



Regulatorowe kwasy rybonukleinowe

Marek Żywicki, Maciej Szymański, Mirosława Z. Barciszewska,
Jan Barciszewski

Instytut Chemii Bioorganicznej, Polska Akademia Nauk, Poznań

Regulatory ribonucleic acids

Summary

In recent years there was a growing number of reports of new non-protein-coding RNAs which are implicated in the regulation of many cellular processes. They differ in many respects from already known housekeeping RNA species involved in protein biosynthesis (tRNA, rRNA) and RNA maturation or modification (RNase P RNA, snRNAs, snoRNAs). Regulatory RNAs (riboregulators) are expressed only in certain cell types, at particular stages of organism development or cell differentiation or in response to biotic and abiotic stimuli. Their expression is usually accompanied by the alteration of patterns of the expression of other genes. The mechanisms employed by riboregulators can affect transcription, pre-mRNA processing and translation. In the post-genomic era, the noncoding regulatory RNAs emerge as key determinants of organismal complexity, providing efficient and highly specific means for integration of various cellular processes.

Key words:

gene expression, regulation, regulatory RNA, non-coding RNA.

Adres do korespondencji

Jan Barciszewski,
Instytut Chemii
Bioorganicznej,
Polska Akademia Nauk,
ul. Noskowskiego 12/14,
61-704 Poznań;
e-mail:
jan.barciszewski@ibch.
poznan.pl

1. Wprowadzenie

Od kilku lat pojawia się coraz więcej doniesień o identyfikacji nowych niekodujących RNA. Coraz więcej uwagi przywiązuje się również do znanych już cząsteczek. Zainteresowanie to związane jest z próbą zrozumienia funkcji ncRNA w komórce. Ich ilość, jak się wydaje, jest większa niż ilość transkrybowanego mRNA. Świadczy to o ogromnym znaczeniu tej grupy cząsteczek dla prawidłowego funkcjonowania komórki, zaprzeczając począt-

kowym przypuszczeniom o pochodzeniu ncRNA od zdegenerowanych niefunkcyjnych genów kodujących białka. Różnorodność funkcji przez nie pełnionych potwierdza tę tezę. Obecne we wszystkich stadiach różnicowania komórek, ncRNA okazały się głównymi czynnikami regulującymi podstawowe procesy biochemiczne takie jak transkrypcja czy translacja. Ich szerokie możliwości regulacyjne nie ograniczają się jedynie do wpływu na metabolizm innych RNA, lecz obejmują również udział w modulacji aktywności białek. Do niedawna wydawało się, że ta właściwość jest wyłączną domeną innych białek. Pośród przyczyn zainteresowania tą grupą cząsteczek nie bez znaczenia pozostaje ich wpływ na procesy patologiczne. Okazało się, że mogą być przyczyną licznych poważnych schorzeń, jak syndromy Prader-Willi i Beckwith-Wiedemann czy liczne choroby nowotworowe.

2. Niekodujące RNA

Głównym celem biologii molekularnej jest zrozumienie ogromnej sieci zależności pomiędzy procesami biochemicznymi zachodzącymi w komórce. Największym problemem są skomplikowane powiązania pomiędzy procesami regulacyjnymi. Możliwość dokładnego zsynchronizowania w czasie i w miejscu tysięcy reakcji chemicznych do dziś jest wyzwaniem dla biologów molekularnych. Przez długi czas uważano, że złożoność i doskonałość tego systemu zależy głównie od białek. Od kilkunastu lat coraz częściej dostrzegana jest inna możliwość. Odkrycie niekodujących RNA (ncRNA) udowodniło, że w złożonych procesach regulacyjnych uczestniczą również kwasy nukleinowe. Obecnie niekodujące RNA są, jak się wydaje, nie mniej powszechnie stosowanymi cząsteczkami regulacyjnymi niż białka. Wiedza na ich temat staje się coraz bardziej jednolita. Pod pojęciem „niekodujące RNA” rozumiemy cząsteczki RNA nie kodujące białek, ale posiadające ściśle określone funkcje w komórce (1).

Wszystkie komórkowe RNA można podzielić na dwie główne grupy: 1) kodujące (matrycowe) RNA – mRNA (*messenger RNA*) oraz 2) RNA, które nie stanowią matrycy dla białek (niekodujące RNA) (tab. 1) (2). Ten rodzaj RNA (ncRNA) ze względu na pełnione funkcje można podzielić na konstytutywne oraz regulatorowe. Konstytutywne RNA ulegają ciągłej ekspresji na stałym poziomie i są niezbędne do przebiegu podstawowych procesów komórkowych. Należą do nich tRNA, rRNA, czy snRNA (tab. 1). Regulatorowe RNA odpowiedzialne są najczęściej za utrzymanie oraz zmianę poziomu ekspresji genów. Ich obecność w komórce związana jest z etapem rozwoju lub czynnikami środowiskowymi. Niekodujące RNA wykazują dużą różnorodność mechanizmu działania w obrębie realizowanych funkcji oraz zaangażowanie w licznych procesach. Wszystkie znane obecnie ncRNA znajdują się w bazie danych niekodujących RNA (<http://biobases.ibch.poznan.pl/ncRNA/>) (3).

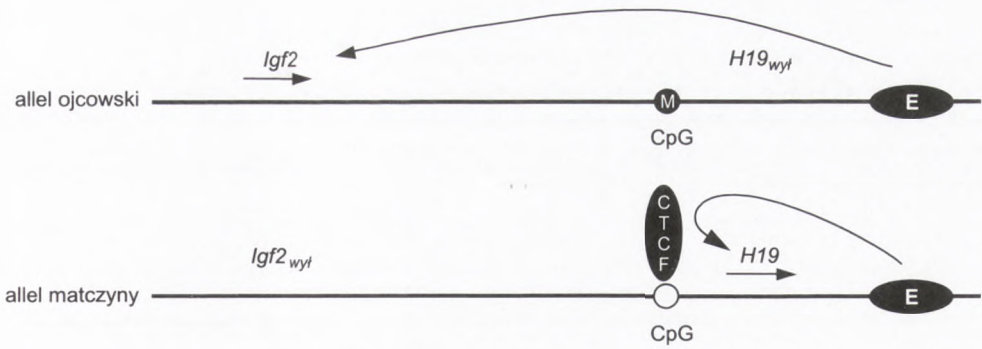
Podział cząsteczek RNA ze względu na właściwości kodujące i funkcjonalne (4)

Kodujące białka	Transkrypty	
	Niekodujące (ncRNA)	
	konstrytutywne RNA	regulatorowe RNA
mRNA	<p>1. Translacja</p> <ul style="list-style-type: none"> – tRNA – translacja informacji genetycznej – rRNA – składnik rybosomu, kataliza formowania wiązań peptydowych – tmRNA – trans-translacja, kontrola jakości translacji <p>2. Dojrzwianie RNA</p> <ul style="list-style-type: none"> – snRNA – splicing pre-mRNA, komponenty spliceosomu – snoRNA – modyfikacje RNA: 2'-O-metylowanie i pseudouracylowanie – RNase MRP – dojrzwianie RNA – RNase P RNA – dojrzwianie 5'-końca pre-tRNA <p>3. Replikacja</p> <ul style="list-style-type: none"> – RNA telomerazy – komponent telomerazy, synteza telomerycznego DNA <p>4. Lokalizacja białek</p> <ul style="list-style-type: none"> – 4.5S RNA, 7SL RNA – translokacja białek przez błony u bakterii i eukariota <p>5. Nieznane funkcje</p> <ul style="list-style-type: none"> – hY RNA 	<p>1. Regulatory transkrypcji</p> <ul style="list-style-type: none"> – regulacja ekspresji genów podlegających imprintingowi (H19, IPW, LIT1) – zmiana struktury chromatyny związana z inaktywacją chromosomu X i kompensacją dawki genów u eukariota (roX, Xist/Tsix) <p>2. Regulatory posttranskrypcyjne</p> <ul style="list-style-type: none"> – interakcje z docelowym mRNA na zasadzie antysensu (DsrA, MicF, lin-4, let-7, microRNAs) <p>3. Modulatory aktywności białek</p> <ul style="list-style-type: none"> – interakcje RNA-białko zmieniające zdolności katalityczne białek (6S RNA, OxyS, SRA RNA) <p>4. Regulatory lokalizacji RNA</p> <ul style="list-style-type: none"> – specyficzny wpływ na lokalizację komórkowych mRNA lub pre-mRNA poprzez specyficzną lokalizację ncRNA (hsr-o, Xlsirt, BC1, BC200)

3. Charakterystyka regulatorowych ncRNA

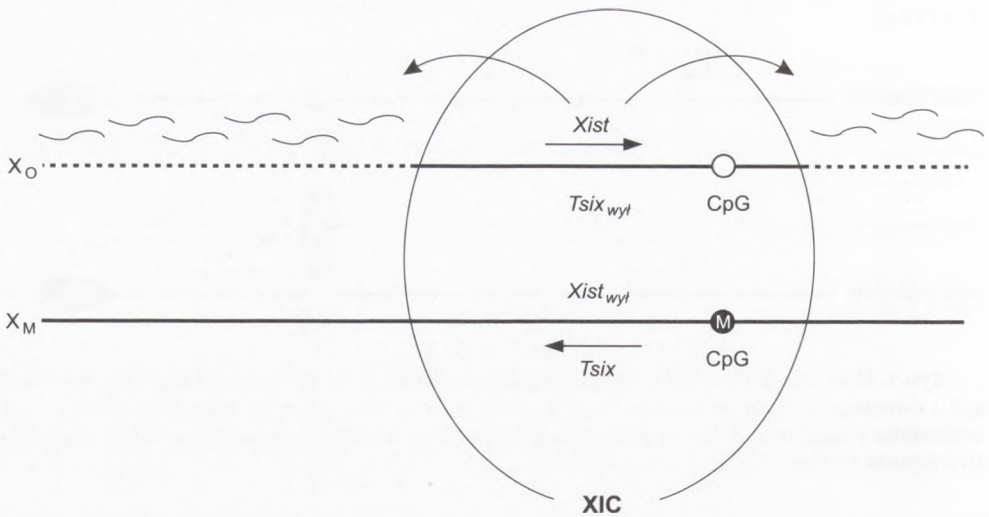
3.1. Regulatory transkrypcji

Do grupy tej należą ncRNA modulujące poziom transkrypcji danego genu. Uczestniczą one w podstawowych szlakach regulacyjnych. Jednym z nich jest udział w specyficznym znakowaniu epigenetycznym (*imprinting*). Dotychczas zidentyfikowano kilka ncRNA uczestniczących w tym procesie, które odgrywają podstawową rolę w genezie złożonych jednostek chorobowych, takich jak syndromy Prader-Willi czy Beckwith-Wiedemann (4,5). Najprawdopodobniej wynikają one z utraty prawidłowego imprintingu (piętna genomowego) określonego segmentu genów. Najle-



Rys. 1. Ekspresja genów *H19* i *Igf2*. *H19* ulega transkrypcji z chromosomu matczyngo, natomiast *Igf2* z chromosomu ojcowskiego. Jest to uzależnione od sekwencji enhancerowych (E) znajdujących się w kierunku 3' względem *H19* oraz od metylacji (M) wysp CpG znajdujących się w kierunku 5' względem *H19* (opis w tekście) (7,8).

