

Instytut Chemii Bioorganicznej Polskiej Akademii Nauk



– Rozprawa doktorska –

Identyfikacja i charakterystyka
elementów strukturalnych genomowego RNA
retrotranspozonu Ty1 oraz jego oddziaływań
z białkiem Gag kluczowych dla dimeryzacji
i pakowania genomu do cząstek wirusopodobnych

– **Julita Gumna** –

Promotor: dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

Praca realizowana w Zakładzie Struktury i Funkcji Retrotranspozonów

Poznań, 2021

*Mojej Promotor, dr hab. Katarzynie Pachulskiej-Wieczorek, prof. ICHB PAN
za opiekę naukową,
za przekazaną wiedzę i wiele cennych rad,
za możliwość rozwoju,
za poświęcony czas i nieocenione wsparcie*

*Koleżankom i Kolegom z ICHB PAN
za pomoc, miłą atmosferę w pracy i życzliwość*

*Rodzicom, Rodzeństwu i Łukaszowi
za ogromne wsparcie i wiarę we mnie*

Dziękuję

Jestem jedną z tych osób, które uważają, że nauka kryje w sobie wielkie piękno. Naukowiec w swym laboratorium jest nie tylko technikiem – jest także dzieckiem stojącym twarzą w twarz ze zjawiskami natury, które budzą jego zachwyty niczym bajka.

– Maria Skłodowska-Curie

SPIS TREŚCI

1. WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ.....	6
2. WYKAZ POZOSTAŁYCH PRAC NAUKOWYCH.....	7
3. STRESZCZENIE.....	8
4. ABSTRACT.....	9
5. WPROWADZENIE.....	10
5.1. Ogólna charakterystyka, budowa i klasyfikacja retrotranspozonów LTR.....	10
5.2. Retrotranspozony LTR w <i>Saccharomyces cerevisiae</i>	11
5.3. Organizacja oraz zarys cyklu replikacyjnego retrotranspozonu Ty1.....	12
5.4. Rola białka Gag Ty1 w cyklu replikacyjnym retrotranspozonu.....	14
5.4.1. Gag Ty1 jako białko strukturalne. Budowa VLP.....	14
5.4.2. Rola białka Gag Ty1 w transporcie gRNA Ty1 do retrosomów.....	15
5.4.3. Gag Ty1 jako białko opiekuńcze wobec kwasów nukleinowych.....	15
5.4.4. Gag Ty1 a mechanizm kontroli liczby kopii retrotranspozonu.....	17
5.5. Oddziaływania RNA/RNA istotne dla retrotranspozycji Ty1.....	17
5.6. Procesy dimeryzacji i pakowania genomowego RNA u retroelementów.....	19
5.6.1. Dimeryzacja i pakowanie genomowego RNA u retrowirusów.....	19
5.6.2. Dimeryzacja i pakowanie genomowego RNA retrotranspozonu Ty1.....	20
6. CEL NAUKOWY PRZEDSTAWIONEGO CYKLU PRAC.....	22
7. SKRÓTOWY OPIS PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ.....	23
7.1. Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA.....	23
7.1.1. Rola białka Gag w dimeryzacji RNA Ty1.....	24
7.1.2. Oddziaływania RNA/RNA istotne dla formowania dimeru RNA Ty1.....	24
7.1.3. Analiza SHAPE monomeru i dimeru RNA mTy1.....	25
7.1.4. Wpływ mutacji w sekwencjach potencjalnie zaangażowanych w dimeryzację RNA Ty1 na efektywność tego procesu <i>in vitro</i>	27
7.1.5. Rola struktury pseudowęzła w dimeryzacji RNA Ty1.....	28
7.1.6. Wpływ mutacji w sekwencjach PAL1, PAL2, PAL3 i IL3 na efektywność retrotranspozycji Ty1 w komórce.....	28
7.1.7. Podsumowanie.....	29
7.2. RNA binding properties of the Ty1 LTR-retrotransposon Gag protein.....	29
7.2.1. Wpływ siły jonowej buforu reakcyjnego na oddziaływanie Gag Ty1/RNA.....	30
7.2.2. Mechanizm wiązania RNA przez Gag Ty1.....	31
7.2.3. Oddziaływania Gag Ty1/RNA w warunkach kompetycyjnych.....	32

7.2.4. Rola struktury pseudowęzła w wiązaniu RNA przez Gag Ty1.....	32
7.2.5. Podsumowanie.....	34
7.3. RNATHOR – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis.....	34
7.3.1. Wprowadzenie.....	35
7.3.2. RNATHOR.....	36
7.3.3. Podsumowanie.....	37
8. PODSUMOWANIE.....	38
9. WYKAZ STOSOWANYCH SKRÓTÓW I NOMENKLATURA.....	40
10. BIBLIOGRAFIA.....	42

ZAŁĄCZNIKI:

Załącznik nr 1 ŻYCIORYS NAUKOWY

Załącznik nr 2 PRACE NAUKOWE WCHODZĄCE W SKŁAD ROZPRAWY DOKTOR-
SKIEJ

Załącznik nr 3 OŚWIADCZENIA OKREŚLAJĄCE WKŁAD KANDYDATKI W POWSTA-
NIE PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

1. WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

Rozprawę doktorską stanowi cykl powiązanych tematycznie publikacji, wymienionych poniżej. Prace te zostały opublikowane w renomowanych czasopismach z listy JCR. Sumaryczny współczynnik wpływu (IF, ang. *impact factor*)^{*} przedstawionego cyklu prac naukowych wynosi 14,513, a liczba punktów MNiSW^{**} – 340. W każdej z prac jestem pierwszym autorem.

[1] Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel, Katarzyna Pachulska-Wieczorek[§] (2019) *Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA*; RNA Biology 16: 1749-1763

IF₂₀₁₉ – **5,350**; pkt MNiSW – **100**

[2] Julita Gumna, Angelika Andrzejewska-Romanowska, David J. Garfinkel and Katarzyna Pachulska-Wieczorek[§] (2021) *RNA binding properties of the Ty1 LTR-retrotransposon Gag protein*; International Journal of Molecular Sciences 22(16):9103

IF₂₀₂₀ – **5,923**; pkt MNiSW – **140**

[3] Julita Gumna, Tomasz Żok, Kacper Figurski, Katarzyna Pachulska-Wieczorek[§], Marta Szachniuk[§] (2020) *RNAtThor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis*; PLoS One 15(10):e0239287

IF₂₀₂₀ – **3,240**; pkt MNiSW – **100**

[§] *autor do korespondencji*

^{*} *IF na podstawie danych w bazie Journal Citation Reports na platformie Web of Science*

^{**} *pkt MNiSW na podstawie ujednoliconego wykazu czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego z dnia 18.02.2021*

Kopie powyższych prac naukowych wraz z materiałami dodatkowymi umieszczono w załączniku nr 2. Oświadczenia, dotyczące wkładu autorki rozprawy doktorskiej w powstanie poszczególnych prac naukowych, znajdują się w załączniku nr 3.

2. WYKAZ POZOSTAŁYCH PRAC NAUKOWYCH

W skład mojego pozostałego dorobku naukowego wchodzi 3 publikacje również dotyczące badań nad retrotranspozonom Ty1. Nie zostały one jednak włączone do cyklu prac naukowych wchodzących w skład rozprawy doktorskiej, ponieważ dotyczą innych aspektów replikacji Ty1, a mój udział w ich powstaniu nie był wiodący. Jeden z rozdziałów publikacji Nishida et al. *NAR*, 2015 przedstawia badania, które realizowałam w ramach mojej pracy magisterskiej.

Angelika Andrzejewska, Małgorzata Zawadzka, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek[§] (2021) *In vivo structure of the Ty1 retrotransposon RNA genome*; *Nucleic Acids Research* 49(5):2878-2893

IF₂₀₂₀^{*} – **16,971**; pkt MNiSW^{**} – **200**

Katarzyna Pachulska-Wieczorek, Leszek Błaszczuk, Julita Gumna, Yuri Nishida, Agniva Saha, Marcin Biesiada, David J. Garfinkel and Katarzyna J. Purzycka[§] (2016) *Characterizing the functions of Ty1 Gag and the Gag-derived restriction factor p22/p18*; *Mobile Genetic Elements* 7;6(2):e1154637

IF₂₀₁₆ – **0**; pkt MNiSW – **70**

Yuri Nishida[†], Katarzyna Pachulska-Wieczorek[†], Leszek Błaszczuk, Agniva Saha, Julita Gumna, David J. Garfinkel[§], Katarzyna J. Purzycka[§] (2015) *Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions*; *Nucleic Acids Research* 43 (15): 7414-7431

IF₂₀₁₅ – **9,202**; pkt MNiSW – **200**

[§] autor do korespondencji

[†] równorzędny pierwszy autor

^{*} IF na podstawie danych w bazie *Journal Citation Reports* na platformie *Web of Science*

^{**} pkt MNiSW na podstawie ujednoliconego wykazu czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego z dnia 18.02.2021

3. STRESZCZENIE

Elementy transpozycyjne występują w większości, jeśli nie we wszystkich, dotychczas scharakteryzowanych genomach eukariotycznych i mają znaczący wpływ na ich organizację, ewolucję i funkcjonowanie. Dużą część mobilnych elementów genetycznych stanowią retrotranspozony typu LTR, które wykazują wiele podobieństw w strukturze i cyklu replikacyjnym do retrowirusów, nie są one jednak infekcyjne. Podczas retrotranspozycji tworzą one cząstki wirusopodobne (VLPs, ang. *virus-like particles*), przypominające wiriony retrowirusów. VLPs zbudowane są głównie z licznych cząsteczek białka Gag i zawierają genomowy RNA (gRNA) retroelementu w formie dimerycznej oraz specyficzne enzymy wymagane do odwrotnej transkrypcji i późniejszej integracji cDNA z genomem gospodarza. Pakowanie gRNA do VLPs jest jednym z kluczowych etapów replikacji retroelementów. Proces ten był intensywnie badany dla retrowirusów i dostępne dane literaturowe wskazują, że pakowanie gRNA jest ściśle związane z jego dimeryzacją, a oba procesy reguluje białko Gag. W odróżnieniu od retrowirusów, proces dimeryzacji i pakowania gRNA retrotranspozonów LTR poznany jest bardzo słabo.

Głównym celem mojej rozprawy doktorskiej było zbadanie procesów dimeryzacji i pakowania gRNA do VLPs u retrotranspozonów LTR. Moim modelem badawczym był retrotranspozon Ty1, naturalnie występujący w genomie *Saccharomyces cerevisiae*. W pracach naukowych, stanowiących moją rozprawę doktorską (Gumna et al. *RNA Biology*, 2019; Gumna et al. *IJMS*, 2021; Gumna et al. *PLoS One*, 2020), wskazałam, że mechanizm dimeryzacji genomowego RNA retrotranspozonu Ty1 wykazuje istotne podobieństwa z analogicznym procesem u retrowirusów, a białko Gag Ty1 jest ważnym czynnikiem regulatorowym dla tego procesu. W celu wyjaśnienia roli Gag Ty1 w selekcji i pakowaniu gRNA retrotranspozonu do VLPs przeanalizowałam właściwości oddziaływań Gag Ty1 z RNA oraz zdefiniowałam motywy strukturalne w RNA Ty1 istotne dla tych interakcji. Swoje doświadczenie w mapowaniu struktury drugorzędowej RNA wykorzystałam do opracowania, we współpracy z Zakładem Bioinformatyki Strukturalnej ICHB PAN, narzędzia obliczeniowego – RNATHor. Narzędzie to służy do w pełni zautomatyzowanej normalizacji, wizualizacji oraz analizy statystycznej danych uzyskanych w eksperymentach próbkowania strukturalnego RNA z wykorzystaniem elektroforezy kapilarnej.

4. ABSTRACT

Transposable elements occur in most, if not all, characterized eukaryotic genomes and significantly influence their organization, evolution, and function. Many of them are LTR retrotransposons, which display similarity in structure and replication cycle to retroviruses but are not infectious. During retrotransposition, they form virus-like particles (VLPs), resembling retroviral virions. VLPs are composed of numerous Gag proteins, contain the retroelement genomic RNA (gRNA) in dimeric form, and the specific enzymes required for reverse transcription and later integration of cDNA into a host genome. gRNA packaging into VLPs is one of the key steps in retroelement replication. This process has been extensively studied for retroviruses, and the available literature data indicate that gRNA packaging is coupled with its dimerization, and both processes are regulated by the Gag protein. In contrast to retroviruses, the process of gRNA dimerization and packaging is poorly understood for LTR retrotransposons.

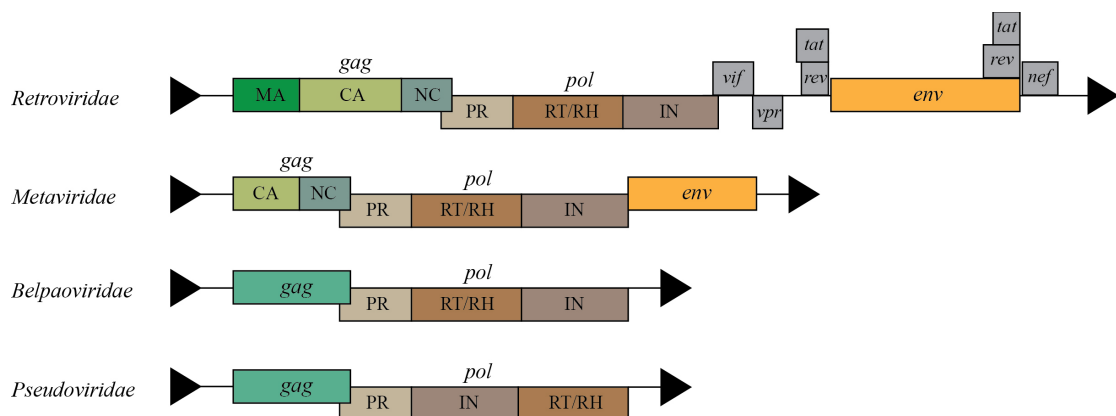
The main goal of my dissertation was to investigate the processes of retrotransposon gRNA dimerization and packaging into VLPs. My model system was Ty1 LTR-retrotransposon, naturally occurring in the genome of *Saccharomyces cerevisiae*. In the publications constituting my dissertation (Gumna et al. *RNA Biology*, 2019; Gumna et al. *IJMS*, 2021; Gumna et al. *PLoS One*, 2020), I indicated that the mechanism of Ty1 gRNA dimerization is similar to the analogous process in retroviruses and regulated by the Ty1 Gag protein. To elucidate the role of Ty1 Gag in the selection and packaging of retrotransposon gRNA into VLPs, I analyzed the properties of the Ty1 Gag interactions with RNA and defined structural motifs in Ty1 RNA relevant for these interactions. I used my experience in RNA secondary structure mapping to develop, in collaboration with the Department of Structural Bioinformatics IBCH PAS, a computational tool – RNAtor. This tool serves for fully automated normalization, visualization, and statistical analysis of data obtained in RNA structural probing experiments and resolved by capillary electrophoresis.

5. WPROWADZENIE

5.1. Ogólna charakterystyka, budowa i klasyfikacja retrotranspozonów LTR

Retrotranspozony LTR to ruchome elementy genetyczne, które mają zdolność przemieszczania się w genomie gospodarza głównie dzięki kodowaniu odwrotnej transkryptazy oraz integrazy. Są one szeroko rozpowszechnione w genomach zwierząt, roślin, alg, grzybów, a także u człowieka [4]. Ta szeroka dystrybucja sugeruje starożytne pochodzenie tych retroelementów, prawdopodobnie towarzyszące pojawieniu się eukariontów [5]. Główne rodziny retrotranspozonów LTR stanowią: *Pseudoviridae* i *Metaviridae*, znane również jako Ty1/Copia i Ty3/Gypsy. Te ostatnie uważane są za przodków retrowirusów [6]. Do autonomicznych retrotranspozonów LTR zalicza się również retrotranspozony BEL/Pao z rodziny *Belpaoviridae* oraz endogenne retrowirusy, które mają swoich przedstawicieli w rodzinie *Retroviridae*.

Pseudoviridae, *Metaviridae*, *Belpaoviridae* i *Retroviridae* wykazują liczne podobieństwa strukturalne oraz funkcjonalne i zostały sklasyfikowane do rzędu *Ortervirales* [7]. Jedną z najważniejszych cech, łączących te cztery rodziny, jest replikacja poprzez cząsteczkę RNA, która w procesie odwrotnej transkrypcji przepisywana jest na dwuniciowy cDNA, ulegający następnie integracji do genomu gospodarza. Jako starter dla odwrotnej transkryptazy wykorzystują cząsteczkę komórkowego tRNA [8]. Ich cDNA posiada na końcach 5' i 3' długie terminalne powtórzenia (LTR, ang. *long terminal repeats*), które są niezbędne do transkrypcji i integracji (Rys. 5.1) [9]. Wszystkie retroelementy zawierają geny *gag* i *pol*, których otwarte ramki odczytu nakładają się.



