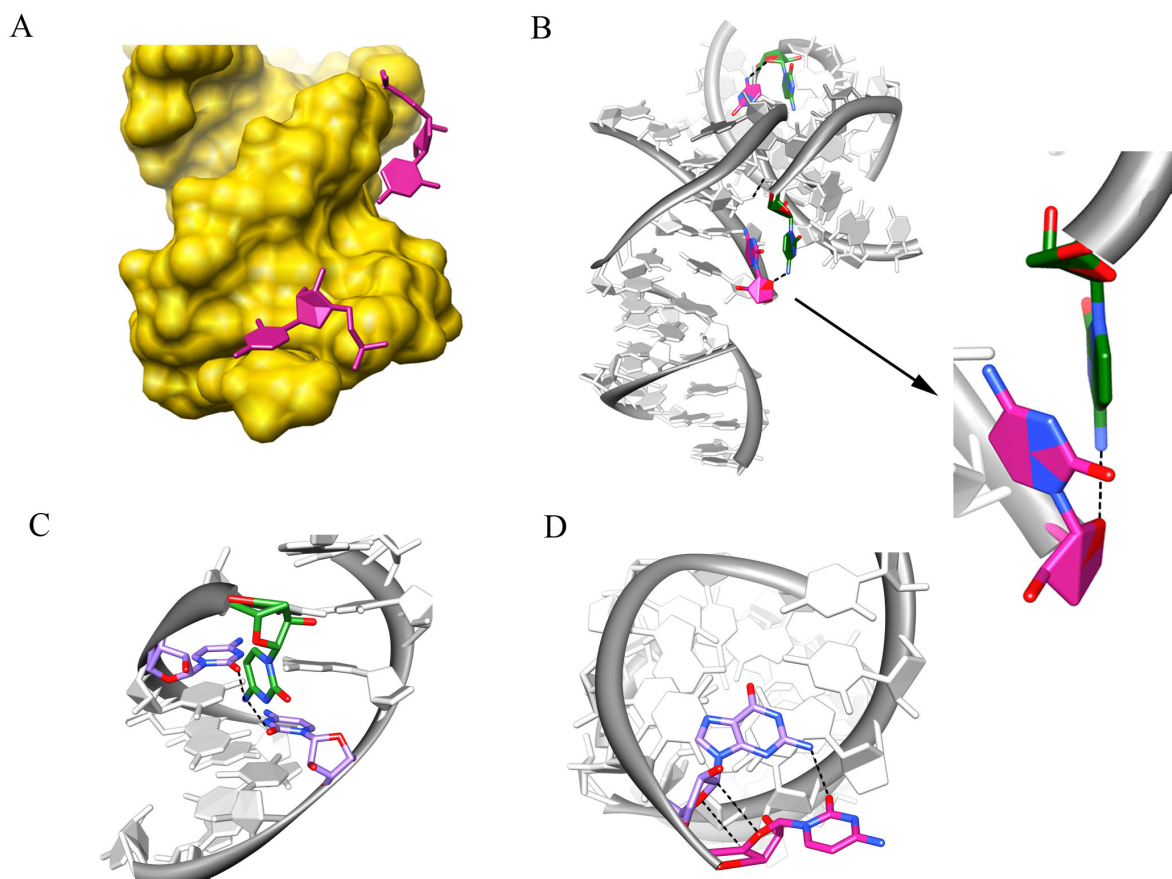
Wspólną cechą par C-C jest to, że tworzą one region dwuniciowy podobnie jak pary G-C. Odległość pomiędzy atomami C1' sparowanych reszt cytydyn wynosi od 10,7–10,9 Å i jest ona nieznacznie większa od średniej odległości C1'-C1' par G-C (10,6 Å) w obu strukturach krystalicznych.

3.4.8. Uporządkowane pary 8C

W duplekcie C+D struktury tetragonalnej reszty 8C są uporządkowane. Obie reszty „wywijają się” na zewnątrz helisy, lokując się w bruzdzie małej (Rysunek 52). Pierścień cukrowy reszty 8C z łańcucha C jest ułożony za zewnątrz w stosunku do bruzdy w kierunku rozpuszczalnika, natomiast jego zasada azotowa skierowana jest w stronę bruzdy małej. Można powiedzieć, że reszta ta niejako „nurkuje” w kierunku powierzchni bruzdy małej. Z kolei druga reszta „leży” swoją stroną 5' na powierzchni bruzdy. Stroną 3' („przód” nukleotydu) oddziałuje natomiast z bruzdą małą symetrycznego dupleksu C'+D'. Wywinięcie się reszt powoduje zmianę biegu łańcucha fosforanowo-cukrowego oraz zmianę konformacyjną pierścienia rybozy reszt 7G z typowej dla A-RNA C3'-*endo* na C2'-*endo* (typowej dla formy B-DNA). Reszta 8C z łańcucha C tworzy wiązania wodorowe z dwiema resztami: 5C z łańcucha D oraz z 6C z łańcucha C. Dwa wiązania wodorowe utworzone są pomiędzy grupą *exo*-aminową wywiniętej reszty a grupami karbonyłowymi cytydyn (Rysunek 52). Druga reszta 8C oddziałuje z sąsiednią resztą 7G za pomocą trzech oddziaływań wodorowych utworzonych pomiędzy grupą karbonylową O2 cytydyny a grupą *exo*-aminową guanozyny oraz pomiędzy atomami pierścieni cukrowych: O4'...C1' i C4'...O4'. Obie reszty 8C oddziałują również ze sobą w sieci krystalicznej. Tworzą one oddziaływania warstwowe typu *edge-to-face* przy czym kąt pomiędzy pierścieniami cytozyn wynosi około 60°. Dodatkowo oddziałują ze sobą wiązaniem wodorowym utworzonym pomiędzy atomem O2' reszty z łańcucha C i atomem N4 drugiej cytydyny (Rysunek 52). Symetryczne dupлексы oddziałują ze sobą w dwóch miejscach resztami 8C i wraz z jeszcze jednym wiązaniem wodorowym, pomiędzy atomami O2' reszt 5C, tworzą miejsce kontaktu między tymi dupleksami w sieci krystalicznej.

3.4.9. Modyfikowane reszty 4G z serii LNA

W strukturze tetragonalnej reszty 4G są nukleotydami z serii LNA. Ze względu na obecność mostka metylenowego ich pierścień cukrowy ma usztywnioną konformację C3'-*endo*, która odpowiada formie A-RNA. Wprowadza to ograniczenia w ruchomości kątów torsyjnych rybozy, natomiast pozostawia w dalszym ciągu labilność wiązania fosfodiesterowego. Kąty torsyjne reszt G^L w łańcuchu C oraz E są typowe natomiast dla reszty G^L w łańcuchu D kąty α (określa rotację wokół wiązania P-O5'), β (określa rotację wokół wiązania O5'-C5') i γ (określa rotację wokół wiązania C5'-C4') mają inne wartości niż typowe dla formy A-RNA. Kąt α wynosi 114° (typowa wartość -68°), β jest równy -120,5° (typowa wartość 178°) a γ ma -156,3° (typowa wartość 54°). Odkształcenie tego fragmentu łańcucha polega na obrocie prawie o 180° wiązania O5'-C5' oraz na translacji atomów P i C5' względem pierścienia rybozy (Rysunek 53). Typowo atomy P, O5', C5' i C4' są niemalże w jednej linii, natomiast w obserwowanej strukturze nie można poprowadzić prostej przez te atomy. Taka konformacja łańcucha fosforanowego wypycha nieco resztę G^L w kierunku bruzdy dużej w stosunku do sąsiadujących z nią po obu stronach reszt cytydynowych (Rysunek 53). Rzutuje to na wartość kąta skrętu helisy, który pomiędzy parą G-C i G^L-C wynosi

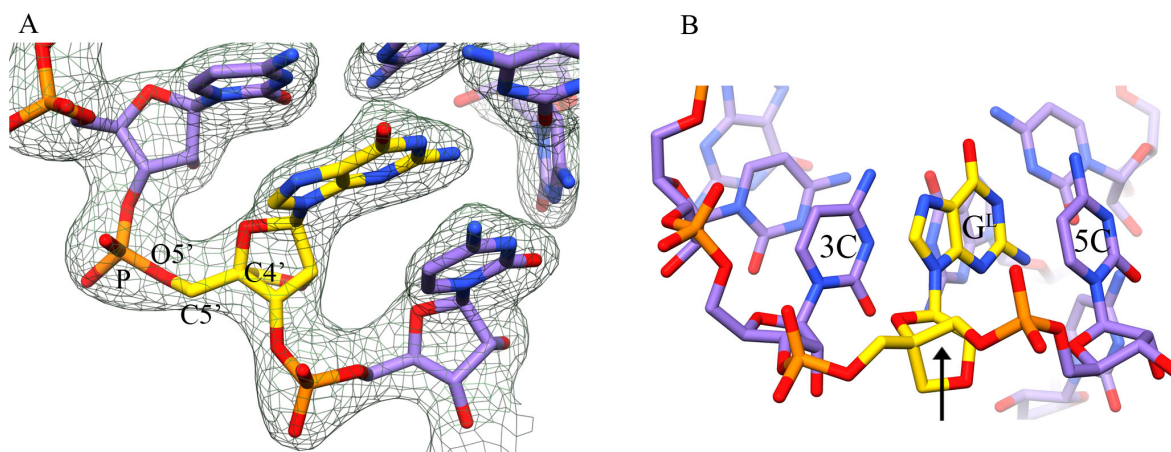


Rysunek 52. Uporządkowane reszty 8C dupletu C+D znajdują się na zewnątrz helisy w bruzdzie małej(A). Nukleotyd 8C z nici C oddziałuje z dwiema resztami cytydynowymi (C), natomiast z nici D tworzy trzy wiązania wodorowe z sąsiednią resztą guanozyny (D). Pomiędzy dwiema symetrycznymi helisami są trzy miejsca kontaktu: dwa oddziaływania warstwowe wywiniętych reszt oraz jedno wiązanie wodorowe pomiędzy grupami hydroksylowymi reszt 5C (B).

zaledwie $14,3^\circ$ (Tabela 20). Lokalne rozkręcenie helisy jest kompensowane w strukturze dupletu w innych miejscach. Jedynym parametrem helikalnym, jednoznacznie skorelowanym z obecnością modyfikowanych reszt, jest *roll* (Tabela 22). Jego wartość, wyznaczona pomiędzy sąsiadującymi parami C-G i G^L -C, jest większa o $10,4^\circ$ od średniej wyznaczonej dla pozostałych skoków (ang. *step*) helisy ($3,8 \pm 2,2^\circ$).

3.4.10. Oddziaływania z rozpuszczalnikiem

Ze względu na słabszą jakość danych krystalograficznych, w obu strukturach powtórzeń CCG wmodelowano niewiele cząsteczek rozpuszczalnika. Niemniej w modyfikowanym duplesie zlokalizowane są dwie cząsteczki jonu siarczanowego. Jeden z nich znajduje się w bruzdzie dużej dupletu C+D i oddziałuje z grupami *exo*-aminowymi reszt cytydynowych w pozycji 2 i 3 łańcucha C oraz jedną cząsteczką wody (Rysunek 54). Drugi anion został zlokalizowany w bruzdzie małej reszty G^L łańcucha D, która ma odkształcony łańcuch fosforanowy (*vide*



Rysunek 53. Reszta $4G^L$ z nici D, której kąty torsyjne α , β i γ przyjmują nietypowe wartości. Atomy fosforu, tlenu $O5'$, węgla $C5'$ i $C4'$ nie znajdują się w jednej linii (A). Atom fosforu jest przesunięty względem pierścienia cukrowego w kierunku sąsiadującej reszty po stronie $5'$. Nastąpił również obrót wokół wiązania $O5'-C5'$. Taka konformacja powoduje, że pierścień modyfikowanej reszty jest przesunięty (strzałka) w kierunku bruzdy dużej (B).

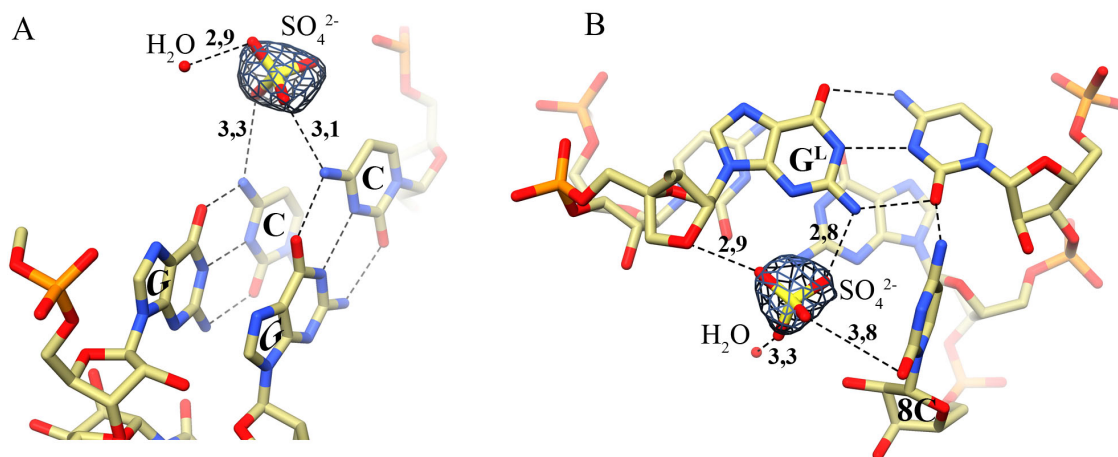
Tabela 22. Zestawienie wartości parametru roll dupleksów zawierających powtórzenia CCG. Odchylenie standardowe oznaczono jako s.d.

Skok	Roll ($^{\circ}$)		
	A+B	C+D	E+E'
CC/GG	10,3	2,5	5,3
$CG^{(L)}/CG$	8	16,2	14,7
$G^{(L)C}/CC$	8,9	6,2	1,3
$CC/G^{(L)C}$	10,4	6,1	1,3
$CG/CG^{(L)}$	3,3	10,4	15,6
GG/CC	6,4	5,9	4,7
średnia	7,9	7,9	7,1
s.d.	2,7	4,8	6,4

podpunkt 3.4.9: Modyfikowane reszty 4G z serii LNA). Anion siarczanowy tworzy dwa wiązania wodorowe z modyfikowanym nukleotydem. Jedno z nich utworzone jest z grupą *exo*-aminową reszty guanozyny, natomiast drugie z atomem $O2'$, który wchodzi w skład mostka metylenowego (Rysunek 54). Jon SO_4^{2-} oddziałuje również z atomem $O4'$ sąsiedniej reszty 5C oraz z cząsteczką wody. Niewykluczone, że słabe wiązanie wodorowe ($3,8 \text{ \AA}$) istnieje również pomiędzy grupą $O2'$ „wywiniętej” reszty 8C z nici C a jonem siarczanowym.

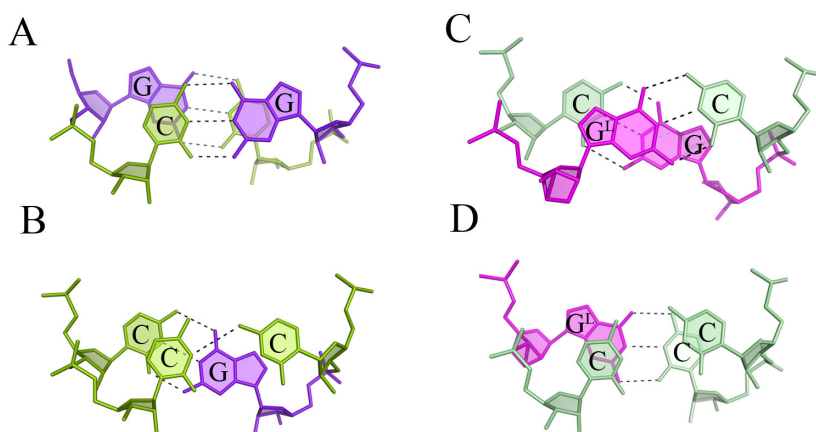
3.4.11. Oddziaływania warstwowe

Pomiędzy modyfikowanym a niemodyfikowanym dupleksem istnieje różnica w rozkładzie oddziaływań warstwowych (Rysunek 55). W przypadku niemodyfikowanej struktury oddziaływania te są typowe dla A-RNA. W parach typu Watsona-Cricka pierścienie zasad znajdują się w znacznym



Rysunek 54. Oddziaływania jonów siarczanowych z dupleksem C+D struktury modyfikowanej. Jeden z anionów oddziałuje z dwiema resztami cytydynowymi łańcucha D (A), natomiast drugi z modyfikowaną resztą G^L z łańcucha C oraz prawdopodobnie z resztą 8C z łańcucha D (B).

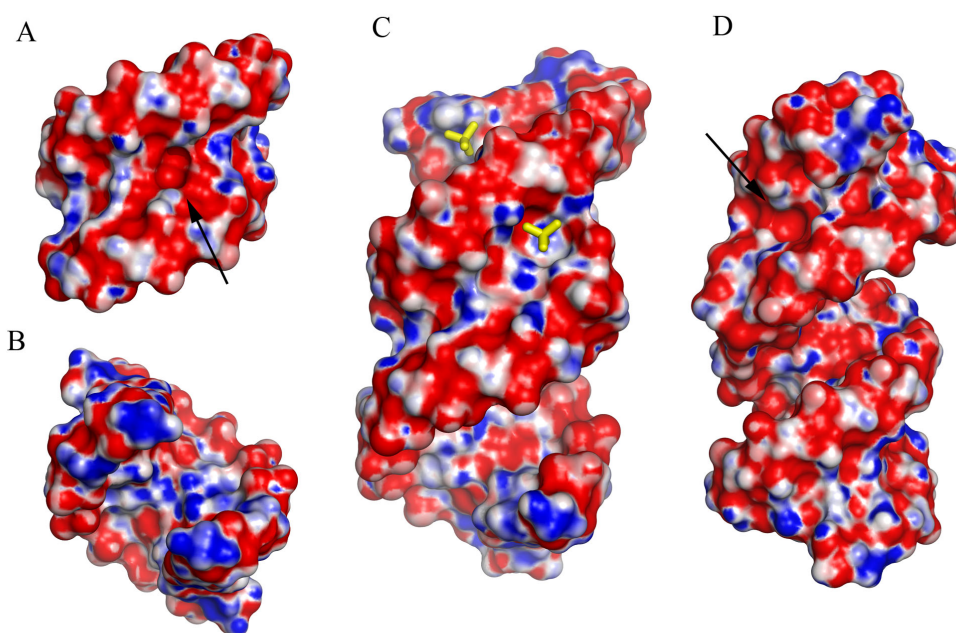
stopniu nad sobą (powierzchnia pokrycia $> 5,3 \text{ \AA}^2$), natomiast pary C-C tworzą oddziaływania warstwowe z sąsiednimi parami w ograniczonym stopniu (maksymalna powierzchnia pokrycia wynosi $0,7 \text{ \AA}^2$). W modyfikowanych dupleksach rozkład asocjujących ze sobą powierzchni zasad jest bardziej równomiernie rozłożony. Największą powierzchnię pokrycia mają reszty G^L , które po stronie 3' oddziałują z cytozyną a po stronie 5' z resztą guanozyny znajdującą się na przeciwnej nici (wyjątkiem jest reszta G^L z nici C, która oddziałuje z sąsiednią resztą cytydyny). Pary C-C, w przeciwieństwie do natywnej struktury, tworzą rozległe oddziaływania z sąsiednimi resztami z tego samego łańcucha (maksymalna powierzchnia pokrycia wynosi $7,7 \text{ \AA}^2$).



Rysunek 55. Przykłady oddziaływań warstwowych w strukturze niemodyfikowanej (A, B) oraz modyfikowanej (C, D). Reszty cytydynowe są zielone, reszty guanozynowe fioletowe a G^L różowe.

3.4.12. Kształt powierzchni i rozkład potencjału elektrostatycznego

Powierzchnie potencjałów elektrostatycznych obu struktur są bardzo podobne do siebie (Rysunek 56). W bruzdzie małej przeplatają się pionowe pasy o potencjale negatywnym i dodatnim równoległe do długiej osi helisy. Pasma potencjałów generowane są przez odpowiednie grupy funkcyjne par G-C i C-G. W środkowej części bruzdy brakuje jednego pasma o potencjale dodatnim, ponieważ znajdują się tam pary C-C, które w tym kierunku eksponują grupy karbonyłowe o potencjale ujemnym. Ponadto, w tym miejscu znajduje się charakterystyczne wgłębienie, które otacza krawędzie Watsona-Cricka obu reszt cytydyn. W dupleksie C+D znajdują się wywinęte reszty 8C, które zaburzają obraz powierzchni potencjału bruzdy małej. Niemniej, wgłębienie i ujemny potencjał wokół par C-C jest zachowany oraz do pewnego stopnia zostały zachowane pionowe naprzemienne pasy potencjałów. Do tego dupleksu wiąże się jon siarczanowy, który znajduje się w zagłębieniu o potencjale dodatnim utworzonym przez resztę 8C z nici C oraz grupę *exo*-aminową 4G^L z łańcucha D. Rozkład potencjału w bruzdzie dużej jest bardziej różnorodny. W strukturze niemodyfikowanej w środkowej części znajduje się duże pasmo o potencjale dodatnim, generowane przez sąsiadujące ze sobą na obu niciach reszty cytydynowe. W modyfikowanych dupleksach obszar ten jest mniej elektrododatni i w mniejszym stopniu zarysowany.



