

Praca doktorska

Struktura regionu powtórzeń
trójnukleotydowych w mRNA genu *IT15*
związanego z chorobą Huntingtona.

Mateusz de Mezer

Promotor: Prof. dr hab. Włodzimierz J. Krzyżosiak

Pracownia Genetyki Nowotworów
Instytut Chemii Bioorganicznej PAN

Poznań 2005

*Panu Profesorowi Włodzimierzowi Krzyżosiakowi
serdecznie dziękuję
za przekazaną wiedzę,
opiekę i pomoc podczas realizacji pracy.*

*Monice
dziękuję za pomoc, cierpliwość i zrozumienie.*

*Koleżankom i Kolegom z Pracowni Genetyki Nowotworów
dziękuję za pomoc, miłą atmosferę i życzliwość.*

1	WSTEP	7
1.1	WPROWADZENIE.	7
1.2	GENOM CZŁOWIEKA.	8
1.2.1	ORGANIZACJA GENOMU CZŁOWIEKA.	8
1.2.2	LICZBA LUDZKICH GENÓW.	10
1.2.3	ZRÓŻNICOWANIE GENETYCZNE – SNP.	11
1.3	MUTACJE W GENOMIE CZŁOWIEKA.	12
1.4	SEKWENCJE POWTARZAJĄCE SIĘ W GENOMIE CZŁOWIEKA.	13
1.5	TRÓJNUKLEOTYDOWE SEKWENCJE MIKROSATELITARNE.	15
1.6	POWTÓRZENIA TRÓJNUKLEOTYDOWE W LUDZKIM TRANSKRYPTOMIE.	16
1.6.1	LICZBA RÓŻNYCH CIĄGÓW TRÓJNUKLEOTYDOWYCH W LUDZKICH MRNA.	17
1.6.2	DŁUGOŚĆ TRAKTÓW TRÓJNUKLEOTYDOWYCH.	18
1.6.3	WYSTĘPOWANIE POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH W FUNKCJONALNYCH CZĘŚCIACH MRNA.	19
1.6.4	KOMÓRKOWE FUNKCJE GENÓW Z POWTÓRZENIAMI TRÓJNUKLEOTYDOWYMI.	21
1.7	WYSTĘPOWANIE I ZNACZENIE NAJCZĘSTSZYCH MOTYWÓW TRÓJNUKLEOTYDOWYCH OBECNYCH W LUDZKIM TRANSKRYPTOMIE.	23
1.7.1	POWTÓRZENIA CAG.	23
1.7.2	POWTÓRZENIA CGG.	24
1.7.3	POWTÓRZENIA CCG.	26
1.7.4	POWTÓRZENIA CUG.	27
1.8	CHOROBY NEUROLOGICZNE CZŁOWIEKA POWODOWANE EKSPANSJĄ POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH.	28
1.8.1	PATOMECHANIZM CHORÓB Z GRUPY TREDs ZWIĄZANYCH Z UTRATĄ FUNKCJI PRZEZ BIAŁKO.	31
1.8.2	PATOMECHANIZM CHORÓB POLIGLUTAMINOWYCH NA PRZYKŁADZIE CHOROBY HUNTINGTONA.	34
1.8.3	PATOMECHANIZM ZAKŁADAJĄCY ROLĘ RNA ZAWIERAJĄCEGO WYDŁUŻONE CIĄGI POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH.	38

<u>2</u>	<u>CEL PRACY</u>	43
<u>3</u>	<u>WYNIKI</u>	44
3.1	POWTÓRZENIA CAG I CCG TWORZĄ STRUKTURĘ TYPU SPINKI.	44
3.2	POLIMORFIZM LICZBY POWTÓRZEŃ CAG I CCG W GENIE <i>IT15</i> ZWIĄZANYM Z CHOROBA HUNTINGTONA.	50
3.3	POWTÓRZENIA CAG WE FRAGMENTCIE mRNA GENU <i>IT15</i> TWORZĄ STRUKTURĘ TYPU SPINKI.	52
3.4	SĄSIADUJĄCE CIĄGI POWTÓRZEŃ CAG I CCG ODDZIAŁUJĄ ZE SOBĄ WE FRAGMENTCIE mRNA GENU <i>IT15</i> .	56
3.5	ANALIZA TERMODYNAMICZNYCH WŁAŚCIWOŚCI STRUKTUR RNA POWSTAJĄCYCH Z ODDZIAŁYWANIA POWTÓRZEŃ CAG I CCG.	59
3.6	WPLYW SUBTELNYCH ZMIAN LICZBY POWTÓRZEŃ CAG I CCG NA STRUKTURĘ REGIONU POWTÓRZEŃ.	62
3.7	WZROST LICZBY POWTÓRZEŃ CAG NIE ZABURZA ODDZIAŁYWANIA POMIĘDZY CAG I CCG W REGIONIE POWTÓRZEŃ mRNA GENU <i>IT15</i> .	66
3.8	POLIMORFIZM POWTÓRZEŃ CAG W GENIE <i>AR</i> ZWIĄZANYM Z RDZENIOWO – OPUSZKOWYM ZANIKIEM MIĘŚNI TYPU KENNEDY’EGO.	68
3.9	CIĄGI POWTÓRZEŃ CAG WE FRAGMENTCIE mRNA GENU <i>AR</i> NIE ODDZIAŁUJĄ ZE SOBĄ.	70
3.10	WZROST LICZBY POWTÓRZEŃ POLIMORFICZNEGO CIĄGU CAG NIE ZMIENIA ZASADNICZO STRUKTURY REGIONU POWTÓRZEŃ mRNA GENU <i>AR</i> .	73
<u>4</u>	<u>DYSKUSJA WYNIKÓW</u>	76
4.1	STRUKTURA SPINKI TWORZONA PRZEZ CZĄSTECZKI RNA ZBUDOWANE Z POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH CAG LUB CCG.	76
4.2	POWTÓRZENIA CAG ODDZIAŁUJĄ Z POWTÓRZENIAMI CCG W TRANSKRYPCIE GENU <i>IT15</i> .	78
4.3	W mRNA GENU <i>IT15</i> DODATKOWE POWTÓRZENIA CAG ODDZIAŁUJĄCE Z CIĄGIEM CCG WYDŁUŻAJĄ STRUKTURĘ SPINKI.	82

4.4	PORÓWNANIE STRUKTURY FRAGMENTU mRNA GENU <i>IT15</i> ZAWIERAJĄCEGO POWTÓRZENIA TRÓJNUKLEOTYDOWE Z MODELEM STRUKTURY PROPONOWANYM WCZEŚNIEJ.	85
4.5	STRUKTURA REGIONU POWTÓRZEŃ mRNA GENU <i>IT15</i> W ODNIESIENIU DO STRUKTURY REGIONÓW POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH INNYCH RNA.	87
4.6	BIOLOGICZNE ZNACZENIE STRUKTURY REGIONU POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH W mRNA GENU <i>IT15</i> .	92
5	PODSUMOWANIE	95
6	MATERIAŁY	99
6.1	MATERIAŁ BIOLOGICZNY.	99
6.2	WAŻNIEJSZE ODCZYNNIKI CHEMICZNE I BIOCHEMICZNE.	99
6.3	ENZYMY.	100
6.4	BUFORY.	101
6.5	ROZTWORY.	102
6.6	ODCZYNNIKI DO CYKLICZNEGO SEKWENCJONOWANIA DNA.	104
6.7	STANDARDY DO OZNACZANIA WIELKOŚCI FRAGMENTÓW DNA.	104
6.8	ZWIĄZKI PROMIENIOTWÓRCZE.	104
6.9	ŻELE.	105
6.10	SYNTETYCZNE OLIGODEOKSYNUKLEOTYDY.	106
6.11	WAŻNIEJSZA APARATURA STOSOWANA W PRACY.	107
7	METODY	108
7.1	AMPLIFIKACJA FRAGMENTÓW GENU METODĄ PCR.	108
7.2	CYKLICZNE SEKWENCJONOWANIE DNA.	111
7.3	REAMPLIFIKACJA FRAGMENTÓW GENÓW <i>IT15</i> I <i>AR</i> NA MATRYCY PRODUKTÓW PCR.	112
7.4	REAMPLIFIKACJA ZE STARTERAMI ZAWIERAJĄCYMI PROMOTOR POLIMERAZY RNA FAGA T7.	113
7.5	OCZYSZCZANIE MATRYC DO TRANSKRYPCJI <i>IN VITRO</i> .	113

7.6	PRZYGOTOWANIE SYNTETYCZNYCH MATRYC DO TRANSKRYPCJI <i>IN VITRO</i> .	113
7.7	TRANSKRYPCJA <i>IN VITRO</i> .	114
7.8	OCZYSZCZANIE TRANSKRYPTU W ŻELU POLIAKRYLOAMIDOWYM I ELUCJA RNA.	115
7.9	ZNAKOWANIE RADIOIZOTOPEM KOŃCA 5' TRANSKRYPTU.	115
7.10	ANALIZA RNA W ŻELACH NIEDENATURUJĄCYCH.	116
7.11	HYDROLIZA CZĄSTECZEK RNA SONDAMI STRUKTURALNYMI.	116
7.12	ANALIZA PRODUKTÓW REAKCJI CZĄSTECZEK RNA SONDAMI STRUKTURALNYMI.	z 117
7.13	ANALIZA WŁAŚCIWOŚCI TERMODYNAMICZNYCH STRUKTUR CZĄSTECZEK RNA.	119
7.14	BIOINFORMATYCZNA IDENTYFIKACJA LUDZKICH MRNA ZAWIERAJĄCYCH POWTÓRZENIA TRÓJNUKLEOTYDOWE.	119
7.15	ANALIZA STRUKTURY RNA ZA POMOCĄ PROGRAMÓW KOMPUTEROWYCH.	120
8	<u>BIBLIOGRAFIA</u>	<u>122</u>

1 WSTĘP

1.1 Wprowadzenie

Powszechne zastosowanie technik szybkiego sekwencjonowania DNA jakie nastąpiło w ostatnim dwudziestoleciu XX wieku przyczyniło się do powstania nowej dyscypliny nauk biologicznych zajmującej się kompleksową analizą całych genomów – genomiki. Genom jest zapisaną w języku DNA informacją o budowie i fizjologii organizmu.

Ze względów praktycznych najważniejsze było poznanie genomu człowieka. Zakończone niedawno badania jego sekwencji znacznie przybliżyły perspektywę poznania normalnych funkcji genów, jak również ich roli w patogenezie tysięcy znanych chorób. Poznanie genomu człowieka pozwoliło nie tylko na pogłębienie wiedzy o części kodującej zawartej w genach, ale także na uzyskanie pełniejszych informacji o występowaniu różnych elementów takich jak niekodujące RNA czy sekwencje powtórzone. Wśród tych ostatnich szczególnie ciekawe są sekwencje mikrosatelitarne, ponieważ związany jest z nimi wyjątkowy typ mutacji dynamicznych polegających na wydłużaniu ciągów powtórzeń krótkich motywów sekwencji. Mutacje dynamiczne zachodzące najczęściej w obrębie ciągów trójnukleotydowych stanowią molekularne podłoże kilkunastu chorób neurodegeneracyjnych. Mechanizm powstawania tych schorzeń jest słabo poznany. Postuluje się, że obecność wydłużonego ciągu powtórzeń trójnukleotydowych może zaburzać funkcjonowanie komórek przez zmianę struktury kodowanego białka lub patogenny wpływ struktury utworzonej przez te powtórzenia w RNA. W tej pracy poza analizą występowania motywów trójnukleotydowych w kodującej części genomu człowieka przedstawiono biochemiczną charakterystykę regionu mRNA genu *IT15* zawierającego ciągi powtórzeń CAG i CCG w kontekście możliwej roli przyjmowanej przez ten fragment struktury w rozwoju choroby Huntingtona. Ponadto przeprowadzono analizę struktury regionów zawierających kilka ciągów powtórzeń trójnukleotydowych w innych mRNA.

1.2 Genom człowieka

Na początku lat dziewięćdziesiątych ubiegłego wieku rozpoczęła się realizacja projektu sekwencjonowania ludzkiego genomu. Do 2001 roku poznano „surową sekwencję” [International Human Genome Sequencing Consortium. (2001), Venter J. C. i wsp. (2001)], a w roku 2004 ogłoszono zakończenie sekwencjonowania genomu człowieka [IHGSC. (2004)]. Obejmuje ona 3,08 miliarda par zasad z czego 2,88 miliarda to chromatyna aktywna. Chromatyna aktywna nazywana też interfazową lub euchromatyną, jest najmniej skondensowaną formą DNA w komórce, ulegającą transkrypcji, lub dostępną dla tego procesu, podczas ekspresji informacji genetycznej.

1.2.1 Organizacja genomu człowieka.

Ludzki genomowy DNA jest podzielony na 24 chromosomy: 22 chromosomy autosomalne i dwa chromosomy płciowe X i Y. Komórka człowieka, z wyjątkiem gamet, zawiera dwie kopie każdego chromosomu. Chromosomy jednej pary nazywane są chromosomami homologicznymi, chromosomami niehomologicznymi są chromosomy płciowe.

Chromosom składa się z pojedynczej cząsteczki DNA, która dzięki oddziaływaniu ze specyficznymi białkami – histonami ulega silnej kondensacji poprzez tworzenie uporządkowanych zwojów i pętli. Taki kompleks nazywany jest chromatyną. Chromosomy różnią się między sobą wielkością, co wynika z różnic wielkości budującej je cząsteczki DNA (tabela 1.1) oraz kształtem. Podczas cyklu życiowego komórki stan kondensacji chromosomów ulega ciągłym zmianom. Poza okresem podziału komórkowego chromosomy ulegają częściowemu rozluźnieniu, co zapewnia właściwą dostępność DNA dla enzymów i innych białek koniecznych do transkrypcji i replikacji.

Każdy chromosom jest odrębną jednostką strukturalną, która poza przechowywaniem „porcji” informacji genetycznej i umożliwianiem ekspresji musi być zdolna do przekazania jej komórkom potomnym. Chromosom musi być zdolny do replikacji i zapewnić prawidłowy podział informacji pomiędzy nowopowstającymi komórkami. Za kontrolę tych podstawowych funkcji chromosomu odpowiedzialne są

trzy typu sekwencji. Pierwszym z nich są miejsca inicjacji replikacji, czyli sekwencje od których rozpoczyna się duplikacja. Drugim rodzajem sekwencji są centromery odpowiedzialne za przekazanie do komórek potomnych jednej z kopii podwojonego chromosomu. Trzecim rodzajem sekwencji obecnej we wszystkich chromosomach są telomery.

Typowa komórka wykorzystuje tylko część informacji genetycznej, a różne typy komórek wykorzystują informację kodowaną przez różne grupy genów. Ekspresję genów zwykle kontrolują zestawy białek regulatorowych. Kondensacja chromatyny jest jednym ze sposobów regulacji ekspresji genów poprzez wpływ na ich dostępność transkrypcyjną. Ściśle upakowana heterochromatyna jest nie dostępna dla białkowych czynników inicjacji transkrypcji, co powoduje zahamowanie ekspresji występujących w jej obrębie genów.

Tabela 1.1. Liczba par zasad sekwencji chromosomów człowieka (wg danych z 31 maja 2005 NCBI) i udział euchromatyny w genomie [IHGSC. (2004)].

Chromosom	Całkowita wielkość (kbp)	Całkowita poznana sekwencja euchromatyny (kbp)	Udział euchromatyny w chromosomie (%)
1	245 000	222 828	91
2	242 000	237 503	98
3	200 000	194 636	97
4	191 000	187 161	98
5	180 000	177 703	99
6	171 000	167 318	98
7	158 000	154 759	98
8	146 000	142 613	98
9	138 000	117 781	85
10	135 000	131 614	97
11	134 000	131 131	98
12	132 000	130 259	99
13	114 000	95 560	84
14	106 000	88 291	83
15	100 000	81 342	81
16	88 000	78 885	90
17	80 000	77 800	97
18	76 000	74 656	98
19	64 000	55 786	87
20	62 000	59 505	96
21	46 000	34 170	74
22	49 000	34 765	71
X	154 000	150 394	98
Y	57 000	24 872	44
Łącznie	3 068 000	2 851 331	93

1.2.2 Liczba ludzkich genów.

Po publikacji „surowej” sekwencji genomowego DNA człowieka często zadawanym pytaniem było pytanie o liczbę zawartych w nim genów, czyli w dużym uproszczeniu funkcjonalnych jednostek informacji genetycznej występujących w *locus* o określonej lokalizacji chromosomowej, determinujących charakterystyczne cechy gatunku jako całości i cechy osobnicze w obrębie gatunku.

Na podstawie informacji o liczbie genów znajdujących się we wcześniej poznanych genomach w chwili przedstawienia w lutym 2001 roku „surowej” sekwencji genomu człowieka [Venter J. C. i wsp. (2001), IHGSC (2001)] szacowano, że powinien on zawierać około 35 tysięcy genów kodujących białko, podczas gdy wcześniejsze źródła sugerowały ich liczbę na poziomie zbliżonym do 75 a nawet 100 tysięcy. Prowadzone przez blisko 3 lata dalsze prace International Human Genome Sequencing Consortium doprowadziły do jeszcze bardziej zaskakującego wyniku. W październiku 2004 roku ogłoszono, że w sekwencji genomu człowieka znajduje się 22287 genów, na które składa się 19438 znanych genów i 2188 prawdopodobnie kodujących białko regionów DNA. Tej liczbie genów odpowiada 34214 transkryptów co oznacza, że z jednego *locus* powstaje statystycznie 1,54 mRNA. Całkowita długość sekwencji kodującej sekwencje białek, czyli ulegającej translacji to blisko 3,4 milionów par zasad, czyli zaledwie 1,2% ludzkiej euchromatyny. Regiony DNA kodujące nie ulegające translacji części transkryptów to dodatkowe 2,1 miliona par zasad (0,7%) euchromatyny. Zidentyfikowano 231667 eksonów, zatem statystyczny gen składa się z 10,4 eksonów, a statystyczny transkrypt wykorzystuje sekwencję zapisaną w 9,1 eksonach. Grupą regionów DNA, której poznanie stało się możliwe po uzyskaniu całkowitej sekwencji ludzkiego genomu są pseudogeny. Są one regionami przypominającymi prawidłowe geny nie kodującymi jednak funkcjonalnych produktów. Pseudogeny są nieaktywne, ponieważ w ich sekwencji występują zaburzenia prawidłowej struktury genu. Przykładowo nie posiadają one sekwencji zgodnej 5' koniecznej do zajścia procesu składania mRNA. Kolejną cechą uniemożliwiającą wytworzenie prawidłowego mRNA może być zmutowany sygnał poliadenylacji. Może także mieć miejsce mutacja kodonu inicjującego ATG na GTG. W pseudogenach zaobserwowano również szereg innych zmian uniemożliwiających prawidłowy przebieg syntezy białka. Przypuszcza się, że ogólna liczba pseudogenów w genomie człowieka jest zbliżona do liczby funkcjonalnych genów.

Przedstawione powyżej dane nie zawierają informacji o liczbie genów kodujących RNA nie ulegające translacji pełniących funkcje regulatorowe takich jak małe jąderkowe RNA (snoRNA – *small nucleolar RNA*) czy mikroRNA oraz geny cząsteczek RNA odpowiedzialnych za ekspresję informacji genetycznej czyli rybosomowych RNA i transferowych RNA. Fakt, że ciągle są odkrywane nowe cząsteczki RNA nie ulegające translacji, a pełniące w komórce istotne funkcje wskazuje na konieczność prowadzenia dalszych badań i nie pozwala na rzetelne podsumowanie całości wiedzy o tej części ludzkiego genomu.

1.2.3 Zróżnicowanie genetyczne – SNP.

Osobniki jednego gatunku są do siebie bardzo podobne ale nie identyczne. Zróżnicowanie fenotypowe, takie jak na przykład różny wygląd zewnętrzny czy zdolności do przystosowania się do warunków otoczenia, wynika z różnic obecnych w ich genomach. Różnice te to polimorfizmy, czyli odstępstwa od sekwencji DNA danego *locus* charakterystycznej dla większości osobników. Nie powodują one utraty zdolności do pełnienia prawidłowej funkcji przez sekwencję, w której występują, a jedynie ją zmieniają. Występują one w populacji z częstością większą niż 3%, przy czym różne progi częstości od 2 do 5% podawane są przez autorów różnych publikacji. Największą i jednocześnie najlepiej poznaną grupą tego rodzaju różnic są polimorfizmy pojedynczego nukleotydu SNP (*Single Nucleotide Polymorphism*). Są to najczęściej różnice polegające na zastąpieniu jednego nukleotydu innym. Istnieje sześć występujących z różną częstością możliwych zmian nukleotydów. Najczęstsze, stanowiące po około 30%, to tranzycje (czyli zastąpienie zasady purynowej lub pirymidynowej inną należącą do tego samego typu) A/G i C/T [Venter J. C. i wsp. (2001)]. Rzadziej obserwowane są zamiany zasady purynowej na pirymidynową lub pirymidynowej na purynową nazwane transwersjami: A/C, A/T, C/G i T/G. Występują one z częstością około 10% każda [Venter J. C. i wsp. (2001)]. Zatem tranzycje występują w genomie człowieka 1,6 raza częściej niż transwersje. SNP są rozłożone w genomie nierównomiernie, w sposób przypadkowy. Aktualnie poznano i scharakteryzowano 9,37 miliona tego rodzaju różnic obecnych w genomie człowieka. Informacje na ich temat zgromadzone są w bazie danych dbSNP (www.ncbi.nlm.nih.gov/SNP). Pod względem wpływu na kodowaną informację genetyczną SNP podzielono na dwie klasy. Pierwszą, zdecydowanie liczniejszą stanowią te, które nie zmieniają sekwencji kodowanych białek lub ze względu na ich

lokalizację wydają się nie mieć wpływu na strukturę i funkcję nie ulegających translacji RNA. Występują one często w niekodujących regionach międzygenowych, gdzie znaleziono 75% poznanych dotąd polimorfizmów, lub w intronach. Stwierdzono, że częstość występowania SNP w intronach jest wyższa niż w sekwencjach międzygenowych. Polimorfizmy pojedynczego nukleotydu należące do tej grupy obecne w regionach ulegających translacji nie powodują zmiany wbudowywanego aminokwasu pomimo zmiany w kodonie, ponieważ są zlokalizowane w pozycji jego trzeciego nukleotydu. Druga klasa to około 1% wszystkich SNP. Do niej należą te, które powodują zmianę w sekwencji białka poprzez spowodowanie zmiany jednego aminokwasu na inny czy też zmianę kodonu „stop” na inny lub wprowadzanie sygnału przedwczesnej terminacji translacji. Pomimo spowodowania zmiany struktury białka nie zaburzają one jego funkcji w sposób na tyle istotny aby prowadziło to do pojawienia się fenotypu chorobowego, co w przypadku dwóch ostatnich zmian jest zjawiskiem raczej rzadkim. Warto zwrócić uwagę, że częstość SNP w częściach eksonowych genów jest wyższa niż w regionach niekodujących [Venter J. C. i wsp. (2001)].

1.3 Mutacje w genomie człowieka

Zmiany sekwencji DNA mające wpływ na zmianę funkcji genu, w którym występują określane są mianem mutacji. W bazie danych zawierającej informacje na temat tego rodzaju zmian występujących w genomie człowieka (Human Gene Mutation Database – www.hgmd.cf.ac.uk) zgromadzono informacje na temat 54562 zmian w 2093 genach i 1893 sekwencjach cDNA nie mających jeszcze swoich genomowych odpowiedników (dane z dnia 08.08.2005). Od połowy lat osiemdziesiątych XX wieku stale wzrasta tempo wykrywania nowych mutacji. W ciągu ostatnich 6 lat osiągnęło ono pułap około 5 tysięcy doniesień o nieznanym wcześniej mutacji pojawiających się w ciągu roku. Można przypuszczać, że łączna liczba mutacji będzie podlegała dalszemu wzrostowi w miarę postępu badań nad związkami genów z chorobami i obejmowania badaniami kolejnych grup osób chorych.

Mutacje podzielono na dwie grupy ze względu na wielkość odcinka DNA jakiego dotyczą. Pierwszą grupą są małe zmiany zachodzące w obrębie kilku nukleotydów. Najliczniejszym rodzajem mutacji są mutacje punktowe polegające na

zastąpieniu w sekwencji kwasu deoksyrybonukleinowego jednej zasady inną co prowadzi do zmiany kodowanego aminokwasu lub wprowadzenia kodonu terminującego translację. Blisko 60% wszystkich dotąd poznanych mutacji stanowią mutacje typu zmiany sensu. Do grupy małych zmian DNA należy także utrata niewielkiego fragmentu cząsteczki – delecja (16%) oraz wbudowanie dodatkowych nukleotydów – insercja (6%). Istotną część tej grupy tworzą mutacje prowadzące do zaburzenia dojrzewania RNA i regulacji ekspresji na innych jej poziomach (10%). Zdecydowanie mniejszą jest druga grupa mutacji, którą tworzą zmiany zachodzące z udziałem dużych fragmentów DNA. Największą częścią tej grupy są delecje dużych odcinków DNA, które stanowią 5% wszystkich mutacji w ludzkim genomie. Poza nimi do grupy „dużych” mutacji należą rearanżacje obejmujące między innymi inwersje, czyli odwrócenia regionów DNA oraz insercje i duplikacje. Osobną grupę stanowią te zmiany, które dotyczą liczby elementów w obrębie sekwencji powtarzających się w genomie człowieka. Dotąd poznano 143 mutacje powodujące skrócenie lub wydłużenie ciągu złożonego z wielokrotnie powtórzonych motywów sekwencji (HGMD 08.08.2005).

1.4 Sekwencje powtarzające się w genomie człowieka.

W genomie człowieka, podobnie jak w genomach innych organizmów stwierdzono występowanie powtarzających się wielokrotnie sekwencji. Nie są to powtórzenia jednego rodzaju motywu. Powtarzane mogą być sekwencje o zróżnicowanej budowie i wielkości. Stanowią one około 50% całej sekwencji genomowej. Podzielono je na kilka klas [IHGSC (2001)] obejmujących między innymi: sekwencje rozproszone gdzie elementy powtarzające się są oddzielone od siebie innymi sekwencjami, nieaktywne kopie genów czyli pseudogeny. Dużą grupę stanowią sekwencje tandemowe, których motywy powtarzające się ułożone są jak „głowa do ogona” i występują obok siebie. Wśród nich wyróżniono powtórzenia związane z regionami centromerowymi i telomerowymi chromosomów, powtórzenia segmentów o wielkości od 10 do 300 tysięcy par zasad. Do powtórzeń tandemowych należą także powtórzenia krótkich kilku lub kilkunastonukleotydowych motywów określane mianem powtórzeń satelitarnych.

Do pierwszej z grup należą sekwencje LINE (*Long Interspersed Nucleotide Elements* – długie rozproszone powtórzenia). Są one jednym z większych motywów

powtarzających się w genomie człowieka. Mają długość do 7 tysięcy par zasad [Singer M. F. i wsp. (1993)]. Innym rodzajem powtórzeń rozproszonych są odcinki o wielkości 100 – 500 par zasad zaliczane do rodziny SINE (*Short Interspersed Nucleotide Elements* – krótkie rozproszone powtórzenia). Do tej grupy należy rodzina sekwencji Alu występująca w genomie człowieka w liczbie około miliona kopii, której długość powtarzanego motywu wynosi niecałe 300 par zasad [Schmid C. W., Jelinek W. R. (1982)].

Sekwencje powtarzające się złożone z motywów mających długość od 10 do 80 par zasad i całkowitej długości wynoszącej od kilkuset do ponad 20 tysięcy par zasad [Weir B.S. (1992)] nazywane są sekwencjami minisatelitarnymi (minisatelity) lub VNTR (*Variable Number of Tandem Repeats* – sekwencje o zmiennej liczbie powtórzeń tandemowych).

Ważną, chociaż niewielką grupę sekwencji powtarzających się stanowią sekwencje zwane mikrosatelitami [Weber J. L., Wong C. (1993)] (inne nazwy spotykane w literaturze to: STRs – *Short Tandem Repeats*, SSRs – *Simple Sequence Repeats*). Sekwencje mikrosatelitarne są grupą sekwencji o powtarzających się motywach długości od 1 do 9 par zasad ułożonych obok siebie i łączących się ze sobą jak „głowa do ogona”. Zawierają one od 6 do 50 powtórzeń motywu [Beckman J. S., Weber J. L. (1992)]. Sekwencje mikrosatelitarne mogą występować jako idealne powtórzenia lub w formie przedzielonych krótkimi, kilkunukleotydowymi wstawkami (*interruptions*) zaburzającymi ciągłość traktów. Istnieją również mikrosatelity złożone z ciągu dwóch lub większej liczby różnych motywów powtarzających się.

Różne sekwencje mikrosatelitarne występują z różną częstością w genomach Eukariota [Stallings R. L. (1994)], przy czym ich rozmieszczenie w materiale genetycznym większości organizmów jest względnie równomierne. W genomie człowieka występują średnio co 6 tysięcy par zasad i stanowią blisko 3% całego DNA [IHGSC (2001)]. Większość STR jest zlokalizowanych w sąsiedztwie końca 3' sekwencji Alu albo rzadziej pomiędzy dwoma monomerami Alu. Zaobserwowano także występowanie STR od strony 5' elementu Alu. Jedna trzecia mikrosatelitów nie jest związana z sekwencjami Alu, jednak około 20% z nich jest zasocjowanych z innymi sekwencjami powtarzającymi się w genomie [Nadir E. i wsp. (1996)].

Stwierdzono zależność długości ciągu powtórzeń mikrosatelitarnych od rodzaju powtarzanego motywu. U człowieka ciągi dwunukleotydowe są najdłuższe

(wyłączając ciągi poly-A) natomiast sześci nukleotydowe najkrótsze. Również częstość występowania ciągów dwunukleotydowych jest największa w porównaniu z innymi STR, jeżeli i w tym przypadku pominie się w ocenie specyficzny motyw jakim jest ciąg poly-A [Nadir E. i wsp. (1996)].

1.5 Trójnukleotydowe sekwencje mikrosatelitarne.

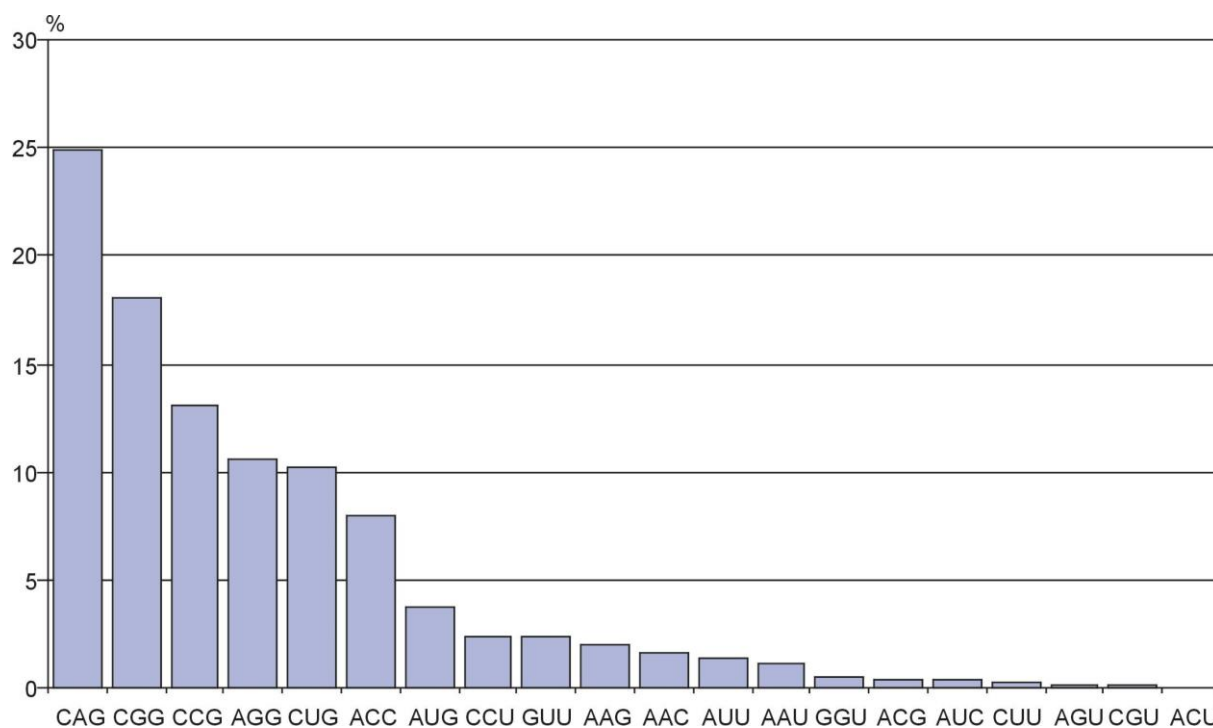
Powtórzenia trójnukleotydowe w genomie człowieka, podobnie jak w genomach innych ssaków, są jedną z mniejszych grup w porównaniu z innymi rodzajami sekwencji mikrosatelitarnych (stanowią około 6% wszystkich mikrosatelitów), mniej jest tylko ciągów o powtarzającym się motywie pięcynukleotydowym [Stallings R. L. (1994), IHGSC (2001)]. Warto zauważyć, że powtórzenia trójnukleotydowe występują najczęściej w sekwencjach genów, rzadko zaś w niekodujących regionach międzygenowych, a ich rozmieszczenie w chromosomach człowieka nie jest równomierne [Stallings R. L. (1994)]. W ludzkich genach powtórzenia trójnukleotydowe występują w prawie równych proporcjach w eksonach i intronach [Stallings R. L. (1994)]. Wiadomo, że częstość występowania tego typu mikrosatelitów waha się w zależności od rodzaju tworzącego je motywu osiągając średnią wartość 11,8 traktu w milionie par zasad. Pozwala to przypuszczać, że w całym genomie człowieka występuje ponad 35 tysięcy ciągów trójnukleotydowych. Najczęściej spotykane są motywy AAT i AAC, które łącznie występują w ponad połowie traktów, natomiast najmniej jest powtórzeń motywów ACC (4%), CAG (2,2%), ACT (1,4%) oraz CGA (0,1%) [IHGSC (2001)]. Długość sekwencji powtarzających się zbudowanych z elementów trójnukleotydowych, podobnie jak w przypadku pozostałych rodzajów sekwencji mikrosatelitarnych, może niekiedy wykazywać polimorfizm długości. Te niewielkie różnice pomiędzy tymi samymi *loci* w genomach różnych osobników tego samego gatunku są istotne z punktu widzenia zmienności genetycznej populacji. Można je wykorzystać jako markery genetyczne służące do identyfikacji osób i określania pokrewieństwa oraz do sporządzania map genetycznych przydatnych do identyfikacji genów związanych z określonymi cechami fenotypowymi. Pomocne mogą być także przy sporządzaniu map filogenetycznych zależności pomiędzy gatunkami [Gastier J. M. i wsp. (1995)].

1.6 Powtórzenia trójnukleotydydowe w ludzkim transkryptomie.

Pomimo poznania pełnej sekwencji genomu człowieka do chwili podjęcia badań objętych tą pracą doktorską nie przedstawiono jeszcze kompletnego zestawienia ludzkich genów zawierających ciągi powtórzeń trójnukleotydydowych. Dążąc do uzupełnienia wiedzy o znaczeniu tego rodzaju mikrosatelitów w ludzkim materiale genetycznym w ramach niniejszej pracy doktorskiej przeanalizowano, pod kątem ich występowania, dostępne w bazie danych GenBank sekwencje mRNA. Poszukiwano sekwencji zawierających 6 i więcej niezaburzonych powtórzeń motywu składającego się z trzech nukleotydów. W prowadzonych badaniach podstawowym pytaniem, na które chciano znaleźć odpowiedź była liczba różnych mRNA zawierających ciągi trójnukleotydydowe. Interesujące było także to, jakiego rodzaju motywy dominują w transkryptomie. Informacji o roli tego typu mikrosatelitów mogło dostarczyć poznanie ich usytuowania w poszczególnych częściach dojrzałego mRNA oraz określenie funkcji zawierających je genów. W celu rozwiązania tak sformułowanego problemu posłużono się programem BLASTn umożliwiającym wyszukanie wśród zdeponowanych sekwencji tych, które zawierają określony motyw. Analizę sekwencji powtórzeń trójnukleotydydowych w RNA, pomimo istnienia 60 możliwych tripletów, można ograniczyć tylko do 20 kombinacji sekwencji nukleotydów o powtarzonym motywie (CAG, CUG, AGG, CCU, AGU, UCA, AUU, UAA, ACG, UGC, CCG, CGG, ACU, UGA, CUU, GAA, GGU, CCA, GUU, CAA). Na przykład zapis $(CAG)_n$ oznacza jednocześnie 3 możliwe trójki nukleotydów: $(CAG)_n$, $(AGC)_n$, $(GCA)_n$. Każdą z zaproponowanych kombinacji układu trzech nukleotydów można by przedstawić w podobny sposób. Rozpatrując teoretycznie wszystkie możliwe kombinacje trójnukleotydydowe należałoby także dodać 4 typy będące powtórzeniami identycznego nukleotydu np. $(AAA)_n$, które jednak są klasyfikowane jako ciągi jednonukleotydydowe, więc zostały w tych badaniach pominięte. Poza ciągami złożonymi z jednego rodzaju motywu w prowadzonych analizach uwzględniono także trakty dłuższe niż 6 powtórzeń zawierające pojedyncze zaburzenia w postaci innego typu trójki nukleotydów. Informacje na temat funkcji białek kodowanych przez zidentyfikowane w ten sposób mRNA pochodzą z baz danych GenBank i OMIM. Geny, których funkcja została określona zakwalifikowano do jednej z 9 kategorii funkcjonalnych.

1.6.1 Liczba różnych ciągów trójnukleotydowych w ludzkich mRNA.

W wyniku przeprowadzonych poszukiwań zidentyfikowano 619 różnych ludzkich mRNA zawierających 718 ciągów powtórzeń trójnukleotydowych [Jasińska A. i wsp. (2003)]. Różnica w liczbie zidentyfikowanych traktów i zawierających je mRNA wynika z faktu, że w 47 przypadkach w obrębie jednej cząsteczki mRNA obecny jest więcej niż 1 (do 6) ciąg powtórzeń oddzielony co najmniej 6 nukleotydami od innych trójnukleotydowych sekwencji powtarzających się. Blisko połowa z nich zawiera różnego typu powtórzenia. Pięć rodzajów motywów obserwowanych jest z częstością przewyższającą 10% (ryc. 1.1). W porównaniu z danymi dotyczącymi sekwencji genomowego DNA zaskakująca jest liczba traktów CAG obecnych w mRNA. Znalezione ich 179, co stanowi 25% wszystkich powtórzeń trójnukleotydowych w transkrypcie, natomiast w sekwencji genomowej stanowią one tylko 2,2%. Najmniej liczne w DNA (0,1%) motywy ACG są obecne w mRNA z częstością (0,4%). Zaś wśród zidentyfikowanych transkryptów nie ma żadnego, który zawiera powtórzenia motywu ACU, mimo że jego odpowiednik w DNA (ACT) stanowi 1,4% ciągów trójnukleotydowych. Zatem różne motywy trójnukleotydowe są



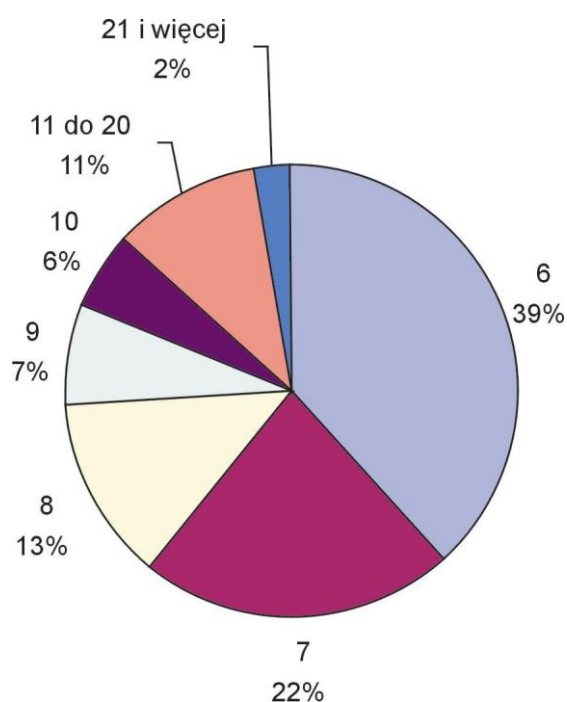
Rycina 1.1. Częstość występowania powtórzeń poszczególnych motywów trójnukleotydowych w ludzkich mRNA.

obecne w transkrypcji z częstością, która znacznie odbiega od obserwowanej w sekwencji genomowej [de Mezer M. dane nie publikowane]. Jest to szczególnie ważne ponieważ, jak już wspomniano występują one najczęściej w sekwencjach genów, rzadko zaś w niekodujących regionach międzygenowych. Można więc przypuszczać, że ich obecność w regionach stanowiących dojrzałą cząsteczkę mRNA biorącą udział w translacji może mieć istotny wpływ na przebieg tego procesu lub dla funkcjonowania białka, w którym mogłyby się znaleźć kodowane przez nie trakty jednego rodzaju aminokwasu.

Wśród zidentyfikowanych ciągów, 72 (10%) zawierają zaburzenia w postaci obecności motywów różnych od podstawowej trójki tworzącej dany trakt [de Mezer M. dane nie publikowane]. Największą liczbę zaburzeń stwierdzono w jednym z ciągów CAG, w którym 21 powtórzeń tego typu było pięciokrotnie przedzielonych innymi motywami. Występowanie zaburzeń może mieć istotne znaczenie dla możliwej funkcji powtórzeń, na co wskazywałby fakt jego występowania w dłuższych ciągach.

1.6.2 Długość traktów trójnukleotydomy.

Istotne znaczenie dla możliwej roli powtórzeń trójnukleotydomy w procesach biologicznych wydaje się mieć liczba powtórzeń określonego motywu sekwencji



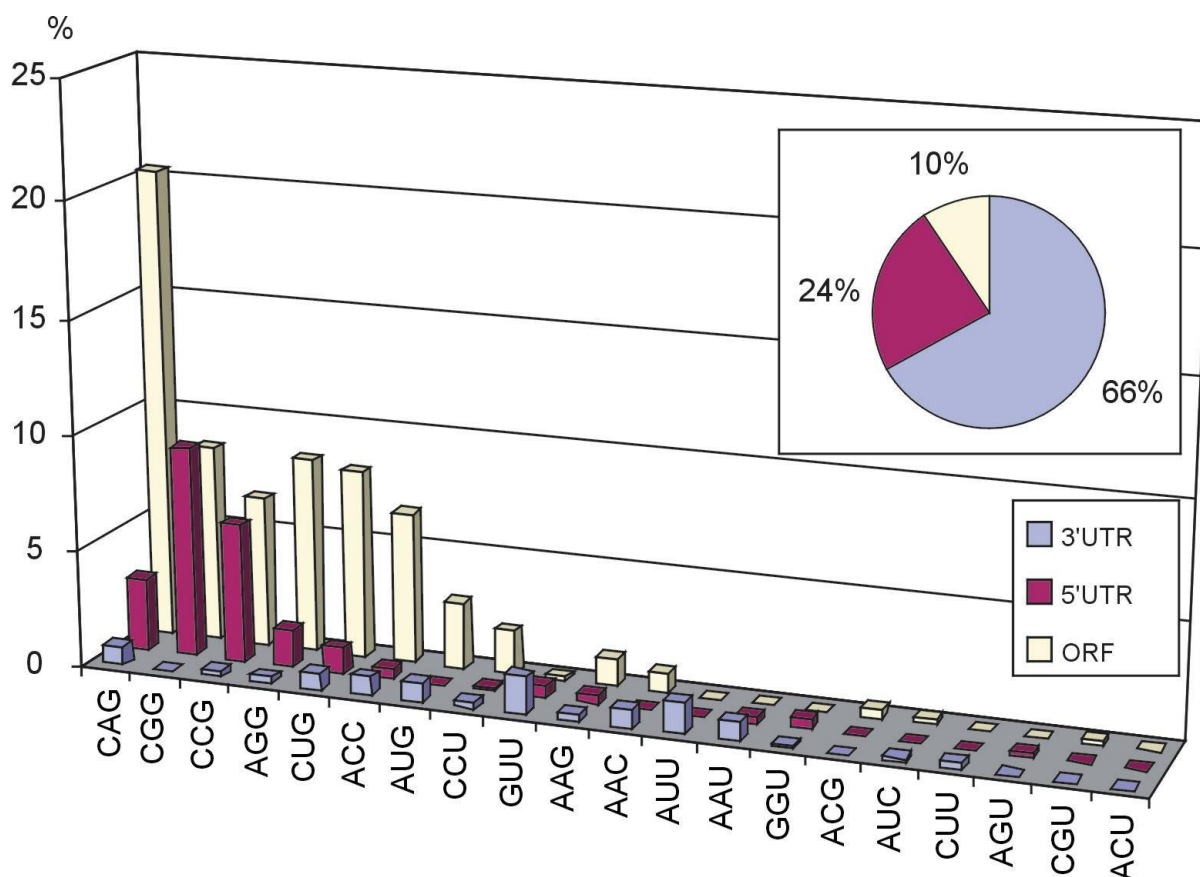
Rycina 1.2. Częstość występowania traktów o określonej liczbie powtórzeń motywu trójnukleotydomy.

tworzących pojedynczy ciąg, a więc jego długość. W puli zidentyfikowanych mRNA zdecydowana większość zawiera krótkie trakty złożone z 6 i 7 powtórzeń (ryc. 1.2). Grupę traktów o długości 11 do 20 powtórzeń tworzy 77 genów (11%), natomiast długich, zawierających więcej niż 21 powtórzeń trójnukleotydowych znaleziono zaledwie 19 (2%) mRNA [Jasińska A. i wsp. (2003)]. Tak wyraźna dysproporcja częstości występowania ciągów trójnukleotydowych różnej długości przypuszczalnie nie odzwierciedla rzeczywistej sytuacji w genomie. Niska liczba zidentyfikowanych traktów zawierających więcej niż 21 powtórzeń elementu trójnukleotydowego może wynikać z faktu ich preferencyjnej delekcji podczas klonowania w bakteriach w procesie sekwencjonowania genomu. Ponadto, część ciągów trójnukleotydowych charakteryzuje się polimorfizmem liczby powtórzeń motywów wchodzących w ich skład. W poznanych mRNA zaledwie w 7% stwierdzono, na podstawie informacji zawartych w bazach sekwencji GenBank i RefSeq, występowanie kilku wariantów długości [Jasińska A. i wsp. (2003)]. Zaznaczyć trzeba, że tylko około 1% ciągów trójnukleotydowych w mRNA opisano jako nie wykazujące polimorfizmu. Należy się zatem spodziewać, że wraz z postępem poznawania osobniczego zróżnicowania informacji genetycznej wzrośnie liczba genów zawierających sekwencje mikrosatelitarne zbudowane z motywów trójkowych o zróżnicowanej długości między poszczególnymi osobnikami.