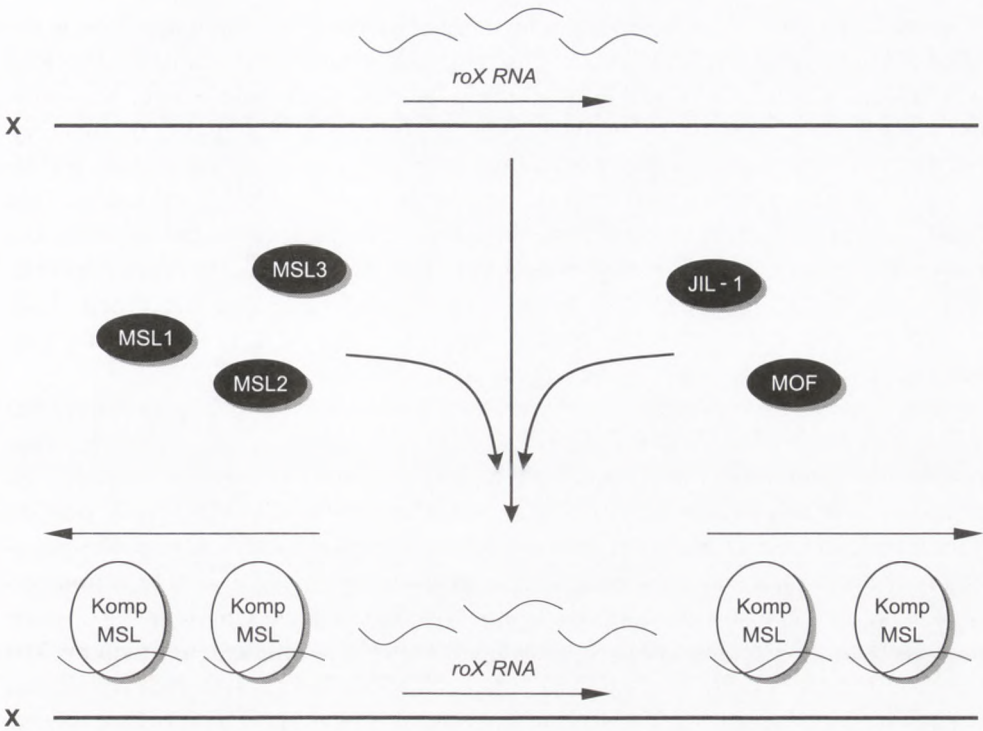
piej poznanym niekodującym RNA mającym wpływ na imprinting innych genów jest *H19* RNA. Ma on długość 2,5 kb. Część ta podlega dojrzewaniu, ulega poliadenylacji i składaniu. Przy końcu 5' posiada m7G-cap. Dojrzały *H19* RNA jest eksportowany z jądra do cytoplazmy. Nie posiada on długiej otwartej ramki odczytu, która umożliwiłaby jego translację do białka. Ekspresja *H19* zachodzi w okresie rozwoju płodowego w tkankach pochodzenia mezo- i ektodermalnego. W okresie późniejszym może ona być przywrócona w tkankach, które wykazywały ekspresję w okresie płodowym i uległy transformacji nowotworowej (6). U człowieka gen *H19* RNA znajduje się w locus 11p15.5 (region homologiczny u myszy występuje na chromosomie 7), około 110-200 kb w kierunku telomeru od genu *IGF2* (*insulin-like growth factor II*), kodującego białko odpowiedzialne za wzrost i podziały komórek oraz regulację ich metabolizmu (7,8). Zaburzenia w syntezie *IGF2* są podstawą procesów nowotworowych (9,10). Pomiędzy poziomem transkrypcji obu genów występuje ścisła zależność (11,12). Geny *H19* i *IGF2* ulegają przeciwnemu imprintingowi: *H19* ulega ekspresji z allelu matczyngo, natomiast *IGF2* z allelu ojcowskiego. Elementami odpowiedzialnymi za ten proces są sekwencje wzmacniające (*enhancer*), oraz region kontroli imprintingu (*imprinting control region, ICR*) znajdujące się w pobliżu *H19* (rys. 1). Liczne wyspy CpG zawarte w ICR podlegają zróżnicowanej metylacji, co stanowi sygnał do włączenia lub wyłączenia ekspresji danego genu. Obniżony poziom metylacji ICR umożliwia przyłączenie się w tym regionie czynnika transkrypcyjnego CTCF będącego swoistym izolotorem (granicą) dla sekwencji wzmacniających. Wynikiem tego są oddziaływania sekwencji wzmacniających na pobliski region promotorowy genu *H19* i jego ekspresja. Natomiast hipermetylacja regionu ICR ogranicza wiązanie CTCF, co powoduje oddziaływanie sekwencji wzmacniających z odległym promotorem genu *IGF2* i jego transkrypcją (13). Mechanizm ten jest charakterystyczny dla komórek niezróżnicowanych. Delecja ICR w komórkach już róż-



Rys. 2. Regulacja ekspresji genów *Xist* i *Tsix* i ich wpływ na inaktywację chromosomu X. Brak metylacji wysp CpG umożliwia transkrypcję *Xist*, który rozciąga się wzdłuż chromosomu X i powoduje jego inaktywację. Metylacja (M) powoduje aktywację genu *Tsix* i inaktywację *Xist*, co prowadzi do pozostawienia chromosomu aktywnym. X_o – chromosom ojcowski, X_M – chromosom matczyzny, linia przerywana wskazuje na brak transkrypcji (7).

nicowanych nie warunkuje ekspresji *H19*. Dlatego postuluje się istnienie drugiego mechanizmu kontroli imprintingu, który ujawnia się dopiero podczas rozwoju zarodkowego. Pod koniec rozwoju płodowego region promotorowy *H19* ulega konstytutywnej hipermetylacji, uniemożliwiając ekspresję genu nawet przy braku metylacji ICR (13). Jest to dodatkowy, jak się wydaje, mechanizm zabezpieczający przed ewentualną aktywacją ekspresji wyciszzonego genu. Zaburzenie tego mechanizmu obserwować można w procesach nowotworowych, w których przywrócona zostaje ekspresja (19).

Przeciwny imprinting genów *H19* i *IGF2* jest tylko jednym z elementów utrzymujących równowagę ilościową pomiędzy transkrybowanymi z nich RNA. Pokazano, że wprowadzenie do komórki dodatkowych kopii *H19* powoduje spadek poziomu zarówno mRNA *IGF2*, jak i samego białka w komórce (11), co sugeruje bezpośrednią wzajemną regulację poziomu transkrypcji. Stwierdzono, że *H19* ma trzy miejsca wiązania białek IMP (*IGF2 mRNA binding protein*), jak również, że mRNA *IGF2*, *H19* i białka IMP znajdują się w tej samej przestrzeni subkomórkowej (12). Białka IMP wiążą się do mRNA *IGF2* przy końcu 5' nie ulegającym transkrypcji i tym samym modulują syntezę tego białka podczas rozwoju. Możliwość wiązania białek IMP do *H19* RNA sugeruje, że może on mieć wpływ na eksport mRNA *IGF2* z jądra, jego lokalizację cytoplazmatyczną oraz translację (12). Szczegółowy mechanizm tego procesu nie jest jednak jeszcze poznany.



Rys. 3. Kompensacja dawki genów z chromosomu X u *Drosophila*. Białka MSL, MOF i JIL-1 tworzą kompleks MSL w miejscu ekspresji genu *roX*. MSL wiąże się do chromosomu powodując wzmożenie ekspresji genów z chromosomu X.

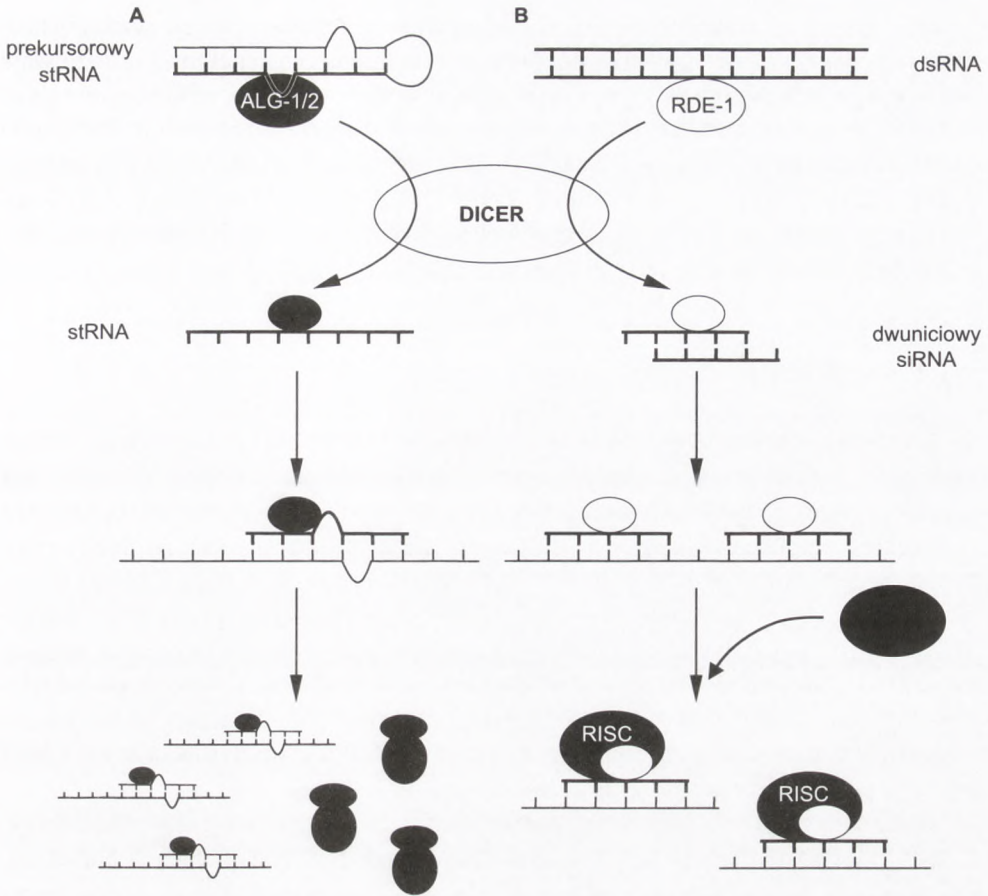
Innym procesem komórkowym, w którym biorą udział ncRNA należące do grupy regulatorów transkrypcyjnych jest kompensacja dawki genów znajdujących się na chromosomie X. Proces ten może odbywać się poprzez wyłączenie jednego z chromosomów u płci posiadającej dwa chromosomy X lub poprzez podwojenie poziomu ekspresji u płci posiadającej jeden chromosom X. Pierwszy z tych mechanizmów charakterystyczny jest dla ssaków, gdzie płeć zdeterminowana jest obecnością lub brakiem chromosomu Y. W przypadku jego braku (u samic) uruchomiona zostaje kaskada sygnałów prowadząca do inaktywacji jednego z chromosomów X, pozostającego w jądrze w postaci skondensowanej i widocznego pod mikroskopem jako ciało Barra (14). Za proces inaktywacji odpowiedzialne jest centrum inaktywacji chromosomu X (Xic) – region chromosomu X, którego delekcja uniemożliwia przebieg całego procesu. Obok licznych genów znajduje się tam również *Xist* (*XIST* u człowieka), którego produktem jest RNA (15-17 kb), podlegający składaniu i poliadenyacji (15). Część ta pełni podstawową rolę podczas inicjacji procesu inaktywacji. Późniejszy brak *Xist* RNA nie powoduje ponownej aktywacji wyciszonego chromosomu. U myszy wybór chromosomu X do wyciszenia odbywa się poprzez im-

printing będący wynikiem zróżnicowanej metylacji wysp CpG znajdujących się w pobliżu 3' końca genu *Xist* RNA (rys. 2). Efektem takiego stanu jest ekspresja *Xist* RNA wyłącznie z chromosomu ojcowskiego, co powoduje jego inaktywację. Na etapie implantacji zarodka w macicy metylacja zostaje zniesiona. Następuje ponowny wybór chromosomu mającego ulec wyciszeniu, który tym razem, jak się wydaje, jest losowy. Duże znaczenie w tym procesie przypisuje się *Tsix* RNA (u człowieka *TSIX* RNA) – cząsteczce RNA transkrybowanej z tego samego locus w przeciwnym kierunku zawierającej duży fragment komplementarny do *Xist* RNA. Do czasu implantacji zarodka *Xist* RNA jest transkrybowany wyłącznie z chromosomu matczyńskiego i, jak się wydaje, dodatkowe zabezpieczenie przed jego inaktywacją, która mogłaby mieć miejsce w przypadku „wycieku” ekspresji *Xist* RNA z tego locus.

