

STRUKTURALNE PODSTAWY PATOGENEZY CHORÓB NEUROLOGICZNYCH WYWOŁYWANYCH EKSPANSJĄ TRÓJNUKLEOTYDOWYCH POWTÓRZEŃ I ICH TERAPII CELOWANEJ W RNA

Włodzimierz Krzyżosiak
Instytut Chemii Bioorganicznej PAN

WPROWADZENIE

Zakończone niedawno badania sekwencji ludzkiego genomu znacznie przybliżyły perspektywę poznania normalnych funkcji genów jak również ich roli w patogenezie tysięcy znanych chorób. Stworzyły też podstawę do opracowania nowych metod ich terapii wycelowanej specyficznie w produkty wadliwie funkcjonujących genów. W ten sposób osiągnięty został pierwszy etap na drodze do realizacji wizji nowoczesnej genetycznej medycyny.

W pierwszej połowie lat 90-tych śledząc rozwój badań nad ludzkim genomem wybrałem kierunki własnych badań dotyczących związków między genami i chorobami. Obiektem zainteresowań mojego zespołu stało się między innymi kilka chorób, których wspólną cechą było ich bardzo interesujące podłoże genetyczne - ekspansja powtarzającego się tandemowo krótkiego trójnukleotydowego motywu sekwencji. Był to nieznan wcześniej typ mutacji odkryty dopiero w 1991 roku. W kolejnych latach podobne podłoże przypisano kilkunastu dalszym chorobom genetycznym, głównie neurologicznym.

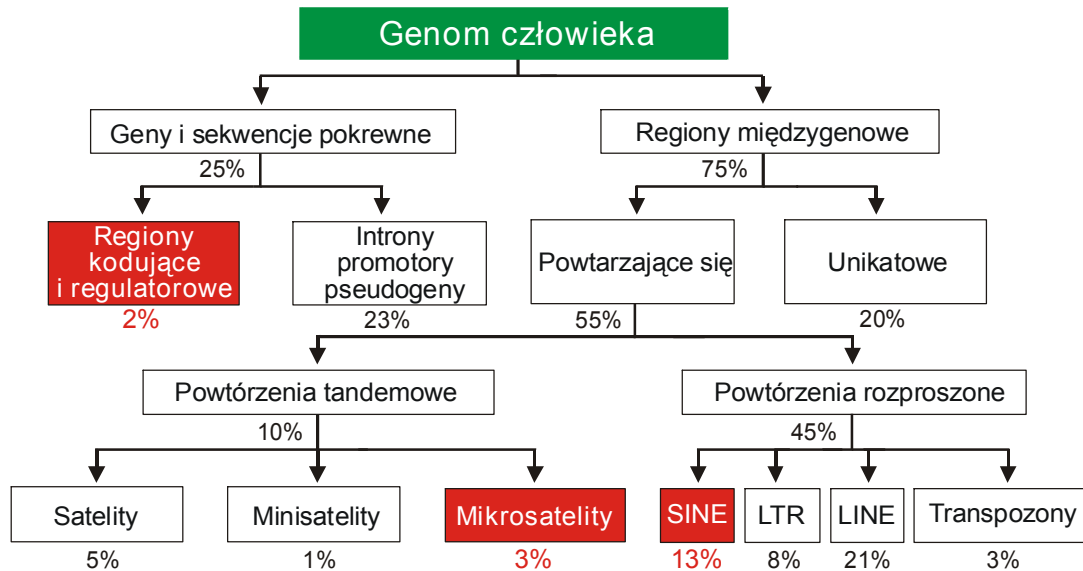
Nasze badania w tej dziedzinie podały w całości nowym kierunkowi nie będącym wówczas przedmiotem zainteresowań innych badaczy. Postawiliśmy hipotezę, że w patogenezie wielu z tych chorób znaczącą, a w niektórych przypadkach dominującą rolę może odgrywać struktura RNA. Gromadziliśmy dowody na poparcie tej hipotezy badając strukturę regionów powtórzeń trójnukleotydowych w transkryptach genów związanych z tymi chorobami. Okazało się, że struktury te dobrze tłumaczą obserwowane zależności między wielkością ekspansji sekwencji powtarzającej się a występowaniem objawów choroby.