Rysunek 56. Powierzchnia potencjału elektrostatycznego niemodyfikowanego (A, B) oraz modyfikowanego dupleksu (C, D). W bruzdzie małej znajduje się zagłębienie (strzałka), które utworzone jest przez pary C-C (A, D). W środkowej części bruzdy dużej struktury romboedrycznej znajduje się obszar o dodatnim potencjale elektrostatycznym generowany przez grupy *exo*-aminowe reszt cytydynowych (B). W strukturze modyfikowanej związały się dwa jony siarczanowe (żółty) (C).

3.5. Opracowanie metody stabilizacji struktur spinkowych RNA w warunkach krystalizacji

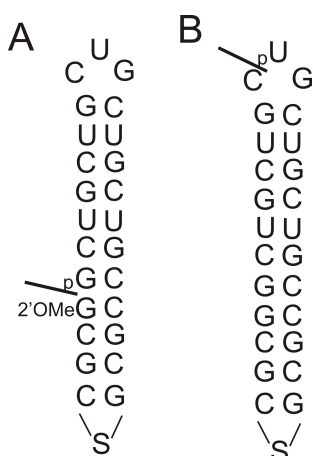
3.5.1. Zaprojektowanie sekwencji oligomerów RNA tworzących strukturę spinki

Do opracowania metody stabilizacji formy spinki tworzonej przez cząsteczki RNA używałam oligomery zawierające powtórzenia CUG. W pierwszym etapie pracy zaprojektowałam dwa typy oligomerów różniące się miejscem ligacji. Sekwencje tych oligomerów są następujące:

5' pGCU GCU GCU GCU GCU GCC GCG SCG CG2'OMe 3' – skrócona nazwa DNAL

5'pUGC UGC UGC CGC GSC GCG GCU GCU GC 3' – skrócona nazwa RNAL, gdzie S oznacza łącznik.

Oligomer DNAL posiada miejsce ligacji w obszarze dwuniciowym (Rysunek 57 panel A), co wymagane jest przy stosowaniu ligazy DNA z faga T4. Dodatkowo aby poprawić wydajność reakcji cyrkularyzacji, ostatnia reszta guanozyny posiada grupę metylową w pozycji 2' zamiast grupy hydroksylowej (132). Obecność tej grupy jest również istotna w reakcji ligacji chemicznej, ponieważ blokuje możliwość powstania nieprawidłowego wiązania fosfodiesterowego 5'-2' faworyzując tym samym typowe 5'-3'. Drugi oligomer posłużył do testowania reakcji ligacji przy użyciu T4 RNA ligazy. Miejsce ligacji znajduje się w obrębie pętli apikalnej (Rysunek 57 panel B). Obydwa oligomery zawierają na końcu 5' grupę monofosforanową wymaganą do reakcji ligacji.



Rysunek 57. Struktura drugorzędowa oligomeru DNAL (A) i RNAL (B) z zaznaczonymi miejscami ligacji.

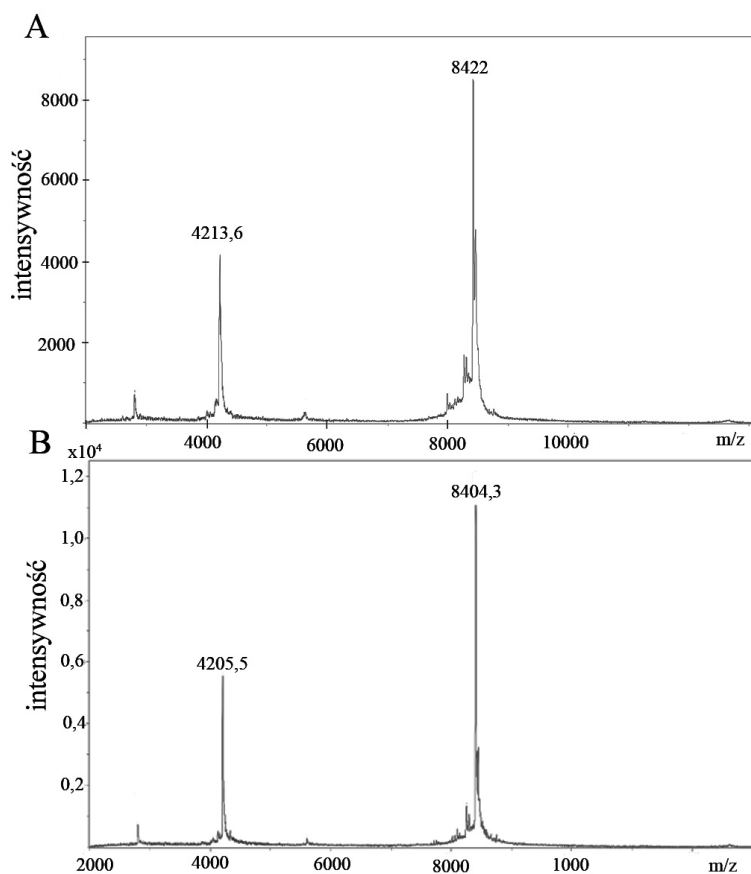
W sekwencję RNA włączane były dwa różne łączniki: glikol heksaetyleny (EG6, spacer 18) oraz pochodna eterowa stilbenu. EG6 jest komercyjnie dostępny, natomiast pochodna stilbenu została zamówiona na Wydziale Chemii Uniwersytetu im. Adama Mickiewicza. Niestety okazało się, że wydajność syntezy oligomeru z łącznikiem stilbenowym była bardzo niska. Następowo

hamowanie reakcji wydłużania po przyłączeniu łącznika do łańcucha oligomeru. Prawdopodobnie przyczyną była niestabilność układu stilbenowego. Na tym etapie zakończyłam próby otrzymania oligomeru z tym łącznikiem.

Po syntezie chemicznej, oligomery były odblokowywane i oczyszczane według standardowej procedury, która składała się z następujących etapów:

- odcięcie od podłoża i zdjęcie zasadolabilnych grup ochronnych,
- zdjęcie grupy silylowej,
- odsalanie przy użyciu kolumniek NAP-25,
- oczyszczanie z wykorzystaniem elektroforezy w 12% żelu poliakryloamidowym w warunkach denaturujących,
- elucja i odsalanie.

Masa oligomerów potwierdzana była za pomocą spektrometrii mas (Rysunek 58).

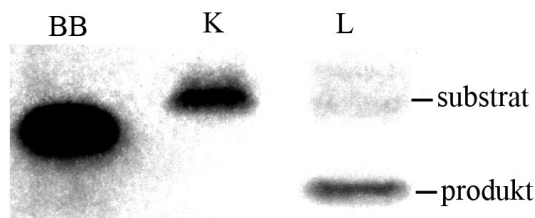


Rysunek 58. Widma masowe oligomerów DNAL (A) o masie 8412,8 daltonów i RNAL o masie 8396,8 daltonów (B).

3.5.2. Testowanie i optymalizacja reakcji cyrkularyzacji oligomeru

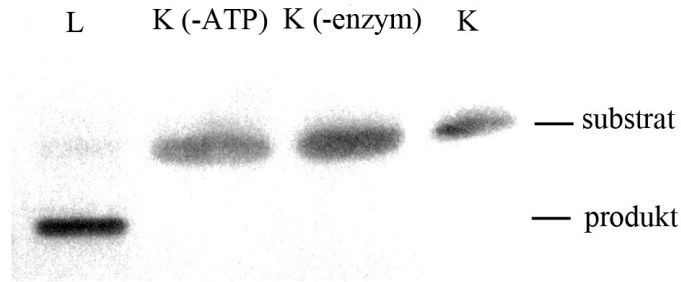
Oligomery DNAL i RNAL zostały wykorzystane do przetestowania reakcji cyrkularyzacji trzema sposobami: chemicznie oraz enzymatycznie z wykorzystaniem ligazy DNA lub RNA z faga

T4. Ligacja chemiczna prowadzona była w obecności jonów niklu, imidazolu oraz bromocyjanu przez noc w temperaturze 4°C (3). Cyrkularyzację enzymatyczną przeprowadzałam według protokołu zawartego w książce Handbook of RNA Biochemistry (133). Najwydajniejszą reakcją (powyżej 80%) okazała się ligacja przy użyciu ligazy RNA (Rysunek 59).

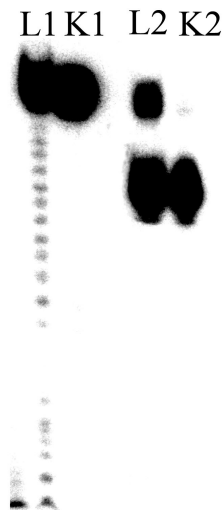


Rysunek 59. Obraz rozdziału elektroforetycznego reakcji ligacji. Zaznaczono wysokości odpowiadające substratowi. BB—barwnik błękit bromofenolowy, który migruje w 8% denaturującym żelu poliakryloamidowym jak oligomer o długości 19nt; K—kontrola reakcji, L—reakcja ligacji.

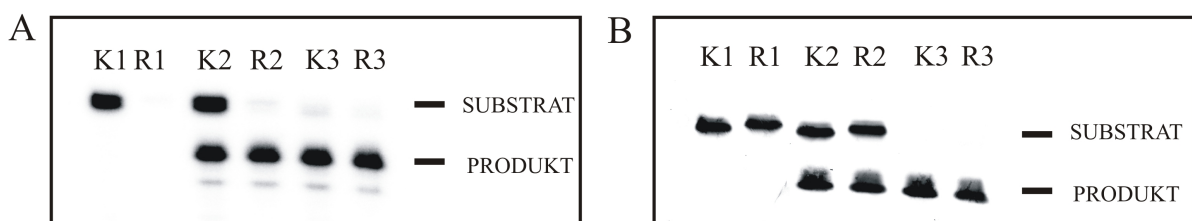
W celu potwierdzenia otrzymania zcyrkularyzowanego produktu przeprowadziłam szereg eksperymentów. Pierwszy z nich polegał na przeprowadzeniu reakcji kontrolnych, w których brakowało jednego z kluczowych komponentów: ligazy RNA lub ATP (Rysunek 60). W reakcjach, w których brakowało enzymu lub ATP nie powstał zcyrkularyzowany produkt. W kolejnym eksperymencie wykorzystałam ograniczoną hydrolizę jonami Mg^{2+} w obecności formamidu, która powoduje przerwanie wiązania fosfodiesterowego w dowolnym miejscu cząsteczki. Warunki reakcji muszą być tak dobrane, aby hydroliza zaszła tylko w jednym miejscu oligomeru. W tym wypadku skutkiem reakcji było uzyskanie substratu z produktu na podstawie porównania migracji w żelu poliakryloamidowym, co potwierdzało otrzymanie zligowanego produktu (Rysunek 61). Przeprowadziłam również eksperyment z wykorzystaniem alkalicznej fosfatazy, która usuwa grupę monofosforanową znajdującą się na końcu 5' cząsteczek RNA. Zarówno substrat jak i produkt wyznakowane były radioaktywnym izotopem ^{32}P , co pozwalało na monitorowanie działania enzymu. Grupa fosforanowa mogła być usunięta jedynie w przypadku substratu, czego efektem byłby zanik sygnału radioaktywnego. Po rozdziale reakcji w żelu poliakryloamidowym, otrzymałam autoradiogram, w którym na wysokości substratu nastąpił zanik sygnału radioaktywnego. Dodatkowo obecność oligomerów, szczególnie tych zdefosforylowanych, monitorowałam poprzez barwienie żelu za pomocą odczynnika Sybr Green II (Rysunek 62). Potwierdziłam również masę zcyrkularyzowanego RNA za pomocą spektrometrii mas MALDI-TOF MS (Rysunek 63). Otrzymany rezultat przy stosunkowo dużej masie oligomeru zawierał duży błąd pomiarowy. Na jego podstawie można jednak było stwierdzić, że oligomer posiadał spodziewaną masę około 8000 daltonów.



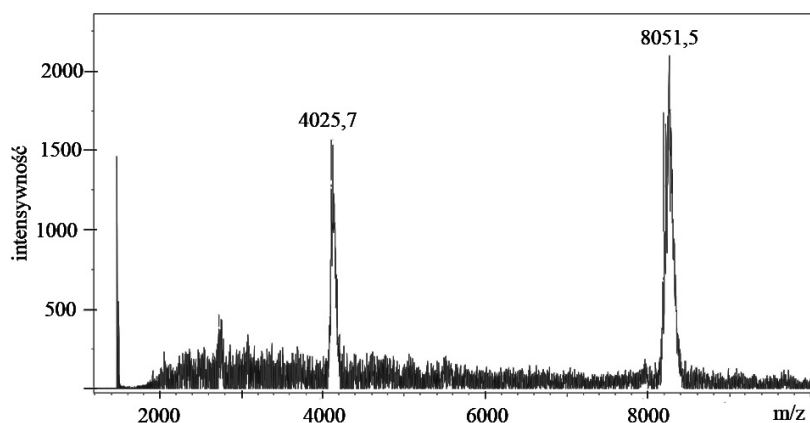
Rysunek 60. Obraz rozdziatu elektroforetycznego szeregu reakcji kontrolnych: L—ścieżka reakcji ligacji; K (-ATP)—reakcja prowadzona bez ATP; K (-enzym)—reakcja prowadzona bez enzymu; K—substrat przed reakcją ligacji.



Rysunek 61. Ograniczona hydroliza jonami Mg^{2+} w formamidzie: substratu (L1) oraz zcyrkularyzowanego oligomeru (L2). Literą K zaznaczono odpowiednio linie kontrolne.



Rysunek 62. Defosforylacja substratu i produktu. Na panelu A przedstawiono autoradiogram, na którym w ścieżkach reakcji defosforylacji z substratem następuje zanik sygnału radioaktywnego. Panel B przedstawia obraz otrzymany po barwieniu tego samego żelu za pomocą odczynnika Sybr Green II. Ścieżki: K1—reakcja kontrolna z substratem bez enzymu; R1—reakcja z substratem; K2—reakcja kontrolna mieszaniny substrat i produkt bez enzymu; R2—reakcja mieszaniny substratu i produktu; K3—reakcja kontrolna produktu bez enzymu; R3—reakcja z produktem.



Rysunek 63. Widmo masowe zcyrkularyzowanego oligomeru.

3.5.3. Opracowanie metody oczyszczania zcyrkularyzowanego oligomeru.

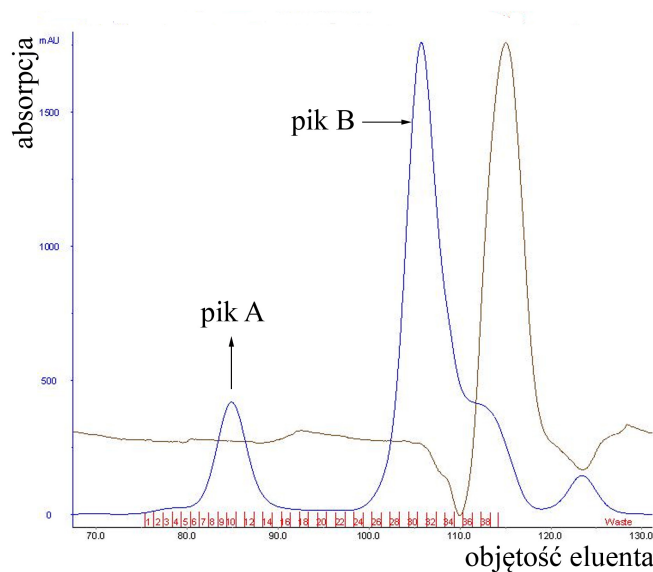
Oczyszczanie RNA jest etapem, w którym obrana metoda powinna być jednocześnie najbardziej wydajna (nie powodować znacznych strat materiału) oraz dawać możliwie najczystszy produkt. Do oczyszczania kwasów nukleinowych najczęściej stosuje się elektroforezę w żelu poliakrylamidowym. Jest to metoda stosunkowo prosta do przeprowadzenia i niedroga. Jej wadą jest niska wydajność elucji cząsteczek z żelu. Alternatywnie, szczególnie w przypadku oczyszczania preparatywnego, można wykorzystać techniki chromatograficzne, które są bardziej wydajne, ale dobór odpowiedniej metody, złożeń czy warunków rozdzielania bywa czasochłonny i trudny do osiągnięcia.

W celu znalezienia najlepszego sposobu oczyszczania zcyrkularyzowanego oligomeru wypróbowałam szereg metod chromatograficznych. Stosowałam następujące metody rozdzielania:

- układ faz odwróconych,
- sączenie molekularne – złożeń: superdex 75, superdex 200,
- chromatografia jonowymienna - złożeń: DEAE celuloza, hydroksypatyt.

Wszystkie próby oczyszczania chromatograficznego zakończyły się niepowodzeniem. Substrat i produkt miały ten sam czas retencji oraz podobne powinowactwo do złożeń. Oznacza to, że wpływały z kolumny w postaci pojedynczego piku (Rysunek 64). Prawdopodobnie oba oligomery posiadały podobny promień hydrodynamiczny oraz taką samą ilość ładunków.

W związku z powyższym podjęłam próby różnicowania substratu od produktu metodami biochemicznymi stosując np. defosforylację niezligowanego oligomeru (zmniejszenie ilości ujemnych grup fosforanowych) czy specyficzną degradację substratu za pomocą nukleazy RNA o aktywności 5'→3' egzonukleazy, która trawi tylko te RNA, które posiadają monofosforan na końcu 5' (Terminator Exonuclease). Oba podejścia również się nie sprawdziły, ponieważ w pierwszym przypadku



Rysunek 64. Przykładowy rozdział chromatograficzny mieszaniny ligacyjnej na kolumnie ze złożem *superdex 200*. Kolorem niebieskim zaznaczono wykres absorpcji, natomiast kolorem żółtym zaznaczono wykres konduktometrii podczas rozdziału. Pikowi A odpowiada mieszanina substrat i produkt. Pik B zawiera ATP i ADP.

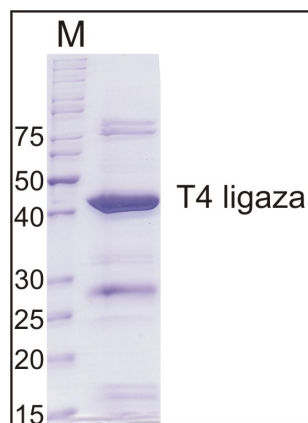
defosforylacja nie różnicowała substratu i produktu, a nukleaza działała niespecyficznie i trawiła zarówno substrat jak i produkt. Z powyższych przyczyn preparatywne oczyszczanie zcyrkularyzowanego RNA przeprowadzałam w żelu poliakryloamidowym w warunkach denaturujących. Jest to jedyna znana metoda, która pozwoliła na rozdzielenie substratu od produktu mieszaniny ligacyjnej.

3.5.4. Otrzymywanie ligazy RNA z faga T4

Do otrzymania, w skali preparatywnej, zcyrkularyzowanego oligomeru potrzebna jest duża ilość enzymu. Przykładowo na 40 nanomoli substratu zużywa się 3300 jednostek enzymu czyli trzy i jedną trzecią fiołki ligazy RNA z firmy Fermentas. Do wstępnej krystalizacji obejmującej 48 różnych warunków potrzeba przynajmniej trzy razy więcej produktu. W celu obniżenia kosztów, postanowiłam otrzymywać enzym poprzez nadekspresję w bakteriach *E. coli*. Wektor zawierający sekwencję genu ligazy RNA z faga T4 otrzymałam od prof. Petera J. Unrau z Uniwersytetu im. Simona Frasera w Kanadzie. Nadekspresję i oczyszczanie enzymu prowadziłam zgodnie z protokołem opracowanym przez Q. S. Wang i P. J. Unrau (134) (Rysunek 65).

Eksprymowane białko posiada znacznik histydynowy, co umożliwia jego oczyszczenie przy zastosowaniu jednego etapu chromatografii powinowactwa do jonów niklu. Przygotowane klarowne supernatanty lizatów komórkowych nanosiłam na kolumnę ze złożem agarozowym Ni-NTA. Enzym eluowałam ze złoża buforem ze wzrastającym stężeniem imidazolu. Frakcje zawierające ligazę RNA identyfikowałam za pomocą elektroforezy SDS-PAGE. Otrzymany enzym był zanieczyszczony

innymi białkami, ale w związku z tym, że był aktywny na poziomie podobnym do komercyjnie dostępnemu enzymu, nie podjęłam dalszych prób doczyszczania białka.



Rysunek 65. Oczyszczony enzym ligazy RNA z faga T4 (masa enzymu 43kDa). Ścieżkę z markerem wysokości (kDa) zaznaczono literą M.

3.5.5. Procedura otrzymywania stabilnej formy spinki RNA

Opracowana przeze mnie procedura otrzymywania stabilnej formy spinki RNA w nanomolarnych ilościach odbywała się według następującego protokołu:

- przeprowadzenie reakcji cyrkularyzacji oczyszczonego substratu za pomocą RNA ligazy z faga T4,
- zagęszczanie reakcji,
- usuwanie enzymu poprzez ekstrakcję fenol/chloroform,
- odsalanie,
- zagęszczanie,
- elektroforeza w żelu poliakryloamidowym,
- elucja,
- zagęszczanie,
- odsalanie i zagęszczanie.