Inaczej proces ten przebiega u człowieka. Chromosomy X trafiają do zygoty bez żadnych informacji co do ich wyciszenia. Wybór chromosomu X mającego ulec inaktywacji następuje tylko raz i jest, jak się wydaje, losowy. Odmienność mechanizmu związana jest ze zmianami w strukturze ludzkiego genu *TSIX* (16,17). W pobliżu końca 3' nie ma wysp CpG będących determinantami inaktywacji u myszy. Nie posiada on również fragmentu sekwencji komplementarnego do *XIST*. Fakty te w połączeniu z odkrytą niedawno możliwością alternatywnego składania mysiego *Tsix*, w wyniku którego powstają izoformy o skróconym końcu 5', komplementarnym do *Xist* (18), zmuszają do rewizji poglądów na temat roli *TSIX* i *Tsix* w procesach wyciszenia i akumulacji *XIST/Xist* RNA (19-21). Obecnie proponowane są cztery możliwe mechanizmy działania *TSIX/Tsix*: 1) maskowanie polegające na wychwytywaniu *XIST/Xist* RNA przez antysensowy *TSIX/Tsix*, 2) interferencja polegająca na możliwości blokowania ekspresji *XIST/Xist* przez transkrypcję *TSIX/Tsix* z tego samego locus, 3) represja polegająca na uzależnionym od transkrypcji *TSIX/Tsix* wiązaniu w pobliżu 3'-końca *XIST/Xist*, czynnika blokującego jego transkrypcję, 4) otwarcie chromatyny polegające na jej rozluźnieniu poprzez transkrypcję *TSIX/Tsix*, co umożliwiłoby asocjację dodatkowych czynników regulacyjnych (19).

Mechanizm działania *Xist* RNA nie jest do końca poznany. Wiadomo jednak, że wiąże się on do chromosomu X z którego uległ ekspresji na całej jego długości i pełni główną rolę w przebiegu procesu inaktywacji. Nieaktywny chromosom X charakteryzują typowe cechy konstytutywnej heterochromatyny: hipocetylacja histonów, hipermetylacja wysp CpG i późna replikacja. Nie jest jednak do końca wyjaśniona rola *Xist* w powstawaniu tych zmian. Postuluje się możliwość wiązania do *Xist* białek zdolnych do wprowadzania stałych zmian w strukturze chromatyny, co tłumaczyłoby silne ograniczenie roli *Xist* RNA do etapu inicjacji (19). Do tej pory nie udało się jednak zidentyfikować białek zdolnych do asocjacji z *Xist*.

Funkcjonowanie niekodujących RNA należących do regulatorów transkrypcyjnych nie sprowadza się jedynie do możliwości obniżania poziomu transkrypcji. Mogą one również odpowiadać za jej dodatkową aktywację. Tak jest w przypadku procesu kompensacji dawki genów z chromosomu X u *Drosophila*. Podobnie jak u ssaków, samce posiadają po jednym chromosomie X i Y, natomiast samice dwa



Rys. 4. Geniza i funkcje stRNA (A) i RNAi (B). Elementem wspólnym obu ścieżek jest enzym DICER hydrolizujący pre-RNA odpowiednio do stRNA i siRNA. Białka odpowiadające za skierowanie danego RNA na drogę stRNA zaznaczone są na czarno (ALG-1/2), natomiast na drogę RNAi na biało (RDE-1) (25).

chromosomy X. Determinacja płci zależy jednak nie od obecności chromosomu Y, lecz od stosunku liczby chromosomów X do liczby autosomów (15). Kompensacja dawki następuje poprzez podwojenie ekspresji pojedynczego chromosomu X u samców. Główną rolę w tym procesie pełnią dwa ncRNA: roX1 i roX2. Obie cząsteczki podlegają dojrzewaniu i poliadenylacji. Ich długość wynosi odpowiednio 3,7 i 0,6 kb. Są one odpowiedzialne za poprawne składanie kompleksu MSL (*male specific lethal*) (22). W skład kompleksu – oprócz wymienionych cząsteczek ncRNA – wchodzi białka MSL, acetylotransferaza histonu H4 (MOF) oraz fosforylaza histonu H3 (JIL-1).

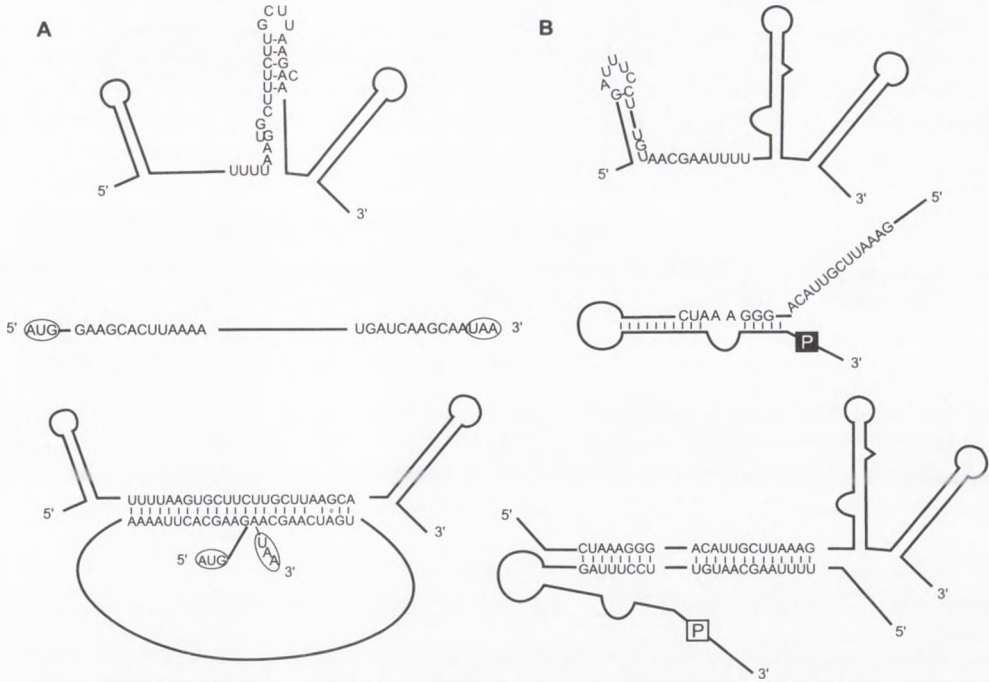
Cząsteczki roX1 i roX2 RNA stanowią swoistą etykietę chromosomu X mającego ulec nadekspresji. Składanie kompleksu MSL zachodzi w miejscu ich transkrypcji na chromosomie X. Miejsca te nazywane są miejscami dostępu do chromatyny (*chroma-*

tin entry sites) (22). Prawidłowo złożone kompleksy przemieszczają się wzdłuż chromosomu. Białka w nich zawarte powodują rozluźnienie chromatyny i umożliwiają wzmożoną ekspresję genów (rys. 3). O głównej roli roX RNA w rozpoznaniu chromosomu świadczy fakt, że insercja genu *roX* do autosomu powoduje związanie białek kompleksu MSL i podwyższenie poziomu ekspresji genów w pobliżu miejsca insercji (23). Do tej pory nie wiadomo, czy roX RNA oprócz identyfikacji chromosomu uczestniczą w procesie rozluźniania chromatyny lub migracji kompleksów MSL wzdłuż chromosomu.

3.2. Regulatory posttranskrypcyjne

Cząsteczki RNA należące do grupy regulatorów posttranskrypcyjnych (w odróżnieniu do regulatorów transkrypcyjnych) odpowiadają za regulację poziomu już zsyntetyzowanego mRNA. Oddziałują one bezpośrednio z docelowym mRNA poprzez tworzenie kompleksu. Możliwe są dwa rodzaje takich oddziaływań: *cis*, kiedy cząsteczka regulatorowego RNA i docelowa transkrybowane są z tego samego locus, lecz w odwrotnym kierunku, oraz *trans*, kiedy obie cząsteczki pochodzą z odrębnych loci. Dla oddziaływań w układzie *cis* charakterystyczne jest tworzenie długich dupleksów poprzez regulatorowy ncRNA i mRNA. Przykładem takiego typu oddziaływań jest postulowana regulacja ekspresji genu *KvLQT1*, kodującego białko kanału potasowego, przez antysensowy transkrypt *LIT1* (24). Efektem parowania jest silne i bardzo specyficzne ograniczenie translacji mRNA.

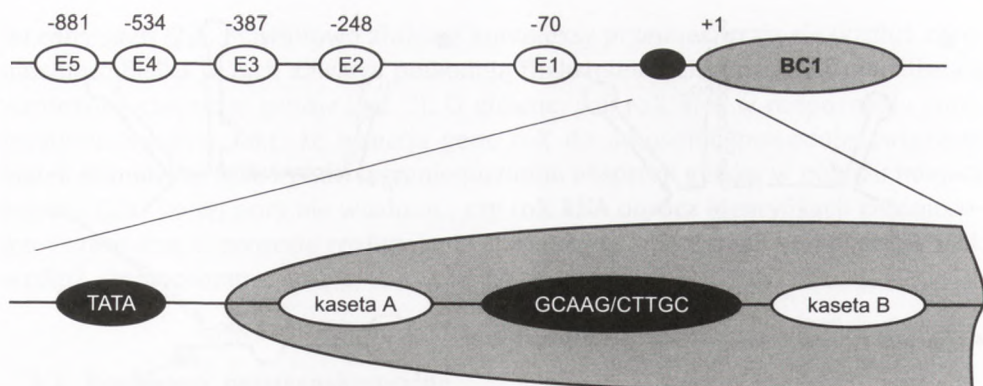
W oddziaływaniach typu *trans* bierze udział zróżnicowana grupa cząsteczek. Charakterystyczne dla nich jest tworzenie dupleksów maksymalną liczbą par zasad, jednak krótkich i nie zawsze dokładnych. Spośród ncRNA należących do tej grupy największą stanowią krótkie (ok. 20 nt) dwuniciowe RNA zwane interferencyjnymi (siRNA, *small interference RNA*). Mechanizm ich działania polega na inicjacji degradacji docelowego RNA za pomocą kompleksu białkowo-nukleinowego RISC (*RNA-induced silencing complex*), którego głównymi składnikami są krótkie dwuniciowe RNA oraz swoista nukleaza. Proces ten nazywany jest RNAi (*RNA interference*). Rola siRNA w tym procesie polega na zapewnieniu wysokiej specyficzności. Cząsteczki siRNA powstają w wyniku hydrolizy długiego fragmentu dwuniciowego prekursorowego RNA przez enzym DICER (25). Następnie zostają do nich przyłączone białka tworząc siRNP, które po aktywacji przekształca się w RISC (rys. 4A). Proces interferencji RNA początkowo wydawał się systemem obronnym skierowanym przeciwko wirusom oraz rozprzestrzeniającym się transpozonom. Wysoka specyficzność degradacji sugeruje jednak, że może mieć on duże znaczenie w procesie regulacji ekspresji genów. Tezę tę popiera wciąż rosnąca liczba identyfikowanych cząsteczek, które potencjalnie mogą funkcjonować wykorzystując mechanizm RNAi (np. niektóre miRNA, *micro RNA*).