Wykorzystując metody bioinformatyki, genomiki strukturalnej i funkcjonalnej poznaliśmy również komórkową scenerię, w której zmutowany transkrypt może w sprzyjających warunkach aktywować proces patogenezy. Mechanizm tego procesu może polegać na konfiskowaniu ważnych regulatorowych białek transkryptom zawierającym powtórzenia normalnej długości przez transkrypt zawierający nadmiernie wydłużone powtórzenia. W ten sposób może dochodzić do zaburzeń metabolizmu wielu cząsteczek RNA w komórce, które skutkują chorobą. Zidentyfikowaliśmy setki ludzkich mRNA zawierających trójnukleotydowe powtórzenia, których funkcja może być upośledzona pod wpływem mutacji. Określiśmy zakres zmienności ich sekwencji powtarzającej się w populacji oraz zbadaliśmy poziom ekspresji tych transkryptów w różnych tkankach.

W ramach badań nad nowymi metodami terapii tych chorób naszym pierwotnym zamysłem było zastosowanie antysensowych oligonukleotydów i rybozymów. Obecnie wykorzystujemy w tym celu zjawisko interferencji RNA. W wyszukiwaniu skutecznych reagentów RNAi kierujemy się wiedzą o przebiegu naturalnego procesu biogenezy mikroRNA. Specyficzna eliminacja produktu zmutowanego genu wywołującego neurodegenerację w obecności produktu genu normalnego i ukazanie potencjału terapeutycznego tego podejścia jest dla nas obecnie dużym wyzwaniem.

POWTÓRZENIA KRÓTKICH MOTYWÓW SEKWENCJI W LUDZKIM GENOMIE

Genom człowieka w ponad połowie składa się z sekwencji powtarzających się. Należą do nich między innymi tandemowe powtórzenia, które stanowią 10% całkowitej sekwencji DNA. Ta grupa obejmuje satelity, minisatelity oraz mikrosatelity, różniące się wielkością powtarzającego się motywu sekwencji. Znacznie mniejszą część genomu, bo tylko 1% zajmują sekwencje kodujące około 22 tysięcy ludzkich genów i drugie tyle stanowią ich sekwencje regulatorowe. Sekwencje powtarzające się występują nie tylko w częściach międzygenowych lecz także w genach, nawet w ich częściach kodujących i mogą brać udział w regulacji ich ekspresji.



Mikrosatelity znane również jako krótkie tandemowe powtórzenia (ang. Short Tandem Repeats (STR) albo powtórzenia prostych sekwencji (ang. Simple Sequence Repeats (SSR) składają się z powtarzających się 1-6 nukleotydowych motywów. Liczba powtórzeń takiej sekwencji zwykle nie przekracza 10, a jeśli jest wyższa to często taki ciąg charakteryzuje się polimorfizmem długości w populacji. W rzadkich przypadkach, najczęściej patologicznych, liczba powtórzeń może osiągać setki a nawet tysiące.

Wszystkie sekwencje typu STR obejmują łącznie 3% ludzkiego genomu. Zarówno zawartość jak i rozmieszczenie różnych powtórzeń tego typu w genomie są silnie zróżnicowane. Większość sekwencji STR występuje znacznie częściej w regionach niekodujących. Wyjątkiem są tutaj ciągi powtórzeń trój- i sześci nukleotydowych, które są niemal dwukrotnie częstsze w eksonach niż w intronach i regionach międzygenowych. Ich pozytywna selekcja w eksonach sugeruje, że mogą one mieć znaczenie funkcjonalne. Wysoka podatność ciągów prostych powtórzeń sekwencji na mutacje i wynikający z niej częsty polimorfizm długości czynią z sekwencji powtarzających się bogate źródło zmienności fenotypowej.