Rysunek 5.1. Uogólniony schemat organizacji genetycznej retroelementów, należących do rodzin *Retroviridae*, *Metaviridae*, *Belpaoviridae* i *Pseudoviridae* (na podstawie [7]). Wszystkie retroelementy posiadają rejony LTR (czarne trójkąty) oraz geny *gag* i *pol*. Różnią się obecnością genu *env* oraz kolejnością kodowania białek PR, RT/RH i IN. DNA *Retroviridae* zawiera dodatkowo geny *vif*, *vpr*, *rev*, *tat* i *nef*, kodujące białka regulatorowe.

Gen *gag* koduje główne białko strukturalne Gag, które jest niezbędne do tworzenia wirionów lub cząstek wirusopodobnych. Gen *pol* koduje specyficzne enzymy potrzebne na róż-

nych etapach cyklu replikacyjnego: proteazę asparaginową (PR), odwrotną transkryptazę z domeną RNazy H (RT/RH) oraz integrazę (IN) [4]. Rodzina Ty1/Copia wyróżnia się kolejnością kodowania tych enzymów (PR-IN-RT), która dla pozostałych omawianych rodzin jest następująca: PR-RT-IN. Jedną z głównych różnic między retrowirusami a retrotranspozonami LTR jest infekcyjność warunkowana obecnością pełnej sekwencji genu *env*, kodującego białka otoczki wirusa [4]. *Retroviridae* posiadają gen *env*, jednak u endogennych retrowirusów nie dochodzi do produkcji białek funkcjonalnych [10]. Retrotranspozony LTR zawierają tylko szczątkowe fragmenty genu *env* lub nie zawierają go wcale, przez co mogą wstawiać nowe kopie swojego genomu do DNA gospodarza jedynie w obrębie tej samej komórki [4].

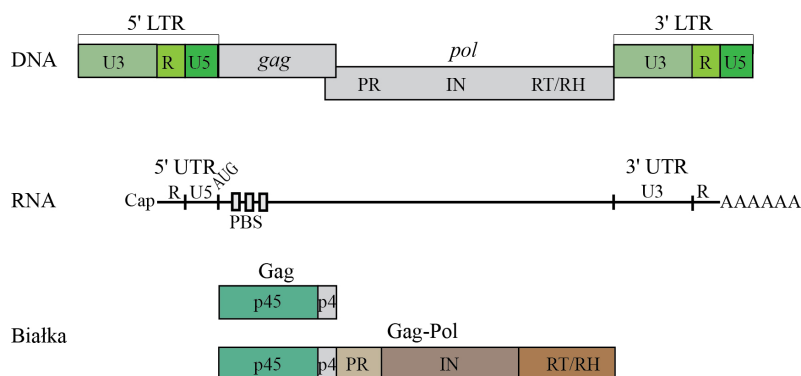
5.2. Retrotranspozony LTR w *Saccharomyces cerevisiae*

W odróżnieniu od innych organizmów eukariotycznych, *Saccharomyces cerevisiae* posiadają niewielką ilość mobilnych elementów genetycznych i stanowią one jedynie około 3,35% genomu [11]. Większość z nich, czyli Ty1, Ty2, Ty4 i Ty5 to retrotranspozony LTR z rodziny Ty1/Copia, natomiast rodzinę Ty3/Gypsy reprezentują elementy Ty3 i Ty3p. Prześledzenie pełnej sekwencji genomu szczepu S288C *S. cerevisiae* pozwoliło wyznaczyć łącznie 483 insercje elementów Ty [11]. Chociaż znaczną ich część stanowią tzw. solo-LTR, to zidentyfikowano również elementy pełnej długości, wśród których z największą częstotliwością występują retrotranspozony Ty1 i Ty2 (odpowiednio 32 i 13 kopii). Dla pozostałych retrotranspozonów Ty liczba kopii o pełnej sekwencji jest dużo niższa Ty3 - 2, Ty4 - 3 i Ty5 - 1. Niektóre kopie Ty zawierają delecje i mutacje, które czynią je transpozycyjnie defektywnymi lub nieautonomicznymi, ale nadal mogą ulegać ekspresji. Wykazano jednak, że część elementów Ty1, Ty2, Ty3 i Ty4 w genomie *S. cerevisiae* ma nienaruszoną strukturę i zachowaną zdolność do transpozycji. W przypadku rodziny Ty5 nie wykryto aktywnych elementów [11,12].

Drożdzowy retrotranspozon Ty1 jest jednym z najlepiej poznanych endogennych retroelementów. Od wielu lat jest szeroko wykorzystywany jako model badawczy w celu zrozumienia w, jaki sposób retroelementy namnażają się w komórce, wchodzą w interakcje z czynnikami komórkowymi, kształtują genom gospodarza oraz jak ich aktywacja wpływa na ekspresję sąsiednich genów. Naturalna retrotranspozycja Ty1 jest bardzo rzadkim zdarzeniem w *S. cerevisiae* i występuje z częstotliwością 10^{-5} - 10^{-7} na element Ty1 na pokolenie [13], co utrudnia odróżnienie retromobilności od zdarzeń rekombinacyjnych w DNA. Dlatego też opracowano szereg narzędzi, ułatwiających scharakteryzowanie procesów retrotranspozycji w *Saccharomyces cerevisiae*. W badaniach wykorzystuje się zmodyfikowane genetycznie szczepy, zawierające wysokokopijne plazmidy z wklonowanym elementem funkcjonalnym Ty1-H3, co zapewnia dużo większą częstotliwość retrotranspozycji w komórce.

5.3. Organizacja oraz zarys cyklu replikacyjnego retrotranspozonu Ty1

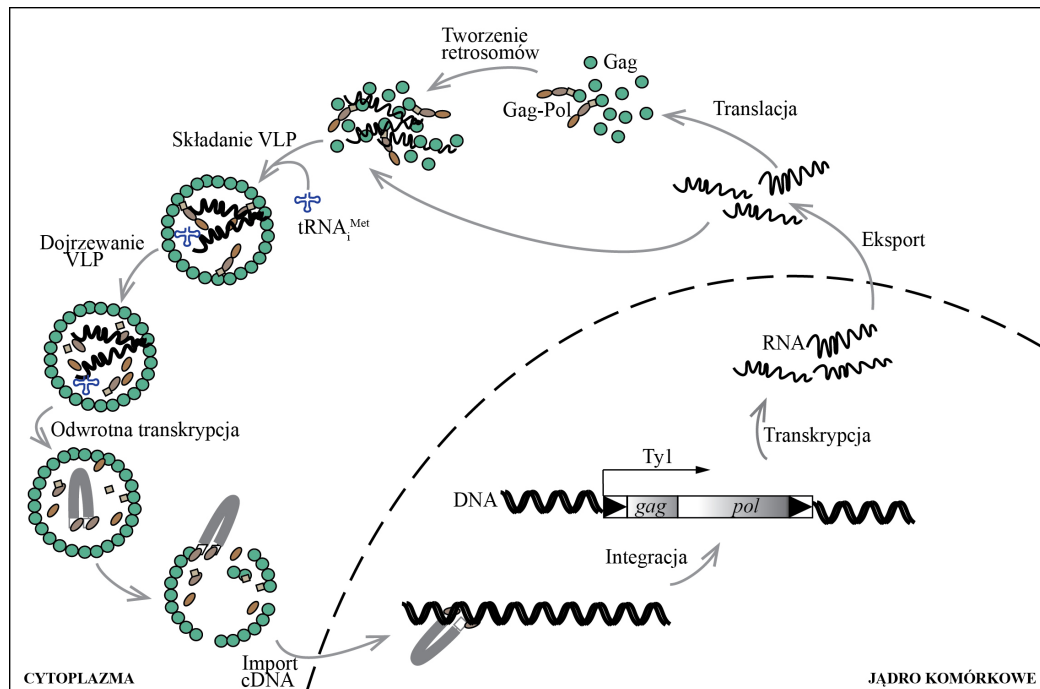
DNA drożdżowego elementu Ty1 razem z sekwencjami LTR o długości 334 pz na każdym końcu, liczy 5918 pz (wszystkie pozycje reszt nukleotydowych podano w niniejszej pracy w oparciu o sekwencję elementu Ty1-H3, GenBank: M18706.1) (Rys. 5.2). Każdy LTR zawiera dwie sekwencje unikatowe: U3 (ang. *unique 3' terminus*) i U5 (ang. *unique 5' terminus*) o długości odpowiednio 240 nukleotydów i 38 nukleotydów oraz 56-nukleotydowy fragment powtórzony R (ang. *repeat*). Pomiędzy rejonami LTR znajdują się sekwencje dwóch genów: *gag* i *pol* (znane w literaturze również jako TYA1 i TYB1). Kodon start dla *gag* stanowią trzy ostatnie nukleotydy rejonu R na końcu 5'. Ramka odczytu genu *pol* znajduje się w pozycji +1 w stosunku do ramki odczytu genu *gag* i zachodzi na jej ostatnie 38 pz. Element Ty1 nie zawiera odpowiednika retrowirusowego genu *env* ani żadnych jego pozostałości [14,15].



Rysunek 5.2. Organizacja elementu Ty1, jego RNA i białek. DNA Ty1 zawiera rejony LTR na końcach 5' i 3' oraz geny *gag* i *pol*. Genomowy RNA Ty1 zawiera rejony nieulegające translacji (UTR), miejsce wiązania startera (PBS), które wiąże inicjatorowe tRNA^{Met} oraz rejon kodujący. Geny *gag* i *pol* kodują białka Gag i Gag-Pol. Poliproteina Gag-Pol zawiera białka enzymatyczne: PR, RT/RH i IN. Gag Ty1 jest cięty potranslacyjnie na Gag-p45 i peptyd p4.

Cykl replikacyjny retrotranspozonu to proces, dzięki któremu retroelement wstawia swoją kopię w nowym miejscu w genomie gospodarza. Rozpoczyna się on w jądrze komórkowym od syntezy RNA na matrycy DNA Ty1 zintegrowanego z genomem gospodarza (Rys. 5.3). Transkrypcja Ty1 prowadzona jest przez polimerazę RNA II i regulowana przez, co najmniej dziewięć komórkowych czynników transkrypcyjnych (tj. Gcr1, Ste12, Tec1, Mcm1, Tea1/Ibf1, Rap1, Gcn4, Mot3 i Tye7) oraz trzy kompleksy remodelujące chromatynę (Swi/Snf, SAGA i ISWI) [14]. Nowopowstały RNA Ty1 zawiera rejon R na obu końcach cząsteczki oraz rejon U5 na końcu 5' i U3 na końcu 3' (Rys. 5.2). RNA Ty1 nie podlega splicingowi, ale udokumentowano obecność w komórce również krótszych, subgenomowych transkryptów Ty1 (Ty1i) [16]. Pomimo faktu, że retrotranspozycja Ty1 jest zjawiskiem rzadkim i tylko część transkryptów Ty1 posiada ogon poli A, to mogą one stanowić nawet do 10% poliadenylowanego mRNA w komórkach *S. cerevisiae* [17]. Najprawdopodobniej wynika to ze znacznie dłuższego okresu

półtrwania (~5h) RNA Ty1 w porównaniu z drożdżowymi mRNA [18]. Wykazano, że istotne dla stabilności RNA retrotranspozonu w komórce są oddziaływania z białkiem Gag Ty1, które dodatkowo uczestniczy w transporcie RNA Ty1 z jądra komórkowego do cytoplazmy (rozdział 5.4.2) [19]. W cytoplazmie transkrypty Ty1 służą jako matryce do translacji poliprotein Gag-p49 oraz Gag-Pol-p199 (Rys. 5.2) [14]. Gag-Pol powstaje w wyniku przesunięcia ramki odczytu, które jest skutkiem pauz translacyjnych z powodu obecności rzadkich kodonów tRNA [20,21]. Wydajność produkcji Gag-Pol szacowana jest między 3% a 20% w porównaniu do Gag-p49 w zależności od stosowanej metody pomiaru i warunków doświadczalnych [20,22,23].



Rysunek 5.3. Schemat cyklu replikacyjnego retrotranspozonu Ty1. Element Ty1 ulega transkrypcji prowadzonej przez maszynę komórki gospodarza. Potem RNA Ty1 jest transportowane do cytoplazmy i poddawane translacji do Gag i Gag-Pol. RNA Ty1, Gag i Gag-Pol kumulują się w cytoplazmie, tworząc retrosomy, w których gRNA Ty1 i tRNA^{Met} są pakowane do VLPs. Cięcia proteolityczne białek Gag i Gag-Pol przez proteazę Ty1 prowadzi do dojrzewania VLPs. Następnie RNA Ty1 podlega odwrotnej transkrypcji, a powstałe cDNA ulega integracji z genomem gospodarza.

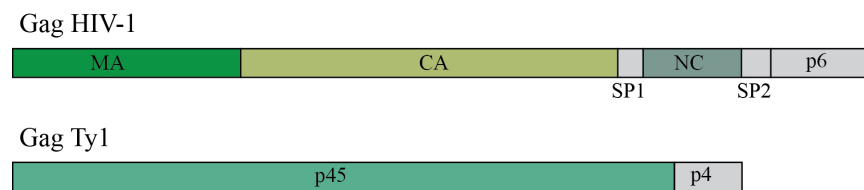
RNA i białka retrotranspozonu kumulują się w cytoplazmie, tworząc retrosomy (zwane inaczej ciążkami T), które są miejscem składania cząstek wirusopodobnych [24,25]. Niedojrzałe VLPs zawierają białka prekursorowe Gag-p49 i Gag-Pol, dwie kopie genomowego RNA Ty1 w formie dimeru, komórkowy tRNA^{Met}, a także pewną ilość innych komórkowych RNA i białek [14]. Podczas dojrzewania VLPs z poliproteiny Gag-Pol autokatalitycznie wycinana jest proteaza PR-p20, która katalizuje wszystkie kolejne cięcia białek prekursorowych. W wyniku cięcia Gag-p49 powstają Gag-p45 i p4, a z Gag-Pol wycinane są dodatkowo integraza IN-p71 oraz odwrotna transkryptaza RT/RH-p63 [26,27]. Następnie zachodzi reakcja odwrotnej transkrypcji RNA Ty1, której etapy są analogiczne do odwrotnej transkrypcji retrowirusowego RNA

w wirionach [28,29]. Powstały cDNA staje się częścią kompleksu nukleoproteinowego, przypominającego kompleks preintegracyjny opisany dla retrowirusów, i jest transportowany do jądra komórkowego. Proces retrotranspozycji kończy się włączeniem cDNA do genomu komórki za pośrednictwem IN lub rzadziej na drodze homologicznej rekombinacji [30]. Specyficznym miejscem integracji Ty1 z chromosomalnym DNA gospodarza jest rejon poprzedzający geny transkrybowane przez polimerazę RNA III [31,32].

5.4. Rola białka Gag Ty1 w cyklu replikacyjnym retrotranspozonu

5.4.1. Gag Ty1 jako białko strukturalne. Budowa VLP.

Retrowirusowe białka Gag zawierają trzy główne domeny, pełniące różne funkcje podczas składowania wirionów: domenę macierzy (MA) zaangażowaną w kierowanie nowopowstającego wirionu do błony komórkowej, domenę kapsydu (CA) pośredniczącą w interakcjach białko/białko oraz domenę nukleokapsydu (NC), która zawiera jeden lub dwa motywy palca cynkowego, umożliwiające specyficzne wiązanie wirusowego RNA (Rys. 5.4) [33]. Podczas dojrzewania wirionu Gag ulega wielokrotnemu cięciu proteolitycznemu w wyniku, którego powstają dojrzałe białka nazywane analogicznie do domen Gag. W dojrzałym wirionie MA znajduje się bezpośrednio pod osłonką, CA w środkowej jego części, formując kapsyd, a NC wiąże wirusowy RNA, tworząc nukleokapsyd zlokalizowany w wewnętrznej części wirionu [34].



Rysunek 5.4. Organizacja białek Gag HIV-1 i Ty1. Gag HIV-1 zawiera domeny macierzy (MA), kapsydu (CA) i nukleokapsydu (NC) oraz peptydy SP1, SP2 i p6. Gag Ty1 składa się z Gag-p45 i krótkiego peptydu p4.

