

**STRUKTURA A PATOGENEZA CHORÓB
ZWIĄZANYCH Z EKSPANSJĄ POWTÓRZEŃ CNG**
STRUCTURE AND PATHOGENESIS OF DISORDERS
RELATED TO CNG REPEATS

Agnieszka Kiliszek, Wojciech Rypniewski*

*Instytut Chemii Bioorganicznej Polskiej Akademii Nauk
ul. Noskowskiego 12/14, 61-704 Poznań
e-mail: wojtekr@ibch.poznan.pl

Abstract

Wykaz stosowanych symboli i oznaczeń

Wprowadzenie

1. Struktura drugorzędowa RNA wydłużonych powtórzeń CNG
2. Powtórzenia CUG
 - 2.1. Struktury krystaliczne RNA powtórzeń CUG
3. Powtórzenia CGG
 - 3.1. Struktury krystaliczne RNA powtórzeń CGG
4. Powtórzenia CAG
 - 4.1. Struktury krystaliczne RNA powtórzeń CAG
5. Powtórzenia CCG
 - 5.1. Struktury krystaliczne RNA powtórzeń CCG
6. Implikacje badań krystalograficznych powtórzeń CNG
 - 6.1. Ogólna struktura powtórzeń CNG
 - 6.2. Oddziaływania par N-N i ich wpływ na strukturę RNA
 - 6.3. Oddziaływania z rozpuszczalnikiem
 - 6.4. Aspekt biologiczny

Uwagi końcowe

Podziękowanie

Piśmiennictwo cytowane

Dr Agnieszka Kiliszek jest absolwentką Wydziału Chemii Uniwersytetu im. Adama Mickiewicza (2006). Studia doktoranckie odbyła w Instytucie Chemii Bioorganicznej PAN w Poznaniu, pod kierunkiem prof. Wojciecha Rypniewskiego. W 2011 roku obroniła z wyróżnieniem rozprawę doktorską pt. „Badania strukturalne powtórzeń CNG w sekwencjach RNA o znaczeniu w patogenezie TREDs”. Rozprawa została nagrodzona przez Premiera RP. Obecnie kontynuuje pracę jako adiunkt w IChB PAN. Jej zainteresowania obejmują badania krystalograficzne patogennych cząsteczek RNA, ich kompleksów z białkami i potencjalnymi terapeutykami.

Prof. dr hab. Wojciech Rypniewski ukończył kierunek biofizyka na Uniwersytecie w York (1983), a następnie doktoryzował się z krystalografii białek na Uniwersytecie w Cambridge (1987). Habilitację (2001) i profesurę (2007) w dziedzinie chemia, specjalności biochemia uzyskał w Polsce. Pracował na uniwersytetach stanowych Arizony i Wisconsin oraz w European Molecular Biology Laboratory (EMBL). Od 2002 roku pracuje w Instytucie Chemii Bioorganicznej PAN w Poznaniu, gdzie kieruje Zespołem Struktury i Funkcji Biomolekuł. Jest autorem lub współautorem blisko siedemdziesięciu prac eksperymentalnych, opublikowanych w międzynarodowych czasopismach naukowych. Jego zainteresowania naukowe obejmują badania krystalograficzne zarówno białek jak i kwasów nukleinowych, w szczególności enzymów chitynolitycznych, glikolitycznych, proteolitycznych oraz patogennych cząsteczek RNA.

ABSTRACT

CNG repeats (N stands for one of the four natural nucleotides) are a special class of microsatellite sequences of the human genome. They are most often found in exons, in their coding parts as well as in the 5' or 3' untranslated regions. Characteristic frequencies of their occurrence within the different parts of the genes suggest that they play a functional role. The number of CNG repeats in a block is usually below 30 but it can undergo abnormal expansion leading to the development of one of approximately 20 neurological diseases known as TREDs (Triplet Repeat Expansion Disorders). One model of pathogenesis proposes that the toxic factor is mRNA containing an expanded run of CNG repeats. The anomaly results in aberrant alternative splicing and/or accumulation of the RNA in the cell nucleus, followed by a sequestration of important regulatory proteins and formation of RNA/protein aggregates known as nuclear foci. This is accompanied by a deregulation of vital cellular processes.

In this paper we have focused on crystallographic studies of RNA oligomers with embedded CNG repeats. We describe briefly diseases associated with each type of repeat and present the crystal structures. All the CNG repeats form stable "hairpins" consisting of a small apical loop and a long double-stranded stem, in which the non-canonical N-N pairs are flanked by the standard C-G and G-C pairs. All CNG repeats form duplexes of type A, characteristic of RNA, but with local deviations from the typical geometry (Fig. 1). The duplexes are stabilised by the strong C-G and G-C Watson-Crick interactions, while the N-N pairs are accommodated within the helical context, each in a characteristic way (Fig. 2). The U-U pairs tend to form just one hydrogen bond, instead of two observed in other contexts. The interactions within the C-C pairs are even weaker, via one very weak hydrogen bond or none. On the other hand, accommodation of the bulky A-A pairs involves pushing the purine rings towards the major groove while in the G-G pairs one of the guanosine residues flips to a syn conformation. The unrealised hydrogen-bonding potential of the N-N pairs is externalised into the major and the minor grooves and can be assessed through interactions with ordered water molecules and other small ligands. The N-N pairs are associated with local distortions of the A-helix (Fig. 1). All the CNG repeats show a characteristic striped pattern of surface electrostatic potential in the minor groove (Fig. 3). Assessment of the different CNG structures allows us to identify the characteristic and the common features (Tab. 1).

Keywords: RNA crystal structure, TREDs, CNG repeats

Słowa kluczowe: struktura krystalograficzna RNA, TREDs, powtórzenia CNG

WYKAZ STOSOWANYCH SKRÓTÓW

A	– reszta adenozyiny
C	– reszta cytydyny
DM	– dystrofia miotoniczna (ang. <i>myotonic dystrophy</i>)
FRAXE	– bezsyndromowe opóźnienie umysłowe powiązane z chromosomem X (ang. <i>non-syndromic X-linked mental retardation</i>)
FXTAS	– zespół drżenia i ataksji związany z łamliwym chromosomem X (ang. <i>fragile X-associated tremor ataxia syndrome</i>)
FXS	– syndrom łamliwego chromosomu X
G	– reszta gwanozyny
HD	– choroba Huntingtona
HDL2	– choroba podobna do płasawicy Huntingtona (ang. <i>Huntington disease like 2</i>)
LNA	– reszty nukleozydowe o zamkniętej konformacji pierścienia cukrowego (ang. <i>locked nucleic acid</i>)
miRNA (mikroRNA)	– jednoniciowe cząsteczki RNA o długości ok. 21–23 nukleotydów, regulujące ekspresję innych genów
ORF	– otwarta ramka odczytu
PDB	– baza danych zawierająca współrzędne atomowe i czynniki struktury makrocząsteczek biologicznych (ang. <i>Protein Data Bank, www.rcsb.org</i>)
RNA	– kwas rybonukleinowy
SBMA	– rdzeniowo-opuszkowy zanik mięśni (choroba Kennedy'ego)
SCA	– ataksja rdzeniowo-mózdkowa
TNR	– powtórzenia trójnukleotydydowe
TREDS	– choroby związane z ekspansją trójnukleotydydowych powtórzeń (ang. <i>trinucleotide repeat expansion diseases</i>)
U	– reszta urydyny
UTR	– region nie ulegający translacji (ang. <i>untranslated region</i>)

WPROWADZENIE

Powtórzenia trinukleotydydowe (TNR) (ang. *trinucleotide repeats*) są specyficzną klasą sekwencji mikrosatelitarnych obejmującą około 32000 sekwencji repetytywnych ludzkiego genomu. Niektóre z nich występują w egzonach, zarówno w obrębie otwartych ramek odczytu, jak i w regionach niekodujących 5' i 3' (5', 3'UTR). Z puli wszystkich możliwych kombinacji powtórzeń trinukleotydydowych najczęściej w egzonach spotykane są powtórzenia typu CNG (gdzie N oznacza jeden z czterech nukleotydów) [1]. Ich nadreprezentacja w egzonach w stosunku do innych części genomu sugeruje, że lokalizacja ta nie jest przypadkowa [1]. Co więcej, wydaje się, że umiejscowienie danego typu powtórzenia w mRNA jest również określone. Powtórzenia CAG i CUG występują najczęściej w obrębie części kodującej (około 80%) natomiast CCG i CGG w regionach 5'UTR (powyżej 50%). Na tej podstawie postuluje się, że są one funkcjonalnymi elementami genetycznymi, których polimorfizm może manifestować się różnymi fenotypami [1, 2].