1.6.3 Występowanie powtórzeń trójnukleotydowych w funkcjonalnych częściach mRNA.

W dojrzałych cząsteczkach mRNA wyróżniamy następujące regiony: otwartą ramkę odczytu (ORF – *Open Reading Frame*) kodującą sekwencję białkową, mającą u człowieka średnią wielkość 1100 nukleotydów oraz regiony nie ulegające translacji (UTR – *UnTranslated Regions*) zlokalizowane na końcach 5' i 3', o przeciętnej długości odpowiednio 240 i 400 nt. W każdej z tych części mogą występować ciągi powtórzeń trójnukleotydowych. W trakcie analizy zwrócono szczególną uwagę na lokalizację tego rodzaju sekwencji powtarzających się w mRNA, ponieważ może ona mieć istotne znaczenie dla poznania ich funkcji. Stwierdzono nierównomierne rozmieszczenie traktów trójnukleotydowych w mRNA. Zdecydowana ich większość – 67% występuje w otwartych ramkach odczytu. Blisko czwarta część zlokalizowana jest w 5' UTR, zaś tylko 9% w 3'UTR (ryc. 1.3). Taki rozkład w stosunku do wielkości

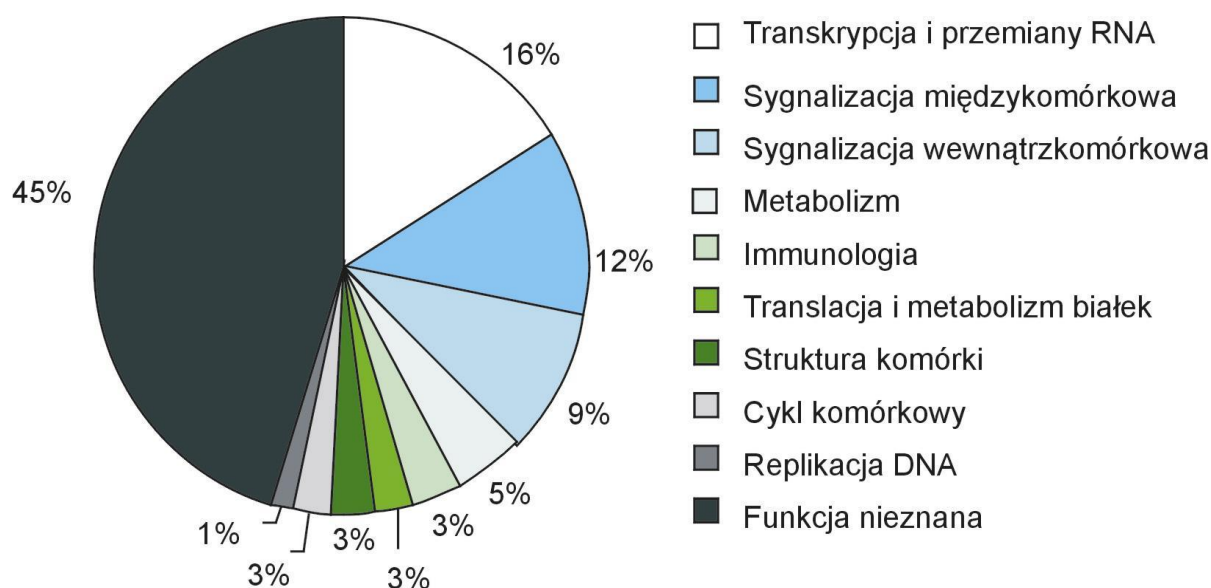
regionów mRNA wskazuje na proporcjonalnie większą częstość powtórzeń w 5'UTR, a zaniżoną ich reprezentację w 3'UTR [Jasińska A. i wsp. (2003)]. Lokalizacja traktów przypuszczalnie wynika z ich funkcji. 5'UTR pełni rolę w regulacji translacji, więc stosunkowo duża liczba powtórzeń CGG, CCG i CAG (ryc. 1.3) sugeruje, że tworzone przez nie elementy struktury drugorzędowej mogą brać udział w regulacji tego procesu stanowiąc na przykład miejsce wiązania czynników białkowych. Natomiast w odpowiedzialnym za stabilność mRNA i ich lokalizację komórkową 3'UTR dominują powtórzenia motywów bogatych w A i U (GUU, AUU, AAU – ryc. 1.3) wykazujące mniejszą zdolność do tworzenia trwałych struktur drugorzędowych. O roli powtórzeń w 3'UTR dodatkowo świadczy fakt, że powtórzenia AUU i CUU nie występują w innych regionach mRNA. Ponadto w regionie tym stwierdzono praktycznie brak motywów charakterystycznych dla regionu 5'UTR (całkowity brak powtórzeń CGG).



Rycina 1.3. Częstość występowania powtórzeń różnych motywów trójnukleotydowych w poszczególnych częściach mRNA.

1.6.4 Komórkowe funkcje genów z powtórzeniami trójnukleotydowymi.

Z uwagi na to, że w znacznej części zidentyfikowanych mRNA powtórzenia trójnukleotydowe występują w otwartej ramce odczytu ulegającej translacji, a pozostałe wydają się mieć znaczenie dla ekspresji informacji genetycznej poprzez wpływ na regulację procesu powstawania białka lub regulację stabilności mRNA, można by sądzić, że istnieje grupa genów o określonej funkcji w komórce, w której częściej występuje ten rodzaj sekwencji mikrosatelitarnych. Wśród 619 genów 54% miało opisaną biologiczną funkcję. Na podstawie klasyfikacji 17000 genów o znanej funkcji zastosowanej w opisie sekwencji genomu człowieka [IHGSC (2001)], geny z powtórzeniami przydzielono do jednej z dziewięciu grup: transkrypcja i przemiany RNA, sygnalizacja międzykomórkowa, sygnalizacja wewnątrzkomórkowa, metabolizm, immunologia, translacja i metabolizm białek, struktura komórki, cykl komórkowy oraz replikacja DNA. Najliczniejszą z nich (16% – ryc. 1.4) jest ta, do której należą geny kodujące czynniki transkrypcyjne i białka biorące udział w dojrzewaniu lub degradacji cząsteczek RNA. Również komunikacja wewnątrz- i międzykomórkowa stanowią istotną część procesów, w które zaangażowane są geny zawierające trakty trójnukleotydowe [de Mezer M. dane nie publikowane].



Rycina 1.4. Funkcja genów zawierających powtórzenia trójnukleotydowe w sekwencji kodującej.

W odniesieniu do lokalizacji ciągu w funkcjonalnych częściach mRNA wśród transkryptów zawierających powtórzenia w otwartej ramce odczytu zachowana jest dominacja (34%) funkcji regulacji transkrypcji. Natomiast te cząsteczki, w których sekwencja mikrosatelitarna zlokalizowana jest w regionach nie ulegających translacji częściej kodują białka uczestniczące w sygnalizacji międzykomórkowej i wewnątrzkomórkowej. Wśród mRNA z ciągami trójnukleotydomi występującymi w 5'UTR stwierdzono także ponad dwukrotny wzrost udziału genów białek metabolizmu porównaniu z ogólną charakterystyką funkcjonalną genomu człowieka. W grupie transkryptów z powtórzeniami zlokalizowanymi w regionie 3'UTR zaobserwowano wzrost udziału mRNA kodujących białka strukturalne i metabolizmu podstawowego oraz biorące udział w odpowiedzi immunologicznej [de Mezer M. dane nie publikowane].

Wydaje się zatem, że powtórzenia trójnukleotydowe, a właściwie kodowane przez nie ciągi jednego typu aminokwasów, mogą mieć znaczenie dla prawidłowego działania białek zaangażowanych w procesy syntezy RNA i jego prawidłowego funkcjonowania, poprzez wpływ na regulację jego dojrzewania oraz degradacji. Natomiast wpływ wywierany przez obecne w regionach niekodujących trakty trójnukleotydowe na poziom translacji oraz stabilność i lokalizację mRNA dotyczy procesów metabolicznych oraz wymiany informacji pomiędzy komórkami.

Najwięcej informacji istotnych dla poznania roli powtórzeń trójnukleotydomi w genach człowieka dostarczyło zestawienie liczby genów zawierających sekwencje mikrosatelitarne o trójelementowych motywach powtarzających się pełniących określoną funkcję z wielkością odpowiedniej klasy funkcjonalnej utworzonej z wszystkich genów człowieka. Porównanie to wykazało większy udział białek związanych z sygnalizacją między- i wewnątrzkomórkową (wzrost z 28% do 39%) oraz czynników regulujących procesy transkrypcji i translacji (wzrost z 24% do 34%). Istotną różnicą jest także mniejsza liczba genów z powtórzeniami zaangażowanych w regulację metabolizmu (spadek z 19% do 9%) [Jasińska A. i wsp. (2003)]. Ponieważ występowanie traktów trójnukleotydomi w sekwencji genu najczęściej uwidacznia się pojawieniem się w białku ciągu aminokwasów jednego typu, w świetle analiz funkcjonalnych można przypuszczać, że obecność STR ma istotne znaczenie dla oddziaływań pomiędzy elementami białkowymi szlaków sygnalizacyjnych oraz ich wiązania do układu błon komórkowych. Natomiast w regulacji procesów ekspresji genów rola powtórzeń trójnukleotydomi, poza wpływem na interakcje między

biorącymi w nich udział białkami, może polegać na regulacji stabilności cząsteczek kwasów nukleinowych, szczególnie mRNA, oraz wymuszeniu przyjęcia przez regiony regulatorowe cząsteczek kwasów nukleinowych określonej struktury drugorzędowej decydującej o przyłączeniu odpowiednich czynników białkowych.

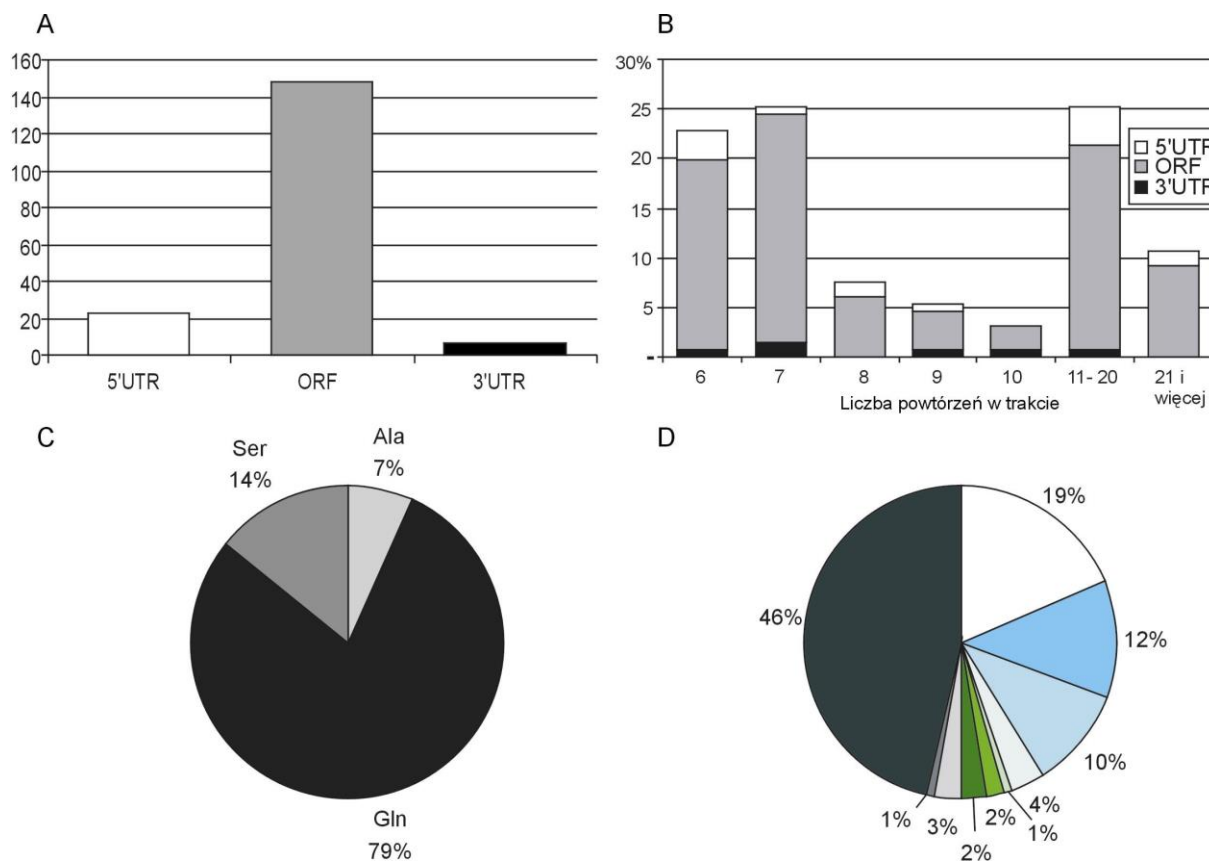
1.7 Występowanie i znaczenie najczęstszych motywów trójnukleotydowych obecnych w ludzkim transkryptomie.

Wśród 5 typów motywów tworzących ciągi trójnukleotydowe występujące w transkryptomie człowieka z częstością przekraczającą 10% większość można określić jako powtórzenia typu CNG. Są one ważną grupą z punktu widzenia poznania roli komórkowej powtórzeń trójnukleotydowych, ponieważ stanowią łącznie 66% wszystkich ciągów zidentyfikowanych w transkryptomie. Istotnym faktem jest ich zdolność do przyjmowania struktury typu „spinki do włosów” w cząsteczkach RNA [Napierała M., Krzyżosiak W. J. (1997)]. Ze względu na przedstawione powyżej powody dokonano szczegółowej charakterystyki występowania i właściwości traktów złożonych z powtórzeń CNG obecnych w mRNA [de Mezer M. dane nie publikowane].

1.7.1 Powtórzenia CAG.

Najczęstszym (24,9%) typem motywu trójnukleotydowego tworzącego sekwencje ciągów powtórzeń w kodujących częściach genów człowieka jest triplet CAG. Powtórzenia tego rodzaju występują głównie w obrębie otwartej ramki odczytu (ryc. 1.5 A) i powodują wbudowanie kolejnych reszt glutaminy do cząsteczki białkowej kodowanej przez mRNA, w którym występują, lub, co ma miejsce zdecydowanie rzadziej, reszt serynowych bądź alaninowych (ryc. 1.5 B). Ciągi powtórzeń CAG można podzielić pod względem liczby tworzących je motywów na dwie grupy. Pierwsza to trakty krótkie, złożone z sześciu lub siedmiu powtórzeń a drugą stanowią trakty dłuższe niż 10 powtórzeń CAG. W analizie funkcjonalnej, poza niewielkim (o 3%) wzrostem udziału genów związanych z procesami syntezy i degradacji cząsteczek RNA nie zaobserwowano istotnych odstępstw od tendencji opisanych dla wszystkich zidentyfikowanych mRNA zawierających powtórzenia trójnukleotydowe. Wydaje się, że podstawowa rola powtórzeń CAG może polegać na

kodowaniu traktów poliglutaminowych (poliQ) pełniących nie do końca poznaną dotąd funkcję w oddziaływaniach między białkami. Trudno także ocenić jakie znaczenie funkcjonalne mogą odgrywać w cząsteczkach mRNA struktury tworzone przez naturalnie występujące trakty CAG.

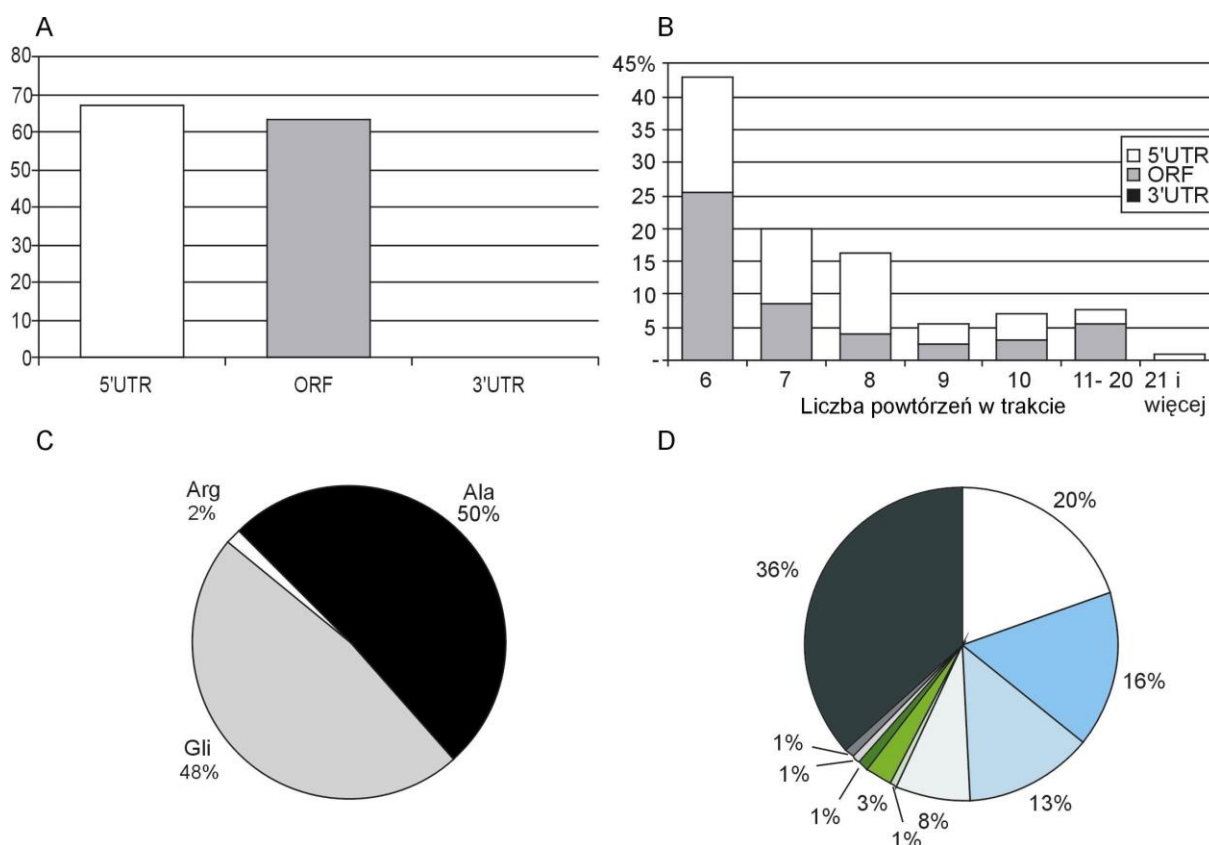


Rycina 1.5. Powtórzenia CAG w transkrypcie człowieka. A. Liczba ciągów powtórzeń CAG występujących w różnych częściach mRNA; B. Liczba powtórzeń motywu CAG w ciągu; C. Aminokwasy kodowane przez powtórzenia CAG; D. Funkcje genów zawierających powtórzenia CAG w części kodującej (oznaczenia jak na rycinie 1.4).

1.7.2 Powtórzenia CGG.

Ciągi powtórzeń CGG są drugą co do wielkości grupą powtórzeń trójnukleotydowych występujących w transkrypcie człowieka. Pierwszą istotną różnicą w porównaniu z powtórzeniami CAG jest ich dystrybucja pomiędzy częściami funkcjonalnymi cząsteczek mRNA. Motywy CGG występują z podobną częstością w otwartej ramce odczytu oraz w regionie 5' nie ulegającym translacji (ryc. 1.6 A). Nie zaobserwowano ich w regionie 3'UTR, co wskazuje na to, że raczej nie pełnią funkcji w regulacji stabilności cząsteczek mRNA. Ciągi powtórzeń CGG, które występują w

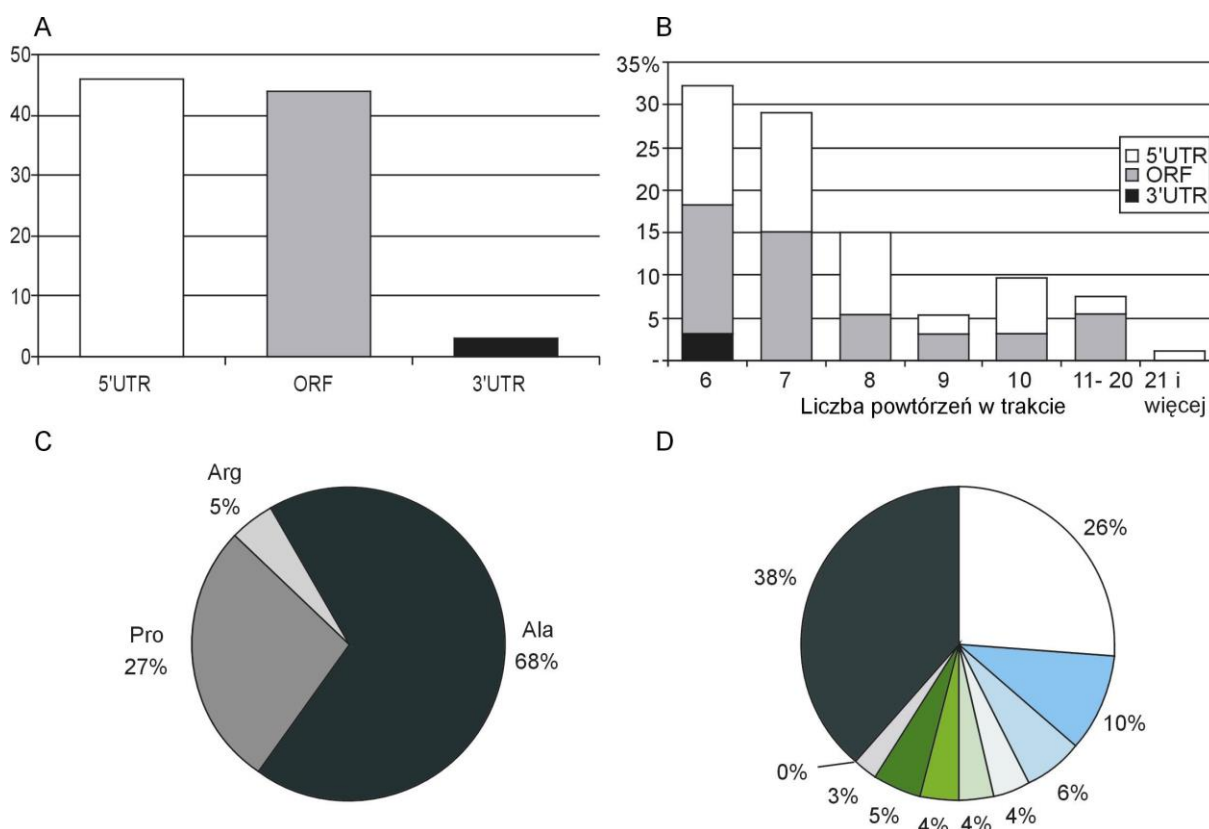
ORF kodują w zasadzie dwa rodzaje aminokwasów: alaninę i glicynę przy zaledwie 2% udziale argininy (ryc. 1.6 C). Wydaje się, że nawet krótkie trakty złożone z reszt argininowych występujące w białkach mogą mieć toksyczne właściwości, skoro kodujące je ciągi trójnukleotydomy praktycznie nie występują. Także analiza długości ciągów powtórzeń CGG wskazuje na silną presję skierowaną na wyeliminowanie z puli mRNA tych z nich, które zawierają większą liczbę powtórzeń tego motywu (ryc. 1.6 B). Pozwala to sądzić, że także dłuższe trakty tego rodzaju obecne w części 5'UTR mogą, poprzez przyjmowaną strukturę drugorzędową, mieć wpływ na obniżenie poziomu ekspresji zawierających je mRNA w sposób szkodliwy dla komórki. Tym bardziej, że wśród transkryptów, w których stwierdzono występowanie powtórzeń motywu CGG zaobserwowano wzrost liczby genów zaangażowanych w tak ważne dla homeostazy procesy translacji i dojrzewania RNA, sygnalizacji wewnątrzkomórkowej i międzykomórkowej oraz regulacji metabolizmu.



Rycina 1.6. Powtórzenia CGG w transkrypcie człowieka. A. Liczba ciągów powtórzeń CGG występujących w różnych częściach mRNA; B. Liczba powtórzeń motywu CGG w ciągu; C. Aminokwasy kodowane przez powtórzenia CGG; D. Funkcje genów zawierających powtórzenia CGG w części kodującej (oznaczenia jak na rycinie 1.4).

1.7.3 Powtórzenia CCG.

Powtórzenia CCG, podobnie jak CGG, występują w zbliżonych proporcjach w otwartej ramce odczytu i regionie 5' nie ulegającym translacji (ryc. 1.7 A). Kilka tego rodzaju traktów złożonych z 6 powtórzeń stwierdzono także w 3'UTR. Analiza długości ciągów zbudowanych z motywów CCG wykazała, że te zawierające więcej niż 20 powtórzeń obecne są tylko w 5'UTR. Pozostałe zakresy wielkości są obserwowane ze zbliżoną częstością zarówno w ORF jak i 5'UTR (ryc. 1.7 B). Taka lokalizacja ciągów o określonej długości pozwala sądzić, że pełnią one podobną do powtórzeń CGG rolę w regulacji procesu translacji na poziomie struktury mRNA. O podobnej roli traktów CCG i CGG występujących w części ORF świadczy częste kodowanie przez oba typy powtórzeń ciągów reszt alaninowych (ryc. 1.7 C).

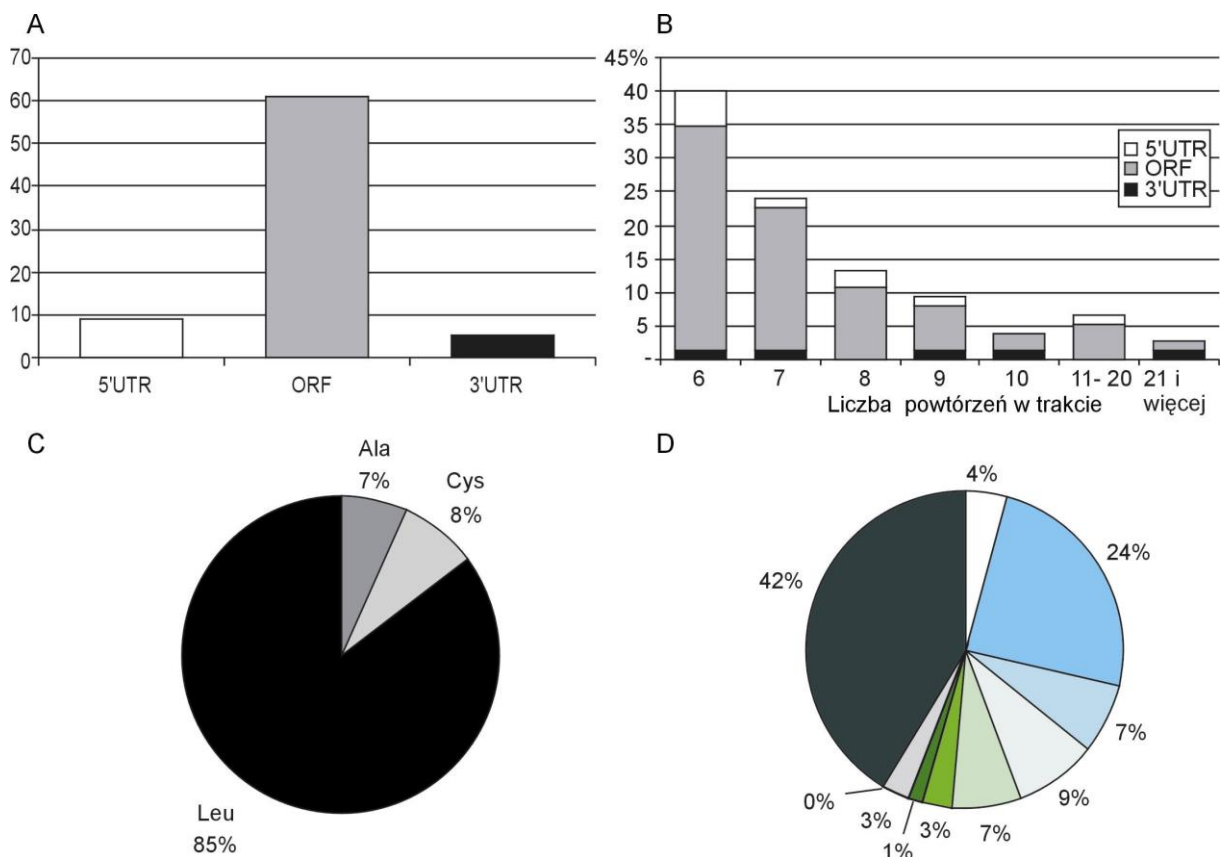


Rycina 1.7. Powtórzenia CCG w transkrypcie człowieka. A. Liczba ciągów powtórzeń CCG występujących w różnych częściach mRNA; B. Liczba powtórzeń motywu CCG w ciągu; C. Aminokwasy kodowane przez powtórzenia CCG; D. Funkcje genów zawierających powtórzenia CCG w części kodującej (oznaczenia jak na rycinie 1.4).

Analiza roli powtórzeń CCG z punktu widzenia funkcji pełnionych przez geny zawierające w swojej sekwencji ten typ powtórzeń trójnukleotydowych, poza potwierdzeniem znaczącego ich udziału w procesach syntezy i dojrzewania cząsteczek kwasu rybonukleinowego nie dostarczyła dodatkowych informacji o roli tego rodzaju sekwencji.

1.7.4 Powtórzenia CUG.

Powtórzenia CUG są najmniej liczną grupą powtórzeń typu CNG występujących w ludzkim transkryptomie. Długość ich ciągów nie odbiega od ogólnych tendencji obserwowanych w przypadku innych powtórzeń trójnukleotydowych obecnych w zidentyfikowanych mRNA (ryc. 1.8 B). Ciekawy jest fakt, że występujące głównie w otwartej ramce odczytu motywy CUG (ryc. 1.8 A) w



Rycina 1.8. Powtórzenia CUG w transkryptomie człowieka. A. Liczba ciągów powtórzeń CUG występujących w różnych częściach mRNA; B. Liczba powtórzeń motywu CUG w ciągu; C. Aminokwasy kodowane przez powtórzenia CUG; D. Funkcje genów zawierających powtórzenia CUG w części kodującej (oznaczenia jak na rycinie 1.4).

85% genów kodują leucynę (ryc. 1.8 C). Tylko w przypadku tego typu powtórzeń stwierdzono tak znaczną preferencję kodowania przez nie jednego rodzaju aminokwasu. Obserwacja ta, wraz z wyraźnym wzrostem udziału genów o funkcjach związanych z sygnalizacją międzykomórkową (ryc.1.8 D) w porównaniu do innych omawianych grup mikrosatelitów pozwala sądzić, że ciągi hydrofobowych reszt leucynowych, poza przyczynianiem się do przyjmowania przez łańcuch aminokwasowy struktury helisy α biorą też udział w oddziaływaniach białek z komórkowymi błonami lipidowymi.

1.8 Choroby neurologiczne człowieka powodowane ekspansją powtórzeń trójnukleotydowych.

Z sekwencjami mikrosatelitarnymi związany jest wyjątkowy rodzaj mutacji [Richards R. I., Sutherland G. R. (1992)] powodującej ponad 20 dziedzicznych chorób neurologicznych. Jest to, określana mutacją dynamiczną, obserwowana tylko u człowieka [Pearson C. E. i wsp.(2005)] niestabilność ciągów powtórzeń charakteryzująca się ekspansją liczby powtórzeń najczęściej trójnukleotydowych motywów przekazywanych z pokolenia na pokolenie. Zjawisko mutacji dynamicznych jest niezgodne z powszechnie uznawaną definicją mutacji, ponieważ prawdopodobieństwo wystąpienia kolejnej zmiany w regionie DNA wydłużonym w wyniku pierwszej mutacji jest większe od prawdopodobieństwa wystąpienia pierwszej mutacji. W procesie tym istotna jest pierwotna liczba powtórzeń w ciągu, ponieważ ciągi dłuższe łatwiej ulegają wydłużeniu niż trakty o mniejszej liczbie powtórzeń trójnukleotydowego motywu. Choroby wywoływane mutacjami dynamicznymi charakteryzują się ścisłym powiązaniem intensywności objawów i wieku ich pojawienia się u osób chorych z liczbą powtórzeń obecnych w odpowiedzialnym za ich wystąpienie ciągu mikrosatelitarnym. Zjawisko wcześniejszego pojawienia się objawów i ich nasilanie się w kolejnych pokoleniach obserwowane tylko w przypadkach chorób związanych z niestabilnością sekwencji STR nazywane jest antycypacją genetyczną [Harper P. S. i wsp. (1992)].

Wśród niestabilnych powtórzeń mikrosatelitarnych najistotniejszą, a zarazem największą grupę stanowią ciągi składające się z powtórzeń trójnukleotydowego motywu. Choroby wywoływane przez ich ekspansję określa się często wspólną nazwą TREDs (*Trinucleotide Repeat Expansion Diseases*) [Mitas M. i wsp. (1995)].

Po raz pierwszy mutację polegającą na wzroście liczby powtórzeń motywu w ciągach trójnukleotydowych opisano w 1991 roku po dokonaniu analizy genu *FMR1* związanego z zespołem łamliwego chromosomu X typu FRAXA [Kremer E. J. i wsp. (1991)]. W tym samym czasie stwierdzono, że mutacją powodującą rdzeniowo – opuszkowy zanik mięśni jest występowanie ekspansji powtórzeń trójnukleotydowych w genie *AR* [La Spada A. R. i wsp. (1991)].

W genach związanych z TREDs liczba powtórzeń trójnukleotydowych w niezmutowanych allelach wykazuje polimorfizm (tabela 1.2). Za mutację uważa się wydłużenie ciągu powyżej określonej dla każdego z genów granicy. W większości przypadków allele zmutowane, których obecność powoduje zmiany fenotypowe, są znacznie (nawet do kilku tysięcy powtórzeń) dłuższe od pierwotnych alleli.

Ciągi powtórzeń trójnukleotydowych ulegające ekspansji można podzielić na trzy grupy. Pierwszą z nich są ciągi złożone tylko z jednego rodzaju motywów. Do drugiej należą te, które są przerywane motywami innego typu. Trzecią zaś stanowią te, które nie są przerywane ale w ich bliskim sąsiedztwie występuje inny ciąg trójnukleotydowy nie ulegający wydłużaniu. Tylko w dwóch przypadkach wśród genów związanych z TREDs obserwowana jest ostatnia sytuacja. W genie *IT15*, związanym z chorobą Huntingtona, ciągowi powtórzeń CAG ulegającemu ekspansji towarzyszy wykazujący niewielki polimorfizm ciąg CCG. Natomiast w genie *AR* w pobliżu traktu CAG, w obrębie którego dochodzi do mutacji będącej przyczyną rdzeniowo – opuszkowego zaniku mięśni obecny jest drugi, krótki ciąg CAG.

Geny, w których obecne są ciągi trójnukleotydowe których zmiana długości jest odpowiedzialna za powstanie choroby, zawierają omawiane sekwencje w części kodującej i w regionach 5' lub 3' UTR. W przypadku jednej z chorób (ataksja Friedreicha) mutacją za nią odpowiedzialną jest wydłużenie ciągu występującego w intronie. Ze względu na lokalizację powtórzeń trójnukleotydowych w genie, choroby spowodowane ich ekspansją można podzielić na dwie podstawowe grupy:

- a) spowodowane ekspansją powtórzeń trójnukleotydowych występujących w niekodujących regionach genów; 6 dotąd poznanych zespołów chorobowych należących do tej grupy charakteryzuje się wystąpieniem dużych zmian długości ciągów trójnukleotydowych wywołujących różnorodne objawy fenotypowe związane z dysfunkcją lub degradacją różnych tkanek [Cummings C. J., Zoghbi H. Y. (2000)].

Choroba	Gen/locus	Lokalizacja chromosomowa	Lokalizacja powtórzeń	Typ powtórzeń	Zaburzenia sekwencji	Zakres liczby powtórzeń	
						Normalny	Mutacyjny
Choroba Huntingtona (HD)	<i>IT15</i>	4p16.3	ORF	CAG	-	6-35	41-200
HDL2	<i>JPH3</i>	16q24.3	ORF	CTG	-	6-39	≥41
Ataksja rdzeniowo-mózdkowa 1 (SCA1)	<i>SCA1</i>	6p23	ORF	CAG	CAT	6-39	39-91
Ataksja rdzeniowo-mózdkowa 2 (SCA2)	<i>SCA2</i>	12q24.1	ORF	CAG	CAA	14-30	36-200
Ataksja rdzeniowo-mózdkowa 3 (SCA3)	<i>SCA3 (MJD1)</i>	14q32.1	ORF	CAG	-	12-47	53-86
Ataksja rdzeniowo-mózdkowa 6 (SCA6)	<i>CACNA1A</i>	19p13	ORF	CAG	-	7-18	20-33
Ataksja rdzeniowo-mózdkowa 7 (SCA7)	<i>SCA7</i>	3p12-13	ORF	CAG	-	7-19	36-460
Ataksja rdzeniowo-mózdkowa 8 (SCA8)	<i>SCA8</i>	13q21	3'UTR	CTA/CTG	-	15-70	71->800
Ataksja rdzeniowo-mózdkowa 12 (SCA12)	<i>PPP2R2B</i>	5q31-33	5'UTR	CAG	-	4-32	51-78
Ataksja rdzeniowo-mózdkowa 17 (SCA17)	<i>TBP</i>	6q27	ORF	CAG	CAA	25-42	49-66
Zwyrodnienie zębato-czerwienne (DRPLA)	<i>DRPLA</i>	12p13.31	ORF	CAG	-	6-35	48-93
Rdzeniowo-opuszkowy zanik mięśni (SBMA)	<i>AR</i>	Xq13-21	ORF	CAG	-	11-34	38-66
Dystrofia miotoniczna 1 (DM1)	<i>DMPK</i>	19q13	3'UTR	CTG	CCG	5-37	50-3000
Zespół łamliwego chromosomu X (FXS)	<i>FMR1</i>	Xq27.3	5'UTR	CGG	AGG	6-58	59-200 (premutacja) 200-1000
Przedwczesna niewydolność jajników (POF)	<i>FMR1</i>	Xq27.3	5'UTR	CGG	AGG	6-58	59-200
Zespół drżenia i ataksji (FXTAS)	<i>FMR1</i>	Xq27.3	5'UTR	CGG	AGG	6-58	59-200
Opóźnienie umysłowe związane z FRAXE	<i>FMR2 (FRAXE)</i>	Xq28	5'UTR	GCC	-	7-35	130-150 230-750
Ataksja Friedreicha	<i>FRDA</i>	9q13-21.1	intron	GAA	GAG	6-34	80; 112-1700
Dystrofia mięśniowa, oczno-gardzielowa (OPMD)	<i>PABPN1</i>	14q11.2-q13	ORF	GCG	-	6	8-13 7
Achondroplazja rzekoma	<i>COMP</i>	19p13.1	ORF	GAC	-	5	7; 4

Tabela 1.2. Choroby związane z ekspansją powtórzeń trójnukleotydowych.

b) spowodowane ekspansją powtórzeń trójnukleotydomowych występujących w kodujących regionach genów. Cechą wspólną większości chorób należących do tej grupy jest fakt, że mutacje za nie odpowiedzialne występują w ciągach powtórzeń CAG, które kodują glutaminę. W związku z tym choroby te nazywane są też chorobami poliglutaminowymi. Rozmiary ekspansji powtórzeń trójnukleotydomowych są mniejsze niż we wcześniej omówionej grupie chorób. U pacjentów w średnim wieku pojawia się ostre upośledzenie funkcji układu nerwowego, którego przyczyną jest postępujące uszkodzenie neuronów. Na podstawie przeprowadzonych badań stwierdzono, że za patologiczne zmiany w komórkach prawdopodobnie jest odpowiedzialna spowodowana mutacją zmiana funkcji pełnionej przez białko [Zoghbi H. Y., Orr H. T. (1999), Gatchel J. R., Zoghbi H. Y. (2005)].

1.8.1 Patomechanizm chorób z grupy TREDs związanych z utratą funkcji przez białko.

Wspólny mechanizm powstawania przypisuje się chorobom powodowanym przez ekspansję powtórzeń w regionie 5' nie ulegającym translacji. Wydłużanie traktu CGG w 5' UTR genu *FMR1* związanego z najczęściej występującym spośród chorób TREDs zespołem łamliwego chromosomu X, powoduje obniżenie poziomu produktu ekspresji tego genu, jakim jest białko FMRP [Oostra B. A., Willemsen R. (2002)]. Gen *FMR1* jest skorelowany z kilkoma chorobami. Jeżeli liczba powtórzeń motywu CGG mieści się w zakresie 55 – 200 (premutacja) rozwijają się objawy przedwczesnej niewydolności jajników (POF) lub zespołu drżenia i ataksji (FXTAS) (tabela 1.2) [Hagerman R. J., Hagerman P. J. (2002), Hagerman P. J., Hagerman R. J. (2004)], natomiast dopiero przekroczenie 200 powtórzeń CGG jest przyczyną zespołu łamliwego chromosomu X. Zmiany poziomu ekspresji białka FMRP są ściśle skorelowane z długością ciągu powtórzeń trójnukleotydomowych w regionie 5'UTR jego mRNA. Liczne badania wskazują na znaczne podwyższenie w komórce poziomu mRNA genu *FMR1* zawierającego wydłużony ciąg powtórzeń CGG u nosicieli premutacji chorych na FXTAS [Tassone F. i wsp. (2000), Kenneson A. i wsp. (2001)]. Jednocześnie obniżony pozostaje poziom białka FMRP co pozwala sądzić, że zaburzenie translacji, będące skutkiem obecności wydłużonego traktu CGG, indukuje

wzrost intensywności transkrypcji jako sposób na zachowanie wystarczającego poziomu białka w komórce [Primerano B. i wsp. (2002), Chen L. S. i wsp. (2003)]. W sytuacji, gdy ulegający ekspansji ciąg powtórzeń CGG osiągnie długość mutacyjną dochodzi do całkowitego zahamowania translacji obserwowanej z zespole łamliwego chromosomu X (FXS). Zakłada się, że zjawisko to jest wynikiem zatrzymania rybosomu na tworzonej przez nie strukturze [Napierała M. i wsp. (2005)] w czasie przesuwania się jego małej podjednostki w poszukiwaniu kodonu start. Potwierdzają to obserwacje polirybosomów związanych z mRNA *FMR1* [Feng Y. i wsp. (1995), Primerano B. i wsp. (2002)].