Cząstki wirusopodobne Ty1 to kuliste, bądź owalne struktury o wielkości ~14 MDa, a ich średnica wynosi około 30 - 80 nm. VLPs retrotranspozonu Ty1 nie posiadają osłonki białkowo-lipidowej, a ich budowa przypomina rdzeń retrowirusowego wirionu [35]. Zbudowane są głównie z Gag Ty1, ale białko to nie wykazuje żadnej oczywistej homologii sekwencyjnej do swojego wirusowego odpowiednika. Gag Ty1 nie posiada domen MA, CA i NC ani motywu strukturalnego palca cynkowego (Rys. 5.4). Prekursorowy Gag-p49 (440 aa) ulega tylko pojedynczemu cięciu proteolitycznemu blisko C-końca do głównego dojrzałego białka strukturalnego Gag-p45 (401 aa) oraz peptydu p4 (39 aa) [14]. Funkcja peptydu p4 nie jest poznana. Zaproponowano jednak, że p4 może odgrywać pewną rolę w morfogenezie cząstek wirusopodobnych [36]. Wykorzystując eksperymenty *in vivo* i wprowadzając delecje w sekwencji kodującej Gag Ty1, wyznaczano rejony białka (aa E41-A62/P114-E147/T223-E287/R330-Q350), które są ko-

nieczne do jego wzajemnego oddziaływania i zachowania zdolności składania struktury VLPs [37,38]. Wprowadzenie mutacji w jednym miejscu nie przyczyniało się jednak do całkowitego zahamowania formowania VLPs, co może świadczyć o elastyczności tego procesu oraz o zaangażowaniu kilku rejonów Gag w oddziaływania białko/białko. Regularny kształt VLPs Ty1 uzyskiwany jest dzięki odpowiedniej orientacji białka Gag. Jest ono usytuowane tak, że rejon N-końca tworzy powierzchnię kapsydu, natomiast, wiążący RNA rejon C-końcowy, skierowany jest do wewnętrznej części VLP [39].

5.4.2. Rola białka Gag Ty1 w transporcie gRNA Ty1 do retrosomów

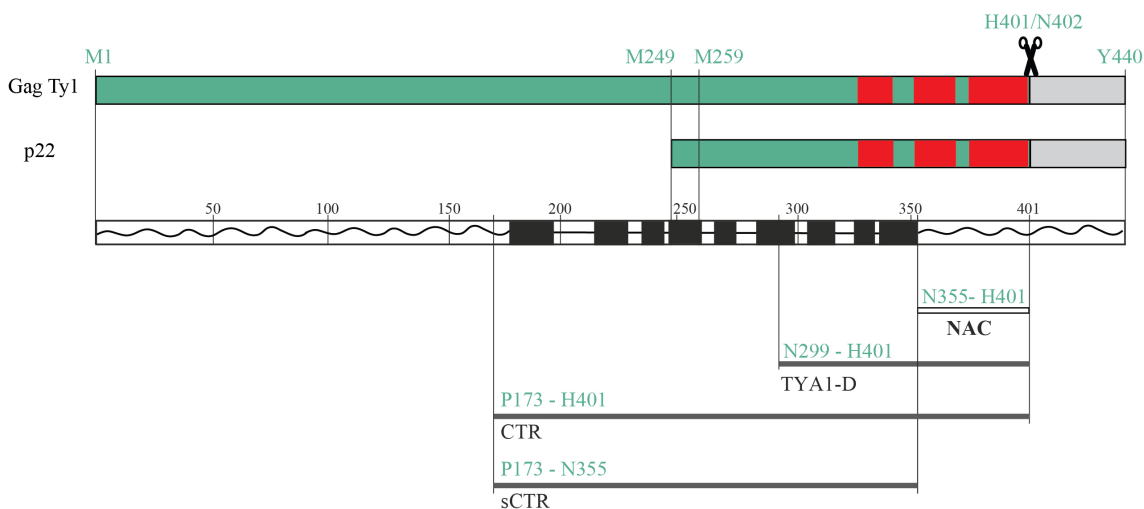
Składanie VLPs poprzedza formowanie retrosomów, które są skupiskami białek i RNA Ty1 w przestrzeni cytoplazmatycznej [24,25]. Mutacje w wiążącej RNA C-końcowej domenie Gag zakłócają tworzenie retrosomów, wskazując, że proces ten wymaga bezpośredniej interakcji między Gag a gRNA Ty1 [19,25]. Wpływ na kolokalizację RNA Ty1 i Gag mają również komórkowe komponenty ciałek P kodowane przez geny *dhh1*, *kem1*, *lsm1* i *pat1* [40]. Nie jest do końca jasne, gdzie zachodzi pierwsza interakcja RNA Ty1 z Gag. W przypadku nieobecności Gag, gRNA Ty1 akumuluje się w jądrze komórkowym i staje się niestabilny, natomiast ekspresja Gag *in trans* przywraca lokalizację RNA Ty1 w retrosomach [19]. Sygnał lokalizacji jądrowej nie został zidentyfikowany w sekwencji Gag Ty1 i nie ma dowodów na to, aby białko to wchodziło do jądra komórkowego. Niemniej jednak zaobserwowano, że Gag Ty1 gromadzi się na obrzeżach jądra komórkowego i wzmacnia eksport jądrowy gRNA, zachodzący przy użyciu szlaku Mex67p [19,24]. Inne badania wskazują, że podczas translacji Gag Ty1 jest kierowany do retikulum endoplazmatycznego (ER) przez cząstkę rozpoznającą sekwencję sygnałową (SRP, ang. *signal recognition particle*) [41], czyli rybonukleoproteinę, będącą częścią ścieżki sekrecyjnej w sortowaniu białek [42]. SRP oddziałuje z rybosomem i specyficznymi sekwencjami w powstającym polipeptydzie Gag, przejściowo wstrzymuje translację i skierowuje kompleks do błony ER, gdzie powstający dalej peptyd jest translokowany do światła ER. W świetle ER Gag Ty1 prawdopodobnie przyjmuje stabilną strukturę przestrzenną, a następnie ulega retrotranslokacji do cytoplazmy, gdzie wiąże się z podlegającym translacji gRNA Ty1. Multimeryzacja Gag związanego z gRNA Ty1 może hamować translację i indukować pakowanie RNA do VLPs. W przypadku braku translokacji do ER, Gag jest syntetyzowany, ale szybko degradowany, a retrosomy nie powstają [41].

5.4.3. Gag Ty1 jako białko opiekuńcze wobec kwasów nukleinowych

Poza oczywistą rolą strukturalną w dojrzałym wirionie, białka Gag retrowirusów pełnią istotne funkcje regulatorowe w replikacji [43]. Uważa się, że wiele funkcji białek Gag oraz białek nukleokapsydu w replikacji retrowirusa wynika z ich aktywności opiekuńczej względem kwasów nukleinowych (aktywności NAC, ang. *nucleic acid chaperone activity*) [44,45]. Jako białka opiekuńcze odgrywają one istotną rolę w dimeryzacji genomowego RNA oraz przyłączaniu

starterowego tRNA. Ponadto, dojrzałe białko NC promuje transfery nici kwasów nukleinowych podczas odwrotnej transkrypcji [44,45].

Proponuje się, że Gag Ty1 jest również białkiem opiekuńczym i pełni podobne funkcje regulatorowe podczas replikacji retrotranspozonu. Na podstawie badań *in vitro* chemicznie zsyntetyzowanego peptydu o długości 103 reszt aminokwasowych (TYA1-D), zaproponowano, że za wiązanie RNA oraz aktywność NAC odpowiedzialny jest rejon C-końcowy Gag Ty1 (aa N299–H401), zawierający trzy zgrupowania aminokwasów zasadowych (Rys. 5.5) [46]. Peptyd TYA1-D zdolny był do wiązania zarówno DNA, jak i RNA *in vitro* oraz do formowania nukleoproteinowych kompleksów o wysokiej masie cząsteczkowej [46].



Rysunek 5.5. Schemat białka Gag Ty1 oraz p22. Na schemacie zaznaczono miejsce cięcia przez proteazę Ty1 oraz trzy zgrupowania aminokwasów zasadowych (czerwone prostokąty). Poniżej zaznaczono przewidziane za pomocą XtalPred rejony nieuporządkowane (linia) oraz rejony α -helikalne (czarne prostokąty) Gag Ty1 [47]. Domena istotna dla aktywności NAC znajduje się w rejonie nieuporządkowanym i obejmuje 46 reszt aminokwasowych (N355-H401) [48]. Poniżej przedstawiono skrócone formy białka Gag Ty1: TYA1-D, CTR, sCTR.

W eksperymentach *in vitro* TYA1-D pozytywnie wpływał na przyłączenie $tRNA_i^{Met}$ do rejonu PBS w cząsteczce RNA Ty1 oraz wspomagał jego elongację przez odwrotną transkryptazę. Ponadto, TYA1-D promował dimeryzację RNA Ty1, ale tylko w obecności $tRNA_i^{Met}$. W naszym laboratorium również badaliśmy aktywność opiekuńczą Gag Ty1, stosując rekombinowane białka, odpowiadające różnym jego rejonom [47]. Białko CTR (ang. *C-terminal region*), obejmujące 228 reszt z C-końca Gag Ty1 i zawierające wszystkie trzy zgrupowania aminokwasów zasadowych, wykazywało silną aktywność opiekuńczą porównywalną z peptydem TYA1-D, podczas gdy skrócona forma CTR (sCTR ang. *short C-terminal region*), zawierająca tylko pierwszą grupę aminokwasów zasadowych, nie była aktywna. Wykorzystując analizy komputerowe ustaliliśmy, że tylko drugie i trzecie zgrupowanie aminokwasów zasadowych znajduje się w rejonie o nieuporządkowanej strukturze (Rys. 5.5) [47]. Badania innych białek opiekuńczych wskazują, że zasadowe reszty aminokwasowe z rejonów nieuporządkowanych, są

szczególnie istotne dla aktywności NAC [49,50]. Nasze obserwacje potwierdzają najnowsze analizy strukturalne, biofizyczne i genetyczne, wskazujące, że rejon nieuporządkowany N355-H401 jest właściwą domeną odpowiedzialną za aktywność NAC białka Gag Ty1 [48].

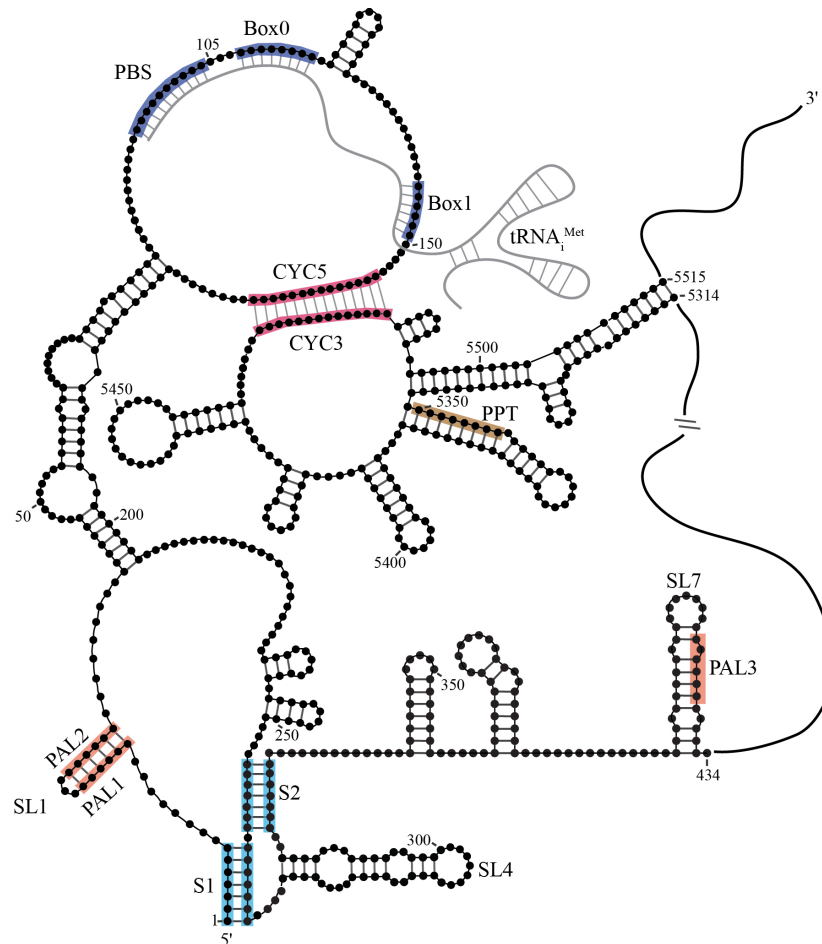
5.4.4. *Gag Ty1 a mechanizm kontroli liczby kopii retrotranspozonu*

Gen *gag* Ty1 oprócz białka Gag koduje czynnik restrykcyjny p22, który jest kluczowym i niezbędnym elementem w kontroli liczby kopii retrotranspozonu (CNC, ang. *copy number control*) w genomie *S. cerevisiae* [16]. p22 ulega translacji z subgenomowego mRNA Ty1 (Ty1i), a jego sekwencja jest identyczna z sekwencją C-końcowej części Gag-p49 (aa M249/M259-Y440; Rys. 5.5). Wykazano, że kontrola liczby kopii retrotranspozonu jest sumarycznym efektem negatywnego wpływu p22 na różne etapy replikacji Ty1, a inhibicja retrotranspozycji wynika najprawdopodobniej z bezpośredniego oddziaływania pomiędzy p22 i Gag [16,47,51]. Poprzez kolokalizację białek p22 i Gag Ty1 zaburzone jest formowanie retrosomów. Czynnik restrykcyjny wpływa też na wydajność tworzenia cząstek wirusopodobnych, ich kompozycję białkową oraz morfologię. Na podstawie analiz strukturalnych i biofizycznych zaproponowano mechanizm, prowadzący do tych zaburzeń, według którego p22 wiąże się z domeną C-końcową Gag Ty1 i w konsekwencji zaburza prawidłowe składanie VLPs, ponieważ nie posiada obecnego w Gag Ty1 rejonu koniecznego do dalszych oddziaływań białko/białko [48].

5.5. Oddziaływania RNA/RNA istotne dla retrotranspozycji Ty1

Badania, mające na celu ustalenie minimalnej sekwencji elementu Ty1 wymaganej do pełnego cyklu replikacyjnego wykazały, że 580 nt z końca 5' gRNA i 357 nt z jego końca 3' jest wystarczające do wydajnej retrotranspozycji w przypadku koekspresji z elementem pomocniczym, kodującym białka retrotranspozonu [52]. Wyznaczone rejony genomowego RNA Ty1 zawierają sekwencje, które pełnią istotne funkcje regulatorowe w replikacji (Rys. 5.6). Na końcu 5' znajduje się domena PBS (ang. *primer binding site*), wiążąca komórkowy tRNA_i^{Met} wykorzystywany jako starter dla odwrotnej transkryptazy. PBS to 10-nukleotydowa sekwencja komplementarna do ramienia akceptorowego tRNA_i^{Met} [53]. Dalsze badania dostarczyły dowodów na obecność dodatkowych sekwencji zlokalizowanych poniżej PBS, które są niezbędne do pakowania tRNA_i^{Met} do VLPs: Box0, Box1 i Box2.1 [54,55]. Te sekwencje są komplementarne do ramion TΨC oraz DHU cząsteczki tRNA_i^{Met} i umożliwiają nie tylko jej pakowanie, ale także odgrywają rolę w inicjacji odwrotnej transkrypcji *in vivo* [56]. Natomiast wyselekcjonowany koniec 3' gRNA Ty1 zawiera sekwencję PPT (ang. *polypurine tract*), która służy jako starter do syntezy nici (+)DNA podczas odwrotnej transkrypcji [57]. Wykazano również, że oddziaływania dalekiego zasięgu pomiędzy rejonem CYC5, sąsiadującym z PBS a komplementarnym rejonem CYC3 na końcu 3' gRNA wspomagają proces syntezy cDNA oraz samej retrotranspozycji [58]. W wyniku tych oddziaływań dochodzi do cyklizacji gRNA Ty1, która prowadzi do przestrzennego zbliżenia jego końców 5' i 3', ułatwiając w ten sposób transfer nici (-)DNA podczas od-

wrotnej transkrypcji. Sekwencja CYC5 jest jednocześnie komplementarna do ramienia DHU cząsteczki tRNA^{Met} (zawiera Box2.1), co stwarza możliwość dodatkowych interakcji. Innym oddziaływaniem dalekiego zasięgu, które jest istotne dla retrotranspozycji Ty1, m.in. dla wydajnej inicjacji odwrotnej transkrypcji, jest parowanie zasad między sekwencjami 1GAGGAGA7 i 264UCUCCUC270 [59].