Powtórzenia CNG występują w wielu różnych genach. Liczba powtórzonych jednostek jest charakterystyczna dla danego genu, przy czym typowo jest mniej niż 30 powtórzeń [3]. Liczba powtórzeń może ulec patogennemu wydłużeniu (50–1000 i więcej powtórzeń), prowadząc do rozwoju nieuleczalnych chorób neurodegeneracyjnych zwanych zbiorowo TREDs (ang. *trinucleotide repeat expansion disorders*) [4]. U zdrowych osób ciągi CNG są krótkie, a ich długość nie ulega zmianie. Natomiast u osób dotkniętych TREDs, liczba powtórzeń jest dłuższa i niestabilna zarówno w komórkach somatycznych jak i rozrodczych [5]. Jak dotąd nie wiadomo co powoduje przejście ze stabilnej do niestabilnej liczby powtórzeń. Być może ma to związek z obecnością i utrzymaniem tzw. sekwencji przerywnikowych [6–9]. W tych przypadkach ciągi powtórzeń CNG są poprzedzielane inną trójką nukleotydów dzielącą je na krótsze bloki. Gdy liczba powtórzeń zostanie zwielokrotniona, ich wydłużanie może odbywać się podczas replikacji, naprawy DNA lub rekombinacji [5, 10]. Co ciekawe, podczas tych procesów możliwe jest również zjawisko kontrakcji czyli skrócenia liczby powtórzeń [10].

TREDs zostały odkryte około 20 lat temu [11–14]. Wówczas po raz pierwszy powiązano zwielokrotnioną liczbę powtórzeń mikrosatelitarnych z objawami chorobowymi. Są to nieuleczalne choroby neurodegeneracyjne, charakteryzujące się bardzo szerokim obrazem klinicznym, który jest specyficzny dla danej jednostki chorobowej. Początkowo objawy są ledwie zauważalne lecz stopniowo pogłębiają się prowadząc w rezultacie do śmierci. Pomimo podobnego defektu genetycznego rozwój choroby może następować w różny sposób. Wyróżnia się trzy główne mechanizmy: nabycie nowej funkcji białka (ang. *protein gain of function*), utrata funkcji genu zawierającego powtórzenia (ang. *gene loss of function*), nabycie nowej funkcji przez mRNA zawierające długie ciągi powtórzeń CNG (ang. mRNA gain of function) [4]. Klasyfikacja ta jest jednak bardzo ogólna i w wielu przypadkach nie wyklucza się współistnienia różnych patomechanizmów lub/oraz zaburzeń powstałych z innych przyczyn.

W niniejszym artykule przedstawimy wyniki badań krystalograficznych w odniesieniu do patogenezы poszczególnych chorób. Zaprezentowane będą głównie struktury krystaliczne cząsteczek RNA zawierających toksyczne powtórzenia CNG. Zostaną omówione ich charakterystyczne cechy strukturalne takie jak forma RNA, sposób oddziaływania niestandardowych par N-N, oddziaływania z ligandami czy rozkład powierzchni potencjału elektrostatycznego. Następnie zostanie przedstawione porównanie powyższych cech oraz przedyskutowane zostaną wnioski z tego implikacje biologiczne.

1. STRUKTURA DRUGORZĘDOWA RNA ZAWIERAJĄCYCH WYDŁUŻONE CIĄGI POWTÓRZEŃ CNG

W mechanizmie nabycia funkcji przez kwasy mRNA zawierające zwielokrotnione ciągi CNG postuluje się, że toksyczny jest RNA [15]. W związku z tym, że powtórzenia CNG są zlokalizowane w obrębie genów, ulegają one transkrypcji. Użycie takich narzędzi biochemicznych do mapowania struktur drugorzędowych jak trawienie nukleazami czy indukowanie cięć łańcucha RNA za pomocą jonów Pb^{2+} wykazało, że fragmenty mRNA zawierające zwielokrotnione ciągi są silnie ustrukturalizowane i niezależnie od typu powtórzenia tworzą strukturę typu spinki [16]. Zbudowana jest ona z długiego trzonu oraz niewielkiej pętli apikalnej. Trzon budują bloki par G-C oraz C-G, pomiędzy którymi znajdują się niestandardowe pary typu N-N (A-A, C-C, G-G lub U-U). Izolowane fragmenty składające się jedynie z ciągów powtórzeń tworzą współistniejące alternatywne spinki różniące się liczbą par w trzonie (tzw. „poślizgnięte” struktury). Dodanie na końcu klamry, czyli sekwencji tworzącej u podstawy trzonu region dwuniciowy, redukuje mikroheterogenność. Naturalne klamry istnieją również w transkryptach niektórych genów np. SCA6 i SCA1 [8, 17, 18].

2. POWTÓRZENIA CUG

Nadmiernie wydłużone trakty powtórzeń CUG są powiązane z trzema różnymi schorzeniami: chorobą podobną do płasawicy Huntingtona (HDL2), ataksją rdzeniowo-mózdkową typu 8 (SCA8) oraz dystrofią miotoniczną typu 1 (DM1) [15]. Patomechanizm został najlepiej poznany w przypadku DM1 i polega on na nabyciu funkcji przez mRNA.

Ekspansję powtórzeń CTG zlokalizowanych w regionie niekodującym 3' genu kinazy białkowej DMPK powiązano z dystrofią miotoniczną typu 1 w 1992 roku [19–21]. Jest to częsta choroba genetyczna występująca raz na 8000 urodzeń, ujawniająca się najczęściej w wieku dojrzałym. Obraz kliniczny jest bardzo szeroki, jednakże typowymi objawami jest miotonia oraz osłabienie wszystkich rodzajów mięśni.

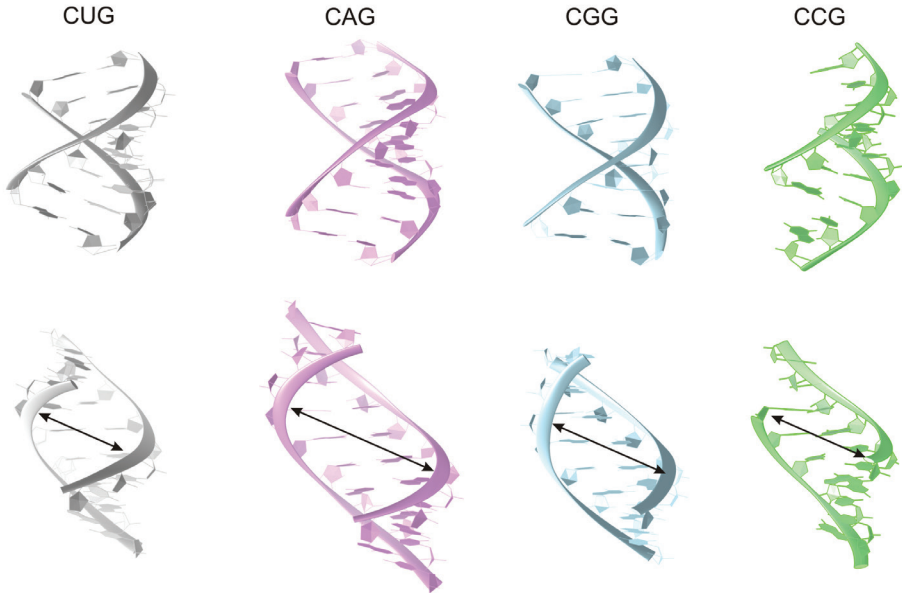
U zdrowych osób liczba powtórzeń CTG wynosi 5–37, natomiast forma patogenna zawiera 50–3000 trójek [20]. Ze względu na ich lokalizację w genie DMPK ulegają one transkrypcji i w mRNA pojawiają się powtórzenia CUG. Fragmenty mRNA zawierające ciągi powtórzeń CUG są toksyczne, gdyż nadmiernie wiążą białka istotne dla prawidłowego funkcjonowania komórki [22]. Wśród nich jest MBNL1 (ang. *muscleblind-like 1 protein*), regulator alternatywnego splicingu transkryptów związanych z rozwojem organizmu [23]. Obniżenie dostępności wolnego białka MBNL1 w komórce powoduje zmianę profilu alternatywnego składania pre-mRNA. Oznacza to, że u chorych powstają naturalne warianty mRNA, lecz nieodpowiednie dla danej tkanki i etapu rozwojowego [24]. Przykładowo, oporność insulinowa u chorych na DM jest wynikiem produkcji wariantu receptora insulinowego (IR) o obniżonej pojemności sygnałnej [15]. 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.

2.1. STRUKTURY KRYSTALICZNE RNA ZAWIERAJĄCEGO CIĄGI POWTÓRZEŃ CUG

Spośród wszystkich rodzajów powtórzeń, najwięcej informacji dostarczyły struktury krystaliczne RNA zawierającego powtórzenia CUG. Do tej pory opublikowano pięć prac dotyczących struktur krystalicznych RNA zawierających te powtórzenia.