TRÓJNUKLEOTYDOWE POWTÓRZENIA W TRANSKRYPTOMIE CZŁOWIEKA

Koncentrując się na trójnukleotydowych powtórzeniach w ludzkim transkryptomie zadaliśmy szereg pytań: ile różnych mRNA zawiera takie powtórzenia w swojej sekwencji? Jaka jest częstość różnych typów trójnukleotydowych powtórzeń? Jaka jest dystrybucja ciągów powtórzeń w różnych regionach funkcjonalnych mRNA? Jaki

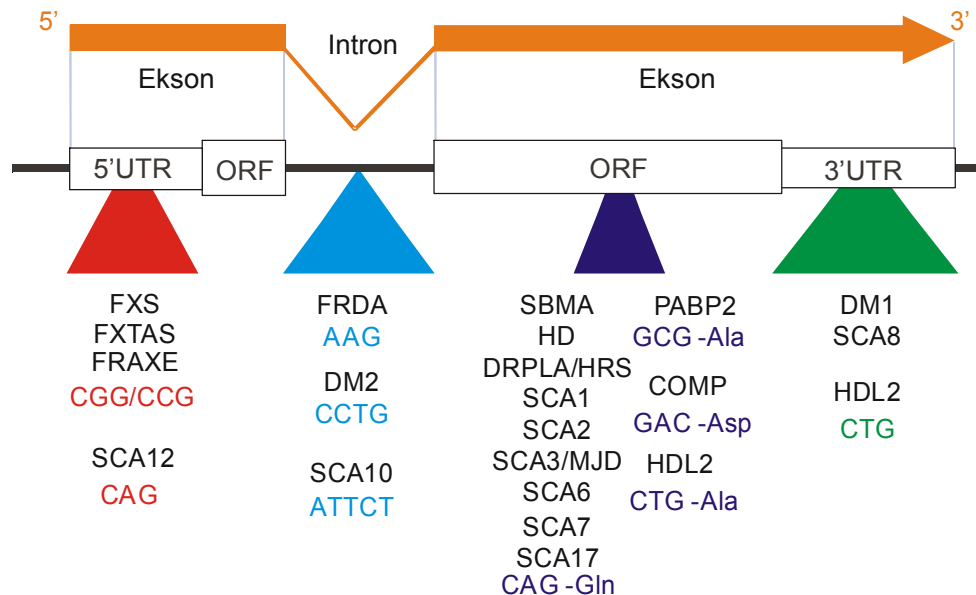
jest zakres zmienności sekwencji powtarzającej się w populacji? Jaki jest poziom ekspresji cząsteczek mRNA zawierających ciągi powtórzeń w różnych tkankach? Aby odpowiedzieć na te pytania, przeszukaliśmy najpierw genomową bazę danych GenBank pod kątem występowania transkryptów zawierających ciągi 20 różnych motywów trójnukleotydomowych składające się z co najmniej 6 powtórzeń danego motywu. W ten sposób znaleźliśmy 718 ciągów trójnukleotydomowych powtórzeń występujących w 619 różnych mRNA. Większość spośród zidentyfikowanych transkryptów (86%) zawierała krótkie ciągi powtórzeń składające się z 6-10 powtórzonych jednostek, transkrypty zawierające 11-20 powtórzonych jednostek stanowiły 11%, a te w których występowały ciągi zawierające ponad 20 powtórzeń tego samego motywu sekwencji - tylko 2%. Najczęściej występującymi motywami powtórzonymi okazały się: CAG, CGG, CCG, CUG, AGG i ACC, natomiast ACG, AUC, CUU, AGU, CGU i ACU należą do najrzadszych w transkryptomie. Większość ciągów powtórzeń występuje w ORF (ang. Open Reading Frame) (67%), mniej w 5'UTR (ang. Untranslated Region) (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.