Rysunek 5.6. Schemat oddziaływań RNA/RNA, zachodzących podczas replikacji retrotranspononu Ty1 na podstawie struktury RNA Ty1 *in vivo* [60]. Dla uproszczenia przedstawiono tylko fragmenty gRNA Ty1 z końca 5' i 3'. Sekwencje istotne dla reakcji odwrotnej transkrypcji są zaznaczone kolorem granatowym (PBS, Box0, Box1) i brązowym (PPT). Miejsce cyklizacji gRNA Ty1 zaznaczono kolorem malinowym. Trzony pseudowęzła (S1 i S2) zaznaczono na niebiesko. Na rysunku zaznaczono spinki SL1, SL4 i SL7, zawierające sekwencje potencjalnie zaangażowane w formowanie dimeru gRNA Ty1.

Modele struktury drugorzędowej gRNA Ty1 opracowane na podstawie próbkowania struktury RNA metodą SHAPE (ang. *selective 2'-hydroxyl acylation and primer extension*) *in vitro*, *in virio*, *ex virio* oraz *in vivo* [60,61] wspierają wiele aspektów proponowanych wcześniej oddziaływań, w tym parowanie tRNA^{Met} z rejonami PBS, Box0 i Box1 oraz cyklizację RNA Ty1 poprzez interakcję CYC5/CYC3 (Rys. 5.6). Badania struktury gRNA Ty1 *in vivo* wskazują, że do tych oddziaływań dochodzi w cytoplazmie lub retrosomach, jeszcze przed spa-

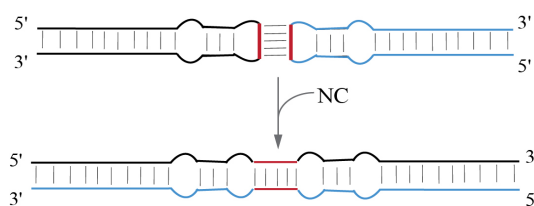
kowaniem gRNA Ty1 do VLPs [60]. Funkcjonalnie zdefiniowane parowanie nukleotydów 1-7 z nukleotydami 264-270 zostało zidentyfikowane jako trzon S1 struktury pseudowężła, stabilizującej koniec 5' RNA Ty1 (Rys. 5.6) [60-62]. Pseudowęzeł w RNA Ty1 zawiązuje się między rejonem jednoniciowym a pętlą wewnętrzną tak, że reszty nukleotydowe 1-7 oraz 256-262 oddziałują odpowiednio z resztami 264-270 oraz 319-325. Tym samym obejmuje on rejon ponad 320 reszt nukleotydowych.

Postuluje się, że w obrębie pierwszych 580 nukleotydów RNA Ty1 znajdują się także sekwencje istotne dla procesów dimeryzacji i pakowania genomowego RNA do VLPs. Ten temat został dokładnie omówiony w rozdziale 5.6.2.

5.6. Procesy dimeryzacji i pakowania genomowego RNA u retroelementów

5.6.1. Dimeryzacja i pakowanie genomowego RNA u retrowirusów

Podczas replikacji retrowirusów do wirionów pakowane są dwie kopie pełnej długości gRNA w postaci niekowalencyjnie połączonego dimeru [63,64]. Genomowy RNA wirusa wybierany jest do pakowania ze środowiska komórkowego, które zawiera znaczny nadmiar komórkowych oraz poddanych splicingowi wirusowych mRNA. Istotną rolę w selekcji wirusowego RNA odgrywają interakcje między domenami NC poliprotein Gag a specyficznymi sekwencjami w gRNA wirusa [43,65]. Sekwencje te nazywane są sygnałami pakowania (Psi, ang. *packaging signal*) i zazwyczaj znajdują się w 5' UTR gRNA lub czasem obejmują również rejon kodujący. U retrowirusów sygnały pakowania nakładają się lub sąsiadują z sekwencjami zaangażowanymi w dimeryzację gRNA i procesy te są ze sobą ściśle powiązane [64]. Rejony gRNA istotne dla dimeryzacji i pakowania charakteryzują się silnym ustrukturyzowaniem z licznymi motywami spinkowymi [66,67]. Pętle apikalne niektórych spinek zawierają krótkie sekwencje palindromowe, umożliwiające międzycząsteczkowe parowanie zasad, tworząc w ten sposób kompleksy *kissing-loop* (lub inaczej *loose dimers*; Rys. 5.7). Sugeruje się, że utworzenie dimeru może hamować translację RNA wirusa i kierować je do spakowania [68-70]. Wykazano, że wirusowy gRNA w formie dimerycznej jest pakowany preferencyjnie w stosunku do monomerów tego RNA [71,72]. Proponuje się, że podczas dojrzewania wirionu nowopowstałe białko NC indukuje zmiany strukturalne w dimerze



Rysunek 5.7. Schemat mechanizmu dimeryzacji gRNA retrowirusów, przedstawiający kompleks *kissing-loop* i jego przejście do bardziej stabilnej formy rozszerzonego międzycząsteczkowego dupleksu w wyniku działania białka NC.

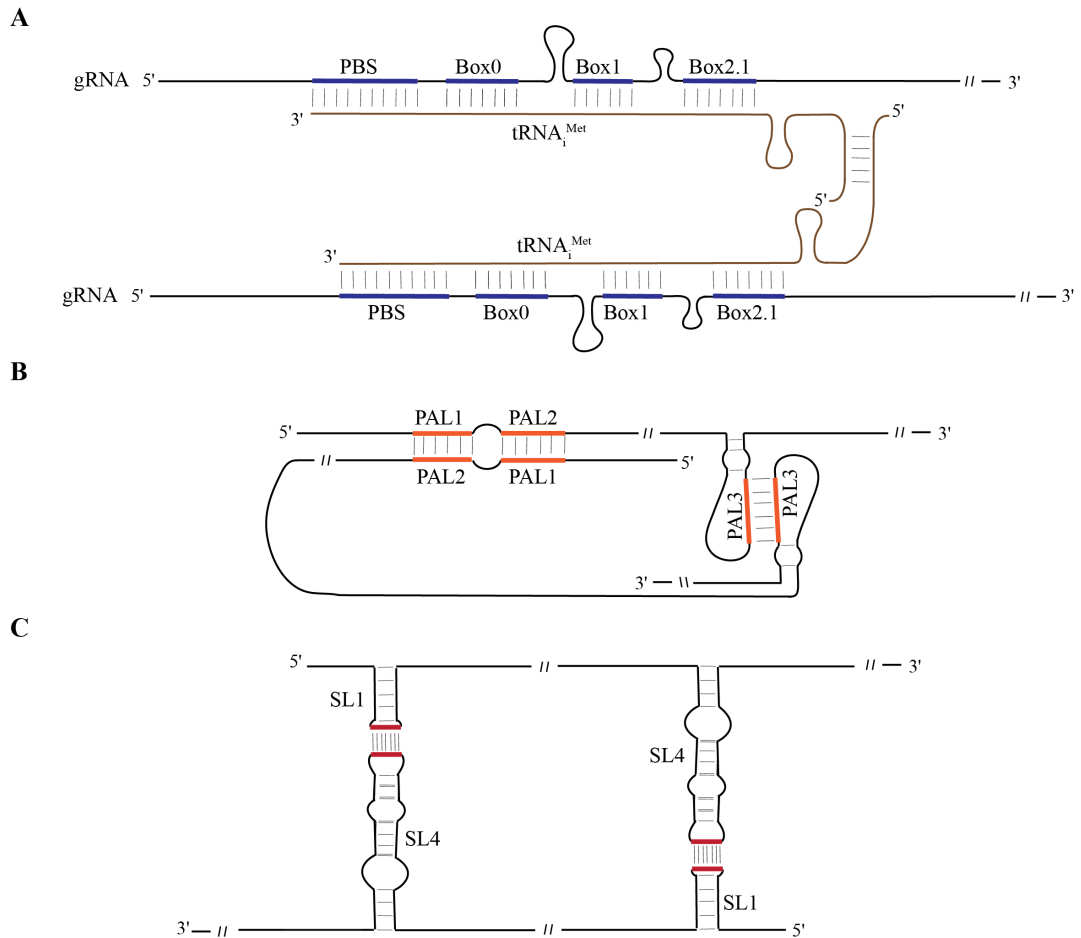
RNA, które rozszerzają międzycząsteczkowe parowanie zasad, tworząc stabilniejszą formę dimeru tzw. *tight dimer* (Rys. 5.7) [67]. Podczas odwrotnej transkrypcji dimeryczny genom ułatwia przeniesienie syntetyzowanej nici cDNA między jedną cząsteczką gRNA a drugą, promu-

jąc różnorodność genetyczną i ewolucję szczepów opornych na leki i takich, które mogą uniknąć odpowiedzi immunologicznej [73,74].

5.6.2. Dimeryzacja i pakowanie genomowego RNA retrotranspozonu Ty1

Sekwencje Psi dla elementu Ty1 nie są precyzyjnie zdefiniowane, jednakże dostępne dane wskazują, że rejon istotny dla pakowania gRNA znajduje się w obrębie nukleotydów 230-580 [52]. Wiadomo również, że pakowanie gRNA Ty1 do VLPs jest promowane przez białko Gag [75], a gRNA Ty1 wewnątrz VLP ma formę dimeryczną [76]. Jednak mechanizm dimeryzacji i pakowania gRNA Ty1 pozostaje niejasny. Nie zostało też rozstrzygnięte, czy między pakowaniem RNA do VLPs a jego dimeryzacją występuje podobna zależność, jak u retrovirusów.

Pierwszy opisany w literaturze model dimeru gRNA Ty1 opiera się na wynikach eksperymentów dimeryzacji *in vitro* z wykorzystaniem transkryptu, reprezentującego rejon końca 5' gRNA Ty1 oraz syntetycznego peptydu TYA1-D (rozdział 5.4.3) [46]. Zaobserwowano, że TYA1-D promuje dimeryzację RNA Ty1 tylko w obecności tRNA_i^{Met} i na tej podstawie zaproponowano, że dimer utrzymywany jest przez parowanie reszt nukleotydowych między dwiema cząsteczkami tRNA_i^{Met} związanymi jednocześnie z sekwencjami PBS, Box0, Box1 i Box2.1 dwóch cząsteczek gRNA Ty1 (Rys. 5.8A). Późniejsze badania sugerują jednak, że analogicznie do retrovirusów, podczas dimeryzacji dochodzi do bezpośrednich oddziaływań między dwiema cząsteczkami genomowego RNA Ty1. Na podstawie analiz strukturalnych gRNA Ty1 *in virio* oraz w formie monomerycznej (*ex virio* oraz *in vitro*) zasugerowano, że w interakcje międzycząsteczkowe w dimerze RNA mogą być zaangażowane trzy sekwencje palindromowe PAL1 (15GUAUA20), PAL2 (24UAUAC29) i PAL3 (423CCUGGG428), które tworzą dwa oddziaływania PAL1/PAL2 i jedno PAL3/PAL3 (Rys. 5.8B) [61]. Według innej hipotezy proces dimeryzacji RNA Ty1 inicjowany jest poprzez tworzenie dimeru typu *loose dimer* [77]. Autorzy modelu zauważyli komplementarność 7-nukleotydowych sekwencji w spinkach SL1a (9AUUCUGU25) i SL3a (293ACAGAAU299) (SL1 i SL4 zgodnie z nazewnictwem stosowanym w prezentowanym cyklu prac naukowych). Mutacje, zaburzające tę komplementarność, zmniejszały stabilność RNA Ty1 *in vivo* i znacząco hamowały retrotranspozycję, a mutacje kompensacyjne niwelowały negatywny wpływ pierwszych substytucji. Na tej podstawie zasugerowano, że inicjacja dimeryzacji gRNA Ty1 odbywa się przez utworzenie dwóch oddziaływań *kissing-loop* między spinkami SL1 i SL4 (Rys. 5.8C). Rola sekwencji palindromowych oraz interakcji SL1/SL4 nie była jednak sprawdzona bezpośrednio w eksperymentach dimeryzacji.



Rysunek 5.8. Schematy proponowanych mechanizmów dimeryzacji RNA Ty1. **(A)** Dimer gRNA Ty1 tworzony jest pośrednio przez oddziaływanie dwóch cząsteczek $tRNA_i^{Met}$ przyhybrydowanych w rejonie PBS. **(B)** Dimer RNA Ty1 formowany jest przez oddziaływanie trzech sekwencji palindromowych PAL1, PAL2 i PAL3. **(C)** Formowanie dimeru inicjowane jest przez utworzenie dwóch oddziaływań *kissing-loop* między pętlami apikalnymi spinek SL1 i SL4.

6. CEL NAUKOWY PRZEDSTAWIONEGO CYKLU PRAC

Głównym celem przedstawionego cyklu prac naukowych było rozpoznanie mechanizmów, regulujących dimeryzację i pakowanie genomowego RNA retrotranspozonów LTR, w tym zbadanie roli białka Gag w tych procesach. Modelem badawczym był drożdżowy retrotranspozon Ty1. Cele szczegółowe moich badań pokrywają się z tematyką trzech publikacji wchodzących w skład cyklu.

1. Celem pracy **Gumna et al. *RNA Biology*, 2019** było wytypowanie sekwencji zaangażowanych w oddziaływanie międzycząsteczkowe podczas dimeryzacji gRNA Ty1 oraz określenie roli białka Gag w tym procesie. Dążyłam również do zidentyfikowania rearanżacji strukturalnych, zachodzących podczas dimeryzacji RNA Ty1 oraz ustalenia struktury dimeru tego RNA.

2. Badania przedstawione w publikacji **Gumna et al. *IJMS*, 2021** były ukierunkowane na analizę oddziaływań białek Gag Ty1 z RNA. Szczególnie interesował mnie mechanizm rozpoznawania gRNA retrotranspozonu, który jest kluczowy dla jego selektywnego pakowania do cząstek wirusopodobnych. Moim kolejnym celem było zbadanie wpływu specyficznych motywów strukturalnych na oddziaływanie RNA Ty1 z Gag.

3. Praca **Gumna et al. *PLoS One*, 2020** jest efektem mojego zaangażowania w rozwój bioinformatycznych narzędzi do analizy danych z eksperymentów mapowania struktury drugorzędowej RNA. Za cel postawiłam sobie zaproponowanie automatycznej metody normalizacji i kontroli jakości danych otrzymanych z eksperymentów próbkowania strukturalnego RNA z wykorzystaniem elektroforezy kapilarnej.

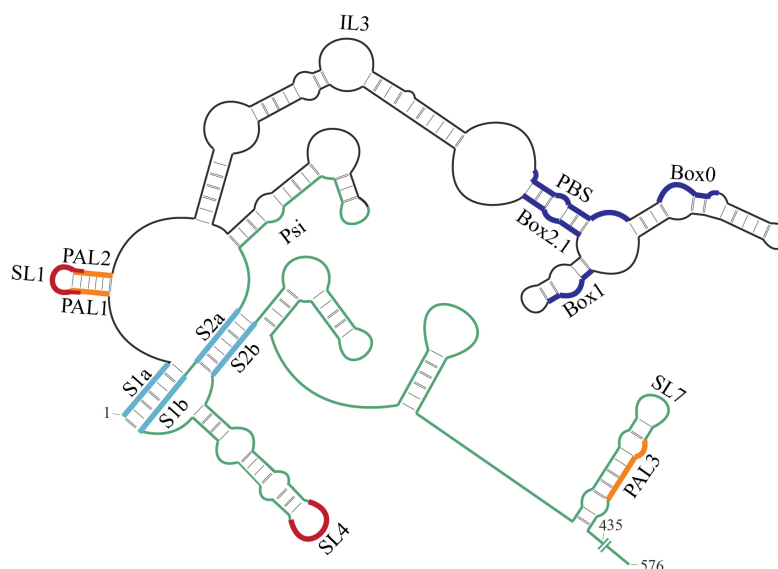
7. SKRÓTOWY OPIS PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

7.1. Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA

Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel, Katarzyna Pachulska-Wieczorek;

RNA Biology, 2019; 16: 1749-1763

Kiedy rozpoczynaliśmy badania przedstawione w powyższej publikacji, mechanizm dimeryzacji genomowego RNA Ty1 był wciąż niejasny. W literaturze funkcjonowało kilka modeli dime-ru RNA Ty1, zakładających zaangażowanie różnych sekwencji nukleotydowych w oddziaływa- nia międzycząsteczkowe [46,61,77]. Celem pracy było wyjaśnienie w, jaki sposób dochodzi do dimeryzacji RNA Ty1, zdefiniowanie kluczowych sekwencji nukleotydowych i określenie roli białka Gag Ty1 w tym procesie. W eksperymentach dimeryzacji RNA Ty1 *in vitro* wykorzysta- łąm RNA mini Ty1 (mTy1), który reprezentuje rejon 5' gRNA retrotranspozonu Ty1 (+1-576 nt). RNA mTy1 zawiera elementy kluczowe dla efektywnej retrotranspozycji, w tym sekwencje proponowane wcześniej jako istotne dla dimeryzacji i pakowania RNA Ty1 do czą- stek wirusopodobnych (Rys. 7.1) [46,52,55,58,59,61,77].