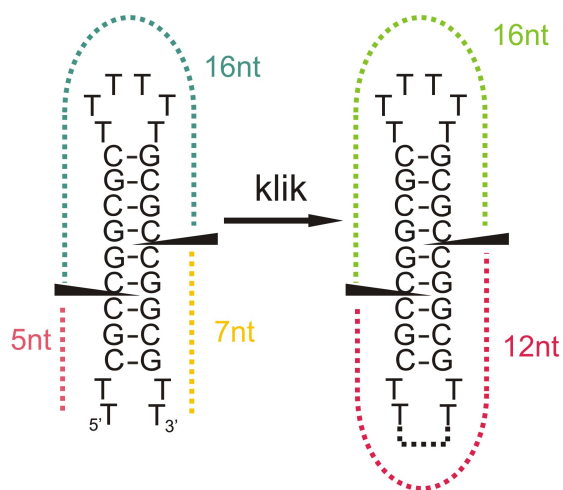
3.5.6. Otrzymanie zcyrkularyzowanego oligomeru metodą klik

Testy reakcji klik przeprowadzałam na oligomerze DNA, ponieważ chemiczna synteza DNA jest tańsza i ma lepszą wydajność. Sekwencje oligomeru zaprojektowałam tak, aby w trzonie spinki składającym się głównie z par G-C znajdowało się miejsce restrykcyjne dla enzymu *HpaII*, które wykorzystywałam do potwierdzenia otrzymania zcyrkularyzowanego oligomeru (Rysunek 66). Pętla apikalna oraz dwa ostatnie nukleotydy na końcu 5' i 3' składały się z reszt tymidynowych. Dodatkowo na końcu 5' znajdowała się zmodyfikowana grupa fosforanowa zawierająca atom bromu, a na końcu 3' fosforan zawierał grupę acylową. Oligomer miał sekwencję: 5' TTC GCC

GGC GCT TTT TTG CGC CGG CGT T 3'.

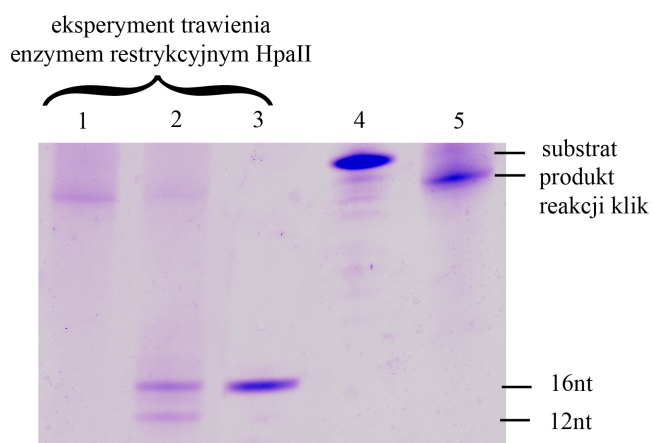
Po syntezie chemicznej, przeprowadzałam reakcję podstawienia atomu bromu grupą azydkową. Następnie oligomer odcinałam od złoża i oczyszczałam go w żelu poliakryloamidowym w warunkach denaturujących.

Reakcję klik przeprowadzałam wg protokołu opracowanego przez Paredes i Dass (135). Standardowo reakcję przeprowadza się z kationami miedzi uprzednio zredukowanymi askorbinianem sodu do pierwszego stopnia utlenienia. W celu stabilizacji zredukowanej miedzi do mieszaniny reakcyjnej zaleca się dodawanie np. acetonitrylu lub PMDETA (N,N,N',N',N''-pentametyldietylenotriamina). Przetestowałam oba związki, ale znacznych różnic w wydajności reakcji nie dostrzegłam.



Rysunek 66. Schemat reakcji klik, w wyniku której końce 5' i 3' substratu zostają połączone wiązaniem kowalencyjnym. Czarne trójkąty wskazują miejsca restrykcyjne dla enzymu *HpaII*. Po trawieniu substratu enzymem otrzymuje się trzy produkty o długości 5 nt (różowy), 7 nt (żółty) i 16 nt (niebieski). Natomiast po trawieniu enzymem produktu otrzymuje się dwa oligomery o długości 12 nt (różowy) i 16 nt (zielony).

W celu potwierdzenia otrzymania zcyrkularyzowanego oligomeru przeprowadzałam cięcie restrykcyjne za pomocą enzymu *HpaII*. Spodziewanym efektem eksperymentu było otrzymanie w przypadku zamkniętego oligomeru – dwóch produktów o długości 16 i 12 nukleotydów, natomiast w przypadku substratu trzy: 5, 7 oraz 16 nukleotydów (Rysunek 67). Po rozdziale mieszaniny reakcyjnej w żelu poliakryloamidowym w ścieżce substratu otrzymałam tylko jeden produkt prawdopodobnie dlatego, że barwnik *stains-all* nie wybarwia tak krótkich oligomerów. Niemniej eksperyment potwierdził otrzymanie zcyrkularyzowanego produktu metodą klik.



Rysunek 67. Rozdział elektroforetyczny produktów trawienia restrykcyjnego oraz reakcji klik. Ścieżki: 1 – kontrola zcyrkularyzowany oligomer; 2 – trawienie zcyrkularyzowanego oligomeru; 3 – trawienie substratu; 4 – substrat, 5 – zcyrkularyzowany oligomer.

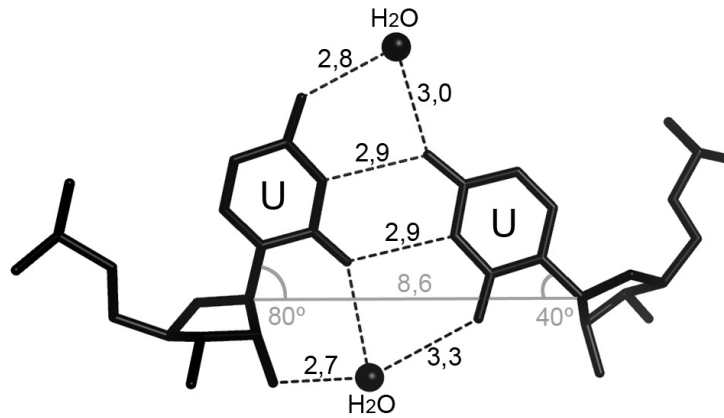
4. Dyskusja

Celem mojej pracy doktorskiej było poznanie struktur krystalicznych wszystkich czterech powtórzeń CNG. Na podstawie otrzymanych wyników zidentyfikowałam wspólne oraz charakterystyczne właściwości fizyko-chemiczne powtórzeń CNG takie jak: powinowactwo do ligandów, charakterystyka powierzchni, dystrybucja potencjału elektrostatycznego czy sieć wiązań wodorowych i hydratacja. Stworzony przeze mnie profil może być użyty jako matryca 3D dla racjonalizacji obserwacji biochemicznych i cytologicznych. Ponadto w literaturze sygnalizowana jest potrzeba szczegółowego poznania struktur RNA w celu racjonalnego projektowania ligandów (ang. *drug design*) wiążących się do powtórzeń CNG (136–138).

4.1. Niestandardowe pary N-N

Najbardziej wyróżniającą się cechą struktur zawierających powtórzenia CNG jest obecność niestandardowych par N-N w otoczeniu silnych par G-C. W przypadku tego typu RNA, niestandardowe pary nadają specyficzne właściwości danemu powtórzeniu. W dupleksach powtórzeń CAG, reszty adenzynowe parują się w sposób unikalny, oddziałując ze sobą jednym wiązaniem wodorowym C-H...N. Tego rodzaju wiązanie jest mało poznane w biomolekułach, ponieważ atom węgla jest słabszym donorem protonu niż atomy tlenu oraz azotu i dlatego to one dominują one w tworzeniu oddziaływań wewnątrz- i międzycząsteczkowych w biomolekułach.

Pary U-U oddziałują ze sobą również jednym wiązaniem wodorowym utworzonym pomiędzy atomami N3 jednej z reszt U oraz O4 nachylonej w kierunku bruzdy małej drugiej reszty urydyny. Takie pary opisane jako $W+C+/W+C+$ *cis* (z jednym wiązaniem wodorowym – 1 H b.b) obecne są w strukturach krystalicznych RNA. Znakomita większość (ponad 70%) z wyszukanych 261 par U-U tego typu za pomocą wyszukiwarki FRABASE (139) wchodzi w skład struktur rybosomalnych. Pozostałe to struktury tRNA glutaminowego w kompleksie z syntetazą aminoacylo-tRNA oraz inne RNA. Dominującą konformacją par U-U w cząsteczkach RNA jest jednak W/W *cis* (typ XVI wg klasyfikacji Seangera) (140) (3143 trafień FRABASE), w której pomiędzy resztami urydynowymi utworzone są dwa wiązania wodorowe O2...N3 i N3...O4 (Rysunek 68) (141). Takie parowanie wymaga zbliżenia reszt cukrowych do siebie (odległość C1'-C1' wynosi około 8,6 Å), niewielkiego nachylenia jednej z reszt w kierunku bruzdy małej ($\lambda = 40^\circ$) oraz wychylenia drugiej reszty w kierunku bruzdy dużej ($\lambda = 80^\circ$).



Rysunek 68. Typowa para U-U wygenerowana za pomocą serwera SwS (141). Pomiędzy parą U-U utworzone są dwa wiązania wodorowe. Para oddziałuje z dwiema cząsteczkami wody. Jedna z nich zlokalizowana jest w bruzdzie dużej a druga w bruzdzia małej.

W powtórzeniach CGG tworzą się pary G-G, w których jedna z reszt guanozynowych przyjmuje konformację *syn* natomiast druga pozostaje w konformacji *anti*. Jest to jedyna niekanoniczna para typu N-N, która ma dwa wiązania wodorowe. Dodatkowo istnieje jeszcze jedno wewnątrz nukleotydowe wiązanie pomiędzy grupą *exo*-aminową G(*syn*) a atomem tlenu jej grupy fosforanowej. Tego typu para jest dobrze poznana w innych strukturach RNA zarówno krystalicznych oraz określonych za pomocą techniki NMR i wydaje się być faworyzowana w regionach helikalnych RNA. Jej cechą charakterystyczną jest możliwość przechodzenia z orientacji *syn-anti* w *anti-syn*. Takie nieuporządkowanie obserwowane jest w dwóch dupleksach struktury [GC(8-BrG)GCGGCGGC]₂, ale również w warunkach prowadzonych podczas pomiaru NMR (97).

Para C-C w powtórzeniach CCG różni się właściwościami od pozostałych trzech niekanonicznych par. Podstawową różnicą jest brak określonej konformacji. Każda z trzech obserwowanych par C-C paruje się w inny sposób. W jednej parujące się reszty cytydynowe nie tworzą żadnych wiązań wodorowych. W kolejnej występuje jedno słabe wiązanie. Trzecia przyjmuje konformację podobną do par U-U, gdzie jedna z reszt cytydynowych nachylona jest w kierunku bruzdy małej ($\lambda = 31^\circ$) a wiązanie wodorowe utworzone jest pomiędzy atomami N3 a O4. Taka labilność konformacyjna jest charakterystyczna dla par tego typu. W dużej podjednostce rybosomalnej eubakterii *Deinococcus radiodurans*, w bardzo podobnym kontekście sekwencyjnym i strukturalnym, znajduje się para 1219C-1253C. Na podstawie wyszukanych za pomocą FRABASE 30 krystalicznych struktur tej cząsteczki można obserwować całe spektrum konformacji, w których kąt λ nachylonej reszty cytydyny przyjmuje wartości od 0–45°.

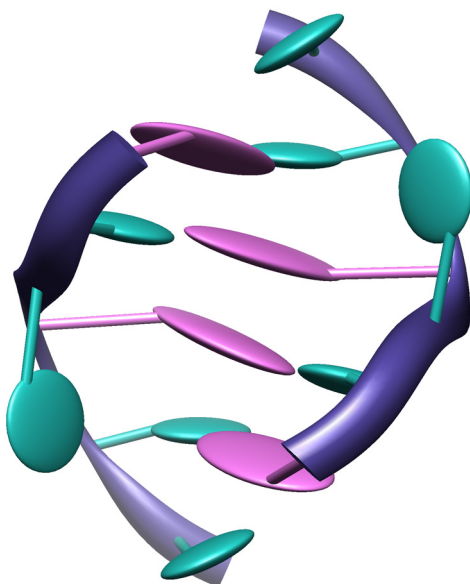
4.2. Akomodacja par N-N w strukturze A-RNA

Sposób parowania się niekanonicznych par jest kompromisem pomiędzy utrzymaniem formy A-RNA narzuconej przez pary G-C i C-G a możliwością utworzenia oddziaływań wodorowych i uzyskaniem optymalnej energii. Każda para N-N wykazuje swój specyficzny sposób akomodacji w uformowanej strukturze. Utworzona „sztywność” strukturalna ujawnia się jako dążenie do zachowania określonej odległości pomiędzy sparowanymi łańcuchami. Miarą tego zjawiska jest dystans (około 10,5–10,8 Å) pomiędzy atomami C1' sparowanych reszt nukleotydowych. Wymiernym efektem tego dążenia jest to, że niekanoniczne pary, tak jak kanoniczne, tworzą region dwuniciowy. Nie tworzą one ani wybrzuszeń ani nie powodują znacznych deformacji łańcucha fosforanowo-cukrowego.

Zasady purynowe i pirymidowe reszt nukleotydowych różnią się między sobą „objętością”. W parach kanonicznych „duża” zasada paruje się z „małą”. Natomiast w strukturach powtórzeń CNG niekanoniczne pary utworzone są przez dwie zasady o tych samych rozmiarach. W przypadku parowania się dwóch „dużych” zasad purynowych problemem jest relatywnie mała dostępna przestrzeń. Na przykładzie par G-G oraz A-A obserwuje się dwa różne sposoby akomodacji do zaistniałych warunków. W parze G-G jedna z reszt „usuwa się” poprzez przyjęcie konformacji *syn*, w wyniku czego „chowa się” nad pierścień rybozy, co tworzy przestrzeń dla drugiej reszty G. Oprócz aranżacji *syn-anti* następuje rotacja wokół wiązania O5'-C5' reszty G(*syn*). Wydaje się, że konformacja ta jest niezbędna do i) optymalizacji wiązań wodorowych pomiędzy krawędziami Watsona-Cricka a Hoogsteena poprzez przesunięcie G(*syn*) w kierunku bruzdy małej ii) uniknięcia zderzenia sterycznego grup karbonylowych O6 obu reszt iii) utworzenia oddziaływania wewnątrznukleotydowego pomiędzy atomem N2 a atomem tlenu grupy fosforanowej reszty G(*syn*) iv) utworzenia oddziaływania warstwowego pomiędzy G(*syn*) a sąsiadującą resztą cytydyny. Z kolei w przypadku par A-A oba nukleotydy przyjmują konformację *anti*. Dwie duże zasady znajdujące się *vis-a-vis* wypychają siebie w kierunku bruzdy dużej (miarą tego efektu jest wysoka wartości kąta λ obydwu reszt). Wypchnięcie jednej z reszt jest tak duże ($\lambda = 87^\circ$), że wystaje ona ponad pozostałe reszty i wygląda tak, jakby pokazywała znak „kciuk do góry” (ang. *thumbs up sign*). Takie nierównomierne wypchnięcie umożliwia utworzenie wiązania wodorowego pomiędzy nimi.

Uracyl jest mniejszy od zasady purynowej. W parze U-U odległość pomiędzy atomami C1' wynosi 10,4 Å. Oznacza to, że jeżeli ustawilibyśmy dwie reszty w takiej odległości naprzeciw siebie to będzie ona zbyt duża, aby utworzyć wiązanie wodorowe. W strukturze powtórzeń CUG reszty U-U oddziałują ze sobą, ponieważ jedna z reszt urydynowych nachyla się w kierunku bruzdy małej, co zmniejsza dystans pomiędzy nimi. Wydaje się, że taka aranżacja jest dosyć łatwa do utworzenia i nie powoduje naprężeń w strukturze RNA.

W przypadku par C-C, akomodacja polega na ogólnym unikaniu parowania się ze sobą reszt cytydynowych. We wszystkich dupleksach powtórzeń CCG, z dwóch możliwych do utworzenia par C-C obecna jest tylko jedna. Sugeruje to, że bilans energetyczny jest korzystniejszy, jeżeli w strukturze znajdują się niesparowane i nieuporządkowane reszty C w pozycji 8 aniżeli druga para C-C. Podobnie, w strukturze NMR dupleksu DNA (CCGCCG)₂ (kod PDB: 1noq) (142) obserwuje się tendencję do minimalizacji ilości par C-C. W tym przypadku nastąpił również „poślizg” nici i utworzenie niesparowanych końców 5' i 3' (Rysunek 69). Dodatkowo reszta 4C w jednym z łańcuchów została wypętlona i w efekcie nie obserwuje się ani jednej pary C-C. Natomiast w obu strukturach krystalicznych powtórzeń CCG sumarycznie są trzy pary C-C, które niezależnie od sposobu parowania się wraz z parami G-C stanowią region dwuniciowy rozumiany jako zachowanie dystansu 10,8 Å pomiędzy atomami C1'. Prawdopodobnie problem akomodacji nie dotyczy tych par i jedynym ograniczeniem ich labilności konformacyjnej jest stopień nachylenia się reszty cytydyny w kierunku bruzdy małej (minimalny kąt $\lambda = 31^\circ$).

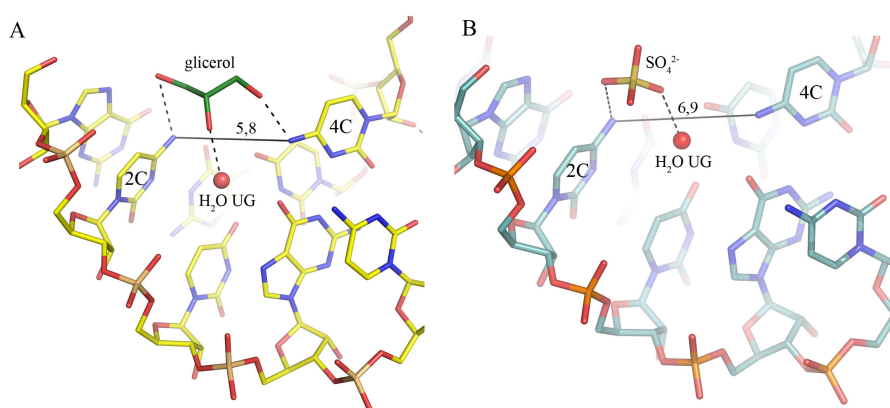


Rysunek 69. Struktura DNA dupleksu (CCGCCG)₂. „Poślizgnięta” helisa zbudowana jest z par G-C i C-G oraz z wypętlonych reszt C (zielony).

4.3. Oddziaływania par N-N z cząsteczkami rozpuszczalnika

Sposób parowania się par N-N nie wysyca ich potencjału do tworzenia oddziaływań, dlatego mają one zdolność do przyciągania ligandów. W parze U-U obie reszty eksponują w kierunku bruzdy małej swoje grupy karbonyłowe O2. Dodatkowo dostępna dla rozpuszczalnika jest grupa amidowa N3 nachylonej reszty U. Utworzoną konformację stabilizuje „woda U-U”, która oddziałuje z obiema resztami jednocześnie. Służy ona jako donor i akceptor wiązania wodorowego. W kierunku bruzdy dużej w „rozciągniętej” parze U-U najbardziej eksponowany jest atom O4 nienachylonej reszty i to z nim oddziałuje woda UG. Wydaje się, że drugi karbonyłowy atom tlenu jest

mniej dostępny i cząsteczka wody preferencyjnie tworzy wiązanie z atomem O4 sąsiadującej guanozyny. Ponadto w strukturze jednoskośnej powtórzeń CUG w bruzdzie dużej związały się glicerol i jon siarczanowy. Wspólną cechą tych cząsteczek jest obecność grup hydroksylowych, które mogą być zarówno akceptorem jak i donorem wiązań wodorowych. Oba ligandy oddziałują z wodą UG i wydaje się, że nie dążą do jej zastąpienia. Na tej podstawie cząsteczki wody UG i U-U można określić jako stabilny element struktury. Drugim wspólnym miejscem oddziaływania ligand-RNA jest grupa *exo*-aminowa reszty 2C. Pomimo wysycenia jej potencjału do tworzenia wiązań wodorowych, w wyniku parowania z resztą guanozyny, istotny jest brak oddziaływania warstwowego z sąsiadującą po jej 3' stronie nachyloną resztą urydyny. Wiązanie się glicerolu, którego grupy funkcyjne są bardziej „rozciągnięte” w przestrzeni niż jonu siarczanowego, koreluje z dużymi skokami wartości kąta skręcenia helisy A+B. Oprócz tego zmniejsza się o 1,1 Å odległość pomiędzy atomami N4 reszt 2C nici A i 4C nici B co sugeruje, iż glicerol wymaga większego dopasowania strukturalnego niż anion siarczanowy (Rysunek 70). W przypadku trzeciego dupleksu E+E' struktury jednoskośnej nie zaobserwowano analogicznego wiązania cząsteczek ligandu. Niewykluczone, że ligandy oddziałują z cząsteczką RNA, lecz gorsza jakość mapy gęstości elektronowej tego obszaru nie pozwala na jego lokalizację.



Rysunek 70. Odległość pomiędzy atomami N4 reszt 2C oraz 4C dupleksu A+B (A) oraz dupleksu C+D (B). Gdy ze strukturą oddziałuje cząsteczka glicerolu odległość między atomami N4 jest mniejsza o 1,1 Å niż gdy związany jest jon siarczanowy.

W parze G-G jedna z reszt guanozynowych przyjmuje konformację *syn*. Lokalizacja zasady nad pierścieniem rybozy powoduje ekspozycję krawędzi Watsona-Cricka w kierunku bruzdy dużej. Krawędź ta ma duży potencjał do tworzenia oddziaływań i w kanonicznych parach zaangażowana jest w tworzenie wiązań wodorowych z resztami cytydynowymi. Wiązanie wodorowe utworzone przez atomy O6 oraz N2 nie wysycza całkowicie potencjału reszty G(*syn*). Co więcej, pierścień zasady G(*syn*) nie tworzy oddziaływań warstwowych z nukleotydem znajdującym się po stronie 3', co najprawdopodobniej jeszcze bardziej wzmacnia efekt niewysycenia potencjału. W strukturach (GCGGCGGC)₂ i [GC(8-BrG)GCGGCGGC]₂ z krawędzią Watsona-Cricka oddziałuje jon siarczanowy, który wydaje się być idealnie dopasowany pod względem chemicznym oraz geometrycznym. Oznacza to, że rozstaw jego atomów tlenu, dobrych akceptorów protonów,

pasuje do rozstawu eksponowanych grup funkcyjnych reszty G(*syn*). W przypadku struktury [GC(8-BrG)GCGGC]₂ w medium krystalizacyjnym nie było jonów siarczanowych, lecz jony wapniowe, które również oddziałują z parą G-G. Kationy mogą być tylko akceptorem elektronów, dlatego wiążą się do jedynej dostępnej i bogatej w elektrony grupy karbonylowej O6 G(*anti*). Na podstawie porównania struktur krystalizujących w różnych warunkach można zaobserwować, że gdy brakuje dogodnego liganda, potencjał pary G-G zostaje wysycony przez cząsteczki wody.