Badając zakres długości różnych ciągów trójnukleotydomowych powtórzeń w ludzkich genach w normalnej populacji zadaliśmy następujące pytania: Jakie czynniki decydują o tym, że niektóre ciągi powtórzeń są polimorficzne a inne nie? Jaka jest częstość zaburzeń regularności sekwencji obrębie tandemowych powtórzeń? Jakiego typu zaburzenia występują najczęściej i w jakich regionach? Aby odpowiedzieć na te pytania zbadaliśmy ciągi powtórzeń w ponad 100 genach zawierających najczęstsze powtórzenia CAG i CTG. Zaobserwowaliśmy, że głównym czynnikiem wpływającym na polimorfizm długości jest liczba niezaburzonych trójnukleotydomowych powtórzeń w najczęstszym w populacji allelicznym wariacie genu. Odnotowaliśmy również, że najczęstszymi zaburzeniami występującymi w ciągach CAG kodujących glutaminę są tryplety CAA kodujące ten sam aminokwas. Na tej podstawie stwierdziliśmy, że presja selekcyjna działa w większym stopniu na geny i transkrypty zawierające długie niezaburzone ciągi powtórzeń niż na białka zawierające długie ciągi zbudowane z tego samego aminokwasu – glutaminy.

W celu lepszego poznania transkryptów zawierających trójnukleotydomowe powtórzenia, poza określeniem ich dystrybucji i osobniczej zmienności, dokonaliśmy charakterystyki ekspresji dużej grupy tych cząsteczek w wielu normalnych ludzkich tkankach. Przeprowadziliśmy analizę bioinformatyczną dostępnych w sieci danych pochodzących z zastosowań metod genomiki funkcjonalnej: sekwencjonowania EST, SAGE i mikromacierzy oligonukleotydomowych wysokiej gęstości. Dane EST okazały się najbardziej odpowiednie do określania poziomów ekspresji różnych transkryptów w każdej z badanych tkanek, natomiast dane mikromacierzowe do porównań poziomów poszczególnych transkryptów między tkankami. Zidentyfikowaliśmy te mRNA, które wnoszą największy wkład w obserwowane poziomy transkryptów zawierających różne typy powtórzeń. Szerszą analizę przeprowadziliśmy dla mRNA zawierających powtórzenia CUG oraz mRNA kodujących białka wiążące się do tych powtórzeń w transkryptach. Wyniki tych badań znacznie poszerzają wiedzę na temat transkryptów, które mogą brać udział w postulowanym mechanizmie patogenezy na poziomie RNA w chorobach człowieka wywoływanych ekspansją trójnukleotydomowych powtórzeń.

ROLA RNA W PATOGENEZIE DZIEDZICZNYCH CHOROÓB NEUROLOGICZNYCH WYWOŁYWANYCH EKSPANSJĄ POWTÓRZEŃ TRÓJNUKLEOTYDOWYCH

Ekspansja niestabilnych ciągów powtórzeń w pojedynczych genach jest przyczyną ponad 20 różnych chorób dziedzicznych znanych pod wspólną angielską nazwą Triplet Repeat Expansion Diseases (TREDs). Należą do nich zespół łamliwego chromosomu X (FXS), dystrofia miotoniczna (DM), choroba Huntingtona (HD) i szereg ataksji rdzeniowo-mózdkowych (SCAs). Ciągi powtórzeń ulegające patogenicznym ekspansjom występują we wszystkich regionach funkcjonalnych genów, zarówno w częściach ulegających translacji, jak i w regionach niekodujących.



Większość genów związanych z TREDs zawiera powtórzenia CAG w regionie kodującym i wywołująca chorobę liczba powtórzeń mieści się najczęściej w zakresie od około 40 do 100. Ekspansje powtórzeń w regionach niekodujących są większe i bardziej zmienne. Powszechnie uważa się, że powtórzenia występujące w regionach ulegających translacji wywołują patologię poprzez zmiany funkcji kodowanego białka, które w większości przypadków zawiera wydłużone ciągi poliglutaminowe. Jednak ich toksyczne działanie może nie być jedynym mechanizmem odpowiedzialnym za zaburzenia funkcji neuronów. Transkrypty z wydłużonymi powtórzeniami w regionach niekodujących nie powodują jakichkolwiek zmian jakościowych w białku, ale mogą ulegać wyciszeniu transkrypcyjnemu lub zahamowaniu translacji jak to ma miejsce w przypadku dwóch zespołów chorobowych związanych z genem *FMR1* (FXS i FXTAS). Mogą one również aktywować inne mechanizmy patologii na poziomie RNA, w tym konfiskowanie specyficznych białek przez zmutowane transkrypty, co stanowi główny szlak patologii w dwóch typach dystrofii miotonicznej związanych z genami *DMPK* i *ZNF9*.