Białko FMRP, zlokalizowane głównie w cytoplazmie komórek nerwowych, pełni rolę regulatora translacji poprzez wiązanie się do polirybosomów w sposób zależny od mRNA [Corbin F. i wsp. (1997)]. Zidentyfikowano wiele mRNA, z którymi FMRP oddziałuje [Brown V. i wsp. (2001)]. Są to mRNA genów mających istotne znaczenie dla wielu procesów komórkowych, takich jak regulacja stabilności cytoszkieletu – białko 1B wiązane z mikrotubulami (MAB1B) czy innych przemian w czasie synaptycznego przekazywania sygnału pomiędzy neuronami [Zhang Y. Q. i wsp. (2001), Xu K. i wsp. (2004)]. Wykazano, że FMRP wiąże mRNA zawierające sekwencję RGG tworzącą motyw strukturalny tetrapleksu „G – quartet” [Darnell J. C. i wsp. (2001)]. Ponadto stwierdzono możliwość wiązania FMRP do innego motywu w mRNA nazwanego *FMRP kissing complex* co zwiększa grupę genów, których ekspresja zależy od tego czynnika [Darnell J. C. i wsp. (2005)]. Innym możliwym procesem regulacji translacji, w którym uczestniczy FMRP jest szlak RNAi. Białko to wiąże się z kluczowymi dla tego procesu czynnikami, którymi są białka Argonaute i kompleks RISC (*RNA – Induced Silencing Complex*) [Gatchel J. R., Zoghbi H. Y. (2005), Caudy A. A. i wsp. (2002), Ishizuka A. i wsp. (2002)]. Zaobserwowano także współwystępowanie FMRP z nukleazą *Dicer* [Jin P. i wsp. (2004)]. Proponowana rola FMRP w szlaku RNAi miałyby polegać na selektywnym wiązaniu cząsteczek RNA i zbliżaniu ich do kompleksu RISC, a tym samym na ułatwianiu mikroRNA specyficznej kontroli translacji ich docelowych cząsteczek [Jin P. i wsp. (2004)].

Tak szerokie spektrum działania FMRP jako czynnika regulującego translację wskazuje na istotne znaczenie jego poziomu dla prawidłowego funkcjonowania komórek nerwowych. Obniżenie jego ilości w wyniku wystąpienia mutacji dynamicznej w ciągu powtórzeń CGG w 5'UTR jego mRNA jest więc podstawową przyczyną opisanych dotąd zmian w zespole łamliwego chromosomu X (FXS).

Należy się spodziewać, że dalsze badania dostarczą informacji o kolejnych dotąd nie poznanych skutkach tej mutacji.

W przypadku genu *FMR2* powtórzenia CCG wydłużane w wyniku mutacji również zlokalizowane są w 5'UTR. Ulegają one hipermetylacji wraz z sąsiadującą wyspą CpG co prowadzi do zahamowania transkrypcji tego genu [Knight S. J. i wsp. (1993), Gecz J. i wsp. (1996), Gu Y. i wsp. (1996)]. Obniżenie komórkowego poziomu białka FMR2 należącego do rodziny białek bogatych w prolinę i serynę prowadzi do powstania choroby nazwanej opóźnieniem umysłowym związanym z FRAXE. Białko FMR2 określa się jako czynnik transkrypcyjny istotny dla procesów rozwoju i różnicowania układu nerwowego [Gecz J. i wsp. (1997)]. Na podstawie badań nad rozwojem embrionalnym *Drosophila melanogaster* przypuszcza się że FMR2 pełni istotną funkcję w regulacji szlaku MAPK (kinazy białkowej aktywowanej mitogenem). Podobna do obserwowanej w przypadku FRAXE hipermetylacja ma również miejsce w obrębie ciągu CGG w genie *FMR1* i także ten proces odgrywa ważną rolę w patogenezie zespołu łamliwego chromosomu X (FXS) [Oberle I. i wsp. (1991)].

Ataksja Friedreicha jest najpowszechniej występującą dziedziczną ataksją. Jej molekularnym podłożem jest ekspansja powtórzeń GAA w pierwszym intronie genu *FRDA* kodującego białko frataksynę związane z wewnętrzną błoną mitochondrialną [Campuzano V. i wsp. (1996), Campuzano V. i wsp. (1997)]. Podobnie jak w wyżej opisanych przypadkach w wyniku elongacji ciągu powtórzeń dochodzi do zahamowania transkrypcji, co prowadzi do obniżenia poziomu białka. Brak frataksyny w mitochondriach powoduje gromadzenie się w nich żelaza i tym samym wzrost podatności na stres oksydacyjny zaburzający działanie wielu białek mitochondrialnych łańcucha oddechowego, takich jak akonidaza [Foury F., Cazzalini O. (1997)]. Przypuszcza się, że proces ten nie stanowi głównego szlaku patogenezy. Nie znany jest jednak dotąd inny proces komórkowy, w którego przebiegu białko genu *FRDA* pełniłoby jakąś funkcję [Seznec H. i wsp. (2005)]. Na podstawie obserwacji, że myszy całkowicie pozbawione funkcji tego białka giną w trakcie rozwoju embrionalnego stwierdzono, że pełni ono rolę w rozwoju organizmu [Cossee M. i wsp. (2000)].

W przypadku wszystkich chorób należących do grupy TREDs, którym przypisywany jest mechanizm patogenezy związany z utratą funkcji przez białko dochodzi w wyniku mutacji do zahamowania ekspresji danego genu na poziomie

transkrypcji lub translacji. Brak białka powoduje zaburzenie szeregu procesów komórkowych, jednakże określenie dokładnego przebiegu rozwoju choroby wymaga poznania prawidłowej funkcji genów tej grupy.

1.8.2 Patomechanizm chorób poliglutaminowych na przykładzie choroby Huntingtona.

Choroby poliglutaminowe to te spośród chorób TREDs, które powodowane są ekspansją powtórzeń CAG zlokalizowanych w otwartej ramce odczytu (tabela 1.2). Do tej grupy należy dziewięć chorób. Są to: choroba Huntingtona (HD), zwyrodnienie zębato – czerwienne (DRPLA), opuszkowo – rdzeniowy zanik mięśni (SBMA) oraz ataksje mózdkowo – rdzeniowe (SCA) typu 1, 2, 3, 6, 7 i 17. Ich wspólną cechą jest ekspresja w tkankach centralnego układu nerwowego genów, w których dochodzi do mutacji dynamicznej. Zróżnicowanie neurodegeneracyjnych objawów charakterystycznych dla poszczególnych chorób wynika z faktu, że w każdej z nich zaburzeniu podlegają funkcje różnych grup neuronów. W przypadku tej grupy chorób mutacja prowadząca do wzrostu liczby powtórzeń motywu CAG w odpowiednim genie powoduje na poziomie białka wydłużenie ciągu reszt glutaminowych (Q) kodowanych przez kodon CAG. Proponowany mechanizm patogenezy chorób poliglutaminowych zakłada, że zaburzenie procesów komórkowych spowodowane jest zmianą funkcji pełnionych przez prawidłowe białko [Zoghbi H. Y. i Orr H. T. (2000)]. Obserwowane na przykład w genie *SCA2* zaburzenia ciągłości traktu powtórzeń CAG nie powodują zmiany kodowanego aminokwasu, wydają się nie mieć więc znaczenia dla przebiegu zmian chorobowych.

Jedną z najczęstszych i najintensywniej badanych chorób poliglutaminowych jest choroba Huntingtona (HD) występująca w populacji kaukaskiej z częstością około 1 na 10 tysięcy [Nance M. A. (1998)]. Jej pierwsze objawy występują zwykle u pacjentów między 40 a 50 rokiem życia [Paulson H. L., Fischbeck K. H. (1996)]. HD prowadzi do śmierci po około 20 latach choroby. Charakterystyczny dla niej zanik neuronów w centralnym układzie nerwowym objawia się postępującym upośledzeniem ruchu i zdolności postrzegania otoczenia oraz zaburzeniami osobowości. Stwierdzono również przypadki charakteryzujące się wcześniejszym pojawieniem się pierwszych symptomów choroby (forma juvenilna) w pierwszym dwudziestolecu życia i jej ostrzejszym przebiegiem [Andrew S. E. i wsp. (1993)]. HD

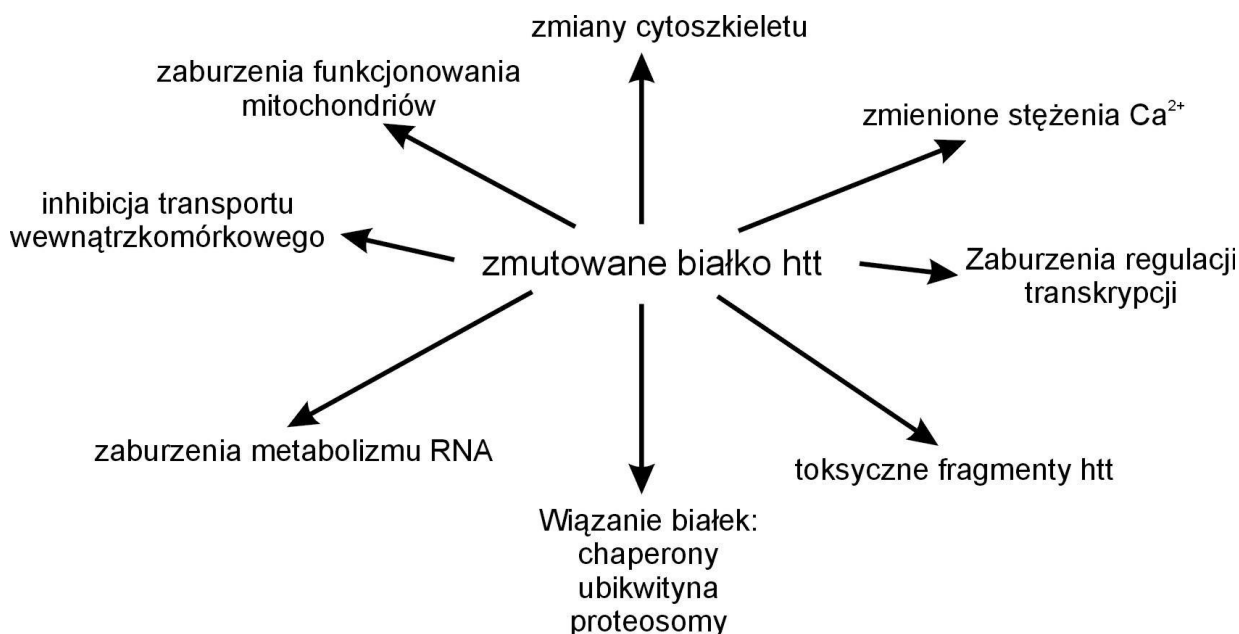
jest chorobą dziedziczną w sposób autosomalny, dominujący [Mitas M. i wsp. (1995)].

Choroba Huntingtona powodowana jest przez wzrost liczby powtórzeń CAG występujących w 1 eksonie genu *IT15*. W przypadku tego genu interesujący jest fakt występowania w bezpośrednim sąsiedztwie dwóch polimorficznych ciągów trójnukleotydowych w obrębie sekwencji kodującej. Ulegający patogennym ekspansjom ciąg CAG leży 12 nukleotydów powyżej słabo polimorficznego ciągu powtórzeń CCG. Liczba powtórzeń CCG waha się od 6 – 12. Natomiast powtórzenia CAG wykazują większą zmienność. Allele normalne zawierają mniej niż 26 powtórzeń motywu CAG. Allele z 27 – 35 powtórzeniami, mimo iż są niestabilne mejotycznie, także uznawane są za niezmutowane. Wśród form powstałych w wyniku mutacji wyróżniamy allele ze zredukowaną (36-40) i pełną penetracją (≥ 41) [Potter N. T. i wsp., (2004)]. Liczba powtórzeń CAG w wariacie mutacyjnym wpływa na nasilenie się objawów i czas ich wystąpienia. Im ciąg powtórzeń jest dłuższy tym pierwsze objawy HD pojawiają się wcześniej i choroba postępuje szybciej.

Gen *IT15* zlokalizowany w pozycji 4q16.3 [Gusella J. F. i wsp. (1983)], zawiera 67 exonów kodujących białko huntingtynę o wielkości 3144 aminokwasów i masie 348 kDa charakteryzujące się obecnością w pobliżu końca aminowego polimorficznej domeny bogatej w reszty glutaminy i proliny [Huntington's Disease Collaborative Research Group. (1993)]. Funkcja prawidłowej huntingtyny (htt), chociaż intensywnie badana, pozostaje ciągle nieznana pomimo, że jej sekwencja zawiera wiele regionów homologicznych do innych znanych białek. htt ulega konstytutywnej ekspresji w wielu tkankach i jest zlokalizowana w jądrze komórkowym oraz części cytoplazmatycznej neuronów gdzie towarzyszy aparatom Golgiego, retikulum endoplazmatycznemu i mikrotubulom [Harjes P., Wanker E. E. (2003), Li S. H. i Li. X.I. (2004)]. Na podstawie poznanych oddziaływań z innymi białkami kompleksu dynaktyny i β -tubuliny przypuszcza się, że może ona być czynnikiem zaangażowanym w pęcherzykowy transport wewnątrzkomórkowy i procesy zakotwiczenia cytoszkieletu. Może także odgrywać pewną rolę w zależnej od klatryny endocytozie oraz sygnalizacji postsynaptycznej. Stwierdzono również, że huntingtyna może chronić neurony przez apoptozą [Rigamonti D. i wsp. (2000)].

Zmutowana huntingtyna, podobnie jak inne zmutowane białka zawierające wydłużone ciągi poliQ tworzy w neuronach nierozpuszczalne agregaty białkowe charakterystyczne dla chorób poliglutaminowych. Ich rola w patogenezie nie została

dotąd jednoznacznie potwierdzona [Bates G. (2003)]. W niektórych chorobach obserwowane są, tak jak w przypadku HD i SBMA skupiska białka zlokalizowane zarówno w jądrze jak i w cytoplazmie, natomiast w innych tylko złoży jądrowe. Poza cząsteczkami białka o pełnej długości w skupiskach znajdują się duże ilości fragmentów białek zawierających trakty poliglutaminowe. N-końcowe fragmenty htt powstają na drodze obróbki białka pełnej długości przez proteazy takie jak kaspaza 3, kalpaina czy nie zidentyfikowane proteazy aspartylowe [Gafni J., Ellerby L. M. (2002)., Wellington C. L. i wsp. (2002)]. Fragmenty te mogą uczestniczyć w procesach patologicznych w podobnym stopniu co białka pełnej długości. W inkluzjach stwierdzono obecność białek chaperonowych odpowiedzialnych za naprawę błędnie sfałdowanych peptydów oraz komponentów proteasomu związanego z zależnym od ubikwityny systemem degradacji białek (UPS – *Ubiquitin – Proteasome System*). Akumulacja zmutowanych białek jest postulowaną przyczyną zaburzeń funkcjonowania systemu UPS [Ciechanover A., Brundin P. (2003), Sakahira H. i wsp. (2002)]. W badaniach *in vitro* stwierdzono [Venkatraman P. i wsp. (2004)], że eukariotyczne proteasomy nie degradują ciągów poliglutaminowych co potwierdza wcześniejszą hipotezę.



Rycina 1.9. Mechanizm patogenezy chorób poliglutaminowych na przykładzie choroby Huntingtona – skutki pojawienia się zmutowanego białka w komórce.

W chorobie Huntingtona, podobnie jak i w innych chorobach poliglutaminowych obserwowano zaburzenia transkrypcji wielu genów. Stwierdzono oddziaływanie białek poliQ z czynnikiem CREB i jego gromadzenie w inkluzjach co prowadziło do obniżenia ekspresji genów zależnych od CREB [Glass M. i wsp. (2000), Sugars K. L. i Rubinsztein D. C. (2003)]. W ten sam sposób wiązane są także inne czynniki transkrypcyjne takie jak TBP (*TATA box binding protein*) czy p53 [Steffan J. S. i wsp. (2000), Dunah A. W. i wsp. (2002)]. Jednocześnie zaobserwowano, że inhibicja transkrypcji następuje także w komórkach, w których nie są jeszcze obecne agregaty białkowe. Stwierdzono, że wynika ona z obecności w środowisku rozpuszczalnych form białek poliQ [Dunah A. W. i wsp. (2002)]. Zmiany ich struktury wywołane wydłużeniem ciągu glutaminowego ułatwiają wiązanie innych białek zawierających takie same motywy sekwencji aminokwasowej, wśród których zidentyfikowano czynniki transkrypcyjne TBP i AR. Geny, odpowiednio *SCA17* i *AR* kodujące oba te białka zawierają ciągi CAG ulegające ekspansji [Nakamura K. i wsp. (2001), La Spada A. R. i wsp. (1991)]. Ponadto huntingtyna może poprzez oddziaływanie z jądrowymi kompleksami represorów transkrypcji pełnić tego rodzaju rolę w komórce. Dotychczasowe badania wyraźnie wskazują na udział inhibicji syntezy mRNA w procesach prowadzących do rozwoju patologicznego fenotypu chorób poliglutaminowych. Także potranskrypcyjne etapy metabolizmu RNA mogą ulegać zmianom w wyniku oddziaływania z białkami zawierającymi trakty poliQ. Stwierdzono, że ataksyna 1 może wiązać się do RNA i brać udział w jego dojrzewaniu [Fernandez – Funez P. i wsp. (2000)]. Ponadto zaobserwowano, że ataksyna 2 wiąże się z białkiem A2BP1 (*ataxin 2 binding protein 1*), które uczestniczy w składaniu mRNA [Shibata H. i wsp. (2000)].

Białka zawierające wydłużone trakty poliQ mogą również negatywnie wpływać na inne procesy komórkowe przebiegające w cytoplazmie. Przykładem może być blokada transportu wzdłuż aksonu stwierdzona w obecności zmutowanej huntingtyny bądź fragmentów receptora androgeny [Gunawardena S. i wsp.(2003), Szebenyi G. i wsp. (2003)]. W badaniach nad zmianami zachodzącymi w komórce w wyniku pojawienia się zmutowanego białka huntingtyny zaobserwowano także zaburzenia działania mitochondriów [Panov A. V. i wsp. (2002)]. Wynikają one z zaangażowania formy htt zawierającej wydłużony trakt poliQ w procesy przekazywania sygnału zależne od wapnia [Tang T. S. i wsp. (2005)].

Stwierdzono również, że wspomniane wcześniej fragmenty białek poliglutaminowych nie tylko uczestniczą w procesach patogenezy chorób na równi z białkami o pełnej długości ale wywierają dodatkowy efekt. Ich toksyczność, szczególnie badana na przykładzie choroby Huntingtona, wynika ze zdolności do indukcji autofagii [Ravikumar B. i wsp. (2004)], czyli procesu w którym dochodzi do zniszczenia przez komórkę białek o długim czasie życia oraz jej organelli. W tworzonych przez trakty poliQ agregatach wiązany jest czynnik mTOR hamujący powstawanie autofagosomów, co prowadzi do obniżenia jego efektywnego stężenia w cytoplazmie, co z kolei prowadzi do indukcji procesu lizy organelli komórkowych a w konsekwencji do apoptozy.

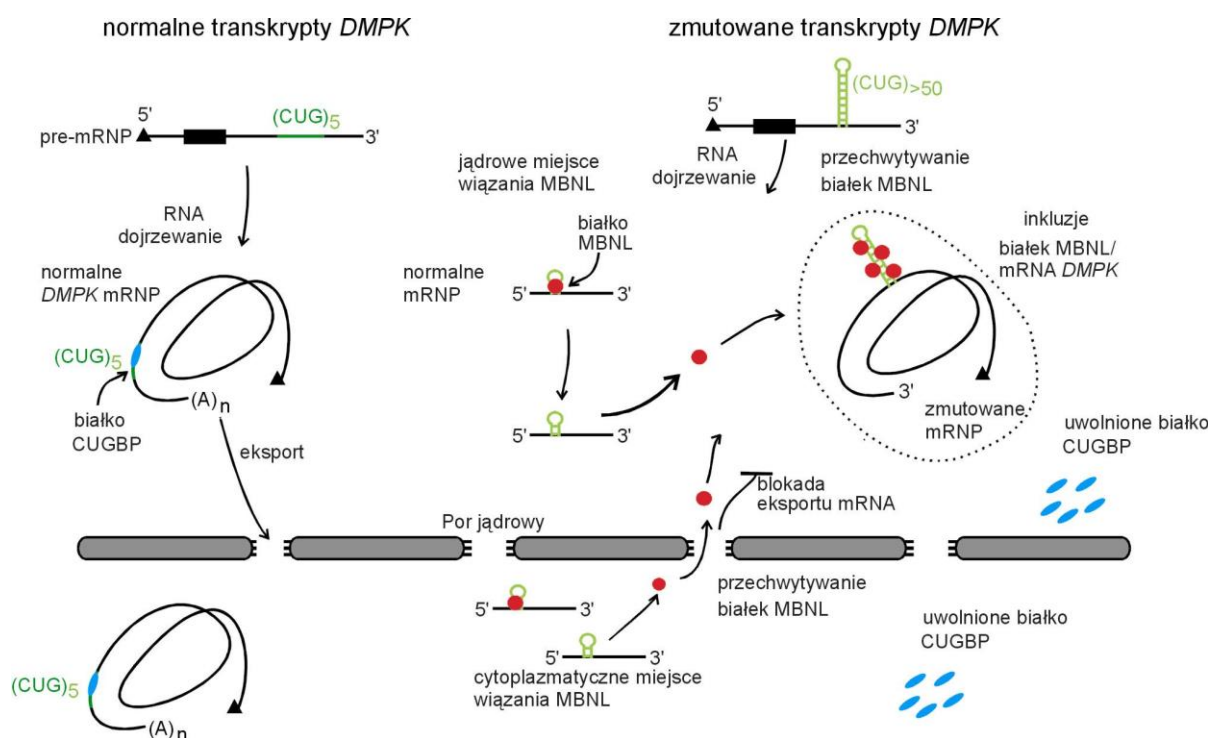
Dotychczasowe badania patomechanizmu chorób z grupy TREDs, w których dochodzi do ekspansji powtórzeń CAG w otwartej ramce odczytu, ukazały złożoność tego procesu wynikającą z uzyskania przez białko nowych właściwości wskutek wydłużenia ciągu reszt glutaminowych. Zaobserwowano także, że istotne znaczenie dla rozwoju choroby mają modyfikacje białek w innych regionach niż trakt poliQ. Wysoką toksyczność wykazują na przykład fosforylacja jednej z seryn w ataksynie 1 [Emamian E. S. i wsp. (2003)] oraz przyłączenie do huntingtyny jednostki SUMO (*small ubiquitin – like modifier*) [Steffan J. S. i wsp. (2004)], która bierze udział w regulacji oddziaływania htt z innymi makrocząsteczkami komórki. Zauważyć należy, że nadal istnieje grupa objawów, których nie można wyjaśnić zmianami na poziomie białka [Peel A. L. i wsp. (2001)].

1.8.3 Patomechanizm zakładający rolę RNA zawierającego wydłużone ciągi powtórzeń trójnukleotydowych.

Patogenezy niektórych chorób z grupy TREDs nie można wyjaśnić zmianą funkcji kodowanego białka, ponieważ ulegające ekspansji ciągi powtórzeń zlokalizowane są w odpowiednich genach w regionach nie ulegających translacji. Nie podlegają też one mechanizmowi, który obejmuje utratę funkcji przez białko. Są to dystrofie miotoniczne typu pierwszego (DM1) i drugiego (DM2) o bardzo podobnych objawach. Dystrofię miotoniczną typu 1 powoduje wzrost liczby powtórzeń trójnukleotydowego motywu CTG w regionie 3' UTR genu *DMPK (Dystrophia Myotonica Protein Kinase)* [Brook J. D. i wsp. (1992), Fu Y. H. i wsp. (1992), Mahadevan M. i wsp. (1992)], natomiast dystrofię miotoniczną typu 2 wydłużenie

traktu czteronukleotydowego CCTG występującego w intronie 1 genu *ZNF9* [Ranum L. P. i wsp. (1998), Liquori C. L. i wsp. (2001)]. W toku badań nad komórkowymi zmianami wywołanymi obecnością wydłużonego ciągu powtórzeń w regionie 3' UTR umieszczono sekwencję 250 powtórzeń motywu CUG w 3' UTR mRNA aktywnej. U tak stransformowanych myszy zaobserwowano rozwój niektórych objawów charakterystycznych dla dystrofii miotonicznej [Mankodi A. i wsp. (2000)]. Obserwacja ta połączona z faktem, że takie same zmiany wywołują mutacje w różnych genach, zlokalizowanych na różnych chromosomach i mające różne otoczenie genowe, nasunęła przypuszczenie, że mechanizm patogenezy polega na zaburzonym, w wyniku ekspansji powtórzeń CUG lub CCUG, wiązaniu czynników białkowych do RNA, powodując zmianę ich dystrybucji w środowisku komórkowym i uniemożliwiając im pełnienie prawidłowej funkcji [Tapscott S. J. i Thronton C. A. (2001), Ranum L. P., i Day J. W. (2004)]. Zidentyfikowano dwie rodziny białek wiążących RNA wykazujących zdolność wiązania powtórzeń CUG i CCUG. Jedną z nich są białka CELF (*CUG-BP and ETR3-Like Faktor*). Drugą stanowią białka MBNL (*muscleblind – like*), które zlokalizowano w jądrowych agregatach tworzonych przez transkrypty zawierające wydłużone ciągi powtórzeń CUG [Miller J. W. i wsp. (2000), Fardaei M. i wsp. (2002)]. Wykorzystując myszy transgeniczne stwierdzono, że skutkiem braku białek MBNL, jest zaburzenie procesów alternatywnego składania genów [Kanada R. N. i wsp. (2003)]. Wykazano, że podobny efekt obserwowany jest także gdy ich ilość w środowisku komórkowym uległa znacznemu zmniejszeniu w wyniku związania ze zmutowanymi RNA. Natomiast białka CUGBP (*CUG binding protein*) nie są gromadzone w jądrowych inkluzjach. W przeciwieństwie do białek MBNL w przypadku białka CUGBP obserwuje się jego podwyższony poziom w obecności zmutowanych ciągów powtórzeń CUG [Timchenko L. T. (1999)]. Wzrost poziomu ekspresji białek CUGBP powoduje charakterystyczne dla DM1 zmiany w procesach dojrzewania mRNA na etapie wycinania intronów. Zidentyfikowano kilkanaście pre-mRNA, których składanie w komórkach pacjentów DM1 jest nieprawidłowe. Wśród nich najważniejsze wydają się transkrypty kanału chlorkowego (CIC1), sercowej troponiny T i receptora insulinowego [Philips A. V. i wsp. (1998), Mankodi A. i wsp. (2002), Savkur R. S. i wsp. (2001), Jiang H. i wsp. (2004)], gdyż ze zmienioną ekspresją tych genów wiązane są zaburzenia pracy serca i niewrażliwość na insulinę, które są typowe u pacjentów chorych na dystrofię miotoniczną. Zaobserwowano także zmiany składania pre-mRNA w obrębie

centralnego układu nerwowego. Dotyczą one mRNA receptora NR1, którego ligandem jest NMDA (*N – methyl – D – aspartate*) oraz mRNA genu *Tau* – neuronalnego białka towarzyszącego mikrotubulom [Jiang H. i wsp. (2004)]. Zakłada się, że zmiany te powstają w wyniku mechanizmu wiązania białek do wydłużonych RNA traktów CUG. Na podstawie badań roli białek CELF i MBNL w procesie patogenezy dystrofii miotonicznych stwierdzono, że w prawidłowych warunkach działają one antagonistycznie względem siebie i z tego powodu komórkach zawierających zmutowane ciągi trójnukleotydom dochodzi do zaburzenia równowagi, będącej konsekwencją związania jednego z nich przez zmutowane cząsteczki RNA. Poza zróżnicowaniem funkcjonalnym białek MBNL i CELF stwierdzono różnice w charakterystyce ich wiązania do powtórzeń CUG w RNA. Wykazano, że krótkie ciągi powtórzeń CUG w strukturze transkryptów występują w formie pojedynczoniowej, zaś zawierające więcej niż 10 powtórzeń tworzą strukturę typu spinki, której stabilność rośnie wraz z liczbą powtórzeń w ciągu [Napierała M., Krzyżosiak W. J. (1997)]. Białka MBNL wiążą się do długich ciągów proporcjonalnie do ich długości [Miller J. W. i wsp. (2000)], są zatem specyficzne dla podwójnoniciowych struktur tworzonych przez ciągi powtórzeń. W wyniku ekspansji



Rycina 1.10. Mechanizm patogenezy dystrofii miotonicznej typu 1 oparty na zmianie struktury RNA regionu powtórzeń CUG. Na podstawie [Miller J. W. i wsp. (2000)] zmienione.

powtórzeń trzon struktury spinki ulega wydłużeniu i wiąże więcej białek MBNL, co prowadzi do konfiskowania ich tym pre-mRNA, które stanowią prawidłowy cel działania tej grupy białek. W konsekwencji dochodzi do niewłaściwego składania pewnych mRNA i zmiany funkcji kodowanych białek. Białka CELF wiążące się do krótkich, pojedynczonicowych ciągów CUG w prawidłowych komórkach pozostają związane z mRNA genu *DMPK* [Timchenko N. A. i wsp. (2001)]. Natomiast w wyniku zmiany struktury regionu mRNA wywołanej zwiększeniem liczby powtórzeń w ciągu dochodzi do uwolnienia do środowiska komórki większej ilości białek CELF i wzrostu ich aktywności jako czynnika odpowiedzialnego za składanie pre-mRNA. Należy pamiętać, że jednocześnie dochodzi do związania białek MBNL.

W przypadku dystrofii miotonicznych sekwencje powtarzające się zlokalizowane są w regionach nie ulegających translacji. Przyjmowana przez nie struktura drugorzędowa pełni funkcję regulatora dojrzewania mRNA poprzez tworzenie miejsca wiązania białek biorących udział w tym procesie. Zmiany tej struktury na skutek wydłużenia ciągu powtórzeń prowadzą do przyłączenia białek o specyficzności do regionów podwójnoniciowych oraz do uwolnienia białek dotąd związanych z regionem pojedynczonicowym do środowiska komórki. Zauważyć trzeba, że istotne znaczenie ma w tej sytuacji stabilność struktury spinki tworzonej przez powtórzenia CUG. Dla spinek tworzonych przez nie zmutowane ciągi składające się z 10 do 20 powtórzeń obserwowana jest równowaga pomiędzy wariantami struktury. Obserwowane są takie formy strukturalne regionu powtórzeń, w których wszystkie powtórzenia zaangażowane są w tworzenie struktury spinki oraz takie, w których część sekwencji powtarzającej się pozostaje w formie pojedynczonicowej towarzyszącej „quasi-stabilnej” strukturze spinki po stronie 3' [Napierała M., Krzyżosiak W. J. (1997)]. Może zatem jednocześnie dochodzić do wiązania białek do regionów pojedynczo- i podwójnoniciowych co pozwala zachować ich równowagę przy pewnym zakresie polimorficznych zmian długości ciągu powtórzeń.

Dystrofie miotoniczne nie są jedynymi spośród chorób TREDs dla których proponuje się udział RNA w patomechanizmie. W przypadku zespołu drżenia i ataksji FXTAS oraz przedwczesnej niewydolności jajników POF związanych ze wzrostem, do zakresu premutacyjnego (60 – 200), liczby powtórzeń CGG w regionie 5' nie ulegającym translacji genu *FMR1* zaobserwowano dochodzący nawet do dziesięciokrotnego wzrost poziomu jego mRNA przy jednoczesnym obniżeniu

poziomu białka FMRP do około połowy w stosunku do komórek prawidłowych [Tassone F. i wsp. (2000)]. W neuronach i astrocytach mózgu osób chorych na FXTAS stwierdzono obecność inkluzji wewnątrzjądrowych [Greco C. M. i wsp. (2002) Tassone F. i wsp. (2004)], w których znaleziono RNA transkryptów *FMR1* [Tassone F. i wsp. (2004)]. Te podobieństwa do zmian obserwowanych w przypadku DM 1 i DM 2 oraz pojawiające się doniesienia o odkryciu białek wiążących się do powtórzeń CGG nasuwają przypuszczenie o możliwości istnienia podobnego mechanizmu powstawania zmian patologicznych w tych chorobach [Gatchel J. R., Zoghbi H. Y. (2005)].

2 CEL PRACY

Ciągi powtórzeń trójnukleotydomowych występują w wielu różnych ludzkich mRNA, a w ponad 600 z nich liczba powtórzonych tripletów wynosi 6 i więcej. Rola tego rodzaju sekwencji jest słabo poznana. Wydaje się, że niektóre z nich mogą pełnić funkcje regulatorowe stanowiąc miejsce wiązania rozmaitych czynników białkowych. Znaczne wydłużanie się ciągów powtórzeń trójnukleotydomowych stanowi podłoże molekularne chorób neurologicznych z grupy TREDs. Jedną z hipotez dotyczących patomechanizmów tych chorób zakłada, że zmiana struktury regionu powtórzeń trójnukleotydomowych w RNA powoduje nabycie przez nie toksycznych właściwości, polegających na nadmiernym wiązaniu niektórych czynników białkowych niezbędnych do prawidłowej regulacji procesów komórkowych.

Celem badań opisanych w niniejszej pracy doktorskiej było scharakteryzowanie struktury powtórzeń trójnukleotydomowych we fragmentach mRNA genu *IT15*, w kontekście potencjalnej jej roli w patogenezie choroby Huntingtona. Cel ten realizowano w następujących etapach:

1. zbadanie struktur RNA tworzonych przez powtórzenia CAG i CCG;
2. określenie wpływu jaki wywiera na strukturę ciągu powtórzeń CAG znajdujący się w bezpośrednim sąsiedztwie ciąg powtórzeń CCG;
3. określenie natury oddziaływania ciągów powtórzeń CAG i CCG;
4. zbadanie wpływu zmian liczby powtórzeń trójnukleotydomowych obu ciągów na strukturę fragmentu mRNA genu *IT15*.

Drugim celem niniejszej pracy było . Zrealizowano go w następujących etapach:

1. zbadanie struktury fragmentu mRNA genu *AR* zawierającego dwa ciągi powtórzeń trójnukleotydu CAG z uwzględnieniem wpływu zmian liczby powtórzeń w polimorficznym ciągu CAG;
2. poszukiwanie mRNA zawierających kilka ciągów powtórzeń trójnukleotydomowych i bioinformatyczna analiza porównawcza przyjmowanych przez nie struktur.

3 WYNIKI

3.1 Powtórzenia CAG i CCG tworzą strukturę typu spinki.

Pierwszym etapem badań prowadzących do poznania struktury regionów zawierających powtórzenia trójnukleotydydowe w mRNA genów *IT15* i *AR* było poznanie struktur tworzonych przez ciągi powtórzeń CAG i CCG. W tym celu zaprojektowano modelowe cząsteczki RNA składające się z powtórzeń pozbawionych naturalnych sekwencji oskrzydlających. Na podstawie dostępnych danych literaturowych i modelowania komputerowego zdecydowano się na analizę cząsteczek zawierających 16 i 17 powtórzeń CAG i CCG, ponieważ były to jedne z najkrótszych traktów zdolnych do tworzenia struktury typu spinki do włosów dające możliwość obserwacji zmian jej architektury. Poza ciągiem powtórzeń w sekwencji cząsteczek modelowych na końcu 5' znajdowały się dwie reszty guanylowe. Ich obecność wynikała z wymagań sekwencji promotora polimerazy RNA faga T7 używanej do reakcji transkrypcji *in vitro*, w której na matrycy oligonukleotydów DNA syntetyzowanych chemicznie uzyskiwano transkrypty do badań strukturalnych.

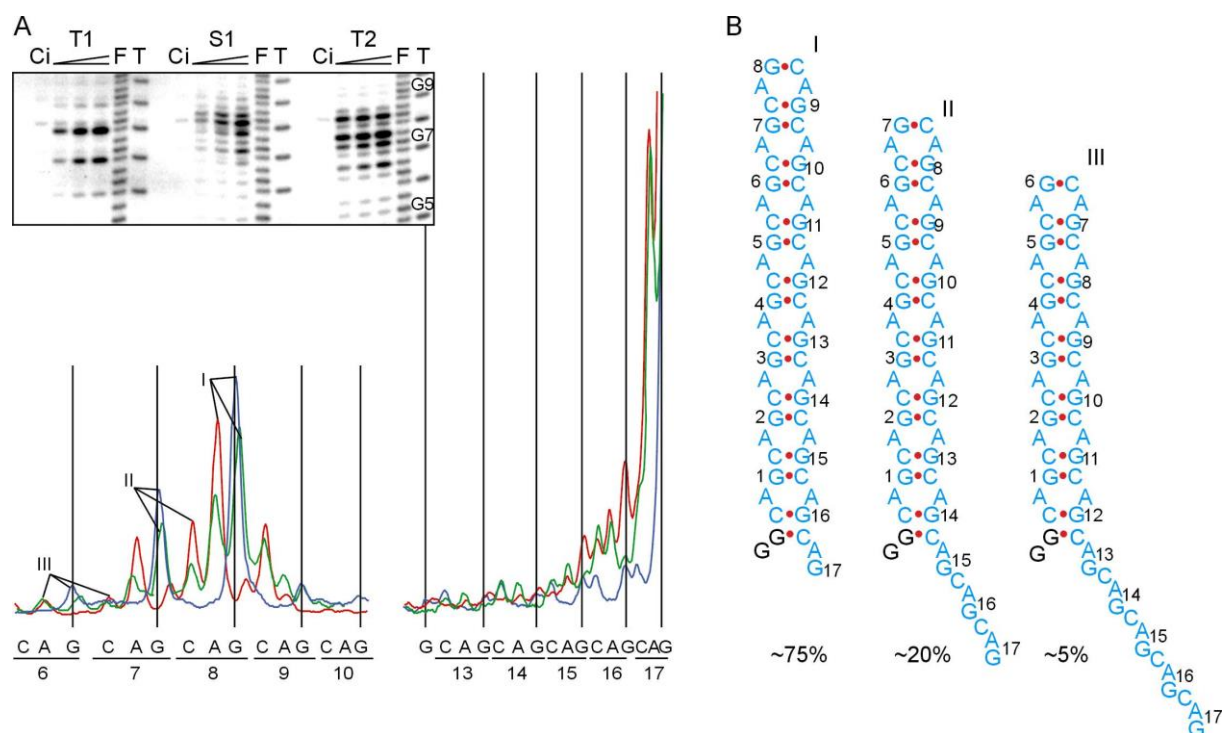
Wzór cięć generowanych przez nukleazy i dwudodatnie jony ołowiu uzyskany w wyniku sondowania struktury transkryptu CAG17 (ryc. 3.1 A – autoradiogram z lewej strony) wskazuje na obecność regionu pojedynczonicowego w centralnej części cząsteczki. Region ten stanowi pętlę terminalną struktury spinki. Obecności niektórych spośród obserwowanych cięć nie można było jednak przypisać żadnej pojedynczej strukturze. Zaobserwowano także grupę cięć poblizu końca 3' cząsteczki (ryc. 3.1 A – autoradiogram z lewej strony). Taki wzór cięć jest charakterystyczny dla współwystępujących w roztworze kilku struktur typu spinki [Napierała M., Krzyżosiak W.J. (1997)]. W celu dokładnego określenia budowy pętli terminalnej badaniami objęto również cząsteczki CAG17_cl i CAG16_cl zawierające poza sekwencją powtórzeń także dodatkowe nukleotydy C i G odpowiednio na końcach 5' i 3' mogące silnie oddziaływać ze sobą tworząc region sparowany o wielkości sześciu par zasad. Oddziaływanie to wymusiło przyjęcie przez region powtórzeń jednej formy struktury, co jest widoczne we wzorze cięć sondami strukturalnymi jako brak cięć w części 3'

transkryptu i mniejsza liczba wiązań internukleotydowych ciętych w regionie centralnym (ryc. 3.1 A – autoradiogram odpowiednio środkowy i z prawej strony). Różnica w homogenności struktury pomiędzy transkryptami CAG17 i CAG17_cl jest też wyraźnie widoczna w wynikach analizy migracji tych cząsteczek w żelu niedenaturującym. Prążek w żelu pochodzący od transkryptu CAG17_cl jest bardziej zwarty, a odpowiadający jemu pik bardziej ostry, co świadczy o tym, że badana cząsteczka przyjmuje jedną strukturę. Dla porównania sygnał odpowiadający prążkowi transkryptu CAG17 jest bardziej „rozmyty” zgodnie z obecnością kilku wariantów struktury różniących się szybkością migracji w żelu niedenaturującym (ryc. 3.1 C). Na podstawie obserwowanych wzorów cięć stwierdzono, że struktura spinki tworzona przez 17 powtórzeń CAG w CAG17_cl charakteryzuje się siedmionukleotydową pętlą terminalną utworzoną przez nukleotydy należące do 8, 9 i 10 powtórzenia motywu CAG, której parą zamykającą jest C z 8 i G z 10 tripletu powtórzeń CAG (ryc. 3.1 B). W przypadku cząsteczki CAG16_cl stwierdzono, że pętla terminalna jest tworzona z nukleotydów AG z 8 i CA z 9 powtórzenia. Jest to zatem pętla czteronukleotydowa. Jest ona bardziej termodynamicznie korzystna niż ta, która jest obecna w przypadku cząsteczek RNA zawierających nieparzystą liczbę powtórzeń (ryc. 3.1 B).

W reakcjach z jonami ołowiu reaktywność wiązań internukleotydowych w regionie trzonu i pętli terminalnej nie jest wyraźnie zróżnicowana. Wskazuje to na rozluźnioną strukturę trzonu cząsteczki (ryc. 3.1 D). Najbardziej reaktywne w stosunku do tej sondy są wiązania CpA, natomiast wiązania GpC prawie nie wykazywały podatności na jej działanie. Podobną preferencję trawienia wiązań internukleotydowych stwierdzono w przypadku RNazy T2. Natomiast przecinająca w obrębie regionów podwójnoniciowych RNaza V1 generuje najmocniejsze cięcia wiązań CpG. Obserwacje te pozwoliły stwierdzić, że trzon struktury spinki powtórzeń CAG składa się z par GC i CG oddzielonych pojedynczymi niesparowaniami AA.