Kolejnymi strukturami, w których niekanoniczna para oddziałuje z jonem siarczanu, są dupleksy zawierające powtórzenia CAG. Ta unikalna para A-A eksponuje swoje grupy funkcyjne znajdujące się na krawędzi Watsona-Cricka w kierunku bruzdy dużej. Pomiedzy te grupy wpasował się jon siarczanowy, którego atom tlenu O1 oddziałuje z obiema resztami będąc jednocześnie akceptorem i donorem wiązania wodorowego. Drugi atom tlenu O2 jest akceptorem wiązania utworzonego z grupą *exo*-aminową bardziej wysuniętej reszty adenozyiny. Obsadzenie jonów siarczanowych w obu strukturach krystalicznych jest połówkowe i prawdopodobnie świadczy o ograniczonej specyficzności do wiązania tego anionu. Z drugiej strony wydaje się, że jon siarczanowy wysyca potencjał reszt adeninowych do tworzenia oddziaływań tak samo jak cząsteczki wody, które wiążą się w sytuacji, gdy go nie ma. W bruzdzie małej do każdego atomu N3 reszt A wiąże się cząsteczka wody. Są one obserwowane dla każdej reszty adenozyiny zarówno w strukturze wysokorozdzielczej jak i niskorozdzielczej (forma trygonalna), gdzie gęstość elektronowa rozpuszczalnika jest obserwowana tylko dla bardzo uporządkowanych atomów. Taka hydratacja jest charakterystyczna również dla kanonicznych par, ale w tym przypadku oddziaływanie to wydaje się być specyficzne i silne, zważywszy na niewysycony potencjał reszt adenozyinowych do tworzenia oddziaływań.

4.4. Wpływ niekanonicznych par na globalną strukturę RNA

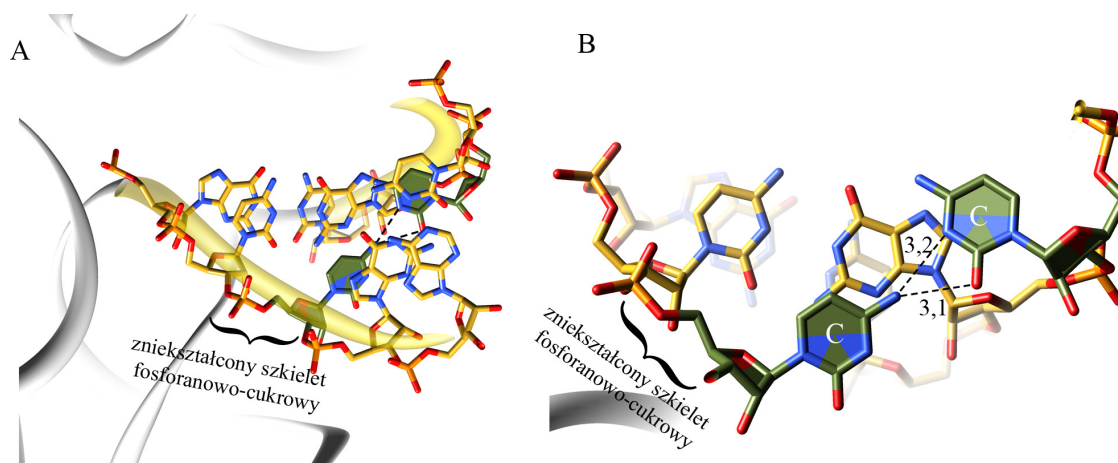
Pomimo liczebnej przewagi par G-C i C-G, które kształtują strukturę powtórzeń CNG, pary N-N mają również na nią wpływ. Pierwszym zauważalnym parametrem jest średnia wartość odległości C1'-C1' wyznaczona dla par C-G i G-C w strukturach powtórzeń CAG, CGG i CUG. W sytuacji, gdy w dupleksie znajdują się niestandardowe pary typu puryna-puryna (G-G lub A-A) helisa jest szersza (C1'-C1' = 10,7 Å) niż gdy są to pary urydyna-urydyna (C1'-C1' = 10,5 Å). Taka korelacja wydaje się być logiczna w kontekście wielkości par N-N. Pary A-A oraz G-G mają większe wymagania steryczne i wpływają na globalną strukturę RNA w bardziej widoczny sposób niż pary U-U. W obu przypadkach w dupleksach znajdują się reszty z rotacją wokół wiązania O5'-C5'. Jest to skorelowane ze zmniejszeniem wartości kąta skrętu helisy. Dla powtórzeń CGG jest to lokalny efekt, który kompensowany jest w innym miejscu w strukturze, gdzie następuje zwiększenie kąta skrętu i sumarycznie średnia wartość tego parametru wynosi 30–32°. W przypadku powtórzeń CAG jest to efekt globalny. Helisa ulega rozkręceniu oraz następuje

poszerzenie i otwarcie bruzdy dużej ($>20 \text{ \AA}$). Z drugiej strony parametry te wydają się być wrażliwe na rodzaj ligandu związanego w bruzdzie dużej. W strukturach powtórzeń CGG obecny jest albo kation Ca^{2+} , albo anion SO_4^{2-} . Szerokość bruzdy dużej dupleksu z jonem wapnia jest mniejsza prawie o 4 \AA od struktur z jonem siarczanowym (szerokość bruzdy wynosi około 18 \AA). Niewykluczone, że obecność anionu stabilizuje zwiększoną szerokość bruzdy, co jest również obserwowane w strukturach $(\text{GUGGUCUGAUGAGGCC})_2$ zawierających parę G-G oddziałującą z jonem siarczanowym (kod: 3czw i 3d0m) (143).

Pary C-C wpływają na strukturę w inny sposób niż pozostałe pary N-N. Przede wszystkim powodują, że tworzy się zupełnie inna struktura niż to było zaplanowane. Jeżeli jednak zaczniemy rozpatrywać wpływ par C-C w tak utworzonej helisie, to wydaje się, że pary te w niemodyfikowanym dupleksie nie powodują znacznych zmian strukturalnych, podobnie jak w przypadku par U-U. W tym dupleksie jedna z reszt jest lekko nachylona w kierunku bruzdy małej ($\lambda = 39^\circ$), natomiast druga reszta C zachowuje typową konformację. Helisa ma standardowe wartości parametrów helikalnych i kątów torsyjnych. W przypadku modyfikowanych dupleksów powtórzeń CCG następują liczne zaburzenia, które powodowane są różnymi konformacjami par C-C. W obu parach płaszczyzny reszt cytozynowych są znacznie skręcone względem siebie (parametr *propeller* wynosi $-19,3^\circ$ oraz $-21,5^\circ$), co rzutuje na sąsiednie pary. Miarą wywołanych zaburzeń są nietypowe wartości lokalnych parametrów wyznaczonych pomiędzy sparowanymi resztami, przede wszystkim *buckle*, oraz wyznaczonych pomiędzy sąsiadującymi parami (kąt skrętu helisy, odległość między sąsiednimi płaszczyznami zasad). Dodatkowo w dupleksie C+D jedna z reszt cytydynowych nachylona jest w kierunku bruzdy małej ($\lambda = 31^\circ$). Ten sposób parowania stabilizowany jest przez oddziaływanie z „wywiniętymi” resztami 8C, które powodują zmianę biegu łańcucha fosforanowo-cukrowego oraz konformacji pierścienia cukrowego reszt 7G z C3'-*endo* na C2'-*edno*. Być może świadczy to o tym, że większe nachylenie się reszty cytydyny w kierunku bruzdy małej oraz utworzenie wiązań wodorowych z drugą resztą cytydyny wprowadza duże napięcia strukturalne, które mogą być utrzymane jedynie w obecności sąsiedniej reszty 4G^L, która ma właściwości stabilizujące.

Dysponując otrzymanymi rezultatami, trudno jest jednoznacznie określić dlaczego reszty C-C nie wykazują tendencji do parowania. Być może najdogodniejszą konformacją jaką preferują pary C-C jest nachylenie się jednej z reszt w kierunku bruzdy małej tak, aby utworzyć za pomocą grupy *exo*-aminowej rozgałęzione wiązanie wodorowe z atomem karbonylowym O2 i grupą amidową N3 drugiej reszty (Rysunek 71). Taka konformacja obserwowana jest w strukturze 1j5a rybosomu eubakterii *Deinococcus radiodurans* (144). Aby doszło do takiego oddziaływania odległość pomiędzy atomami C1' musi wynosić powyżej $11,1 \text{ \AA}$ oraz musi nastąpić odkształcenie łańcucha fosforanowo-cukrowego nachylonej cytydyny. Z nieznanymi powodami warunek ten wydaje się być niemożliwy do spełnienia w przypadku powtórzeń CCG. Z drugiej strony, niezwykle jest to, że

reszty cytydynowe w strukturach powtórzeń CCG parują się w taki sposób, że istnieje pomiędzy nimi niekorzystny efekt steryczny. Źródłem obserwowanych efektów jest grupa *exo*-aminowa cytydyny, która jest objętościowo większa niż grupa karbonylowa reszty urydyny.



Rysunek 71. Para C-C (zielony) w strukturze rybosomu 1j5a. Para ta znajduje się w regionie dwuniciowym (A). Jedna z reszt cytydynowych nachylona jest w kierunku bruzdy małej i tworzy rozgałęzione wiązanie wodorowe z drugą resztą C (B). Powoduje to odkształcenie łańcucha fosforanowo-cukrowego nachylonej reszty C (A, B).

4.5. Efekt wprowadzonych modyfikacji

W trzech strukturach krystalicznych zostały wprowadzone modyfikacje. Pierwsza z nich to 8-bromoguanozyna, która znajduje się w dupleksach $[GC(8-BrG)GCGGCGG]_2$ i $[GC(8-BrG)GCGGC]_2$. Z kolei do oligomeru $GCCG^LCCGC$ został wprowadzony nukleotyd guanozynowy z serii LNA. W obu przypadkach nie ma dowodów na to, że zmieniają one strukturę w porównaniu do natywnego RNA. Natywny i bromowany dupleks może być nałożony na siebie z odchyleniem rmsd równym około 1 Å. Podobieństwo to jest również odzwierciedlone w parametrach helikalnych. Porównując oddziaływania z rozpuszczalnikiem, atom bromu wydaje się zastępować cząsteczkę wody, która zlokalizowana jest 3,2 Å od atomu C8 natywnej reszty G(*syn*). Niemniej jego głównym efektem jest tworzenie zawady przestrzennej, która promuje konformację *syn* i jednocześnie ogranicza wolność konformacyjną tej reszty nukleotydowej. Wprowadzenie atomu bromu zmienia wartość pK_a guanozyny z 9,3 do 8,4 (145), co może wpływać na zdolność do tworzenia wiązań wodorowych, aczkolwiek w tym przypadku jest to raczej niezauważalne. Z kolei reszta G^L posiada usztywnioną konformację C3'-*endo* pierścienia cukrowego, co bardzo dobrze koresponduje z formą A-RNA. Właściwością nukleotydów z serii LNA jest przede wszystkim wzrost stabilności struktur kwasów nukleinowych, w skład których wchodzi. W jednym z dupleksów w nici D nastąpiło odkształcenie szkieletu fosforanowo-cukrowego modyfikowanej reszty G^L (Rysunek 53). Gdyby w tej pozycji była niemodyfikowana reszta guanozynowa, prawdopodobnie nie utrzymałby konformacji pierścienia cukrowego C3'-*endo*, co możliwe jest w przypadku G^L . Zgodnie z powyższym, istnieje możliwość, że tego typu nukleotyd nie zmienia struktury, lecz umożliwia reorganizację

szkieletu fosforanowo-cukrowego. W kontekście oddziaływań z rozpuszczalnikiem, zamiast grupy 2' OH znajduje się w tym miejscu atom tlenu tworzący mostek metylenowy. W związku z tym, jego charakter chemiczny uległ zmianie ale nadal może być akceptorem wiązania wodorowego. W bruzdzie małej dupleksu GCCG^LCCGC znajduje się jon siarczanowy, który oddziałuje z grupą *exo*-aminową G^L ale również z atomem O2'. Wiązanie się anionu wskazuje na miejsce o wysokim potencjale do tworzenia oddziaływań, które powstało najprawdopodobniej na skutek zaistniałych deformacji (naprężeń strukturalnych). Prawdopodobnie anion siarczanowy stabilizuje w ten sposób modyfikowaną resztę nukleotydową.

4.6. Dystrybucja konformacji par N-N w dupleksach

Interesującą cechą par N-N jest to, że pomimo symetryczności chemicznej jaką prezentują, są one strukturalnie asymetryczne (wyjątek stanowi para C-C z dupleksu C+D). W strukturach powtórzeń CUG przyjęcie danej konformacji przez pary U-U jest niezależne od tworzenia się innych par U-U. Oznacza to, że dystrybucja konformacji w dupleksach jest dowolna (Rysunek 23). Na jednej nici mogą znajdować się tylko nachylone urydyny, które mogą być również rozłożone równomiernie na obu łańcuchach helisy. W strukturach powtórzeń CGG i CAG faworyzowana jest raczej aranżacja symetryczna (naprzemienna), tzn. gdy np. pierwsza para to G(*syn*)-G(*anti*) to drugą będzie G(*anti*)-G(*syn*) (Tabela 18). Oznacza to, że pomiędzy niekanonicznymi parami w danym dupleksie istnieje kooperatywność. W przypadku powtórzeń CGG, gdzie liczba obserwowanych par jest reprezentatywna, istnieją przykłady, w których obserwowana jest aranżacja asymetryczna. Świadczy to o możliwości uzyskania takiej dystrybucji, chociaż prawdopodobnie jest ona mniej korzystna energetycznie. To może tłumaczyć dlaczego w strukturze [GC(8-BrG)GCGGCGG]₂ w dwóch z trzech dupleksów środkowa para jest nieuporządkowana. Obecność reszty 8-BrG narzuca konformację skrajnych par G-G i wówczas dystrybucja konformacji jest następująca: G(*syn*)-G(*anti*), G(^{*syn*}*anti*)-G(^{*anti*}*syn*) i G(*anti*)-G(*syn*). W takiej sytuacji nie jest możliwe, niezależnie od konformacji środkowej pary, uzyskanie aranżacji naprzemiennnej. O ile nieuporządkowanie może być wytłumaczone, to jaki czynnik decyduje o tym, że w trzecim dupleksie jest uporządkowana para G-G a w niemodyfikowanych dupleksach są przykłady asymetrycznej dystrybucji? Być może potrzebne są do tego dodatkowe elementy stabilizujące, np. kontakt z symetrycznymi cząsteczkami w sieci krystalicznej lub oddziaływania z cząsteczkami rozpuszczalnika. Prawdopodobnie w przypadku par A-A możliwa jest również asymetryczna aranżacja aczkolwiek nie jest ona obserwowana. Inną kwestią wartą rozpatrzenia jest to czy ma znaczenie, która konformacja występuje pierwsza w dupleksie. W strukturze (CGGCGGC)₂ G(*anti*) występuje bliżej końca 5' nieco częściej niż G(*syn*) (8 z 14 dupleksów prezentujących aranżację naprzemienną). Natomiast w powtórzeniach CAG, reszta w konformacji „kciuk do góry” jest zawsze bliżej końca 3'. Sugeruje to, że ta reszta z niekanonicznej pary, która ma bardziej „typową” konformację jest faworyzowana bliżej końca 5'.

Kooperatywność, która występuje pomiędzy parami typu puryna-puryna w dupleksach powtórzeń CNG, powiązana jest z ich wpływem na strukturę przestrzenną, który jest o wiele bardziej widoczny niż w przypadku par U-U. Prawdopodobnie generowane przez nie zmiany w strukturze są lepiej tolerowane, gdy układ konformacji reszt purynowych jest równomiernie rozmieszczony na obu niciach aniżeli skumulowany na jednej.

4.7. Badania strukturalne a badania termodynamiczne

Stabilność termodynamiczna jest bardzo istotną właściwością kwasów nukleinowych, jednakże trudną do badania metodą krystalograficzną. Pewne kalkulacje mogą jednak być dokonane na podstawie liczby wiązań wodorowych czy zakresu oddziaływań warstwowych. W tym kontekście struktura powtórzeń CGG jest najbardziej stabilna termodynamicznie, co jest zgodne z badaniami kalorymetrycznymi. Pary G-G posiadają najwięcej utworzonych wiązań wodorowych (2 + 1 wewnątrznukleotydowe). W chwili, gdy 8-bromoguanozyna zastępuje jedną z reszt z pary G-G następuje wzrost stabilności takiego dupleksu. Temperatura topnienia [GC(8-BrG)GCGGC]₂ jest wyższa o 14 °C niż (GCGGCGGC)₂ (5). Ze strukturalnego punktu widzenia prawdopodobnie związane jest to z ograniczeniem konformacyjnej wolności, które wprowadza atom bromu i narzucenie aranżacji 8-BrG(*syn*)-G(*anti*) oraz wyłączenie możliwości utworzenia pary G(*anti*)-G(*anti*).

Powtórzenia CUG są drugą pod względem stabilności strukturą. Jest to skutkiem istnienia jednego wiązania wodorowego pomiędzy parami U-U. W badaniach termodynamicznych zauważono, że gdy dupleks składa się z dwóch powtórzeń CUG to jego energia swobodna ΔG°_{37} jest relatywnie niższa ($5,08 \pm 0,18$ kcal/mol) niż gdy ma trzy i cztery powtórzenia ($\Delta G^{\circ}_{37} > 7$ kcal/mol). Trudno wytłumaczyć jednoznacznie ten efekt. Być może jest to spowodowane ograniczonymi oddziaływaniami warstwowymi w krótszej cząsteczce. Z drugiej strony, jeżeli spojrzymy na dupleks składający się z dwóch powtórzeń CCG, to jego relatywnie wysoka stabilność termodynamiczna ($6,09 \pm 0,48$ kcal/mol) może być efektem „poślizgu” czyli utworzenia struktury zawierającej jedynie pary G-C i niesparowany koniec 3'. Natomiast, gdy liczba powtórzeń CCG rośnie, to przy czterech następuje obniżenie stabilności dupleksu ($6,05 \pm 0,35$ kcal/mol). Może to świadczyć o utworzeniu się par C-C, które destabilizują dupleks.

Powtórzenia CAG są trzecie w szeregu stabilności. Pomiedzy parami A-A jest słabe wiązanie wodorowe C-H...N. Energia wiązania C-H...X jest szacowana na około 1 kcal/mol i mniej, przy czym C-H...N jest słabsze niż C-H...O. Jednak pomimo niskiej energii, obecność wiązania tego typu odpowiada mierzalnemu efektowi termicznej stabilności dupleksu. W pomiarach termodynamicznych wprowadzenie do dupleksu 8-bromoadenozyny, która przyjmuje konformację *syn*, powoduje wzrost energii swobodnej ΔG°_{37} o 0,7 kcal/mol, co odpowiada obniżeniu temperatury topnienia o 4 °C. Efekt ten wydaje się być addytywny kiedy więcej par A-A jest

modyfikowanych.

Interesującą obserwacją jest to, że łatwość i powtarzalność krystalizacji oligomerów pokrywa się z szeregiem stabilności termodynamicznej powtórzeń CNG. Oligomery zawierające powtórzenia CGG krystalizowały dosyć szybko (czasem w przeciągu kilku godzin), ale trudno było otrzymać monokryształ prawdopodobnie z powodu mobilności konformacyjnej par G-G. Szczepienie kryształów okazało się bardzo dobrym posunięciem, które porządkowało dystrybucję par G-G. Z kolei w przypadku powtórzeń CUG już podczas próbnych krystalizacji wyrosły kryształy odpowiednie do pomiaru. Powtórzenia CAG i CCG sprawiły największe trudności podczas krystalizacji. Powtarzalność krystalizacji była przypadkowa, a próby optymalizacji nie dawały rozsądnych rezultatów. Uważam, że przypadkowy wpływ na bardzo delikatną równowagę w układzie krystalizacyjnym pozwolił otrzymać unikalne kryształy. Oligomer GCCG^LCCGC krystalizował w wielu warunkach co wskazuje, że przyczyną trudności krystalizacji powtórzeń CCG mogła być ich mała stabilność termodynamiczna.

4.8. Aspekt biologiczny

Powtórzenia CNG są unikalną grupą sekwencji mikrosatelitarnych obecnych w mRNA niektórych genów. Tworzą one zarówno regiony jednoniciowe jak i dwuniciowe (struktury spinkowe), co może mieć istotne znaczenie dla ich funkcji w komórce. Obserwowane struktury krystaliczne powtórzeń CAG, CGG i CUG prawdopodobnie odzwierciedlają trzony spinek. Nie obserwuje się efektu rozplatania końców helisy ze względu na sposób upakowania dupleksów w sieci krystalicznej. Pomiedzy sąsiednimi cząsteczkami istnieją oddziaływania warstwowe, które są istotne w tworzeniu pseudonieskończonej helisy. W przypadku powtórzeń CCG obserwowana struktura jest najprawdopodobniej efektem długości krystalizowanego oligomeru. Badania biochemiczne struktury drugorzędowej sugerują bowiem, że długie ciągi powtórzeń CCG tworzą podobną strukturę do pozostałych powtórzeń. Jeżeli tak jest, to istnieje prawdopodobieństwo, że pary C-C będą się parować w sposób podobny do tego jaki obserwuje się w dupleksie niemodyfikowanym, który ma najmniej zaburzoną formę A-RNA.