Rys. 5. Mechanizm oddziaływania DsrA RNA z docelowym mRNA. A) Wyciszenie mRNA H-NS poprzez wiązanie okolicy promotora i terminatora translacji. B) Aktywacja mRNA RpoS poprzez uwolnienie promotora translacji (P). Sekwencje homologiczne wyróżnione literowo (26).

Duże zainteresowanie RNAi jest związane z możliwościami wykorzystania tego procesu jako nowej techniki badawczej. Umożliwia on kontrolowane okresowe lub ciągłe wyciszenie zdefiniowanych genów, co pozwala na badanie ich funkcji. Jedynym wymogiem stosowania tej metody jest znajomość sekwencji genu mającego ulec wyciszeniu. Na jej podstawie można zsyntetyzować odpowiednie dwuniciowe siRNA, które zostaną wprowadzone do komórki. Tę technologię wykorzystano dotychczas do badania funkcji takich białek jak CD4, CD8, czynników transkrypcyjnych HTF i PSKH1 oraz badań białek podstawowych w cyklu życiowym wirusa HIV, takich jak Tsg101 (26). Możliwe jest również, jak się wydaje, wykorzystanie RNAi do potwierdzenia identyfikacji genów w zsekwencjonowanych genomach oraz badania funkcji zakodowanych w nich białek.

Bardzo podobne do RNAi są cząsteczki stRNA (*small temporal RNA*), np. *lin-4*, *let-7*. Pomimo zbliżonej wielkości i mechanizmu tworzenia, ich oddziaływanie z docelowym mRNA różni się od RNAi. Cząsteczki te tworzą z mRNA dupлексы niedokładne, zawierające błędne pary zasad lub niewielkie wypętlania istotne dla przebiegu procesu wyciszenia (27,28). Główną funkcją stRNA nie jest degradacja mRNA, lecz inhi-



Rys. 6. Organizacja sekwencji regulatorowych genu *BC1*. Pozycje kaset E zaznaczono jako E1-E5. Powiększony fragment przedstawia sekwencję inicjatorową TATA oraz elementy promotora wewnętrzne: kasety A i B oraz zachowawczą sekwencję GCAAG/CTTGC (35).

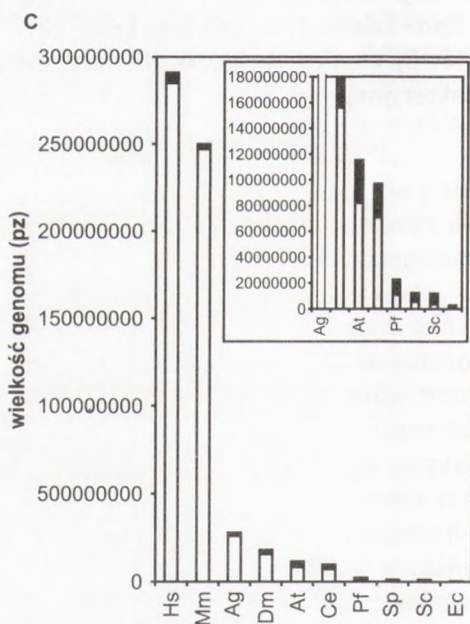
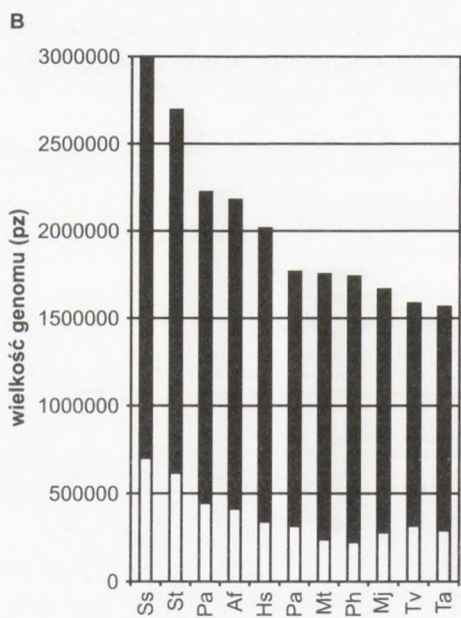
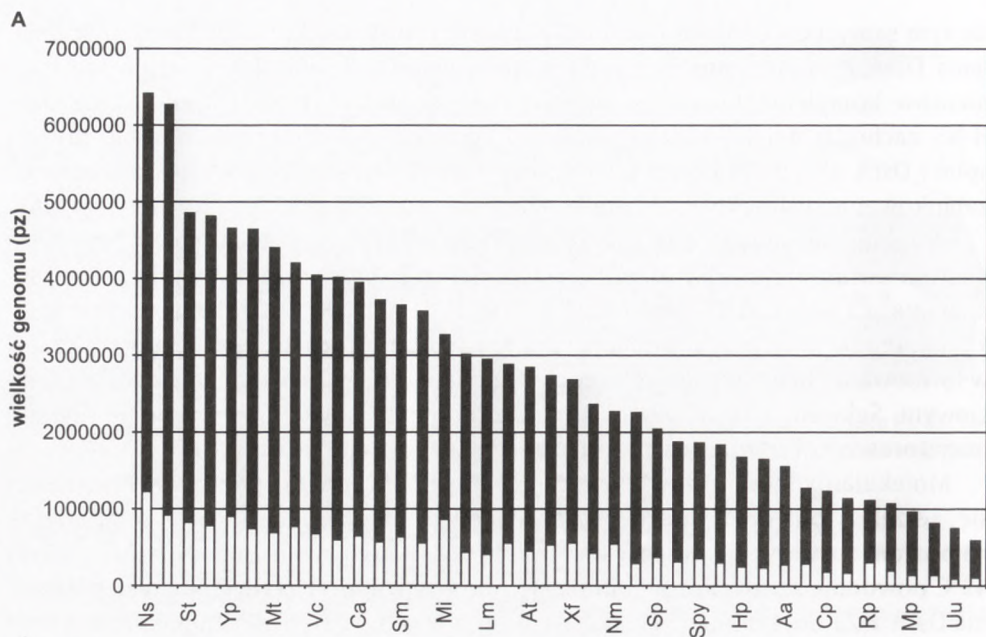
bicia jego translacji poprzez wiązanie się do regionów 3'-UTR. Proces tworzenia stRNA posiada wspólne etapy z procesem tworzenia siRNA (rys. 4). Widoczne jednak są wyraźne różnice pozwalające komórce na odróżnienie dwóch rodzajów cząsteczek o tej samej długości. Znakiem odróżniającym są wypętlenia w prekursorowym stRNA (rys. 4A). Są one rozpoznawane przez specyficzne białka (ALG 1/2). Po hydrolizie prekursorowego stRNA przez enzym DICER białka ALG 1/2 blokują przyłączenie białek kompleksu RISC, umożliwiając tym samym oddziaływanie krótkiego jednociowego RNA według mechanizmu stRNA. Inne białka (RDE1) asocjują do prekursorowych dupleksów nie zawierających wypętleń (rys. 4B). Modułują one aktywność enzymu DICER umożliwiając powstanie dwuniciowych siRNA. Pozostając zasocjowane z siRNA odpowiadają za przyłączenie kolejnych białek wchodzących w skład kompleksu RISC, odpowiedzialnego za degradację mRNA w procesie RNAi.

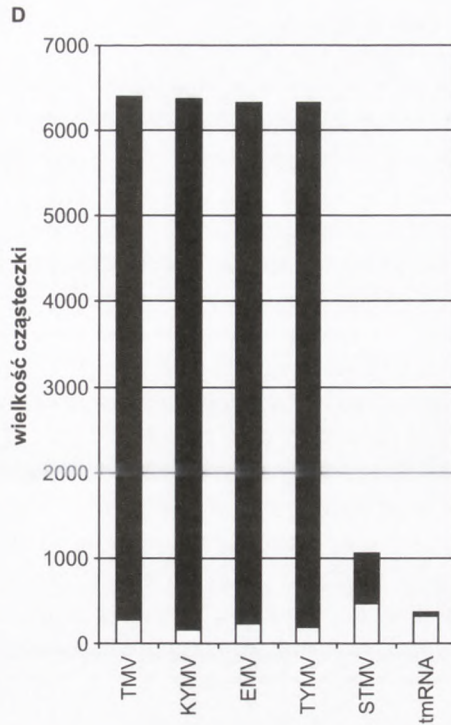
Inną grupą regulatorowych antysensownych RNA działających w układzie *trans* są dłuższe cząsteczki posiadające jedynie krótkie fragmenty komplementarne do docelowego mRNA. Mechanizm ich działania polega najczęściej na ograniczeniu lub stymulacji translacji mRNA poprzez wiązanie się do rejonów nie ulegających translacji, a odpowiedzialnych za jej inicjację (UTR, *untranslated regions*). Przykładem takiej cząsteczki jest DsrA RNA *Escherichia coli*. Jest to cząsteczka, o długości 87 nt uczestnicząca w regulacji poziomu dwóch podstawowych białkowych regulatorów transkrypcji: H-NS, będącego globalnym represorem transkrypcji oraz RpoS, czynnika transkrypcyjnego sigma o działaniu antagonistycznym do H-NS. RpoS ulega ekspresji podczas fazy stacjonarnej wzrostu bakterii oraz w odpowiedzi na stresowe warunki środowiska. Odpowiedzialny jest za zniesienie działania H-NS i zwiększenie poziomu transkrypcji. DsrA RNA wpływa na poziom obu tych cząsteczek w różny sposób: powoduje wzrost ilości RpoS oraz obniżenie poziomu H-NS przyczyniając

się tym samym do ogólnego wzrostu poziomu transkrypcji (29,30). Możliwość działania DsrA RNA jako supresor i aktywator zachodzi dzięki obecności dwóch fragmentów komplementarnych – odpowiednio do mRNA H-NS i RpoS. Wyciszenie H-NS zachodzi dzięki oddziaływaniom fragmentu położonego w obrębie drugiej spinki DsrA RNA z regionami inicjacji i terminacji translacji mRNA H-NS. Następuje zablokowanie tych sekwencji i uniemożliwienie przyłączenia do rybosomu (rys. 5A).

W odmienny sposób zachodzi proces regulacji translacji mRNA RpoS. Jego niski poziom związany jest z tworzeniem krótkich regionów dwuniciowych, które blokują miejsca wiązania do rybosomu (rys. 5B). W DsrA RNA, w obrębie pierwszej spinki, występują dwa krótkie odcinki komplementarne do sekwencji uczestniczącej w formowaniu helis oraz do sekwencji znajdującej się w sąsiednim odcinku jednoniciowym. Sekwencje te wiążą się do mRNA RpoS, co powoduje uwolnienie regionu inicjatorowego i umożliwia przyłączenie mRNA do rybosomu (30).

Molekularny mechanizm aktywacji ekspresji *DsrA* nie jest poznany. Przypuszcza się, że zależy ona od czynników środowiskowych. Stwierdzono wpływ temperatury na poziom transkrypcji, oraz na stabilność DsrA. Podwyższenie temperatury z 25 do 42°C powoduje sześciokrotne obniżenie transkrypcji, oraz skrócenie czasu półtrwania DsrA z 23 do 4 minut (31). Skutkiem tego jest trzydziestokrotnie mniejsza ilość obserwowanego transkryptu. Za taką zależność poziomu transkrypcji od temperatury odpowiedzialny jest fragment sekwencji promotorowej o długości zaledwie 36 pz. Wykazano, że wzrost ilości RpoS po szoku osmotycznym jest uzależniony od obecności DsrA. Na aktywność DsrA RNA wpływa również białko Hfq. Jest to czynnik regulacyjny posiadający domenę wiążącą kwasy nukleinowe, niezbędny m.in. dla replikacji wirusa RNA Q β . Jest on również konieczny dla regulacji ekspresji H-NS i RpoS zależnej od DsrA RNA (32). Wszystkie te fakty sugerują, że DsrA RNA stanowi główny element w złożonym systemie regulacji poziomu transkrypcji w komórce bakteryjnej.