Pierwszym doniesieniem o strukturze przestrzennej powtórzeń CNG była praca opublikowana przez Mooers i in. [25]. Autorzy określili strukturę dupleksu zawierającego sześć powtórzeń CUG z rozdzielczością 1,6 Å (kod PDB: 1ZEV). Oligomer o sekwencji (CUG)₆ tworzy dupleks, który przyjmuje formę A-RNA o typowych parametrach helikalnych. Helisa zawiera pary C-G i G-C oraz znajdujące się pomiędzy nimi niestandardowe pary U-U. Obecność niekanonicznych par wydaje się nie mieć wpływu na strukturę helisy, jednak opis szczegółów dotyczących oddziaływań w parze U-U czy obserwacja cząsteczek rozpuszczalnika jest utrudniona ze względu na nałożenie dwóch modeli w sieci krystalicznej.

Kolejną poznaną strukturą był dupleks oligomeru G(CUG)₂C (rozdzielczość 1,23 Å, kod PDB 3G1P) [26]. W modelu tym, dzięki jednoznacznej i wysokorozdzielczej mapie gęstości elektronowej, możliwa jest analiza oddziaływań wewnątrz dupleksu oraz precyzyjne zlokalizowanie cząsteczek rozpuszczalnika. W niezależnej części komórki elementarnej znajdują się trzy dupleksy. Wszystkie są podobne do siebie i przyjmują formę A-RNA, ze średnią wartością kąta skrętu helisy 32–34° (Rys. 1).

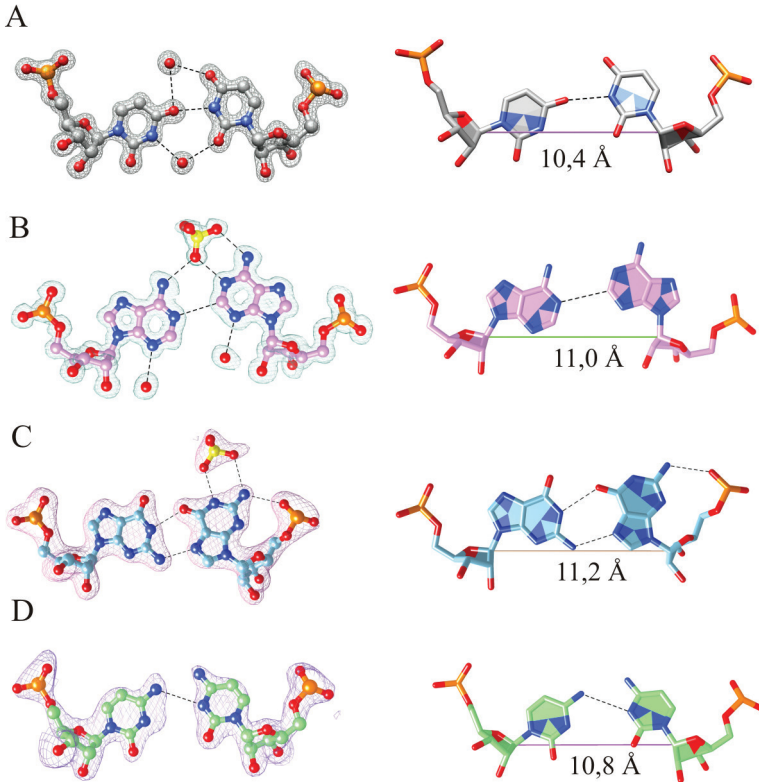


Rysunek 1. Helisy dupleksów każdego z powtórzeń CNG. W górnej części panelu zostały zaprezentowane bruzdy małe a w dolnej części bruzdy duże. Poszerzenie dużej bruzdy oraz rozplecenie helisy jest widoczne dla powtórzeń CAG i CGG

Figure 1. Double helices of each type of CNG repeat. The major grooves are shown in the upper part of the panel, the minor grooves in the lower part. Widening of the major groove and helix unwinding can be seen in CAG and CGG repeats

W parach C-G i G-C tworzą się typowe oddziaływania Watsona-Cricka, podczas gdy w parach U-U zasady oddziałują ze sobą w sposób, który wcześniej nie był obserwowany. W strukturze trójwymiarowej RNA symetria chemiczna niestandardowej pary zostaje złamana i w parze U-U tworzy się jedno wiązanie wodorowe. Jedna z reszt urydyny nachyla się w kierunku małej bruzdy i jej karbonylowy atom O4 łączy się z grupą aminową N3 drugiej reszty (Rys. 2).

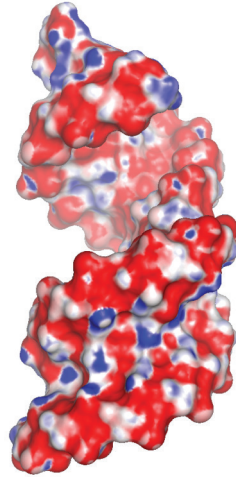
Takie oddziaływanie wygląda na „rozciągnięte” w porównaniu do typowej pary U-U, w której tworzą się dwa wiązania wodorowe [27]. Rozciągnięcie pary U-U odzwierciedlone jest w odległości pomiędzy atomami C1' reszt cukrowych urydyn. W powtórzeniach CUG wynosi ono 10,4 Å, natomiast dla typowej pary U-U odległość ta równa jest 8,6 Å. Dla porównania pomiędzy parami G-C w helisie A-RNA odległość C1'-C1' wynosi 10,6–10,7 Å. Struktura rozciągniętych par U-U jest dodatkowo stabilizowana przez oddziaływania z cząsteczkami wody. Jedna z nich znajduje się w małej bruzdzie, natomiast druga w dużej bruzdzie. Cząsteczki te mogą być traktowane jako część struktury RNA.



Rysunek 2. Niestandardowe pary N-N struktur krystalicznych dupleksów utworzonych przez powtórzenia CNG: U-U (A), A-A (B), G-G (C) oraz C-C (D). Przedstawiono mapy gęstości elektronowej 2Fo-Fc (kontur 1 σ) dla każdej pary oraz oddziałujących z nimi cząsteczek wody lub jonów siarczanowych, na podstawie struktur o następujących kodach PDB: 3GLP, 3NJ6, 3R1C oraz 4E59. Odległość pomiędzy nimi RNA została wyznaczona jako dystans pomiędzy atomami C1' sparowanych reszt (prawa strona)

Figure 2. Non-canonical N-N pairs in crystal structures of duplexes formed by CNG repeats: U-U (A), A-A (B), G-G (C) and C-C (D). Electron density maps 2Fo-Fc (contoured at 1 σ) are shown for each pair together with any associated water molecules or sulphate anions, based on structures indexed under the following PDB codes: 3GLP, 3NJ6, 3R1C and 4E59. The distance between the RNA strands is defined as the distance between the C1' atoms of the paired residues (right)

Dupleksy zawierające powtórzenia CUG mają charakterystyczny motyw rozkładu potencjału elektrostatycznego (Rys. 3).



Rysunek 3. Powierzchnia potencjału elektrostatycznego dwóch kolejnych dupleksów struktury powtórzeń CUG. Rozkład potencjału w bruzdzie małej jest wspólny i charakterystyczny dla wszystkich powtórzeń. Bruzda duża posiada głównie potencjał ujemny (czerwony) z obszarami o potencjale dodatnim (niebieski)

Figure 3. The electrostatic surface potential of two consecutive CUG repeats. The distribution of the potential in the minor groove is characteristic and common to all the repeats. The potential in the major groove is mainly negative (red) with positive patches (blue)

Na powierzchni małej bruzdy, wzdłuż osi helisy, przeplatają się pasy o potencjale dodatnim i ujemnym. Na powierzchni dużej bruzdy rozkład potencjału jest niejednorodny i przeważa potencjał ujemny. Dodatni potencjał generowany jest przez egzo-aminowe grupy reszt cytydyny oraz w mniejszym stopniu przez krawędzie Hoogsteena reszt urydyny.

Kolejną poznaną strukturą RNA zawierającą powtórzenia CUG był dupleks, którego pojedyncza nić zawierała trzy powtórzenia [28]. W tym przypadku oligomer użyty do krystalizacji, oprócz sekwencji repetytywnej posiadał dodatkowe reszty nukleotydowe (pięć reszt na końcu 5' oraz trzy reszty na końcu 3'). W efekcie na końcu 5' znajdowały się dwie niesparowane reszty urydyny (ang. *dangling ends*). Dla tego dupleksu otrzymano dwie formy krystaliczne różniące się upakowaniem cząsteczek w sieci krystalicznej (kod PDB 3SZX, 3SYW). W pierwszej z nich dwie reszty na końcu 5' utworzyły pary U-U z symetryczną cząsteczką RNA, natomiast w drugiej formie te reszty skierowane są do wnętrza helisy i umiejscowione w dużej bruzdzie. W każdej formie krystalicznej w cząsteczce RNA występują trzy niestandardowe pary U-U i środkowa para jest zawsze taka sama. Reszty urydyny ułożone są symetrycznie naprzeciw siebie, prawdopodobnie zbyt daleko, aby tworzyć wiązania wodorowe. W przypadku zewnętrznych par U-U, w dupleksie w którym końcowe niesparowane reszty urydyny są zawinięte do wnętrza helisy, ułożone są one w dużej bruzdzie w bliskim sąsiedztwie pary U-U. Szerokość dużej bruzdy w tym miejscu ulega powiększeniu, a odległość C1'-C1' jest na tyle mała (8,8 Å), że w parze U-U powstają dwa wiązania między zasadami. W dupleksie z drugiej formy

krystalicznej, jedna niestandardowa para U-U tworzy jedno wiązanie wodorowe, druga oddziałuje bardzo słabo (odległość 3,6Å).