Rysunek 7.1. Schemat struktury drugorzędowej monomeru RNA mTy1 (+1-576) *in vitro*. Miejsca potencjalnie zaangażowane w tworzenie dime-ru zaznaczone są kolorem pomarańczowym (sekwencje palindromowe PAL1, PAL2 w SL1 i PAL3 w SL7) oraz czerwonym (sekwencje komplementarne w spinkach SL1 i SL4). RNA mTy1 zawiera również miejsce wiązania starterowego tRNA (PBS, Box0, Box1, Box2.1) zaznaczone kolorem granatowym. Rejon wymagany dla pakowania RNA Ty1 (Psi) do VLPs zaznaczony jest kolorem zielonym. Sekwencje, tworzące trzony pseudowęzła (S1 i S2) zaznaczone są na niebiesko.

RNA mTy1 zwijany *in vitro* przyjmuje prawidłową (natywną) strukturę drugorzędową [62] zgodną ze strukturą *in virio* [61] oraz *in vivo* [60]. Wykorzystanie krótszych transkryptów, reprezentujących wybrane fragmenty gRNA retroelementów jest powszechnym podejściem w badaniach procesów dimeryzacji, ponieważ pełnej długości gRNA tworzy dimery w warunkach *in vitro* bardzo niewydajnie lub wcale [78-81].

7.1.1. Rola białka Gag w dimeryzacji RNA Ty1

U retrowirusów dimeryzacja genomowego RNA promowana jest przez poliproteinę Gag lub białko NC, jednak retrowirusowe transkrypty, zawierające sekwencję 5' UTR zdolne są do spontanicznego tworzenia dimerów *in vitro* również przy braku czynników białkowych [78,79]. W przypadku RNA mTy1 nie dochodzi do spontanicznej dimeryzacji *in vitro*, ale pokazano, że proces ten może zachodzić w obecności peptydu TYA1-D lub innych krótszych form Gag Ty1, zawierających domenę wiążącą RNA [46,47]. Zdolność promowania dimeryzacji nie była jednak zbadana dla kompletnego białka Gag Ty1. W moich badaniach wykorzystywałam dojrzałą formę tego białka, czyli Gag-p45. Zaobserwowałam, że dimer RNA mTy1 tworzył się tylko w obecności białka, a wraz ze wzrostem jego stężenia wzrastała również wydajność procesu dimeryzacji (Rys. 2 w [1]). Maksymalna wydajność promowanej przez Gag-p45 dimeryzacji RNA mTy1 *in vitro* wynosiła około 50% i była ona znacznie wyższa niż w analogicznych eksperymentach z wykorzystaniem krótszych form Gag Ty1 [46,47].

7.1.2. Oddziaływania RNA/RNA istotne dla formowania dimeru RNA Ty1

W pierwszym etapie badań zweryfikowałam hipotezę, zakładającą, że dimer gRNA Ty1 nie powstaje przez bezpośrednie oddziaływanie cząsteczek gRNA Ty1, lecz poprzez oddziaływanie dwóch cząsteczek komórkowego tRNA_i^{Met} przyhybrydyzowanych do gRNA retrotranspozonu w rejonie PBS (Rys. 5.8A) [46]. W eksperymentach *in vitro* nie zaobserwowałam pozytywnego lub negatywnego wpływu tRNA_i^{Met} na wydajność indukowanej przez białko Gag-p45 dimeryzacji RNA mTy1. Można, zatem wnioskować, że dimer gRNA retrotranspozonu Ty1 jest tworzony podobnie jak u retrowirusów, przez bezpośrednie oddziaływanie dwóch cząsteczek gRNA.

Jednocześnie, w celu ustalenia, jakiego typu oddziaływanie utrzymuje dwie cząsteczki RNA mTy1 w dimerze, przeprowadziłam analizę tworzenia dimerów RNA w różnych warunkach elektroforetycznych, stosując bufor z jonami Mg²⁺ lub bez (odpowiednio TBM oraz TB). Mniej stabilne dimery RNA utrzymywane przez oddziaływanie *kissing-loop* są stabilizowane przez jony Mg²⁺ i można je wykryć jedynie w ich obecności. Podczas elektroforezy żelowej w buforze pozbawionym jonów Mg²⁺ dochodzi do ich dysocjacji i obserwuje się tylko trwalsze dimery powstałe przez tworzenie krótkich międzycząsteczkowych dupleksów [79,82,83]. W moich badaniach zaobserwowałam bardzo podobne ilości dimerów RNA mTy1 w buforze TB i TBM (Rys. 2 w [1]), co sugeruje, że dimer RNA retrotranspozonu Ty1 nie jest utrzymywa-

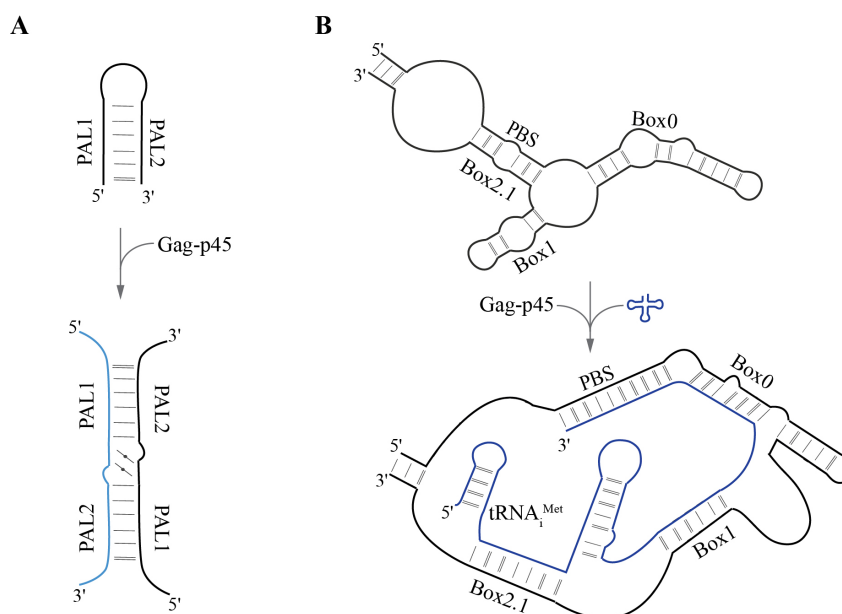
ny przez oddziaływania *kissing-loop*. Tym samym nie znalazłam potwierdzenia dla hipotezy, zakładającej, że inicjacja dimeryzacji RNA Ty1 odbywa się przez międzycząsteczkowe oddziaływania pomiędzy SL1 i SL4 (Rys. 5.8C) [77]. Wyniki kolejnych badań, które opisuję poniżej, również wykluczają tę hipotezę.

7.1.3. Analiza SHAPE monomeru i dimeru RNA mTy1

W kolejnym etapie badań analizowałam strukturę drugorzędową dimeru RNA mTy1. W tym celu wykorzystałam metodę SHAPE (dokładny opis metody znajduje się w rozdziale 7.3.1). Ponieważ dimeryzacja oraz przyłączenie starterowego tRNA, to dwa kluczowe oddziaływania RNA/RNA w cyklu replikacyjnym retrotranspozonu, podczas eksperymentów próbkowania strukturalnego w mieszaninie reakcyjnej obecne były także cząsteczki tRNA_i^{Met}. Maksymalna wydajność dimeryzacji promowanej białkiem Gag-p45 wynosiła około 50%, dlatego podczas analizy reaktywności SHAPE należało wziąć pod uwagę fakt, iż w roztworze RNA poddawanym modyfikacji chemicznej znajdowała się mieszanina monomeru i dimeru RNA. Dlatego zastosowałam matematyczną dekonwolucję danych opartą na założeniu, iż wartość reaktywności otrzymanych dla mieszaniny konformerów RNA jest równa sumie wartości reaktywności dla konformerów składowych (monomeru i dimeru), których waga jest zgodna z wkładem procentowym każdego z nich w całej populacji RNA [84]. Przeprowadzona analiza porównawcza uzyskanych w ten sposób reaktywności SHAPE pozwoliła mi wyznaczyć różnice strukturalne między monomerem a dimerem RNA, a co za tym idzie, wytypować potencjalne miejsca oddziaływań międzycząsteczkowych w dimerze RNA mTy1.

Ogólnie profile reaktywności SHAPE dla monomeru i dimeru RNA mTy1 były do siebie podobne, a zaobserwowane różnice miały charakter lokalny (Rys. 4a w [1]). Wskazuje to, że podczas dimeryzacji dochodzi do zmian strukturalnych tylko w miejscach potencjalnych kontaktów międzycząsteczkowych z zachowaniem motywów strukturalnych, które nie są zaangażowane w ten proces. Najsilniejszy spadek reaktywności SHAPE dla dimeru względem monomeru, zaobserwowałam dla reszt nukleotydowych 21UCU23, łączących sekwencje palindromowe PAL1 i PAL2 (Rys. 4a, c w [1]). W monomerze RNA mTy1 reszty nukleotydowe 21UCU23 są wysoce reaktywne, ponieważ tworzą pętlę apikalną spinki SL1, której trzon utworzony jest przez wewnątrzcząsteczkowe oddziaływanie pomiędzy PAL1 i PAL2 (Rys. 7.2A). Silny spadek reaktywności SHAPE dla 21UCU23 jest najprawdopodobniej wynikiem rozplecenia SL1 i utworzenia międzycząsteczkowego dupleksu poprzez dwa oddziaływania PAL1/PAL2. Ponieważ sekwencje PAL1 i PAL2 są sparowane również w monomerze RNA mTy1, zmiany reaktywności SHAPE w dimerze dla samych palindromów były znacznie mniejsze. Przewidywania struktury międzycząsteczkowego dupleksu z wykorzystaniem programów RNAComposer [85,86] oraz RNAVista [87] wykazały, że jest on tworzony nie tylko przez parowania typu Watsona-Cricka, lecz również przez dwie niekanoniczne pary C-U, angażujące nukleotydy U21

i C22 (Rys. 7.2A; Rys. 5 w [1]). Zaproponowany model dupleksu PAL1/PAL2 bardzo dobrze tłumaczy zmiany reaktywności SHAPE obserwowane dla dimeru RNA mTy1. Wcześniejsze badania, sugerowały, że potencjalne miejsca kontaktu w dimerze stanowią trzy sekwencje palindromowe: PAL1, PAL2 oraz PAL3 (Rys. 5.8B) [61]. W monomerze sekwencja PAL3 znajduje się w trzonie spinki SL7 (Rys. 7.1). Utworzenie oddziaływania PAL3/PAL3 powinno zatem spowodować silny wzrost reaktywności dla nukleotydów, tworzących drugą część trzonu tej spinki. Jednak w moich badaniach nie zaobserwowałam żadnych istotnych zmian w reaktywnościach SHAPE, które mogłyby świadczyć o utworzeniu międzycząsteczkowego oddziaływania PAL3/PAL3 w dimerze RNA mTy1 (Rys. 4a, c w [1]).



Rysunek 7.2. Schemat zmian strukturalnych wywołanych powstaniem oddziaływań RNA/RNA podczas procesów dimeryzacji RNA mTy1 (A) oraz przyłączania tRNA_i^{Met} (B) promowanych przez białko Gag-p45, zaproponowany na podstawie danych z eksperymentów próbkowania chemicznego RNA metodą SHAPE.

Zmian nie zaobserwowałam również dla spinki SL4. Reszty nukleotydowe w pętli apikalnej SL4 pozostawały wysoce reaktywne zarówno dla konformacji monomerycznej, jak i dimerycznej RNA mTy1 (Rys. 4a, c w [1]), co przeczy zaangażowaniu SL4 w dimery *kissing-loop*. Nie wielki, ale istotny statystycznie spadek reaktywności SHAPE dla dimeru w porównaniu do monomeru odnotowałam za to dla krótkiego palindromu 67AAUU70 w pętli wewnętrznej IL3.

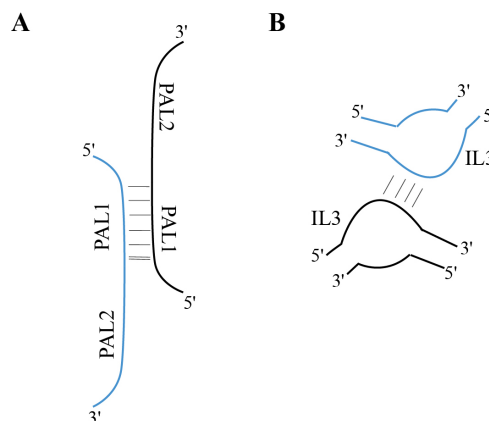
Różnice w reaktywnościach SHAPE dla dimeru i monomeru wskazują, że do zmian strukturalnych w RNA mTy1 doszło również w rejonie PBS (Rys. 4a w [1]). Kontrolne eksperymenty próbkowania dimeru RNA mTy1 wykonane bez dodatku tRNA_i^{Met} potwierdziły, że zmiany reaktywności SHAPE w tym rejonie nie są skutkiem dimeryzacji, lecz przyłączenia startowego tRNA. Moje badania potwierdzają, zatem rolę białka Gag Ty1 nie tylko w dimeryzacji RNA Ty1, ale też w promowaniu hybrydyzacji komórkowego tRNA_i^{Met} (Rys. 7.2B).

7.1.4. Wpływ mutacji w sekwencjach potencjalnie zaangażowanych w dimeryzację RNA Ty1 na efektywność tego procesu *in vitro*

W celu potwierdzenia roli sekwencji PAL w tworzeniu dimeru przeprowadziłam eksperymenty dimeryzacji *in vitro* cząsteczek RNA mTy1, zawierających mutacje w wytypowanych sekwencjach palindromowych (dalej nazywane: mutant PAL1, mutant PAL2 oraz mutant PAL3). Substytucje w PAL1 lub PAL2 obniżyły wydajność procesu dimeryzacji o około 50% (Rys. 6a w [1]), potwierdzając tym samym istotną rolę oddziaływań PAL1/PAL2 dla formowania dimeru. Dimeryzacja nie została jednak zahamowana całkowicie, co sugerowało obecność innych, dodatkowych miejsc kontaktu w dimerze. W celu ich zidentyfikowania, przeprowadziłam analizę struktury monomerów i dimerów mutantów RNA mTy1, stosując metodę SHAPE. Jedyną istotną zmianą zaobserwowaną dla dimeru mutantu PAL2 względem monomeru był bardzo silny spadek reaktywności SHAPE w palindromie PAL1 (Rys. 6c w [1]). W monomerze, na skutek zaburzenia komplementarności z sekwencją PAL2, PAL1 był wysoce reaktywny. Na tej podstawie zaproponowałam, że w tym przypadku doszło do utworzenia międzycząsteczkowego oddziaływania PAL1/PAL1 (Rys. 7.3A). Natomiast różnice w reaktywnościach SHAPE między monomerem a dimerem mutantu PAL1 wskazują, że mutacje w sekwencji PAL1 indukują dime-

ryzację przez oddziaływanie międzycząsteczkowe pomiędzy sekwencjami palindromowymi 67AAUU70 zlokalizowanymi w pętli wewnętrznej IL3 (Rys. 7.3B; Rys. 6c w [1]). Spadek reaktywności SHAPE w tym rejonie dla dimeru mutantu PAL1 był znacznie silniejszy niż dla dimeru RNA mTy1 typu dzikiego (Rys. 4a, c w [1]). Zaobserwowałam również, że substytucje w pętli IL3 (mutant IL3) obniżają wydajność dimeryzacji *in vitro*, ale w mniejszym stopniu niż mutacje w sekwencjach PAL1 lub PAL2 (Rys. 6b w [1]). Dla mutantu PAL3 nie zaobserwowałam zmian w wydajności procesu dimeryzacji w porównaniu do RNA mTy1 typu dzikiego (Rys. 6a w [1]), co potwierdziło, że PAL3 nie jest zaangażowany w formowanie dimeru RNA Ty1.

Ponadto, aby zweryfikować hipotezę, zakładającą, że dimer RNA Ty1 formowany jest przez oddziaływania *kissing-loop* między spinkami SL1 i SL4 [77], przeprowadziłam również eksperymenty z RNA mTy1, zawierającym mutacje w pętli apikalnej SL4 (mutant SL4). Mutacja ta nie wpłynęła jednak istotnie na wydajność dimeryzacji (Rys. 6b w [1]).