Poznanie struktury regionu pętli terminalnej w transkryptach zawierających dodatkowe sekwencje stabilizujące strukturę pojedynczej spinki utworzoną przez powtórzenia CAG w transkryptach CAG17_cl i CAG16_cl oraz zastosowanie do analiz ilościowych oprogramowania skanera PhosphorImager Typhoon pozwoliło na pełniejszą interpretację wzoru cięć uzyskiwanych w cząsteczce CAG17. Zauważono pewne podobieństwo w lokalizacji i intensywności cięć centralnego regionu transkryptu CAG17 do obrazu cięć w pętli terminalnej transkryptu CAG16_cl (ryc. 3.1

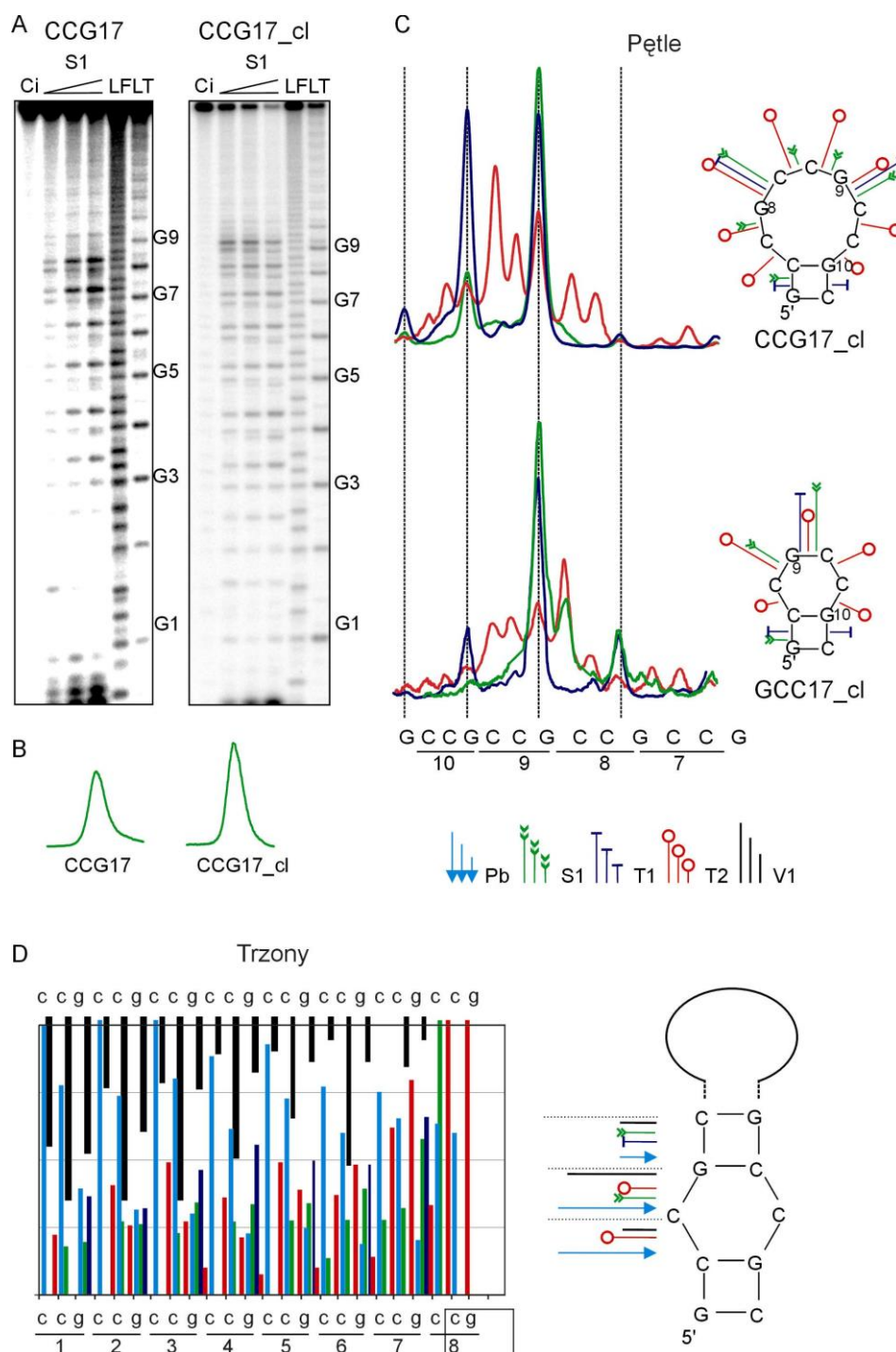
A). Przyjęto, że złożony obraz trawień w regionie pętli terminalnej transkryptu CAG17 jest wynikiem współwystępowania wariantów struktury o identycznej budowie pętli terminalnej, w której tworzeniu biorą jednak udział sąsiadujące ze sobą powtórzenia. Dało to podstawę do określenia struktury poszczególnych wariantów i oznaczenia ich udziału w mieszaninie (ryc. 3.2). Dominującą formą jest przyjmowana przez około 75% cząsteczek RNA CAG17 struktura spinki utworzonej z 16 powtórzeń motywu CAG. Jej struktura różni się od struktury transkryptu CAG16_cl jedynie brakiem regionu stabilizującego, zaś dodatkowe powtórzenie nie biorące udziału w oddziaływaniu stanowi krótki region pojedynczoniciowy po stronie 3'. Kolejne dwa warianty występujące w mieszaninie z częstością 20 i 5% różnią się od formy dominującej przesunięciem pętli terminalnej w kierunku końca 5' o odpowiednio dwa i cztery powtórzenia CAG. Wraz ze skróceniem trzonu struktury spinki wydłużeniu ulega region pojedynczoniciowy zlokalizowany na końcu 3' cząsteczki.



Rycina 3.2. Warianty struktury transkryptu CAG17. A. Wzór cięć generowanych przez nukleazy specyficzne w stosunku do regionów pojedynczoniciowych oraz wykresy intensywności trawień w regionie pętli terminalnej i w pobliżu końca 3' transkryptu CAG17. Zaznaczono cięcia charakteryzujące pętle poszczególnych wariantów struktury. B. Warianty struktury transkryptu CAG17. Poniżej podano ich szacunkowy udział w mieszaninie.

W badaniach transkryptów złożonych z powtórzeń CCG wykorzystano informacje wynikające z analiz struktury powtórzeń CAG. W przypadku transkryptu CCG17, podobnie jak CAG17, obserwowana była grupa cięć generowanych przez sondy specyficzne w stosunku do regionów pojedynczonicowych. Ten wzór cięć odpowiadał mieszaninie niewiele różniących się od siebie wariantów strukturalnych. Dodatkowym podobieństwem była obecność cięć w pobliżu końca 3' (ryc. 3.3 A). Wobec analogicznej sytuacji, jak ta która wystąpiła podczas analizy powtórzeń CAG przystąpiono do badań transkryptu CCG17_cl. Obserwowana zmiana profilu cięć (ryc. 3.3 A) jak i sposobu migracji w żelu niedenaturującym potwierdziła słuszność założenia, że dodatkowe nukleotydy oddziałują ze sobą u podstawy struktury spinki i spełniają rolę stabilizującą. W transkrypcie stabilizowanym CCG17_cl wzór cięć w obrębie 8 i 9 oraz 10 powtórzenia potwierdza istnienie spodziewanej siedmionukleotydydowej pętli terminalnej (ryc. 3.3 C). Przeprowadzona analiza właściwości strukturalnych trzonu potwierdziła występujący w nim periodyczny układ motywów sparowanych GC CG oddzielonych od siebie niesparowaniami CC (ryc. 3.3 D). Ze względu na charakter sekwencji powtórzeń transkryptu CCG17 możliwe jest, w wyniku parowania z dodatkowymi nukleotydami GG znajdującymi się na końcu 5', powstanie alternatywnego ułożenia struktury w sposób odmienny od poślizgu obserwowanego w przypadku powtórzeń CAG. Odpowiada on strukturze „spiętej” u podstawy sekwencji GCC17_cl i charakteryzuje się występowaniem czteronukleotydydowej pętli terminalnej przy zachowanej strukturze trzonu (ryc. 3.3 C). Tak więc zmiana „orientacji” powtórzonych motywów CCG na GCC wywołała zmianę struktury podobną do obserwowanej między parzystą a nieparzystą liczbą powtórzeń w ciągu CAG.

Wykazano zatem, że RNA powtórzenia CAG jak i CCG występują w formie kilku wariantów struktury typu spinki o czteronukleotydydowej pętli terminalnej różniących się długością pojedynczonicowego odcinka złożonego z 3' końcowych powtórzeń. Natomiast w sytuacji, kiedy sekwencje otaczające tworzą region sparowany wymuszający przyjęcie przez ciąg powtórzeń trójnukleotydydowych pojedynczej struktury spinki, o wielkości jej pętli terminalnej decyduje parzysta lub nieparzysta liczba powtórzeń.



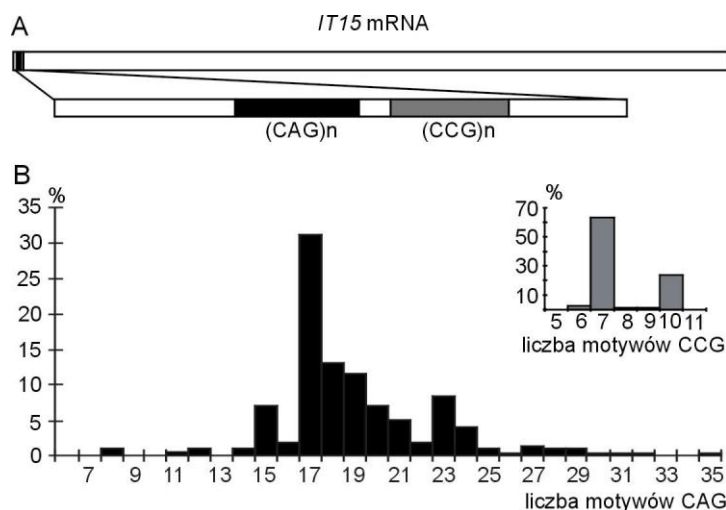
Rycina 3.3. Analiza struktury transkryptów złożonych z powtórzeń trójnukleotydu CCG. A. Wzór cięć generowanych przez nukleazę S1 w obrębie badanych transkryptów CCG17 (autoradiogram z lewej) i CCG17_cl (autoradiogram z prawej). B. Analiza jednorodności struktur tworzonych przez cząsteczki RNA złożone z powtórzeń CCG przez porównanie różnicowania szybkości ich migracji w żelu niedenaturującym z wykorzystaniem oprogramowania skanera – PhosphorImager. C. Lokalizacja i porównanie intensywności cięć generowanych przez użyte nukleazy: S1 (linia zielona), T1 (linia niebieska) i T2 (linia czerwona) w regionach pętli terminalnych transkryptów CCG17_cl i GCC17_cl. D. Struktura trzonu spinki tworzonej przez powtórzenia CCG. Wykres prezentuje ilościowe porównanie aktywności zastosowanych sond strukturalnych obserwowanej w obrębie pierwszych 8 powtórzeń CCG cząsteczki CCG17_cl (nukleotydy zaangażowane w tworzenie pętli terminalnej zaznaczono w ramce). W jednej skali przedstawiono aktywność sond w stosunku do regionów pojedynczonicowych. Dla silniejszych cięć nukleazy V1 (rozpoznającej nukleotydy sparowane) zastosowano inną skalę. Kolory słupków odpowiadających poszczególnym nukleazom i jonom Pb^{2+} są zgodne z przedstawioną na rycinie legendą.

3.2 Polimorfizm liczby powtórzeń CAG i CCG w genie *IT15* związanym z chorobą Huntingtona.

Po określeniu właściwości struktur tworzonych przez same powtórzenia trójnukleotydowe CAG i CCG przystąpiono do ich zbadania w otoczeniu naturalnych sekwencji. Zaplanowane analizy miały na celu uzyskanie odpowiedzi na pytanie czy identyczne struktury mogą występować również w naturalnym otoczeniu sekwencji *IT15* mRNA, czy też dochodzi do wzajemnego oddziaływania tych sekwencji występujących w bezpośrednim sąsiedztwie w sekwencji transkryptu.

Przed przystąpieniem do badań struktury fragmentów mRNA eksonu 1 genu *IT15* konieczne było uzyskanie szeregu różnych wariantów transkryptu różniących się liczbą powtórzeń CAG i CCG. Przyjęto, że najlepszymi modelami będą takie transkrypty, które występują w naturze. W związku z tym przeprowadzono badania polimorfizmu długości, jaki wykazują występujące w mRNA *IT15* dwa ciągi powtórzeń trójnukleotydowych CAG i CCG w populacji polskiej (ryc. 3.4).

W celu określenia zakresu polimorficznych zmian liczby powtórzeń CAG i CCG w genie *IT15* w populacji polskiej wykonano analizę długości fragmentów produktów amplifikacji regionu zawierającego oba ciągi trójnukleotydowe. Ponieważ



Rycina 3.4. Częstość występowania ciągów o określonej liczbie powtórzeń motywów CAG i CCG w genie *IT15* w populacji polskiej. A. Schemat sekwencji regionu powtórzeń trójnukleotydowych we fragmencie genu *IT15* (kolor biały – sekwencje specyficzne; kolor czarny – ciąg powtórzeń CAG; kolor szary – ciąg powtórzeń CCG). B. Częstość występowania różnej długości ciągów powtórzeń CAG w eksonie 1 genu *IT15* na podstawie analizy 300 chromosomów (N = 300). Wewnątrz częstość występowania różnej długości ciągów powtórzeń CCG w eksonie 1 genu *IT15* (N = 300).

oba analizowane ciągi powtórzeń wykazują polimorfizm konieczne było ich rozdzielenie. Region rozdzielający ciągi jest krótką sekwencją pseudopowtórzeń, której wykorzystanie jako miejsca przyłączenia startera w przypadku amplifikacji z użyciem genomowego DNA jako matrycy nie byłoby skuteczne. Z tego względu proces amplifikacji regionu powtórzeń genu *IT15* prowadzono w dwóch etapach. Najpierw amplifikowano cały region powtórzeń a produkt pierwszego etapu PCR użyto jako matrycę do drugiego etapu, w którym amplifikowano oddzielnie oba ciągi trójnukleotydydowe. W ten sposób przeprowadzono niezależną analizę długości produktów reamplifikacji ciągu CAG oraz analizę długości fragmentów produktów reamplifikacji regionu zawierającego ciąg CCG. Populację polską reprezentowało w tych badaniach 150 zdrowych, nie spokrewnionych osób zamieszkujących region Wielkopolski i Górnego Śląska nie selekcionowanych pod kątem historii rodzinnej chorób neurologicznych. Uzyskane wyniki wskazują na unimodalny rozkład zmian liczby powtórzeń CAG w zakresie od 8 do 35 powtórzeń (ryc. 3.4). Najczęściej występuje allel z 17 powtórzeniami CAG. Natomiast w ciągu powtórzeń CCG dominują dwa allele o długości 7 i 10 powtórzeń przy niewielkim zakresie zmienności (6 – 10 CCG). Dane o zakresie zmienności długości ciągu CAG nieznacznie różnią się w doniesieniach mówiących o liczbie powtórzeń CAG w allelach genu *IT15* i częstości ich występowania w różnych populacjach. Na przykład w populacji holenderskiej stwierdzono najczęstsze występowanie allelu z 16 powtórzeniami CAG [De Rooij K. E. i wsp. (1993)]. Wyniki takie otrzymano również podczas badań populacji szkockiej [Simpson S. A. i wsp. (1993)]. W populacji francuskiej stwierdzono najwyższą częstość allelu zawierającego 20 powtórzeń [Dode C. i wsp. (1993)]. W populacjach nieeuropejskich allel najczęściej występujący zawiera podobną liczbę powtórzeń CAG. Przykładowo wśród Chińczyków zamieszkujących Tajwan dominuje allel z 17 powtórzeniami [Soong B. W., Wang J. T. (1995)], a w populacji meksykańskiej z 18 powtórzeniami [Alonso M. E. i wsp. (1997)]. Fakt dominacji allelu z 7 powtórzeniami CCG jest zgodny z wynikami dotychczasowych badań innych populacji. Na podstawie zebranych informacji o rozkładzie długości traktów trójnukleotydydowych w populacji wybrano modelowe cząsteczki RNA o charakterystyce przedstawionej w tabeli 3.1. Wybrane cząsteczki stwarzają możliwość odpowiedzi na pytanie czy sąsiadujące ze sobą powtórzenia CAG i CCG tworzą niezależne struktury, czy też oddziałują ze sobą. Pozwalają także na zaobserwowanie ewentualnych zmian struktury RNA pod wpływem zmian liczby

poszczególnych powtórzeń trójnukleotydom. Interesująca jest też aranżacja struktury regionu rozdzielającego trakt CAG i CCG oraz jej zmiany wraz ze zmianami liczby powtórzeń w obu traktach.

Tabela 3.1. Modelowe transkrypty odpowiadające fragmentom genu *IT15* użyte do badań struktury RNA regionu powtórzeń trójnukleotydom występujących w tym genie.

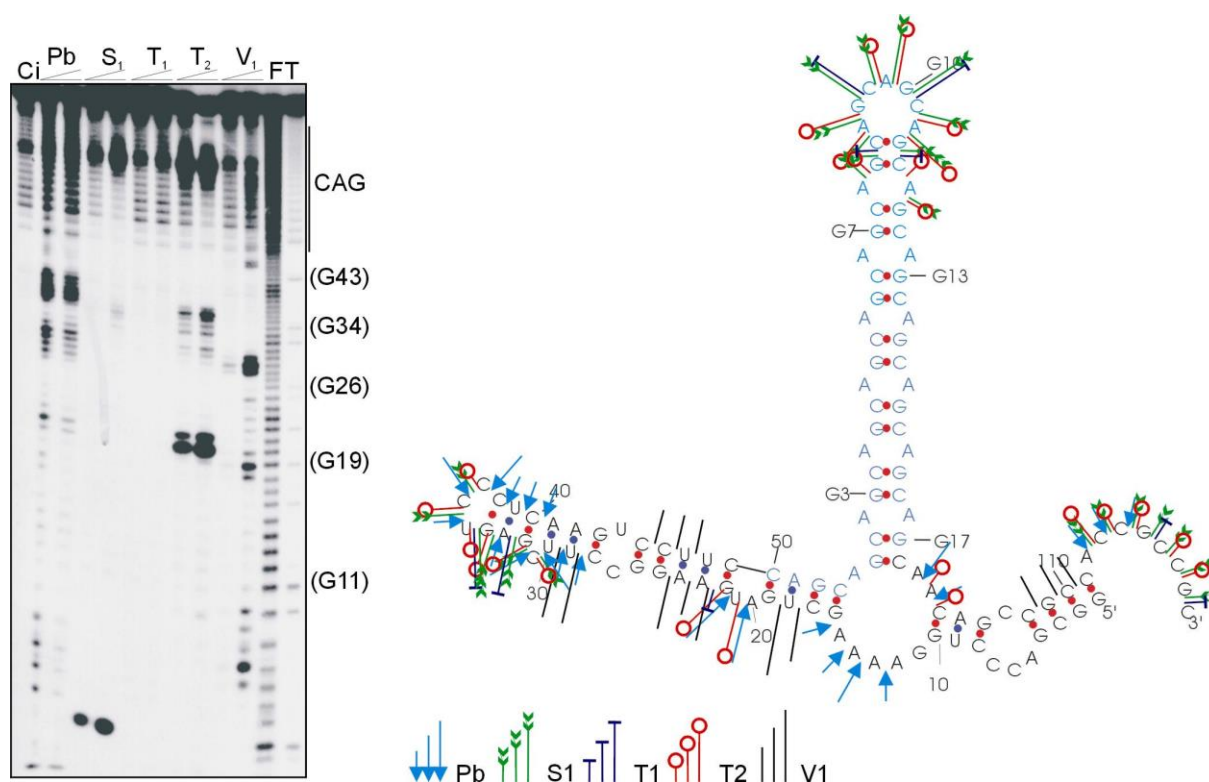
Nazwa transkryptu	Liczba powtórzeń CAG	Liczba powtórzeń CCG	Uwagi i cel badań
IT15–17	17	-	Określenie struktury najczęstszego w populacji ciągu CAG w obecności naturalnych sekwencji
IT15–10	10	-	Określenie struktury krótkiego ciągu CAG
IT15 10-10	10	10	Stwierdzenie możliwości oddziaływania ciągów powtórzeń CAG i CCG jednakowej długości w otoczeniu naturalnych sekwencji
IT15 17-10	17	10	Drugi co do częstości allel w populacji; Wykazanie wpływu zmiany liczby powtórzeń CAG na strukturę drugorzędową
IT15 17-7	17	7	Najczęstszy allel w populacji polskiej (~20%) zawierający oba ciągi powtórzeń, Wykazanie wpływu małych zmian liczby powtórzeń CCG na strukturę drugorzędową
IT15 23-7	23	7	Wykazanie wpływu wzrostu liczby powtórzeń CAG na strukturę drugorzędową
IT15 24-7	24	7	Wykazanie wpływu parzystej liczby powtórzeń CAG na strukturę drugorzędową
IT15 44-7	44	7	Mutacja; określenie struktury RNA fragmentu allelu zmutowanego

3.3 Powtórzenia CAG we fragmencie mRNA genu *IT15* tworzą strukturę typu spinki.

W celu odpowiedzi na pytanie o możliwość występowania struktury spinki przyjmowanej przez powtórzenia trójnukleotydom CAG w otoczeniu naturalnej sekwencji specyficznej wykonano badania struktury transkryptu IT15–17. Jest to fragment mRNA genu *IT15* zawierający poza ciągiem powtórzeń CAG sekwencję o długości 49 nukleotydów po jego stronie 5'. Odcinek sekwencji specyficznej po stronie 3' zawiera 19 nukleotydów. Jego długość wynika z obecności drugiego ciągu powtórzeń CCG pomijanego na tym etapie badań. Transkrypt ten umożliwiał

bezpośrednie porównanie uzyskanych wyników z danymi dotyczącymi transkryptów CAG17 i CAG17_cl.

Uzyskany profil trawień transkryptu IT15–17 (ryc. 3.5) generowanych przez zastosowane sondy strukturalne specyficzne dla regionów pojedynczoniciowych charakteryzuje się występowaniem dwóch grup silnych cięć. Pierwsza z nich jest zlokalizowana w obrębie sekwencji specyficznej po stronie 5' ciągu powtórzeń. Jej lokalizacja pozwala stwierdzić, że znaczna część regionu otaczającego 5' przyjmuje w transkrypcie IT15–17 formę spinki (ryc. 3.5), której trójnukleotydoma pętla terminalna jest utworzona przez nukleotydy U35, C36 i C37. Większa liczba cięć zlokalizowanych pomiędzy U30 a C40 wskazuje na rozluźnienie struktury u podstawy tej małej pętli. Może ono wynikać z destabilizującego wpływu pętli zwiększonego dodatkowo przez obecność w niewielkiej odległości od niej jednonukleotydowego wybrzuszenia C31. Utworzenie takiej rozluźnionej struktury zwiększa podatność wiązań internukleotydowych na specyficzne działanie zastosowanych sond.



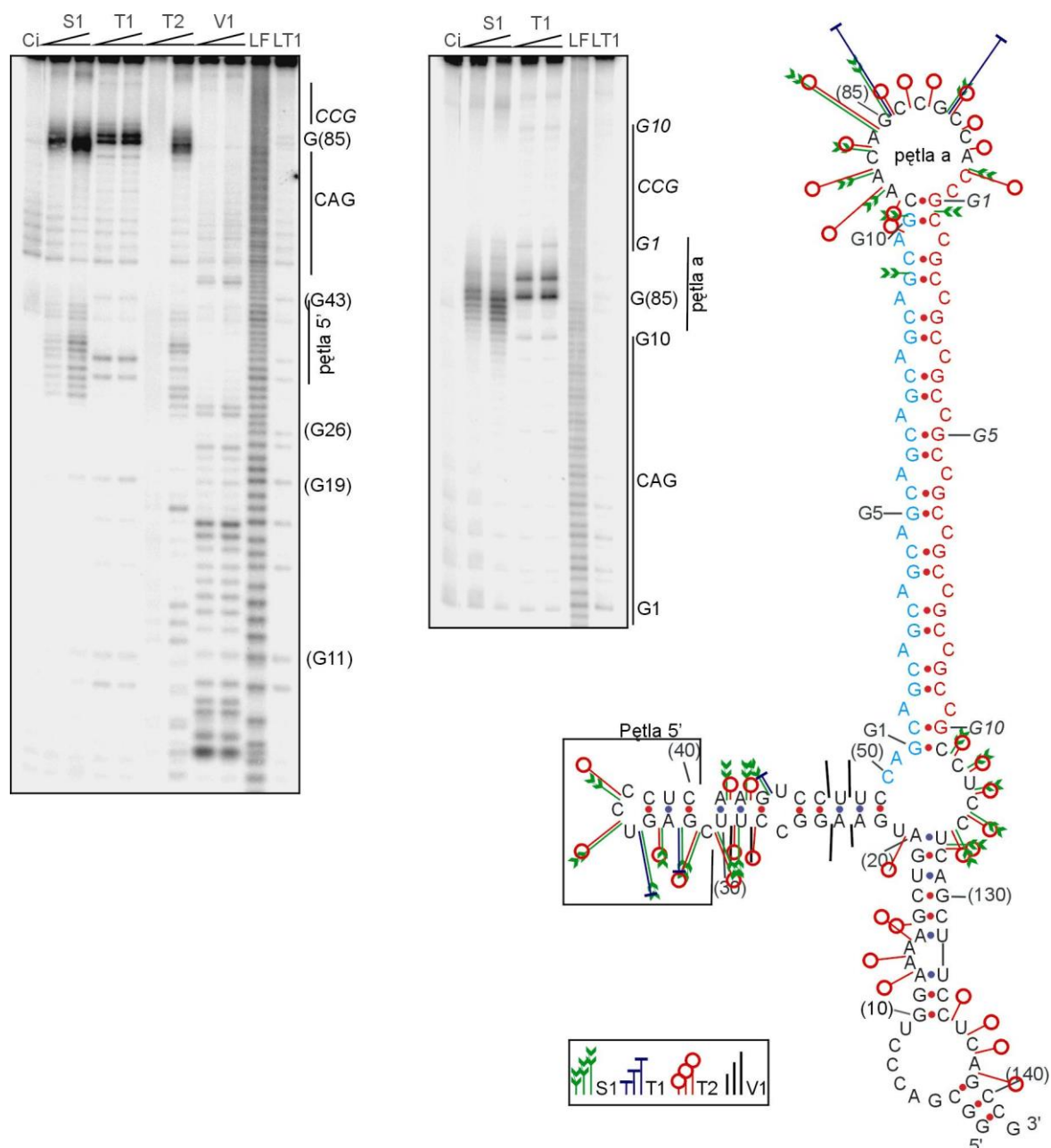
Rycina 3.5. Powtórzenia CAG w obecności naturalnej sekwencji otaczającej zlokalizowanej po stronie 5' przyjmują strukturę typu spinki. A – wzór cięć generowanych przez sondy strukturalne; B – struktura sekwencji 17 powtórzeń CAG w obecności sekwencji otaczającej mRNA genu *IT15*. Lokalizację i intensywność cięć poszczególnymi sondami zaznaczono symbolami przedstawionymi w legendzie.

W strukturze regionów sekwencji specyficznej otaczającej ciąg powtórzeń CAG dochodzi do oddziaływania sekwencji 5' i 3'. W obrębie sekwencji powtórzeń przecinane są wiązania pomiędzy nukleotydami 9, 10 i 11 tripletu. Ich lokalizacja i intensywność są bardzo podobne do wzoru cięć charakteryzującego siedmionukleotydową pętlę transkryptu CAG17_cl. Zatem w obecności sekwencji otaczającej został zahamowany „poślizg” sekwencji powtarzającej się. Nie stwierdzono istnienia regionu podwójnoniciowego u podstawy struktury spinki jednak wydaje się możliwe, że rolę stabilizatora mogą pełnić oddziaływania warstwowe pomiędzy trzonami struktur spinek sekwencji specyficznej 5' oraz tworzonej przez powtórzenia. Przesunięcie pętli terminalnej o dwa powtórzenia w kierunku końca 5' traktu wynika z zaangażowania pierwszego powtórzenia CAG i reszty cytydylowej drugiego powtórzenia w oddziaływanie z sekwencją otaczającą 5', z którą nukleotydy te tworzą dokładnie sparowany region o długości czterech par zasad. Inny charakter oddziaływań w jakich uczestniczy fragment 5' ciągu powtórzeń CAG jest widoczny w różnicy reaktywności jego wiązań internukleotydowych w porównaniu z innymi wiązaniami internukleotydowymi w regionie trzonu spinki utworzonej z samych powtórzeń CAG (ryc. 3.5).

Po stwierdzeniu, że struktura 17 powtórzeń CAG w obecności fragmentu sekwencji otaczającej zasadniczo nie odbiega od schematu obserwowanego na przykładzie transkryptu CAG17_cl interesujące było określenie czy minimalna długość ciągu tego rodzaju powtórzeń występująca w populacji jest wystarczająca do utworzenia struktury typu spinki. Do odpowiedzi na to pytanie wykorzystano transkrypt IT15–10 zawierający 10 powtórzeń CAG i fragment sekwencji otaczającej 5' taki jak w transkrypcie IT15–17. Wyniki reakcji sondowania struktury transkryptu IT15–10 wskazują, że także w tym przypadku jest przyjmowana struktura typu spinki przez powtórzenia CAG (ryc. 3.6 B). Jej pętla terminalna (AGCA) rozpoznawana przez zastosowane sondy enzymatyczne (ryc. 3.6 A) jest tworzona przez nukleotydy 6 i 7 powtórzeń i ma budowę typową dla struktury tworzonej przez ciąg zawierający parzystą liczbę powtórzeń. Trzon podobnie, jak w cząsteczce CAG17 i IT15–17 jest złożony ze sparowanych regionów CG GC oddzielonych niesparowaniami AA. Obserwowane przesunięcie regionu pętlowego w kierunku końca 3' ciągu powtórzeń powodowane jest, obserwowanym także w transkrypcie IT15–17, zaangażowaniem 4 pierwszych nukleotydów ciągu powtórzeń w oddziaływanie z regionem sekwencji specyficznej występującej po stronie 5' u podstawy struktury typu spinki tworzonej

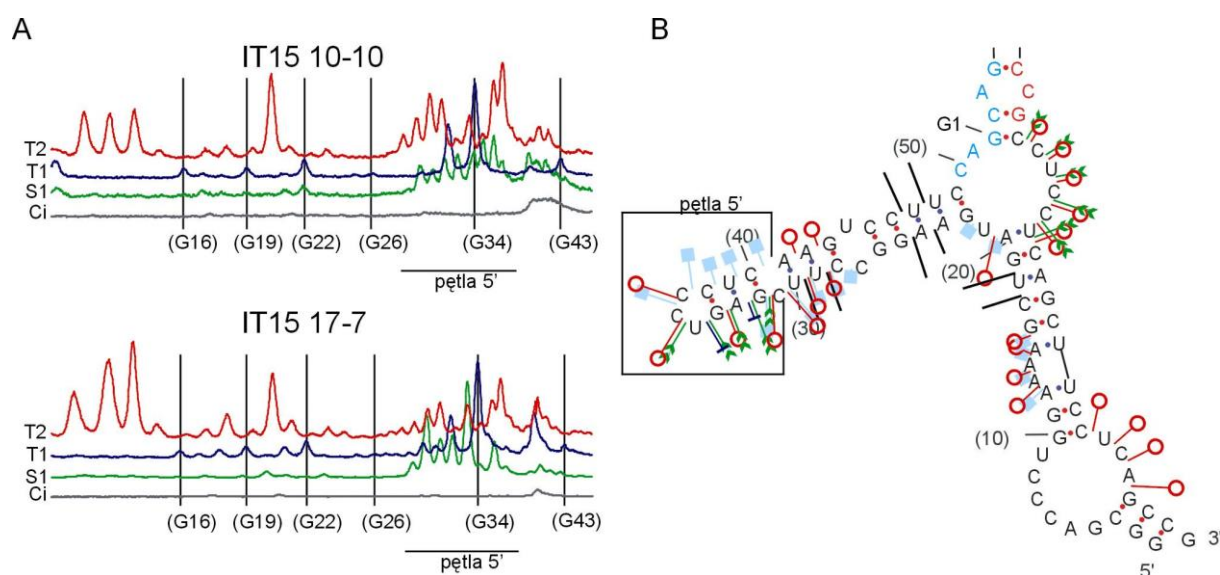
3.4 Sąsiadujące ciągi powtórzeń CAG i CCG oddziałują ze sobą we fragmencie mRNA genu *IT15*.

W celu uzyskania odpowiedzi na pytanie o możliwość oddziaływania powtórzeń CAG i CCG występujących w mRNA genu *IT15* przeprowadzono analizę struktury transkryptu *IT15* 10-10 (tab. 3.1). Zawiera on 10 powtórzeń w każdym ciągu. Są one oddzielone od siebie dwunastonukleotydową sekwencją CAACAGCCGCCA. Specyficzna sekwencja po stronie 5' powtórzeń ma długość 49 nukleotydów, a po stronie 3' 21 nukleotydów. Obserwacja szybkości migracji w żelu niedenaturującym transkryptu *IT15* 10-10 wskazuje na jednorodność struktury przyjmowanej przez tę cząsteczkę. Nukleazy specyficzne w stosunku do regionów pojedynczoniciowych użyte do analizy struktury generują w obrębie transkryptu *IT15* 10-10 grupy intensywnych cięć (ryc. 3.7). Pierwsza z nich jest zlokalizowana w obrębie sekwencji specyficznej po stronie 5' regionu powtórzeń. Wzór cięć (ryc. 3.7 – pętla 5') pozwala określić wielkość regionu pojedynczoniciowego na 10 nukleotydów. Jednakże słabsze trawienie wiązań internukleotydowych zlokalizowanych w tym regionie, pomiędzy resztami które mogą tworzyć 3 pary zasad typu Watson – Crick, wskazuje na zaangażowanie ich w tego rodzaju oddziaływanie. W konsekwencji region ten tworzy pętlę o wielkości trzech nukleotydów (ryc. 3.7 i 3.8 – pętla 5'). Stanowi ona pętlę terminalną struktury spinki tworzonej przez 28 nukleotydów specyficznej sekwencji 5' otaczającej region powtórzeń, której trzon wykazuje u podstawy wysoką stabilność i jest rozpoznawany przez RNazę V1 specyficzną w stosunku do regionów podwójnoniciowych. Struktura spinki tworzona przez sekwencję otaczającą 5' jest połączona po stronie 3' ze strukturą regionu powtórzeń trójnukleotydowych poprzez niesparowane nukleotydy CA pierwszego powtórzenia CAG (ryc. 3.7; szczegóły ryc. 3.8). Natomiast po stronie 5' trzon struktury spinki utworzonej przez sekwencję specyficzną połączony jest poprzez jeden niesparowany nukleotyd z regionem sparowanym powstającym w wyniku oddziaływania ze sobą sekwencji otaczającej 5' i 3'. Ten podwójnoniciowy region tworzy 9 par zasad. W odległości 3 par zasad od jej końca 5' zlokalizowane jest dwunukleotydowe wybrzuszenie AA dokładnie mapowane przez jony Pb^{2+} i RNazę T2 (ryc. 3.8). Pierwszy nukleotyd sekwencji otaczającej region powtórzeń po stronie 3' paruje z G



Rycina 3.7. Analiza struktury fragmentu mRNA genu *IT15* zawierającego 10 powtórzeń CAG i 10 CCG. Autoradiogram po lewej stronie pokazuje profil cięć generowanych przez sondy enzymatyczne w całym badanym transkrypcie. Autoradiogram po prawej stronie pokazuje miejsca przecinania regionu powtórzeń trójnukleotydowych przez nukleazę S1 i RNazę T1. Po prawej stronie przedstawiono strukturę przyjmowaną przez fragment mRNA genu *IT15* z 10 powtórzeniami w obu ciągach. Lokalizację i intensywność cięć poszczególnymi sondami zaznaczono symbolami przedstawionymi w legendzie. Zaznaczono kolejne reszty G powtórzeń CAG (litery niebieskie) zaś reszty G powtórzeń CCG (litery czerwone) oznaczono kursywą. W nawiasach podano numer kolejnych, liczonych od końca 5', nukleotydów cząsteczki.

pierwszego powtórzenia CAG tworząc pierwszą parę trzonu. Następne 4 nukleotydy stanowią region łączący (ryc. 3.8) struktury spinki regionu powtórzeń trójnukleotydowych z regionem podwójnoniciowym tworzonym przez oddziałujące ze sobą fragmenty sekwencji otaczających. Poza cięciami widocznymi w obrębie specyficznej sekwencji po stronie 5' regionu powtórzeń trójnukleotydowych (pętla 5') transkryptu IT15 10-10 w regionie zawierającym powtórzenia obserwowana jest grupa intensywnych cięć (ryc. 3.7) zlokalizowanych w specyficznej sekwencji rozdzielającej ciąg CAG od CCG (pętla a). Nukleaza S1 przecina kolejne wiązania fosfodiesterowe począwszy od wiązania pomiędzy resztą G dziesiątego powtórzenia CAG aż do G85 sekwencji specyficznej ze wzrastającą intensywnością (ryc. 3.7 – autoradiogram po prawej stronie). Także rybonukleaza T2 intensywnie trawi wiązania pomiędzy nukleotydami od A z 10 powtórzenia CAG do A91 sekwencji specyficznej poprzedzającej ciąg CCG. We wzorze trawień generowanych przez ten enzym wyraźnie widoczna jest jego nukleotydowa specyficzność w stosunku do reszt adenylowych. Rybonukleaza T1 przecina wiązania G85pC i G88pC w sekwencji



Rycina 3.8. Struktura specyficznej sekwencji otaczającej jest niezależna od zmian w regionie powtórzeń. A. Porównanie lokalizacji i intensywności cięć generowanych przez sondy strukturalne w sekwencji otaczającej 5' fragmentów mRNA genu *IT15*. W części górnej – transkrypt IT15 10-10, w części dolnej – transkrypt IT15 17-7. B. Struktura drugorzędowa sekwencji specyficznej otaczającej region powtórzeń z zaznaczoną lokalizacją i intensywnością cięć generowanych przez zastosowane sondy. Opracowana została na podstawie analiz wszystkich użytych w tej pracy transkryptów odpowiadających fragmentom genu *IT15* zawierającym oba ciągi powtórzeń trójnukleotydowych. Oznaczenia jak na rycinie 3.7.

rozdzielającej ciągi powtórzeń (ryc. 3.7 – autoradiogram po prawej stronie). Obserwowany wzór trawień pozwala stwierdzić, że odpowiada on pętli terminalnej struktury typu spinki tworzonej przez oddziałujące ze sobą ciągi trójnukleotydowe (ryc. 3.7). Jest to duża, trzynastonukleotydowa pętla, której parę zamykającą tworzą pierwszy nukleotyd sekwencji specyficznej C80 i reszta G pierwszego powtórzenia CCG. Pojedyncze cięcia obserwowane w części terminalnej trzonu struktury spinki są przypuszczalnie efektem rozluźnienia jego struktury pod wpływem tej pętli. Profil cięć generowanych przez nukleazy w regionie trzonu struktury pozwala określić jego strukturę jako układ krótkich regionów podwójnoniciowych tworzonych przez pary GC CG oddzielonych niesparowaniami AC.

3.5 Analiza termodynamicznych właściwości struktur RNA powstających z oddziaływania powtórzeń CAG i CCG.

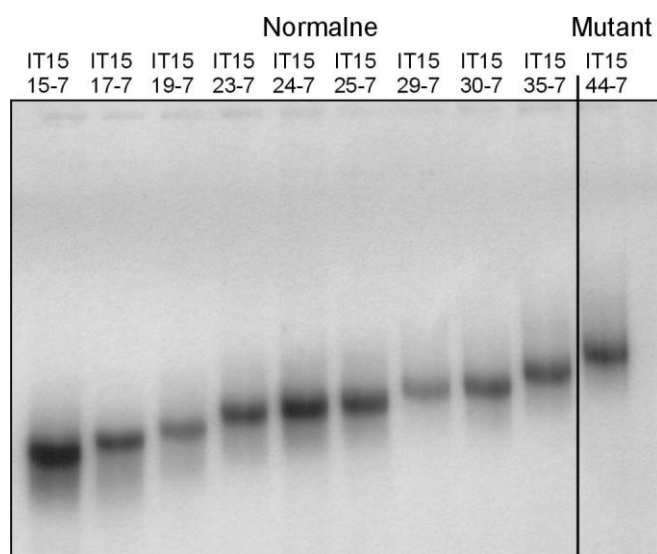
Na podstawie obserwacji tworzenia samodzielnej, stabilnej struktury spinki przez powtórzenia CAG zarówno w cząsteczkach modelowych jak i we fragmentach naturalnego mRNA genu *IT15* (rozdz. 3.1 i 3.3) oraz po stwierdzeniu możliwości oddziaływania powtórzeń CAG i CCG w transkrypcie *IT15* 10-10 (rozdz. 3.4) sformułowano hipotezę, że struktura utworzona w wyniku oddziaływania dwóch ciągów trójnukleotydowych w mRNA genu *IT15* jest termodynamicznie bardziej korzystna niż utworzenie dwóch niezależnych struktur typu spinki.

W celu zweryfikowania tej hipotezy po raz kolejny przystąpiono do analizy struktury modelowych transkryptów o sekwencji składającej się z motywów powtarzających się. Tym razem badano transkrypty zawierające 20 powtórzeń CAG – (CAG)₂₀, 20 powtórzeń CCG – (CCG)₂₀ oraz transkrypt z 10 powtórzeniami CAG i 10 CCG – (CAG)₁₀(CCG)₁₀. We wszystkich przypadkach stwierdzono, że badane cząsteczki przyjmują strukturę spinki, której centralnie zlokalizowana pętla terminalna złożona jest z czterech nukleotydów (ryc. 3.9). Dodatkowe cięcia generowane przez nukleazę S1 po stronie 5' pętli terminalnej (ryc. 3.9 A) wskazują na występowanie drugiej alternatywnej formy struktury typu spinki o pętli przesuniętej w kierunku 5'. Występowanie takiego, opisanego na przykładzie powtórzeń CAG, a typowego także dla innych CNG [Sobczak K. i wsp. (2003)], „poślizgniętego” wariantu struktury potwierdzają cięcia obserwowane w obrębie 6 ostatnich nukleotydów badanych

transkryptów. W cząsteczkach zbudowanych z jednego rodzaju powtórzeń, zarówno CAG jak i CCG pętla terminalna dominującego wariantu struktury tworzona jest przez nukleotydy 10 i 11 powtórzenia. W cząsteczce zbudowanej z dwóch rodzajów powtórzeń jest ona tworzona z nukleotydów AG ostatniego powtórzenia CAG i CC pierwszego powtórzenia CCG, zaś trzon jest zbudowany z oddziałujących ze sobą powtórzeń CAG i CCG. Interesujący jest także fakt zróżnicowanego stopnia dostępności poszczególnych reszt guanylowych badanych cząsteczek dla RNazy T1 w reakcji w warunkach semidenaturujących, w której jest przygotowany marker długości (ryc. 3.9 A – linie LT1). W przypadku cząsteczek złożonych z jednego typu powtórzeń rozpoznawane były wszystkie reszty G w cząsteczce, natomiast w transkrypcie (CAG)₁₀(CCG)₁₀ w linii przedstawiającej intensywność trawień w markerze długości (LT1) widoczne są tylko reszty występujące w regionie pętli. Obserwacja ta pozwala sądzić, że stabilność proponowanych struktur w dużym stopniu zależy od rodzaju tworzącego je motywu trójnukleotydowego. W celu precyzyjnego określenia termodynamicznych właściwości obserwowanych struktur przeprowadzono ich topnienie monitorowane spektrofotometrycznie. Uzyskane krzywe topnienia (ryc. 3.9 C) potwierdzają wcześniejsze obserwacje. Najtrwalszą strukturę tworzy transkrypt zawierający dwa oddziałujące ze sobą ciągi powtórzeń trójnukleotydowych. Jej temperaturę topnienia określono na 77,5 °C a energię swobodną $\Delta G_{37^\circ\text{C}} = -16,5$ kcal/mol. Struktura cząsteczki RNA zbudowanej z samych powtórzeń CAG charakteryzuje się wartością energii swobodnej $\Delta G_{37^\circ\text{C}} = -6,1$ kcal/mol i temperaturą topnienia 62,4 °C. Natomiast cząsteczka złożona z powtórzeń CCG przyjmuje najmniej stabilną strukturę spinki o $T_m = 58,2$ °C i $\Delta G_{37^\circ\text{C}} = -2,5$ kcal/mol. Analizowane transkrypty różnią się w swojej strukturze budową trzonu, którego podwójnoniciowe regiony GC CG są oddzielone od siebie pojedynczymi niesparowaniami. W najbardziej stabilnej cząsteczce (CAG)₁₀(CCG)₁₀ występują niesparowania AC. Natomiast w transkryptach (CAG)₂₀ i (CCG)₂₀ są to odpowiednio niesparowania AA i CC. Te różnice wydają się być podstawą obserwowanego zróżnicowania stabilności tworzonych struktur typu spinki.

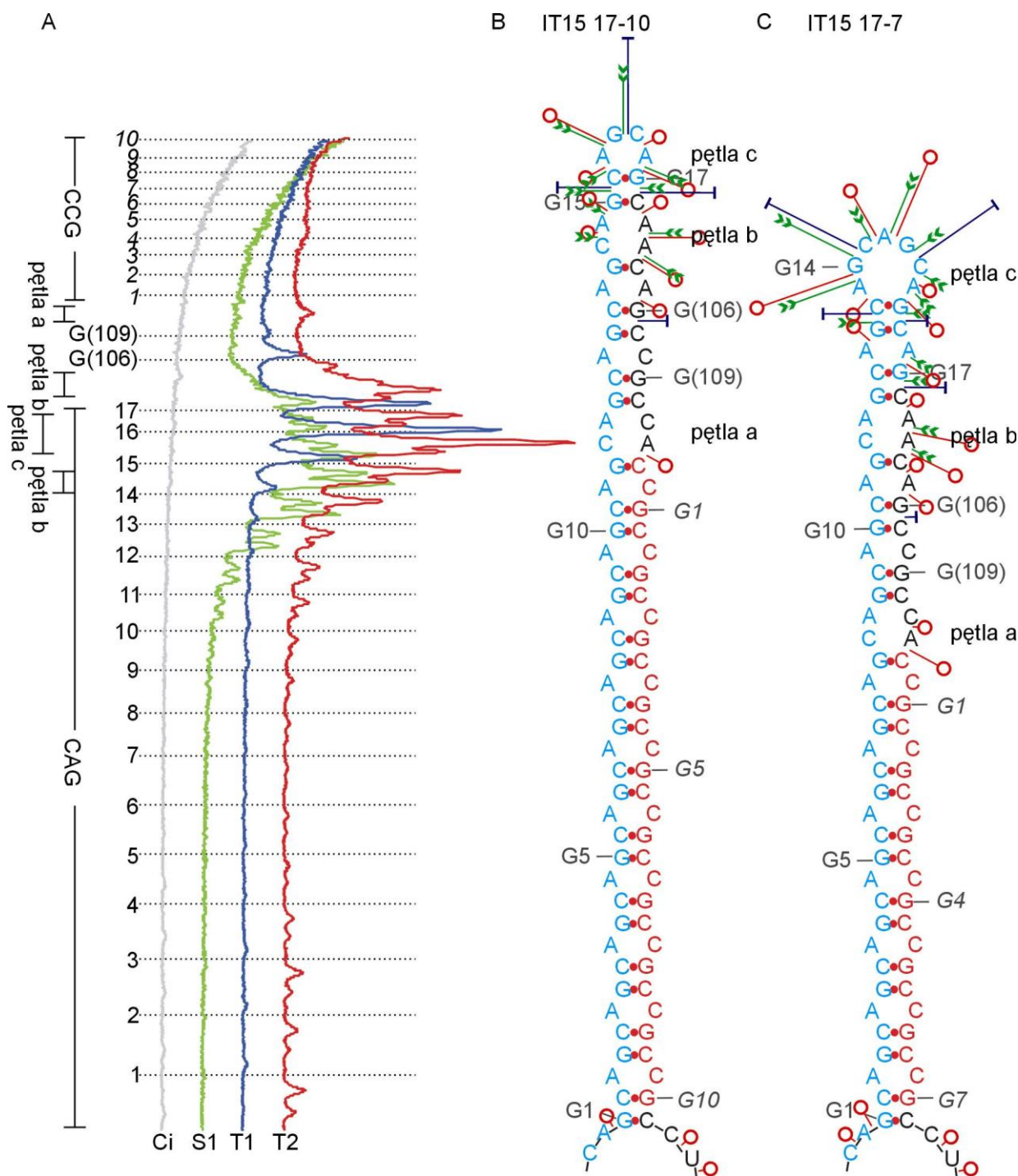
3.6 Wpływ subtelnych zmian liczby powtórzeń CAG i CCG na strukturę regionu powtórzeń.