Kolejne badania strukturalne dotyczyły dupleksu oligomeru $G(CUG)_6C$ [29]. W kryształach znajdują się dwa symetrycznie niezależne duplekisy, przy czym jeden z nich ma symetrię osi dwukrotnej, a więc symetrycznie niezależna jest jego pojedyncza nić (kod PDB 4E48). W modelu struktury $G(CUG)_6C$ znajduje się sumarycznie dziewięć niezależnych par U-U, co pozwala na obserwację rozkładu konformacyjnego pary niestandardowej. Dominuje konformacja rozciągniętej pary U-U (siedem przypadków), której towarzyszą dwie semi-konserwatywne cząsteczki wody (cząsteczki te nie są obserwowane dla wszystkich par U-U). W pozostałych dwóch przypadkach reszty przyjmują konformację opisaną przez autorów jako symetryczne nieoddziałujące parowanie U-U (ang. *symmetric H-nonbonded U-U pairing*).

Najnowsza praca o powtórzeniach CUG opisuje strukturę syntetycznej spinki RNA zawierającej pętlę GAAA w tandemie z jej receptorem. Elementy te tworzą wierzchołek i górną część spinki, zaś dolną część trzonu spinki stanowią dwa dwuniciowe powtórzenia CUG (PDB code 4FNJ)[30]. Kombinacja pętli apikalnej z sekwencją receptorową ułatwiła krystalizację, a to z kolei pozwoliło na opis doczepionych dwóch powtórzeń CUG. Część trzonu zbudowana z powtórzeń przyjmuje formę A-RNA. Jedna z niestandardowych par tworzy jedno wiązanie wodorowe, podczas gdy druga, znajdująca się bliżej podstawy trzonu, przyjmuje konformację typowej pary U-U z dwoma wiązaniami wodorowymi. Na tej podstawie autorzy postulują, że pary te są dynamiczne.

Uważana analiza map gęstości elektronowych oraz kontekst sekwencyjny muszą być brane pod uwagę podczas interpretacji uzyskanych danych strukturalnych. Dodanie „nienaturalnych” sekwencji z jednej strony ułatwia krystalizację, ale może mieć znaczny wpływ na strukturę przestrzenną RNA. Wówczas uzyskany rezultat może być obciążony błędem w wyniku obecności artefaktu.

W świetle wszystkich określonych struktur krystalicznych, powtórzenia CUG, pomimo obecności niestandardowych par, tworzą helisę A-RNA. Pary U-U otoczone kanonicznymi parami G-C i ZC-G zachowują odległość $C1'-C1'$ charakterystyczną dla formy A-RNA. Obserwuje się pewną dowolność w tworzeniu wiązań wodorowych, jednakże z tendencją do tworzenia jednego wiązania $N3-H\cdots O4$ w parze. W odniesieniu do struktur krystalicznych trudno jest jednoznacznie określić stopień dynamiczności danej pary zasad. Dlatego w przypadku niestandardowych par U-U można raczej stwierdzić, że wykazują one pewną dowolność do tworzenia wiązań wodorowych wynikającą z parametrów termodynamicznych danej konformacji niż, że ich oddziaływanie ma charakter dynamiczny.

3. POWTÓRZENIA CGG

Powtórzenia CGG występują w regionie niekodującym 5' genu FMR1 (ang. *fragile X mental retardation*) leżącym na chromosomie X. Ich występowanie związane jest z kilkoma różnymi jednostkami chorobowymi [31]. W zdrowej populacji, liczba powtórzeń trójek CGG w genie FMR1 wynosi 5–54 [31, 32], przy czym jej górny zakres 45–54 jest definiowany jako „szara strefa” (ang. *grey zone*). Istnieje wówczas większe prawdopodobieństwo dalszej ekspansji powtórzeń [31, 33]. Trakty CGG o długości 55–200 powtórzeń są premutacjami, których obecność u mężczyzn powoduje progresywną chorobę neurodegeneracyjną nazywaną zespołem drżenia i ataksji związanej z łamliwym chromosomem X (FXTAS) [34, 35]. U kobiet będących nosicielkami premutacji może dojść do rozwoju niewydolności jajników [36]. Obecność więcej niż 200 powtórzeń CGG jest określana jako pełna mutacja i stanowi przyczynę łamliwości chromosomu X (FXS) [37]. Jest to najczęstszy dziedziczny syndrom opóźnienia umysłowego u mężczyzn.

Pomimo faktu, że obie choroby dotyczą tych samych powtórzeń CGG genu FMR1, mają one inne podłoże molekularne oraz różny obraz kliniczny. Łamliwość chromosomu X spowodowana jest brakiem ekspresji białka FMRP, które normalnie syntetyzowane jest na wysokim poziomie w neuronach [38–40]. W przypadku FXTAS rozwój choroby następuje na poziomie RNA [35, 40–42]. Podobnie jak w przypadku dystrofii miotonicznej typu 1 u chorych pacjentów obserwowane są złogi jądrowe, w skład których, oprócz patogennego RNA, wchodzi ponad 20 różnych białek [43]. Jeden z proponowanych patomechanizmów FXTAS zakłada, że dochodzi do zaburzenia ogólnej równowagi i dostępności białek regulatorowych wiążących RNA. Natomiast inny opiera się na tym, że kluczowym momentem rozpoczęcia choroby jest związanie do ciągów CGG białka Sam68, które jest zaangażowane w proces alternatywnego składania pre-mRNA [44]. Dochodzi wówczas do zmiany profilu tego procesu. Niewykluczone, że białko Sam68 indukuje tworzenie rozległych złogów jądrowych także poprzez oddziaływanie z innymi białkami.

Trzeci, najnowszy, proponowany mechanizm FXTAS związany jest z sekwestracją przez wydłużone ciągi powtórzeń CGG białka DGCR8 oraz współdziałającej z nim RNazy DROSHA [45]. Białka te zaangażowane są w proces dojrzewania miRNA. Takie RNA transkrybowane są przez II polimerazę RNA jako początkowy pri-miRNA. W jądrze komórkowym transkrypty te przekształcane są przez kompleks enzymatyczny DROSHA i DGCR8 do pre-miRNA. Następnie pre-miRNA transportowane są do cytoplazmy, gdzie ostatecznie przekształcane są do formy dojrzałej (miRNA) przez enzym DICER. U chorych na FXTAS białko DGCR8, wiążące dwuniciowy RNA, rozpoznaje długi dwuniciowy trzon spinki powtórzeń CGG. Tworzy się kompleks DROSHA-DGCR8, który ulega sekwestracji. W efekcie zmniejsza się ilość niezwiązanego kompleksu a w konsekwencji również ilości dojrzałego miRNA. Ostatecznie prowadzi to do dysfunkcji i degeneracji komórek nerwowych.

3.1. STRUKTURY KRystaliczne RNA Zawierającego Powtórzenia CGG

Z dwóch prac na temat struktur krystalicznych RNA zawierającego powtórzenia CGG, pierwsza opisuje trzy struktury: natywny dupleks $(GCGGCGGC)_2$ oraz dwa dupleksy $[GC(8-BrG)GCGGC]_2$ i $[GC(8-BrG)GCGGCGGC]_2$ zawierające modyfikowaną resztę guanozyny 8-BrG, która w pozycji 8 puryny zamiast atomu wodoru ma atom bromu [46]. W części asymetrycznej komórki elementarnej natywnej struktury znajduje się aż 18 niezależnych cząsteczek. Są one uporządkowane w przestrzeni w taki sposób, że można je ułożyć w jedną pseudonieskończoną helisę. We wszystkich trzech strukturach powtórzenia CGG tworzą helisę A-RNA, która posiada pewne lokalne odkształcenia (Rys. 1). Pomiędzy kanonicznymi parami G-C i C-G znajdują się pary G-G, w których jedna guanozyna jest zawsze w konformacji *syn* natomiast druga w *anti* (Rys. 2). W przypadku modyfikowanej reszty 8-BrG wprowadzony atom bromu wymusza konformację *syn*. Powoduje to narzucenie danego ułożenia pary G-G oraz zmniejsza swobodę konformacyjną modyfikowanej reszty. We wszystkich parach G-G występują dwa wiązania wodorowe, jedno pomiędzy tlenem karbonylowym a N1 oraz drugie między atomem N7 a grupą *egzo*-aminową. Dodatkowo konformacja G(*syn*) stabilizowana jest przez wewnątrzcząsteczkowe oddziaływanie grupy *egzo*-aminowej z atomem tlenu grupy fosforanowej. Kąty torsyjne α (określa rotację wokół wiązania P-O5') i γ (określa rotację wokół wiązania C5'-C4') reszty G(*syn*) są nietypowe, co powoduje lokalne rozplecenie helisy, które jest kompensowane w innym miejscu dupleksu. Niekanoniczne pary G-G charakteryzują się specyficznym systemem solwatacji oraz zdolnością do przyciągania jonów, zwłaszcza przez eksponowaną do dużej bruzdy krawędź Watsona-Cricka. Jonami oddziałującymi bezpośrednio z resztami guanozyny są aniony siarczanowe oraz kationy wapnia, które były obecne w medium krystalizacyjnym. Prezentowane przez pary G-G ułożenie *syn-anti* wydaje się być optymalne dla uniknięcia kolizji grup funkcyjnych zlokalizowanych na krawędziach Watsona-Cricka oraz dla zachowania odległości C1'-C1' typowej dla helisy A-RNA, która jest nieco dłuższa (średnio 11,3 Å) niż dla par kanonicznych C-G i G-C.