Nasze badania dowiodły, że powtórzenia CUG i CCUG, których ekspansja powoduje dystrofię miotoniczną typu 1 i 2 tworzą stabilne struktury typu spinki do włosów, co oznacza że konfiskowane przez nie białka wiążą się z dwuniciowymi regionami tych struktur. Takie białka zostały wyizolowane przez innych badaczy i wykazano ich kolokalizację ze zmutowanymi transkryptami w formie skupisk jądrowych.

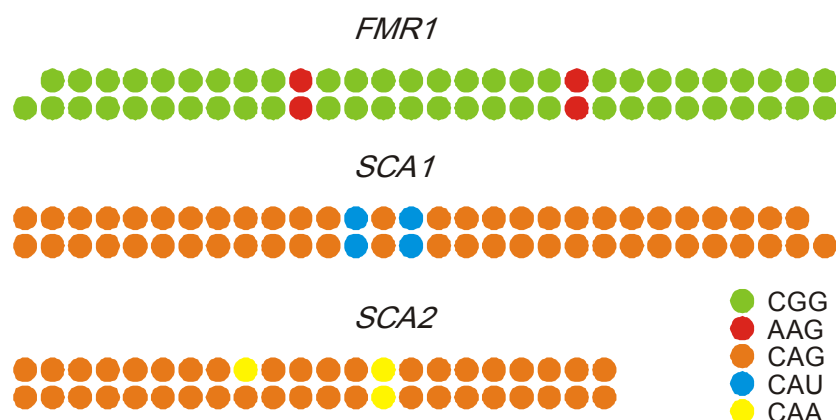
spinki (CGG, CCG, CAG, CUG, CGA i CGU). Wśród nich znalazły się wszystkie powtórzenia typu CNG, które jak wykazały nasze analizy bioinformatyczne są najczęstszym typem powtórzeń w transkryptach. Dwa transkrypty (AGG i UGG) tworzyły stabilne struktury tetrapleksu. Kolejne sześć transkryptów (UAG, AUG, CUA, CAU, UUA i UAA) wykazywało mniejszą tendencję do tworzenia struktur spinki, a pozostałe nie wykazywały skłonności do parowania zasad.

Trzon struktury spinki tworzonej przez powtórzenia CNG składa się z występujących periodycznie par C-G i G-C oraz niesparowań N-N. Jeśli w transkrypcie nie występują oddziałujące ze sobą sekwencje oskrzydłające powtórzenia wtedy sekwencje (CNG)_n wykazują tendencję do przyjmowania kilku wariantowych ułożeń. Cechą struktury spinki tworzonej przez powtórzenia identycznego trypletu CNG jest wzrost jej sztywności wraz z długością. Takie struktury tworzone przez powtórzenia CUG i CAG przyjmują geometrię typową dla formy A-RNA.

Struktury RNA tworzone przez samą sekwencję powtarzającą się oraz te tworzone przez powtórzenia występujące w kontekście sekwencji transkryptu mogą się różnić. Nasze przewidywania struktur szeregu mRNA związanych z TREDs: FMR1, FMR2, SCA1, AR, SCA6, SCA7 i SCA12 pokazały, że sekwencje oskrzydłające region powtórzeń mogą w istotny sposób wpływać na stabilność struktur tworzonych przez powtórzenia. Wyniki tych przewidywań potwierdziliśmy eksperymentalnie, dowodząc że struktury spinki są faktycznie główną formą występowania regionu powtórzeń w transkryptach genów związanych z TREDs. Jeśli mechanizm patogenezы na poziomie RNA dominowałby w komórkach osób chorych, to zróżnicowany efekt wywierany przez sekwencję otaczającą powtórzenia mógłby wyjaśniać dlaczego w chorobach wywoływanych mutacjami w różnych genach istnieją różnice progowej patogenicznej liczby powtórzeń.