Końcowym efektem określenia struktur przestrzennych powtórzeń CNG jest wskazanie ich wspólnych cech, które wynikają z podobnej sekwencji, oraz różnic wynikających z obecności niestandardowych par N-N. Szczegółowe zestawienie podobieństw i różnic zostało zebrane w Tabeli 23. Na podstawie sporządzonego profilu można przypuszczać, że białko MBNL1, które w dystrofii miotonicznej wchodzi w skład złogów w jądrze komórkowym, rozpoznaje powtórzenia CUG oraz CAG i CCG (59–61) ze względu na obecność par N-N. We wszystkich trzech przypadkach pomiędzy sparowanymi resztami jest maksymalnie jedno wiązanie wodorowe. Niewykluczone, że MBNL1, które prawdopodobnie wiąże jednoniciowe RNA (63), wymaga rozplecenia helisy

utworzonej przez powtórzenia. W przypadku powtórzeń CGG, pomiędzy resztami guanozyn są dwa wiązania wodorowe i ich rozplecenie przez MBNL1 staje się niemożliwe. Niewykluczone też, że o wiązaniu RNA przez białko MBNL1 decydują inne cechy strukturalne, np. rozkład potencjału elektrostatycznego, który dla wszystkich powtórzeń CNG jest bardzo charakterystyczny i powtarzalny, czy też fakt, że pomimo obecności niestandardowych par N-N przyjmują one formę A-RNA. Wymaga to jednak dalszych badań, które mam nadzieję w przyszłości kontynuować.

4.9. Metodyka otrzymywania stabilnych struktur spinkowych RNA

Otrzymanie stabilnej formy spinki cząsteczek RNA w warunkach krystalizacyjnych było drugim celem mojej pracy doktorskiej. Krystalizacja tego typu struktur RNA należy do bardzo trudnych zadań, ponieważ w warunkach krystalizacji zamiast spinki zwykle otrzymuje się kryształy dupletu (99, 143). Stanowi to barierę dla poznawania takich struktur metodą krystalograficzną, ale również, choć w mniejszym stopniu metodą NMR. Podejmując się realizacji tego projektu moim celem było zrobienie kroku w kierunku rozwoju badań strukturalnych RNA prowadzonych na szeroką skalę i jednocześnie zmniejszenie dysproporcji pomiędzy liczbą poznanych struktur białkowych a RNA.

Prace nad tym zadaniem ukończyłam na etapie opracowania zoptymalizowanego protokołu otrzymywania zcyrkularyzowanych cząsteczek RNA metodą enzymatyczną oraz cząsteczek DNA za pomocą reakcji klik. Metodologia jest uniwersalna, łatwa w użyciu i nie wymaga specjalistycznego sprzętu. Niemniej dwukrotne oczyszczanie produktów w żelu poliakryloamidowym (po syntezie chemicznej i po reakcji cyrkularyzacji) powoduje duże straty. Być może w celu uniknięcia strat dobrym rozwiązaniem byłoby zastosowanie techniki chromatografii faz odwróconych na etapie oczyszczania oligomerów po syntezie chemicznej. Czystość oligomeru uzyskana tym sposobem jest zazwyczaj niższa, ale powinna być wystarczająca do przeprowadzenia kolejnych etapów opracowanej metody. Pozostałych zanieczyszczeń można prawdopodobnie pozbyć się na etapie oczyszczania zcyrkularyzowanego produktu w żelu poliakryloamidowym.

Metoda wykorzystująca reakcję klik wydaje się być bardziej wydajna, jeśli chodzi o etap cyrkularyzacji. Natomiast, jej ograniczeniem jest możliwość wprowadzenia tylko jednego typu łącznika. Niemniej chciałabym również przetestować reakcję klik w celu cyrkularyzacji oligomerów RNA. Sądzę, że pomimo różnic między DNA i RNA nie powinno mieć to dużego znaczenia. Reakcja klik zachodzi bowiem w obrębie grup funkcyjnych połączonych z grupami fosforanowymi, które w obu rodzajach cząsteczek są takie same. Mam również nadzieję, że w przyszłości powiedzie mi się wykrystalizowanie i określenie struktury zcyrkularyzowanego oligomeru oraz, że stworzona przeze mnie metodyka okaże się przydatna dla innych badaczy zajmujących się strukturami RNA.

Tabela 23. Profil powtórzeń CNG—zestawienie cech strukturalnych.

Rodzaj powtórzenia	CGG	CAG	CUG	CCG
Forma RNA	A	A	A	A
Kąt skrętu (°)	30–32	28,5 ± 5,7	33,6 ± 4,1	34–35 (rdzeń 29–34)
Szerokość bruzdy dużej (Å)	17,9 ± 0,9, 17,8 ± 2,5, 14,3 (trzy modele)	23,8 ± 0,2	12,7 ± 2,3	15 (niemodyfikowany), 15,8 (modyfikowany)
Szerokość bruzdy małej (Å)	15,8 ± 0,5, 15,4 ± 0,5, 16,1	15,3 ± 0,1	15,7 ± 0,4	15,7 oraz 14,3
Średni dystans C1'-C1' pary N-N (i pozostałych par) (Å)	11,3 (10,7)	11,0 (10,7)	10,4 (10,5)	10,7–10,9 (10,6)
Wpływ par N-N na strukturę RNA	Lokalne rozkręcenie helisy kompensowane w innym miejscu dupleksu	Rozkręcenie helisy	brak	Brak lub lokalne rozkręcenie helisy kompensowane w innym miejscu dupleksu
Oddziaływania par N-N	N1H...O6 i <i>exo</i> -N2H...N7 i wewnętrznukleotydowe: <i>exo</i> -N2H...O2	C2-H2...N1	N3-H3...O4	Brak jednego wspólnego wzoru parowania pary C-C
Sposób akomodacji par N-N (na podstawie wartości kąta λ)	Jedna reszta G ma konformację <i>syn</i> (nachylona w kierunku bruzdy małej)	Obie reszty przesunięte w kierunku bruzdy dużej	Jedna z reszt U nachylona jest w kierunku bruzdy małej	Brak lub skrócenie płaszczyzn zasad względem siebie
Wpływ par N-N na konformację sąsiedniej pary N-N	Faworyzowana jest aranżacja symetryczna	Aranżacja asymetryczna	brak: każda z par U-U przyjmuje jedną z dwóch możliwych konformacji	Nie określono
Profil elektrostatyczny	Naprzemienne pasy o potencjale dodatnim i ujemnym	Naprzemienne pasy o potencjale dodatnim i ujemnym	Naprzemienne pasy o potencjale dodatnim i ujemnym	Naprzemienne pasy o potencjale dodatnim i ujemnym; w obrębie par C-C potencjał ujemny
Powinowactwo do cząsteczek ligandu	Jon siarczanowy lub wapniowy związane w bruzdzie dużej	Jon siarczanowy związany w bruzdzie dużej	Jon siarczanowy lub glicerol związany poprzez cząsteczkę wody UG	Jon siarczanowy związany w bruzdzie dużej i małej
Ekspozowane grupy funkcyjne par N-N:				
-bruzda duża	G(<i>anti</i>) O6 karbonylowy G(<i>syn</i>) O6 karbonylowy, N1 aminowy, N2 <i>exo</i> -aminowy	Jedna reszta A: N6 aminowy; Druga reszta A: N1 iminowy, N6 aminowy	Jedna reszta U: karbonylowy atom O4	Brak lub jedna reszta C: karbonylowy atom N4
-bruzda mała	G(<i>anti</i>) N3 iminowy, N2 <i>exo</i> -aminowy	Jedna z reszt A: N3 iminowy; Druga reszta A: N3 iminowy	Jedna z reszt U: O2 karbonylowy; Druga reszta U: O2 karbonylowy, N3 aminowy	Brak lub jedna z reszt C: O2 karbonylowy; Druga reszta C: O2 karbonylowy, N3 aminowy

5. Materiały i Metody

5.1. Materiały

5.1.1. Odczynniki chemiczne i biochemiczne

Większość odczynników pochodziła z firm Sigma-Aldrich, Fluka, Lab Empire, Molekula, Bio-Rad lub PoCH. Odczynniki były najwyższej klasy czystości i oznaczone jako wolne od rybonukleaz. Wszystkie filtry zakupione były w firmie Millipore lub Sartorius, kolumny NAP-25 w GE Healthcare a kolumny Sep-Pak C18 w Waters. Amidofosforyny do syntezy chemicznej z blokadą 2'-O-tert butyloдимetylosililową (TBDMS) oraz łącznik C18 zakupione były w firmach Glen Research, Azco, Proligo. Promieniotwórczy roztwór [γ ^{32}P] ATP (4000-5000 Ci/mmol) dostarczany był przez Hartmann Analytic.

5.1.2. Enzymy

Kinaza polinukleotydomowa T4, alkaliczna fosfataza FastAPTM, enzym restrykcyjny HpaII, T4 RNA ligaza, T4 DNA ligaza — zakupione w firmie MBI Fermentas. Terminator Exonuclease – Epicenter.

5.1.3. Stosowane bufony, roztwory i pożywki

Bufor do elektroforezy w warunkach denaturujących 10 × TBE

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

40% roztwór żelu poliakryloamidowego 29 : 1 (w/w)

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

Żel do elektroforezy RNA w warunkach denaturujących

	8%	12%	20%
40% roztworu akryloamid/bisakryloamidu 29:1 (w/w)	100 ml	150 ml	250 ml
10 × TBE (końcowe stężenie 1 ×)	50 ml	50 ml	50 ml
Mocznik (końcowe stężenie 8 M)	240 g	240 g	240 g
H ₂ O	do 500 ml	do 500 ml	do 500 ml

Roztwór dejonizowano z wykorzystaniem żywicy jonowymiennej AG® 501-X8 (Bio-Rad) i filtrowano na filtrach z membraną 0,22 μm . W trakcie przygotowywania żelu dodawano: 0,04% (v/v) TEMED, 0,07% (w/v) nadsiarczanu amonu.

Roztwory obciążające próbki zawierające RNA lub DNA

BB: 8 M mocznik, 0,02% (w/v) błękit bromofenolowy; XC: 8 M mocznik, 0,02% (w/v) ksylencjanol

0,1% roztwór stains-all

30 mg stains-all (bromek 1-etylo-2-[3-(1-etylnafto[1,2-d]) rozpuszczony w 30 ml formamidu

0.005% roztwór stains-all do barwienia żeli poliakrylamidowych

15 ml 0,1% roztworu stains-all, 15 ml formamidu, 75 ml izopropanolu, H₂O do 300 ml

Roztwór 1 × Sybr green II do barwienia żeli poliakrylamidowych

15 μl stężonego 10,000 × roztworu SYBR® Green II RNA w DMSO, 1 × TBE do 150 ml

30% roztwór żelu białkowego 37,5:1 (w/w)

30 g akrylamidu, 0,8 g bisakrylamidu, H₂O do 100 ml

4 × bufor warstwy zagęszczającej żelu białkowego

6,06 g Tris, 4 ml 10% SDS, HCl do pH 6,8, H₂O do 100 ml

4 × bufor warstwy rozdzielającej żelu białkowego

18,17 g Tris, 4 ml 10% SDS, HCl do pH 8,8, H₂O do 100 ml

Żel poliakrylamidowy do elektroforezy białek w warunkach denaturujących

5% żel zagęszczający:

0,83 ml 30% żelu białkowego

2,94 ml H₂O

40 μl nadsiarczanu amonu

9 μl TEMED

końcowa objętość: 5 ml

12% żel rozdzielający:

4 ml 30% żelu białkowego

3,4 ml H₂O

80 μl nadsiarczanu amonu

18 μl TEMED

końcowa objętość: 10 ml

Bufor do elektroforezy białek w warunkach denaturujących

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

Roztwór do denaturacji i obciążania próbek białkowych 4 × SB

200 mM Tris-HCl pH 6,8, 40 mM DTT, 8% SDS, 40% glicerol, 0,08% (w/v) błękit bromofenolowy

Roztwór do dehydratacji żeli białkowych

50% etanol, 10% kwas octowy

Roztwór do barwienia żeli białkowych

50% metanol, 10% kwas octowy, 0,05% barwnik kumasya Brilliant Blue R-250

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 dodawana tuż przed rozpoczęciem hodowli komórek bakteryjnych

Pożywka LB stała

5 g bakto trypton, 2,5 g ekstrakt drożdżowy, 5 g NaCl, 1 ml 1 M NaOH, 7,5 g agaru, H₂O do 500 ml, ampicylina 100 µg/ml dodawana do pożywki o temperaturze najwyżej 50 °C tuż przed zestaleniem.

Bufor podstawowy 2 × stężony używany podczas procesu oczyszczania T4 ligazy RNA

100 mM buforu fosforanowego pH 8,0, 600 mM NaCl

Bufor do lizy komórek bakteryjnych

1 × stężony bufor podstawowy, 10 mM Na₄P₂O₇, 5 mM imidazol

Bufory W1 – W6

	W1	W2	W3	W4	W5	W6
2 × bufor podstawowy	1 ×	1 ×	1 ×	1 ×	1 ×	1 ×
Na ₄ P ₂ O ₇ (mM)	10	-	-	-	-	-
imidazol (mM)	10	20	40	60	100	500

Bufor do przechowywania ligazy T4 RNA

25 mM HEPES-NaOH pH 7,5, 2 mM DTT, 0,1 mM EDTA, 50% glicerol

10 × bufor do ligacji RNA za pomocą T4 RNA ligazy

500 mM Tris-HCl pH 7,5, 100 mM MgCl₂, 100 mM DTT

Plazmid pRHT4

Wektor pRHT4 zawierający gen ligazy T4 (E.C. 6.5.1.3.) został podarowany przez prof. Petera J. Unrau z Uniwersytetu Simona Frasera w Kanadzie. Plazmid posiada sekwencję promotorową (PT7) i terminatorową dla polimerazy RNA T7, miejsce wiązania rybosomu (RBS) i gen (*ble*) oporności na ampicylinę. Pomędzy kodonem start a sekwencją enzymu znajdują się kodony dla 6 reszt histydynowych, zatem eksprymowane białko (63 reszty aminokwasowe; 43 kDa) posiada znacznik histydynowy wykorzystywany podczas oczyszczania.

Szczepy bakteryjne

E. coli BL21 Star (DE3)—Invitrogen

Szczep bakteryjny BL21 Star (DE3) posiada mutację w genie RNazy E co powoduje redukcję degradacji mRNA w komórce. Ponadto zawiera gen polimerazy T7 RNA z faga DE3, który jest pod kontrolą promotora lacUV5. Promotor jest aktywowany za pomocą IPTG (izopropylotiogalaktozyd).

5.2. Metody

5.2.1. Elektroforeza kwasów nukleinowych w żelu poliakrylamidowym w warunkach denaturujących

żel 8% — oczyszczanie reakcji znakowania radioizotopowego; oczyszczanie reakcji ligacji przeprowadzanej w skali analitycznej lub preparatywnej,

żel 12% — oczyszczanie oligomerów po syntezie chemicznej, analiza produktów specyficznej degradacji cząsteczek RNA,

żel 20% — oczyszczanie po reakcji klik prowadzonej w skali analitycznej, analiza produktów specyficznej degradacji cząsteczek RNA.

Elektroforezę w skali preparatywnej prowadzono w żelach o grubości 2 mm, natomiast w skali analitycznej stosowano żele o grubości 0,2 mm lub 2 mm.

Każda elektroforeza prowadzona była w buforze $1 \times$ TBE i poprzedzana była preelektroforezą prowadzoną przy natężeniu prądu 10 lub 15 mA. Przed nałożeniem na żel, próbki mieszano w stosunku 1:1 z roztworem obciążającym XC lub BB. Elektroforezę prowadzono przy natężeniu prądu, które stopniowo zwiększano do uzyskania wartości 25 mA (żel 15×15 cm) lub 50 mA (żel 30×40 cm).

5.2.2. Elucja kwasów nukleinowych

Po rozdziale na żelu poliakrylamidowym pasma odpowiadające właściwym fragmentom RNA wizualizowano w świetle UV lub przez autoradiografię, wycinano z żelu i eluowano metodą „*crash and soak*” przez inkubację w dejonizowanej wodzie przez noc, w temperaturze 4°C , z łagodnym wytrząsaniem.

5.2.3. Strącanie kwasów nukleinowych

Oligomery strącano w obecności 300 mM octanu sodu o pH 5,2, 10 μ g glikogenu oraz z 3 objętościami 96% etanolu. Próbki strącano w temperaturze -20°C przez noc a następnie wirowano przez 25 minut przy 14000 r.p.m. w temperaturze 4°C . Po usunięciu cieczy, osad suszono i rozpuszczano w dejonizowanej wodzie. Stężenie kwasów nukleinowych oznaczano na podstawie pomiarów absorpcji przy długości fali $\lambda = 260$ nm z wykorzystaniem spektrofotometru UV/VIS NanoDrop-1000.

5.2.4. Barwienie żeli poliakrylamidowych z mocznikiem za pomocą roztworu stains-all

Po skończonej elektroforezie żel umieszczano w 0,005% roztworze stains-all na przynajmniej 30 minut. Barwienie odbywało się bez dostępu światła. Następnie żel przekładano na folię i po 10–15 minutowej ekspozycji na światło następowało częściowe odbarwienie żelu, który następnie skanowano.

5.2.5. Elektroforeza białek w żelu poliakrylamidowym w obecności SDS (SDS-PAGE)

Elektroforezę prowadzono w żelach o rozmiarze 10×10 cm i grubości 1 mm, w układzie: 5% żel zagęszczający i 12% żel rozdzielający. Przed elektroforezą, do każdej analizowanej próbki białka dodawano bufor $4 \times \text{SB}$ w stosunku 3:1 i inkubowano przez 5 min. w 95°C . Próbki zwirowywano i nanoszono na żel. W celu łatwiejszego monitorowania elektroforezy, na żel nanoszono wybarwiony marker wielkości białek. Napięcie prądu wynosiło 100 V dla warstwy zagęszczającej, natomiast dla żelu rozdzielającego 150 V. Po skończonej elektroforezie żel umieszczano na 10 minut w roztworze do dehydratacji a następnie barwiono przez 30 minut w roztworze do barwienia. Żel odbarwiano roztworem do odbarwiania a następnie skanowano z użyciem skanera komputerowego.

5.2.6. Ekstrakcja fenolem

Procedurę tą wykorzystywano do oczyszczania produktów po reakcji ligacji prowadzonej w skali preparatywnej. Do próbki zawierającej RNA dodawano 1 objętość fenolu o pH 4,5 i wytrząsano przez 2 minuty w temperaturze pokojowej. Mieszaninę wirowano przez 1 minutę przy 10000 r.p.m. Warstwę wodną przenoszono do nowej próbki. W celu pozbycia się resztek fenolu do fazy wodnej dodawano 1 objętość mieszaniny chloroform:alkohol izoamylowy w stosunku 24:1 (v/v) i wytrząsano 1 minutę w temperaturze pokojowej, po czym wirowano przez 1 minutę przy 10000 r.p.m. Fazę wodną następnie oczyszczano od resztek fenolu i chloroformu na kolumnkach NAP-25.

5.2.7. Odsalanie na kolumnkach NAP-25

Zasoloną lub zanieczyszczoną małowcząsteczkowymi związkami próbkę kwasu nukleinowego oczyszczano na handlowo dostępnych kolumnkach NAP-25 wypełnionych złożem Sephadex G-25. Na kolumnę nakładano 1–1,2 ml próbki a następnie przemywano dejonizowaną wodą. Pierwsze dwie frakcje zbierano po 1 ml a następne po 500 μ l (razem 6 ml). Frakcje zawierające kwas nukleinowy identyfikowano na podstawie pomiaru absorpcji przy długości fali $\lambda = 260$ nm.

5.2.8. Defosforylacja

Do reakcji usuwania grup fosforanowych na końcu 5' oligomeru RNA używano alkalicznej fosfatazy FastAPTM. Reakcję prowadzono w dostarczonym z enzymem buforze ze 160 pmolami RNA w objętości 50 μ l przez 10 minut w temperaturze 37 °C. Enzym inaktywowano inkubując mieszaninę przez 10 minut w temperaturze 75 °C. RNA wytrącano etanolem lub przeprowadzano elektroforezę analityczną w 12% żelu poliakryloamidowym w warunkach denaturujących.

5.2.9. Znakowanie RNA na końcu 5' radioaktywnym izotopem ³²P

Okolo 160 pmoli RNA znakowano za pomocą kinazy polinukleotydowej T4. Przed reakcją, RNA denaturowano przez 5 minut w temperaturze 90 °C a następnie inkubowano przez 10 minut na lodzie. Do tak przygotowanej próbki dodawano bufor dostarczony z enzymem, 1 μ l [γ ³²P] ATP o aktywności 4000–5000 Ci/mmol (0,02 mCi) oraz 10 U enzymu. Reakcję prowadzono przez 30 minut w temperaturze 37 °C. Produkty reakcji oczyszczano w denaturującym 8% żelu poliakryloamidowym. Produkty reakcji wizualizowano przez autoradiografię. Poziom radioaktywności wyznakowanych cząsteczek oznaczano za pomocą licznika scyntylicyjnego.