Powierzchnia potencjału elektrostatycznego małej bruzdy jest podobna do powtórzeń CUG, z charakterystycznymi przeplatającymi się pasami o potencjale dodatnim i ujemnym. W dużej bruzdzie potencjał jest głównie ujemny z obszarem o potencjale dodatnim generowanym przez eksponowaną krawędź Watsona-Cricka reszty G(*syn*) oraz poprzedzającą ją grupę *egzo*-aminową reszty cytydyny.

Druga praca o strukturze traktów CGG opisuje dupleks zawierający trzy takie powtórzenia (kod PDB 3SJ2) [47]. Użyty oligomer, tak samo jak w przypadku powtórzeń CUG, zawiera dodatkowe sekwencje flankujące z niesparowanymi dwiema resztami urydyny na końcu 5'. Struktura ta jest bardzo podobna do tej opisaną przez Kiliszek i in. (2011). Obserwowane pary G(*syn*)-G(*anti*) są takie same i wykazują nietypową konformację szkieletu fosforanowo-cukrowego reszt G(*syn*) oraz rozplecenie lokalnej helisy. Autorzy nie zaobserwowali jednak żadnych jonów w pobliżu par G-G mimo, że zarówno aniony siarczanowe i kationy były obecne w roztworze krystalizacyjnym.

4. POWTÓRZENIA CAG

Największa liczba chorób TREDs związana jest z ekspansją powtórzeń CAG [4]. Należą do nich m.in. choroba Huntingtona (HD), różnego rodzaju ataksje rdzeniowo-mózdkowe (SCA) czy zależny od chromosomu X rdzeniowo-opuszkowy zanik rdzeniowy (choroba Kennedy'ego – SBMA). Ze względu na fakt, że powtórzenia CAG zlokalizowane są w części kodującej, ulegają one translacji na sekwencję aminokwasową białka. Kodon CAG odpowiada glutaminie, dlatego białko powstające ze zmutowanego genu posiada wydłużone trakty glutaminowe. Z tego powodu choroby związane z powtórzeniami CAG nazywane są poliglutaminowymi (ang. *poly-Q lub polyglutamine diseases*) [4, 48]. W chorobach tych główną przyczyną jest niepoprawnie sfałdowane białko. W ostatnich latach pojawiły się jednak doniesienia, iż w patogenezie tych chorób możliwy jest również udział zmutowanego mRNA, co przyczynia się do zwiększenia intensywności zaistniałych procesów degeneracyjnych [49]. Wykazano, że RNA zawierające wydłużone trakty CAG mają zdolność wiązania białek, w tym MBNL1 [50, 51]. W przypadku powtórzeń CAG nie zauważono zaburzeń w procesie składania pre-mRNA, jednakże wprowadzenie ich do regionu nieulegającego translacji genu SCA3 u transgenicznej muszki owocowej powodowało degenerację omatidów i dysfunkcje neuronalne [52]. Co ciekawe, zmutowane powtórzenia CAG tworzą złogi jądrowe, które kolokalizują z MBNL1, ale ich rozmiar jest mniejszy niż w dystrofii miotonicznej [50, 52, 53].

5.1. STRUKTURY KRystaliczne RNA ZAWIERAJĄCEGO POWTÓRZENIA CAG

Poznano strukturę dwóch form krystalicznych tego samego oligomeru o sekwencji GGCAGCAGCC [54]. Jedną z nich to wysokorozdzielcza struktura (0,95 Å), która mieści w części asymetrycznej komórki elementarnej jedną nić (kod PDB 3NJ6). Dupleks generowany jest przez oś dwukrotną, prostopadłą do helisy RNA. Druga struktura została określona z rozdzielczością 1,9 Å i zawiera trzy niezależne cząsteczki (kod PDB 3NJ7). Dupleksy obu struktur są bardzo podobne do siebie i przyjmują formę A-RNA (Rys. 1). Helisa ta, tak jak w przypadku powtórzeń CGG, ulega rozkręceniu. Obie reszty adenozynowe niekanonicznych par znajdują się naprzeciwko siebie i przyjmują konformację anti. Dwa duże pierścienie purynowe zasad mieszczą się jednak w przestrzeni wyznaczonej przez formę A-RNA (odległość C1'-C1' wynosi 11 Å), ponieważ reszty adenozynowe są wypchnięte w kierunku dużej bruzdy, przy czym jedna z nich jest wypchnięta bardziej (Rys. 2). Wystaje ona ponad inne reszty, przyjmując konformację „kciuk do góry” (ang. *thumbs up*). Pomiędzy tak ułożonymi resztami adenozyny tworzy się jedno słabe wiązanie wodorowe między atomem wodoru węgla C2 oraz atomem azotu N1 (C2-H2...N1). Taki rodzaj parowania nie został wcześniej opisany a wiązanie wodorowe tego typu rzadko są opisywane w przypadku biocząsteczek. Niewysycony potencjał reszt adenozynowych do tworzenia wiązań wodorowych realizowany jest w obszarze dużej

bruzdy w której para ta wiąże anion siarczanowy obecny w medium krystalizacyjnym. Rozkład potencjału elektrostatycznego dla bruzdy małej wygląda podobnie jak w cząsteczkach RNA zawierających powtórzenia CUG i CGG. Bruzda duża wykazuje większe zróżnicowanie w dystrybucji potencjałów. Obszary o potencjale dodatnim generowane są przez grupy egzo-aminowe reszt cytydyn oraz adenozyń.

Kolejną strukturą krystaliczną RNA zawierającego powtórzenia CAG jest oligomer zawierający trzy powtórzenia i identyczną, jak dla powtórzeń CGG i CUG, sekwencję flankującą (kod PDB 4J50) [55]. W środkowej parze A-A obie reszty przyjmują konformację anti. Ułożone są jednak względem siebie symetrycznie, a więc prawdopodobieństwo wystąpienia wiązania C2-H2...N1 jest znikome. Autorzy twierdzą, że w przypadku zewnętrznych niekanonicznych par jedna z reszt jest w konformacji syn a druga w anti. Jednak inspekcja map gęstości elektronowej nie potwierdza jednoznacznie tego opisu, gdyż dla par A-A jest ona stosunkowo niskiej jakości. Zewnętrzne reszty adenozyń, szczególnie te w konformacji syn, są raczej nieuporządkowane i dlatego taka interpretacja autorów wydaje się zbyt jednoznaczna i niepoprawna. Ponadto w dużej bruzdzie, przy środkowej parze A-A widoczny jest duży pozytywny pik gęstości elektronowej. Sugeruje to, że znajduje tam jon siarczanowy, który nie został uwzględniony w modelu struktury.

5. POWTÓRZENIA CCG

Powtórzenia CCG są nadreprezentowane w egzonach. Zlokalizowane są zarówno w regionach 5'UTR jak i w otwartych ramkach odczytu [1]. Ich rola w patogenezie chorób TREDs nie jest do końca poznana w porównaniu do pozostałych rodzajów powtórzeń. Powtórzenia CCG powiązано dotychczas z trzema chorobami: dystrofią miotoniczną typu 1 [56], chorobą Huntingtona [57] oraz bezsyndromowym opóźnieniem umysłowym powiązanym z chromosomem X (FRAXE) [58]. W przypadku ostatniego syndromu ujawnia się on, gdy osoba posiada więcej niż 200 powtórzeń CCG w genie FMR2, który znajduje się około 6000 par zasad od genu FMR1 [4, 59]. Jak dotąd nie stwierdzono fenotypu neurodegeneracyjnego charakterystycznego dla powtórzeń CCG o długości premutacyjnej (60–200), ale powtórzenia te wprowadzone do genomu muszki owocowej ujawniły właściwości toksyczne [60]. Zauważono także pewną korelację pomiędzy długością powtórzeń CCG a łagodnymi opóźnieniami umysłowymi czy chorobą Parkinsona lub przynajmniej niektórymi jej objawami [61, 62]. Z tego powodu nie wyklucza się istnienia, analogicznie do powtórzeń CGG, neurodegeneracyjnych właściwości powtórzeń CCG o długości premutacyjnej.