ROLA ZABURZEŃ CIĄGÓW POWTÓRZEŃ W GENACH I TRANSKRYPTACH

Normalne warianty trzech genów związanych z TREDs: *SCA1*, *SCA2* i *FMR1*, zawierają specyficzne zaburzenia w ciągach powtórzeń, które nie występują lub rzadko występują w zmutowanych allelach. Przeprowadziliśmy analizę regionów



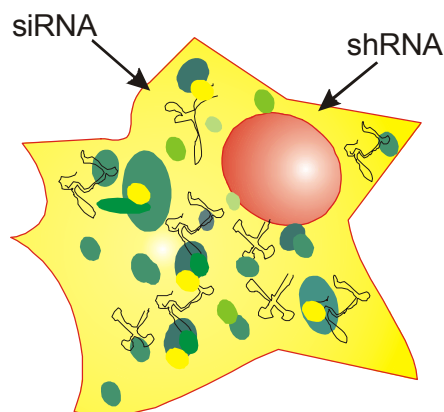
powtórzeń w genach *SCA1* i *SCA2* w populacji polskiej i określiliśmy układy zaburzeń w obrębie ciągów powtórzeń w tych genach. Na tej podstawie zaproponowaliśmy mechanizm ekspansji sekwencji powtarzającej się i związek obecności różnych wariantów polimorficznych genów w różnych populacjach z częstością występowania w nich odpowiednich chorób. Obecność zaburzeń sekwencji powtarzającej się

wydaje się być strategią obronną komórki przed ekspansją powtórzeń mogącą nastąpić w procesie replikacji DNA.

Poznaliśmy też rolę strukturalną różnych typów zaburzeń ciągów powtórzeń w transkryptach stosując chemiczne i biochemiczne sondowanie ich struktur. Nasze analizy wykazały, że rola tych zaburzeń polega na osłabianiu lub skracaniu struktur spinki tworzonych przez czyste powtórzenia, co prawdopodobnie zmniejsza ich zdolność do wiązania białek oddziałujących z regularnymi dwuniciowymi strukturami tworzonymi przez powtórzenia CAG i CGG. Zaburzenia, zależnie od ich liczby i lokalizacji w obrębie powtórzeń, mogą w różny sposób destabilizować strukturę tworzoną przez sekwencję powtórzoną. Substytucja pojedynczego nukleotydu może powodować powiększenie pętli terminalnej, powstawanie pętli wewnętrznych lub wymusza tworzenie się dwóch mniejszych struktur spinkowych. Struktury RNA zaburzonych regionów powtórzeń CAG występujące u niektórych chorych na SCA2 lepiej wyjaśniają ich fenotyp kliniczny niż długości ich ciągów poliglutaminowych. Również struktury RNA zaburzonych ciągów powtórzeń CGG u nosicieli zaburzonych alleli premutacyjnych *FMR1* wydają się dobrze korelować z fenotypem podobnie jak struktury zaburzonych ciągów CAG u chorych na SCA1.

PODEJŚCIA DO TERAPII TREDs Z WYKORZYSTANIEM INTERFERENCJI RNA

Większość chorób wywoływanych ekspansją trójnukleotydowych powtórzeń jest warunkowana w sposób autosomalny dominujący czyli wywołuje je mutacja w jednym z dwóch wariantów genu. Produkt zmutowanego genu, transkrypt lub białko, albo obydwa są toksyczne dla komórki. Zatem usunięcie zmutowanego transkryptu lub białka powinno przeciwdziałać chorobie. Terapia TREDs jest dużym wyzwaniem ze względu na miejsce gdzie rozwijają się główne objawy choroby czyli centralny układ nerwowy oraz konieczność allelo-specyficznego eliminacji produktów zmutowanych genów.



Celem allelo-specyficznego wyciszenia ekspresji genu może być albo sam region powtórzeń, który różni się długością i strukturą RNA między normalnym a zmutowanym wariantem mRNA, albo jak się wydaje bardziej odpowiednie do zastosowania RNAi miejsca występowania polimorfizmu pojedynczych nukleotydów SNP. Znajdowanie skutecznych allelo-specyficznych reagentów RNAi (np. siRNA) jest trudniejsze niż znajdowanie siRNA dla typowych zastosowań RNAi. Podejście takie wydaje się jednak możliwe i jest obecnie sprawdzane w naszym laboratorium.