5.2.10. Hydrolizat formamidowy

Znakowany izotopem ³²P RNA mieszano z formamidem w stosunku 1:5 w obecności 10 mM MgCl₂. Mieszaninę inkubowano w temperaturze 100 °C przez 10 minut po czym dodawano 1 objętość roztworu formamidu z barwnikami elektroforetycznymi i przenoszono próbkę do suchego lodu. Produkty hydrolizy rozdzielano w żelu poliakryloamidowym z 8 M mocznikiem. Następnie wizualizowano produkty reakcji metodą autoradiografii. Do tego celu wykorzystywano skaner radioaktywności Fluorescent Image Analyzer FLA-5100 — FujiFilm.

5.2.11. Ograniczone trawienie RNA rybonukleazą T1 w warunkach semidenaturujących

Do roztworu zawierającego RNA wyznakowany izotopem ³²P dodawano bufor zawierający 50 mM cytrynian sodu o pH 5,3 i 7 M mocznik. Mieszaninę inkubowano w temperaturze 100 °C przez 2 minuty po czym przenoszono do suchego lodu. Następnie dodawano 0,2 U rybonukleazy

T1. Reakcję prowadzono przez 10 minut w temperaturze 55 °C. Reakcję zatrzymywano poprzez zamrożenie próbki w suchym lodzie. Produkty hydrolizy rozdzielano w żelu poliakryloamidowym z 8M mocznikiem po czym wizualizowano metodą autoradiografii z wykorzystaniem skanera radioaktywności Fluorescent Image Analyzer FLA-5100 — FujiFilm.

5.2.12. Reakcja cyrkularyzacji RNA z użyciem ligazy DNA

Mieszanina reakcyjna zawierała 160 pmoli RNA, 1× stężony bufor dostarczony z enzymem, 5% PEG 4000, 5 U enzymu. Przed reakcją roztwór RNA denaturowano przez 5 minut w temperaturze 90 °C w obecności 50 mM KCl a następnie schładzano przez 20 minut do temperatury pokojowej. Reakcję prowadzono w temperaturze 22 °C przez noc po czym RNA strącano lub przeprowadzano rozdział produktów w żelu poliakryloamidowym w warunkach denaturujących.

5.2.13. Ligacja chemiczna

Do reakcji używano 60 pmoli RNA, które uprzednio denaturowano w temperaturze 90 °C przez 2 minuty a następnie schładzano przez 20 minut do temperatury pokojowej. Następnie dodawano NiCl₂ do stężenia końcowego 50 mM, imidazol (końcowe stężenie 100 mM) oraz bromocyjan (końcowe stężenie 80 mM). Reakcję prowadzono przez noc w temperaturze 4 lub 22 °C po czym RNA strącano.

5.2.14. Reakcja klik

Mieszanina reakcyjna zawierała 1 nmol oligomeru (stężenie końcowe 10 μM), 200 mM NaCl, 500 μM CuSO₄ (50-krotny nadmiar w stosunku do oligomeru), 5 mM przygotowany na świeżo askrobinianu sodu (500-krotny nadmiar w stosunku do oligomeru) oraz 10% aceton lub 500 μM PMDETA (N,N,N',N',N''-pentametyldietylenotriamina), które stabilizują miedź na I stopniu utlenienia. Przed reakcją oligomer denaturowano przez 5 minut w temperaturze 90 °C a następnie schładzano przez 30 minut do temperatury pokojowej.

5.2.15. Trawienie enzymem restrykcyjnym HpaII

Do reakcji prowadzonej w 20 μl używano 180 pmoli oligomeru, bufor Tango dostarczony wraz z enzymem oraz 20 U enzymu *HpaII*. Reakcję prowadzono przez 8 godzin w temperaturze 37 °C, strącano a produkty trawienia restrykcyjnego rozdzielano w żelu poliakryloamidowym w warunkach denaturujących. Żel barwiono roztworem stains-all i skanowano.

5.2.16. Oczyszczanie oligomerów RNA i DNA

Oligomery były syntetyzowane chemicznie i oczyszczane we współpracy z Pracownią Chemii RNA prowadzoną przez prof. dr. hab. Ryszarda Kierzka. Syntezę przeprowadzano na syntetyzerze Applied Biosystems z wykorzystaniem metody amidofosforynowej na podłożu stałym — CPG (ang. *controlled pore glass*). Syntezę prowadzono w skali 1 μ mola przy użyciu handlowo dostępnych amidofosforynów z blokadą 2'-O-TBDMS. Po syntezie odblokowywano grupy zasadolabilne i jednocześnie odcinano oligomery od podłoża inkubując złoże z syntetyzowanym RNA w roztworze amoniaku z etanolem w stosunku 3:1 (v/v) przez noc w temperaturze 55 °C. W przypadku oligomerów DNA był to 32% roztwór amoniaku. Mieszaninę schładzano przez 20 minut w temperaturze -20 °C a następnie przenoszono do kolumny z sączkiem. Odsączony płyn odparowywano do sucha (wyparka wirówkowa). Na tym etapie oligomery DNA oczyszczano w 12% żelu poliakryloamidowym w warunkach denaturujących. W przypadku oligomeru RNA kolejnym etapem oczyszczania było zdjęcie blokady siliolowej. Uzyskany osad zalewano mieszaniną fluorku tetrabutylamonowego w DMF i inkubowano do całkowitego rozpuszczenia przez 2-3 godziny (w zależności od długości oligomeru) w temperaturze 55 °C. Oligomery o długości do 11 nukleotydów były strącane w obecności n-butanolu przez godzinę w -20 °C, zwirowywane w 4 °C przez 10 minut przy 5000 r.p.m. Ciecz zlewano znad osadu, który suszono pod próżnią. W międzyczasie przygotowywano kolumny Sep-pak C18 poprzez przemycie ich 10 ml acetonitrylu i zrównoważenie złoża 10 ml 10 mM octanu amonu. Wysuszony osad rozpuszczano w 10 ml 10 mM octanu amonu i nakładano na kolumnę po czym przemycano ją 10 ml 10 mM octanu amonu. Oligomer wymywano 4 ml roztworu acetonitryl/woda w stosunku 3:7 (v/v) i odparowywano do sucha. Odsolony osad oligomeru rozpuszczano w 150 μ l wody i rozdzielano metodą chromatografii cienkowarstwowej na płytkach TLC w fazie 1-propanol/amoniak/woda w stosunku 55:35:10 (v/v/v). Płytkę wizualizowano w świetle UV i wyskrobywano z niej pasmo odpowiadające danemu oligomerowi, który eluowano 5 ml wody.

Oligomery dłuższe niż 12 nukleotydów, po odblokowaniu, zalewano 1 ml 1 M Tris-HCl (pH 7,5) i pozostawiano na 30 min. w temperaturze pokojowej, aby usunąć 2'-hemiacetale. Następnie usuwano DMF poprzez odparowanie mieszaniny do połowy jej objętości początkowej. Oligomery odsalano na kolumnie NAP-25 wypełnionej złożem Sephadex G-25. Oligomer eluowano sterylną, dejonizowaną H₂O. Zebrane frakcje łączono i zatężano (wyparka wirówkowa). Główny produkt syntezy oczyszczano elektroforetycznie w 12% żelu poliakryloamidowym w warunkach denaturujących. Po wizualizacji metodą wygaszania fluorescencji, z żelu wycinano pasma zawierające oligomery o pożądanej długości i eluowano 8 ml sterylnej, dejonizowanej wody przez noc w temperaturze 4 °C. Na koniec oligorybonukleotyd zatężano i odsalano na kolumnkach NAP-25 (elucja sterylną, dejonizowaną H₂O).

Odsolone oligomery RNA i DNA liofilizowano lub zagęszczano i przechowywano w -20°C . Ilość oligomerów w próbce oznaczano przy użyciu spektrometru UV/VIS. Analizę jakościową oligomeru przeprowadzano metodą spektrometrii mas MALDI-TOF MS.

5.2.17. Otrzymywanie ligazy RNA z faga T4

Transformacja komórek *E. coli* wektorem pRHT4

Plazmid zawierający gen ligazy RNA z faga T4 wprowadzano do komórek BL21 Star (DE3) metodą szoku cieplnego. Komórki kompetentne ($50\ \mu\text{l}$) rozmrażano na lodzie przez 30 minut. Następnie dodawano 20 ng wektora pRHT4 i dalej inkubowano na lodzie przez kolejne 30 minut. Zawiesinę bakteryjną umieszczano na 30 sekund w 42°C po czym przenoszono na 2 minuty do lodu, dodawano $250\ \mu\text{l}$ płynnej pożywki LB i wytrząsano przez 1 godzinę przy 200 r.p.m. w temperaturze 37°C . Zawiesinę stransformowanych komórek (50 , 100 lub $150\ \mu\text{l}$) rozprowadzano na powierzchni szalek Petriego ze stałą pożywką LB z antybiotykiem. Hodowlę prowadzono przez 12 godzin w temperaturze 37°C .

Nadekspresja ligazy RNA

Pojedynczą kolonię przenoszono do płynnej pożywki LB z antybiotykiem. Hodowlę prowadzono w objętości 25 ml w temperaturze 37°C przy 230 obrotach/minutę do osiągnięcia gęstości optycznej $\text{OD}_{600} = 0,6$. Zawiesinę zaszczipiano 1 litr płynnej pożywki LB z antybiotykiem, którą następnie rozdzielano po 250 ml do kolb stożkowych o pojemności 1 litra. Inkubacja kontynuowana była do osiągnięcia ponownie wartości $\text{OD}_{600} = 0,6$, po czym indukowano ekspresję za pomocą IPTG, o stężeniu końcowym 1 mM. Po 2 godzinach hodowlę przenoszono do probówek o objętości 0,25 l. Masę komórkową zbierano poprzez wirowanie w 4°C , przy 4000 r.p.m. przez 20 minut i dekantację. Osad komórkowy zamrażano w -80°C .

Podczas pierwszej hodowli komórek przygotowano stoki bakteryjne. Tuż przed indukcją za pomocą IPTG pobrano 4 ml zawiesiny bakteryjnej, do której dodano glicerolu do stężenia końcowego 20%. Zawiesinę rozdzielono na frakcje po $200\ \mu\text{l}$ i przechowywano w -80°C . Stoki bakteryjne były wykorzystywane przy kolejnych hodowlach do zaszczipiania 25 ml płynnej pożywki LB z pominięciem etapu transformacji i hodowli na szalkach Petriego.

Oczyszczanie ligazy T4

Ekstrakt komórkowy rozpuszczano w buforze do lizy: na 5 g osadu dodawano 10 ml buforu z dodatkiem 1 mg/ml lizozymu, 1 tabletki zawierającej inhibitory proteaz bez EDTA oraz 5 mM β -merkaptoetanol. Liza komórek prowadzona była na lodzie przez 30 minut, po czym zawiesinę sonikowano (Ultrasonic Processor, Model CV33, Fisher) przez 3×15 sekund przy amplitudzie 70% mocy. Ekstrakt wirowano w temperaturze 4°C przy 12000 r.p.m. przez 30 minut. Ciecz z osadu przenoszono do czystej probówki o pojemności 250 ml. Do supernatantu lizatu komórkowego dodawano około 1 ml agarozowego złoża niklowego typu Ni-NTA (Qiagen), zrównoważonego

wcześniej buforem do lizy. Całość inkubowano przez 1 godzinę na lodzie delikatnie mieszając, po czym przenoszono do kolumny chromatograficznej o objętości 10 ml, która podłączona była do pompy perystaltycznej. Rozdział wykonywano w temperaturze 4 °C przy przepływie około 0,7 ml/min. Złoże przepłukiwano buforami: 50 ml buforu W1, następnie W2–W6 po 10 ml. Bufory W2–W6 zawierały wzrastające stężenie imidazolu. Frakcje zawierające ligazę T4 identyfikowano elektroforetycznie, łączono i dializowano przez noc względem buforu do przechowywania.

5.2.18. Oznaczanie stężenia białka

Stężenie enzymu oznaczano metodą spektroskopii VIS przy użyciu odczynnika Bradford (Bio-Rad). W zależności od stężenia preparatu białkowego pobierano różne objętości od 1–10 μ l i dopełniano wodą do objętości 0,8 ml. Rozcieńczony enzym mieszano z 200 μ l odczynnika Bradford, mierzono absorpcję przy długości 595 nm a uzyskane wartości porównywano z krzywą wzorcową, przygotowaną w oparciu o standard białkowy — BSA (albumina z surowicy bydlęcej).

5.2.19. Test aktywności T4 RNA ligazy

250 pmoli RNA zawieszano w 60 mM KCl. Mieszaninę ogrzewano przez 2 minuty w temperaturze 90 °C, schładzano przez 20 minut w temperaturze pokojowej, po czym dodawano 10 \times stężonego buforu do ligacji (końcowe stężenie 1 \times), ATP do stężenia 0,5 mM oraz T4 RNA ligazę (końcowe stężenie wynosiło 1 U/ μ l). Reakcję prowadzono przez 4 godziny w temperaturze 37 °C. Następnie próbki rozdzielano elektroforetycznie w żelu poliakrylamidowym z 8 M mocznikiem. Żel barwiono roztworem stains-all a obraz żelu skanowano.

5.2.20. Otrzymywane zcyrkularyzowanych oligomerów RNA w skali preparatywnej

Reakcję prowadzono w objętości 3–4 ml, gdzie stężenie końcowe RNA wynosiło 10 pmol/ μ l. Oligomer mieszano z KCl (końcowe stężenie 60 mM) i inkubowano przez 10 minut w temperaturze 95 °C a następnie schładzano przez 30 minut w temperaturze pokojowej. Dodawano 10 \times stężonego buforu do ligacji (końcowe stężenie 1 \times), ATP do stężenia 0,5 mM oraz T4 RNA ligazę (końcowe stężenie wynosiło 1 U/ μ l). Reakcję prowadzono przez 4 godziny w 37 °C po czym mieszaninę zagęszczano do 1,5 ml (wyparka wirówkowa). Enzym usuwano metodą ekstrakcji fenolem, odsalano chromatograficznie na kolumnach NAP-25, zagęszczano i oczyszczano w 8% żelu poliakrylamidowym z 8 M mocznikiem.

5.2.21. Przygotowanie oligomerów RNA do krystalizacji

Liofilizowane oligorybonukleotydy rozpuszczane były w wodzie i filtrowane na sączkach wirówkowych o średnicy 0,1 μ m Millipore Ultrafree-MC przy szybkości 12000 r.p.m. przez 15

minut w temperaturze pokojowej. Następnie dodawano roztworu soli np. KCl lub octanu amonu, mieszaninę denaturowano przez 10 minut w temperaturze 65 °C lub 95 °C i schładzano do temperatury pokojowej przez co najmniej 30 minut.

5.2.22. Krystalizacja

Proces krystalizacji polega na wytrąceniu substancji w postaci kryształu. Efekt ten można osiągnąć przykładowo poprzez zagęszczanie próbki znajdującej się w roztworze krystalizacyjnym. Najpopularniejszą techniką w biokrystalografii jest metoda dyfuzji par w układzie wiszącej kropli. Proces ten prowadzony jest w zamkniętym układzie np. na płytkach typu Limbro. Posiadają one studzienki, do których nalewa się roztworu krystalizacyjnego, nazywanego też roztworem rezerwurowym. Otwór studzienki zamyka się szczelnie szkiełkiem nakrywkowym, na którym znajduje się wisząca kropla. Zawiera ona krystalizowaną substancję wymieszaną z roztworem rezerwurowym w określonych proporcjach. Roztwór rezerwurowy w kropli jest rozcieńczony, ale następuje wyrównywanie stężeń i potencjałów pomiędzy rozcieńczonym roztworem rezerwurowym w kropli a roztworem rezerwurowym znajdującym się w studziencie. Proces ten zachodzi na drodze dyfuzji par i powoduje zagęszczanie wszystkich substancji znajdujących się w kropli. Pożądanym efektem zagęszczania jest wytrącenie się kryształów.

Jest wiele czynników determinujących wzrost kryształów. Do najważniejszych i/lub dających się najłatwiej kontrolować jest temperatura, stężenie i czystość substancji krystalizowanej, siła jonowa, rodzaj substancji strącającej znajdującej się w roztworze krystalizacyjnym oraz rodzaj jonów. Każda krystalizowana substancja wymaga określenia indywidualnych warunków, w których będzie krystalizować. Niekiedy, kryształy odpowiednie do pomiarów dyfrakcyjnych rosną w roztworze z komercyjnie dostępnych zestawów do krystalizacji. Częściej jednak określa się wstępne warunki krystalizacji, które należy optymalizować. Wiąże się to z systematyczną modyfikacją poszczególnych zmiennych. Podczas optymalizacji można stosować różne rodzaje dodatków (ang. *additives*) lub szczepienie, gdy krystalizuje dużo małych lub pozrastanych kryształów. Szczepienie polega na przeniesieniu zarodków krystalizacji do przesyconego roztworu substancji krystalizowanej. Bardzo ważne jest, aby kropla znajdowała się w obszarze metastabilnym (możliwy jest wzrost zarodków lecz ich powstawanie już nie), a nie labilnym (zarodki tworzą się spontanicznie) lub nienasyconym (zarodki rozpuszczają się). Tylko wówczas następuje optymalny wzrost zarodków krystalizacyjnych, które zostały wprowadzone do kropli.

W moich badaniach stosowałam metodę wiszącej lub siedzącej kropli z dyfuzją par. Wstępne warunki krystalizacji ustalałam przy użyciu komercyjnie dostępnych zestawów roztworów krystalizacyjnych Natrix1, Natrix 2 lub mini Screen (Hampton Research). Typowo, RNA mieszałam z roztworem krystalizacyjnym w stosunku 1:1. Krystalizację oligomeru zawierającego powtórzenia CCG nastawiałam za pomocą robota do krystalizacji Gryphon Art Robbins.

Stosowanie robota pozwala na szybkie przetestowanie różnych warunków krystalizacji, obniżenie ilości oligomeru i roztworów krystalizacyjnych zużywanych w krystalizacji (najmniejsze możliwe do nastawienia krople mają objętość 0,2 μ l). Dodatkową zaletą jest precyzyjność i zminimalizowanie błędów np. podczas pipetowania. Podczas optymalizacji krystalizacji oligomeru GCGGCGGC stosowałam szczepienie kryształów techniką „*streak-seeding*”, w której zarodki krystalizacji przenoszone są za pomocą włosa.

5.2.23. Pomiary dyfrakcyjne

Dane dyfrakcyjne rejestrowane były przy użyciu promieniowania synchrotronowego w następujących ośrodkach:

- EMBL w Hamburgu, wiązka X11 i X13
- BESSY w Berlinie, wiązka BL14.1 i BL14.2

Kryształy transportowano do ośrodka synchrotronowego na płytkach lub zamrożone w zbiorniku utrzymującym temperaturę poniżej 100 K. Pomiar wykonywany był w warunkach kriogenicznych (100 K). Obniżona temperatura ogranicza przede wszystkim uszkodzenia radiacyjne, które mogą powstać na skutek dyfuzji wolnych rodników w kryształach. Dodatkowo zmniejsza nieuporządkowanie dynamiczne oraz poprawia wytrzymałość kryształu (można użyć promieniowania o większej intensywności lub wydłużyć czas naświetlania), co zwykle przekłada się na lepszą jakość zarejestrowanych danych (146). Do zamrażania kryształów stosuje się roztwór krioprotekcyjny, który zawiera roztwór macierzysty oraz krioprotektant (np.: glicerol, PEG 400, 2-metylopentano 2,4-diol, 1,6 heksanediol). Pozwala to na uniknięcie powstawania kryształów lodu podczas zamrażania i utrzymanie rozpuszczalnika w fazie amorficznej.

W trakcie rejestrowania jednego obrazu dyfrakcyjnego kryształ naświetlany jest promieniowaniem X przez określony czas – od kilku do kilkudziesięciu sekund, w zależności od intensywności wiązki i zdolności kryształu do rozpraszania. Co więcej, podczas pojedynczego naświetlania, kryształ obracany jest o kilka stopni ($0,5\text{--}2^\circ$) wokół osi prostopadłej do wiązki promieniowania, przez co więcej refleksów przechodzi przez sferę Ewalda i jest rejestrowanych. Kąt obrotu zależy od odległości płaszczyzn sieci krystalicznej, mozaikowości kryształu oraz od jakości wiązki. Dane rejestruje się do momentu uzyskania maksymalnej kompletności.

5.2.24. Przetwarzanie i skalowanie danych dyfrakcyjnych

Zarejestrowane dane były procesowane i skalowane za pomocą pakietu programów HKL2000 (122): XdisplayF – wizualizacja obrazów dyfrakcyjnych, DENZO – wskaźnikowanie i integracja danych i SCALEPACK – uśrednianie i skalowanie intensywności refleksów.

W pierwszym etapie wybierałam zbiór odpowiednio silnych refleksów, na podstawie których program DENZO wstępnie wskaźnikował oraz obliczał współczynniki niedopasowania względem różnych sieci Bravais’ego. Parametry komórki oraz grupę przestrzenną wybierałam biorąc pod uwagę najwyższą symetrię z niskim współczynnikiem niedopasowania. Następnie, słabsze refleksy były wskaźnikowane a parametry kryształu (rotacja kryształu, parametry komórki elementarnej) i układu pomiarowego (pozycja wiązki, dystans) udokładniane. Po uzyskaniu najlepszej zgodności pomiędzy rzeczywistymi pozycjami refleksów a przewidywanymi, dane integrowałam. Na tym etapie następuje wydzielenie sygnału od tła przy założeniu, że profil sygnału silnych i słabych refleksów jest podobny, przy czym profil tworzony jest na podstawie silnych refleksów. Następnie powierzchnia pod krzywą profilu jest sumowana jako intensywność refleksu. Procedura ta wykonywana jest dla każdego obrazu osobno. W kolejnym etapie używałam programu SCALEPACK, za pomocą którego otrzymywałam spójny zbiór wywskaźnikowanych refleksów i ich intensywności (I) wraz z odchyleniem standardowym $\sigma(I)$. Podczas konwersji danych do wspólnej skali, program SCALEPACK oblicza dla każdego obrazu odpowiedni czynnik skali. W ten sposób uwzględniane są zmiany w intensywności rozpraszania promieniowania rentgenowskiego spowodowane wahaniami intensywności wiązki pierwotnej, różnym czasem naświetlania lub uszkodzeniami radiacyjnymi. Czynniki skali dobierane są poprzez porównywanie intensywności symetrycznych refleksów zakładając, że są identyczne w zakresie błędu statystycznego.