5.1. STRUKTURY KRystaliczne RNA Zawierającego Powtórzenia CCG

Dotychczas została opublikowana jedna praca dotycząca struktury krystalicznej powtórzeń CCG [63]. Dotyczy ona dwóch oligomerów tworzących dupleksy: $(GCCGCCGC)_2$ (kod PDB 4E59) oraz $(GCCGLCCGC)_2$ (kod PDB 4E58), gdzie reszta (GL) jest modyfikowaną resztą guanozyny typu LNA. Obydwa dupleksy przyjmują formę A-RNA (Rys. 1) lecz parowanie się nici jest zaskakujące. W strukturze niemodyfikowanej nici oligomeru są przesunięte („poślizgnięte”) względem siebie w kierunku 5'. Oznacza to, że końce 5' helisy mają wystające reszty. W strukturze modyfikowanej obserwowany jest odwrotny efekt: ślizg nici nastąpił w kierunku 3'. W obu przypadkach spowodowało to redukcję liczby spodziewanych niestandardowych par C-C z dwóch do jednej. Sumarycznie obserwowane są trzy niezależne niekanoniczne pary: jedna w strukturze natywnej i dwie w strukturze zawierającej modyfikacje. W dwóch parach C-C tworzą się wiązania wodorowe (Rys. 2), natomiast w trzeciej takiego wiązania brak. Niestandardowe pary C-C nie zwiększają stabilności dupleksu, co objawia się tym, że układ dąży do ich eliminowania przez ślizg nici. Rezultaty te są zgodne z wynikami termodynamicznymi potwierdzając stosunkowo niską stabilność powtórzeń CCG. Wprowadzenie modyfikacji LNA wydaje się nie mieć widocznego efektu na parametry helikalne w porównaniu do natywnej struktury jednakże obserwowane jest zwiększenie oddziaływań warstwowych.

6. IMPLIKACJE BADAŃ KRystalograficznych Powtórzeń CNG

Końcowym efektem określenia struktur przestrzennych powtórzeń CNG jest wskazanie ich wspólnych, jak i unikalnych cech oraz odniesienie do aspektu biologicznego i medycznego. Szczegółowe zestawienie podobieństw i różnic zostało przedstawione w Tabeli 1.

Tabela 1. Profil powtórzeń CNG – zestawienie cech strukturalnych
Table 1. Structural profile of CNG repeats – a collation of characteristic features

Rodzaj powtórzeń	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

Rodzaj powtórzenia	CGG	CAG	CUG	CCG
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 wewnętrzznukleotydowe: <i>exo</i> -N2H...O2	C2-H2...N1	N3-H3...O4	Brak jednego wspólnego wzoru parowania się 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 się 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 siarczynowy związany w bruzdzie dużej lub 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

6.1. OGÓLNA STRUKTURA POWTÓRZEŃ CNG

Najbardziej charakterystyczną cechą wszystkich struktur utworzonych z powtórzeń trinukleotydowych jest tworzenie helisy A-RNA (Rys. 1). Stabilizowana jest ona przez kanoniczne pary C-G i G-C, które stanowią swego rodzaju rusztowanie. Z kolei niestandardowe pary N-N są elementem różnicującym poszczególne struktury. Każda para N-N wykazuje swój specyficzny sposób dopasowania w utworzonej strukturze, przy czym zakres tej akomodacji jest ograniczony do zachowania określonej odległości pomiędzy sparowanymi łańcuchami. W tym przypadku jest to odległość 10,5–10,8 Å pomiędzy atomami C1' par zasad sparowanych reszt nukleotydowych (Rys. 2). W efekcie na całej długości helisy utrzymana jest forma dwuniciowa niezależnie czy tworzą ją pary kanoniczne, czy niekanoniczne. Istotne jest również to, że nie obserwuje się wybrzuszeń ani znacznych deformacji łańcucha fosforanowo-cukrowego.

6.2. ODDZIAŁYWANIA PAR N-N I ICH WPŁYW NA STRUKTURĘ RNA

Specyfiką dopasowania par N-N w strukturze jest to, że w odróżnieniu od par kanonicznych, gdzie zasada purynowa tworzy parę z pirymidyną, zawiązują się pary typu puryna-puryna lub pirymidyna-pirymidyna (Rys. 2). Podczas parowania się dwóch zasad purynowych problemem jest ograniczona przestrzeń między niemi RNA. Na przykładzie par G-G oraz A-A obserwuje się dwa różne sposoby dostosowania do zaistniałych warunków. Z kolei pary pirymidynowe mają tej przestrzeni zbyt dużo. W powtórzeniach CUG reszty U-U znajdują się w odległości 10,4 Å. Jest to dystans zbyt duży, aby utworzyć wiązanie wodorowe. W większości obserwowanych przypadków dążą one jednak do tego, aby zmniejszyć tę odległość poprzez nachylenie się jednej z reszt urydyny w kierunku małej bruzdy. Akomodacja par C-C polega na unikaniu parowania się ze sobą reszt cytydyny. Gdy jednak znajdują się one naprzeciw siebie w łańcuchu RNA, mogą tworzyć słabe oddziaływania wodorowe. Niewykluczone, że reszty C-C mogą parować się tak jak reszty U-U. Byłoby to jednak związane z dużą barierą energetyczną bądź steryczną czy też zmianą konformacyjną łańcucha RNA, która jest jednak trudna do pokonania.

Pomimo liczebnej przewagi par G-C i C-G, które kształtują strukturę powtórzeń CNG, obecność pary N-N ma również na nią wpływ. Dla powtórzeń CGG oraz CAG jest on bardzo widoczny: helisy są „rozkręcone” a szerokości ich dużych bruzd są zauważalnie większe (Rys. 1). Związane jest to obecnością w każdej z helis reszty, w której nastąpiła rotacja wokół wiązania O5'-C5'. W efekcie znajduje się ona niemalże w tej samej płaszczyźnie co sąsiadująca z nią po stronie 5' reszta nukleotydowa, zmniejszając kąt skrętu helisy. Dla powtórzeń CGG jest to lokalny efekt, który kompensowany jest w innym miejscu w strukturze, natomiast w przypadku powtórzeń CAG jest to efekt bardziej globalny. Dodatkowo średnia wartość odległości C1'-C1' wyznaczona dla par C-G i G-C jest większa niż dla par U-U. Różnica

ta jest niewielka (wynosi jedynie 0,2 Å), lecz korelacja ta jest logiczna w kontekście wielkości par N-N. Pary A-A oraz G-G mają większe wymagania steryczne i bardziej wpływają na globalną strukturę RNA niż pary U-U.

Obecność par C-C może spowodować, że nici RNA łączą się inaczej niż by to wynikało z analizy sekwencji. Jeżeli jednak rozpatrywać wpływ par C-C na strukturę helisy, to w niemodyfikowanym duplesie nie powodują one znacznych zmian strukturalnych, podobnie jak w przypadku par U-U.

6.3. ODDZIAŁYWANIA Z ROZPUSZCZALNIKIEM

Możliwość obserwacji cząsteczek rozpuszczalnika w strukturach krystalicznych jest niezwykle ważnym atutem krystalografii. Umożliwia lokalizację ligandów i opis oddziaływania z biomolekułami. Na tej podstawie można badać m.in. mechanizmy reakcji, ale również projektować ligandy, które potencjalnie mogą wiązać się do cząsteczki, której struktura określona została metodami krystalograficznymi. To ostatnie podejście stosowane jest przy projektowaniu leków w oparciu o strukturę (ang. *structure-based drug design*).

W strukturach RNA zawierających powtórzenia CNG sposób parowania się par N-N nie wysycza ich potencjału do oddziaływań, dlatego mają one zdolność do przyłączania ligandów (Rys. 2). Największy potencjał wykazują grupy funkcyjne zlokalizowane na krawędzi Watsona-Cricka, czyli tej, która w kanonicznych parach zaangażowana jest w parowanie. Pary U-U mają eksponowane grupy funkcyjne zarówno w kierunku małej bruzdy (dwie grupy karbonylowe O2 oraz amidowa N3) jak i w kierunku dużej bruzdy (grupa O4). W efekcie z parą tą oddziałują dwie cząsteczki wody. Z kolei pary G-G w zależności od tego, co było obecne w roztworze krystalizacyjnym, tworzą wiązania wodorowe z hydratowanym kationem Ca^{2+} lub z jonem siarczanowym. W przypadku braku liganda miejsce wiązania zastępują cząsteczki wody. Anion siarczanowy oddziałuje z krawędzią Watsona-Cricka reszty G(*syn*) i wydaje się być idealnie dopasowany zarówno pod względem chemicznym jak i geometrycznym podobnie jest w przypadku par A-A. Ich grupy funkcyjne są inaczej eksponowane do bruzdy dużej, lecz pomimo tego pomiędzy krawędzie obu reszt wpasował się jon siarczanowy. Miejsca wiązania jonów, we wszystkich przypadkach, bardzo dobrze korelują z rozkładem potencjału elektrostatycznego struktur krystalicznych powtórzeń CNG.