Rys. 7. Porównanie wielkości części kodującej (czarne słupki) i niekodującej (białe słupki) genomów różnych organizmów: A – eubakterii: Ns – *Nostoc* sp. PCC 7120, Pa – *Pseudomonas aeruginosa*, St – *Salmonella typhimurium* LT2, Se – *Salmonella enterica* subsp. *enterica* serovar Typhi, Yp – *Yersinia pestis*, Ec – *Escherichia coli*, Mt – *Mycobacterium tuberculosis* CDC1551, Bh – *Bacillus halodurans*, Vc – *Vibrio cholerae*, Cc – *Caulobacter crescentus*, Ca – *Clostridium acetobutylicum*, Rs – *Ralstonia solanacearum*, Sm – *Sinorhizobium meliloti*, SP – *Synechocystis* PCC6803, Ml – *Mycobacterium leprae*, Li – *Listeria innocua*, Lm – *Listeria monocytogenes* EGD-e, Sa – *Staphylococcus aureus* subsp. *aureus* Mu50, At – *Agrobacterium tumefaciens*, Dr – *Deinococcus radiodurans*, Xf – *Xylella fastidiosa*, Ll – *Lactococcus lactis* subsp. *lactis*, Nm – *Neisseria meningitidis*, Pm – *Pasteurella multocida*, Sp – *Streptococcus pneumoniae*, Tm – *Thermotoga maritima*, Spy – *Streptococcus pyogenes*, Hi – *Haemophilus influenzae* Rd, Hp – *Helicobacter pylori* 26695, Cj – *Campylobacter jejuni*, Aa – *Aquifex aeolicus*, Rc – *Rickettsia conorii*, Cp – *Chlamydomonas reinhardtii* AR39, Tp – *Treponema pallidum*, Rp – *Rickettsia prowazekii*, Mp – *Mycoplasma pulmonis*, Mpn – *Mycoplasma pneumoniae*, Uu – *Ureaplasma urealyticum*, Mg – *Mycoplasma genitalium*; B – archeobakterii: Ss – *Sulfolobus solfataricus*, St – *Sulfolobus tokodaii*, Pa – *Pyrobaculum aerophilum*, Af – *Archaeoglobus fulgidus*, Hs – *Halobacterium* sp. NRC-1, Pab – *Pyrococcus abyssi*, Mt – *Methanobacterium thermoautotrophicum*, Ph – *Pyrococcus horikoshii*, Mj – *Methanococcus jannaschii*, Tv – *Thermoplasma volcanium*, Ta – *Thermoplasma acidophilum*; C – eukariota: Hs – *Homo sapiens*, Mm – *Mus musculus*, Ag – *Anopheles gambiae*, Dm – *Drosophila melanogaster*, At – *Arabidopsis thaliana*, Ce – *Caenorhabditis elegans*, Pf – *Plasmodium falciparum*, Sp – *Schizosaccharomyces pombe*, Sc – *Saccharomyces cerevisiae*, Ec – *Encephalitozoon cuniculi*, D – wirusów: TMV – *Tobacco mosaic virus*, KYMV – *Kennedy yellow mosaic virus*, EMV – *Eggplant mosaic virus*, TYMV – *Turnip yellow mosaic virus*, STMV – *Satellite tobacco mosaic virus*, tmRNA – *Escherichia coli* tmRNA.

3.3. Modulatory aktywności białek

Wysoka specyficzność mechanizmów regulacyjnych opartych na tworzeniu dwuniciowych kompleksów RNA:RNA uzyskiwana jest dzięki unikatowej sekwencji. Jednak regulatorowe ncRNA są w stanie oddziaływać z wysoką specyficznością z innymi cząsteczkami niż RNA, nie posiadającymi komplementarnych sekwencji. Tak jest w przypadku ncRNA zdolnych do bezpośredniej modulacji aktywności białek. Reprezentatywnym przykładem takiej cząsteczki jest SRA RNA (*steroid receptor activator*). Cząsteczka ta występuje w kilku tkankowospecyficznych izoformach (33). Forma I o długości około 0,7-0,85 kb syntetyzowana jest w największej ilości, natomiast w mniejszych ilościach występują formy II i III o długości około 1,3-1,5 kb. Wszystkie izoformy mają wspólny rdzeń o długości 687 pb oraz są poliadenylowane, różnią się natomiast fragmentami sekwencji przy końcach 3' i 5'. W obrębie rdzenia znajduje się otwarta ramka odczytu, sugerująca możliwość translacji, jednak dotychczas nie zidentyfikowano żadnego produktu białkowego oraz wykazano funkcjonowanie SRA jako RNA. Wprowadzanie dodatkowych kodonów STOP oraz obecność cykloheksimidu nie powodowały utraty funkcjonalności.

Funkcje SRA realizuje jako kompleks białkowo-rybonukleinowy SRC-1 (*steroid receptor coactivator 1*) (33,34). Wykazano, że w tej postaci wiąże się on do N-końcowej domeny receptorów steroidowych. Przypuszcza się, że obecność SRA RNA w kompleksie powoduje wysoką specyficzność tego wiązania. Pełni on również istotną rolę w regulacji zwrotnej procesu aktywacji genów zależnych od hormonów steroidowych. SRA jest wiązany przez SHARP – białkowy represor receptorów steroidowych, ulegający ekspresji pod wpływem hormonów steroidowych (35). Zidentyfikowano motywy strukturalne w obrębie SRA RNA odpowiedzialne za asocjacje tych białek (36). SRA stanowi również, jak się wydaje, istotny element w przebiegu procesów nowotworowych. Poziom SRA jest podwyższony w komórkach pobranych z nowotworu piersi (37). Co więcej, zmniejszenie poziomu SRA w komórkach nowotworowych poprzez wprowadzenie antysensowych RNA powoduje zahamowanie ich proliferacji (38).

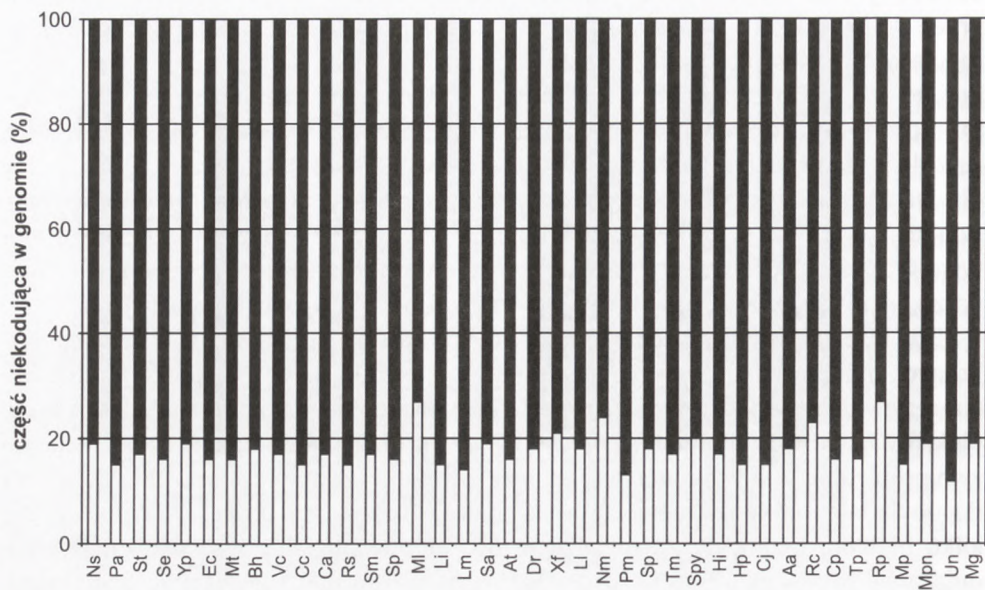
3.4. Regulatory lokalizacji RNA

Inną znaną obecnie możliwością realizacji funkcji regulacyjnych przez ncRNA jest wpływ na dystrybucję mRNA w komórce. Ma to szczególne znaczenie w komórkach o dużej specjalizacji poszczególnych przedziałów subkomórkowych. Najlepiej poznaną cząsteczką wykazującą taką aktywność jest występujący w tkance nerwowej gryzoni BC1 RNA oraz jego homolog występujący u naczelnych BC200. Obie te cząsteczki wykazują podobieństwo zarówno budowy, ekspresji jak i mechanizmu działania, chociaż potwierdzone jest ich zróżnicowane pochodzenie ewolucyjne (39). Oba RNA są produktem transkrypcji w późnych stadiach rozwojowych tkanki

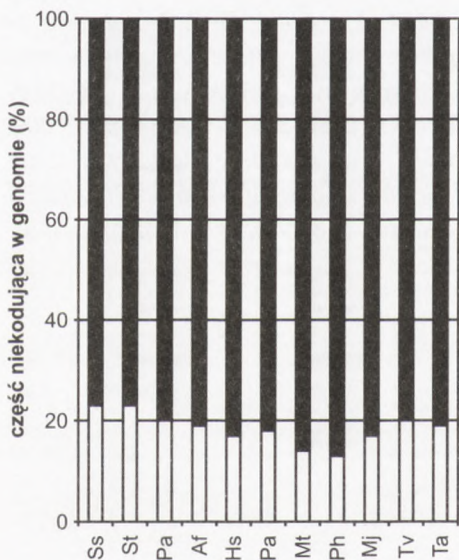
nerwowej oraz ulegają poliadenylacji. Pomimo że transkrypcja katalizowana jest przez polimerazę III, w sekwencji promotorowej obu genów występują sekwencje regulatorowe charakterystyczne dla mechanizmu transkrypcji realizowanego przez polimerazę II. Do elementów tych zaliczyć można m.in. znajdujące się w regionie promotorowym sekwencje wzmacniające (zorganizowane w kasety E) (rys. 6) wraz ze zidentyfikowanymi, wiążącymi się do nich białkami (40). Oprócz nich niezbędny dla transkrypcji jest charakterystyczny dla polimerazy III promotor wewnętrzny, składający się z kaset A i B oraz znajdującego się pomiędzy nimi specyficznego motywu GCAAG/CTTGC (rys. 6). Ostatnia z wymienionych sekwencji jest odpowiedzialna za regulację poziomu BC1 RNA zależną od poziomu transkryptu (41).