Rysunek 7.3. Schemat oddziaływań międzycząsteczkowych w dimerze RNA mutantu PAL2 (A) i PAL1 (B) zaproponowany na podstawie reaktywności SHAPE.

7.1.5. Rola struktury pseudowężła w dimeryzacji RNA Ty1

Analizy struktury gRNA Ty1 *in virio* i *in vivo* wykazały, że jego koniec 5' jest stabilizowany przez strukturę pseudowężła, który zawiązuje się dzięki utworzeniu dwóch trzonów S1 oraz S2 (Rys. 5.6) [60,61]. Pseudowężel jest też zachowany w strukturze RNA mTy1 (Rys. 7.1) [62], ale nie w pełnej długości gRNA Ty1 *in vitro* [60]. Mutacje, zaburzające strukturę pseudowężła, negatywnie wpływają na retrotranspozycję Ty1, wskazując na jego biologicznie ważną funkcję [62,77,88]. W celu zbadania roli pseudowężła w dimeryzacji, przeprowadziłam eksperymenty dimeryzacji *in vitro* RNA mTy1 z delecją pierwszych pięciu nukleotydów (mutant $\Delta S1a$). Mutacja ta uniemożliwia utworzenie trzonu S1, a tym samym powoduje destabilizację całego motywu pseudowężła (Rys. 7e w [1]). Zaobserwowałam, że wydajność dimeryzacji mutantu $\Delta S1a$ nie była znacząco obniżona (Rys. 6b w [1]), co wskazuje, że utworzenie struktury pseudowężła nie jest istotne dla formowania dimeru RNA Ty1. Dodatkowym potwierdzeniem tego wniosku są wyniki otrzymane dla mutantu SL4. Analiza SHAPE struktury 2D monomeru tego mutantu wykazała, że substytucje w pętli apikalnej SL4 zaburzają parowania reszt nukleotydowych w trzonach pseudowężła (Rys.7f w [1]), ale mimo to nie wpływały znacząco na dimeryzację *in vitro*. Mutacje wprowadzone w pętli SL4 były identyczne z tymi zastosowanymi wcześniej w pracy, sugerującej inicjację dimeryzacji RNA Ty1 przez oddziaływania SL1/SL4 [77]. Jednak moje wyniki podważają tę propozycję i wskazują, że obniżony poziom retrotranspozycji Ty1 po zmianie sekwencji w spince SL4 wynika z nieprawidłowej struktury pseudowężła, a nie z zaburzenia procesu dimeryzacji RNA Ty1.

7.1.6. Wpływ mutacji w sekwencjach PAL1, PAL2, PAL3 i IL3 na efektywność retrotranspozycji Ty1 w komórce

W finalnym etapie badań sprawdziliśmy wpływ sekwencji PAL1, PAL2, PAL3 oraz IL3 na aktywność retroelementu Ty1 *in vivo*. W tym celu przeprowadziliśmy funkcjonalne testy retrotranspozycji w komórkach drożdży pozbawionych naturalnie występujących elementów Ty (*S. paradoxus*, szczep DG3852). Badania *in vivo* zostały przeprowadzone we współpracy z zespołem badawczym, którego kierownikiem jest prof. D. J. Garfinkel (University of Georgia, USA). W eksperymentach tych zastosowaliśmy system dwuplazmidowy (helper-Ty1/mini-Ty1) (Rys. 8a w [1]), pozwalający na niezależną ekspresję białek i RNA retrotranspozonu [52,59,77,89]. Plazmid helper-Ty1 koduje mRNA dla białek Gag i Gag-Pol, ale ze względu na mutacje w sekwencji PBS oraz brak 3' UTR, nie może być matrycą w odwrotnej transkrypcji. Plazmid mini-Ty1 zawiera sekwencje elementu Ty1 niezbędne do dimeryzacji, pakowania oraz odwrotnej transkrypcji, ale pozbawiony jest większości rejonu kodującego białka retrotranspozonu. Sekwencje te są wystarczające, aby RNA mini-Ty1 mógł funkcjonować, jako aktywny retrotranspozon, gdy białka Ty1 są dostarczane *in trans*. Ponadto plazmid mini-Ty1 posiada gen *his3AI*, pozwalający na ocenę mobilności Ty1 w komórkach drożdży przy zastosowaniu histy-

dyny, jako czynnika selekcyjnego w pożywce [13]. W przypadku mutantów PAL1 i PAL2 zaobserwowaliśmy dwukrotny spadek mobilności Ty1 w porównaniu do typu dzikiego. Natomiast mutacje w PAL3 lub IL3 nie miały wpływu na efektywność retrotranspozycji (Rys. 8b w [1]). Wynik ten potwierdza ważną rolę oddziaływania PAL1/PAL2 w retrotranspozycji Ty1 i jednocześnie wskazuje, że oddziaływanie IL3/IL3 nie jest istotne *in vivo*, kiedy możliwe jest tworzenie dimeru przez główne miejsca kontaktu.

7.1.7. Podsumowanie

Podsumowując, dzięki analizom strukturalnym i funkcjonalnym uzyskałam dowody na to, że oddziaływania wzajemnie komplementarnych sekwencji palindromowych PAL1 i PAL2 są istotne dla tworzenia dimeru gRNA Ty1. Mutacje zaburzające komplementarność PAL1/PAL2, znacznie ograniczały dimeryzację *in vitro* oraz retrotranspozycję Ty1 w komórce. Wyniki moich badań nie potwierdziły udziału tRNA^{Met} oraz międzycząsteczkowych oddziaływań PAL3/PAL3 i SL1/SL4 w tworzeniu dimeru RNA Ty1. Na tej podstawie opracowałam model dimeru RNA mTy1, który utrzymywany jest poprzez dwa krótkie, międzycząsteczkowe duplekisy PAL1/PAL2 (Rys. 4b w [1]). Pokazałam, że w przypadku zaburzenia oddziaływania PAL1/PAL2 dimeryzacja Ty1 RNA może zachodzić poprzez alternatywne miejsca kontaktu, które nie są naturalnie wykorzystywane. Podobne zjawisko zaobserwowano u retrowirusów. Potwierdziłam też, że białko Gag promuje procesy dimeryzacji RNA Ty1 oraz przyłączania startowego tRNA.

7.2. RNA binding properties of the Ty1 LTR-retrotransposon Gag protein

Julita Gumna, Angelika Andrzejewska-Romanowska, David J. Garfinkel and Katarzyna Pachulska-Wieczorek;

International Journal of Molecular Sciences, 2021; 22(16):9103

W przypadku retrowirusów, jak i retrotranspozonów LTR do efektywnej replikacji i pakowania genomowego RNA do wirionów lub VLPs potrzebne są bezpośrednie interakcje między białkiem Gag i gRNA [90,91]. W przeciwieństwie do szeroko przebadanych retrowirusowych białek Gag, znacznie mniej wiadomo na temat oddziaływań Gag retrotranspozonów LTR z RNA. Celem mojej kolejnej pracy było przeprowadzenie analizy specyficzności wiązania RNA przez białko Gag retrotranspozonu Ty1. W badaniach wykorzystywałam trzy transkrypty: RNA mTy1, Psi⁻ RNA Ty1 (w publikacji nazywane non-Psi Ty1 RNA) i 18S rRNA *S. cerevisiae*. RNA mTy1 zawiera rejon Psi istotny dla pakowania do VLPs (Rys. 7.1) [52]. Psi⁻ RNA Ty1 obejmuje fragment sekwencji kodującej Ty1 (+1000-1616). Ten RNA odpowiada naturalnie występującemu transkryptowi Ty1i, który nie jest pakowany do VLPs i koduje czynnik restrykcyjny p22

[16,47,92]. Fragment drożdżowego 18S rRNA (+1-576) zastosowałam jako dodatkową kontrolę specyficzności wiązania. W eksperymentach porównałam właściwości białek Gag Ty1 w formie prekursorowej (Gag-p49) i dojrzałej (Gag-p45). Ponieważ transport RNA Ty1 do retrosomów i pakowanie do VLPs następuje przed dojrzewaniem Gag [14], szczególnie skupiłam się na właściwościach Gag-p49.

Główną metodą stosowaną przeze mnie w badaniach interakcji Gag/RNA była termoforeza w skali mikro (MST, ang. *microscale thermophoresis*), która zapewnia szybki i czuły sposób analizy oddziaływań międzycząsteczkowych [93]. Metoda ta opiera się na detekcji ruchu cząsteczek w mikroskopijnych gradientach temperatury, który zależy od wielkości, ładunku i zmian powłoki hydratacyjnej cząsteczki. Dodatkowo wyniki analiz MST zweryfikowałam, wykorzystując technikę elektroforetycznej opóźnionej migracji kompleksów w żelu (EMSA, ang. *electrophoretic mobility shift assay*). W celu oszacowania średnich wartości stałej dysocjacji (K_D) dla kompleksów białko/RNA wyniki pomiarów były dopasowywane do modelu równania Hilla.

7.2.1. Wpływ siły jonowej buforu reakcyjnego na oddziaływanie Gag Ty1/RNA

W pierwszym etapie badań analizowałam tworzenie kompleksów białko/RNA w różnych stężeniach jonów jednowartościowych (Na^+). Zmiany K_D wraz ze wzrostem stężenia tych jonów dostarczają informacji na temat charakteru oddziaływań w kompleksie. Wysokie stężenie soli eliminuje oddziaływania niespecyficzne (elektrostatyczne), nie powinno jednak wpływać na kompleksy, w których dominują oddziaływania specyficzne, takie jak wiązania wodorowe [94,95]. Badania właściwości retrowirusowych białek Gag pokazały, że w fizjologicznym stężeniu soli (150 mM) wiążą one Psi^+ RNA i Psi^- RNA z podobnym powinowactwem, ale swoistość dla Psi^+ RNA jest wyraźna, gdy powinowactwo wiązania mierzone jest w buforach o wysokiej sile jonowej [96-98].

Moje badania wykazały, że w fizjologicznym stężeniu NaCl, Gag-p49 wiąże RNA mTy1, Psi^- RNA Ty1 oraz 18S rRNA z wysokim powinowactwem i odnotowałam tylko niewielką różnicę w wartościach K_D dla tych RNA (Rys. 2A, B; S1; Tab. 1 w [2]). Wzrost siły jonowej buforu stosowanego w reakcji (300, 400 i 500 mM NaCl) powodował silniejszą redukcję wiązania białka Gag-p49 do Psi^- RNA Ty1 niż do RNA mTy1. Jednak w przypadku oddziaływania Gag-p49 z 18S rRNA, wyższe stężenie NaCl powodowało wzrost K_D porównywalny do otrzymanego dla RNA mTy1 (Rys. 2C; Tab. 1 w [2]). Podobny efekt obserwowałam dla dojrzałej formy Gag Ty1 (Rys. 2D, E; Tab. 1 w [2]). W fizjologicznym stężeniu soli Gag-p45 wiązał RNA mTy1 i 18S rRNA z bardzo podobnym powinowactwem, a wzrost stężenia NaCl powodował wzrost wartości K_D dla oddziaływań Gag-p45 zarówno z RNA mTy1, jak i z 18S rRNA. Wyniki te wskazują, że wiązanie obu form Gag Ty1 do RNA zawierającego Psi nie jest bardziej odporne na wzrastające stężenie soli w porównaniu do innych RNA. Niemniej jednak, przy każ-

dym testowanym stężeniu soli, stała dysocjacji dla kompleksów Gag Ty1 z RNA mTy1 była niższa niż dla kompleksów z Psi⁻ RNA Ty1 lub 18S rRNA, wskazując na pewną niewielką preferencję wiązania do RNA zawierającego sygnały pakowania. Na podstawie danych otrzymanych w eksperymentach MST przeprowadziłam analizę, pozwalającą oszacować względny udział oddziaływań elektrostatycznych (Z_{eff}) i nieelektrostatycznych ($K_{D(1M)}$) w kompleksach Gag Ty1/RNA (Rys. 2F; Tab. 1 w [2]). Wartości $K_{D(1M)}$ dla wiązania Gag-p49 z RNA mTy1 oraz 18S rRNA były do siebie podobne, co wskazuje na taki sam udział oddziaływań nieelektrostatycznych w tworzeniu tych kompleksów. Większe zróżnicowanie dla tych RNA odnotowałam w interakcjach z Gag-p45. Dla kompleksów Gag-p49/Psi⁻ RNA Ty1 zaobserwowałam najslabsze oddziaływania nieelektrostatyczne. Natomiast, siła oddziaływań elektrostatycznych była podobna dla wszystkich kompleksów białek Gag Ty1 z RNA (Tab. 1 w [2]). Obserwacje te wyraźnie różnią się od tych dla poliprotein Gag retrowirusów, gdzie wiązanie do Psi⁺ RNA charakteryzowało się znacznie silniejszymi (nawet kilkuset krotnie) oddziaływaniami nieelektrostatycznymi w porównaniu do wiązania z Psi⁻ RNA [98,99].

Analizę interakcji białek Gag-p49 i Gag-p45 z RNA mTy1 oraz 18S rRNA w różnych stężeniach soli przeprowadziłam dodatkowo za pomocą metody EMSA (Rys. 4A; Tab. S1 w [2]). Warunki buforowe odpowiadały tym stosowanym w analizach MST, przy czym wykonałam dodatkowe testy w 800 mM NaCl. W eksperymentach EMSA również obserwowałam osłabienie wiązania białek Gag Ty1 zarówno do RNA mTy1, jak i 18S rRNA wraz ze wzrostem siły jonowej buforu reakcyjnego, ale obserwowane zmiany w wartościach K_D były znacznie mniejsze niż te wykryte przy użyciu MST. Kompleksy Gag-p49 z mTy1 RNA i 18S rRNA obserwowałam nawet przy zastosowaniu 800 mM NaCl. Chociaż wartości stałych dysocjacji dla badanych kompleksów były różne od tych otrzymanych za pomocą metody MST, to ogólny trend trochę silniejszego powinowactwa białek Gag Ty1 do RNA mTy1 niż do 18S rRNA był zachowany. Zebrane dane sugerują, że oba białka Gag Ty1 wiążą różne RNA z wysokim powinowactwem, ale niską specyficznością.

7.2.2. Mechanizm wiązania RNA przez Gag Ty1

Współczynnik Hilla (n_H) był większy niż 1 przy każdym stosowanym stężeniu soli dla obu białek Gag Ty1 w interakcjach ze wszystkimi testowanymi RNA (Tab. 1 w [2]), co wskazuje, że wiązanie RNA przez Gag-p49 lub Gag-p45 *in vitro* ma charakter kooperatywny niezależnie od rodzaju cząsteczki RNA. Podobny mechanizm działania zaobserwowano dla białek Gag retrowirusów [96]. Można, zatem wnioskować, że podobnie jak w przypadku składania wirionów [100], każdy RNA może służyć jako rusztowanie dla tworzenia cząstek wirusopodobnych Ty1 i napędzać oddziaływania wielu cząsteczek Gag ze sobą, a także z RNA.

7.2.3. Oddziaływania Gag Ty1/RNA w warunkach kompetycyjnych

W celu dokładniejszego zbadania specyficzności wiązania Gag-p49 do RNA, przeprowadziłam testy MST z różnymi kompetycyjnymi RNA dodanymi w nadmiarze w stosunku do znakowanego RNA. Skuteczność konkurencji między cząsteczkami RNA oceniałam na podstawie zdolności kompetytora do hamowania tworzenia kompleksów znakowanego RNA z białkiem, obserwowanej w zmianie wartości stałej dysocjacji dla tych kompleksów.

W pierwszym typie eksperymentów, stężenia RNA znakowanego i nieznakowanego (kompetytora) były stałe, a wzrastało stężenie białka Gag-p49. Początkowo w charakterze kompetytora zastosowałam tRNA z *E. coli*. W przypadku Gag HIV-1 dodanie nadmiaru tRNA blokuje niespecyficzne oddziaływanie domen MA i NC z RNA przez dodatnio naładowane aminokwasy i wzmacnia specyficzną interakcję z RNA zawierającym Psi przez motywy palca cynkowego [96,101]. Natomiast, dla Gag Ty1 zaobserwowałam, że obecność nadmiaru tRNA w reakcji skutkuje podobnym wzrostem wartości K_D dla kompleksów z RNA zawierającym Psi (RNA mTy1) oraz 18S rRNA, niezależnie od zastosowanego stężenia tRNA (Rys. 3A; S2; Tab. 2 w [2]). Następnie w roli kompetytorów dla znakowanego RNA mTy1 zastosowałam dłuższe transkrypty: Psi⁻ RNA Ty1 i 18S rRNA. Aby ustalić maksymalny poziom kompetycji możliwy do zaobserwowania w stosowanym układzie eksperymentalnym, jako kompetytora użyłam także nieznakowany RNA mTy1. Dodanie w nadmiarze nieznakowanego RNA mTy1 silnie hamowało tworzenie kompleksów Gag-p49 ze znakowanym RNA mTy1 i obserwowałam ośmiokrotny wzrost stałej dysocjacji (Rys. 3B; Tab. 2 w [2]). Natomiast, gdy Psi⁻ RNA Ty1 lub 18S rRNA współzawodniczyły z RNA mTy1 o oddziaływanie z Gag-p49, odnotowałam odpowiednio, czterokrotnie i dwukrotnie mniejszy wzrost wartości K_D . Analogiczne testy kompetycji między RNA mTy1 i 18S rRNA o wiązanie z Gag-p49 przeprowadziłam również za pomocą metody EMSA (Rys. 4B; Tab. S2 w [2]). Podobnie, jak w przypadku MST, hamowanie oddziaływania Gag-p49/RNA mTy1 przez 18S rRNA było o 50% słabsze niż przez nieznakowany RNA mTy1.