Innym ważnym czynnikiem, który powinien być brany pod uwagę w ocenie zdolności par N-N do przyciągania ligandów, są oddziaływania warstwowe. W kwasach nukleinowych są one bardzo powszechne i, obok wiązań wodorowych, mają swój znaczny udział w tworzeniu struktury przestrzennej. Miarą tego oddziaływania jest wartość powierzchni nakrywania wyrażana w Å². Im większa jest jej wartość tym oddziaływania pomiędzy zasadami są silniejsze. Typowo w regionach helikalnych pierścienie zasad par kanonicznych znajdują się w znacznym stopniu nad sobą. W przypadku powtórzeń CNG reguła ta została potwierdzona, gdy asocjują ze

sobą pary G-C i C-G (powierzchnia nakrywania jest zazwyczaj większa niż 4 \AA^2). W przypadku niestandardowych par N-N oddziaływania warstwowe są znacznie ograniczone.

Wspólną cechą wszystkich par N-N są nieznaczne oddziaływania warstwowe oraz niewysyczone miejsca o wysokim potencjale do tworzenia wiązań. Specyficzność jest generowana przez indywidualny charakter chemiczny oraz przestrzenne ułożenie danej niekanonicznej pary, co decyduje o tym, jaki ligand jest wiązany.

6.4. ASPEKT BIOLOGICZNY

Powtórzenia CNG są unikalną grupą sekwencji mikrosatelitarnych obecnych w mRNA niektórych genów. Występowanie nadmiernej liczby powtórzeń stanowi przyczynę ponad dwudziestu chorób neurodegeneracyjnych. mRNA zmutowanych genów zawierają struktury typu spinki utworzone z powtórzeń CNG. Obserwowane struktury krystaliczne powtórzeń CAG, CGG, CUG i CCG prawdopodobnie odzwierciedlają struktury trzonów patogennych spinek. We wszystkich opisanych strukturach dupleksów, helisy układają się w sieci krystalicznej jedna nad drugą tworząc długie równoległe kolumny. Najdłuższą pseudohelisę zaobserwowano dla oligomeru GCGGCGGC. W części asymetrycznej tej struktury znajduje się aż 18 niezależnych dupleksów. Są one uporządkowane w przestrzeni w taki sposób, że mogą utworzyć jedną pseudonieskończoną helisę składającą się ze wszystkich 18 symetrycznie niezależnych cząsteczek. Pomiędzy sąsiednimi cząsteczkami istnieją oddziaływania warstwowe, dlatego nie obserwuje się efektu rozplatania końców helisy.

Białko MBNL1 jest istotnym czynnikiem w patogenezie dystrofii miotonicznej typu 1. Rozpoznaje zmutowane ciągi powtórzeń CUG, do których nadmiernie się wiąże powodując zmianę profilu alternatywnego składania szeregu pre-mRNA. MBNL1 może również oddziaływać z powtórzeniami CAG i CCG [50, 51, 64]. Pomimo, iż znana jest struktura krystaliczna domen białka MBNL1 oraz jedna z domen w kompleksie z oligomerem RNA, nie wiadomo jak białko to rozpoznaje wydłużone ciągi powtórzeń CNG. Na podstawie sporządzonego profilu strukturalnego można przypuszczać, że kluczowe dla rozpoznania są niestandardowe pary N-N. W przypadku powtórzeń CUG, CAG i CCG w niekanonicznych parach jest maksymalnie jedno wiązanie wodorowe. Białko MBNL1 prawdopodobnie wiąże jednoniciowe RNA. Ciągi CNG tworzą strukturę dwuniciową, którą to białko może rozpleść w przypadku CUG, CAG i CCG i przyłączyć pojedynczą nić. W powtórzeniach CGG, pomiędzy resztami guanozyn występują dwa wiązania wodorowe, zatem ich rozplecenie staje się znacznie trudniejsze. 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 helisa przyjmuje formę A-RNA.

UWAGI KOŃCOWE

Badania krystalograficzne umożliwiły nie tylko poznanie struktury przestrzennej powtórzeń typu CNG, ale również określenie ich wspólnych oraz specyficznych właściwości fizyko-chemicznych takich jak: powinowactwo do ligandów, charakterystyka powierzchni, dystrybucja potencjału elektrostatycznego czy sieć wiązań wodorowych i hydratacja. Uzyskany profil 3D może być użyty dla racjonalizacji obserwacji biochemicznych i cytologicznych a także do racjonalnego projektowania ligandów (ang. *drug design*) wiążących się do powtórzeń CNG [65–67].

PODZIĘKOWANIA

Finansowanie: projekt badawczy Narodowego Centrum Nauki UMO-2011/01/B/NZ1/04429 oraz projekt MNiSZ IUVENTUS PLUS 0450/IPI/2013/72.

PIŚMIENNICTWO CYTOWANE

- [1] P. Kozłowski, M. de Mezer, W.J. Krzyżosiak, *Nucleic Acids Res.*, 2010, **38**, 4027.
- [2] A. Jasińska, G. Michlewski, M. de Mezer, K. Sobczak, P. Kozłowski, M. Napierała, W.J. Krzyżosiak, *Nucleic Acids Res.*, 2003, **31**, 5463.
- [3] G. Sicot, M. Gomes-Pereira, *Biochim. Biophys. Acta*, 2013, **1832**, 1390.
- [4] H.T. Orr, H.Y. Zoghbi, *Annu. Rev. Neurosci.*, 2007, **30**, 575.
- [5] C.E. Pearson, K. Nichol Edamura, J.D. Cleary, *Nat. Rev. Genet.*, 2005, **6**, 729.
- [6] A. Lopez Castel, J.D. Cleary, C.E. Pearson, *Nat. Rev. Cell Biol.*, 2010, **11**, 165.
- [7] M. Napierała, D. Michałowski, M. de Mezer, W.J. Krzyżosiak, *Nucleic Acids Res.*, 2005, **33**, 451.
- [8] K. Sobczak, W.J. Krzyżosiak, *J. Biol. Chem.*, 2004, **279**, 41563.
- [9] K. Sobczak, W.J. Krzyżosiak, *J. Biol. Chem.*, 2005, **280**, 3898.
- [10] S.M. Mirkin, *Nature*, 2007, **447**, 932.
- [11] M.V. Bell, M.C. Hirst, Y. Nakahori, R.N. MacKinnon, A. Roche, T.J. Flint, P.A. Jacobs, N. Tommerup, L. Tranebjaerg, U. Froster-Iskenius i in., *Cell*, 1991, **64**, 861.
- [12] I. Oberle, F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boue, M.F. Bertheas, J.L. Mandel, *Science*, 1991, **252**, 1097.
- [13] A.J. Verkerk, M. Pieretti, J.S. Sutcliffe, Y.H. Fu, D.P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F.P. Zhang i in., *Cell*, 1991, **65**, 905.
- [14] S. Yu, M. Pritchard, E. Kremer, M. Lynch, J. Nancarrow, E. Baker, K. Holman, J.C. Mulley, S.T. Warren, D. Schlessinger i in., *Science*, 1991, **252**, 1179.
- [15] L.P. Ranum, T.A. Cooper, *Annu. Rev. Neurosci.*, 2006, **29**, 259.
- [16] K. Sobczak, M. de Mezer, G. Michlewski, J. Król, W.J. Krzyżosiak, *Nucleic Acids Res.*, 2003, **31**, 5469.
- [17] W.J. Krzyżosiak, K. Sobczak, M. Napierała, R.D. Wells, T. Ashizawa, *Chapter 45 – Structural Characteristics of Trinucleotide Repeats in Transcripts*, Academic Press, Burlington, 2006.
- [18] G. Michlewski, W.J. Krzyżosiak, *J. Mol. Biol.*, 2004, **340**, 665.
- [19] C. Aslanidis, G. Jansen, C. Amemiya, G. Shutler, M. Mahadevan, C. Tsiflidis, C. Chen, J. Alleman, N.G. Wormskamp, M. Vooijs i in., *Nature*, 1992, **355**, 548.