Analizowane dotąd modelowe transkrypty pozwoliły stwierdzić występowanie oddziaływania powtórzeń CAG i CCG w tworzonej strukturze drugorzędowej RNA. Obecność tego oddziaływania stwierdzono również w transkrypcie IT15 10-10 odpowiadającym rzadko występującemu w populacji allelowi. Tym bardziej interesująca stała się odpowiedź na pytanie o możliwość wystąpienia takiej struktury w transkryptach reprezentujących dominujące w populacji allele, w których liczba powtórzeń CAG znacznie przewyższa liczbę powtórzeń motywu CCG. W celu określenia stopnia homogenności strukturalnej przeanalizowano szybkość migracji w warunkach żelu niedenaturującego (ryc. 3.10) cząsteczek RNA zawierających powtórzenia trójnukleotydydowe w liczbie odpowiadającej pełnemu zakresowi zmienności w populacji. Elektroforeza w żelu buforowanym zarówno 0,5x TB (6.4.2) (ryc. 3.10) jak i w warunkach sondowania struktury, nie wykazała zmian tempa migracji niezgodnych ze zróżnicowaną długością analizowanych cząsteczek co pozwala stwierdzić, że analizowane transkrypty przyjmują jedną formę strukturalną



Rycina 3.10. Analiza homogenności struktury transkryptów odpowiadających fragmentom mRNA genu *IT15* zawierającym region powtórzeń trójnukleotydydowych. Badane transkrypty reprezentują pełen zakres zmienności liczby powtórzeń motywu CAG obserwowany w populacji polskiej. Transkrypty poddawano elektroforezie w warunkach niedenaturujących w celu porównania szybkości cząsteczek w 10% żelu poliakryloamidowym (29:1) buforze 0,5x TB (6.4.2).

lub przyjmowane różne warianty struktury są bardzo podobne do siebie. Podstawowym modelem wybranym do potwierdzenia występowania postulowanego oddziaływania pomiędzy ciągami powtórzeń i udzielenia odpowiedzi na pytanie jakiego typu strukturę utworzą nie oddziałujące z ciągiem CCG „nadmiarowe” powtórzenia CAG został transkrypt IT15 17-10 odpowiadający jednemu z najczęstszych w populacji polskiej (ryc. 3.4 B) allelowi zawierającemu 17 powtórzeń CAG i 10 CCG (tab. 3.1). W profilu trawień regionu powtórzeń tego transkryptu przez nukleazy specyficzne dla regionów pojedynczonicowych widoczne są intensywne cięcia zlokalizowane w obrębie ostatnich powtórzeń CAG i pierwszych nukleotydów sekwencji rozdzielającej powtórzenia (ryc. 3.11). Najintensywniejsze z nich występujące pomiędzy 15 a 17 powtórzeniem CAG determinują charakter struktury pętli terminalnej. Nukleaza S1 silnie przecina wiązania internukleotydowe za resztami A i G 16 powtórzenia CAG, a także słabiej wiązanie w miejscu G powtórzenia 15 (ryc. 3.11 A linia zielona). Spośród stosowanych sond strukturalnych nukleotydy niesparowane od sparowanych najlepiej różnicują w tym przypadku rybonukleazy o specyficzności sekwencyjnej. RNaza T1 najsilniej trawi wiązanie po G powtórzenia 16 wskazując ją jako element zdecydowanie pojedynczonicowy (ryc. 3.11 A linia niebieska). Słabsze cięcia po G15 i G17 wskazują na możliwość ich zaangażowania w oddziaływanie rozluźnionej struktury trzonu. Podobnie RNaza T2 generuje najsilniejsze cięcia w miejscach C i A powtórzeń 16 i 17 (ryc. 3.11 A – czerwona linia), przy czym wyraźnie widoczna jest specyficzność tego enzymu w stosunku do A oraz całkowity brak aktywności w stosunku do G. Zdecydowanie słabsze cięcia o podobnej charakterystyce w obrębie powtórzenia 14 i pierwszych nukleotydów sekwencji rozdzielającej wskazują na obecność symetrycznej pętli wewnętrznej (pętla b) tworzonej głównie przez niesparowane reszty A. Na podstawie takiego rozkładu cięć można stwierdzić że pętla terminalna jest pętlą czteronukleotydową AGCA utworzoną z elementów 16 i 17 powtórzenia CAG z parą zamykającą CG. Pętla wewnętrzna (pętla a) tworzona przez 12 powtórzenie CAG i ostatnie nukleotydy sekwencji rozdzielającej jest słabo widoczna w profilu cięć. Wydaje się to wynikać z faktu silnej samodegradacji wiązań CpA w obrębie traktu CAG mogącej maskować cięcia rybonukleazy T2.



Rycina 3.11. Struktura regionu powtórzeń trójnukleotydydowych we fragmentach mRNA genu *IT15* zawierających różną liczbę powtórzeń CCG. A. Intensywność cięć generowanych przez sondy strukturalne w regionie powtórzeń transkryptu zawierającego 17 powtórzeń CAG i 10 CCG. B. Struktura regionu zawierającego 17 powtórzeń CAG i 10 CCG we fragmencie mRNA genu *IT15*. C. Struktura regionu zawierającego 17 powtórzeń CAG i 7 CCG we fragmencie mRNA genu *IT15*. Oznaczenia jak na rycinie 3.7.

Trzon struktury spinki tworzonej przez region powtórzeń trójnukleotydowych w transkrypcie IT15 17-10 stanowią oddziałujące ze sobą 10 pierwszych powtórzeń CAG i 10 powtórzeń CCG. Jego struktura wewnętrzna jest, tak jak w przypadku cząsteczki IT15 10-10 złożona z krótkich podwójnoniciowych regionów GC CG połączonych niesparowaniami AC.

Wpływ zmian liczby powtórzeń motywu CCG na strukturę regionu powtórzeń trójnukleotydowych mRNA genu *IT15* badano na przykładzie transkryptu IT15 17-7 (tab. 3.1) reprezentującego najczęstszy allel w populacji polskiej (ryc. 3.4 B). Grupa najintensywniejszych cięć generowanych przez sondy specyficzne w stosunku do regionów pojedynczoniciowych jest zlokalizowana pomiędzy 13 a 16 powtórzeniem CAG i definiuje pętlę terminalną (ryc. 3.11 C – pętla c). Jest ona pętlą siedmionukleotydową AGCAGCA utworzoną z elementów 14, 15 i 16 powtórzenia CAG z parą zamykającą CG. Słabe cięcia rybonukleazą T1 w obrębie trzonu w pobliżu pętli terminalnej (w miejscu reszt G13, G16 i G17) wskazują na rozluźnienie jego struktury tym regionie. Strukturę trzonu u podstawy pętli terminalnej, w którym obecne są niesparowane reszty A charakteryzują cięcia RNazy T2 w obrębie powtórzeń 13 i 17. Druga grupa cięć we wzorze trawień regionu powtórzeń transkryptu IT15 17-7 jest zlokalizowana w regionie sekwencji oddzielającej ciągi trójnukleotydowe (ryc. 3.11 C – pętla b i a). Sekwencja ta oddziałując z powtórzeniami CAG tworzy dwie czteronukleotydowe pętle wewnętrzne. Ich regiony 5' tworzone z powtórzeń CAG nie są wyraźnie widoczne w profilu cięć. Wydaje się to wynikać, z obserwowanej także w transkrypcie IT15 17-10, samodegradacji wiązań CpA w obrębie ciągu powtórzeń CAG transkryptu IT15 17-7, mogących maskować cięcia rybonukleazy T2. W porównaniu do transkryptu IT15 10-10 stopień samodegradacji wiązań CpA w transkrypcie IT15 17-7 jest znacznie większy. Trzon struktury spinki tworzonej przez region powtórzeń trójnukleotydowych w transkrypcie IT15 17-7 stanowią oddziałujące ze sobą 7 pierwszych powtórzeń CAG i 7 powtórzeń CCG. Jego struktura wewnętrzna jest taka jak opisano dla transkryptów IT15 10-10 i IT15 17-10. Skrócenie ciągu CCG powoduje, że pętla terminalna ulega oddaleniu od regionu oddziaływania powtórzeń CAG z sekwencją rozdzielającą poprzez utworzenie pomiędzy pętlą b i c trzonu z dwóch powtórzeń CAG.

3.7 Wzrost liczby powtórzeń CAG nie zaburza oddziaływania pomiędzy CAG i CCG w regionie powtórzeń mRNA genu *IT15*.

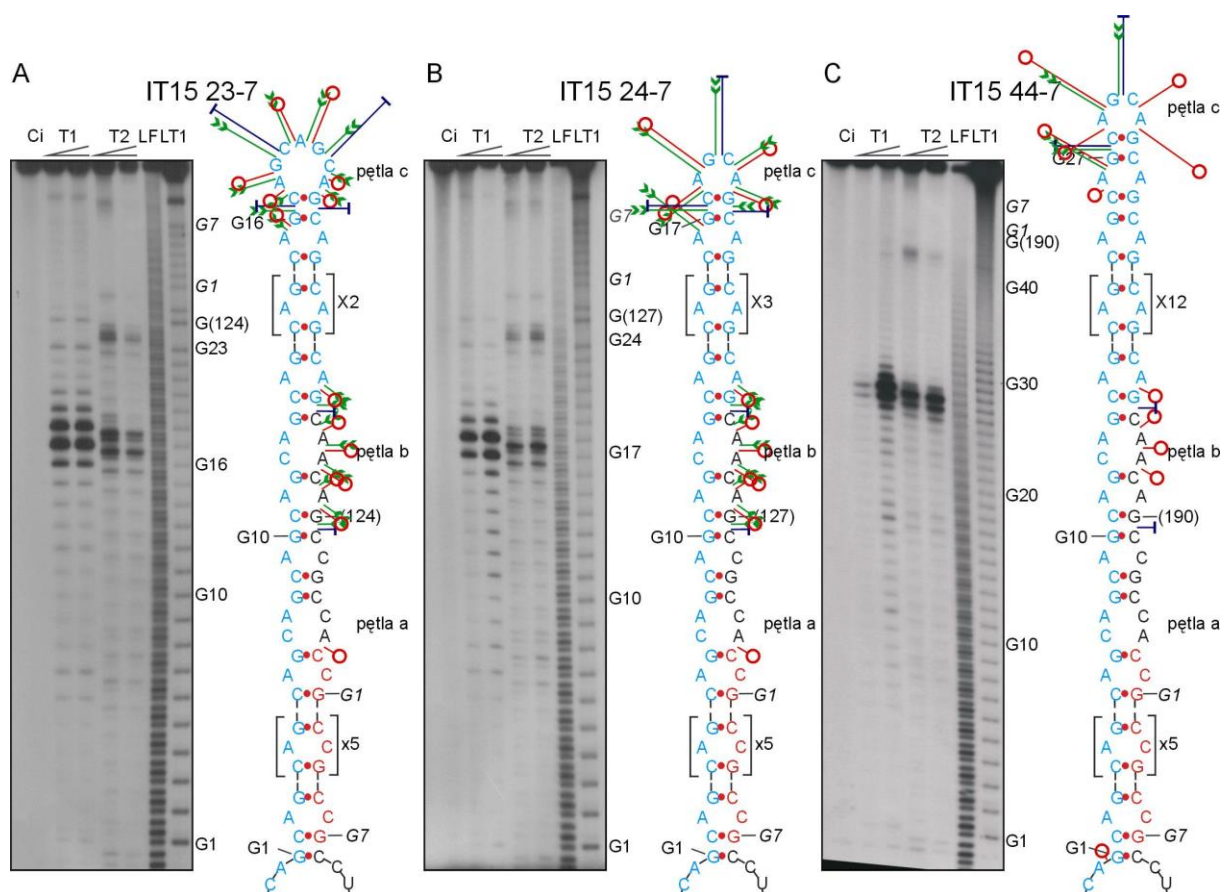
Wyniki przeprowadzonych analiz wpływu niewielkich zmian liczby powtórzeń CCG i CAG w mRNA genu *IT15* pokazują, że prowadzą one do istotnych zmian wielkości pętli terminalnej struktury spinki tworzonej przez powtórzenia. Struktura trzonu i zasada jego tworzenia nie ulega zmianom. W tym świetle ważne staje się zadanie pytania jaki wpływ na opisaną strukturę ma dalszy wzrost liczby powtórzeń CAG? W celu udzielenia odpowiedzi na to pytanie przeprowadzono analizę struktury transkryptów *IT15* 23-7 i *IT15* 24-7 odpowiadających wydłużonym allelom normalnym oraz transkryptu *IT15* 44-7, który jest fragmentem mRNA *IT15* powstającego ze zmutowanego allelu genu *IT15*.

Również w tych cząsteczkach stwierdzono taką, jak opisano wcześniej strukturę sekwencji otaczających region powtórzeń trójnukleotydowych, co ponownie potwierdziło wcześniejsze obserwacje niezależności struktury sekwencji specyficznej od zmian długości ciągów w badanym mRNA. We wzorach trawień sondami strukturalnymi obu transkryptów *IT15* 23-7 i *IT15* 24-7 widoczne są dwie grupy cięć (ryc. 3.12 – A i B, autoradiogramy). Jedna z nich występuje w obrębie ciągu powtórzeń CAG a druga tuż za jego końcem 3' w obrębie nukleotydów sekwencji rozdzielającej. Podobnie jak w innych cząsteczkach analizowanych w tej pracy region powtórzeń CCG nie jest trawiony przez żadną ze stosowanych sond. Wzór cięć wewnątrz ciągu powtórzeń CAG wskazuje na obecność pętli terminalnej (ryc. 3.12 – pętla c). W transkrypcie *IT15* 23-7 jest to siedmionukleotydowa pętla tworzona przez elementy powtórzeń CAG od 17 do 19. Natomiast w transkrypcie *IT15* 24-7 występuje czteronukleotydowa pętla terminalna zlokalizowana w 18 i 19 powtórzeniu. W transkrypcie *IT15* 44-7 rozkład miejsc trawień jest podobny do obserwowanego w cząsteczce *IT15* 24-7. Najbardziej reaktywny jest region nukleotydów od G27 do A29 co odpowiada schematowi trawień regionu czteronukleotydowej pętli terminalnej struktury spinki tworzonej przez „nadmiarowe” powtórzenia CAG.

Druga grupa cięć widocznych na rycinie 3.12 odpowiada stronie 3' symetrycznej pętli wewnętrznej (pętli b) tworzonej przez reszty CA 12 powtórzenia CAG i AA będące 2 i 3 nukleotydem sekwencji rozdzielającej ciągu trójnukleotydowe przyjmujące strukturę analogiczną do obserwowanej w transkrypcie *IT15* 17-7. Kolejna czteronukleotydowa pętla wewnętrzna (pętla a) jest tworzona przez elementy 8 powtórzenia CAG oraz dwa ostatnie nukleotydy sekwencji rozdzielającej. Nie jest

ona rozpoznawana w wyraźny sposób przez żadną ze stosowanych sond strukturalnych. Różnica pomiędzy trawieniem pętli wewnętrznych może wynikać z różnicy ich składu nukleotydowego. Pętla nie rozpoznawana złożona jest z niesparowań CA i AC.

Fakt występowania w transkryptach IT15 23-7, IT15 24-7 i IT15 44-7 trawień w powtórzeniach sąsiadujących z pętlą terminalną wskazuje na możliwość alternatywnego usytuowania tej pętli, w którym jest ona przesunięta o jedno powtórzenie w kierunku 5'. Na występowanie takiej „poślizgniętej” formy strukturalnej wskazują także cięcia w obrębie ostatniego powtórzenia CAG. Reszty A i G tego powtórzenia stanowią w tej formie fragment pętli wewnętrznej – pętla b, która uległa



Rycina 3.12. Struktura regionu powtórzeń trójnukleotydowych zawierających rosnącą liczbę powtórzeń CAG we fragmencie mRNA genu *IT15*. A. Wzór cięć w regionie powtórzeń transkryptu IT15 23-7. Obok autoradiogramu przedstawiono strukturę regionu powtórzeń tego transkryptu. B. Wzór cięć i struktura regionu powtórzeń trójnukleotydowych w transkrypcie IT15 24-7. C. Wzór cięć i struktura regionu powtórzeń trójnukleotydowych w transkrypcie IT15 44-7 odpowiadającym fragmentowi zmutowanego mRNA genu *IT15*. We wszystkich przypadkach pokazano profil cięć generowanych przez RNazy T1 i T2. Oznaczenia jak na rycinie 3.7

przekształceniu w pętlę niesymetryczną (2 na 4). Podobne „poślizgnięte” formy strukturalne obserwowano już w analizie struktury powtórzeń (CAG)_n, która nie była stabilizowana przez silne oddziaływanie regionów otaczających u podstawy trzonu struktury spinki.

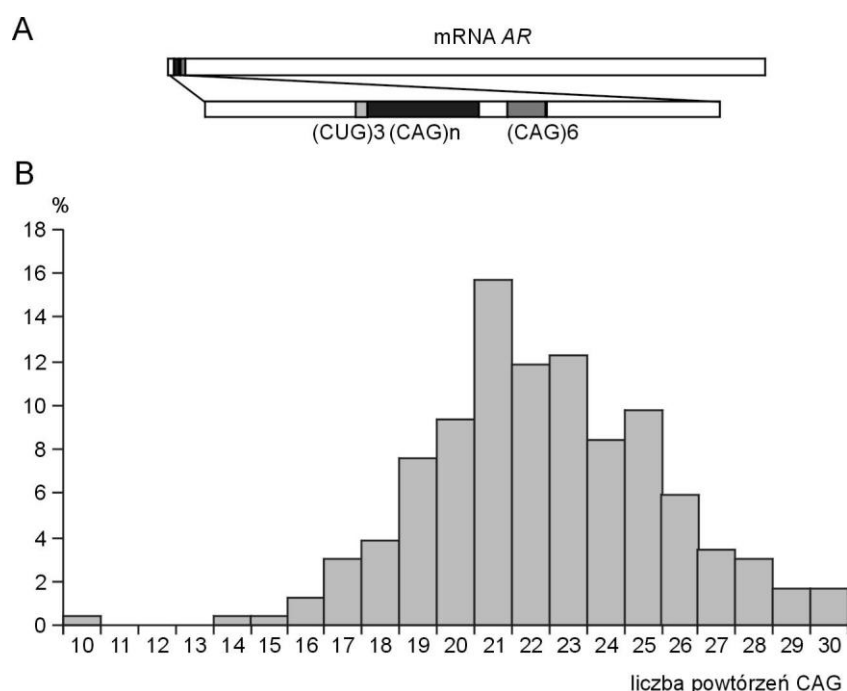
Występowanie pętli terminalnej w obrębie powtórzeń CAG w około 2/3 długości tego ciągu świadczy o tym, że w transkryptach zawierających wydłużony ciąg tych powtórzeń zachowane jest oddziaływanie powtórzeń CAG i CCG. Struktura trzonu powstałego w wyniku oddziaływania powtórzeń CAG z CCG jest taka jak proponowana dla wcześniej opisanych cząsteczek. Świadczy o tym fakt mniejszej intensywności cięć RNazy T2 w regionie 1 do 7 powtórzenia CAG w stosunku do mocniejszego trawienia wiązań CpA i ApG w dalszej części tego ciągu. Jest to widoczne na przykładzie transkryptu IT15 24-7 (fig. 3.12 – autoradiogram). Charakterystyka ta potwierdza fakt nie uczestniczenia reszt A w tworzeniu par w strukturze trzonu. Zróznicowanie to wskazuje także na inny charakter struktury spinki w jej dalszej części. Wynika to z faktu, że powtórzenia CAG nie oddziałujące z traktem CCG tworzą typową dla tego rodzaju sekwencji strukturę spinki.

3.8 Polimorfizm powtórzeń CAG w genie *AR* związanym z rdzeniowo – opuszkowym zanikiem mięśni typu Kennedy’ego.

Wśród genów, w których występują mutacje dynamiczne powodujące choroby neurodegeneracyjne należące do grupy TREDs, poza genem *IT15* znajduje się jeszcze jeden gen zawierający dwa ciągi powtórzeń trójnukleotydowych zlokalizowane blisko siebie. Jest nim gen receptora androgenu – *AR*. W pierwszym eksonie tego genu, w otwartej ramce odczytu znajduje się polimorficzny ciąg powtórzeń CAG. 18 par zasad w kierunku 3’ rozpoczyna się drugi ciąg powtórzeń tego samego motywu trójnukleotydowego (ryc. 3.13 A). Ten drugi ciąg składa się z 6 powtórzeń motywu CAG i nie wykazuje polimorfizmu. Dodatkową różnicą pomiędzy genami *IT15* i *AR* jest długość sekwencji rozdzielającej ciągi powtórzeń. W genie receptora androgenu jest o sześć par zasad dłuższa niż ta która oddziela powtórzenia CAG od CCG w genie *IT15*. Ponadto jest ona całkowicie specyficznym odcinkiem DNA, nie przypominającym w najmniejszym stopniu regionu pseudopowtórzeń obecnego w *IT15*.

W celu określenia zakresu liczby powtórzeń CAG w genie *AR* w populacji polskiej przeprowadzono analizę długości produktów amplifikacji regionu DNA zawierającego ten ciąg wraz z sekwencjami otaczającymi. Populację polską reprezentowało w tych badaniach 120 zdrowych, nie spokrewnionych osób zamieszkujących region Wielkopolski i Górnego Śląska nie selekcionowanych pod kątem historii rodzinnej chorób neurologicznych. Otrzymane wyniki wskazują na unimodalny rozkład liczby powtórzeń CAG, w zakresie od 9 do 30 (ryc. 3.13 B). Najczęściej występuje allel z 21 powtórzeniami CAG. Uzyskane dane wskazują na występowanie w populacji polskiej alleli krótszych niż najkrótsze obserwowane w innych populacjach. Natomiast w zakresie dłuższych ciągów nie różnią się one od stwierdzanych w innych populacjach. Na przykład w populacji skandynawskiej liczba powtórzeń CAG mieści się w zakresie od 15 do 29.

Dodatkową cechą sekwencji regionu powtórzeń CAG w genie *AR* jest obecność trzech tripletów CTG powyżej traktu polimorficznego. Z punktu widzenia struktury mRNA mogą one mieć istotne znaczenie ze względu na całkowitą komplementarność do sekwencji ciągu CAG.



Rycina 5.13. Częstość występowania ciągów CAG o określonej liczbie powtórzeń w genie *AR* w populacji polskiej. A. Schemat sekwencji regionu powtórzeń trójnukleotydujących we fragmencie genu *AR* (kolor biały – sekwencje specyficzne; kolor czarny – polimorficzny ciąg powtórzeń CAG; kolor szary – ciąg powtórzeń (CAG)⁶; kolor jasnoszary – (CTG)³). B. Częstość występowania różnej długości ciągów powtórzeń CAG w eksonie 1 genu *AR* (N = 240).

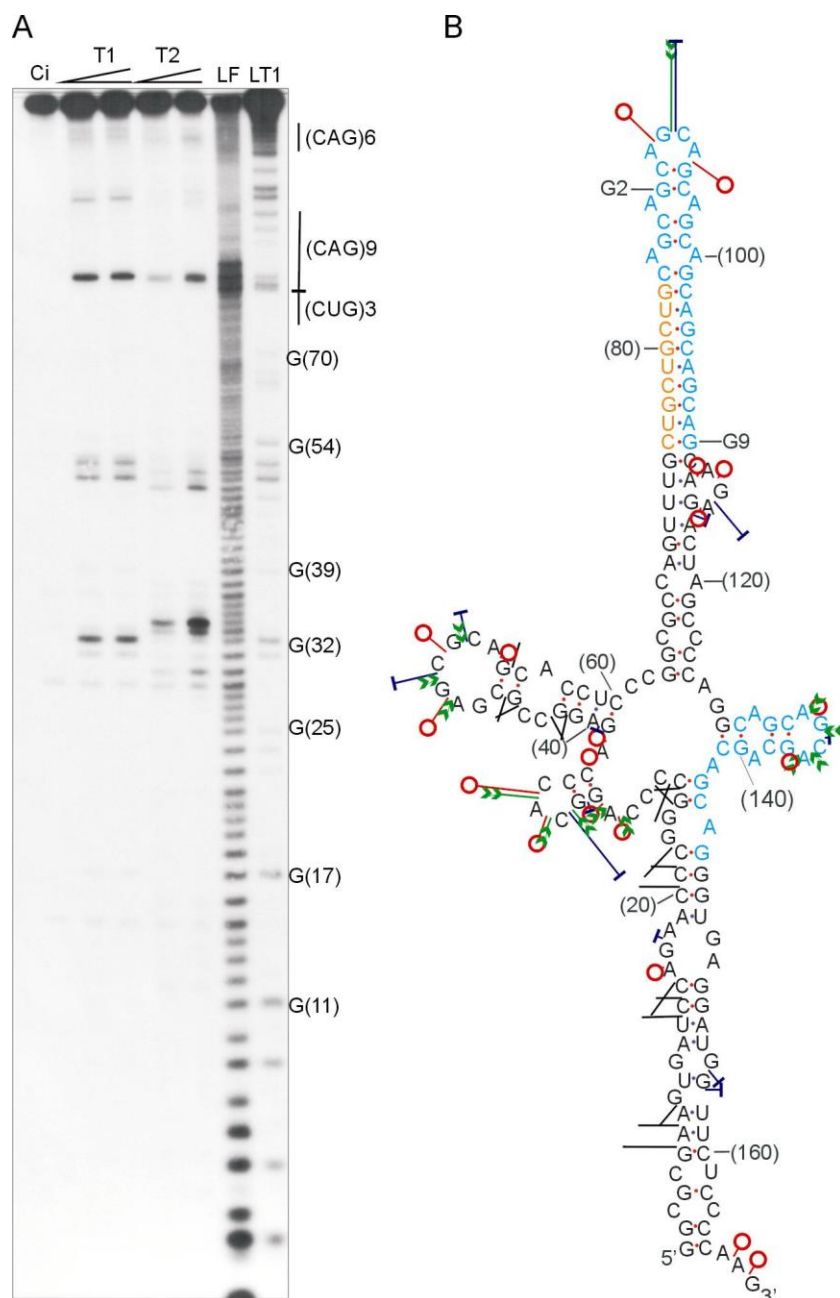
Na podstawie poznanego zróżnicowania w populacji dokonano wyboru alleli których transkrypty użyto do badania struktury regionu powtórzeń mRNA genu *AR*. Wybrane transkrypty stwarzają możliwość odpowiedzi na pytanie czy sąsiadujące ze sobą ciągi powtórzeń CAG oddziałują ze sobą czy tworzą niezależne struktury?

3.9 Ciągi powtórzeń CAG we fragmencie mRNA genu *AR* nie oddziałują ze sobą.

W celu zbadania możliwości oddziaływania ze sobą ciągów powtórzeń CAG we fragmencie mRNA genu *AR* przeprowadzono sondowanie struktury transkryptu *AR-CAG9*. Zawiera on 9 powtórzeń CAG w ciągu polimorficznym. Specyficzna sekwencja po stronie 5' ma długość 83 nukleotydów. Zawiera ona trzy powtórzenia motywu CUG poprzedzające bezpośrednio sekwencję powtórzeń trójnukleotydowych CAG. Natomiast sekwencja otaczająca po stronie 3' ma długość 58 nukleotydów. Zawiera ona niepolimorficzny ciąg powtórzeń (CAG)₆ rozpoczynający się 18 nukleotydów powyżej ciągu polimorficznego.

Zastosowane do badań sondy enzymatyczne specyficzne w stosunku do regionów pojedynczoniowych generują w analizowanym transkrypcie wyraźne grupy cięć (ryc. 3.14). Najintensywniej trawiony jest fragment sekwencji specyficznej po stronie 5' ciągu powtórzeń. Wiązania pomiędzy nukleotydami od 29 do 35 są przecinane ze średnią intensywnością przez nukleazę S1. Użyte RNazy pokazują silne zróżnicowanie reaktywności poszczególnych wiązań tego odcinka. RNaza T1 przecina wiązanie fosfodiesterowe tylko w miejscu G32 nie wykazując aktywności w stosunku do G31. Natomiast RNaza T2 generuje słabe cięcia w miejscach C30 i A31, zachowując wyraźną specyficzność w stosunku do A. Najintensywniej enzym ten przecina wiązanie po A34. Taki wzór cięć wskazuje na istnienie w tym regionie transkryptu odcinka nie oddziałującego z innym fragmentem cząsteczki. Nie jest on całkowicie pojedynczoniowy, ponieważ występujące w nim dwie reszty G31 i G32 tworzą pary z oddalonymi od nich o trzy nukleotydy dwiema resztami C36 i C37. Fragment sekwencji specyficznej po stronie 5' tej cząsteczki oddziałuje z sekwencją otaczającą ze strony 3' (ryc. 3.14 B). Kolejna grupa cięć widocznych w obrębie specyficznej sekwencji 5' (ryc. 3.14 A) występuje około 20 nukleotydów dalej w kierunku 3'. Silne trawienie wiązań w miejscach G49 i G51 przez RNazę T1 oraz A48

i równie silne w miejscu C50 przez RNazę T2 wraz z aktywnością nukleazy S1 wskazują na zlokalizowanie w tym regionie siedmionukleotydu pętli terminalnej struktury spinki tworzonej przez nukleotydy od G39 do C60. Jej trzon, pomimo że występuje w nim niesymetryczna pętla wewnętrzna trawiony jest przez rybonukleazę V1 specyficzną dla regionów podwójnoniciowych.



Rycina 3.14 Polimorficzny ciąg powtórzeń CAG we fragmencie mRNA genu *AR* przyjmuje strukturę spinki stabilizowaną oddziaływaniem z sekwencją otaczającą. A. Autoradiogram przedstawiający wzór cięć generowanych w całej cząsteczce AR-CAG9 przez RNazy T1 i T2. B. Schemat struktury drugorzędowej transkryptu AR-CAG9; kolorem niebieskim zaznaczono nukleotydy powtórzeń CAG, kolor czerwony – sekwencja (CUG)3. Pozostałe oznaczenia jak na ryc. 3.7

W kolejnym reaktywnym regionie występuje pojedyncze silne cięcie w miejscu G92 przez nukleazę S1 i rybonukleazę T1 oraz równie silne cięcie rybonukleazą T2 w miejscu A91 i A94. Jest to jedyne miejsce dostępne dla tych sond w sekwencji „polimorficznego” ciągu CAG o długości 9 powtórzeń. Taki wzór cięć oraz odporność reszt G powtórzeń CUG na trawienie RNazą T1 w warunkach semidenaturujących wskazuje na tworzenie przez powtórzenia CAG struktury spinki. Jest ona silnie stabilizowana w wyniku oddziaływań trzech ostatnich, 3' końcowych powtórzeń CAG z sekwencją (CUG)₃ znajdującą się po stronie 5' ciągu CAG. W wyniku tego oddziaływania oraz oddziaływania specyficznych sekwencji otaczających 5' i 3' powstaje region podwójnoniciowy o długości 12 par zasad. Zaangażowanie części powtórzeń CAG w oddziaływanie z sekwencją otaczającą powoduje, że typową dla powtórzeń CAG strukturę spinki z trzonem złożonym z krótkich odcinków podwójnoniciowych GC CG oddzielonych od siebie niesparowaniami AA tworzy jedynie 6 z 9 powtórzeń. Czteronukleotydomowa pętla terminalna tej struktury ma budowę typową dla cząsteczek zawierających parzystą liczbę powtórzeń trójnukleotydowych.

Sekwencje otaczające region powtórzeń po stronie 5' i 3' oddziałują ze sobą tworząc odcinek sparowany przedłużający region podwójnoniciowy o kolejne 10 par zasad. W strukturze tej występują dwie symetryczne pętle wewnętrzne oraz niesymetryczna pętla wewnętrzna (ryc. 3.14).

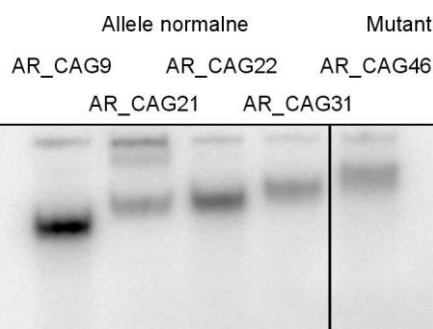
Ciąg (CAG)₆ także przyjmuje strukturę spinki. Jest ona tworzona przez zaledwie cztery powtórzenia. Sondy strukturalne przecinają wiązania w obrębie drugiego i trzeciego powtórzeń w sposób charakterystyczny dla czteronukleotydomowej pętli terminalnej. Sytuacja ta wynika z faktu zaangażowania ostatniego powtórzeń CAG w oddziaływanie z sekwencją otaczającą po stronie 5' oraz tworzenie łącznika pomiędzy podwójnoniciowymi strukturami powtórzeń i sekwencji specyficznej.

W transkrypcje AR-CAG₉ polimorficzne powtórzenia CAG przyjmują strukturę typu spinki silnie stabilizowaną przez znajdujący się u jej podstawy region podwójnoniciowy tworzony przez 3' końcowy odcinek traktu i część sekwencji otaczającej po stronie 5'. Pozostała część cząsteczki przyjmuje zwartą strukturę złożoną z pięciu regionów podwójnoniciowych połączonych krótkimi odcinkami pojedynczoniciowymi. Niepolimorficzny ciąg (CAG)₆ tworzy strukturę niewielkiej spinki.

3.10 Wzrost liczby powtórzeń polimorficznego ciągu CAG nie zmienia zasadniczo struktury regionu powtórzeń mRNA genu *AR*.

Na podstawie określonej struktury transkryptu AR-CAG9 można sądzić, że w przypadku regionu powtórzeń CAG mRNA genu *AR* ich wydłużanie nie będzie prowadzić do zasadniczych zmian strukturalnych poza wydłużeniem trzonu struktury spinki. W celu sprawdzenia tych przypuszczeń przeprowadzono najpierw badania transkryptów reprezentujących fragmenty mRNA genu *AR* zawierające krótki, średniej długości oraz długi ciąg powtórzeń CAG w warunkach elektroforezy w żelu niedenaturującym. Nie zaobserwowano istotnych zmian tempa migracji niezgodnych ze zróżnicowaną długością analizowanych cząsteczek co pozwala sądzić, że analizowane transkrypty przyjmują jedną formę strukturalną lub przyjmowane różne warianty struktury są bardzo podobne do siebie (ryc. 3.15).

Przeprowadzone biochemiczne analizy struktury transkryptu AR-CAG21 odpowiadającego najczęstszemu allelowi w populacji polskiej wykazały, że wydłużenie polimorficznego ciągu powtórzeń (CAG)_n nie spowodowało zmian w sposobie aranżacji struktury regionów otaczających w porównaniu z transkryptem AR-CAG9. Zachowane zostało ułożenie sekwencji specyficznej po stronie 5' oraz 3'. Zarówno struktura niezmiennego ciągu (CAG)₆ jak i pozostałych elementów strukturalnych nie uległa zmianom. Oddziaływanie ostatnich trzech powtórzeń CAG z

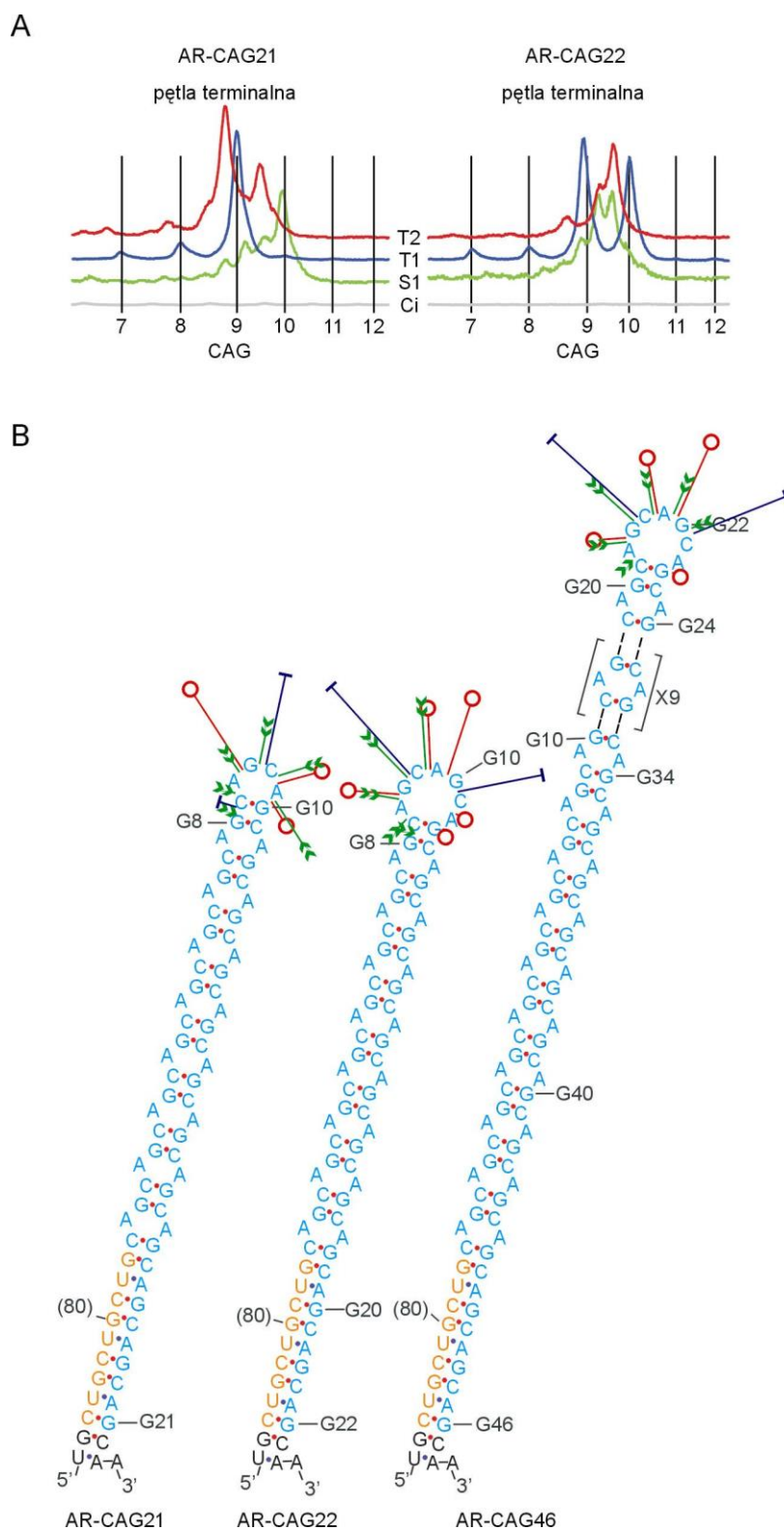


Rycina 3.15. Analiza homogenności struktury transkryptów odpowiadających fragmentom mRNA genu *AR* zawierającym region powtórzeń trójnukleotydowych, na podstawie szybkości ich migracji podczas elektroforezy w warunkach żelu niedenaturującego (10% (29:1) żel w buforze identycznym ze stosowanym w warunkach sondowania struktury (bufor Mg 1/K – 6.4.6.). Badane transkrypty obejmują pełen zakres zmienności liczby powtórzeń CAG obserwowany w populacji polskiej.

sekwencją (CUG)³ także w tym transkrypcie tworzy silny region sparowany stabilizujący strukturę spinki, w którą układają się pozostałe powtórzenia CAG. Ma ona typową strukturę trzonu złożonego ze sparowanych odcinków GC CG przedzielonych niesparowaniami AA (ryc. 3.16 B). Pętla terminalna jest rozpoznawana przez sondy strukturalne w sposób bardzo precyzyjny. RNaza T1 generuje tylko jedno cięcie w miejscu G z 9 powtórzenia CAG, zaś RNaza T2 z podobną siłą trawi wiązanie w miejscu A tego samego powtórzenia i słabiej w miejscach C i A kolejnego (ryc. 3.16 A). Jedynie nukleaza S1 przecina więcej wiązań, wśród których najsilniej trawione jest to w miejscu A z 10 powtórzenia motywu CAG. Nie trawi ona jednak wiązań pomiędzy innymi nukleotydami niż te, które tworzą czteronukleotydową pętlę terminalną.

W celu zaobserwowania zmian wynikających z wydłużenia ciągu o jedno powtórzenie CAG zbadano strukturę transkryptu AR-CAG22. Jediną jakościową różnicą stwierdzoną we wzorze cięć generowanych z użyciem sond enzymatycznych były cięcia w obrębie pętli terminalnej. Dwa silne cięcia w miejscach reszt guanylowych z 9 i 10 powtórzenia generowane przez RNazę T1 i najmocniejsze cięcia RNazy T2 w miejscach reszt C i A z 10 powtórzenia CAG wskazują na zwiększenie jej rozmiarów do siedmiu nukleotydów (ryc. 3.16 A i B).

Wzór cięć uzyskanych podczas analizy struktury transkryptu AR-CAG46 wskazuje, że jedyne miejsca w strukturze „polimorficznego” ciągu powtórzeń CAG dostępne dla nukleaz o specyficzności do regionów pojedynczonicowych zlokalizowane są pomiędzy resztą C z 21 a resztą A z 23 powtórzenia,. Lokalizacja silnych cięć generowanych przez RNazę T1 po G z 21 i G z 22 powtórzenia oraz cięć nukleazy S1 i RNazy T2 jest bardzo podobna do wzoru cięć obserwowanych w siedmionukleotydowej pętli terminalnej transkryptu AR-CAG22. W obrębie regionu powtórzeń CAG transkryptu AR-CAG46 nie stwierdzono innych cięć generowanych przez użyte sondy enzymatyczne. Zatem uzyskane wyniki sondowania struktury transkryptu AR-CAG46 wskazują, że wydłużanie ciągu powtórzeń CAG w mRNA genu *AR* do zakresu mutacyjnego powoduje jedynie wydłużanie trzonu stabilnej struktury typu spinki (ryc. 3.16 B). Jednocześnie ekspansja powtórzeń CAG nie powoduje zmian struktury naturalnej sekwencji otaczającej w transkryptach genu *AR*.



Rycina 3.16. Struktura regionu powtórzeń trójnukleotydu zawierającego rosnącą liczbę powtórzeń CAG we fragmentach mRNA genu *AR*. A. Intensywność cięć generowanych przez sondy strukturalne w regionie pętli terminalnych transkryptów AR-CAG21 i AR-CAG22. B. Struktura regionu polimorficznych powtórzeń trójnukleotydu w transkryptach odpowiadających fragmentom mRNA genu *AR*. Transkrypty AR-CAG21 i AR-CAG22 pochodzą z alleli najczęstszych w populacji, transkrypt AR-CAG46 pochodzi z allelu zmutowanego. Oznaczenia jak na rycinie 3.7 i 3.14.

4 Dyskusja wyników

RNA są jednymi z najważniejszych makrocząsteczek biologicznych występujących w komórkach organizmów żywych. Funkcje różnych rodzajów RNA są związane z koniecznością występowania oddziaływań z białkami i innymi wysoko- i niskocząsteczkowymi ligandami. W oddziaływaniach tych RNA przyjmuje strukturę zapewniającą pełnienie odpowiedniej funkcji. Za powstanie struktur drugo- i trzeciorzędowych odpowiedzialne są wiązania wodorowe pomiędzy grupami nukleotydów z różnych regionów cząsteczki, oddziaływania warstwowe między resztami zasad poszczególnych nukleotydów, oraz wiązania koordynacyjne utworzone z udziałem kationów jedno- i dwuwartościowych.