Jakość przetworzonych danych oceniana jest na podstawie takich parametrów jak kompletność, R_{merge} oraz I/σ . Sumaryczna kompletność danych powinna wynosić więcej niż 85%, przy czym w zakresach niższych rozdzielczości (np. 20–8 Å) powinna być wysoka, gdyż ma to wpływ na jakość mapy gęstości elektronowej. Z kolei stosunek $I/\sigma(I)$ w zakresie najwyższej rozdzielczości powinien być większy od 2, ponieważ wówczas intensywność refleksu jest dwukrotnie większa od błędu jego pomiaru. Wartość R_{merge} mówi o tym jaka jest różnica pomiędzy intensywnościami symetrycznych refleksów, które powinny być identyczne (wzór 5.1). Wartość tego parametru powinna być jak najniższa ale należy pamiętać, że zależy ona od krotności (ang. *redundancy*). Oznacza to, że jeśli mamy refleks zmierzony tylko raz to wówczas R_{merge} wynosi 0.

$$R_{merge}(I) = \frac{\sum_{hkl} \sum_I |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_I I_i(hkl)} \quad (5.1)$$

$I_i(hkl)$ – wartość intensywności obserwacji i refleksu hkl

5.2.25. Obliczanie i analiza statystyczna amplitud czynników struktury

Program SCALEPACK generuje plik o rozszerzeniu .sca, który zawiera zbiór intensywności refleksów wraz z ich odchyleniami standardowymi. Kolejnym krokiem jest przekształcenie inten-

sywności refleksów na amplitudy czynników struktury (na podstawie zależności przedstawionej we wzorze 5.2 za pomocą programu SCALEPACK2MTZ i TRUNCATE (147) znajdujących się w pakiecie programów CCP4i (148). W związku z tym, że niektóre wartości intensywności są ujemne oraz intensywności słabych refleksów są niedoszacowane wyznacza się na podstawie wykresu Wilsona czynniki struktury umieszczając wszystkie dane na absolutnej skali. Powoduje to, że wszystkie obserwacje mają wartości dodatnie a słabe refleksy są „doszacowane”.

$$I(hkl) = |F(hkl)|^2 \quad (5.2)$$

Na tym etapie wydzielano też zbiór refleksów używanych do obliczenia czynnika R_{free} . Następnie przeprowadzano analizę statystyczną przetworzonych danych z użyciem programu PHENIX.XTRIAGE. Na jej podstawie można określić obecność symetrii niekryystalograficznej (NCS) (ang. *non-crystallographic symmetry*), bliźniaczenie a także poprawność wybranej grupy przestrzennej.

5.2.26. Rozwiązywanie struktury – podstawienie cząsteczkowe

Rozwiązanie struktury polega na rozwiązaniu problemu fazowego poprzez przypisanie refleksom wstępnych faz. Umożliwia to obliczenie mapy gęstości elektronowej (wzór 5.3), na podstawie której buduje się model cząsteczkowy.

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \quad (5.3)$$

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

Przy rozwiązywaniu struktury oligomerów RNA posłużyłam się metodą podstawienia cząsteczkowego używając programu PHASER (112), który działał w trybie domyślnym. W tej metodzie wstępne fazy czynników struktury uzyskuje się z faz innej znanej struktury cząsteczki – z modelu fazowego. Dlatego też powodzenie tej metody zależy w dużej mierze od podobieństwa strukturalnego obu cząsteczek. Przyjmuje się, że w przypadku białek wystarczy 30% podobieństwo sekwencyjne (www-structmed.cimr.cam.ac.uk/Course/MolRep/molrep.html).

Aby przypisać fazy odpowiednim refleksom, orientacja i położenie modelu fazowego w komórce elementarnej musi pokrywać się z rozwiązywanym modelem. W tym celu potrzebna jest macierz rotacji oraz wektor translacji, które oblicza stosowny do tego celu program. PHASER używa algorytmów opartych na teorii największej wiarygodności (ang. *maximum likelihood*) i stosuje do tego odpowiednie funkcje translacyjne i rotacyjne (149). Wstępne określenie poprawności rozwiązania odbywa się na podstawie dwóch parametrów: LLG (przyrost logarytmu wiarygodności) (ang. *log-likelihood-gain*) oraz z-score. Pierwszy z nich mówi o ile bardziej mogą być przewidziane dane z otrzymanego modelu niż z modelu składającego się z przypadkowo porozrzucanych atomów. Wartość tego parametru powinna być dodatnia i rosnąć wraz z kolejnymi etapami orientowania pełnego modelu. Parametr z-score jest powiązany z przyrostem logarytmu wiarygodności. Mówi o liczbie standardowych odchyłeń powyżej średniej dla danego parametru LLG. Istotne jest, aby jego wartość była jak największa. Uważa się, że z-score powyżej 8 oznacza znalezienie rozwiązania. Niemniej najważniejsza jest inspekcja mapy elektronowej obliczonej z otrzymanych wstępnych faz, ponieważ niekiedy przy niskiej wartości z-score (np. 6) otrzymuje się model, który jest częściowo prawidłowy i można go wykorzystać do budowy poprawnego modelu.

5.2.27. Natywna mapa Pattersona

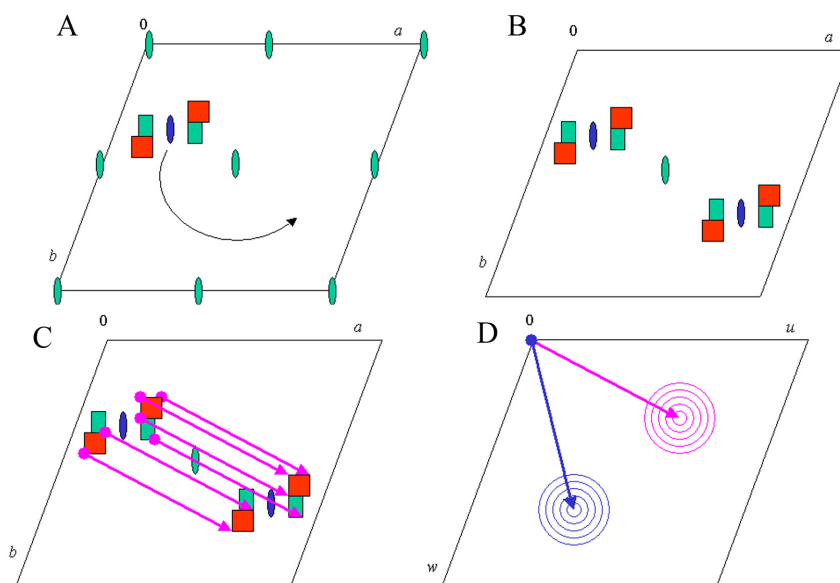
Mapę Pattersona oblicza się jedynie na podstawie amplitudy czynników struktury, które uzyskuje się z eksperymentu dyfrakcyjnego (150). W tym przypadku nie ma potrzeby znajomości fazy. Mapę wylicza się na podstawie wzoru funkcji Pattersona (wzór 5.4).

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 e^{-2\pi i(hu+kv+lw)} \quad (5.4)$$

- P – wartość gęstości elektronowej w punkcie xyz
- $|F(hkl)|$ – amplituda czynnika struktury dla refleksu (hkl)
- u, v, w – współrzędne w przestrzeni rzeczywistej
- h, k, l – współrzędne w przestrzeni odwrotnej

Otrzymana mapa posiada piki, które znajdują się w pozycjach odpowiadających końcom wektorów międzyatomowych (Rysunek 72). Oznacza to, że jeżeli mamy atom w pozycji x_1 oraz drugi atom w pozycji x_2 to wówczas na mapie Pattersona otrzymamy dwa piki w pozycji x_2-x_1 oraz x_1-x_2 ([http : //www – structmed.cimr.cam.ac.uk/Course/Basic_phasing/Phasing.html](http://www-structmed.cimr.cam.ac.uk/Course/Basic_phasing/Phasing.html)). Wysokość piku jest proporcjonalna do liczby elektronów wchodzących w skład atomów tworzących wektor międzyatomowy. Jeżeli w strukturze znajduje się kilka atomów, to jest możliwe obliczenie na podstawie mapy Pattersona ich pozycji w sieci krystalicznej. Niestety przy większej liczbie

atomów staje się to już niemożliwe. Niemniej mapa Pattersona jest bardzo pomocna np. w procesie rozwiązywania struktury czy do określenia obecności niekrystalograficznej symetrii.



Rysunek 72. Schemat przedstawiający powiązanie niekrystalograficznej osi symetrii z mapą Pattersona (według http://www.ruppweb.org/Xray/Patterson/Native_Patterson.htm). Dwie cząsteczki powiązane są niekrystalograficzną osią dwukrotną (niebieska elipsa) (A). Po przekształceniu ich za pomocą osi dwukrotnej, w komórce znajdują się cztery cząsteczki (B). Pomiedzy cząsteczkami niepowiązаныmi ze sobą żadnym elementem symetrii możemy narysować wektory translacyjne (różowe strzałki). Oznacza to, że za ich pomocą możemy przesunąć jedną cząsteczkę w miejsce drugiej (C). Na mapie Pattersona są zaznaczone dwa piki (D). Jeden z nich znajduje się w miejscu końca wektora translacyjnego (różowe strzałki) zawieszonoego w początku układu współrzędnych. Drugi pik reprezentuje wektory wyznaczone pomiedzy cząsteczkami powiązаныmi elementem symetrii.

5.2.28. Udokładnianie modeli RNA

Celem udokładniania struktury jest zmodyfikowanie modelu cząsteczki otrzymanego po rozwiązaniu problemu fazowego w taki sposób, aby najwierniej odzwierciedlał dane eksperymentalne. Zgodność pomiędzy obliczonymi, na podstawie modelu, i obserwowanymi czynnikami struktury monitorowana jest wartościami czynników rozbieżności R i R_{free} (wzory 5.5, 5.6). Zbiór refleksów R_{free} nie jest używany podczas procesu udokładniania. Jego wartość oblicza się na podstawie udokładnionego modelu, co pozwala na monitorowanie poprawności tego procesu. Przykładowo po poprawnym ulepszeniu modelu (np. wprowadzeniu anizotropowych czynników temperaturowych) powinien nastąpić spadek wartości obu czynników R . Dla idealnego modelu (brak nadinterpretacji danych), wartości obu czynników R powinny być podobne ale zwykle R_{free} jest większe od R .

Procedura udokładniania obejmuje dwa, cyklicznie powtarzane etapy: ręczną przebudowę modelu w oparciu o mapy gęstości elektronowej oraz automatyczną minimalizację funkcji opisującej rozbieżność między obliczonymi a obserwowanymi amplitudami czynników struktury (151).

Dokonyje się tego poprzez zmianę czynnika temperaturowego (np. izotropowo lub anizotropowo) oraz położenia i obsadzenia każdego atomu. Dodatkowo w procesie udokładniania biomolekuł bardzo często stosuje się dane stereochemiczne ustalone na podstawie wysokorozdzielczych struktur małowzrostkowych (152). Jeżeli zastosuje się je jako więzy miękkie (ang. *restrains*) to wówczas do obliczeń stosowane są dodatkowe równania (obserwacje), w których niektóre zmienne (np. długości wiązań) dążą do „idealnych” wartości z zachowaniem pewnej wariancji. Natomiast, gdy zostaną użyte jako więzy sztywne (ang. *constrains*), redukuje się liczbę parametrów poprzez wprowadzenie dla nich ustalonych wartości. Ponadto stosuje się korekcję obszarów zajętych przez nieuporządkowane cząsteczki rozpuszczalnika (ang. *bulk solvent correction*) oraz parametry TLS (ang. *translation, libration, screw motion*), które częściowo odzwierciedlają anizotropię skorelowanych drgań całych grup atomów. Stosowanie tych metod zwiększa efektywność procesu udokładniania poprzez poprawę stosunku liczby obserwacji do liczby parametrów.

$$R = \frac{\sum_h w_h ||F_o(h)| - k|F_c(h)||}{\sum_h |F_o(h)|} \quad (5.5)$$

$$R_{free} = \frac{\sum_{h \in T} w_h ||F_o(h)| - k|F_c(h)||}{\sum_{h \in T} |F_o(h)|} \quad (5.6)$$

R – krystalograficzny liniowy wskaźnik rozbieżności

R_{free} – testowy wskaźnik rozbieżności

$|F_o(h)|$ – moduł obserwowanego czynnika struktury

$|F_c(h)|$ – moduł obliczonego czynnika struktury

h – współrzędne węzła sieci odwrotnej hkl

T – zbiór wskaźników refleksów testowych

k – czynnik skali

w_h – waga przypisana obserwacji hkl

Wszystkie struktury udokładniano za pomocą programów REFMAC5 (113), PHENIX (117) lub SHELXL (116). Ostatni z nich opiera się na metodzie najmniejszych kwadratów (ang. *least squares*) opisaną jako prawdopodobieństwo przez intuicję (ang. *„likelihood by intuition”*). Pozostałe programy stosują metodę największej wiarygodności. Metoda ta jest mniej wrażliwa na błędy w modelu oraz na niekompletność struktury. Z kolei inspekcję map gęstości elektronowych i przebudowę modelu dokonywałam za pomocą programu COOT (114).

5.2.29. Pozostałe programy użyte w pracy

ARP/wARP – program służący do automatycznego budowania modelu oraz udokładniania. Za jego pomocą dodawane były cząsteczki wody. Program działał w trybie domyślnym (115).

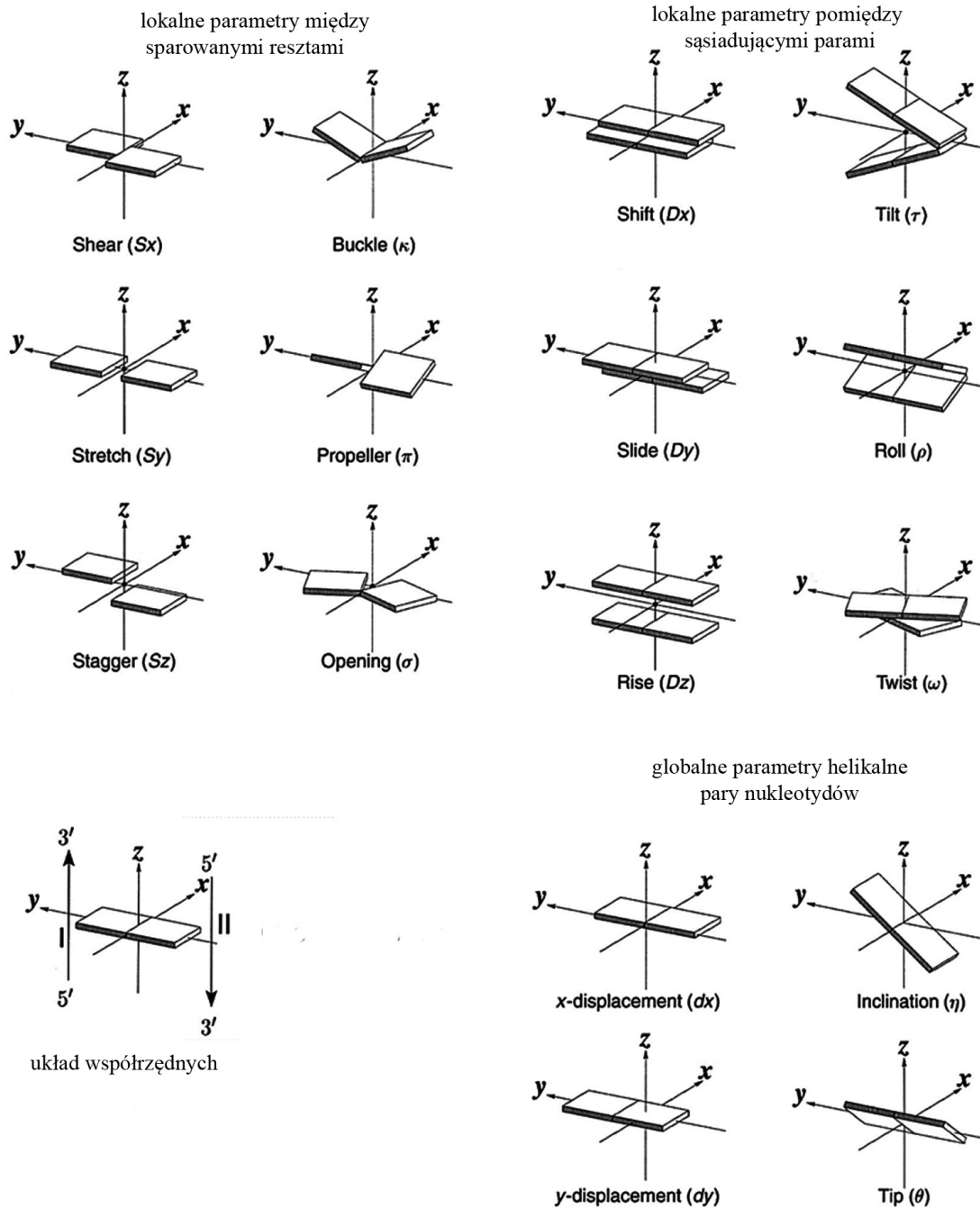
Lsqkab – wykorzystywany do porównywania struktur trzeciorzędowych. Jeden zbiór współrzędnych jest przesuwany, tak aby nakładał się na referencyjny zbiór współrzędnych. Dla kwasów nukleinowych możliwe jest jedynie nakładanie na siebie atomów. W przypadku białek możliwe jest nakładanie atomów węgla C α lub łańcuchów głównych i bocznych (153).

CHARMM GUI – program służący do obliczania mapy potencjału elektrostatycznego (154). Adres WWW: <http://www.charmm-gui.org/>.

PYMOL (155) oraz UCSF Chimera (156) – programy do wizualizacji struktur krystalicznych.

3DNA – wykorzystywany do analizy struktur kwasów nukleinowych (126). Oblicza szereg parametrów heliakalnych: lokalne parametry między sparowanymi resztami, lokalne parametry pomiędzy sąsiadującymi parami, globalne parametry helikalne pary nukleotydów wyznaczone względem osi helisy (Rysunek 73). Ponadto wyznacza wartości kątów torsyjnych, formę kwasu nukleinowego, obszar wiązań warstwowych, oddziaływania pomiędzy parami. Parametry helikalne jak przesunięcie (ang. *displacement*), kąt (ang. *angle*), kąt skrętu (ang. *twist*) oraz odległość pomiędzy płaszczyznami zasad (ang. *rise*) są obliczane na podstawie niezależnych od sekwencji wektorów wyznaczanych przy użyciu pozycji atomów C1' sparowanych reszt nukleotydowych (157). Parametr przesunięcie oznacza przesunięcie środkowego punktu linii łączącej atomy C1' sparowanych reszt nukleotydowych od osi helisy. Parametr kąt oznacza kąt pomiędzy wektorem C1'-C1' a osią helisy (odjęty od 90°). Parametr kąt skrętu jest to kąt pomiędzy wektorami dwóch sąsiednich wektorów C1'-C1'. Parametr odległość pomiędzy płaszczyznami sąsiednich par zasad jest długością wektora powstałego po zrzutowaniu środkowych punktów dwóch sąsiednich wektorów C1'-C1' na oś helisy.

FRABASE – wyszukiwarka motywów strukturalnych w cząsteczkach RNA zdeponowanych w PDB (<http://rnafrabase.ibch.poznan.pl>) (139).



Rysunek 73. Parametry helikalne wyznaczone przez program 3DNA.

Literatura

1. Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M. i Kashi, Y. (2000) Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Res.*, **10**, 62–71.
2. Toth, G., Gaspari, Z. i Jurka, J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, **10**, 967–981.
3. Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.*, **5**, 435–445.
4. Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. i Lathrop, M. (1992) A second-generation linkage map of the human genome. *Nature*, **359**, 794–801.
5. Broda, M., Kierzek, E., Gdaniec, Z., Kulinski, T. i Kierzek, R. (2005) Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry*, **44**, 10873–10882.
6. Krzyzosiak, W.J., Sobczak, K. i Napierala, M. (2006) In Wells, R. D. and Ashizawa, T. (eds.), Genetic Instabilities and Neurological Diseases. *Elsevier*, pp. 705–716.
7. Kozłowski, P., de Mezer, M. i Krzyzosiak, W.J. (2010) Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.*, **38**, 4027–4039.
8. Sobczak, K., Michlewski, G., de Mezer, M., Kierzek, E., Krol, J., Olejniczak, M., Kierzek, R. i Krzyzosiak, W.J. (2010) Structural diversity of triplet repeat RNAs. *J. Biol. Chem.*, **285**, 12755–12764.
9. Orr, H.T. i Zoghbi, H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
10. Mirkin, S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.
11. Lopez Castel, A., Cleary, J.D. i Pearson, C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.*, **11**, 165–170.
12. Napierala, M., Michalowski, D., de Mezer, M. i Krzyzosiak, W.J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.*, **33**, 451–463.
13. Sobczak, K. i Krzyzosiak, W.J. (2004) Imperfect CAG repeats form diverse structures in SCA1 transcripts. *J. Biol. Chem.*, **279**, 41563–41572.
14. Sobczak, K. i Krzyzosiak, W.J. (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J. Biol. Chem.*, **280**, 3898–3910.
15. Pearson, C.E., Nichol Edamura, K. i Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.*, **6**, 729–742.
16. La Spada, A.R. i Taylor, J.P. (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.*, **11**, 247–258.
17. Ranum, L.P. i Cooper, T.A. (2006) RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.*, **29**, 259–277.
18. Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., Warren, S.T. i in. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.