W drugim typie eksperymentów MST zastosowałam rosnące stężenie kompetycyjnego RNA przy stałych stężeniach znakowanego RNA mTy1 i Gag-p49. Przy najwyższym stężeniu kompetycyjnego RNA mTy1 (100-krotny nadmiar) obserwowałam całkowite zahamowanie oddziaływania Gag-p49 ze znakowanym RNA mTy1. Wpływ 18S rRNA lub Psi⁻ RNA Ty1 na tworzenie kompleksów Gag-p49/RNA mTy1 był natomiast znacznie słabszy. Co ciekawe, skuteczność Psi⁻ RNA Ty1 była o 50% mniejsza niż ta obserwowana dla 18S rRNA.

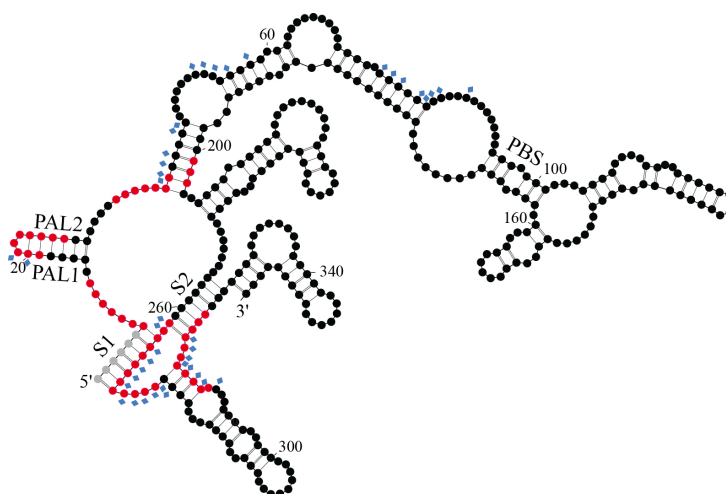
7.2.4. Rola struktury pseudowęzła w wiązaniu RNA przez Gag Ty1

Badania, które prowadziłam w ramach mojej pracy magisterskiej wykazały, że trzony pseudowęzła, znajdującego się na końcu 5' RNA Ty1 oraz rejony im przyległe, stanowią główne miejsca wiązania białek CTR i Gag-p45 [1,47]. Miejsca wiązania prekursorowej formy Gag (p49)

nie były jednak znane. Niejasna pozostała również kwestia, czy pseudowęzeł ułatwia wiązanie Gag, czy też Gag Ty1, jako białko opiekuńcze, wspomaga formowanie tego motywu strukturalnego.

W pierwszej kolejności zbadalam, zatem oddziaływanie Gag-p49 z mutantem RNA mTy1 – $\Delta S1a$ (opis mutantu znajduje się w rozdziale 7.1.5). Powinowactwo wiązania Gag-p49 do mutantu $\Delta S1a$ w 150 mM NaCl było obniżone o około 20% w stosunku do RNA mTy1 typu dzikiego (Rys. 5A; S1; Tab. 1 w [2]). Ponadto w eksperymentach kompetycyjnych zaobserwowałam, że skuteczność konkurencji RNA $\Delta S1a$ o wiązanie z Gag-p49 jest średnio o około 20% niższa niż ta odnotowana dla RNA mTy1 typu dzikiego użytego w charakterze kompetytora (Rys. 5B, C; Tab. 2 w [2]). Dane te wskazują, że destabilizacja pseudowęzła osłabia interakcje między Gag-p49 i RNA Ty1.

W celu dokładniejszego zbadania różnic w interakcjach Gag-p49 z RNA mTy1 i mutantem $\Delta S1a$ postanowiłam porównać miejsca wiązania białka w tych dwóch cząsteczkach RNA i zastosowałam technikę mapowania miejsc wiązania białka za pomocą rodników hydroksylowych. W eksperymencie tym dochodzi do cięcia łańcucha RNA w miejscach wyeksponowanych do roztworu, niezależnie od struktury drugorzędowej RNA [102]. Porównanie profilu cięcia dla RNA w stanie wolnym z profilem cięcia otrzymanym dla RNA w kompleksie z białkiem pozwala wyznaczyć miejsca, w których RNA było chronione przez białko przed działaniem rodników hydroksylowych. W przypadku RNA mTy1 miejsca wiązania Gag-p49 zidentyfikowałam w pobliżu reszt nukleotydowych tworzących trzony pseudowęzła, w sekwencjach palindromowych PAL1 i PAL2 oraz w przyległych im rejonach (Rys. 7.4; Rys. 6A w [2]),



Rysunek 7.4. Model struktury 2D rejonu +1–358 RNA mTy1 z zaznaczeniem pozycji chronionych przed cięciem rodnikami hydroksylowymi w obecności białka Gag-p49 (czerwony). Nieanalizowane nukleotydy są oznaczone kolorem szarym. Pozycje nukleotydów, które odpowiadają miejscom wiązania białka Gag-p45 zaznaczono niebieskimi rombami [1].

co jest zgodne z wynikami, które uzyskałam dla Gag-p45 [1]. Co ciekawe, w przypadku mutantu $\Delta S1a$, profil cięcia rodnikami hydroksylowymi wolnego RNA był bardzo podobny do tego, który uzyskałam dla RNA w kompleksie z Gag-p49 (Rys. 6B w [2]). Zaobserwowałam jedynie niewielki spadek efektywności cięcia RNA przez rodniki hydroksylowe wzdłuż prawie całej długości cząsteczki, a więc, pomimo wciąż wysokiego powinowactwa Gag-p49 do mutantu $\Delta S1a$, nie byłam w stanie wyznaczyć preferencyjnych miejsc wiązania białka. Wyniki te sugerują, że prawidłowo uformowana struktura pseudowęzła może ułatwiać wiązanie Gag Ty1 do genomowego RNA.

7.2.5. Podsumowanie

Moje badania są pierwszą szczegółową analizą oddziaływania białek Gag-p49 i Gag-p45 retrotranspozonu Ty1 z RNA. Wyniki eksperymentów w zmiennych warunkach solnych sugerują, że białka Gag Ty1 wiążą RNA z wysokim powinowactwem, ale mniejszą specyficznością niż białka Gag retrovirusów. Wynika to zapewne z braku w Gag Ty1 typowej domeny nukleokapsydu z motywem palca cynkowego oraz oddziaływania z RNA głównie przez dodatnio naładowane reszty aminokwasowe. Zależne od obecności sygnału pakowania (Psi) zróżnicowanie w wiązaniu Gag-p49 do RNA jest jednak widoczne w eksperymentach kompetycyjnych, sugerując, że niewiele większe powinowactwo Gag do RNA Ty1 może pomóc w odróżnieniu go od innych komórkowych RNA podczas składania VLPs. Oba białka Gag Ty1 wiążą podobne sekwencje w RNA Ty1, a ponadto wydaje się, że Gag Ty1 może rozpoznawać konkretne motywy strukturalne. Prawidłowa struktura gRNA Ty1 może promować optymalną prezentację sekwencji Psi dla interakcji z Gag i ułatwiać wydajne pakowanie gRNA Ty1 do VLPs. Wykryte różnice w wiązaniu RNA przez Gag Ty1 *in vitro* wyjaśniają jednak tylko częściowo proces selektywnego pakowania gRNA Ty1 do VLPs *in vivo*.

7.3. RNAtor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis

Julita Gumna, Tomasz Żok, Kacper Figurski, Katarzyna Pachulska-Wieczorek, Marta Szachniuk;

PloS one; 2020; 15(10):e0239287

W ramach współpracy z Zakładem Bioinformatyki Strukturalnej ICHB PAN opracowaliśmy nowe narzędzie - RNAtor, do w pełni zautomatyzowanej normalizacji, wizualizacji oraz analizy statystycznej danych uzyskanych w eksperymentach próbkowania strukturalnego RNA z wykorzystaniem elektroforezy kapilarnej (CE, ang. *capillary electrophoresis*). Jest ono częścią platformy RNapolis dedykowanej wieloaspektowej analizie struktury RNA [103].

7.3.1. Wprowadzenie

Próbkowanie chemiczne RNA pozwala nam badać podatność nukleotydów na modyfikację chemiczną, a uzyskane informacje można wykorzystać do przewidywania struktury drugorzędowej RNA. Obecnie najpopularniejszymi odczynnikami chemicznymi wykorzystywanymi do próbkowania struktury drugorzędowej RNA *in vitro* i *in vivo* są odczynniki SHAPE (np. NMIA - bezwodnik kwasu 1-metyloizatoinowego) oraz DMS (siarczan dimetylu) [104,105]. Odczynniki SHAPE reagują z grupą 2'-OH rybozy dowolnej reszty nukleotydu, tworząc kowalencyjne addukty, przy czym bardziej elastyczne, niesparowane nukleotydy są modyfikowane częściej niż nukleotydy ograniczone przez parowanie zasad lub trzeciorzędową interakcją [106-108]. DMS natomiast metyluje centra N1 reszt adenozynowych oraz N3 reszt cytydynowych, które nie uczestniczą w tworzeniu wiązań wodorowych [109]. Detekcja miejsc modyfikacji w RNA odbywa się za pomocą reakcji odwrotnej transkrypcji. W zależności od zastosowanych warunków reakcji powstaje pula cDNA, która może zawierać cząsteczki tej samej długości, ale z mutacjami w sekwencji wprowadzonymi w miejscach modyfikacji RNA (metoda MaP, ang. *mutational profiling*; [110,111]) lub cząsteczki o różnej długości [106,109]. W tym drugim przypadku synteza cDNA zostaje zakończona w pierwszej pozycji przyłączenia grupy acylowej (SHAPE) lub metylowej (DMS), gdyż stanowi ona zawadę przestrzenną dla enzymu. Długość syntetyzowanych fragmentów cDNA odpowiada pozycji modyfikacji w RNA, a ilość danego fragmentu cDNA odzwierciedla częstość modyfikacji (reaktywność SHAPE/DMS), a więc lokalną dostępność strukturalną reszt nukleotydowych w RNA. Miejsca zatrzymań odwrotnej transkryptazy można odczytać za pomocą elektroforezy płytowej, elektroforezy kapilarnej lub sekwencjonowania nowej generacji.

Przy zastosowaniu elektroforezy kapilarnej eksperymentator otrzymuje surowe dane w postaci elektroferogramów, prezentujących intensywność sygnału fluorescencyjnego dla każdej reszty nukleotydu w badanym fragmencie RNA. Intensywność sygnału przekłada się na częstość modyfikacji w danej pozycji w RNA. Pozyskanie danych ilościowych z elektroferogramów CE jest trudne i wymaga skomplikowanej, wieloetapowej analizy sygnałów fluorescencyjnych. Istnieje kilka narzędzi obliczeniowych do przetwarzania elektroferogramów, z których najpopularniejsze to ShapeFinder [112] i QuShape [113]. Po otrzymaniu danych ilościowych z próbkowania strukturalnego i przed ich wprowadzeniem do programów przewidujących strukturę drugorzędową RNA należy jednak wykonać dodatkową normalizację wartości reaktywności oraz kontrolę jakości danych. Etap ten zwykle wykonywany jest ręcznie, przez co uzależniony jest od wiedzy, umiejętności i doświadczenia badacza.

7.3.2. RNATHor

W celu ułatwienia procedury przetwarzania danych z mapowania strukturalnego RNA z eksperymentów skoncentrowanych na badaniu pojedynczych cząsteczek RNA zdecydowaliśmy się

opracować nowe narzędzie - RNAtbor, do w pełni zautomatyzowanej normalizacji danych uzyskanych w eksperymentach próbkowania strukturalnego RNA metodą SHAPE-CE lub DMS-CE i wstępnie przeanalizowanych w programie ShapeFinder lub QuShape. RNAtbor to serwer sieci WWW (<https://rnatbor.cs.put.poznan.pl>) z prostym i intuicyjnym interfejsem oraz przejrzystą wizualizacją wyników analizy. RNAtbor przeprowadza normalizację w oparciu o ugruntowany protokół zaproponowany przez twórców metody SHAPE [114]. Użytkownik otrzymuje wyniki normalizacji w postaci arkusza kalkulacyjnego (.xlsx) oraz wykresów kolumnowego i schodkowego, przedstawiających profile reaktywności. Ponadto RNAtbor wykorzystuje algorytm programu RNAstructure [115] do przewidywania struktury drugorzędowej RNA z uwzględnieniem wprowadzonych danych eksperymentalnych. Uzyskana struktura 2D RNA jest przedstawiona za pomocą notacji kropkowo-nawiasowej oraz diagramu graficznego. Na diagramie narysowanym przy użyciu dostosowanego algorytmu RNApuzler [116] poszczególne nukleotydy są pokolorowane zgodnie ze schematem oznaczania reaktywności SHAPE/DMS zdefiniowanym przez użytkownika w ustawieniach wejściowych. Wyniki analizy ze znormalizowanymi wartościami reaktywności są również gromadzone w plikach tekstowych w formacie .shape, zapewniając interoperacyjność z programem RNAstructure.

Całkowitą nowością jest zaproponowany przeze mnie algorytm identyfikowania niewiarogodnych danych, które wynikają z przedwczesnego zakończenia reakcji wydłużania startera z powodów innych niż modyfikacja odczynnikami SHAPE lub DMS (np. istniejące wcześniej cięcia w sekwencji RNA). Zautomatyzowaną procedurę detekcji niespecyficznych, silnych zatrzymań odwrotnej transkryptazy (Rys. 2 w [3]) opracowałam w oparciu o doświadczenie zdobyte podczas licznych analiz danych z eksperymentów próbkowania chemicznego różnych cząsteczek RNA (w tym RNA retrotranspozonu Ty1) *in vitro* i *in vivo*. Algorytm eliminuje dane, które spełniają jedno z następujących kryteriów:

- wartość reaktywności przed normalizacją jest ujemna;
- powierzchnia pików tła jest, co najmniej pięć razy większa niż średnia powierzchnia pików tła;
- różnica w powierzchni pików między tłem a reakcją jest mniejsza niż 35% średniej powierzchni pików tła i jednocześnie powierzchnia pików tła w tej pozycji jest równa lub większa niż ta średnia.

Przedstawiona w publikacji walidacja proponowanej metody normalizacji danych była przeprowadzona z wykorzystaniem danych z próbkowania struktury RNA mTy1. Opracowany przeze mnie algorytm został zoptymalizowany pod kątem analizy danych z eksperymentów SHAPE i DMS. Istnieje jednak możliwość jego wyłączenia, dzięki czemu RNAtbor może być również z powodzeniem stosowany do analizy danych otrzymanych za pomocą innych metod próbkowania strukturalnego RNA. RNAtbor wykorzystywałam między innymi do analizy danych z eksperymentów próbkowania kompleksów RNA/białko rodnikami hydroksylowymi (Rys. 6 w [2]).

Korzystanie z naszego narzędzia jest całkowicie bezpłatne oraz nie wymaga logowania, jednakże utworzenie konta daje dostęp do dodatkowych funkcji, jakimi są analiza porównawcza i statystyczna. Opcja „Generate chart” prowadzi do wygenerowania profilów reaktywności dla wybranych eksperymentów porównanych na jednym wykresie (Rys. 4A, B w [3]). Podczas analizy statystycznej wykonywany jest nieparametryczny test Manna-Whitneya (jeśli użytkownik wybrał 2 eksperymenty) lub test Kruskala-Wallisa (jeśli użytkownik wybrał od 3 do 5 eksperymentów). W rezultacie użytkownik otrzymuje dane liczbowe oraz graficzne – wykres pudełkowy i skrzypcowy (Rys. 4C w [3]). Oferowane przez RNAuthor funkcje ułatwiają analizę wielu eksperymentów próbkowania strukturalnego, pozwalają ocenić zgodność między eksperymentami i porównać całe zestawy danych dla RNA próbkowanych w różnych warunkach eksperymentalnych.