Najczęściej stosowanymi metodami biochemicznymi określania struktury RNA w roztworze wodnym są metody wykorzystujące zdolność niektórych nukleaz do specyficznego rozpoznawania nukleotydów występujących w obszarach pojedynczoniciowych czy podwójnoniciowych [Knapp G. (1989)]. Często wykorzystuje się również modyfikację chemiczną grup funkcyjnych nukleotydów nie uczestniczących w tworzeniu wewnątrzcząsteczkowych lub międzycząsteczkowych oddziaływań [Ehresmann C. i wsp. (1987)]. Przydatna w badaniu struktury RNA jest także specyficzna hydroliza cząsteczek RNA indukowana jonami metali [Krzyżosiak W. J. i wsp. (1988)].

4.1 Struktura spinki tworzona przez cząsteczki RNA zbudowane z powtórzeń trójnukleotydowych CAG lub CCG.

Ciągi powtórzeń trójnukleotydowych, których ekspansja jest genetycznym podłożem szeregu ludzkich chorób neurologicznych stanowią od wielu lat przedmiot intensywnych badań. Duże zainteresowanie budzi między innymi struktura jaką te motywy sekwencji przyjmują występując w transkryptach genów związanych z TREDs i jej znaczenie dla procesów patogenezy.

Prezentowane w tej pracy biochemiczne badania struktury, do których użyto sond specyficznym rozpoznających obszary pojedynczoniciowe (nukleazy S1, RNazy T1 i T2 oraz jony Pb^{2+}) pozwalających na określenie lokalizacji i rozmiarów pętli terminalnych, pętli wewnętrznych, wybruszeń i węzłów, dowiodły, że transkrypty składające się z powtórzeń CAG lub CCG przyjmują strukturę typu spinki do włosów.

Charakteryzuje się ona czteronukleotydową pętlą terminalną, zaś jej trzon składa się z krótkich regionów podwójnoniciowych złożonych z par GC CG oddzielonych od siebie pojedynczymi niesparowaniami AA lub CC, w zależności od rodzaju sekwencji powtarzającej się. Stabilność struktury tego typu zależy od rodzaju niesparowania i ma na nią wpływ sekwencja otaczająca. W sytuacji kiedy u podstawy spinki utworzonej przez powtórzenia występuje klamra złożona z kilku par GC i CG dochodzi do „zamrożenia” struktury spinki i dodatkowej jej stabilizacji. W tej sytuacji wielkość pętli terminalnej zależy od liczby powtórzeń. Kiedy ta przyjmuje wartości parzyste obserwowana jest pętla czteronukleotydowa, zaś gdy nieparzyste dochodzi do jej powiększenia o „nadmiarowe” powtórzenie do siedmiu nukleotydów. Natomiast gdy brakuje stabilizującego wpływu otoczenia możliwe jest przyjęcie przez region powtórzeń kilku alternatywnych form strukturalnych identycznych pod względem charakteru pętli terminalnej i trzonu, lecz różniących się liczbą powtórzeń zaangażowanych w tworzenie trzonu. Część powtórzeń znajdujących się na końcu 3' ciągu może pozostać w formie pojedynczoniciowej co skutkuje skróceniem trzonu spinki i przesunięciem pętli w kierunku 5'. Takie charakterystyczne dla powtórzeń trójnukleotydowych struktury określane są mianem „poślizgniętych” struktur typu spinki do włosów [Napierała M., Krzyżosiak W. J. (1997)]. Badania termodynamiczne właściwości struktur spinki tworzonych przez ciągi o długości 6 i 7 powtórzeń trójnukleotydowych potwierdzają obserwację, że struktura charakteryzująca się czteronukleotydową pętlą terminalną i pojedynczoniciowymi nukleotydami na końcu 3' jest bardziej stabilna niż spinka z pętlą siedmionukleotydową [Broda M. i wsp. (2005)].

W badaniach prowadzonych w Pracowni Genetyki Nowotworów IChB PAN w Poznaniu, których część opisuje niniejsza praca doktorska stwierdzono, że wszystkie powtórzenia typu CNG przyjmują strukturę typu spinki [Sobczak K. i wsp. (2003)]. Ponadto, w badaniach fragmentów mRNA różnych genów związanych z TREDs wykazano, że specyficzne sekwencje otaczające mogą nie wpływać na charakter tej struktury [Napierała M., Krzyżosiak W. J. (1997)] lub zwiększać jej stabilność [Michlewski G., Krzyżosiak W. J. (2004), Napierała M. i wsp. (2005)]. Natomiast jednonukleotydowe zaburzenia ciągu powtórzeń trójnukleotydowych, występujące w niektórych genach związanych z TREDs, znajdowane są w strukturze RNA w regionach pętli terminalnych lub rozgałęzień struktury typu spinki, wpływają więc na

skrócenie jej trzonu o regularnej strukturze [Sobczak K., Krzyżosiak W. J. (2004), Sobczak K., Krzyżosiak W. J. (2005), Napierała M. i wsp. (2005)].

W niniejszej pracy zbadano przy zastosowaniu metod biochemicznych, strukturę fragmentów mRNA genów *IT15* i *AR*, w których w bezpośrednim sąsiedztwie występują dwa ciągi powtórzeń trójnukleotydowych. W genie *IT15* są to dwa polimorficzne ciągi powtórzeń motywów CAG i CCG, natomiast w genie *AR* obok polimorficznego ciągu powtórzeń CAG występuje drugi, niepolimorficzny ciąg złożony z 6 powtórzeń trójnukleotydu CAG.

4.2 Powtórzenia CAG oddziałują z powtórzeniami CCG w transkrypcie genu *IT15*.

W celu określenia struktury regionu zawierającego ciągi trójnukleotydowe w transkrypcie genu *IT15* przeprowadzono badania cząsteczki zawierającej taką samą liczbę 10 powtórzeń każdego z trójnukleotydów CAG i CCG w obecności specyficznej sekwencji otaczającej. Obserwowany wzór cięć generowanych przez użyte nukleazy pozwolił stwierdzić oddziaływanie powtórzeń CAG z powtórzeniami CCG i tworzenia przez nie trzonu struktury typu spinki, którego pętlę terminalną stanowi dwunastonukleotydowa sekwencja specyficzna występująca pomiędzy nimi. Za taką strukturą przemawiał brak trawień w obrębie ciągów powtórzeń zarówno CAG jak i CCG (ryc. 3.8).

Natomiast w transkryptach składających się z ciągu CAG i naturalnej sekwencji otaczającej po stronie 5' stwierdzono tworzenie przez powtórzenia struktury typu spinki. Jest ona w transkrypcie *IT15*–17 (ryc. 3.6) tworzona przez piętnaście z siedemnastu powtórzeń, a co ciekawsze, w transkrypcie *IT15*–10 przez osiem z dziesięciu powtórzeń CAG (ryc. 3.7.). Charakteryzuje się ona o typową dla powtórzeń CNG budową trzonu i czteronukleotydową pętlą terminalną [Sobczak K. i wsp. (2003)]. Stwierdzenie tworzenia struktury spinki przez ciąg złożony z 8 powtórzeń motywu trójnukleotydowego w zestawieniu z obserwacją braku trawień w obrębie ciągów CAG i CCG w transkrypcie *IT15* 10-10 wskazuje na oddziaływanie ze sobą sekwencji powtarzających się.

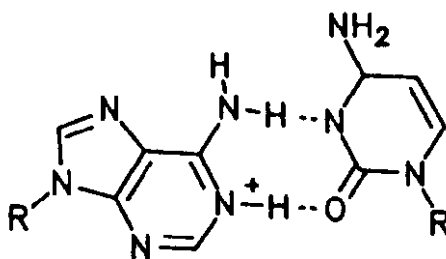
Dalszych dowodów potwierdzających oddziaływanie ze sobą ciągów CAG i CCG dostarczyły badania struktury modelowych cząsteczek RNA zawierających 20 powtórzeń motywu CAG – (CAG)₂₀, po 10 powtórzeń motywów CAG i CCG –

(CAG)₁₀(CCG)₁₀ oraz 20 CCG – (CCG)₂₀. Transkrypty te pozwoliły na zbadanie tworzonych przez nie struktur z pominięciem wpływu naturalnych sekwencji otaczających i regionu rozdzielającego występującego w transkrypcie genu *IT15*. Analiza z zastosowaniem nukleaz potwierdziła we wszystkich analizowanych cząsteczkach przyjmowanie typowej dla powtórzeń (CNG)_n struktury spinki. Jej trzon złożony jest z krótkich regionów sparowanych GC CG oddzielonych pojedynczymi niesparowaniami AA dla transkryptu (CAG)₂₀ i CC dla (CCG)₂₀, natomiast w przypadku transkryptu (CAG)₁₀(CCG)₁₀ występuje niesparowanie AC (ryc. 3.9).

Jak wykazały wcześniejsze badania struktury tworzonej przez 130 powtórzeń CUG przeprowadzone przy użyciu mikroskopii elektronowej [Michalowski S. i wsp. (1999)], a przede wszystkim dane o strukturze RNA tego rodzaju powtórzeń uzyskane metodą krystalografii [Mooers B. H. i wsp. (2005)], struktura trzonu złożonego z par GC CG przedzielonych niesparowaniami UU jest bardzo podobna do struktury typowej helisy A – RNA. Ze względu na obserwowane właściwości transkryptów utworzonych z powtórzeń CNG wcześniej sugerowano [Sobczak K. i wsp. (2003)], że architektura tworzonych przez nie struktur spinki jest bardzo podobna. Autorzy pracy krystalograficznej uważają jednak, że większe podobieństwo może dotyczyć struktur, w których niesparowanymi zasadami są pirymidyny. Niesparowane reszty zasad pirymidynowych lokalizują się wewnątrz struktury podwójnoniciowej nie powodując tym samym zaburzenia helisy A – RNA [Mooers B. H. i wsp. (2005)]. Natomiast powtórzenia trójnukleotydydowe motywów zawierających w drugiej pozycji zasadę purynową (CAG i CGG) nie mogą, według tych autorów przyjmować struktury helisy A – RNA ze względu na ułożenie niesparowanych zasad na zewnątrz łańcucha [Mooers B. H. i wsp. (2005)].

Podobnie jak we wcześniejszych badaniach transkryptu CAG₁₇ [Sobczak K. i wsp. (2003)] również w przypadku transkryptów (CAG)₂₀, (CAG)₁₀(CCG)₁₀ i (CCG)₂₀ stwierdzono występowanie dwóch form strukturalnych wynikające z „poślizgu” sekwencji prowadzącego do przesunięcia pętli terminalnej w kierunku końca 5’ cząsteczki. Szczególną uwagę zwrócono na obserwowane różnice w stabilności struktury cząsteczek widoczne między innymi w zróżnicowaniu intensywności cięć generowanych przez rybonukleazę T1 w warunkach semidenaturujących. Na podstawie dostępności reszt guanylowych stwierdzono, że najbardziej stabilną strukturą jest ta, której trzon jest utworzony w wyniku oddziaływania powtórzeń CAG i CCG. Fakt ten potwierdzają również informacje o

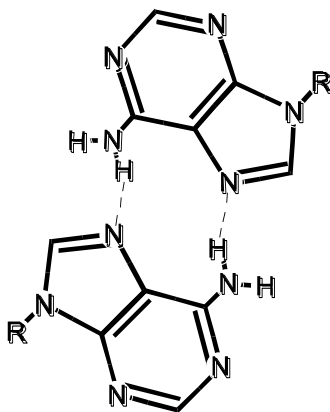
stabilizującym wpływie niekanonicznych oddziaływań AC [Limmer S. (1997), Durant P. C. i Davis D. R. (1999)]. Powstanie stabilizującego strukturę oddziaływania AC jest możliwe po protonacji reszty adeniny w pozycji N1 (ryc. 6.1). Stabilność tego typu oddziaływania zależy od pH, co pokazuje zróżnicowanie temperatury topnienia, która w pH = 7,2 wynosi 50,6 °C, a w pH = 5,0 56,8 °C [Durant P. C. i Davis D. R. (1999)]. Protonowana adenina w pozycji N1 jest donorem protonu w procesie tworzenia wiązania wodorowego w parze AC typu wobble i jej wartość pKa jest podwyższona do około 6,0. Energia swobodna oddziaływania AC zależy także od kontekstu otaczającej je sekwencji. Na przykład: w duplesie GAGCGAG3'/CUCACUC5' $\Delta G^{0}_{37} = -0,35 \pm 0,09$ kcal/mol, zaś w duplesie GAGAGAG3'/CUCCCUC5' $\Delta G^{0}_{37} = 0,31 \pm 0,11$ kcal/mol [Kierzek R. i wsp. (1999)]. W proponowanych modelach struktury badanych cząsteczek kontekst sekwencji otaczającej niekanoniczne oddziaływanie AC jest podobny do przytoczonego w pierwszym przykładzie. W związku z tym można przypuszczać, że jego energia swobodna jest zbliżona do $\Delta G^{0}_{37} = -0,35$ kcal/mol. Występowanie tej niestandardowej pary przed parą GC tworzącą trzon struktury spinki, jak ma to miejsce w proponowanej strukturze transkryptu (CAG)10(CCG)10, nie wpływa znacząco na geometrię helisy [Limmer S. (1997)].



Rycina. 6.1. Schemat wiązań wodorowych w niestandardowej parze zasad A+C typu wobble.

Dla porównania wpływ oddziaływania AA (ryc. 6.2) na stabilność duplesu RNA zależy w dużym stopniu od jego położenia w helisie. Wraz ze wzrostem odległości od końca helisy rośnie destabilizujący wpływ jego obecności: gdy jest drugą parą helisy $\Delta G^{0}_{37} = -1,30 \pm 0,24$ kcal/mol, a gdy trzecią $\Delta G^{0}_{37} = 0,37 \pm 0,08$ kcal/mol. Zdecydowanie mniejszy wpływ na zmiany energii swobodnej oddziaływania AA ma natomiast kontekst sekwencji otaczającej w zależności od której ΔG^{0}_{37} przyjmuje

wartości od 0,67 kcal/mol do $1,13 \pm 0,25$ kcal/mol [Kierzek R. i wsp. (1999)]. W porównaniu z oddziaływaniem AC oddziaływanie AA wpływa zwykle bardziej destabilizująco na strukturę, w obrębie której występuje, w prawie każdym kontekście sekwencji.



Rycina 6.2. Schemat wiązań wodorowych w niestandardowej parze zasad AA typu trans Watson–Crick/Hoogsteen.

Ponadto krystalograficzne dane dotyczące struktury powtórzeń CUG wskazują, że oddziaływania warstwowe pomiędzy parami zasad GC/CG są głównym czynnikiem stabilizującym, natomiast oddziaływania warstwowe zasad CU i UG mają niewielki wpływ destabilizujący [Mooers B. H. i wsp. (2005)]. W świetle tych informacji możliwość wystąpienia pary AC typu wobble wydaje się także zwiększać stabilność regionu podwójnoniciowego, w którym występuje, poprzez możliwy udział w oddziaływaniach warstwowych. Można też przypuszczać, że para AC nie będzie powodowała zaburzenia struktury podwójnoniciowej helisy zbliżonej do helisy A – RNA.

Kolejnych dowodów na poparcie tezy o preferencyjnym oddziaływaniu powtórzeń CAG z sąsiadującymi powtórzeniami CCG i korzystnym wpływie niesparowania AC na stabilność dostarczyła termodynamiczna charakterystyka struktur transkryptów (CAG)₂₀, (CAG)₁₀(CCG)₁₀ i (CCG)₂₀ (ryc. 3.9). Różnica trwałości obserwowanych struktur typu spinki wyrażona jako różnica temperatury topnienia (T_m) wynosi $\Delta T_m = 15,1$ °C między transkryptami (CAG)₁₀(CCG)₁₀ i (CAG)₂₀. Natomiast pomiędzy bardziej trwałą strukturą utworzoną z 20 powtórzeń motywu CAG a mniej trwałą strukturą (CCG)₂₀ różnica ta wynosi $\Delta T_m = 4,2$ °C. Podobne zróżnicowanie widoczne jest w przypadku porównania energii swobodnej

badanych struktur. Różnica pomiędzy wartościami energii swobodnej $\Delta G_{37^\circ\text{C}}$ charakteryzującej stabilniejszą strukturę transkryptu (CAG)₁₀(CCG)₁₀ i strukturę (CAG)₂₀ to – 10,4 kcal/mol. O podobnej stabilności struktur tworzonych przez powtórzenia CAG lub CCG świadczą także wyniki badań właściwości termodynamicznych dupleksów utworzonych przez bardzo krótkie ciągi tych powtórzeń oraz struktur spinki tworzonych przez ciągi o dłuższe, złożone z 6 i 7 powtórzeń CAG lub CCG [Broda M. i wsp. (2005)]. Zatem z określonych struktur transkryptów modelowych oraz ich charakterystyki termodynamicznej wynika, że oddziaływanie powtórzeń CAG z powtórzeniami CCG występującymi w bezpośrednim sąsiedztwie jest bardziej korzystne od sytuacji, w której każdy z traktów tworzy niezależnie struktury. Wysoka stabilność trzonu powstającego w wyniku tego typu interakcji wynika z postulowanego wcześniej stabilizującego wpływu oddziaływania AC.

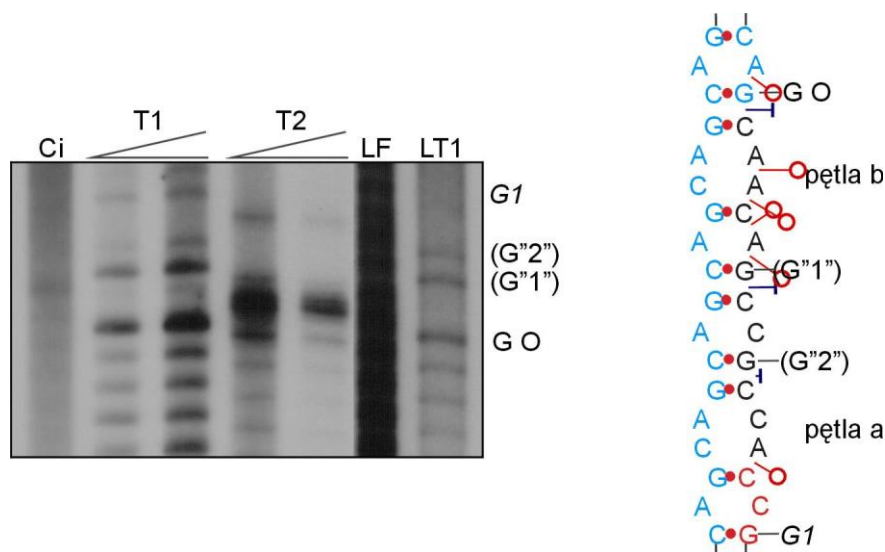
Na podstawie komputerowego przewidywania struktury z wykorzystaniem programu *mfold* określono wpływ trzynastonukleotydowej pętli terminalnej tworzonej przez sekwencję specyficzną łączącą trakty powtórzeń w transkrypcie IT15 10-10. W porównaniu z czteronukleotydową pętlą obserwowaną w modelowej cząsteczce (CAG)₁₀(CCG)₁₀ obniża ona energię swobodną struktury spinki tworzonej przez oddziałujące powtórzenia CAG i CCG o 1,2 kcal/mol.

Analiza struktur modelowych transkryptów i ich właściwości termodynamiczne potwierdzają zatem oddziaływanie powtórzeń CAG i CCG obserwowane w strukturze drugorzędowej fragmentu mRNA genu *IT15* zawierającego dwa ciągi trójnukleotydowe.

4.3 W mRNA genu *IT15* dodatkowe powtórzenia CAG nie oddziałujące z ciągiem CCG wydłużają strukturę spinki.

Stwierdzenie tworzenia przez powtórzenia CAG i CCG jednej struktury typu spinki w modelowych cząsteczkach RNA oraz we fragmencie naturalnego transkryptu IT15 10-10 nie odpowiada jeszcze na pytanie o strukturę regionu zawierającego większą liczbę powtórzeń CAG w transkrypcie. Badania transkryptu IT15 17-10 odpowiadającego jednemu z najczęstszych w populacji allelowi genu *IT15* potwierdziły tworzenie przez region powtórzeń jednej struktury typu spinki (ryc. 3.11). Najistotniejszą jej część stanowi oddziaływanie powtórzeń CAG z traktem

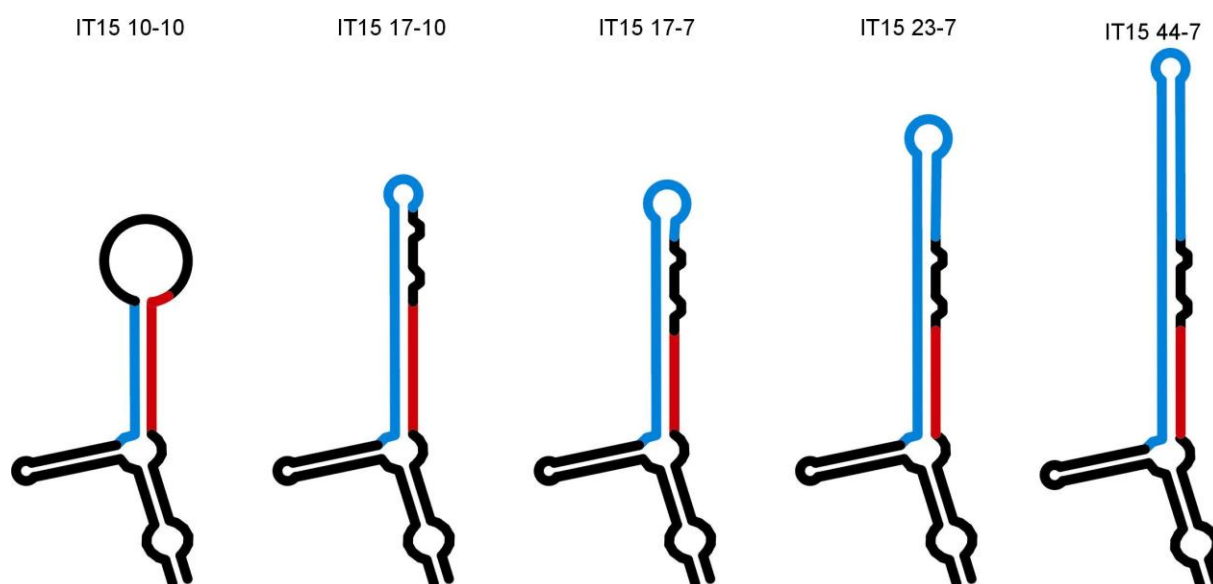
CCG. Powstający w jego wyniku region trzonu silnie stabilizuje pozostałą część struktury a przede wszystkim wymusza wystąpienie oddziaływania kolejnych powtórzeń motywu CAG ze specyficzną sekwencją rozdzielającą. Region rozdzielający ciągi powtórzeń trójnukleotydu w genie *IT15* ma sekwencję CAACAGCCGCCA, która może być uznana za pseudopowtórzenia. Dzięki takiej budowie może on oddziaływać z powtórzeniami CAG, tworząc dwie czteronukleotydowe pętle wewnętrzne (pętla a i b – ryc. 4.3). Pętla a składa się z dwóch niesparowań CA i AC nie mających destabilizującego wpływu na strukturę, co uwidacznia się wyraźnie poprzez brak w tym regionie silnych cięć generowanych przez sondy strukturalne o specyficzności do regionów pojedynczonicowych. Natomiast w przypadku pętli b (CA/AA) widoczna jest większa podatność na działanie nukleaz po stronie 3' (AA) (ryc. 4.3). Zróżnicowanie to jest kolejnym argumentem potwierdzającym korzystny dla stabilności struktury wpływ oddziaływań typu AC. Struktura sekwencji rozdzielającej ciągi powtórzeń CAG i CCG w mRNA genu *IT15* opisana dla transkryptu *IT15* 17-10 jest zachowana w przypadku pozostałych analizowanych transkryptów zawierających więcej powtórzeń CAG niż CCG. Pozostałe dwa „nadmiarowe” powtórzenia CAG transkryptu *IT15* 17-10 nie biorące udziału w oddziaływaniu z traktem CCG ani z regionem rozdzielającym tworzą czteronukleotydową pętlę terminalną (ryc. 3.11 B – pętla c) o strukturze typowej dla powtórzeń CAG. Można było zatem przypuszczać, że w przypadku



Rycina 4.3. Struktura sekwencji rozdzielającej ciągi powtórzeń trójnukleotydu we fragmentach mRNA genu *IT15*. Oznaczenia jak na rycinie 3.7.

większego zróżnicowania liczby powtórzeń motywu CAG i CCG „nadmiarowe” powtórzenia CAG nie zaangażowane w oddziaływanie z innymi sekwencjami utworzą strukturę typu spinki przedłużającą trzon utworzony w wyniku oddziaływania pierwszych powtórzeń CAG z traktem CCG i sekwencją rozdzielającą (ryc. 4.4).

Analiza transkryptu IT15 17-7, odpowiadającego najczęstszemu w polskiej populacji allelowi genu *IT15*, pokazuje słuszność tego przypuszczenia. W strukturze tworzonej przez region powtórzeń tej cząsteczki obecna jest siedmionukleotydowa pętla terminalna (ryc. 3.11 B – pętla c) zbudowana z samych powtórzeń CAG. Jest ona oddzielona od pętli wewnętrznej (ryc. 3.11 B – pętla b) krótkim regionem trzonu złożonego także z samych powtórzeń CAG. Dalszy wzrost długości ciągu powtórzeń CAG, przy zachowanej stałej liczbie 7 powtórzeń CCG, powoduje wydłużanie regionu trzonu pomiędzy pętlą wewnętrzną b a pętlą terminalną – pętlą c. Zatem zwiększenie liczby powtórzeń motywu CAG do 23 widoczne jest w strukturze drugorzędowej RNA jako wydłużenie trzonu złożonego z samych powtórzeń CAG (ryc. 4.4) o 4 regiony podwójnoniciowe par GC CG połączone niesparowaniami AA, przy zachowanej siedmionukleotydowej pętli terminalnej (ryc. 3.12 A). W przypadku parzystej liczby powtórzeń CAG pętla terminalna jest pętlą czteronukleotydową, jak to ma miejsce w transkrypcie IT15 24-7 (ryc. 3.12 B).



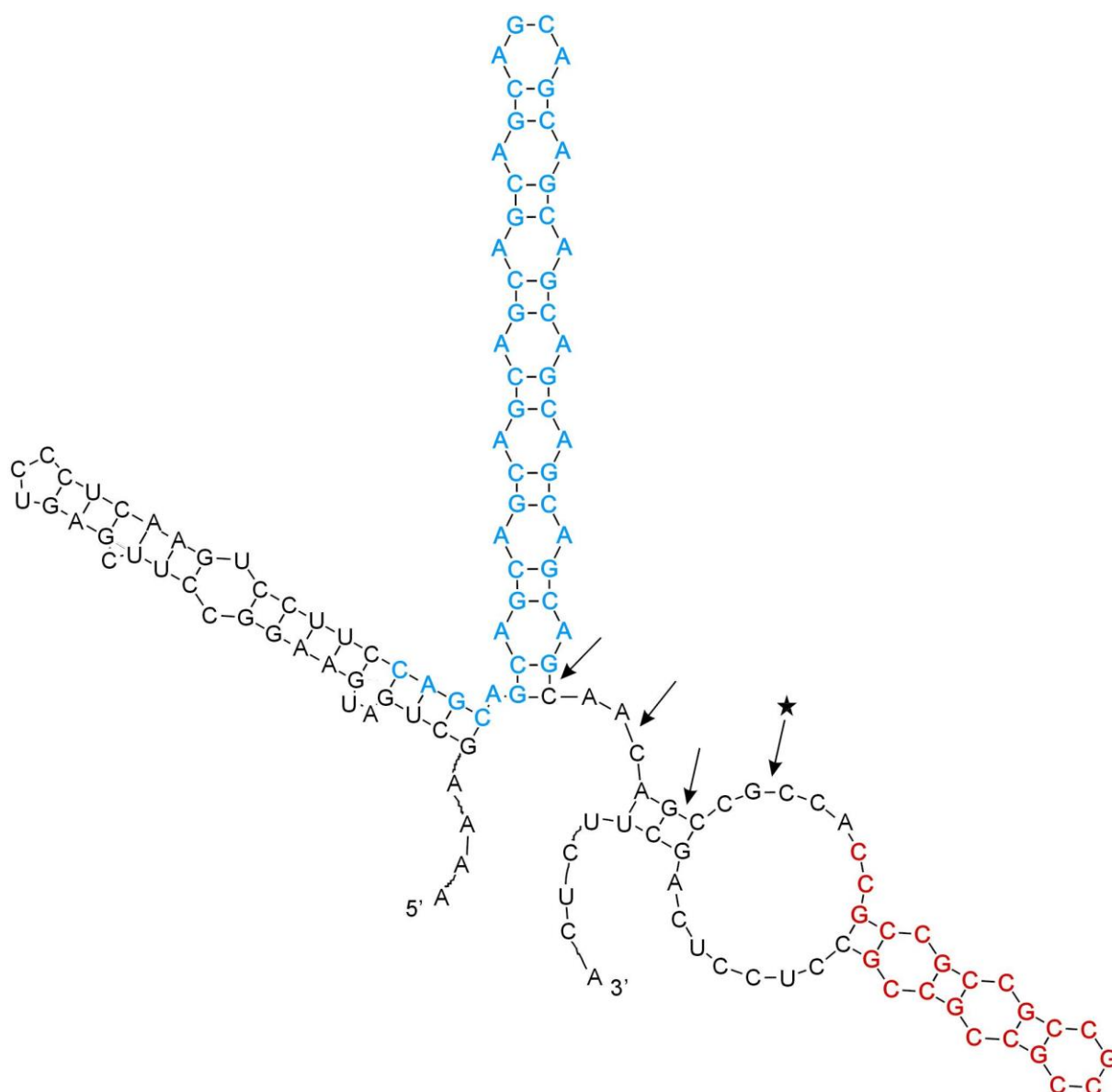
Rycina 4.4. Schemat zmian struktury regionu powtórzeń trójnukleotydowych w mRNA genu *IT15* wynikających ze zmian liczby powtórzeń motywu CAG lub CCG

Badania struktury transkryptu IT15 44-7 będącego przykładem formy zmutowanej, wydłużonej w wyniku ekspansji alleli genu *IT15* pokazały, że nawet przy tak dużej przewadze powtórzeń CAG zachowane są cechy charakterystyczne dla struktury drugorzędowej regionu powtórzeń trójnukleotydowych mRNA genu *IT15* (ryc. 4.4) stwierdzone na przykładzie cząsteczek RNA reprezentujących allele normalne. Podstawa trzonu struktury typu spinki tworzona jest przez pierwsze siedem powtórzeń motywu CAG oddziałujących z ciągiem CCG. Zachowany jest także region zawierający pętle wewnętrzne (ryc. 4.4 oraz ryc. 4.3 i ryc. 3.12 C – pętle a i b) w obrębie oddziaływania kolejnych powtórzeń CAG z sekwencją rozdzielającą. Pozostałe powtórzenia CAG tworzą na istniejącej podstawie strukturę spinki o czteronukleotydowej pętli terminalnej. Zatem, jedyną zmianą w porównaniu z transkrypcją IT15 24-7 jest dalsze wydłużenie trzonu utworzonego z „nadmiarowych” powtórzeń CAG (ryc. 4.4).

4.4 Porównanie struktury fragmentu mRNA genu *IT15* zawierającego powtórzenia trójnukleotydowe z modelem struktury proponowanym wcześniej.

Prezentowany w tej pracy model struktury fragmentu cząsteczki mRNA genu *IT15* zawierającego ciągi powtórzeń CAG i CCG nie jest pierwszym modelem struktury tego mRNA. We wcześniejszej publikacji podjęto próbę określenia struktury regionu powtórzeń w mRNA genu *IT15* [Yen L. i wsp. (1999)]. Podstawowym narzędziem stosowanym w przywoływanej pracy było komputerowe przewidywanie struktury z użyciem programu *mfold* [Yen L. i wsp. (1999)]. Do potwierdzenia lokalizacji pojedynczoniowych regionów cząsteczki mRNA zastosowano również mapowanie struktury z wykorzystaniem cząsteczek DNA o właściwościach katalitycznych [Yen L. i wsp. (1999)]. Użyte w dyskutowanej pracy „nukleazy DNA” lub inaczej DNA–zymy to oligodeoksynukleotydy mające zdolność rozcinania RNA. Zawierają one 15-nukleotydowe centrum katalityczne oraz dwa regiony długości kilkunastu nukleotydów (na końcach 5' i 3') komplementarne do sekwencji substratu RNA. Do swojego działania oligonukleotydy te wymagają możliwości specyficznego związania się z regionami pojedynczoniowymi cząsteczki substratu [Santoro S. W., Joyce G. F. (1998)] a miejsce rozcinania RNA jest wyznaczane przez strukturę obu zhybrydowanych cząsteczek (substratu i DNA–zymu). Sekwencje regionów

otaczających centrum katalityczne „nukleazy DNA” zaprojektowano na podstawie sekwencji pojedynczoniowych fragmentów cząsteczki mRNA genu *IT15* określonych za pomocą komputerowego przewidywania struktury [Yen L. i wsp. (1999)]. Na podstawie lokalizacji miejsc cięcia oraz wyników modelowania komputerowego zaproponowano strukturę cząsteczki z 16 powtórzeniami CAG i 7 powtórzeniami CCG, w której każdy z ciągów powtórzeń tworzy niezależną strukturę typu spinki zaś sekwencja rozdzielająca stanowi pojedynczoniowe połączenie



Rycina. 4.5. Wcześniej proponowany model struktury fragmentu mRNA *IT15* $(CAG)_{16}(CCG)_7$ według Yen i wsp. (1999). Strzałkami oznaczono miejsca cięć indukowanych „nukleazami DNA” – gwiazdka oznacza najsilniejsze cięcie.

między nimi [Yen L. i wsp. (1999)] (ryc. 4.5). Jak opisano wcześniej, wyniki biochemicznego sondowania struktury i analiz termodynamicznych różnych transkryptów reprezentujących zakres polimorficznej zmienności traktów trójnukleotydowych (w tym także transkryptu zawierającego (CAG)¹⁶ i (CCG)⁷ – dane nie prezentowane) jednoznacznie wskazują na oddziaływanie ze sobą ciągów powtórzeń trójnukleotydowych w mRNA genu *IT15*, zaprzeczają więc modelowi zaproponowanemu w pracy Yen L. i wsp. (1999). W określonej w tej pracy doktorskiej strukturze drugorzędowej regionu powtórzeń mikrosatelitarnych w mRNA genu *IT15* miejsca przecinania łańcucha RNA przez DNA–zymy opisane przez Yen L. i wsp. (1999) zlokalizowane są w obrębie regionu pętli wewnętrznych i w bezpośrednim sąsiedztwie pętli terminalnej, czym można wyjaśnić ich dostępność DNA–zemu.

4.5 Struktura regionu powtórzeń mRNA genu *IT15* w odniesieniu do struktury regionów powtórzeń trójnukleotydowych innych RNA.

Sekwencja powtórzeń CAG w genie *IT15* tym się różni od innych ulegających ekspansji ciągów mikrosatelitarnych związanych z chorobami z grupy TREDs, że jest zlokalizowana w bezpośrednim sąsiedztwie innego ciągu trójnukleotydowego, z którym, jak wykazano w niniejszej pracy, oddziałuje w strukturze transkryptu. Jak wynika z badań mRNA genu *AR*, w którym występują również dwa blisko leżące ciągi trójnukleotydowe, w tym przypadku ciągi te nie oddziałują ze sobą. W zbadanych dotąd strukturach transkryptów genów z grupy TREDs zawierających ciągi powtórzeń (CNG)_n stwierdzono występowanie struktur typu spinki, które można podzielić na dwa rodzaje. Pierwszy z nich to, tworzona przez sekwencję powtarzającą się nie zawierającą zaburzeń, „poślizgnięta” struktura spinki charakteryzująca się możliwością współwystępowania kilku form strukturalnych różniących się między sobą lokalizacją pętli terminalnej. Może być ona przesunięta o jedno do trzech powtórzeń w porównaniu z jej lokalizacją w dominującej formie struktury. Tego rodzaju heterogenność struktury występuje kiedy sekwencje otaczające region powtórzeń nie oddziałują ze sobą lub ich oddziaływanie ma miejsce na niewielkim odcinku i nie zapewnia stabilności wystarczającej do „zamrożenia” jednej formy przestrzennej [Napierała M., Krzyżosiak W. J. (1997), Sobczak K., Krzyżosiak W. J. (2003), Michlewski G., Krzyżosiak W. J. (2004)]. Jak wykazano na podstawie

przedstawionej w tej pracy struktury regionu powtórzeń w transkrypcie genu *AR*, w przypadku kiedy sekwencja otaczająca tworzy silnie sparowany region u podstawy struktury spinki regionu powtórzeń nie ma możliwości alternatywnych ułożeń regionu powtórzeń. Drugim rodzajem struktury przyjmowanej przez regiony zawierające sekwencje powtórzeń są rozgałęzione struktury typu trzon i pętla. Powstają one kiedy w ciągu powtórzeń trójnukleotydomowych występują zaburzenia w postaci innych rodzajów tripletów. Zaburzenia lokalizują się w strukturze RNA w regionach pojedynczonicowych pętli terminalnych, pętli wewnętrznych lub miejscach rozgałęzień trzonu. Ich obecność przeciwdziała powstawaniu pojedynczej spinki a sprzyja wytworzeniu się struktur rozgałęzionych [Sobczak K., Krzyżosiak W. J. (2004), Sobczak K., Krzyżosiak W. J. (2005), Napierała M. i wsp. (2005)].

W świetle tych informacji struktura przyjmowana przez region powtórzeń CAG i CCG w mRNA genu *IT15* stanowi nowy rodzaj struktury tworzonej przez sekwencje powtarzające się w mRNA genów związanych z TREDs. Wyjątkowe dla niej oddziaływanie dwóch typów powtórzeń, które ma silnie stabilizujący wpływ na strukturę całego regionu utrudnia przyjmowanie alternatywnych ułożeń przez fragment spinki tworzony z samych powtórzeń CAG. Takie alternatywne ułożenie struktury jest jednak możliwe, na co wskazują dodatkowe cięcia w regionie pętli terminalnej obserwowane w wynikach analizy transkryptów *IT15* 23-7, *IT15* 24-7, a przede wszystkim transkryptu *IT15* 44-7 (ryc. 3.12). Ich wystąpieniu sprzyja charakter sekwencji rozdzielającej ciągu CAG i CCG, która może być uznana za pseudopowtórzenia. Powstające alternatywne struktury „nadmiarowych” powtórzeń CAG nie powodują jednak istotnych zmian architektury regionu powtórzeń w transkrypcie genu *IT15*.

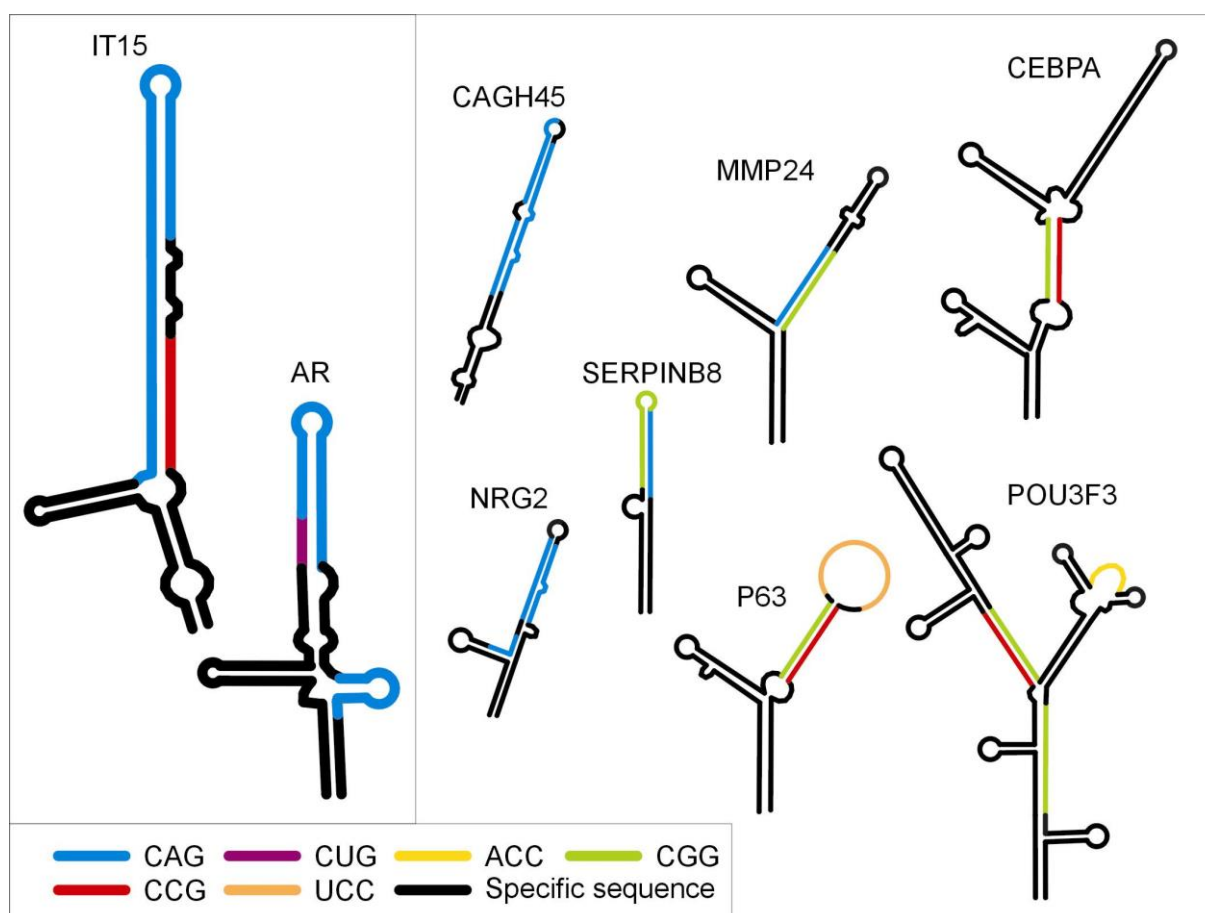
W celu sprawdzenia czy oddziaływanie występujących w niewielkiej odległości od siebie ciągów powtórzeń trójnukleotydomowych stwierdzone w mRNA genu *IT15* ma również miejsce w transkrypcie innego genu związanego z TREDs określono strukturę regionu zawierającego dwa ciągi CAG w mRNA genu *AR*. Przeprowadzono badania struktury transkryptu zawierającego 9 powtórzeń w polimorficznym ciągu CAG w obecności specyficznej sekwencji otaczającej, która w odległości 18 nukleotydów poniżej ciągu polimorficznego zawierała niepolimorficzny ciąg (CAG)₆. Stwierdzono, że tworzą one niezależne motywy strukturalne o architekturze spinki (ryc. 3.14). Wzrost liczby powtórzeń CAG w ciągu polimorficznym zarówno w zakresie normalnej zmienności obserwowanej w populacji jak i do rozmiarów mutacji

powodującej wystąpienie objawów rdzeniowo-opuszkowego zaniku mięśni nie powoduje innych zmian struktury niż wydłużanie trzonu spinki (ryc. 3.15). Ciekawą obserwacją jest fakt, że polimorficzny ciąg powtórzeń CAG w mRNA genu *AR* wykazuje, w porównaniu do ciągu CAG badanego we fragmentach mRNA genu *IT15*, większą trwałość chemiczną widoczną jako brak „samodegradacji” wiązań CpA tak charakterystycznej dla niektórych analizowanych wcześniej sekwencji RNA zawierających powtórzenia CAG [Sobczak K. i wsp. (2003), Michlewski G., Krzyżosiak W. J. (2004), de Mezer M dane nie publikowane]. Brak odpowiednio zaplanowanych badań porównawczych umożliwiających wykluczenie innych czynników niż struktura i jej stabilność uniemożliwia określenie szczegółowych przyczyn tego zróżnicowania. Wysoka stabilność struktury spinki tworzonej przez polimorficzny ciąg powtórzeń CAG w mRNA genu *AR* wynika z obecności u jej podstawy regionu podwójnonicowego o długości 12 par zasad tworzonego przez sekwencję specyficzną 5' zawierającą trzy triplety CUG i ostatnie powtórzenia CAG oraz inne nukleotydy sekwencji otaczającej. Badania transkryptów CAG17_cl (rozdział 3.1) wykazały, że już sześć par zasad u podstawy trzonu powoduje stabilizację spinki i „zamrożenie” jednego z jej wariantów. Wydaje się zatem, że to obecność struktury podwójnonicowej stabilizowanej wieloma kanonicznymi parami zasad stabilizuje strukturę polimorficznego ciągu CAG w mRNA genu *AR* i decyduje o braku oddziaływania pomiędzy traktami powtórzeń trójnukleotydowych w tym regionie.