19. Dombrowski, C., Levesque, S., Morel, M.L., Rouillard, P., Morgan, K. i Rousseau, F. (2002) Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.*, **11**, 371–378.
20. Zhong, N., Ju, W., Pietrofesa, J., Wang, D., Dobkin, C. i Brown, W.T. (1996) Fragile X "gray zone" alleles: AGG patterns, expansion risks, and associated haplotypes. *Am. J. Med. Genet.*, **64**, 261–265.
21. Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B. i Hagerman, P.J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, **57**, 127–130.
22. Jacquemont, S., Hagerman, R.J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J.A., Greco, C., Des Portes, V., Jardini, T., Levine, R. i in. (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.*, **72**, 869–878.
23. Sherman, S.L. (2000) Premature ovarian failure in the fragile X syndrome. *Am. J. Med. Genet.*, **97**, 189–194.
24. Glass, I.A. (1991) X linked mental retardation. *J. Med. Genet.*, **28**, 361–371.
25. Gecz, J., Gedeon, A.K., Sutherland, G.R. i Mulley, J.C. (1996) Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat. Genet.*, **13**, 105–108.
26. Sofola, O.A., Jin, P., Botas, J. i Nelson, D.L. (2007) Argonaute-2-dependent rescue of a Drosophila model of FXTAS by FRAXE premutation repeat. *Hum. Mol. Genet.*, **16**, 2326–2332.
27. Annesi, G., Nicoletti, G., Tarantino, P., Cutuli, N., Annesi, F., Marco, E.V., Zappia, M., Morgante, L., Arabia, G., Pugliese, P. i in. (2004) FRAXE intermediate alleles are associated with Parkinson's disease. *Neurosci. Lett.*, **368**, 21–24.
28. Costa, A., Gao, L., Carrillo, F., Caceres-Redondo, M.T., Carballo, M., Diaz-Martin, J., Gomez-Garre, P., Sobrino, F., Lucas, M., Lopez-Barneo, J. i in. (2011) Intermediate alleles at the FRAXA and FRAXE loci in Parkinson's disease. *Parkinsonism Relat. Disord.*, **17**, 281–284.
29. Murray, A., Youings, S., Dennis, N., Latsky, L., Linehan, P., McKechnie, N., Macpherson, J., Pound, M. i Jacobs, P. (1996) Population screening at the FRAXA and FRAXE loci: molecular analyses of boys with learning difficulties and their mothers. *Hum. Mol. Genet.*, **5**, 727–735.
30. Youings, S.A., Murray, A., Dennis, N., Ennis, S., Lewis, C., McKechnie, N., Pound, M., Sharrock, A. i Jacobs, P. (2000) FRAXA and FRAXE: the results of a five year survey. *J. Med. Genet.*, **37**, 415–421.
31. Philips, A.V., Timchenko, L.T. i Cooper, T.A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science*, **280**, 737–741.
32. Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. i Singer, R.H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.*, **128**, 995–1002.
33. Raca, G., Siyanova, E.Y., McMurray, C.T. i Mirkin, S.M. (2000) Expansion of the (CTG)(n) repeat in the 5'-UTR of a reporter gene impedes translation. *Nucleic Acids Res.*, **28**, 3943–3949.
34. Sasagawa, N., Saitoh, N., Shimokawa, M., Sorimachi, H., Maruyama, K., Arahata, K., Isiura, S. i Suzuki, K. (1996) Effect of artificial (CTG) repeat expansion on the expression of myotonin protein kinase (MtPK) in COS-1 cells. *Biochim. Biophys. Acta*, **1315**, 112–116.
35. Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N.G., Vooijs, M. i in. (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature*, **355**, 548–551.

36. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T. i in. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **69**, 385.
37. Harley, H.G., Brook, J.D., Rundle, S.A., Crow, S., Reardon, W., Buckler, A.J., Harper, P.S., Housman, D.E. i Shaw, D.J. (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature*, **355**, 545–546.
38. Ranum, L.P., Rasmussen, P.F., Benzow, K.A., Koob, M.D. i Day, J.W. (1998) Genetic mapping of a second myotonic dystrophy locus. *Nat. Genet.*, **19**, 196–198.
39. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M. i Thornton, C.A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769–1773.
40. Jansen, G., Groenen, P.J., Bachner, D., Jap, P.H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J.J. i in. (1996) Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat. Genet.*, **13**, 316–324.
41. Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N. i Tapscott, S.J. (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell*, **20**, 483–489.
42. Klesert, T.R., Otten, A.D., Bird, T.D. i Tapscott, S.J. (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet.*, **16**, 402–406.
43. Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H. i Housman, D.E. (1997) Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. U S A*, **94**, 7388–7393.
44. de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N.A., Ashizawa, T., Cooper, T.A. i Botas, J. (2006) MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. *Hum. Mol. Genet.*, **15**, 2138–2145.
45. Seznec, H., Agbulut, O., Sergeant, N., Savouret, C., Ghestem, A., Tabti, N., Willer, J.C., Ourth, L., Duros, C., Brisson, E. i in. (2001) Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum. Mol. Genet.*, **10**, 2717–2726.
46. Osborne, R.J. i Thornton, C.A. (2006) RNA-dominant diseases. *Hum. Mol. Genet.*, **15 Spec No 2**, R162–169.
47. Miller, J.W., Urbinati, C.R., Teng-Umuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A. i Swanson, M.S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO*, **19**, 4439–4448.
48. Ho, T.H., Charlet, B.N., Poulos, M.G., Singh, G., Swanson, M.S. i Cooper, T.A. (2004) Muscleblind proteins regulate alternative splicing. *EMBO*, **23**, 3103–3112.
49. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. i Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805–814.
50. Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S. i Thornton, C.A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.*, **10**, 2165–2170.
51. Timchenko, L.T., Timchenko, N.A., Caskey, C.T. i Roberts, R. (1996) Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum. Mol. Genet.*, **5**, 115–121.

52. Ladd, A.N., Stenberg, M.G., Swanson, M.S. i Cooper, T.A. (2005) Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev. Dyn.*, **233**, 783–793.
53. Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W. i Swanson, M.S. (2003) A muscleblind knockout model for myotonic dystrophy. *Science*, **302**, 1978–1980.
54. Ho, T.H., Bundman, D., Armstrong, D.L. i Cooper, T.A. (2005) Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum. Mol. Genet.*, **14**, 1539–1547.
55. Dansithong, W., Paul, S., Comai, L. i Reddy, S. (2005) MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J. Biol. Chem.*, **280**, 5773–5780.
56. Lin, X., Miller, J.W., Mankodi, A., Kanadia, R.N., Yuan, Y., Moxley, R.T., Swanson, M.S. i Thornton, C.A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.*, **15**, 2087–2097.
57. Todd, P.K. i Paulson, H.L. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.*, **67**, 291–300.
58. Pascual, M., Vicente, M., Monferrer, L. i Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, **74**, 65–80.
59. Ho, T.H., Savkur, R.S., Poulos, M.G., Mancini, M.A., Swanson, M.S. i Cooper, T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell Sci.*, **118**, 2923–2933.
60. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N. i Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495–507.
61. Yuan, Y., Compton, S.A., Sobczak, K., Stenberg, M.G., Thornton, C.A., Griffith, J.D. i Swanson, M.S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.*, **35**, 5474–5486.
62. Warf, M.B. i Berglund, J.A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*, **13**, 2238–2251.
63. Teplova, M. i Patel, D.J. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.*, **15**, 1343–1351.
64. Chen, L.S., Tassone, F., Sahota, P. i Hagerman, P.J. (2003) The (CGG)_n repeat element within the 5' untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum. Mol. Genet.*, **12**, 3067–3074.
65. Primerano, B., Tassone, F., Hagerman, R.J., Hagerman, P., Amaldi, F. i Bagni, C. (2002) Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *RNA*, **8**, 1482–1488.
66. Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E. i Hagerman, P.J. (2000) Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
67. De Rubeis, S. i Bagni, C. (2010) Fragile X mental retardation protein control of neuronal mRNA metabolism: Insights into mRNA stability. *Mol. Cell Neurosci.*, **43**, 43–50.
68. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P. i in. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.

69. Kenneson, A., Zhang, F., Hagedorn, C.H. i Warren, S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.*, **10**, 1449–1454.
70. Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del Bigio, M.R., Jacquemont, S., Leehey, M. i Hagerman, P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.
71. Iwahashi, C.K., Yasui, D.H., An, H.J., Greco, C.M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C.B., Hagerman, R.J. i Hagerman, P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256–271.
72. Jin, P., Zarnescu, D.C., Zhang, F., Pearson, C.E., Lucchesi, J.C., Moses, K. i Warren, S.T. (2003) RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*. *Neuron.*, **39**, 739–747.
73. Van Dam, D., Errijgers, V., Kooy, R.F., Willemsen, R., Mientjes, E., Oostra, B.A. i De Deyn, P.P. (2005) Cognitive decline, neuromotor and behavioural disturbances in a mouse model for fragile-X-associated tremor/ataxia syndrome (FXTAS). *Behav. Brain Res.*, **162**, 233–239.
74. Willemsen, R., Hoogeveen-Westerveld, M., Reis, S., Holstege, J., Severijnen, L.A., Nieuwenhuizen, I.M., Schrier, M., van Unen, L., Tassone, F., Hoogeveen, A.T. i in. (2003) The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. *Hum. Mol. Genet.*, **12**, 949–959.
75. Garcia-Arocena, D. i Hagerman, P.J. Advances in understanding the molecular basis of FXTAS. *Hum. Mol. Genet.*, **19**, R83–89.
76. Jin, P., Duan, R., Qurashi, A., Qin, Y., Tian, D., Rosser, T.C., Liu, H., Feng, Y. i Warren, S.T. (2007) Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome. *Neuron.*, **55**, 556–564.
77. Sofola, O.A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., Nelson, D.L. i Botas, J. (2007) RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS. *Neuron.*, **55**, 565–571.
78. Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R.K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D.J. i in. (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO*, **29**, 1248–1261.
79. Chawla, G., Lin, C.H., Han, A., Shiue, L., Ares, M., Jr. i Black, D.L. (2009) Sam68 regulates a set of alternatively spliced exons during neurogenesis. *Mol. Cell Biol.*, **29**, 201–213.
80. Lukong, K.E. i Richard, S. (2008) Motor coordination defects in mice deficient for the Sam68 RNA-binding protein. *Behav. Brain Res.*, **189**, 357–363.
81. Gecz, J., Bielby, S., Sutherland, G.R. i Mulley, J.C. (1997) Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. *Genomics*, **44**, 201–213.
82. Nilson, I., Reichel, M., Ennas, M.G., Greim, R., Knorr, C., Siegler, G., Greil, J., Fey, G.H. i Marschalek, R. (1997) Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. *Br. J. Haematol.*, **98**, 157–169.
83. Gu, Y., McIlwain, K.L., Weeber, E.J., Yamagata, T., Xu, B., Antalffy, B.A., Reyes, C., Yuva-Paylor, L., Armstrong, D., Zoghbi, H. i in. (2002) Impaired conditioned fear and enhanced long-term potentiation in *Fmr2* knock-out mice. *J. Neurosci.*, **22**, 2753–2763.

84. Katsuno, M., Adachi, H., Kume, A., Li, M., Nakagomi, Y., Niwa, H., Sang, C., Kobayashi, Y., Doyu, M. i Sobue, G. (2002) Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. *Neuron.*, **35**, 843–854.
85. Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y. i Orr, H.T. (1998) Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell*, **95**, 41–53.
86. Wheeler, V.C., Gutekunst, C.A., Vrbanac, V., Lebel, L.A., Schilling, G., Hersch, S., Friedlander, R.M., Gusella, J.F., Vonsattel, J.P., Borchelt, D.R. i in. (2002) Early phenotypes that presage late-onset neurodegenerative disease allow testing of modifiers in Hdh CAG knock-in mice. *Hum. Mol. Genet.*, **11**, 633–640.
87. Li, L.B., Yu, Z., Teng, X. i Bonini, N.M. (2008) RNA toxicity is a component of ataxin-3 degeneration in *Drosophila*. *Nature*, **453**, 1107–1111.
88. de Mezer, M., Wojciechowska, M., Napierala, M., Sobczak, K. i Krzyzosiak, W.J. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res.*, **39**, 3852–3863.
89. Michlewski, G. i Krzyzosiak, W.J. (2004) Molecular architecture of CAG repeats in human disease related transcripts. *J. Mol. Biol.*, **340**, 665–679.
90. Jasinska, A., Michlewski, G., de Mezer, M., Sobczak, K., Kozłowski, P., Napierala, M. i Krzyzosiak, W.J. (2003) Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Res.*, **31**, 5463–5468.
91. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. i Krzyzosiak, W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.*, **31**, 5469–5482.
92. Napierala, M. i Krzyzosiak, W.J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.*, **272**, 31079–31085.
93. Sobczak, K. i Krzyzosiak, W.J. (2004) Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Hum. Mutat.*, **24**, 236–247.
94. Khateb, S., Weisman-Shomer, P., Hershco, I., Loeb, L.A. i Fry, M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)_n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
95. Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludwig, A.L. i Fry, M. (2007) The tetraplex (CGG)_n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.*, **35**, 5775–5788.
96. Ofer, N., Weisman-Shomer, P., Shklover, J. i Fry, M. (2009) The quadruplex r(CG G)_n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
97. Zumwalt, M., Ludwig, A., Hagerman, P.J. i Dieckmann, T. (2007) Secondary structure and dynamics of the r(CG G) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. *RNA Biol.*, **4**, 93–100.
98. Mooers, B.H., Logue, J.S. i Berglund, J.A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. U S A*, **102**, 16626–16631.
99. Holbrook, S.R., Cheong, C., Tinoco, I., Jr. i Kim, S.H. (1991) Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature*, **353**, 579–581.

100. Bevilacqua, P.C. i Blöse, J.M. (2008) Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem.*, **59**, 79–103.
101. Svoboda, P. i Di Cara, A. (2006) Hairpin RNA: a secondary structure of primary importance. *Cell Mol. Life Sci.*, **63**, 901–908.
102. Pils, W. i Micura, R. (2000) Flexible non-nucleotide linkers as loop replacements in short double helical RNAs. *Nucleic Acids Res.*, **28**, 1859–1863.
103. Lewis, F.D., Wu, Y. i Liu, X. (2002) Synthesis, structure, and photochemistry of exceptionally stable synthetic DNA hairpins with stilbene diether linkers. *J. Am. Chem. Soc.*, **124**, 12165–12173.
104. Rumney, S. i Kool, E.T. (1995) Structural Optimization of Non-Nucleotide Loop Replacements for Duplex and Triplex DNAs. *J. Am. Chem. Soc.*, **117**, 5635–5646.
105. Junker, H.D., Hoehn, S.T., Bunt, R.C., Marathius, V., Chen, J., Turner, C.J. i Stubbe, J. (2002) Synthesis, characterization and solution structure of tethered oligonucleotides containing an internal 3'-phosphoglycolate, 5'-phosphate gapped lesion. *Nucleic Acids Res.*, **30**, 5497–5508.
106. Kozerski, L., Mazurek, A.P., Kawecki, R., Bocian, W., Krajewski, P., Bednarek, E., Sitkowski, J., Williamson, M.P., Moir, A.J. i Hansen, P.E. (2001) A nicked duplex decamer DNA with a PEG(6) tether. *Nucleic Acids Res.*, **29**, 1132–1143.
107. Egli, M., Tereshko, V., Mushudov, G.N., Sanishvili, R., Liu, X. i Lewis, F.D. (2003) Face-to-face and edge-to-face pi-pi interactions in a synthetic DNA hairpin with a stilbenediether linker. *J. Am. Chem. Soc.*, **125**, 10842–10849.
108. Carriero, S. i Damha, M.J. (2003) Synthesis of lariat-DNA via the chemical ligation of a dumbbell complex. *Org. Lett.*, **5**, 273–276.
109. Wang, S. i Kool, E.T. (1994) Circular RNA oligonucleotides. Synthesis, nucleic acid binding properties, and a comparison with circular DNAs. *Nucleic Acids Res.*, **22**, 2326–2333.
110. Lietard, J., Meyer, A., Vasseur, J.J. i Morvan, F. (2008) New strategies for cyclization and bicyclization of oligonucleotides by click chemistry assisted by microwaves. *J. Org. Chem.*, **73**, 191–200.
111. Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
112. Storoni, L.C., McCoy, A.J. i Read, R.J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 432–438.
113. Murshudov, G.N., Vagin, A.A. i Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
114. Emsley, P. i Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
115. Lamzin, V.S. i Wilson, K.S. (1993) Automated refinement of protein models. *Acta Crystallogr. D Biol. Crystallogr.*, **49**, 129–147.
116. Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr.*, **A64**, 112–122.
117. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. i Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
118. Dauter, Z. (2003) Twinned crystals and anomalous phasing. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 2004–2016.

119. Padilla, J.E. i Yeates, T.O. (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 1124–1130.
120. Dickerson, R.E. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.*, **17**, 1797–1803.
121. Leontis, N.B. i Westhof, E. (1998) Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.*, **31**, 399–455.
122. Otwinowski, Z.M., W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–325.
123. Vagin, A. i Teplyakov, A. (1997) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.
124. Kissinger, C.R., Gehlhaar, D.K. i Fogel, D.B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 484–491.
125. Painter, J. i Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 439–450.
126. Lu, X.J. i Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
127. Auffinger, P., Bielecki, L. i Westhof, E. (2004) Anion binding to nucleic acids. *Structure*, **12**, 379–388.
128. Sheldrick, G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 479–485.
129. Cowtan, K. (1994) An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.
130. Schuttelkopf, A.W. i van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 1355–1363.
131. Bluckburn, G., Gait, M. J., Loakes, D. i Williams, D. M. (2006) *Nucleic acids in Chemistry and Biology*, RSC Publishing, Cambridge, Anglia.
132. Saikia, M., Dai, Q., Decatur, W.A., Fournier, M.J., Piccirilli, J.A. i Pan, T. (2006) A systematic, ligation-based approach to study RNA modifications. *RNA*, **12**, 2025–2033.
133. Hartmann R.K., Bindereif A., Schön A. i Westhof E. (2005) *Handbook of RNA Biochemistry*. Wiley-VCH, Weinheim, Niemcy.
134. Wang, Q.S. i Unrau, P.J. (2002) Purification of histidine-tagged T4 RNA ligase from *E. coli*. *Biotechniques*, **33**, 1256–1260.
135. Paredes, E. i Das, S.R. Click chemistry for rapid labeling and ligation of RNA. *ChemBiochem.*, **12**, 125–131.
136. Arambula, J.F., Ramisetty, S.R., Baranger, A.M. i Zimmerman, S.C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl. Acad. Sci. U S A*, **106**, 16068–16073.
137. Lee, M.M., Childs-Disney, J.L., Pushechnikov, A., French, J.M., Sobczak, K., Thornton, C.A. i Disney, M.D. (2009) Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: targeting the RNAs that cause myotonic muscular dystrophy. *J. Am. Chem. Soc.*, **131**, 17464–17472.
138. Pushechnikov, A., Lee, M.M., Childs-Disney, J.L., Sobczak, K., French, J.M., Thornton, C.A. i Disney, M.D. (2009) Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease:

- application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J. Am. Chem. Soc.*, **131**, 9767–9779.
139. Popenda, M., Szachniuk, M., Blazewicz, M., Wasik, S., Burke, E.K., Blazewicz, J. i Adamiak, R.W. (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinformatics*, **11**, 231.
140. Saenger, W. (1984) Principles of Nucleic Acid Structure. *Springer-Verlag*, New York.
141. Auffinger, P. i Hashem, Y. (2007) SwS: a solvation web service for nucleic acids. *Bioinformatics*, **23**, 1035–1037.
142. Zheng, M., Huang, X., Smith, G.K., Yang, X. i Gao, X. (1996) Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J. Mol. Biol.*, **264**, 323–336.
143. Rypniewski, W., Adamiak, D.A., Milecki, J. i Adamiak, R.W. (2008) Noncanonical G(syn)-G(anti) base pairs stabilized by sulphate anions in two X-ray structures of the (GUGGUCUGAUGAGGCC) RNA duplex. *RNA*, **14**, 1845–1851.
144. Schlunzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A. i Franceschi, F. (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, **413**, 814–821.
145. Ioele, M., Bazzanini, R., Chatgililoglu, C. i Mulazzani, Q.G. (2000) Chemical Radiation Studies of 8-Bromoguanosine in Aqueous Solutions. *J. Am. Chem. Soc.*, **122**, 1900–1907.
146. Garman, E. (1999) Cool data: quantity AND quality. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 1641–1653.
147. French, G.S. i Wilson, K.S. (1978) On the treatment of negative intensity observations. *Acta Cryst.*, **A34**, 517–525.
148. The CCP4 suite: programs for protein crystallography. (1994) *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
149. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. i Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
150. Patterson, A.L. (1935) A direct method for the determination of the components of interatomic distances in crystals. *Z. Krist.*, (**A**)**90**, 517–542.
151. Tronrud, D.E. (2004) Introduction to macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2156–2168.
152. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A.T. i Berman, H.M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 57–64.
153. Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, **A32**, 923.
154. Jo, S., Kim, T., Iyer, V.G. i Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.*, **29**, 1859–1865.
155. DeLano, W.L. (2002) The PyMOL Molecular Graphics System. *DeLano Scientific*, Palo Alto, CA, USA.
156. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. i Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
157. Dickerson, R.E. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dyn.*, **6**, 627–634.