W wyborze reagentów do RNAi posługujemy się informacjami o biogenezie mikroRNA z ich naturalnych komórkowych prekursorów. W ostatnich latach okazało się że

ponad 1% ludzkich genów koduje grupę mikroRNA, które odgrywają ważną rolę w regulacji ekspresji znacznej części genów. Pierwotne transkrypty genów mikroRNA czyli pri-mikroRNA podlegają przekształceniu w jądrze komórkowym do pre-mikroRNA o długości około 60nt tworzących nieregularne struktury typu spinki. Są one następnie transportowane do cytoplazmy, gdzie ulegają dalszej obróbce przez rybonukleazę Dicer tworząc dojrzałe mikroRNA.

Kierując się mechanizmem obróbki prekursorów mikroRNA i podobieństwem między szlakami biogenezy mikroRNA i RNAi, można planować aktywne dupleksy siRNA. Ustalono, że projektowanie siRNA tak aby przypominały dwuniciowe produkty pośrednie w biogenezie mikroRNA, pozwala uzyskać wysoce efektywne reagenty. Nasz wkład w planowanie efektywnych siRNA polegał na wskazaniu możliwości zwiększenia skuteczności tego procesu. Wykorzystując metody biochemiczne określiliśmy struktury 10 ludzkich prekursorów mikroRNA i 8 spośród tych struktur różniło się od struktur przewidywanych komputerowo. Nasze wyniki pokazały, że przewidywanie funkcjonalnych mikroRNA na podstawie przewidywanych struktur prekursorów może dawać niejednoznaczne, a niekiedy nawet błędne rezultaty. Skuteczność tego procesu można zwiększyć, biorąc pod uwagę eksperymentalnie określone struktury prekursorów mikroRNA. Obecnie badamy kolejne etapy biogenezy mikroRNA i poszukujemy zależności między transkryptami zawierającymi trójnukleotydydowe powtórzenia i komponentami szlaków mikroRNA i RNAi.

WYBRANE PRACE BADAWCZE WSPÓŁFINANSOWANE ZE ŚRODKÓW SUBSYDIUM

Sobczak K., Krzyzosiak W.J.,
RNA structure analysis assisted by capillary electrophoresis.
Nucleic Acids Res. 30, e124, (2002)

Sobczak K., Krzyzosiak W.J.,
Structural determinants of BRCA1 translational regulation.
J. Biol. Chem. 27, 17349-17358, (2002)

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

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

Michlewski G., Krzyzosiak W.J.,
Molecular architecture of CAG repeats in human disease related transcripts.
J. Mol. Biol., 340, 665-679 (2004)

Jasinska A., Krzyzosiak W.J.,
Repetitive sequences that shape the human transcriptome.
FEBS Lett. 567, 136-141 (2004)

Sobczak K., Krzyzosiak W.J.,
Imperfect CAG repeats form diverse structures in SCA1 transcripts.
J. Biol. Chem. 279, 41563-41572 (2004)

Krol J., Sobczak K., Wilczynska U., Drath M., Jasinska A., Kaczynska D., Krzyzosiak W.J.,
Structural features of microRNA precursors and their relevance to miRNA biogenesis and siRNA/shRNA design.
J. Biol. Chem. 279, 42230-42239 (2004)

Sobczak K., Krzyzosiak W.J.,
Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability.
Hum. Mutat. 24, 236-247 (2004)

Sobczak K., Krzyzosiak W.J.,
CAG repeats containing CAA interruptions form branched hairpin structures in SCA2 transcripts.
J. Biol. Chem. 280, 3898-3910 (2005)

Napierala M., Michalowski D., de Mezer M., Krzyzosiak W.J.,
Facile FMR1 mRNA structure regulation by interruptions in CGG repeats.
Nucleic Acids Res., 33, 451-463 (2005)