- [20] J.D. Brook, M.E. McCurrach, H.G. Harley, A.J. Buckler, D. Church, H. Aburatani, K. Hunter, V.P. Stanton, J.P. Thirion, T. Hudson, *in in.*, *Cell*, 1992, **68**, 799.
- [21] H.G. Harley, J.D. Brook, S.A. Rundle, S. Crow, W. Reardon, A.J. Buckler, P.S. Harper, D.E. Housman, D.J. Shaw, *Nature*, 1992, **355**, 545.
- [22] J.W. Miller, C.R. Urbinati, P. Teng-Umnuay, M.G. Stenberg, B.J. Byrne, C.A. Thornton, M.S. Swanson, *EMBO J*, 2000, **19**, 4439.
- [23] T.H. Ho, B.N. Charlet, M.G. Poulos, G. Singh, M.S. Swanson, T.A. Cooper, *EMBO*, 2004, **23**, 3103.
- [24] R.J. Osborne, C.A. Thornton, *Hum. Mol. Genet.*, 2006, **15 Spec No 2**, R162.
- [25] B.H. Mooers, J.S. Logue, J.A. Berglund, *Proc. Natl. Acad. Sci. U S A*, 2005, **102**, 16626.
- [26] A. Kiliszek, R. Kierzek, W.J. Krzyżosiak, W. Rypniewski, *Nucleic Acids Res.*, 2009, **37**, 4149.
- [27] P. Auffinger, Y. Hashem, *Bioinformatics*, 2007, **23**, 1035.
- [28] A. Kumar, H. Park, P. Fang, R. Parkesh, M. Guo, K.W. Nettles, M.D. Disney, *Biochemistry*, 2011, **50**, 9928.
- [29] J. Tamjar, E. Katorcha, A. Popov, L. Malinina, *J. Biomol. Struct. Dyn.*, 2012, **30**, 505.
- [30] L.A. Coonrod, J.R. Lohman, J.A. Berglund, *Biochemistry*, 2012, **51**, 8330.
- [31] Y.H. Fu, D.P. Kuhl, A. Pizzuti, M. Pieretti, J.S. Sutcliffe, S. Richards, A.J. Verkerk, J.J. Holden, R.G. Fenwick, Jr., S.T. Warren *in in.*, *Cell*, 1991, **67**, 1047.
- [32] C. Dombrowski, S. Levesque, M.L. Morel, P. Rouillard, K. Morgan, F. Rousseau, *Hum. Mol. Genet.*, 2002, **11**, 371.
- [33] N. Zhong, W. Ju, J. Pietrofesa, D. Wang, C. Dobkin, W.T. Brown, *Am. J. Med. Genet.*, 1996, **64**, 261.
- [34] R.J. Hagerman, M. Leehey, W. Heinrichs, F. Tassone, R. Wilson, J. Hills, J. Grigsby, B. Gage, P.J. Hagerman, *Neurology*, 2001, **57**, 127.
- [35] S. Jacquemont, R.J. Hagerman, M. Leehey, J. Grigsby, L. Zhang, J.A. Brunberg, C. Greco, V. Des Portes, T. Jardini, R. Levine, E. Berry-Kravis, W.T. Brown, S. Schaeffer, J. Kissel, F. Tassone, P.J. Hagerman, *Am. J. Hum. Genet.*, 2003, **72**, 869.
- [36] S.L. Sherman, *Am. J. Med. Genet.*, 2000, **97**, 189.
- [37] I.A. Glass, *J. Med. Genet.*, 1991, **28**, 361.
- [38] L.S. Chen, F. Tassone, P. Sahota, P.J. Hagerman, *Hum. Mol. Genet.*, 2003, **12**, 3067.
- [39] B. Primerano, F. Tassone, R.J. Hagerman, P. Hagerman, F. Amaldi, C. Bagni, *RNA*, 2002, **8**, 1482.
- [40] F. Tassone, R.J. Hagerman, A.K. Taylor, L.W. Gane, T.E. Godfrey, P.J. Hagerman, *Am. J. Hum. Genet.*, 2000, **66**, 6.
- [41] R.J. Hagerman, P.J. Hagerman, *Mol. Genet. Metab.*, 2001, **74**, 89.
- [42] A. Kenneson, F. Zhang, C.H. Hagedorn, S.T. Warren, *Hum. Mol. Genet.*, 2001, **10**, 1449.
- [43] D. Garcia-Arocena, P.J. Hagerman, *Hum. Mol. Genet.*, 2010, **19**, R83.
- [44] C. Sellier, F. Rau, Y. Liu, F. Tassone, R.K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemssen, D.J. Elliott, P.J. Hagerman, N. Charlet-Berguerand, *EMBO*, 2010, **29**, 1248.
- [45] C. Sellier, F. Freyermuth, R. Tabet, T. Tran, F. He, F. Ruffenach, V. Alunni, H. Moine, C. Thibault, A. Page, F. Tassone, R. Willemssen, M.D. Disney, P.J. Hagerman, P.K. Todd, N. Charlet-Berguerand, *Cell Rep.*, 2013, **3**, 869.
- [46] A. Kiliszek, R. Kierzek, W.J. Krzyżosiak, W. Rypniewski, *Nucleic Acids Res.*, 2011, **39**, 7308.
- [47] A. Kumar, P. Fang, H. Park, M. Guo, K.W. Nettles, M.D. Disney, *Chembiochem.*, 2011, **12**, 2140.
- [48] A.R. La Spada, J.P. Taylor, *Nat. Rev. Genet.*, 2010, **11**, 247.
- [49] P. Gałka-Marciniak, M.O. Urbanek, W.J. Krzyżosiak, *Biol. Chem.*, 2012, **393**, 1299.
- [50] T.H. Ho, R.S. Savkur, M.G. Poulos, M.A. Mancini, M.S. Swanson, T.A. Cooper, *J. Cell Sci.*, 2005, **118**, 2923.
- [51] Y. Yuan, S.A. Compton, K. Sobczak, M.G. Stenberg, C.A. Thornton, J.D. Griffith, M.S. Swanson, *Nucleic Acids Res.*, 2007, **35**, 5474.
- [52] L.B. Li, Z. Yu, X. Teng, N.M. Bonini, *Nature*, 2008, **453**, 1107.

- [53] M. de Mezer, M. Wojciechowska, M. Napierała, K. Sobczak, W.J. Krzyżosiak, *Nucleic Acids Res.*, 2011, **39**, 3852.
- [54] A. Kiliszek, R. Kierzek, W.J. Krzyżosiak, W. Rypniewski, *Nucleic Acids Res.*, 2010, **38**, 8370.
- [55] I. Yildirim, H. Park, M.D. Disney, G.C. Schatz, *J. Am. Chem. Soc.*, 2013, **135**, 3528.
- [56] C. Braidia, R.K. Stefanatos, B. Adam, N. Mahajan, H.J. Smeets, F. Niel, C. Goizet, B. Arveiler, M. Koenig, C. Lagier-Tourenne, J.L. Mandel, C.G. Faber, C.E. de Die-Smulders, F. Spaans, D.G. Monckton, *Hum. Mol. Genet.*, 2010, **19**, 1399.
- [57] B.R. Zhang, J. Tian, Y.P. Yan, X.Z. Yin, G.H. Zhao, Z.Y. Wu, W.H. Gu, K. Xia, B.S. Tang, *J. Neurol. Sci.*, 2012, **312**, 92.
- [58] Y. Gu, Y. Shen, R.A. Gibbs, D.L. Nelson, *Nat. Genet.*, 1996, **13**, 109.
- [59] J. Gecz, A.K. Gedeon, G.R. Sutherland, J.C. Mulley, *Nat. Genet.*, 1996, **13**, 105.
- [60] O.A. Sofola, P. Jin, J. Botas, D.L. Nelson, *Hum. Mol. Genet.*, 2007, **16**, 2326.
- [61] G. Annesi, G. Nicoletti, P. Tarantino, N. Cutuli, F. Annesi, E.V. Marco, M. Zappia, L. Morgante, G. Arabia, P. Pugliese, F. Condino, S. Carrideo, D. Civitelli, M. Caracciolo, N. Romeo, P. Spadafora, I.C. Candiano, A. Quattrone, *Neurosci. Lett.*, 2004, **368**, 21.
- [62] A. Costa, L. Gao, F. Carrillo, M.T. Caceres-Redondo, M. Carballo, J. Diaz-Martin, P. Gomez-Garre, F. Sobrino, M. Lucas, J. Lopez-Barneo, P. Mir, E. Pintado, *Parkinsonism Relat. Disord.*, 2011, **17**, 281.
- [63] A. Kiliszek, R. Kierzek, W.J. Krzyżosiak, W. Rypniewski, *Nucleic Acids Res.*, 2012, **40**, 8155.
- [64] Y. Kino, D. Mori, Y. Oma, Y. Takeshita, N. Sasagawa, S. Ishiura, *Hum. Mol. Genet.*, 2004, **13**, 495.
- [65] J.F. Arambula, S.R. Ramisetty, A.M. Baranger, S.C. Zimmerman, *Proc. Natl. Acad. Sci. U S A*, 2009, **106**, 16068.
- [66] M.M. Lee, J.L. Childs-Disney, A. Pushechnikov, J.M. French, K. Sobczak, C.A. Thornton, M.D. Disney, *J. Am. Chem. Soc.*, 2009, **131**, 17464.
- [67] A. Pushechnikov, M.M. Lee, J.L. Childs-Disney, K. Sobczak, J.M. French, C.A. Thornton, M.D. Disney, *J. Am. Chem. Soc.*, 2009, **131**, 9767.

Praca wpłynęła do Redakcji 29 stycznia 2014