Występowanie w jednym mRNA kilku traktów trójnukleotydowych jest zjawiskiem dość częstym w ludzkim transkryptomie. Spośród ponad 600 mRNA zawierających ten typ sekwencji powtarzającej się [Jasinska A. i wsp. (2003)] aż w 10% (62 mRNA) przypadków stwierdzono obecność więcej niż jednego ciągu. Ponad połowa z nich to mRNA, w których sekwencje powtarzające się obecne są w dwóch miejscach cząsteczki, zaś w blisko jednej czwartej mRNA aż w trzech miejscach. mRNA zawierające większą liczbę ciągów są nieliczne. Tylko w jednym przypadku powtórzenia trójnukleotydowe występują aż w sześciu miejscach sekwencji tej samej cząsteczki dojrzałego mRNA. Podobnie jak w całej puli ludzkich mRNA, także wśród sekwencji powtarzających się występujących w większej liczbie w pojedynczym transkrypcie dominują powtórzenia motywu CAG, które są obecne w połowie tych mRNA i tworzą 25% wszystkich zawartych w nich ciągów. Drugim co do częstości występowania powtarzającym się motywem jest trójka CGG, co również jest zgodne

z tendencją obserwowaną w całym transkryptomie. Zatem, nie ma znaczącej preferencji występowania określonego rodzaju powtórzeń w mRNA zawierających wiele sekwencji powtórzonych. Poza obserwacją, że blisko połowa mRNA tej grupy zawiera ciągi złożone z tego samego rodzaju motywu, nie stwierdzono innych zależności we współwystępowaniu poszczególnych rodzajów powtórzeń.

W kontekście badań przedstawionych w tej pracy sprawą interesującą było jaką strukturę drugorzędową mogą przyjąć regiony mRNA zawierające kilka ciągów trójnukleotydowych. W celu uzyskania wstępnej odpowiedzi na to pytanie przeprowadzono modelowanie struktury cząsteczek RNA przy użyciu programu *mfold 3,1* [Zuker M. (2003)]. W wyniku modelowania fragmentów mRNA 64 genów zawierających ciągi motywów trójnukleotydowych oraz otaczające sekwencje specyficzne o długości 100 nukleotydów stwierdzono, że zaledwie w 14 przypadkach w strukturze drugorzędowej program przewiduje możliwość oddziaływania pomiędzy traktami powtórzeń.



Rycina 4.6. Modelowania struktury zawierających kilka traktów powtórzeń trójnukleotydowych fragmentów ludzkich mRNA.

Spośród 20 różnych typów powtórzeń trójnukleotydowych stwierdzono, że powtórzenia CAG, CUG, CCG, CGG, CGA oraz CGU są zdolne do tworzenia trwałych struktur typu spinki [dane nie opublikowane]. Mniej trwałe struktury mogą być utworzone przez powtórzenia motywów AUG, AUC, AUU, AAU, AGU i ACU. Komputerowe przewidywanie struktury mRNA zawierających kilka ciągów powtórzeń wykazało, że do oddziaływania może dochodzić tylko pomiędzy ciągami, które występując samodzielnie mają możliwość utworzenia struktury typu spinki. We wszystkich przypadkach przynajmniej jeden z uczestników oddziaływania należał do powtórzeń typu (CNG)_n (ryc. 4.6). Najbardziej stabilne regiony sparowane tworzą w pełni komplementarne ciągi powtórzeń CCG i CGG (ryc. 4.6). Oddziaływania takie wydają się mieć istotne znaczenie dla możliwości rearanżacji struktury w procesie dojrzewania mRNA, ponieważ program przewiduje ich oddziaływanie nawet wtedy kiedy występują w znacznej odległości od siebie jak ma to miejsce w transkrypcie genu *CEBPA* (ryc. 4.6), w którym trakty te są oddalone od siebie o 230 nukleotydów. Mogą też zapewniać stabilizację pewnych regionów pojedynczoniciowych jak ma to miejsce w transkrypcie genu *P63* (ryc. 4.6) i w ten sposób przypuszczalnie wpływać na ich potencjalną rolę biologiczną.

Mniej stabilne i występujące w wielu formach są struktury utworzone z kilku ciągów CAG, co zaobserwowano na przykładzie transkryptu genu *CAGH45* (ryc. 4.6). W żadnym z analizowanych fragmentów mRNA różnych genów nie zaobserwowano opisanego w tej pracy oddziaływania ciągu CAG i CCG. Sytuacja ta może wynikać z faktu, iż w żadnym z analizowanych transkryptów trakty te nie występowały w takiej bliskości jak w transkrypcie genu *IT15*. Tak niewielkie oddalenie od siebie ciągów powtórzeń o różnym typie powtarzających się motywów trójnukleotydowych jak to, które występuje w genie *IT15* obserwowano tylko dla powtórzeń typu CAG i CGG w mRNA genów *MMP24* i *SERPIN8*, które według przewidywań programu *mfold* oddziałują ze sobą w tych transkryptach (ryc. 4.6).

Przeprowadzone modelowanie struktur drugorzędowych cząsteczek RNA zawierających kilka ciągów powtórzeń trójnukleotydowych sugerowały, że aby doszło do oddziaływania regionów powtórzeń muszą zostać spełnione pewne warunki. Po pierwsze, sekwencje powtórzone muszą się składać z motywów, które umożliwiają im utworzenie oddziaływania. Drugim warunkiem jest odpowiednia sekwencja specyficzna, której struktura umożliwia kontakt pomiędzy powtórzeniami. Nie

powinny one także zawierać regionów o większym powinowactwie do któregoś z ciągów zdolnych do wzajemnych oddziaływań.

4.6 Biologiczne znaczenie struktury regionu powtórzeń trójnukleotydowych w mRNA genu *IT15*.

Zarówno wyniki eksperymentalne, jak i dane bioinformatyczne przedstawione w tej pracy wskazują na wyjątkowość struktury tworzonej przez region powtórzeń trójnukleotydowych mRNA genu *IT15* związanego z chorobą Huntingtona. W świetle dostępnych informacji jest ona jedyną występującą w komórce strukturą typu spinki utworzoną częściowo z oddziałujących ze sobą motywów CAG i CCG, a w części z samych powtórzeń CAG. Można zadać pytanie czy tak specyficzny motyw strukturalny może mieć znaczenie dla przebiegu któregoś z procesów komórkowych.

Nie poznano dotąd funkcji jaką pełni w komórce region powtórzeń mRNA *IT15*. Odkrycie białka o masie około 63 kD wiążącego się do ciągów CAG występujących w RNA stało się powodem spekulacji na ten temat [McLaughlin B. A. i wsp. (1996)]. Jego współwystępowanie z mRNA genu *IT15* w tkankach ulegających degeneracji w procesie patogenezy choroby Huntingtona pozwala sądzić, że wiązanie białka 63-kD do mRNA genu *IT15* ma znaczenie dla indukcji lub przyspieszenia procesu apoptozy komórek. Poza stwierdzeniem wzrostu ilości białka związanego z RNA, zależnego od długości ciągu powtórzeń CAG, nie dokonano innych obserwacji pozwalających określić jakie znaczenie dla tych interakcji ma struktura drugorzędowa regionu powtórzeń w RNA. Białkiem o wielkości zbliżonej do 63-kD jest kinaza białkowa zależna od dwuniciowego RNA – PKR (69 kDa), której aktywację zaobserwowano w wyniku wiązania do mRNA genu *IT15* [Peel A. L. i wsp. (2001)]. Kinaza PKR jest jednym z elementów inicjacji odpowiedzi interferonowej komórki na infekcję wirusową. Po związaniu dsRNA ulega autofosforylacji i stymuluje aktywację licznych szlaków komórkowej odpowiedzi na stres wśród których głównym jest apoptoza. Przeprowadzone badania aktywności tego enzymu w mózgach osób chorych na chorobę Huntingtona wykazały jego podwyższoną aktywność w regionach, które ulegają zmianom w procesie patogenezy [Peel A. L. i wsp. (2001)]. Obserwacja braku korelacji pomiędzy ilością zmutowanego białka htt a poziomem apoptozy komórek przy jednocześnie podwyższonej aktywności PKR pozwala sądzić, że za niektóre aspekty patogenezy mogą być odpowiedzialne oddziaływania

RNA regionu powtórzeń trójnukleotydowych z białkami komórkowymi, a szczególnie z PKR. Informacji o charakterze oddziaływań tego białka z powtórzeniami RNA dostarczyły badania wiązania czynników białkowych do ciągów powtórzeń CUG ulegających ekspansji w 3' UTR genu *DMPK* [Tian B. i wsp. (2000)]. Minimalna liczba powtórzeń motywu CUG, które przyjmują strukturę spinki [Napierała M., Krzyżosiak W. J. (1997)], konieczna do aktywacji kinazy zależnej od dsRNA to 15 CUG, a wydłużanie ciągu powodowało wzrost aktywności enzymu [Tian B. i wsp. (2000)]. Znaczenie struktury utworzonej przez powtórzenia CUG w patogenezie dystrofii miotonicznej polega na nadmiernym lub niespecyficznym wiązaniu białek prowadzącym do zaburzenia ich komórkowych funkcji regulacji takich procesów jak transkrypcja czy dojrzewanie mRNA, a w konsekwencji do zachwiania homeostazy wewnątrzkomórkowej [Timchenko N. A. i wsp. (2001), Charlet-B. N. i wsp. (2002)]. Dla powtórzeń CAG, a co bardziej istotne dla mRNA genu *IT15*, nie określono dotąd minimalnej liczby powtórzeń koniecznej do aktywacji PKR. Warte podkreślenia jest również to, że nie stwierdzono interakcji kinazy PKR z innymi transkryptami z grupy TREDs zawierającymi powtórzenia CAG. Także dla mRNA genu *FMR1*, w którym ulegający ekspansji trakt CGG zlokalizowany jest w regionie 5' UTR wykazano, że nie dochodzi do aktywacji PKR [Handa V. i wsp. (2003)] pomimo przyjmowania przez powtórzenia struktury spinki [Napierała M. i wsp. (2005)].

Występowanie w chorobie Huntingtona objawów, których nie można wyjaśnić tylko zaburzeniem prawidłowej funkcji białka sugeruje odgrywanie istotnej roli w rozwoju choroby przez inny czynnik. Podobieństwo struktury regionu powtórzeń jak i wiązanie do niego tego samego czynnika białkowego – PKR, obserwowane w przypadku DM1 i HD pozwala na wysunięcie hipotezy, że struktury spinki utworzone przez region powtórzeń CUG w mRNA genu *DMPK* oraz oddziałujące ze sobą trakty CAG i CCG w mRNA genu *IT15* spełniają warunki konieczne do aktywacji szlaku apoptozy komórek. Dotychczasowy brak informacji na temat aktywacji PKR przez struktury spinki regionów powtórzeń trójnukleotydowych w mRNA genów grupy TREDs z wyjątkiem transkryptów *DMPK* i przypuszczalnie *IT15*, pozwala przypuszczać, że właściwości struktury regionu powtórzeń CAG i CCG mogą mieć podobne znaczenie dla wiązania RNA – białko jak struktura spinki powtórzeń CUG. Prawdopodobnie z powtórzeniami CAG mogą oddziaływać inne białka niż z powtórzeniami CUG. Zaobserwowano, że wiążące się do długich ciągów CUG białko MBNL (*muscleblind*) nie wiążą się do ciągów CAG [Kino Y. i wsp. (2004)].

Stwierdzono także, że nie ulegające translacji długie ciągi powtórzeń CAG mają właściwości toksyczne i wywołują w organizmie myszy zmiany patologiczne zgodnie z modelem zmiany funkcji RNA [Pan H. i wsp. Poster AJHG (2005)]. Obserwacja ta sugeruje, że także w organizmie człowieka ciągi powtórzeń CAG mogą odgrywać istotną rolę w patogenezie chorób neurologicznych.

Uzyskane dotąd informacje o strukturze i roli regionu powtórzeń CAG i CCG w mRNA genu *IT15* wskazują na możliwy udział RNA w procesach patogenezy według uogólnionego dla większości TREDs mechanizmu zakładającego toksyczność struktury RNA powtórzeń trójnukleotydomych. Jednakże do precyzyjnego określenia typu zaburzeń wynikających z wydłużenia trzonu struktury spinki spowodowanego ekspansją powtórzeń CAG konieczne jest zidentyfikowanie czynników białkowych mających możliwość wiązania się do niej. Poznanie białek oddziałujących ze strukturą powtórzeń i ich funkcji komórkowej pozwoli na określenie skutków ich konfiskaty z właściwych, fizjologicznych miejsc wiązania. Wiedza ta może przyczynić się do opracowania metod leczenia, na przykład poprzez uzupełnianie ilości czynników białkowych związanych przez zmutowany RNA. Określenie struktury regionu powtórzeń trójnukleotydomych mRNA *IT15* stanowi lepszą podstawę do dalszych poszukiwań takich czynników biorących przypuszczalnie udział w patogenezie choroby Huntingtona.

5 Podsumowanie

W niniejszej pracy doktorskiej przeprowadzono analizę występowania w ludzkim transkryptomie powtórzeń trójnukleotydowych. Eksperymentalnie zbadana została struktura RNA powtórzeń CAG i CCG występujących zarówno niezależnie jak i łącznie, najpierw pozbawionych sekwencji otaczających, a następnie we fragmentach mRNA genów *IT15* i *AR*. Ponadto, zanalizowano *in silico* strukturę fragmentów mRNA zawierających kilka, oddzielonych od siebie ciągów powtórzeń trójnukleotydowych.

1. Genomową bazę sekwencji nukleotydowych GenBank przeszukiwano pod kątem występowania transkryptów zawierających ciągi powtórzeń 20 różnych motywów trójnukleotydowych składające się z co najmniej 6 powtórzeń danego motywu. Ustalono, że:
 - 718 ciągów trójnukleotydowych powtórzeń występuje w 619 różnych mRNA;
 - najczęściej występującymi motywami powtórzonymi okazały się: CAG, CCG, CUG, AGG i ACC stanowiące łącznie 83%, natomiast ACG, AUC, CUU, AGU, CGU i ACU należą do najrzadszych w transkryptomie;
 - większość ciągów powtórzeń występuje w otwartej ramce odczytu – ORF (67%), mniej w regionie 5' nie ulegającym translacji – 5'UTR (24%) i najmniej w 3'UTR (9%);
 - powtórzenia bogate w GC przeważają w 5'UTR, a bogate w AU są częstsze w 3'UTR;
 - biorąc pod uwagę różne długości obydwu regionów mRNA nie ulegających translacji częstość powtórzeń jest znacznie wyższa w sekwencjach 5'UTR, co może świadczyć o ich roli w regulacji translacji, natomiast trakty występujące w 3'UTR mogą odgrywać rolę w regulacji stabilności i lokalizacji komórkowej transkryptów;
 - większość spośród zidentyfikowanych transkryptów (86%) zawiera krótkie ciągi powtórzeń składające się z 6 – 10 powtórzonych jednostek, transkrypty zawierające 11 – 20 powtórzonych jednostek stanowią 11%, a te w których występują ciągi zawierające ponad 20 powtórzeń tego samego motywu sekwencji – tylko 2%;

- porównanie znanych funkcji genów, których mRNA zawierają ciągi trójnukleotydydowe w stosunku do funkcji wszystkich poznanych genów wykazuje większy udział genów związanych z regulacją procesów sygnalizacji wewnątrz- i międzykomórkowej (z 28 do 39%) oraz transkrypcją i translacją (z 24 do 34%), natomiast mniejszy udział genów regulujących metabolizm podstawowy (z 19 do 9%).
2. Analiza eksperymentalna struktury RNA powtórzeń CAG i CCG pozbawionych naturalnego kontekstu sekwencji wykazała, że:
- przyjmują one regularną strukturę spinki, której trzon jest zbudowany z krótkich regionów sparowanych GC CG oddzielonych od siebie pojedynczymi niesparowaniami odpowiednio AA i CC;
 - jeśli w transkrypcie nie występują oddziałujące ze sobą sekwencje oskrzydlające powtórzenia wtedy sekwencje (CAG)_n i (CCG)_n wykazują tendencję do przyjmowania kilku wariantowych ułożeń dążąc do utworzenia bardziej korzystnej termodynamicznie czteronukleotydydowej pętli terminalnej;
 - występowanie u podstawy trzonu dodatkowego regionu sparowanego utworzonego przez oddziałujące ze sobą sekwencje oskrzydlające zwiększa stabilność struktury spinki tworzonej przez powtórzenia motywów CAG i CCG i określa wielkość pętli terminalnej. Cząsteczki z parzystą liczbą powtórzeń charakteryzują się pętlą czteronukleotydydową, natomiast z nieparzystą liczbą powtórzeń tworzą pętlę siedmionukleotydydową.
3. Na podstawie wyników badań struktury fragmentów mRNA genu *IT15* związanego z chorobą Huntingtona stwierdzono, że:
- w obecności naturalnej sekwencji otaczającej nie obejmującej ciągu CCG powtórzenia CAG w liczbie 10 i 17 tworzą strukturę typu spinki o budowie trzonu i pętli terminalnej zgodnej z opisaną wcześniej dla cząsteczek modelowych;

- ciągi CAG i CCG zawierające taką samą liczbę powtórzeń oddziałują ze sobą tworząc trzon struktury spinki, której pętlę terminalną stanowi odcinek sekwencji specyficznej rozdzielającej powtórzenia;
 - większa stabilność trzonu spinki utworzonej przez powtórzenia CAG i CCG wynika z możliwości wystąpienia niekanonicznych oddziaływań AC pomiędzy krótkimi regionami sparowanymi GC CG;
 - zmiana liczby powtórzeń CCG w ciągu powoduje zmianę długości trzonu struktury zbudowanego z obu rodzajów powtórzeń;
 - „nadmiarowe” powtórzenia CAG nie oddziałują z krótszym ciągiem powtórzeń CCG w mRNA genu *IT15* przyjmują strukturę spinki o architekturze typowej dla tego typu powtórzeń, której trzon stanowi przedłużenie struktury regionu oddziaływań CAG – CCG;
 - struktura utworzona przez same powtórzenia CAG jest stabilizowana przez trzon CAG – CCG;
 - w przypadku mutacyjnie wydłużonego ciągu CAG w strukturze dochodzi do wydłużenia trzonu tworzonego przez same powtórzenia CAG bez zmiany oddziaływania ich części z powtórzeniami CCG i sekwencją rozdzielającą;
 - specyficzne sekwencje otaczające przyjmują strukturę niezależną od zmian liczby powtórzeń trójnukleotydowych;
 - można przypuszczać, że znaczne wydłużenie trzonu struktury spinki w wyniku ekspansji powtórzeń CAG oraz jej stabilizacja przez oddziaływanie z występującym w bezpośrednim sąsiedztwie traktem CCG może być toksyczne dla komórki, gdyż może powodować nadmierne wiązanie przez taką strukturę RNA niezidentyfikowanych dotąd czynników białkowych.
4. Badania struktury regionu polimorficznego ciągu powtórzeń CAG i sąsiadującego z nim krótkiego szeregu (CAG)₆ występujących w mRNA genu *AR* związanego z rdzeniowo – opuszkowym zanikiem mięśni wykazały, że:
- dwa ciągi powtórzeń CAG w mRNA genu *AR* nie oddziałują ze sobą;

- polimorficzny ciąg CAG genu *AR* przyjmuje w mRNA strukturę spinki typową dla tego typu sekwencji. Jest ona stabilizowana przez znajdujący się u jej podstawy region podwójnoniciowy tworzony przez fragment sekwencji specyficznej 5' składający się z trzech powtórzeń CUG komplementarnych do ostatnich powtórzeń ciągu CAG;
 - zmiany liczby powtórzeń motywu CAG w mRNA genu *AR* powodują jedynie zmiany długości trzonu tworzonej przez nie struktury spinki;
 - można przypuszczać, że znaczne wydłużenie trzonu struktury spinki w wyniku ekspansji powtórzeń CAG może powodować jej udział w procesach patogenezy analogicznych do postulowanych dla transkryptu IT15.
5. Na podstawie analiz bioinformatycznych obejmujących także przewidywania *in silico* struktury fragmentów transkryptów, w których zaobserwowano obecność kilku ciągów powtórzeń trójnukleotydowych stwierdzono, że:
- występowanie w sekwencji mRNA sąsiadujących ze sobą ciągów trójnukleotydowych jest zjawiskiem dość powszechnym w transkrypcie człowieka;
 - blisko połowa mRNA tej grupy zawiera ciągi powtórzeń złożone z tego samego rodzaju powtarzającego się motywu;
 - częstość występowania poszczególnych typów powtórzeń wśród mRNA zawierających kilka ciągów powtórzeń nie odbiega od obserwowanej w całym transkrypcie;
 - możliwość oddziaływania ze sobą regionów powtórzeń trójnukleotydowych zależy od sekwencji powtarzających się motywów oraz od właściwości ich sekwencji otaczających i struktury przez nie przyjmowanej.

6 MATERIAŁY

6.1 Materiał biologiczny

Materiałem wyjściowym do badań był DNA izolowany z leukocytów krwi obwodowej osób zdrowych pochodzący z banku próbek genomowego kwasu deoksyrybonukleinowego, którym dysponuje Pracownia Genetyki Nowotworów IChB PAN w Poznaniu. Preparaty DNA osób chorych na chorobę Huntingtona i rdzeniowo – opuszkowym zanik mięśni uzyskano dzięki współpracy z Kliniką Neurologii Akademii Medycznej w Poznaniu. Wybrane do badań preparaty pochodziły od osób, które wyraziły świadomą zgodę na wykorzystanie ich DNA do badań.

6.2 Ważniejsze odczynniki chemiczne i biochemiczne.

- Agaroza (Sigma)
- Agaroza Metaphore (Serva)
- Akryloamid (Serva)
- Bisakryloamid (Serva)
- Bromek etydyny (Serva)
- Chlorek magnezu (Serva)
- Chlorek potasu (Serva)
- Chlorek sodu (Serva)
- Chloroform cz. d. a. (POCH Gliwice)
- DMSO (Serva)
- dNTP (Promega)
- DTT (Epicentre Technologies)
- EDTA (Serva)
- Etanol (Polmos Poznań)
- Fenol (Aldrich - Chemie)
- Formamid (Fluka)
- Glicerol (Merck)
- Guanozyna (Pharma Waldhof)
- Izopropanol (POCH Gliwice)

- Kwas borowy cz. d. a. (POCh Gliwice)
- Kwas solny (Fluka)
- Kwas octowy lodowaty (Aldrich - Chemie)
- Mocznik cz. d. a. (POCh Gliwice)
- Nadsiarczan amonu - APS (Serva)
- NTP (Promega)
- Octan potasu (Serva)
- Octan sodu (Serva)
- Octan ołowiu (Serva)
- Sacharoza (Serva)
- SDS (Serva)
- Siarczan cynku (Serva)
- Stains – all (Serva)
- TEMED (Serva)
- Tris (Sigma)
- Wodorotlenek sodu (POCh Gliwice)

6.3 Enzymy

Kinaza polinukleotydomowa faga T4 10000 U/ml (New England Biolabs)

Nukleaza S1 150000 U/ml (Amersham)

Polimeraza RNA faga T7 200 U/ μ l (Epicentre Technologies)

Polimeraza RNA faga T7 200 U/ μ l (Ambion)

Polimeraza Taq DNA 5 U/ μ l (Promega)

Rybonukleaza T1 10000 U/ml (Pharmacia)

Rybonukleaza T2 20 U/ μ l (Invitrogen)

Rybonukleaza V1 0,1 U/ μ l (Ambion)

6.4 Bufory.

6.4.1 TBE 10 x stężony pH 8.3.

890 mM Tris – HCl
890 mM kwas borowy
20 mM EDTA

6.4.2 TB 10 x stężony pH 8.3.

890 mM Tris – HCl
890 mM kwas borowy

6.4.3 Bufor do rybonukleazy T1 2 x stężony.

20 mM cytrynian sodu pH 4.5
1mM EDTA
7M mocznik

6.4.4 Bufor Mg 10 do denaturacji / renaturacji RNA 10 x stężony.

100 mM Tris – HCl pH 7.2
400 mM NaCl
100 mM MgCl₂

6.4.5 Bufor Mg 1 do denaturacji / renaturacji RNA 10 x stężony.

100 mM Tris – HCl pH 7.2
400 mM NaCl
10 mM MgCl₂

6.4.6 Bufor Mg 1/ K do denaturacji / renaturacji RNA 10 x stężony.

100 mM Tris – HCl pH 7.2

1350 mM KCl

150 mM NaCl

10 mM MgCl₂

6.4.7 Bufor do analiz właściwości termodynamicznych 10 x stężony.

100 mM Tris – HCl pH 7.2

400 mM NaCl

6.4.8 Bufor do ekstrakcji DNA z agarozy Metaphore.

20 mM Tris – HCl pH 8.0

1 mM EDTA pH 8.0

6.4.9 Bufory dostępne komercyjnie.

Bufor do polimerazy DNA Taq 10 x stężony (Promega).

Bufor do transkrypcji *in vitro* 10 x stężony (Epicentre Technologies).

Bufor do kinazy polinukleotydowej faga T4 10 x stężony (New England BioLabs).

6.5 Roztwory.

6.5.1 Elu RNA do elucji RNA z żelu poliakryloamidowego.

0,3 M octan sodu

0,5 M EDTA

0,1% (w/o) SDS

6.5.2 Octan sodu.

3 M CH₃COONa pH 5.2

6.5.3 Roztwór formamidowy do hydrolizy RNA.

0,5 mM MgCl₂ w 100% formamidzie

6.5.4 "Stains – all" (Serva) - roztwór do barwienia RNA / DNA w żelach akryloamidowych.

a) roztwór podstawowy (0,1%) uzyskany przez rozpuszczenie 100 mg barwnika w 100 ml formamidu

b) roztwór roboczy (0,005%)

10 ml roztworu podstawowego

10 ml formamidu

50 ml izopropanolu

1 ml 3 M Tris – HCl pH 8.8

uzupełniony do 200 ml H₂O

6.5.5 Octan ołowiu do hydrolizy RNA.

20 mM Pb(CH₃COO)₂ (Serva)

6.5.6 Roztwór obciążający do nakładania próbek na żele agarozowe 6x stężony.

30% (w/o) glicerol

10 mM EDTA

6.5.7 Roztwór obciążający do nakładania próbek na niedenaturujące żele poliakryloamidowe 5 x stężony.

14 % (w/o) sacharoza

10 mM Tris – HCl pH 8.0

0,4 % (w/o) błękit bromofenolowy (BB)

0,4 % (w/o) fiolet ksylenowy (XC)

6.5.8 Roztwór UED do nakładania próbek na denaturujące żele poliakryloamidowe 2 x stężony.

- 7.5 M mocznik
- 5 mM EDTA
- 0,04 % (w/o) błękit bromofenolowy (BB)
- 0,04 % (w/o) fiolet ksylenowy (XC)

6.5.9 Roztwór formamidowy do nakładania próbek na denaturujące żele poliakryloamidowe 2 x stężony.

- 95 % (o/o) formamid
- 5 mM EDTA
- 0,04 % (w/o) błękit bromofenolowy (BB)
- 0,04 % (w/o) fiolet ksylenowy (XC)

6.6 Odczynniki do cyklicznego sekwencjonowania DNA.

Do cyklicznego sekwencjonowania DNA wykorzystywano komercyjnie dostępny zestaw "Femtomol Sequencing Kit" (Promega).

6.7 Standardy do oznaczania wielkości fragmentów DNA.

6.7.1 Do analiz w żelach agarozowych stosowano 1 Kb DNA Ladder (Gibco BRL).

6.7.2 Do analiz w żelach poliakryloamidowych stosowano produkt sekwencjonowania DNA faga M13mp18 (USB).

6.8 Związki promieniotwórcze.

Do radioizotopowego znakowania badanych cząsteczek wykorzystywano [γ - ^{32}P] ATP (Amersham) o aktywności 5000 Ci/mmol.

6.9 Żele.

6.9.1 Żele agarozowe.

- a) Do analizy produktów PCR używano zwykle 1% żelu agarozowego z 1 x stężonym buforem TBE. Do 100 ml żelu dodano 5 μ l bromku etydyny (10 mg/ml) w celu uwidocznienia DNA.
- b) Do oczyszczania produktów PCR stosowano 1,5% agarozę Metaphore w 0.5 x stężonym buforze TBE. Stężenie bromku etydyny w żelu było takie jak podano wyżej.

6.9.2 Żele poliakryloamidowe.

a) Żele denaturujące.

Roztworem podstawowym był 20% roztwór akryloamidu i bisakryloamidu zmieszanych w stosunku wagowym 19:1 oraz 7,5 M mocznik rozpuszczone w buforze 1 x TBE. Żele o wymaganym stężeniu uzyskiwano poprzez zmieszanie roztworu podstawowego z 7,5 M mocznikiem (w 1 x TBE) w następujących proporcjach (podano na 80 ml żelu):

6% 24 ml roztworu podstawowego i 56 ml roztworu mocznika

10% po 40 ml każdego z roztworów

12% 48 ml roztworu podstawowego i 32 ml roztworu mocznika.

W celu zainicjowania polimeryzacji do mieszaniny żelującej dodawano następujące katalizatory: 600 μ l 10% roztworu nadsiarczanu amonu i 50 μ l TEMED – u na 80 ml żelu.

b) Żele niedenaturujące.

Podstawowym żelem niedenaturującym stosowanym w przeprowadzonych badaniach był żel 10%, stosunek akryloamidu i bisakryloamidu (49:1) w buforze 0,5 x TB. Część eksperymentów wykonano z zastosowaniem żeli 8% (29:1) w buforze 1x lub 0,5x stężonym.

Poza buforami standartowymi do przygotowania żeli niedenaturujących stosowano też bufory o składzie takim jak bufory do denaturacji / renaturacji (6.4.4, 6.4.5, 6.4.6)

W celu polimeryzacji dodawano takie same ilości katalizatorów jak w przypadku denaturujących żeli poliakryloamidowych.

6.10 Syntetyczne oligodeoksynukleotydy.

Syntetyczne oligonukleotydy stosowane były jako startery w amplifikacji DNA metodą PCR prowadzącej między innymi do uzyskania matryc do transkrypcji *in vitro*. Syntetyzowano je w Instytucie Chemii Bioorganicznej PAN oraz zamawiano w firmie MWG - Biotech GmbH lub Instytucie Biochemii i Biofizyki PAN w Warszawie. Ich sekwencje i długości podano w tabeli 6.1.

Tabela 6.1 Syntetyczne oligodeoksynukleotydy. F – starter „forward”; R – starter „reverse”; podkreślono sekwencję promotora polimerazy RNA faga T7.

L.p.	Oznaczenie	Oligodeoksynukleotyd (5' → 3')	Długość (nt)
1.	Had 1 (F)	GGACGCCCTGGAAAAGCTGATGA	23
2.	Had 1 T7 (F)	<u>TAATACGACTCACTATAGGACGCCCTGG</u> AAAAGCTGATGA	40
3.	Had 2 (R)	TGAGGCAGCAGCGGCTGT	18
4.	B1 (F)	AGCAGCAGCAGCAACAGCC	19
5.	C1 (R)	CGGCGGTGGCGGCTGTTG	18
6.	M13 (F)	TTCCAGTCACGACGTT	17
7.	AR1 (F)	TGCGCGAAGTGATCCAGAA	19
8.	AR2 (R)	CTTGGGGAGAACCATCCTCT	20
9.	AR1 T7 (F)	<u>TAATACGACTCACTATAGGCGCGAAGTG</u> ATCCAGAA	37
10.	T7GG	TAATACGACTCACTATAGG	
11.	CAG17	(CTG) ₁₇ <u>CCTATAGTGAGTCGTATTA</u>	70
12.	CAG17_cl	GGCCC(CTG) ₁₇ GGGCCT <u>TATAGTGAGTCG</u> TATTA	76
13.	CAG16_cl	GGCCC(CTG) ₁₆ GGGCCT <u>TATAGTGAGTCG</u> TATTA	73
14.	CCG17	(CGG) ₁₇ <u>CCTATAGTGAGTCGTATTA</u>	70
15.	CCG17_cl	GGCCC(CGG) ₁₇ GGGCCT <u>TATAGTGAGTCG</u> TATTA	76
16.	GCC17_cl	GGCCC(GGC) ₁₇ GGGCCT <u>TATAGTGAGTCG</u> TATTA	76
17.	(CAG) ₂₀	(CTG) ₂₀ <u>CCTATAGTGAGTCGTATTA</u>	79
18.	(CAG) ₁₀ (CCG) ₁₀	(CGG) ₁₀ (CTG) ₁₀ <u>CCTATAGTGAGTCGTATT</u> A	79
19.	(CCG) ₂₀	(CGG) ₂₀ <u>CCTATAGTGAGTCGTATTA</u>	79

6.11 Ważniejsza aparatura stosowana w pracy.

- Aparat do elektroforezy agarozowej (Gibco BRL model Horizon 58)
- Aparat do elektroforezy w żelu poliakryloamidowym (Gibco BRL model S2)
- Aparat do elektroforezy pionowej (Kucharczyk)
- Kasety do autoradiografii (Sigma) z ekranem wzmacniającym (DUPONT)
- Licznik scyntylicyjny (Beckmann β -mate II)
- pH - metr (Fisher)
- Speed-vac SC (Savant)
- Spektrofotometr (UV / VIS Perkin Elmer model Lambda Bio 10)
- Spektrometr Beckmann DU 640 z urządzeniem kontrolującym temperaturę
- Suszarka do żeli poliakryloamidowych (Biorad model 583)
- Termocykler (Perkin Elmer Model 480)
- Termocykler (Perkin Elmer System 9600)
- Termocykler (Perkin Elmer System 9700)
- Termomikser (Eppendorf model 5436)
- Wirówka (Eppendorf model 5402)
- Zasilacz do elektroforezy niskonapięciowej (Gibco BRL model 200)
- Zasilacz do elektroforezy wysokonapięciowej (Pharmacia model EPS 3500)
- PhosphorImager „Typhoon” (Amersham Pharmacia)

7 METODY

7.1 Amplifikacja fragmentów genu metodą PCR.

7.1.1 PCR ze starterami nie znakowanymi radioizotopowo.

Skład mieszaniny reakcyjnej na 10 reakcji w objętości 10 μ l przedstawiono w tabeli 7.1.

Do probówek na 200 μ l podano po 9 μ l mieszaniny reakcyjnej oraz 1 μ l (100ng) genomowego DNA.

Reakcję amplifikacji fragmentów genu *IT15* prowadzono w następujących warunkach:

95°C 10 min. wstępnej denaturacji

35 cykli:

95°C/20 s denaturacja

60°C/20 s hybrydyzacja starterów

72°C/45 s wydłużanie starterów

72°C/10 min. końcowe wydłużanie.

Tabela 7.1. Skład mieszaniny reakcyjnej PCR ze starterami nie znakowanymi radioizotopowo (na 100 μ l mieszaniny reakcyjnej).

Objętość	Odczynnik	Stężenie końcowe
59,5 μ l	H ₂ O (dejonizowana)	-
10 μ l	10x stężony bufor do PCR	1 x
2 μ l	dATP	200 μ M
2 μ l	dCTP	200 μ M
2 μ l	dGTP	200 μ M
2 μ l	dTTP	200 μ M
1 μ l	Starter "forward"	100 μ M
1 μ l	Starter "reverse"	100 μ M
10 μ l	DMSO	10%
0,5 μ l	Polimeraza Taq	2,5U/100 μ l
90 μ l	Suma	

Reakcję amplifikacji fragmentów genu *AR* prowadzono w następujących warunkach:

95°C 5 min. wstępnej denaturacji

30 cykli:

95°C/10 s denaturacja

57°C/10 s hybrydyzacja starterów

72°C/20 s wydłużanie starterów

72°C/10 min. końcowe wydłużanie.

Po zakończeniu procedury amplifikacji DNA do 10µl mieszaniny poreakcyjnej dodawano 2µl roztworu obciążającego do nakładania na żel agarozowy (6.5.6) i całość nanoszono na 1% żel agarozowy z bromkiem etydyny (6.9.1). Elektroforezę prowadzono pod napięciem 100V przez 20 minut. Obecność produktu reakcji obserwowano w świetle UV.

7.1.2 PCR ze starterami znakowanymi radioizotopowo na końcu 5'.

PCR ze starterem znakowanym radioizotopowo przeprowadzono w celu określenia dokładnej długości fragmentów genu powstających na matrycy różnych alleli oraz w celu rozdzielenia ich w wypadku użycia heterozygotycznego genomowego DNA jako matrycy.

7.1.2.1 Znakowanie starterów na końcu 5'.

Do przeprowadzenia reakcji znakowania radioizotopem końca 5' startera „forward” sporządzono mieszaninę o składzie przedstawionym w tabeli 7.2.

Tabela 7.2. Skład mieszaniny reakcyjnej znakowania starterów na końcu 5' (na 10µl mieszaniny reakcyjnej).

Objętość	Odczynnik	Stężenie końcowe
6,5µl	H ₂ O (dejonizowana)	-
1µl	Starter “forward”	10µM
1µl	10x stężony bufor do kinazy	
0,5µl	[γ- ³² P] ATP	5 µCi
1µl	Kinaza polinukleotydowa faga T4	10U

Mieszaninę inkubowano w temperaturze 37°C przez 30 minut. Następnie inaktywowano enzym 10 min. w 80°C.

7.1.2.2 Reakcja amplifikacji fragmentów DNA z wykorzystaniem starterów znakowanych radioizotopowo.

Amplifikację DNA przeprowadzano w sposób opisany w rozdziale 7.1.1 z tym, że zamiast startera „forward” podano całość mieszaniny (10µl) po reakcji znakowania (7.1.2.1) co spowodowało konieczność zmniejszenia objętości wody do 49,5µl.

Po reakcji do każdej probówki z mieszaniną poreakcyjną dodano 10µl formamidowego roztworu do nakładania na żele poliakrylamidowe (6.5.9) i po 3 minutowej denaturacji w 95°C pobrano 3µl w celu nałożenia na 6% żel poliakrylamidowy / 7,5 M mocznik (6.9.2a). Jako standard długości wykorzystano produkt sekwencjonowania DNA faga M13mp18.

7.1.2.3 Przygotowanie żelu poliakrylamidowego i elektroforeza.

Płyty szklane o wymiarach 42 na 34 cm umyto oraz spłukano etanolem i rozdzielono przekładkami o grubości 0,4 mm. Po dodaniu związków katalizujących polimeryzację (nadsiarczanu amonu i TEMEDu) roztwór akrylamidu wylano między płyty i uformowano grzebieniem kieszonki. Po spolimeryzowaniu i usunięciu grzebienia przepłukano kieszonki 1x stężonym buforem TBE. Przed nałożeniem próbek żel poddano preelektroforezie prądem o napięciu 1500V i natężeniu 45mA przez 45 minut. Elektroforezę prowadzono przez 2 – 5 godzin w zależności od długości rozdzielanych fragmentów w warunkach 40mA/2000V. Po zakończeniu rozdzielania fragmentów DNA żele przenoszono z płyty na bibułę Whatman 3MM, przykrywano folią i suszono 45 minut do 1 godziny w suszarce do żeli. Suchy żel po przyklejeniu fluorescencyjnych znaczników (Stratagene) przykładano do kliszy rentgenowskiej, którą naświetlano 12 – 24 godzin w kasecie z ekranem wzmacniającym w – 70°C.

7.1.2.4 Elucja produktów PCR z żelu poliakryloamidowego w celu uzyskania homogennych fragmentów genów *IT15* lub *AR*, których używano jako matryc do reamplifikacji.

Po zakończeniu autoradiografii z żelu poliakryloamidowego sterylnym skalpelem wycięto fragmenty zawierające produkty PCR na podstawie obrazu na kliszy. Fragmenty żelu umieszczone w probówkach Eppendorf o pojemności 1,5ml zalano 250 μ l sterylnej dejonizowanej wody w celu eluowania z nich DNA. Elucję prowadzono przez 12 godzin. Po zwirowaniu z dna fragmentów żelu zebrano 200 μ l roztworu, który używano następnie jako matrycy do reamplifikacji. Roztwór przechowywano w – 20°C.

7.2 Cykliczne sekwencjonowanie DNA.

Produkt cyklicznego sekwencjonowania DNA faga M13mp18 był stosowany jako wzorzec pomocny w określeniu długości fragmentów alleli genu *IT15* lub *AR* rozdzielanych w żelach poliakryloamidowych.

7.2.1 Znakowanie na końcu 5' startera M13 używanego w reakcji sekwencjonowania.

Skład mieszaniny reakcyjnej przedstawiony został w tabeli 7.3.

Reakcję znakowania przeprowadzano w warunkach opisanych w rozdziale 7.1.2.1.

Tabela 7.3. Skład mieszaniny reakcyjnej do znakowania na końcu 5' startera M13 używanego w reakcji sekwencjonowania faga M13 (na reakcję w objętości 10 μ l).

Objętość	Odczynnik	Stężenie końcowe
5,5 μ l	H ₂ O (dejonizowana)	-
2 μ l	Starter M13	10 μ M
1 μ l	10x stężony bufor do kinazy	1x
0,5 μ l	[γ - ³² P] ATP	5 μ Ci
1 μ l	Kinaza polinukleotydowa faga T4	10U

7.2.2 Cykliczne sekwencjonowanie DNA faga M13.

Po wyznakowaniu startera (rozdział 7.2.1) przygotowano mieszaninę reakcyjną (skład w tabeli 7.4), którą następnie podzielono na cztery części po 8 μ l. Każdą z nich umieszczono w oddzielnej probówce (o pojemności 200 μ l) oznaczonej A, C, G lub T. Do probówek podano także po 2 μ l mieszaniny terminującej (ddNTP/dNTP w odpowiednim stosunku).

Reakcję prowadzono w warunkach:

95°C/3 min wstępna denaturacja

20 cykli:

90°C/10 s denaturacja

60°C/10 s hybrydyzacja i wydłużanie startera.

Po zakończeniu reakcji do każdej probówki dodano 3 μ l roztworu STOP. Po 3 minutach denaturacji w 95°C pobrano po 3 μ l i nakładano na 6% żel poliakryloamidowy / 7,5 M mocznik (6.9.2a). Elektroforezę prowadzono zgodnie z opisem w rozdziale 7.1.2.3.

Tabela 7.4. Skład mieszaniny reakcyjnej cyklicznego sekwencjonowania faga M13 (na reakcję w objętości 10 μ l).

Objętość	Odczynnik	Stężenie końcowe
1 μ l	H ₂ O (dejonizowana)	-
5 μ l	znakowany starter M13	10 μ M
5 μ l	5x stężony bufor do sekwencjonowania	1x
5 μ l	DNA	200ng
1 μ l	polimeraza Taq	5U

7.3 Reamplifikacja fragmentów genów *IT15* i *AR* na matrycy produktów PCR.

Reakcję reamplifikacji fragmentów genów *IT15* i *AR* przeprowadzano w takich samych warunkach jak reakcje opisane w rozdziale 7.1.1. Różnica polegała na tym, że zamiast genomowego DNA jako matrycy używano roztworu uzyskanego po elucji (7.1.2.3) w ilości 1 μ l na 10 μ l reakcji. Produkty analizowano w żelu agarozowym zgodnie procedurą stosowaną w przypadku amplifikacji (rozdział 7.1.1).

7.4 Reamplifikacja ze starterami zawierającymi promotor polimerazy RNA faga T7.

Fragmenty genów *IT15* i *AR* uzyskane według procedury opisanej w rozdziałach 7.1 i 7.3 poddano reamplifikacji ze starterami odpowiednio Had 1 T7 i Had 2 oraz AR1 T7 i AR2 (tabela 6.1). Stosowano takie warunki reakcji jak w przypadku amplifikacji (rozdział 7.1.1). Startery Had 1 T7 oraz AR T7 posiadały na końcu 5' promotor polimerazy RNA faga T7. Produkty PCR powstałe z wykorzystaniem takich zestawów starterów były matrycą do transkrypcji *in vitro* w celu uzyskania fragmentów mRNA eksonu 1 genów *IT15* i *AR* zawierających ciągi powtórzeń trójnukleotydowych.