BC1 i BC200 w komórkach nerwowych wpływają na dendrytyczną lokalizację wybranych mRNA, chociaż są również doniesienia o roli BC1 w transporcie mRNA wazopresyny do aksonów (42). Oba rodzaje RNA występują w komórce w postaci kompleksów białkowo-rybonukleinowych. W rozwiązaniu mechanizmu transportu dużym problemem jest wyjaśnienie sposobu przenoszenia kompleksu BC1/BC200-mRNA do dokładnie zdefiniowanego miejsca. Istotną rolę w tym procesie przypisuje się kilku białkom BC1 RNP. Za lokalizację BC1 odpowiedzialna jest sekwencja znajdująca się przy końcu 5' tej cząsteczki. Wiążą się do niej specyficzenie białka uważane za regulatory replikacji i translacji DNA – pur α i pur β (43). Białka te charakteryzują się zdolnością do asocjacji z mikrotubulami, co świadczy o ich potencjalnych możliwościach udziału w transporcie cytoplazmatycznym. Oddziaływania pur α i pur β z BC1 okazały się słabe i okresowe (nie są one częścią BC1 RNP). Innym białkiem o podobnych właściwościach będącym częścią BC1 RNP jest translatyna. Białko to, będące represorem translacji, wiąże mRNA zawierające motywy Y i H do mikrotubul (44). Motywy takie odnaleziono również w BC1 RNA. Na podstawie tych obserwacji zaproponowano model inicjacji transportu mRNA przez BC1 RNA. Zakłada się w nim, że białka pur pełnią istotną rolę w procesie przyłączania BC1 RNP do mikrotubul, następnie oddysocjują, a rolę „uchwyty” przejmuje translatyna. Nie wiadomo jednak, jakie czynniki umożliwiają migrację kompleksu wzdłuż mikrotubul. Nie wyjaśniony pozostaje również sposób tak precyzyjnego określenia miejsca docelowego transportu.

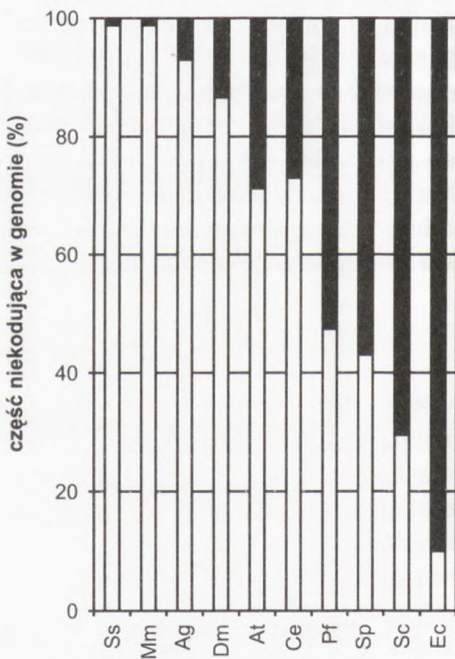
A

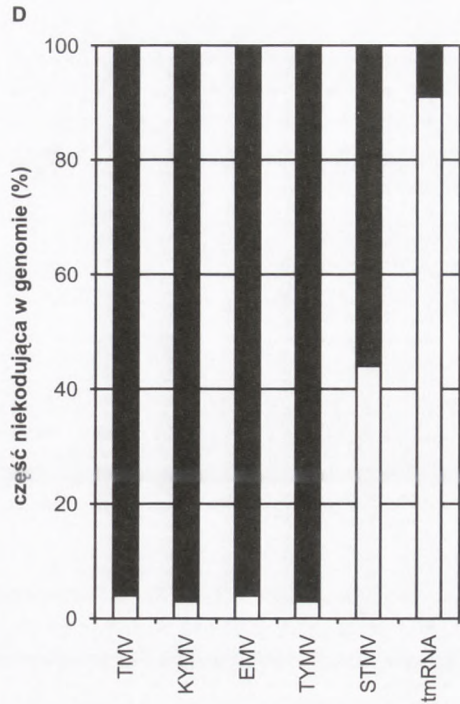


B



C





Rys. 8. Porównanie udziału części kodującej (czarne słupki) i niekodującej (białe słupki) w genomach różnych organizmów: A – eubakterie; B – archeobakterie; C – eukariota, D – wirusy. Nazwy organizmów jak w opisie do rysunku 7.

4. Zawartość ncRNA w genomie

Organizmy wykazują ogromne zróżnicowanie wielkości genomów (tab. 2). Na pierwszy rzut oka widać, że rozmiar genomu nie jest skorelowany ze stopniem skomplikowania organizmu. Od czego zatem zależy złożoność organizmu? Próbując odpowiedzieć na to pytanie wykonano analizę wybranych cech zsekwencjonowanych dotąd genomów (rys. 7-10). Pierwszym wnioskiem, jaki się nasuwa analizując rysunki zawierające zestawienia wielkości części kodujących i niekodujących (rys. 7,8) jest wyraźna korelacja pomiędzy wielkością genomu a udziałem w nim części niekodującej u eukariota, oraz jej brak w obrębie organizmów prokariotycznych. Istnieje kilka możliwych wytłumaczeń wzrostu udziału części niekodującej w obrębie eukariota. Zwiększona ilość DNA nie ulegającego translacji może stanowić swistą ochronę przed wpływem czynników mutagennych, poprzez zmniejszenie prawdopodobieństwa powstania mutacji w genie istotnym dla funkcjonowania komórki, co jest szczególnie istotne w przypadku komórek rozrodczych. Jednocześnie regiony międzygenowe mogą stanowić materiał do rearanzacji, w wyniku których mogą powstawać nowe funkcjonalne geny. Wreszcie wzrost zawartości procentowej części niekodującej u eukariota może świadczyć o zwiększonej liczbie genów kodujących ncRNA i ich istotności dla rozwoju ewolucyjnego organizmów (45). Jednak

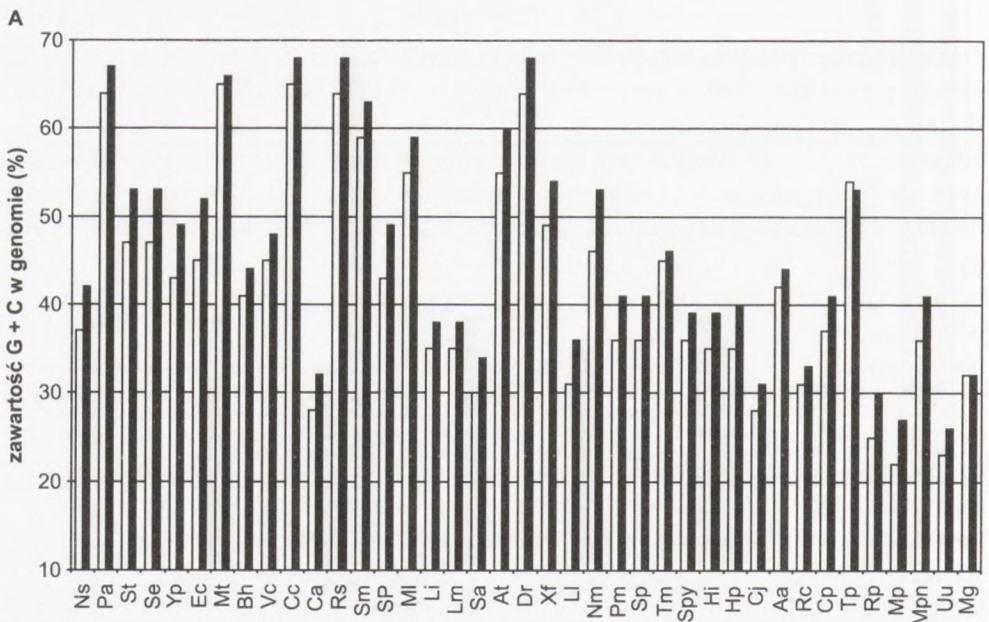
sam wzrost udziału sekwencji niekodujących w genomie nie jest wystarczającym parametrem, aby wyznaczyć rzeczywisty wzrost liczby genów ncRNA.

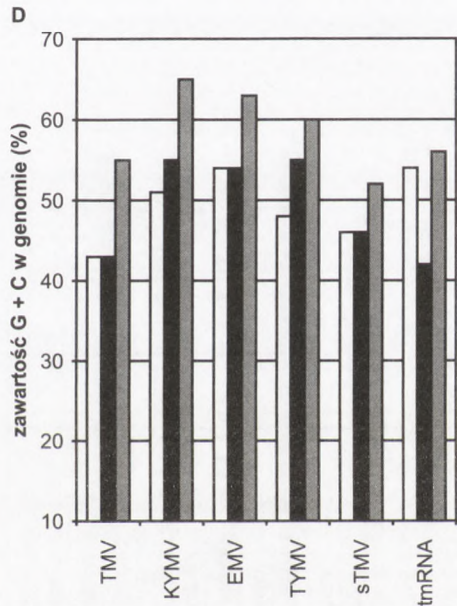
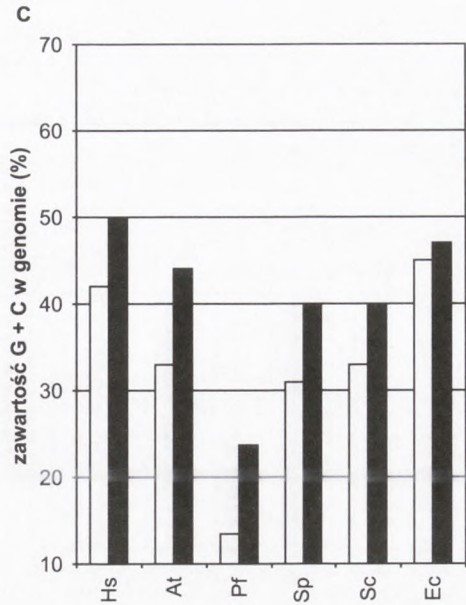
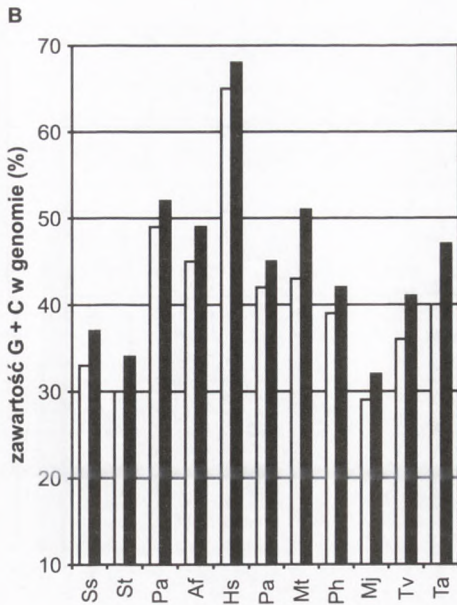
Tabela 2

Zróżnicowanie wielkości genomów różnych organizmów na wybranych przykładach

Organizm	Wielkość genomu (bp)	Liczba chromosomów
<i>Amoeba dubia</i>	670 000 000 000	kilkaset
lilia (<i>Lilium longiflorum</i>)	90 000 000 000	12
mysz (<i>Mus musculus</i>)	3 454 200 000	20
człowiek (<i>Homo sapiens</i>)	3 200 000 000	23
karp (<i>Cyprinus carpio</i>)	1 700 000 000	49
kura (<i>Gallus gallus</i>)	1 200 000 000	39
mucha domowa (<i>Musca domestica</i>)	900 000 000	6
pomidor (<i>Lycopersicon esculentum</i>)	655 000 000	12
drożdże (<i>Saccharomyces cerevisiae</i>)	12 000 000	16
<i>Escherichia coli</i>	4 639 221	1
<i>Agrobacterium tumefaciens</i>	2 841 581	1
<i>Mycoplasma genitalium</i>	580 074	1

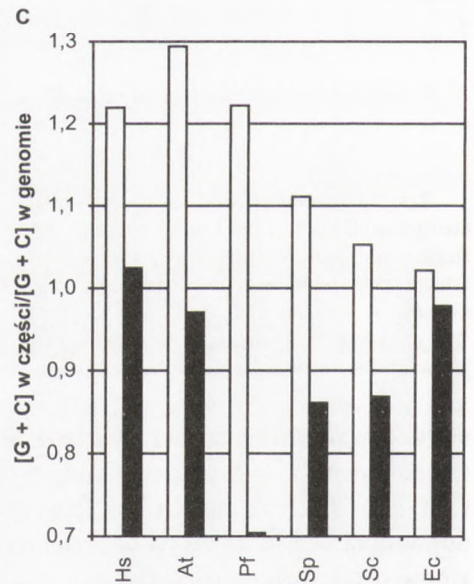
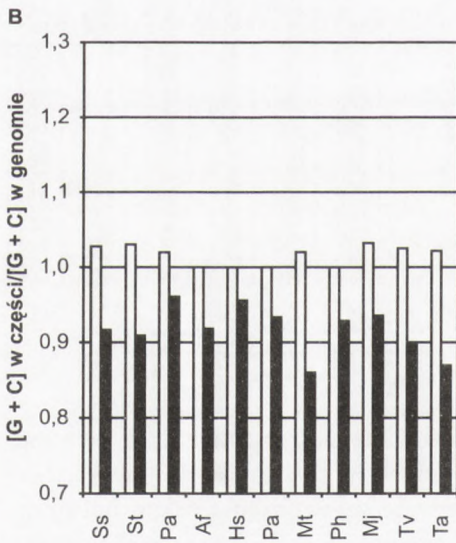
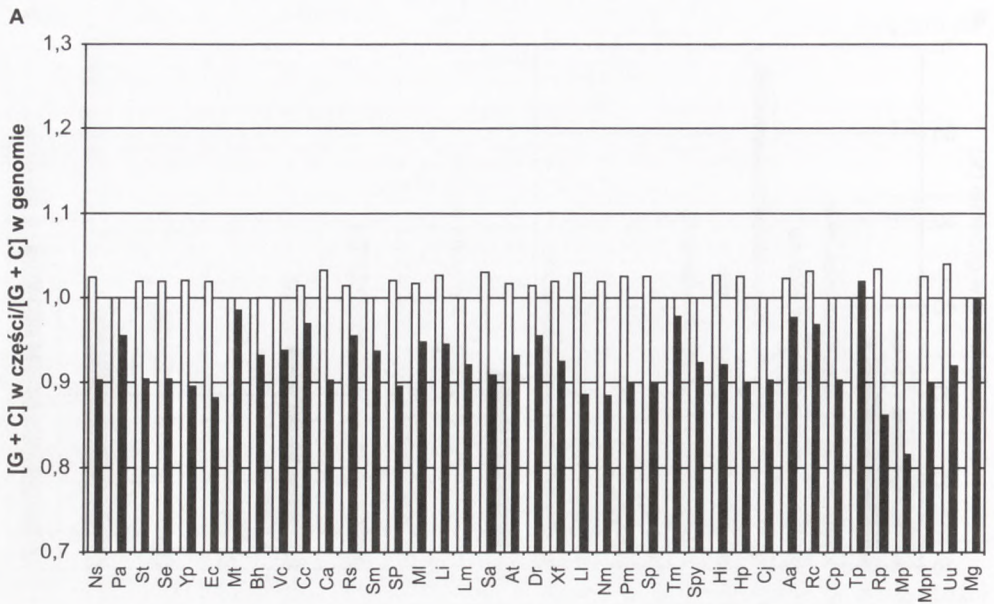
Pewnych dodatkowych wskazówek może dostarczyć analiza rozkładu zawartości G+C pomiędzy częścią kodującą i niekodującą genomu (rys. 9,10). W genomach prawie wszystkich organizmów zawartość G+C w części kodującej jest wyższa niż w części niekodującej. W obrębie genomów bakteryjnych występują o wiele większe wahania zawartości G+C w obu częściach niż u eukariota. Jest to związane z większym zróżnicowaniem warunków życia poszczególnych mikroorganizmów,



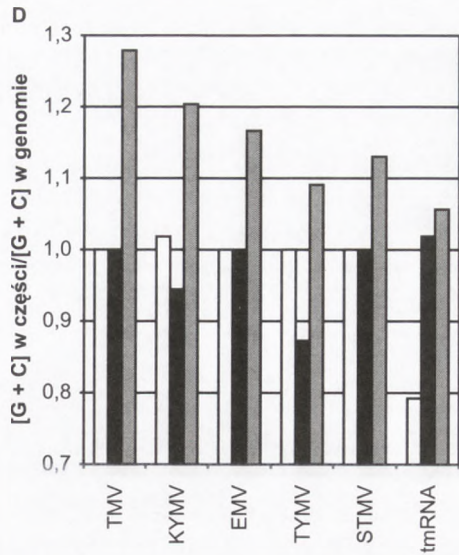


Rys. 9. Porównanie rozkładu zawartości nukleotydów G+C w części niekodującej (białe słupki), kodującej (czarne słupki) oraz strukturalnej (szare słupki) w genomach różnych organizmów: A – eubakterie; B – archeobakterie; C – eukariota; D – wirusy. Nazwy organizmów jak w opisie do rysunku 7.

niż w przypadku eukariota. Warunki obfitujące w czynniki uszkodzające DNA (np. radioaktywność – *Deinococcus radiodurans*) wymuszają znaczny wzrost zawartości G+C (do 68%). Tak dużych fluktuacji nie obserwuje się w obrębie eukariota, prawdopodobnie ze względu na bardziej jednolite warunki życia wymagane dla organizmów o tak dużej złożoności.



Rys. 10. Porównanie rozkładu zawartości nukleotydów G+C w części niekodującej (białe słupki), kodującej (czarne słupki) oraz strukturalnej (kratkowane słupki) w odniesieniu do zawartości G+C w całych genomach różnych organizmów: A – eubakterie; B – archeobakterie; C – eukariota; D – wirusy. Nazwy organizmów jak w opisie do rysunku 7.



Bardzo interesująca okazała się analiza zawartości G+C w genomach wybranych wirusów RNA oraz w tmRNA. Wszystkie wybrane cząsteczki wirusowe posiadają przy końcu 3' domenę podobną do struktury tRNA. W przypadku tmRNA domena taka jest tworzona wspólnie przez 3' i 5' końce. Przedmiotem analizy w obrębie tych cząsteczek, oprócz części kodującej i niekodującej była również wyodrębniona część strukturalna (domena podobna do tRNA). Okazało się, że część strukturalna zawiera znacznie więcej nukleotydów G+C niż pozostała część genomu (cząsteczki) (rys. 9D,10D). Jest to zgodne z obserwacjami przeprowadzonymi wcześniej, mającymi na celu znalezienie metody wyszukiwania genów ncRNA na podstawie dystrybucji nukleotydów (46). Wyjątkowo wysoka zawartość G+C w całej części niekodującej tmRNA jest, jak się wydaje, związana z tym, że w przypadku tej cząsteczki domena tRNA tworzona jest przez oba końce. W takim układzie istotna dla funkcjonalności jest, jak się wydaje, struktura całej cząsteczki, która umożliwi zbliżenie końców 3' i 5' oraz utworzenie domeny tRNA-podobnej. Dlatego dla zachowania struktury całej cząsteczki wymagany jest wzrost poziomu G+C w obrębie całej części niekodującej, a nie tylko na końcach cząsteczki. Popiera to tezę o preferencyjnym rozłożeniu guaniny i cytozyny w obrębie genów kodujących strukturalne RNA (46).

Na podstawie tych danych można postawić hipotezę, że zawartość G+C w części niekodującej odniesiona do zawartości G+C w części kodującej (rys. 10) oraz różnica tych zawartości mogą wskazywać na względną gęstość genów ncRNA w genomie. Jeżeli przyjąć, że liczba G+C w części strukturalnej jest zawsze wyższa niż w pozostałych częściach, to wzrost zawartości genów strukturalnych pociągałby za sobą wzrost zawartości G+C w części nie kodującej białek, a także zbliżanie się tej war-

tości do zawartości G+C w części kodującej. Dobrym przykładem obrazującym tę hipotezę jest zawartość G+C w genie *Arabidopsis thaliana* na tle innych eukariota (rys. 9C, 10C). Na wykresie widać wyraźnie zwiększoną zawartość G+C w części kodującej, która nie pociągnęła za sobą aż tak wyraźnego wzrostu G+C w części niekodującej ($\Delta G+C=11$). Nasuwałoby to wniosek o niskiej zawartości genów kodujących strukturalne RNA. W przypadku *A. thaliana* jest tak w rzeczywistości. Rośliny znane są z obecności ogromnych intronów i regionów intergenowych. Jednak aby zweryfikować tę hipotezę trzeba by dysponować co najmniej dwoma zsekwencjonowanymi genomami, które posiadałyby odnotowane wszystkie geny ncRNA. Na razie jest to niemożliwe.

5. Poszukiwanie genów ncRNA

Sukces technologii sekwencjonowania genomów otworzył nowe wyzwania, szczególnie w dziedzinie bioinformatyki. Jednym z nich było opracowanie sposobu składania sekwencji z ich fragmentów w całe chromosomy. Innym zagadnieniem, o wiele trudniejszym jest poszukiwanie nowych genów. O ile w przypadku genomów prokariotycznych problem jest stosunkowo łatwy do rozwiązania, to dla genomów eukariotycznych nadal nie znane są skuteczne metody przewidywania genów. Jest to spowodowane ogromną różnorodnością ich struktury. W tej chwili można jedynie w przybliżeniu określić prawdopodobną liczbę genów w genie danego organizmu, chociaż mało precyzyjnie. Metody przewidywania genów ulegają jednak bardzo szybkiemu rozwojowi, co odzwierciedla się w coraz to nowych wersjach adnotacji genomów. W genie ludzkim, w którym po zsekwencjonowaniu przewidywano obecność około 29 000 genów białkowych aktualnie szacuje się na około 22 000 genów.

Bardziej skomplikowana jest sytuacja w przypadku poszukiwania genów RNA. Obecnie nie ma metod na oszacowanie ich liczby w genie. Jest to związane z brakiem cech strukturalnych takiego genu, które umożliwiłyby jego rozpoznanie (jak np. otwarta ramka odczytu, czy zachowawcze sekwencje inicjacji translacji w przypadku genów białkowych). Stosowane obecnie metody opierają się głównie na trzech podejściach (56): a) pierwsze z nich podobne jest do tego stosowanego w przypadku poszukiwania genów kodujących białka. Polega ono na próbie znalezienia charakterystycznych wspólnych cech genów (cząsteczek) RNA, które umożliwiłyby ich identyfikację (57); b) drugie z nich odnosi się wyłącznie do poszukiwania jednostek transkrypcyjnych, czyli sekwencji inicjacji i terminacji transkrypcji. Wyszukiwane są jednostki nie posiadające otwartej ramki odczytu. Metody te nie są jednak dokładne ze względu na ogromne różnice w genach strukturalnych RNA, jak i obszarze promotorowym. NcRNA mogą ulegać transkrypcji przez wszystkie trzy typy polimeraz RNA; c) trzecie podejście opiera się na próbie porównywania zachowawczości sekwencji i struktury drugorzędowej homologicznych fragmentów sekwencji między-

genowych ze spokrewnionych organizmów (58). Najbardziej obiecująca, jak się wydaje, jest obecnie metoda zaproponowana przez S.E. Eddy i zastosowana w programie Q-RNA (59). Oprócz poszukiwania homologii uwzględnia ona kilka kryteriów. Analizie poddane są mutacje występujące w obrębie homologicznego fragmentu. Metoda zakłada trzy rodzaje rozkładu mutacji: losowe – charakterystyczne dla rejonów niefunkcyjnych; kompensacyjne charakterystyczne dla rejonów kodujących białka, polegające na zmianie nukleotydu, która nie wpływa jednak na zmianę odczytywanego z tego miejsca aminokwasu oraz kompensacyjne charakterystyczne dla rejonów kodujących RNA, polegająca na zamianie nukleotydu lub pary nukleotydów z zachowaniem struktury drugorzędowej kodowanego RNA. Dzięki temu homologiczne sekwencje można podzielić na trzy grupy: a) o dużym potencjale do kodowania białka, b) o dużym potencjale do kodowania ncRNA oraz c) „inne”. W testach programu przy użyciu sekwencji 7SL RNA wykazano, że najlepsze wyniki uzyskiwane są dla sekwencji o homologii 65-85%. Czulość programu w tym przypadku wynosi około 80,7-97,3% przy specyficzności 100-79%. Trochę gorsze wyniki uzyskano podczas testów przeprowadzonych na całych genomach. Metoda ta nie jest doskonała. Jej główną wadą jest duży spadek parametrów (czulości i specyficzności) dla sekwencji o homologii spoza przedziału 65-85%.