7.5 Oczyszczanie matryc do transkrypcji *in vitro*.

Produkty reamplifikacji z zastosowaniem starterów zawierających promotor polimerazy RNA faga T7 rozdzielano w 1,5% żelu agarozowym o obniżonej temperaturze topnienia (agarozą Metaphore 6.1.9b). Po trwającej 45 minut elektroforezie prowadzonej pod napięciem 100V wycinano fragment żelu zawierający produkt PCR. Po umieszczeniu w 250 μ l buforu do ekstrakcji DNA z agarozy (6.7.4) ogrzewano go w 75°C przez 6 minut wytrząsając w celu ułatwienia elucji DNA. Po schłodzeniu do temperatury 25°C roztwór ekstrahowano równą objętością fenolu o pH 8.0 silnie wytrząsając przez minutę. Oddzieloną od warstwy fenolowej przez 10 minutowe wirowanie warstwę wodną poddawano następnie dwukrotnie ekstrakcji chloroformem prowadzonej w takich samych warunkach jak opisano wcześniej. Uzyskany wodny roztwór DNA poddano precypitacji trzema objętościami 96% etanolu w obecności 0,3 M octanu sodu o pH 5.2 (6.5.2) przez 30 min. do 12 godzin w temperaturze – 70°C. Po zwirowaniu, dwukrotnym przemyciu 70% etanolem osad wysuszono i rozpuszczono w 10 – 50 μ l H₂O uzyskując matrycę DNA do transkrypcji *in vitro*.

7.6 Przygotowanie syntetycznych matryc do transkrypcji *in vitro*.

Oligomery (tabela 6.1) zawierające sekwencję komplementarną do transkryptu i regionu promotorowego dla polimerazy RNA faga T7 na końcu 3' zostały poddane oczyszczaniu na żelu poliakryloamidowym (analogicznie do procedury oczyszczania

transkryptów uzyskanych *in vitro*). Podwójnoniciowe cząsteczki DNA uzyskano na drodze dwuetapowej reakcji wydłużania startera. 200pmoli oczyszczonego oligomeru oraz 1mmol oligomeru T7GG zostało zawieszono w komercyjnym buforze do PCR wraz z trójfosforanami deoksyrybonukleotydów A, C, G, T w stężeniu 200 μ M każdy i polimerazą DNA Taq (0,5U na 100 μ l reakcji). Mieszanina ta była inkubowana przez 50 cykli w warunkach 94°C/15s i 45°C/15s. Po zakończeniu reakcji jej produkty zostały oczyszczone przy użyciu filtrów Microcon YM30 (Millipore) zgodnie z procedurą zalecaną przez producenta.

7.7 Transkrypcja *in vitro*.

Matrycą w reakcji były przygotowane w sposób opisany w rozdziałach 7.4 i 7.5 fragmenty DNA z dodaną sekwencją promotora polimerazy RNA faga T7 lub chemicznie syntetyzowane oligonukleotydy poddane hybrydyzacji w celu uzyskania regionów dwuniciowych (rozdział 7.6). Skład mieszaniny reakcyjnej przedstawiono w tabeli 7.5.

Guanozynę, którą stosowano w celu umożliwienia znakowania końca 5' powstałych fragmentów RNA bez konieczności ich wcześniejszej defosforylacji, dokładnie rozpuszczano ogrzewając 3 minuty w temperaturze 65°C przed podaniem do mieszaniny reakcyjnej.

Tabela 7.5. Skład mieszaniny reakcyjnej transkrypcji *in vitro* (na 50 μ l mieszaniny reakcyjnej).

Objętość	Odczynnik	Stężenie końcowe
3,5 μ l	H ₂ O (dejonizowana)	-
10 μ l	Matrycy DNA	0,5 – 1,5 μ M
10 μ l	5x stężony bufor do transkrypcji	1x
5 μ l	DTT	10mM
1,25 μ l	100mM ATP	2,5mM
1,25 μ l	100mM CTP	2,5mM
1,25 μ l	100mM GTP	2,5mM
1,25 μ l	100mM UTP	2,5mM
15 μ l	10mM guanozyna	3mM
1,5 μ l	Polimeraza RNA faga T7	300U

Transkrypcję prowadzono 5 godzin w temperaturze 37°C. Po jej zakończeniu do 2µl mieszaniny poreakcyjnej dodawano 3µl H₂O i 1µl roztworu obciążającego do nakładania próbek na żele agarozowe (6.5.6) i analizowano w 1% żelu agarozowym w warunkach opisanych w rozdziale 7.1.1.

7.8 Oczyszczanie transkryptu w żelu poliakryloamidowym i elucja RNA.

RNA po transkrypcji *in vitro* strącono jak opisano wcześniej trzema objętościami 96% etanolu w obecności 0,3M octanu sodu o pH 5.2 (6.5.2) przez 0,5 do 12 godzin w temperaturze – 70°C. Po zwirowaniu i dwukrotnym przemyciu 70% etanolem wysuszony osad rozpuszczono w 15µl H₂O oraz 15µl roztworu UED (6.5.8). W celu oczyszczenia transkryptu tak przygotowany roztwór nałożono na 6% lub 10% (w zależności od długości oczyszczanej cząsteczki) żel poliakryloamidowy / 7,5M mocznik (6.9.2a) o grubości 1 mm. Elektroforezę prowadzono przez 2 – 4 godzin w warunkach 1500V/40mA. Fragmenty żelu zawierające produkty transkrypcji wybarwione w wyniku 30 do 60 minutowej inkubacji w roztworze "Stains – All" (6.5.4b) wycięto. RNA eluowano z żelu poliakryloamidowego 250µl roztworu Elu RNA (6.5.1) przez 3 – 12 godzin w temperaturze pokojowej. Po zakończeniu elucji RNA, dokładnie zebrano roztwór z nad fragmentów żelu i poddano precypitacji 750µl 96% etanolu. Osad uzyskany po zwirowaniu przechowywano w temperaturze – 70°C, rozpuszczano w wodzie bezpośrednio przed reakcją znakowania.

7.9 Znakowanie radioizotopem końca 5' transkryptu.

Reakcję znakowania prowadzono inkubując mieszaninę, której skład podano w tabeli 7.6 w temperaturze 37°C przez 10 minut. Zakończano ją dodając 10µl roztworu UED (6.5.8) i produkt oczyszczano w żelu poliakryloamidowym jak opisano w rozdziale 7.7. Po elektroforezie na żelu umieszczono fluorescencyjne znaczniki (Stratagene) i przyłożono kliszę rentgenowską na 1 – 5 minut. Po wywołaniu autoradiogramu wycięto z żelu fragmenty zawierające znakowany radioizotopem transkrypt, który następnie eluowano według procedury opisanej w rozdziale 7.7. Po zakończeniu elucji RNA poddano precypitacji przeprowadzonej zgodnie ze schematem zawartym w rozdziale 7.7. Na każdym etapie, począwszy od wycięcia fragmentu żelu dokonywano pomiaru poziomu radioaktywności za pomocą licznika

scyntylicyjnego. Po zakończeniu całej procedury uzyskiwano $0,65 - 25 \times 10^6$ cpm znakowanego RNA.

Tabela 7.6. Skład mieszaniny reakcyjnej do znakowania na końcu 5' cząsteczek RNA uzyskanych drogą transkrypcji *in vitro* (na 20 μ l mieszaniny reakcyjnej).

Objętość	Odczynnik	Stężenie końcowe
12 μ l	Transkrypt rozpuszczony w H ₂ O (dejonizowanej)	-
2 μ l	10x stężony bufor do kinazy	1x
3 μ l	[γ - ³² P] ATP	30 μ Ci
1 μ l	Kinaza polinukleotydowa faga T4	10U

7.10 Analiza RNA w żelach niedenaturujących.

Cząsteczki RNA analizowano w 8% i 10% żelach niedenaturujących (6.9.2b) o grubości 0,4 mm lub 1 mm i wymiarach 150mm x 150mm. Przed nałożeniem na żel do roztworu RNA zawieszono w buforze do denaturacji / renaturacji (6.4.4, 6.4.5, 6.4.6) dodawano roztwór obciążający do nakładania próbek na niedenaturujące żełe poliakryloamidowe (6.5.7). Elektroforezę prowadzono przez 4 – 6 godzin przy mocy prądu 10 W.

7.11 Hydroliza cząsteczek RNA sondami strukturalnymi

Z roztworu wodnego oczyszczonych znakowanych na końcu 5' cząsteczek RNA o określonej aktywności pobierano porcje stanowiące substrat do reakcji, których produkty używane były jako znaczniki długości (rozdziały 7.12.1 i 7.12.2) w czasie elektroforetycznej analizy produktów reakcji hydrolizy RNA z sondami

Tabela 7.7 Skład mieszaniny poddawanej procedurze denaturacji i renaturacji przed reakcjami RNA z sondami strukturalnymi (na 10 reakcji).

Objętość	Odczynnik
10 μ l	Znakowany RNA
11 μ l	10x stężony bufor do denaturacji / renaturacji (3.4.4 lub 3.4.5 lub 3.4.6)
34 μ l	H ₂ O (dejonizowana)

strukturalnymi (rozdział 7.12.3). Pozostałą większą część RNA wykorzystywano do przygotowania mieszaniny reakcyjnej (jej skład przedstawiono w tabeli 7.7), którą poddawano wstępnej procedurze denaturacji / renaturacji, polegającej na ogrzewaniu w temperaturze 75°C przez 3 minuty a następnie inkubacji 10 minut w temperaturze 37°C. Standardowym buforem do denaturacji / renaturacji był bufor Mg1 (6.4.5) zawierający 1mM MgCl₂, natomiast bufony Mg10 (6.4.4) i Mg1/K (6.4.6) stosowano w celu określenia wpływu zmienionych warunków środowiska na badaną strukturę cząsteczki RNA.

Reakcje hydrolizy przeprowadzano dodając do 5µl mieszaniny RNA równą objętość roztworu sondy strukturalnej o stężeniu podanym tabeli 7.8. Jako reakcję kontrolną (Ci) traktowano próbkę, w której zamiast roztworu sondy podano taką samą objętość H₂O. Po 15 minutach inkubacji w temperaturze 37°C reakcje przerywano dodaniem 10µl roztworu UED (6.5.8) i zamrożeniem mieszaniny.

Odstępstwa od podanego w tym rozdziale schematu analizy strukturalnej podano w opisach rycin przedstawiających wyniki eksperymentów.

Tabela 7.8 Stężenia stosowanych sond strukturalnych.

Sonda	stężenie		
	1	2	3
Pb ²⁺	0,5mM	1mM	2mM
RNaza T1	0,035U/µl	0,075U/µl	0,15U/µl
Nukleaza S1 ^b	2,5U/µl	5U/µl	10U/µl
RNaza T2	0,05U/µl	0,1U/µl	0,2U/µl
RNaza V1	0,05U/ml	0,1U/ml	0,2U/ml

7.12 Analiza produktów reakcji cząsteczek RNA z sondami strukturalnymi.

Precyzyjne określenie reaktywnych wiązań fosfodwuestrowych było możliwe dzięki elektroforetycznemu rozdzielaniu w żelach poliakryloamidowych z 7,5M mocznikiem produktów właściwych reakcji sondowania struktury wraz z produktami reakcji opisanych w rozdziałach 7.12.1 i 7.12.2 umożliwiającymi określenie długości rozdzielanych cząsteczek otrzymanych po rozcięciu substratu.

7.12.1 Statystyczna hydroliza RNA roztworem formamidowym.

Do reakcji użyto 2 μ l roztworu zawierającego znakowany RNA. Do próbówki podano także 9 μ l 1x stężonego roztworu formamidowego do hydrolizy RNA (6.5.3). Mieszaninę reakcyjną ogrzewano przez 10 minut w 100°C. Reakcję przerywano przez dodanie 9 μ l roztworu UED (6.5.8) i zamrożenie.

7.12.2 Trawienie RNA w warunkach semidenaturujących rybonukleazą T1.

Sporządzoną mieszaninę reakcyjną, w której skład wchodziło 2 μ l roztworu wodnego znakowanego transkryptu i 3 μ l 2x stężonego buforu do rybonukleazy T1 (6.4.4) ogrzewano w temperaturze 100°C a następnie schłodzono do temperatury 55°C. Właściwą reakcję trawienia prowadzono przez 15 minut w temperaturze 55°C po dodaniu 1U rybonukleazy T1 i przerywano ją przez dodanie 14 μ l roztworu UED (6.5.8) i zamrożenie.

7.12.3 Elektroforetyczny rozdział produktów reakcji z sondami strukturalnymi.

Po zakończeniu reakcji 3 μ l produktu ogrzewano 3 minuty w 95°C w celu zdenaturowania. Następnie наносzono w zależności od długości badanych fragmentów na 6% lub 12% żel poliakrylamidowy z 7,5M mocznikiem. Elektroforezę i autoradiografię prowadzono w sposób opisany w rozdziale 7.1.2.3. Aby umożliwić ilościowe porównanie wzoru cięć powstających w cząsteczkach RNA w wyniku działania sondy zwracano dużą uwagę na wyrównanie ilości używanego do reakcji znakowanego radioizotopem RNA i standaryzacji warunków elektroforezy i czasu ekspozycji błony rentgenowskiej podczas autoradiografii.

7.12.4 Analiza produktów reakcji sondowania struktury RNA z wykorzystaniem skanera PhosphorImager „Typhoon”.

Poza autoradiografią do wizualizacji wyników reakcji sondowania struktury wykorzystywano skaner PhosphorImager „Typhoon”. Obraz uzyskiwano po zeskanowaniu powierzchni ekranu będącego na wyposażeniu tego urządzenia. Ekran naświetlano poddając go ekspozycji na promieniowanie fosforu ^{32}P znajdującego się w wysuszonym żelu poliakrylamidowym, w którym rozdzielano

produkty reakcji sondowania struktury RNA. Naświetlanie trwało od 6 do 12 godzin w zależności od intensywności promieniowania.

Oprogramowanie urządzenia PhosphorImager „Typhoon” ImageQuant 5.1 dostępne w pakiecie programów „Molecular Dynamics” umożliwia analizę uzyskanych elektroforegramów. Przy jego zastosowaniu wykonano analizę ilościową produktów reakcji sondowania struktury polegającą na bezpośrednim porównaniu intensywności poszczególnych cięć widocznych na obrazie w formie prążków. Użycie programu ImageQuant pozwoliło na porównanie cięć widocznych w pojedynczej ścieżce elektroforezy oraz pomiędzy różnymi ścieżkami. Uzyskane ten sposób bezwzględne wartości intensywności wszystkich obserwowanych cięć pozwalały na dokonanie właściwej ich oceny, a także porównań sposobu rozpoznawania odpowiednich regionów różnych cząsteczek RNA.

7.13 Analiza właściwości termodynamicznych struktur cząsteczek RNA.

Oczyszczone nie znakowane modelowe transkrypty (CAG)₂₀, (CAG)₁₀(CCG)₁₀ i (CCG)₂₀ były poddane topieniu tworzonej przez nie struktury po zawieszeniu w buforze do analiz termodynamicznych (6.4.7). Zmiany struktury badanych transkryptów były określane na podstawie pomiaru absorpcji fali o długości 260nm w czasie ogrzewania próbek od 0 do 90°C z szybkością 1°C na minutę przy użyciu spektrofotometru Beckmann DU 640 z systemem kontroli temperatury. Do obliczenia krzywych topnienia i parametrów termodynamicznych wykorzystano oprogramowanie MeltWin 3.5.

7.14 Bioinformatyczna identyfikacja ludzkich mRNA zawierających powtórzenia trójnukleotydydowe.

W celu zidentyfikowania ludzkich mRNA zawierających powtórzenia trójnukleotydydowe przeprowadzono analizy sekwencji zgromadzonych w bazach danych RefSeq i GenBank. Wykorzystano algorytm BLASTn (*Basic Local Alignment Search Tool – nucleotide*) dostępnych na stronach internetowych NCBI (*National Centre for Biotechnology*) (<http://www.ncbi.nlm.nih.gov/BLAST>). Celem poszukiwań było znalezienie ciągów trójnukleotydydowych złożonych z sześciu i więcej motywów. Zastosowano metodę wyszukiwania sekwencji zawierających określony motyw

sekwencyjny. Ze względu na sposób zapisu sekwencji w formie ciągu dziesięciu nukleotydów stosowany analizowanych bazach danych poszukiwano sekwencji o długości 6 nukleotydów odpowiadającej 2 powtórzeniom motywu trójnukleotydowego. Przy pomocy dostępnych funkcji oprogramowania zawężono pulę przeszukiwanych sekwencji do sekwencji ludzkich. Nie można było zastosować ograniczenia do sekwencji mRNA ze względu na fakt, że w bazie GenBank istnieje pewna grupa sekwencji zapisanych w formie cDNA nie mających swoich odpowiedników w zapisie mRNA.

Po otrzymaniu pierwszych wyników spośród wybranych sekwencji usunięto rekordy odpowiadające sekwencjom bakteryjnym zawartym w zdeponowanych sekwencjach klonów i wielokrotnie powtórzone sekwencje fragmentów tych samych genów pozostawiając do dalszych analiz możliwie najbardziej pełne sekwencje pojedynczych ludzkich mRNA lub cDNA. Kolejnym krokiem było wyeliminowanie sekwencji zawierających ciągi krótsze niż sześć powtórzeń trójnukleotydowych oraz nie spełniające warunku lokalizacji w mRNA. Ostatnim etapem analiz bioinformatycznych było określenie funkcji genów zawierających powtórzenia trójnukleotydowe w sekwencji mRNA. Informacje na ten temat pochodziły z opisów genów zawartych w przeszukiwanych bazach danych a w przypadku ich braku poszukiwano odpowiednich opisów w bazie OMIM (*Online Mendelian Inheritance in Man*) dostępnej także na internetowych stronach NCBI (<http://www.ncbi.nlm.nih.gov/OMIM>).

7.15 Analiza struktury RNA za pomocą programów komputerowych.

Komputerową analizę drugorzędowych struktur RNA prowadzono za pomocą programu: *mfold* wersja 3.1 dostępnego poprzez sieć Internet (<http://mfold2.wustl.edu/~mfold/rna>). Wykorzystuje on algorytmy minimalizacji energii swobodnej ΔG i metody najbliższego sąsiedztwa. W niniejszej pracy komputerowe modelowanie struktury drugorzędowej RNA zastosowano w dwóch podejściach.

Pierwszym z nich było poszukiwanie struktury najbardziej zgodnej z danymi eksperymentalnymi. Modelowano sekwencję cząsteczek RNA badanych metodami biochemicznymi. Wykorzystywano także możliwości programu pozwalające na określenie, które regiony modelowanej cząsteczki powinny występować w strukturze

w formie pojedynczoniowej. Wyboru takich niesparowanych zasad dokonywano na podstawie wyników sondowania struktury badanych transkryptów.

Drugim zastosowaniem programu *mfold* było określenie prawdopodobnej struktury fragmentów ludzkich mRNA zidentyfikowanych jako zawierające kilka ciągów powtórzeń trójnukleotydowych. Modelowano struktury sekwencji RNA obejmującej regiony ciągów powtórzeń trójnukleotydowych i sekwencji specyficznej oddzielającej je od siebie otoczone po stronie 5' i 3' fragmentami o długości 100 nukleotydów.

8 Bibliografia

Adachi H., Katsuno M., Minamiyama M., Waza M., Sang C., Nakagomi Y., Kobayashi Y., Tanaka F., Doyu M., Inukai A., Yoshida M., Hashizume Y., Sobue G. (2005). Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain*. 128(Pt 3);659-70.

Alonso M. E., Yescas P., Cisneros B., Martinez C., Silva G., Ochoa A., Montanez C. (1997). Analysis of the (CAG)_n repeat causing Huntington's disease in a Mexican population. *Clin Genet*. 51;225-30.

Andrew S. E., Goldberg Y. P., Kremer B., Telenius H., Theilmann J., Adam S., Starr E., Squitieri F., Lin B., Kalchman M. A., i wsp. (1993). The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat Genet*. 4;398-403.

Bates G. (2003). Huntingtin aggregation and toxicity in Huntington's disease. *Lancet* 361;1642 – 1644.

Bates G., Leirach H. (1994). Trinucleotide repeat expansions and human genetic disease. *Bioessays*. 16;277-84.

Beckman J. S., Weber J. L. (1992). Survey of human and rat microsatellites. *Genomics*. 12;627-31.

Broda M., Kierzek E., Gdaniec Z., Kuliński T., Kierzek R. (2005). Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry*. 16;10873 – 10882.

Brook J. D. i wsp. (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.

Brown C. J., Goss S. J., Lubahn D. B., Joseph D. R., Wilson E. M., French F. S., Willard H. F. (1989) Androgen receptor locus on the human X chromosome; regional

localization to Xq11-12 and description of a DNA polymorphism. *Am J Hum Genet.* 44(2);264-9.

Brown V., Jin P., Ceman S., Darnell J. C., O'Donnell W. T., Tenenbaum S. A., Jin X., Feng Y., Wilkinson K. D., Keene J. D., Darnell R. B., Warren S. T. (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell.* 107;477 – 487.

Campuzano V., Montermini L., Molto M. D., Pianese L., Cossee M., Cavalcanti F., Monros E., Rodius F., Duclos F., Monticelli A., Zara F., Canizares J., Koutnikova H., Bidichandani S. I., Gellera C., Brice A., Trouillas P., De Michele G., Filla A., De Frutos R., Palau F., Patel P. I., Di Donato S., Mandel J. L., Coccozza S., Koenig M., Pandolfo M. (1996). Friedreich's ataxia; autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271;1423 – 1427.

Campuzano V., Montermini L., Lutz Y., Cova L., Hindelang C., Jiralerspong S., Trottier Y., Kish S. J., Faucheux B., Trouillas P., Authier F. J., Durr A., Mandel J. L., Vescovi A., Pandolfo M., Koenig M. (1997). Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum. Mol. Genet.* 6;1771 – 1780.

Cady A. A., Myers M., Hannon G. J., Hammond S. M. (2002). Fragile X – related protein and VIG associate with the RNA interference machinery. *Genes Dev.* 16;2491 – 2496.

Chamberlain N. L., Driver E. D., Miesfeld R. L. (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res.* 22(15);3181-6.

Charlet-B. N., Savkur R. S., Singh G., Philips A. V., Grice E. A., Cooper T. A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell.* 10(1);45-53.

Chen L. S., Tassone F., Sahota P., Hagerman P. J. (2003). The (CGG)_n repeats element within 5' untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of the downstream reporter. *Hum. Mol. Gen.* 12;3067 – 3074.

Ciechanover A., Brundin P. (2003). The ubiquitin proteasome system in neurodegenerative diseases; sometimes the chicken, sometimes the egg. *Neuron* 40;427 – 446.

Corbin F., Bouillon M., Fortin A., Morin S., Rousseau F., Khandjian E. W. (1997). The fragile X mental retardation protein is associated with poly(A)⁺ mRNA in actively translating polyribosomes. *Hum. Mol. Genet.* 6;1465 – 1472.

Cossee M., Puccio H., Gansmuller A., Koutnikova H., Dierich A., LeMeur M., Fischbeck K., Dolle P., Koenig M. (2000). Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation. *Hum. Mol. Genet.* 9;1219 – 1226.

Cummings C. J., Zoghbi H. Y. (2000). Fourteen and counting; unraveling trinucleotide repeat diseases. *Hum Mol Genet.* 9;909 – 16.

Darnell J. C., Jensen K. B., Jin P., Brown V., Warren S. T., Darnell R. B. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell.* 107;489 – 499.

Darnell J. C., Fraser C. E., Mostovetsky O., Stefani G., Jones T. A., Eddy S. R., Darnell R. B. (2005). Kissing complex RNAs mediate interaction between the fragile – X mental retardation protein KH2 domain and brain polyribosomes. *Genet. Dev.* 19;903 – 918.

De Rooij K. E., De Koning Gans P. A., Skraastad M. I., Belfroid R. D., Vegter-Van Der Vlis M., Roos R. A., Bakker E., Van Ommen G. J., Den Dunnen J. T., Losekoot M. (1993). Dynamic mutation in Dutch Huntington's disease patients; increased paternal repeat instability extending to within the normal size range. *J Med Genet.* 30;996-1002.

Dode C., Durr A., Pecheux C., Mouret J. F., Belal S., Bachner L., Agid Y., Kaplan J. C., Brice A., Feingold J. i wsp. (1993) Huntington's disease in French families; CAG repeat expansion and linkage disequilibrium analysis. *C R Acad Sci III.* 316(11);1374-80.

Dunah A. W., Jeong H., Griffin A., Kim Y. M., Standaert D. G., Hersch S. M., Mouradian M. M., Young A. B., Tanese N., Krainc D. (2002). Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* 396;2238 – 2243.

Durant P. C. i Davis D. R. (1999) Stabilization of the anticodon stem-loop of tRNA^{Lys,3} by an A+C base-pair and by pseudouridine. *J Mol Biol.* 8;115-31.

Ehresmann C., Baudin F., Mougel M., Romby P., Ebel J. P., Ehresmann B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15;9109-28.

Emamian E. S., Kaytor M. D., Duvick L. A., Zu T., Tousey S. K., Zoghbi H. Y., Clark H. B., Orr H.T. (2003). Serine 776 of ataxin – 1 is critical for polyglutamine – induced disease in SCA1 transgenic mice. *Neuron* 38;375 – 387.

Fardaei M., Rogers M. T., Thorpe H. M., Larkin K., Hamshere M. G., Harper P. S., Brook J. D. (2002). Three proteins, MBNL, MBLL and MBXL, co – localize *in vitro* with nuclear foci of expanded – repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* 11;805 – 814.

Feng Y., Zhang F., Lokey L. K., Chastain J. L., Lakkis L., Eberhart D., Warren S. T. (1995). Translational suppression by trinucleotide repeats expansion at FMR1. *Science* 268;731 – 734.

Foury F., Cazzalini O. (1997). Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria. *FEBS Lett.* 411;373 – 377.

Fernandez-Funez P., Nino-Rosales M. L., de Gouyon B, She W. C., Luchak J. M., Martinez P., Turiegano E., Benito J., Capovilla M., Skinner P. J., McCall A., Canal I., Orr H. T., Zoghbi H. Y., Botas J. (2000). Identification of genes that modify ataxin – 1 – induced neurodegeneration. *Nature* 408;101 – 106.

Fu Y. H., Pizzuti A., Fenwick R. G. Jr, King J., Rajnarayan S., Dunne P. W., Dubel J., Nasser G. A., Ashizawa T., de Jong P. i wsp. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255;1256 – 1258.

Gafni J., Ellerby L. M. (2002). Calpain activation in Huntington's disease. *J Neurosci.* 22(12);4842-9

Gastier J. M., Pulido J. C., Sunden S., Brody T., Buetow K. H., Murray J. C., Weber J. L., Hudson T. J., Sheffield V. C., Duyk G. M. (1995). Survey of trinucleotide repeats in the human genome; assessment of their utility as genetic markers. *Hum Mol Genet.* 4;1829-36.

Gatchel J. R., Zoghbi H. Y. (2005). Diseases of unstable repeat expansion; mechanisms and common principles. *Nat. Rev. Genetics.* 6;743 – 755.

Gecz J., Gedeon A. K., Sutherland G. R., Mulley J. C. (1996). Identification of the gene *FMR2*, associated with FRAXE mental retardation. *Nature Genet.* 13;105 – 108.

Gecz J., Bielby S., Sutherland G. R., 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.

Glass M., Dragunow M., Faull R. L. (2000). The pattern of neurodegeneration in Huntington's disease; a comparative study of cannabinoid, dopamine, adenosine and GABA(A) receptor alterations in the human basal ganglia in Huntington's disease. *Neuroscience* 97;505 – 519.

Greco C. M., Hagerman R. J., Tassone F., Chudley A. E., Del Bigio M. R., Jacquemont S., Leehey M., Hagerman P. J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor / ataxia syndrome among fragile X carriers. *Brain.* 125;1760 – 1771.

Gu Y., Shen Y., Gibbs R. A., Nelson D. L. (1996). Identification of *FMR2*, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nature Genet.* 13;109 – 113.

Gunawardena S., Her L. S., Bruschi R. G., Laymon R. A., Niesman I. R., Gordesky-Gold B., Sintasath L., Bonini N. M., Goldstein L. S. (2003). Disruption of axonal transport by loss of huntingtin or expression of pathogenic polyQ proteins in *Drosophila*. *Neuron* 40;25 – 40.

Gusella J. F., Wexler N. S., Conneally P. M., Naylor S. L., Anderson M. A., Tanzi R. E., Watkins P. C., Ottina K., Wallace M. R., Sakaguchi A. Y. i wsp. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. 306;234-8.

Hagerman P. J., Hagerman R. J. (2004). The fragile – X premutation; a maturing perspective. *Am. J. Hum. Genet.* 74(5);805 – 816.

Hagerman R. J., Hagerman P. J. (2002). The fragile X premutation; into the phenotypic fold. *Curr. Opin. Genet. Dev.* 12;278 – 273.

Handa V., Saha T., Usdin K. (2003) The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic Acids Res.* 31(21);6243-8.

Harjes P., Wanker E. E. (2003). The hunt for huntingtin function; interaction partners tell many different stories. *Trends Biochem Sci.* 28(8);425-33.

Harper P.S., Harley H. G., Reardon W., Shaw D. J. (1992). Anticipation in myotonic dystrophy; new light on an old problem. *Am J Hum Genet.* 51;10-6.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. (2004) *Nature* 431; 931-945

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. (2001) *Nature* 409; 860-921

Ishizuka A., Siomi M. C., Siomi H. (2002). *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.* 16;2497 – 2508.

Jasińska A., Michlewski G., de Mezer M., Sobczak K., Kozłowski P., Napierała M., Krzyżosiak W. J. (2003). Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Res.* 31;5463-5468.

Jiang H., Mankodi A., Swanson M. S., Moxley R. T., Thornton C. A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* 13;3079 – 3088.

- Jin P., Zarnescu D. C., Ceman S., Nakamoto M., Mowrey J., Jongens T.A., Nelson D.L., Moses K., Warren S.T. (2004). Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nature Neurosci.* 7;113 – 117.
- Kanadia R. N., Johnstone K. A., Mankodi A., Lungu C., Thornton C. A., Esson D., Timmers A. M., Hauswirth W. W., Swanson M. S. (2003). A muscleblind knockout model for myotonic dystrophy. *Science* 302;1978 – 1980.
- Kazemi-Esfarjani P., Trifiro M. A., Pinsky L. (1995) Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor; possible pathogenetic relevance for the (CAG)_n-expanded neuropathies. *Hum Mol Genet.* 4(4);523-7.
- Kennedy W. R., Alter M., Sung J. H. (1968). Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait. *Neurology.* 18(7);671-80.
- Kenneson A., Zhang F., Hagedorn C. H., Warren S. T. (2001). Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate – length and permutation carriers. *Hum. Mol. Gen.* 10;1449 – 1454.
- Kierzek R., Burkard M. E., Turner D. H. (1999). Thermodynamics of single mismatches in RNA duplexes. *Biochemistry.* 38;14214-23.
- Kino Y., Mori D., Oma Y., Takeshita Y., Sasagawa N., Ishiura S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum Mol Genet.* 13(5);495-507.
- Knapp G. (1989). Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* 180;192-212.
- Knight S. J., Flannery A. V., Hirst M. C., Campbell L., Christodoulou Z., Phelps S. R., Pointon J., Middleton-Price H. R., Barnicoat A., Pembrey M. E., i wsp. (1993). Trinucleotide repeats amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* 74;127 – 134.

Kremer E.J., Yu S., Pritchard M., Nagaraja R., Heitz D., Lynch M., Baker E., Hyland V. J., Little R. D., Wada M., i wsp. (1991). Isolation of a human DNA sequence which spans the fragile X. *Am J Hum Genet.* 49;656-61.

Krzyżosiak W. J., Marciniak T., Wiewiorowski M., Romby P., Ebel J. P., Giege R. (1988). Characterization of the lead(II)-induced cleavages in tRNAs in solution and effect of the Y-base removal in yeast tRNAPhe. *Biochemistry.* 27;5771-7.

La Spada A. R., Wilson E. M., Lubahn D. B., Harding A. E., Fischbeck K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature.* 352(6330);77-9.

Li S. H., Li X.I. (2004). Huntingtin – protein interactions and the pathogenesis of Huntington's disease. *Trends Genet.* 20;146 – 154.

Limmer S. (1997). Mismatch base pairs in RNA. *Prog Nucleic Acid Res Mol Biol.* 57;1-39.

Liquori C. L., Ricker K., Moseley M. L., Jacobsen J. F., Kress W., Naylor S. L., Day J. W., Ranum L. P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of *ZNF9*. *Science* 293;864 – 867.

Mahadevan M., Tsilfidis C., Sabourin L., Shutler G., Amemiya C., Jansen G., Neville C., Narang M., Barcelo J., O'Hoy K., i wsp. (1992). Myotonic dystrophy mutation; an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255;1253 – 1255.

Mankodi A., Takahashi M. P., Jiang H., Beck C. L., Bowers W. J., Moxley R. T., Cannon S. C., Thornton C. A. (2002). Expanded CUG repeats trigger aberrant splicing of *ClC-1* chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol Cell.* 10(1);35-44.

Mankodi A., Logigian E., Callahan L., McClain C., White R., Henderson D., Krym M., Thornton C. A. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* 289;1769 – 1773.

- McLaughlin B. A., Spencer C., Eberwine J. (1996). CAG trinucleotide RNA repeats interact with RNA-binding proteins. *Am J Hum Genet.* 59;561-9.
- McPhaul M. J. (1999) Molecular defects of the androgen receptor. *J Steroid Biochem Mol Biol.* 69(1-6);315-22.
- Michalowski S., Miller J. W., Urbinati C. R., Paliouras M., Swanson M. S., Griffith J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res.* 27(17);3534-42.
- Michlewski G., Krzyżosiak W. J. (2004). Molecular architecture of CAG repeats in human disease related transcripts. *J. Mol. Biol.* 340,665-679.
- Miller J. W., Urbinati C. R., Teng-Umuay P., Stenberg M. G., Byrne B. J., Thornton C. A., Swanson M. S. (2000). Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J.* 19;4439 – 4448.
- Mitas M., Yu A., Dill J., Haworth I. S. (1995). The trinucleotide repeat sequence d(CGG)₁₅ forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry.* 34;12803-11.
- Mooers B. H., Logue J. S., 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(46);16626-31.
- Nadir E., Margalit H., Gallily T., Ben-Sasson S. A. (1996). Microsatellite spreading in the human genome; evolutionary mechanisms and structural implications. *Proc Natl Acad Sci U S A.* 93;6470-5.
- Nakamura K., Jeong S. Y., Uchihara T., Anno M., Nagashima K., Nagashima T., Ikeda S., Tsuji S., Kanazawa I. (2001). SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA – binding protein. *Hum. Mol. Genet.* 10;1441-1448.
- Nance M. A. (1998). Huntington disease; clinical, genetic, and social aspects. *J Geriatr Psychiatry Neurol.* 11;61-70.

Napierała M., Michałowski D., de Mezer M., Krzyżosiak W. J. (2005). Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.* 33(2);451-63.

Napierała M., Krzyżosiak W. J. (1997). CUG repeats present in myotonin kinase RNA form metastable „slippery” hairpins. *J Biol Chem.* 272;31079-085.

Oberle I., Rousseau F., Heitz D., Kretz C., Devys D., Hanauer A., Boue J., Bertheas M. F., Mandel J. L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 252;1097 – 1102.

Oostra B. A., Willemsen R. (2002). The X chromosome and fragile X mental retardation. *Cytogenet. Genome Res.* 99;257 – 264.

Pan H., Hsu R.-J., Li C.-Y. Hsiao K.-M. Long tract of untranslated CAG repeats is pathogenic in transgenic mice. Poster AJHG (2005)

Panov A. V., Gutekunst C. A., Leavitt B. R., Hayden M. R., Burke J. R., Strittmatter W. J., Greenamyre J. T. (2002). Early mitochondrial calcium defects in Huntington's disease are a direct effect of polyglutamines. *Nature Neurosci.* 5;731 – 736.

Paulson H.L., Fischbeck K.H. (1996). Trinucleotide repeats in neurogenetic disorders. *Annu Rev Neurosci.* 19;79-107.

Pearson C. E., Edamura K. N., Cleary J. D. (2005). Repeat instability; mechanism of dynamic mutations. *Nat. Rev. Genetics.* 6; 729 – 742.

Peel A. L., Rao R. V., Cottrell B. A., Hayden M. R., Ellerby L. M., Bredesen D. E. (2001). Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Hum Mol Genet.* 10(15);1531-8.

Philips A. V., Timchenko L. T., Cooper T. A. (1998). Disruption of splicing regulated by a CUG – binding protein in myotonic dystrophy. *Science* 280;737 – 741.

Potter N. T., Spector E. B., Prior T. W. (2004). Technical standards and guidelines for Huntington disease testing. *Genet Med.* 6(1);61-5.

Primerano B., Tassone F., Hagerman R. J., Hagerman P., Amaldi F., Bagni C. (2002). Reduced FMR1 mRNA translation efficiency in fragile X patients with permutations. *RNA* 8;1482 – 1488.

Ranum L. P., Rasmussen P. F., Benzow K. A., Koob M. D., Day J. W. (1998). Genetic mapping of a second myotonic dystrophy locus. *Nature Genet.* 19;196 – 198.

Ranum L. P., i Day J. W. (2004). Pathogenic RNA repeats; an expanding role in genetic disease. *Trends Genet.* 20;506 – 512.

Ravikumar B., Vacher C., Berger Z., Davies J. E., Luo S., Oroz L. G., Scaravilli F., Easton D. F., Duden R., O'Kane C. J., Rubinsztein D. C. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nat. Genet.* 36;585 – 595.

Richards R. I., Sutherland G. R. (1992). Dynamic mutations; a new class of mutations causing human disease. *Cell.* 70;709-12.

Rigamonti D., Bauer J. H., De-Fraja C., Conti L., Sipione S., Sciorati C., Clementi E., Hackam A., Hayden M. R., Li Y., Cooper J. K., Ross C. A., Govoni S., Vincenz C., Cattaneo E. (2000). Wild – type huntingtin protect from apoptosis upstream of caspase – 3. *J. Neurosci.* 20;3705 – 3713.

Sakahira H., Breuer P., Hayer-Hartl M. K., Hartl F. U. (2002). Molecular chaperones as modulators of polyglutamine protein aggregation and toxicity. *Proc. Natl. Acad. Sci. USA.* 99 (suppl 4);16412 – 16418.

Santoro SW, Joyce GF. (1998). Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry.* 37;13330-42.

Savkur R. S., Philips A. V., Cooper T. A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nature Genet.* 29;40 – 47.

Schmid C.W., Jelinek W.R. (1982). The Alu family of dispersed repetitive sequences. *Science.* 216;1065-70.

Seznec H., Simon D., Bouton C., Reutenauer L., Hertzog A., Golik P., Procaccio V., Patel M., Drapier J. C., Koenig M., Puccio H. (2005). Friedreich ataxia; the oxidative stress paradox. *Hum. Mol. Genet.* 14;463 – 474.

Shibata H., Huynh D. P., Pulst S. M. (2000). A novel protein with RNA – binding motifs interacts with ataxin – 2. *Hum. Mol. Genet.* 9;1303 – 1313.

Simpson S. A., Davidson M. J., Barron L. H. (1993). Huntington's disease in Grampian region; correlation of the CAG repeat number and the age of onset of the disease. *J Med Genet.* 30;1014-7.

Singer M. F., Krek V., McMillan J. P., Swergold G. D., Thayer R. E. (1993). LINE-1; a human transposable element. *Gene.* 135;183-8.

Sobczak K, Krzyżosiak W. J. (2005). CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J Biol Chem.* 280(5);3898-910. Epub 2004 Nov 8.

Sobczak K., Krzyżosiak W. J. (2004) Imperfect CAG repeats form diverse structures in SCA1 transcripts. *J Biol Chem.* 279(40);41563-72.

Sobczak K., de Mezer M., Michlewski G., Król J., Krzyżosiak W. J. (2003). RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* 31,5469-5482.

Soong B. W., Wang J. T. (1995). A comparison of the Huntington's disease associated trinucleotide repeat between Chinese and white populations. *J Med Genet.* 32;404-5.

Stallings R. L. (1994). Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence; implications for human genetic diseases. *Genomics.* 21;116-21.

Steffan J. S., Kazantsev A., Spasic-Boskovic O., Greenwald M., Zhu Y. Z., Gohler H., Wanker E. E., Bates G. P., Housman D. E., Thompson L. M. (2000). The Huntington's disease protein interacts with p53 and CREB – binding protein and represses transcription. *Proc. Natl. Acad. Sci USA* 97;6763 – 6768.

Steffan J. S., Agrawal N., Pallos J., Rockabrand E., Trotman L. C., Slepko N., Illes K., Lukacsovich T., Zhu Y. Z., Cattaneo E., Pandolfi P. P., Thompson L. M., Marsh J. L. (2004). SUMO modification of huntingtin and Huntington disease pathology. *Science* 304;100 – 104.

Sugars K. L., Rubinsztein D. C. (2003). Transcriptional abnormalities in Huntington disease. *Trends Genet.* 19;233 – 238.

Szebenyi G., Morfini G. A., Babcock A., Gould M., Selkoe K., Stenoien D. L., Young M., Faber P. W., MacDonald M. E., McPhaul M. J., Brady S. T. (2003). Neuropathogenic forms of huntingtin and androgen receptor inhibit fast axonal transport. *Neuron* 40;41 – 52.

Tang T. S., Slow E., Lupu V., Stavrovskaya I. G., Sugimori M., Llinas R., Kristal B. S., Hayden M. R., Bezprozvanny I. (2005). Disturbed Ca⁺ signaling and apoptosis of medium spiny neurons in Huntington's disease. *Proc. Natl. Acad. Sci. USA* 102;2602 – 2607.

Tapscott S. J., Thornton C. A. (2001) Biomedicine. Reconstructing myotonic dystrophy. *Science.* 293(5531);816-7.

Tassone F., Hagerman R. J., Taylor A. K., Gane L. W., Godfrey T. E., 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.

Tassone F., Hagerman R. J., Chamberlain W. D., Hagerman P. J. (2000). Transcription of the *FMR1* gene in individuals with fragile X syndrome. *Am. J. Med. Genet.* 97;195 – 203.

Tassone F., Hagerman R. J., Garcia – Arocena D., Khandjian E. W., Greco C. M., Hagerman P. J. (2004). Intranuclear inclusions in neural cells with permutation alleles in fragile X associated tremor / ataxia syndrome. *J. Med. Genet.* 41;e43

Tassone F., Iwahashi C., Hagerman P. J. (2004). *FMR1* RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biology* 1;103 – 105.

The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*. 72;971-83.

Tian B., White R. J., Xia T., Welle S., Turner D. H., Mathews M. B., Thornton C. A. (2000). Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*. 6(1);79-87.

Timchenko L.T. (1999). Myotonic dystrophy; the role of RNA CUG triplet repeats. *Am J Hum Genet*. 64;360-4.

Timchenko N. A, Iakova P., Cai Z. J., Smith J. R., Timchenko L. T. (2001). Molecular basis for impaired muscle differentiation in myotonic dystrophy. *Mol Cell Biol*. 21(20);6927-38.

Timchenko N. A., Cai Z. J., Welm A. L., Reddy S., Ashizawa T., Timchenko L.T. (2001). RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem*. 276;7820 – 7826.

Venkatraman P., Wetzel R., Tanaka M., Nukina N., Goldberg A. L. (2004). Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine – containing proteins. *Mol. Cell* 14;95 – 104.

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., Gocayne J. D., The sequence of human genome. (2001) *Science* 291, 1304-1351

Weber J. L., Wong C. (1993). Mutation of human short tandem repeats. *Hum Mol Genet*. 2;1123-8.

Weir B. S. (1992). Independence of VNTR alleles defined as floating bins. *Am J Hum Genet*. 51;992-7.

Wellington C. L., Ellerby L. M., Gutekunst C. A., Rogers D., Warby S., Graham R. K., Loubser O., van Raamsdonk J., Singaraja R., Yang Y. Z., Gafni J., Bredesen D., Hersch S. M., Leavitt B. R., Roy S., Nicholson D. W., Hayden M. R. (2002) Caspase

cleavage of mutant huntingtin precedes neurodegeneration in Huntington's disease. *J Neurosci.* 22(18);7862-72.

Xu K., Bogert B. A., Li W., Su K., Lee A., Gao F. B. (2004). The fragile X – related gene affects the crawling behavior of *Drosophila* larvae the mRNA level of the DEG/EnaC protein pickpocke1. *Curr. Biol.* 14;1025 – 1034.

Yeap B. B., Wilce J. A., Leedman P. J. (2004) The androgen receptor mRNA. *Bioessays.* 26(6);672-82.

Yen L., Strittmatter S. M., Kalb R. G. (1999). Sequence-specific cleavage of Huntingtin mRNA by catalytic DNA. *Ann Neurol.* 46;366-73.

Zhang Y. Q., Bailey A. M., Matthies H. J., Renden R. B., Smith M. A., Speese S. D., Rubin G. M., Broadie K. (2001). *Drosophila* fragile X – related gene regulates the *MAB1B* homolog *Futsch* to control synaptic structure and function. *Cell.* 107;591 – 603.

Zoghbi H. Y., Orr H.T. (2000). Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23;217 – 247.

Zoghbi H.Y., Orr H.T. (1999). Polyglutamine diseases; protein cleavage and aggregation. *Curr Opin Neurobiol.* 9;566-70.

Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13);3406-15.