6. Podsumowanie

Niekodujące RNA, jak się wydaje, są niezwykle istotnym elementem w systemie regulacji ekspresji w komórce. Różnorodność procesów w których biorą udział wskazuje na ich uniwersalność zbliżoną do białek. Specyficzność reakcji w wielu przypadkach jest wyższa, niż byłyby to w stanie osiągnąć białka. Liczba znanych obecnie cząsteczek ncRNA stale rośnie. W projekcie mającym na celu zidentyfikowanie kompletnego transkryptomu myszy ujawniono 15 815 nowych kandydatów na ncRNA, spośród których wyłoniono 4280 najsilniejszych (60). Grupa ta zawierała 2431 (1573 wcześniej nie zidentyfikowanych) par ncRNA-mRNA, mogących oddziaływać na zasadzie antysensu. Uzasadnione są, jak się wydaje, pytania dotyczące rzeczywistego udziału ncRNA w genomie, a co się z tym wiąże, ogromu ich znaczenia w regulacji procesów komórkowych. Znajomość całego transkryptomu otwiera szansę na poznanie odpowiedzi na te pytania. Prawdziwy obraz potęgi ncRNA ujawni się jednak dopiero, gdy zostanie zbadana rola nowych ncRNA oraz dokładnie poznane zostaną mechanizmy kierujące ich działaniem. Dokładna wiedza w tym zakresie rokuje nadzieje na nowe metody terapeutyczne licznych schorzeń, w które uwikłane są niekodujące RNA, jak syndrom Beckwith-Wiedemann, czy różne odmiany nowotworów.

Literatura:

1. Eddy S. R., (2002), *Cell*, 109, 137-140.
2. Szymański M., Barciszewski J., (2002), *Genome Biol.*, 3, 00051-00058.
3. Szymanski M., Erdmann V. A., Barciszewski J., (2003), *Nucleic Acids Res.*, 31, 421-431.
4. Muralidhar B., Marney A., Butler M. G., (1999), *Genet Med.*, 1, 141-145.
5. DeBaun M. R., Niemitz E. L., McNeil D. E., Brandenburg S. A., Lee M. P., Feinberg A. P., (2002), *Am. J. Hum. Genet.*, 70, 604-611.
6. Ariel I., Ayesh S., Perlman E. J., Pizov G., Tanos V., Shneider T., Erdmann V. A., Podeh D., Komitowski D., Quasem A. S., de Groot N., Hochberg A., (1997), *Mol. Pathol.*, 50, 34-44.
7. Constancia M., Hemberger M., Hughes J., Dean W., Ferguson-Smith A., Fundele R., Stewart F., Kelsey G., Fowden A., Sibley C., Reik W., (2002), *Nature*, 417, 945-948.
8. Ukkola O., Sun G., Bouchard C., (2001), *Diabetologia*, 44, 2231-2236.
9. Kim S. J., Park S. E., Lee C., Lee S. Y., Jo J. H., Kim J. M., Oh Y. K., (2002), *Biochim. Biophys. Acta*, 1586, 307-315.
10. Ravenel J. D., Broman K. W., Perlman E. J., Niemitz E. L., Jayawardena T. M., Bell D. W., Haber D. A., Uejima H., Feinberg A. P., (2001), *J. Natl. Cancer Inst.*, 93, 1698-1703.
11. Wilkin F., Paquette J., Ledru E., Hamelin C., Pollak M., Deal C. L., (2000), *Eur. J. Biochem.*, 267, 4020-4027.
12. Runge S., Nielsen F. C., Nielsen J., Lykke-Andersen J., Wewer U. M., Christiansen J., (2000), *J. Biol. Chem.*, 275, 29562-29569.
13. Srivastava M., Hsieh S., Grinberg A., Williams-Simons L., Huang S. P., Pfeifer K., (2000), *Genes Dev.*, 14, 1186-1195.
14. Krzanowska H., (1998), *Genetyka molekularna*, red. Węgleński P., 307-321, Wyd. Nauk. PWN, Warszawa.
15. Kelley R. L., Kuroda M. I., (2000), *Cell*, 103, 9-12.
16. Migeon B. R., Chowdhury A. K., Dunston J. A., McIntosh I., (2001), *Am. J. Hum. Genet.*, 69, 951-960.
17. Migeon B. R., Lee C. H., Chowdhury A. K., Carpenter H., (2002), *Am. J. Hum. Genet.*, 71, 286-293.
18. Shibata S., Lee J. T., (2003), *Hum. Mol. Genet.*, 12, 125-136.
19. Avner P., Heard E., (2001), *Nat. Rev. Genet.*, 2, 59-67.
20. Morey C., Arnaud D., Avner P., Clerc P., (2001), *Hum. Mol. Genet.*, 10, 1403-1411.
21. Luikenhuis S., Wutz A., Jaenisch R., (2001), *Mol. Cell Biol.*, 21, 8512-8520.
22. Kelley R. L., Meller V. H., Gordadze P. R., Roman G., Davis R. L., Kuroda M. I., (1999), *Cell*, 98, 513-522.
23. Meller V. H., Gordadze P. R., Park Y., Chu X., Stuckenzol C., Kelley R. L., Kuroda M. I., (2000), *Current Biol.*, 10, 136-143.
24. Mitsuya K., Meguro M., Lee M. P., Katoh M., Schulz T. C., Kugoh H., Yoshida M. A., Niikawa N., Feinberg A. P., Oshimura M., (1999), *Hum. Mol. Genet.*, 8, 1209-1217.
25. Zamore P. D., (2001), *Nat. Struct. Biol.*, 8, 746-750.
26. McManus M. T., Sharp P., (2002), *Nat. Rev. Genet.*, 3, 737-747.
27. Ha I., Wightman B., Ruvkun G., (1996), *Genes Dev.*, 10, 3041-3050.
28. Banarjee D., Slack F., (2002), *Bioessays*, 24, 119-129.
29. Majdalani N., Cunnig Ch., Sledjeski D., Elliot T., Gottesman S., (1998), *Proc. Natl. Acad. Sci. USA*, 95, 12462-12467.
30. Lease R. A., Belford M., (2000), *Proc. Natl. Acad. Sci. USA*, 97, 9919-9924.
31. Repoila F., Gottesman S., (2001), *J. Bacteriol.*, 183, 4012-4023.
32. Sledjeski D. D., Whitman Ch., Zhang A., (2001), *J. Bacteriol.*, 183, 1997-2005.
33. Lanz R. B., McKenna N. J., Onate S. A., Albrecht U., Wong J., Tsai S. Y., Tsai M. J., O'Malley B. W., (1999), *Cell*, 97, 17-27.
34. Watanabe M., Yanagisawa J., Kitagawa H., Takeyama K., Ogawa S., Arai Y., Suzawa M., Kobayashi Y., Yano T., Yoshikawa H., Masuhiro Y., Kato S., (2001), *EMBO J.*, 20, 1341-1352.
35. Shi Y., Downes M., Xie W., Kao H. Y., Ordentlich P., Tsai C. C., Hon M., Evans R. M., (2001), *Genes Dev.*, 15, 1140-1151.

36. Lanz R. B., Razani B., Goldberg A. D., O'Malley B., (2002), Proc. Natl. Acad. Sci. USA, 99, 16081-16086.
37. Murphy L. C., Simon S. L., Parkes A., Leygue E., Dotzlaw H., Snell L., Troup S., Adeyinka A., Watson P. H., (2000), Cancer Res., 60, 6266-6271.
38. Cavarretta I. T., Mukopadhyay R., Lonard D. M., Cowser L. M., Bennett C. F., O'Malley B. W., Smith C. L., (2002), Mol. Endocrinol., 16, 253-270.
39. Kuryshev V. Y., Skryabin B. V., Kremerskothen J., Jurka J., Brosius J., (2001), J. Mol. Biol., 309, 1049-66.
40. Kobayashi S., Anzai K., (1998), Biochem. Biophys. Res. Commun., 245, 59-63.
41. Kobayashi S., Anzai K., (1997), Biochem. Biophys. Res. Commun., 239, 407-411.
42. Trembleau A., Melia K. R., Bloom F. E., (1995), Eur. J. Neurosci., 7, 2249-2260.
43. Ohashi S., Kobayashi S., Omori A., Ohara S., Omae A., Muramatsu T., Li Y., Anzai K., (2000), J. Neurochem., 75, 1781-1790.
44. Muramatsu T., Ohmae A., Anzai K., (1998), Biochem. Biophys. Res. Commun., 247, 7-11.
45. Mattick J. S., (2001), EMBO Rep., 21, 986-991.
46. Schattner P., (2002), Nucleic Acids Res., 30, 2076-2082.
47. The Arabidopsis Genome Initiative, (2000), Nature, 408, 796-815.
48. The *C. elegans* Sequencing Consortium, (1998), Science, 282, 2012-2018.
49. Adams M. D., et al. (2000), Science, 287, 2185-2195.
50. Katinka M. D., Duprat S., Cornillot E., Metenier G., Thomarat F., Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P., Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M., Weissenbach J., Vivares Ch. P., (2001), Nature, 414, 450-453.
51. Venter J. C., et al., (2001), Science, 291, 1304-1351.
52. Wood V., et al., (2002), Nature, 415, 871-880.
53. Holt R. A., et al., (2002), Science, 298, 129-149.
54. Gardner M. J., et al., (2002), Nature, 419, 498-511.
55. Mouse Genome Sequencing Consortium, (2002), Nature, 420, 520-562.
56. Eddy S. R., (2002), Cell, 109, 137-140.
57. Carter J. R., Dubchak I., Holbrook S. R., (2001), Nucleic Acids Res., 29, 3928-3938.
58. Rivas E., Klein R. J., Jones T. A., Eddy S. R., (2001), Curr. Biol., 11, 1369-1373.
59. Rivas E., Eddy S. R., (2001), BMC Bioinformatics, 2, 8.
60. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team, (2002), Nature, 420, 563-573.

Ostatnio zidentyfikowano nowy mechanizm kontroli ekspresji genu *Makorin1* przez jego pseudogen *Makorin1-p1*. Transkrypt pseudogenu nie ma potencjału kodującego i jest niezbędny dla zapewnienia wysokiej ekspresji genu *Makorin1*. Postuluje się dwa możliwe mechanizmy tego procesu: współzawodnictwo ilościowe o czynnik destabilizujący mRNA (np. RNaza) lub o wiązanie represorów transkrypcji. W obu przypadkach następuje zmniejszenie stężenia białka w cytoplazmie poprzez jego związanie do niekodującego transkryptu pseudogenu. Odkrycie to rodzi pytanie o rzeczywistą rolę około 20 000 pseudogenów znajdujących się w genomie człowieka. Ich liczba, zbliżona do liczby genów białkowych sugeruje możliwość ich funkcjonowania jako niekodujące RNA w procesie regulacji ekspresji genów białkowych.

Hirotsune S., Yoshida N., Chen A., Garret L., Sugiyama F., Takahashi S., Yagami K., Wynshaw-Boris A., Yoshiki A., (2003), Nature, 423, 91-96.

Lee J.T., (2003), Nature, 423, 26-28.