7.3.3. Podsumowanie

W porównaniu z ręcznym lub półautomatycznym przetwarzaniem danych, RNAuthor znacznie skraca czas potrzebny na analizę i zamienia żmudną pracę w kilkuminutowe zadanie. RNAuthor łączy w sobie następujące funkcje: (1) normalizacja danych z mapowania struktury 2D RNA z kilku eksperymentów jednocześnie, (2) automatyczne wykrywanie niewiarygodnych danych, (3) wizualizacja znormalizowanych wartości reaktywności, (4) analiza statystyczna i porównawcza kilku zestawów danych jednocześnie, (5) przewidywanie struktury drugorzędowej RNA z uwzględnieniem dostarczonych danych eksperymentalnych. Tym samym RNAuthor może zastąpić nawet kilka programów wykorzystywanych przez badaczy. Chociaż istnieją programy, które mogą przetwarzać wyniki z próbkowania strukturalnego RNA otrzymane z wykorzystaniem elektroforezy kapilarnej, to żaden z nich, w przeciwieństwie do naszego narzędzia, nie oferuje identyfikacji niewiarygodnych danych, a ten etap analizy był zwykle wykonywany ręcznie.

8. PODSUMOWANIE

Przeprowadzone przeze mnie analizy strukturalne oraz funkcjonalne *in vitro* i *in vivo* poszerzają wiedzę na temat dimeryzacji i pakowania RNA retrotranspozonu Ty1. Dzięki temu moje badania w znacznym stopniu przyczyniają się do lepszego poznania molekularnych mechanizmów replikacji endogennych retroelementów z rodziny *Pseudoviridae*. Opisywane wcześniej w literaturze mechanizmy dimeryzacji i pakowania RNA prezentowane są, bowiem głównie z perspektywy infekcyjnych retrowirusów (*Retroviridae*). Najważniejsze wyniki moich badań to:

- ustalenie struktury końca 5' RNA Ty1 w formie dimeru stabilizowanego przez krótkie międzycząsteczkowe dupleksy [1];
- wykazanie kluczowej roli sekwencji palindromowych PAL1 i PAL2 w procesie dimeryzacji RNA Ty1 [1];
- wykazanie, że RNA Ty1 może formować dimer również przez alternatywne miejsca kontaktu [1];
- scharakteryzowanie po raz pierwszy oddziaływań białka Gag retrotranspozonu LTR z RNA [2];
- wykazanie zależności między strukturą RNA a specyficznym wiązaniem przez białko Gag Ty1 [2].

Warto podkreślić, że badania dotyczące retrotranspozonów LTR są ważne nie tylko ze względu na ich ewolucyjne pokrewieństwo z retrowirusami. Obecnie wiadomo, że endogenne retroelementy, których znaczną część stanowią retrotranspozony LTR to nie „śmieciowe” DNA, lecz elementy, odgrywające ważną rolę w utrzymaniu architektury genomów, ich funkcjonowaniu oraz ewolucji. Transkrypcja RNA tych retroelementów, translacja ich białek, a nawet czasem też produkcja VLPs towarzyszy różnym chorobom nowotworowym i autoimmunologicznym człowieka. Ponadto, w niektórych przypadkach sekwencje kodujące białka retrotranspozonów zostały zaadoptowane i aktualnie kodują funkcjonalne białka istotne dla prawidłowego działania wielu procesów u organizmów eukariotycznych. W ludzkim genomie zidentyfikowano między innymi grupę ponad 80 genów, pochodzących od genu *gag* retrotranspozonów z rodziny Ty3/Gypsy. Dodatkowo retrotranspozony są ważnymi narzędziami w biologii syntetycznej. Przykładem może być tutaj białko Gag Ty1 wykorzystywane w produkcji bifunkcyjnych szczepionek typu VLP.

Metodyka moich badań, dotyczących retrotranspozonu Ty1, w znacznym stopniu opiera się na eksperymentach próbkowania chemicznego struktury RNA. Przeprowadzenie wielu analiz danych z tych eksperymentów wskazało mi słabe strony istniejących metod ich przetwarzania i zainspirowało do stworzenia nowego narzędzia bioinformatycznego, ułatwiającego ten etap badań. Efektem mojej pracy jest:

- opracowanie wraz z Zakładem Bioinformatyki Strukturalnej ICHB PAN wielofunkcyjnego programu RNATHOR do przetwarzania danych z eksperymentów mapowania struktury drugorzędowej RNA [3];
- zaproponowanie pierwszej zautomatyzowanej metody identyfikacji niewiarygodnych danych w eksperymentach próbkowania struktury drugorzędowej RNA metodą SHAPE-CE lub DMS-CE [3].

9. WYKAZ STOSOWANYCH SKRÓTÓW I NOMENKLATURA

aa	aminokwas
CA	domena kapsydu, występująca w poliproteinie Gag retrowirusów (ang. <i>capsid</i>)
cDNA	komplementarny DNA zsyntetyzowany na matrycy RNA
CE	elektroforeza kapilarna (ang. <i>capillary electrophoresis</i>)
CNC	mechanizm kontroli liczby kopii retrotranspozonu w komórce gospodarza (ang. <i>copy number control</i>)
DNA	kwas deoksyrybonukleinowy
DMS	siarczan dimetylu
EMSA	technika elektroforetyczna opóźnionej migracji kompleksów w żelu (ang. <i>electrophoretic mobility shift assay</i>)
ER	retikulum endoplazmatyczne
gRNA	genomowy RNA
IL	pętla wewnętrzna w strukturze drugorzędowej RNA (ang. <i>internal loop</i>)
IN	integraza
K_D	stała dysocjacji
LTR	dłgie terminalne powtórzenia na obu końcach DNA retrowirusów i retrotranspozonów LTR (ang. <i>long terminal repeats</i>)
MA	domena macierzy, występująca w poliproteinie Gag retrowirusów (ang. <i>matrix</i>)
MST	termoforeza w skali mikro (ang. <i>microscale thermophoresis</i>)
mRNA	informacyjny RNA (ang. <i>messenger RNA</i>)
NAC	aktywność opiekuńcza względem kwasów nukleinowych (ang. <i>nucleic acid chaperone activity</i>)
NC	domena nukleokapsydu, występująca w poliproteinie Gag retrowirusów (ang. <i>nucleocapsid</i>)
NMIA	bezwodnik kwasu 1-metyloizatoinowego
nt	nukleotyd
PBS	sekwencja RNA, wiążąca tRNA wykorzystywanego jako starter podczas odwrotnej transkrypcji retroelementów (ang. <i>primer binding site</i>)
PR	proteaza
Psi	rejon RNA retroelementów istotny dla pakowania RNA do wirionów/VLPs zwany sygnałem pakowania (ang. <i>packaging signal</i>)
RH	RNaza H
RNA	kwas rybonukleinowy
rRNA	rybosomalny RNA (ang. <i>ribosomal RNA</i>)
RT	odwrotna transkryptaza (ang. <i>reverse transcriptase</i>)

SHAPE	selektywna acylacja grup 2'-OH rybozy analizowana metodą odwrotnej transkrypcji (ang. <i>selective 2'-hydroxyl acylation analyzed by primer extention</i>)
SL	struktura spinki do włosów w strukturze drugorzędowej RNA, składająca się z trzonu, w którym reszty nukleotydydowe tworzą pary oraz z pętli apikalnej z nieparowanymi nukleotydami (ang. <i>stem loop</i>)
SRP	cząstka rozpoznająca sekwencję sygnałową (ang. <i>signal recognition particle</i>)
TB	bufor wykorzystywany m.in. do elektroforezy na żelu agarozowym i poliakrylamidowym kwasów nukleinowych, zawierający Tris-HCl oraz H ₃ BO ₃
TBM	bufor wykorzystywany m.in. do elektroforezy na żelu agarozowym i poliakrylamidowym kwasów nukleinowych, zawierający Tris-HCl, H ₃ BO ₃ oraz MgCl ₂
tRNA	transportujący RNA (ang. <i>transfer RNA</i>)
UTR	niepodlegające translacji końcowe rejony RNA (ang. <i>untranslated region</i>)
VLPs	cząstki wirusopodobne (ang. <i>virus-like particles</i>)

10. BIBLIOGRAFIA

1. Gumna, J.; Purzycka, K.J.; Ahn, H.W.; Garfinkel, D.J. and Pachulska-Wieczorek, K. Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA biology* **2019**, *16*, 1749-1763, doi: 10.1080/15476286.2019.1657370.
2. Gumna, J.; Andrzejewska-Romanowska, A.; Garfinkel, D.J. and Pachulska-Wieczorek, K. RNA Binding Properties of the Ty1 LTR-Retrotransposon Gag Protein. *International journal of molecular sciences* **2021**, *22*, doi: 10.3390/ijms22169103.
3. Gumna, J.; Zok, T.; Figurski, K.; Pachulska-Wieczorek, K. and Szachniuk, M. RNAtor - fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis. *PLoS one* **2020**, *15*, e0239287, doi: 10.1371/journal.pone.0239287.
4. Eickbush, T.H. and Jamburuthugoda, V.K. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus research* **2008**, *134*, 221-234, doi: 10.1016/j.virusres.2007.12.010.
5. Koonin, E.V.; Dolja, V.V. and Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **2015**, *479-480*, 2-25, doi: 10.1016/j.virol.2015.02.039.
6. Koonin, E.V. and Dolja, V.V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiology and molecular biology reviews : MMBR* **2014**, *78*, 278-303, doi: 10.1128/MMBR.00049-13.
7. Krupovic, M.; Blomberg, J.; Coffin, J.M.; Dasgupta, I.; Fan, H.; Geering, A.D.; Gifford, R.; Harach, B.; Hull, R.; Johnson, W. *et al.* *Ortervirales*: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. *Journal of virology* **2018**, *92*, doi: 10.1128/JVI.00515-18.
8. Menendez-Arias, L.; Sebastian-Martin, A. and Alvarez, M. Viral reverse transcriptases. *Virus research* **2017**, *234*, 153-176, doi: 10.1016/j.virusres.2016.12.019.
9. Benachenhou, F.; Sperber, G.O.; Bongcam-Rudloff, E.; Andersson, G.; Boeke, J.D. and Blomberg, J. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA* **2013**, *4*, 5, doi: 10.1186/1759-8753-4-5.
10. Hayward, A. and Katzourakis, A. Endogenous retroviruses. *Current biology : CB* **2015**, *25*, R644-646, doi: 10.1016/j.cub.2015.05.041.
11. Carr, M.; Bensasson, D. and Bergman, C.M. Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. *PLoS one* **2012**, *7*, e50978, doi: 10.1371/journal.pone.0050978.
12. Kim, J.M.; Vanguri, S.; Boeke, J.D.; Gabriel, A. and Voytas, D.F. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome research* **1998**, *8*, 464-478, doi: 10.1101/gr.8.5.464.
13. Curcio, M.J. and Garfinkel, D.J. Single-step selection for Ty1 element retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 936-940, doi: 10.1073/pnas.88.3.936.

14. Curcio, M.J.; Lutz, S. and Lesage, P. The Ty1 LTR-retrotransposon of budding yeast, *Saccharomyces cerevisiae*. *Microbiology spectrum* **2015**, *3*, 1-35, doi: 10.1128/microbiolspec.MDNA3-0053-2014.
15. Boeke, J.D.; Eichinger, D.; Castrillon, D. and Fink, G.R. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. *Molecular and cellular biology* **1988**, *8*, 1432-1442, doi: 10.1128/mcb.8.4.1432-1442.1988.
16. Saha, A.; Mitchell, J.A.; Nishida, Y.; Hildreth, J.E.; Ariberre, J.A.; Gilbert, W.V. and Garfinkel, D.J. A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. *Journal of virology* **2015**, *89*, 3922-3938, doi: 10.1128/JVI.03060-14.
17. Curcio, M.J.; Hedge, A.M.; Boeke, J.D. and Garfinkel, D.J. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in *Saccharomyces cerevisiae*. *Molecular & general genetics : MGG* **1990**, *220*, 213-221, doi: 10.1007/BF00260484.
18. Munchel, S.E.; Shultzaberger, R.K.; Takizawa, N. and Weis, K. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Molecular biology of the cell* **2011**, *22*, 2787-2795, doi: 10.1091/mbc.E11-01-0028.
19. Checkley, M.A.; Mitchell, J.A.; Eizenstat, L.D.; Lockett, S.J. and Garfinkel, D.J. Ty1 gag enhances the stability and nuclear export of Ty1 mRNA. *Traffic* **2013**, *14*, 57-69, doi: 10.1111/tra.12013.
20. Kawakami, K.; Pande, S.; Faiola, B.; Moore, D.P.; Boeke, J.D.; Farabaugh, P.J.; Strathern, J.N.; Nakamura, Y. and Garfinkel, D.J. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in *Saccharomyces cerevisiae*. *Genetics* **1993**, *135*, 309-320.
21. Xu, H. and Boeke, J.D. Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 8360-8364, doi: 10.1073/pnas.87.21.8360.
22. Clare, J.J.; Belcourt, M. and Farabaugh, P.J. Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 6816-6820, doi: 10.1073/pnas.85.18.6816.
23. Harger, J.W. and Dinman, J.D. An in vivo dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *Rna* **2003**, *9*, 1019-1024, doi: 10.1261/rna.5930803.
24. Malagon, F. and Jensen, T.H. The T body, a new cytoplasmic RNA granule in *Saccharomyces cerevisiae*. *Molecular and cellular biology* **2008**, *28*, 6022-6032, doi: 10.1128/MCB.00684-08.
25. Malagon, F. and Jensen, T.H. T-body formation precedes virus-like particle maturation in *S. cerevisiae*. *RNA biology* **2011**, *8*, 184-189, doi: 10.4161/rna.8.2.14822.
26. Merkulov, G.V.; Lawler, J.F., Jr.; Eby, Y. and Boeke, J.D. Ty1 proteolytic cleavage sites are required for transposition: all sites are not created equal. *Journal of virology* **2001**, *75*, 638-644, doi: 10.1128/JVI.75.2.638-644.2001.

27. Garfinkel, D.J.; Hedge, A.M.; Youngren, S.D. and Copeland, T.D. Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1. *Journal of virology* **1991**, *65*, 4573-4581, doi: 10.1128/JVI.65.9.4573-4581.1991.
28. Garfinkel, D.J.; Boeke, J.D. and Fink, G.R. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **1985**, *42*, 507-517, doi: 10.1016/0092-8674(85)90108-4.
29. Hughes, S.H. Reverse Transcription of Retroviruses and LTR Retrotransposons. *Microbiology spectrum* **2015**, *3*, MDNA3-0027-2014, doi: 10.1128/microbiolspec.MDNA3-0027-2014.
30. Perlman, P.S. and Boeke, J.D. Molecular biology. Ring around the retroelement. *Science* **2004**, *303*, 182-184, doi: 10.1126/science.1093514.
31. Devine, S.E. and Boeke, J.D. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes & development* **1996**, *10*, 620-633, doi: 10.1101/gad.10.5.620.
32. Qi, X.; Daily, K.; Nguyen, K.; Wang, H.; Mayhew, D.; Rigor, P.; Forouzan, S.; Johnston, M.; Mitra, R.D.; Baldi, P. *et al.* Retrotransposon profiling of RNA polymerase III initiation sites. *Genome research* **2012**, *22*, 681-692, doi: 10.1101/gr.131219.111.
33. Bell, N.M. and Lever, A.M. HIV Gag polyprotein: processing and early viral particle assembly. *Trends in microbiology* **2013**, *21*, 136-144, doi: 10.1016/j.tim.2012.11.006.
34. Coffin, J.M.; Hughes, S.H.; Varmus, H.E. and editors. *Retroviruses*; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1997.
35. Roth, J.F. The yeast Ty virus-like particles. *Yeast* **2000**, *16*, 785-795, doi: 10.1002/1097-0061(20000630)16:9<785::AID-YEA550>3.0.CO;2-L.
36. Merkulov, G.V.; Swiderek, K.M.; Brachmann, C.B. and Boeke, J.D. A critical proteolytic cleavage site near the C terminus of the yeast retrotransposon Ty1 Gag protein. *Journal of virology* **1996**, *70*, 5548-5556, doi: 10.1128/JVI.70.8.5548-5556.1996.
37. Brachmann, C.B. and Boeke, J.D. Mapping the multimerization domains of the Gag protein of yeast retrotransposon Ty1. *Journal of virology* **1997**, *71*, 812-817, doi: 10.1128/JVI.71.1.812-817.1997.
38. Martin-Rendon, E.; Marfany, G.; Wilson, S.; Ferguson, D.J.; Kingsman, S.M. and Kingsman, A.J. Structural determinants within the subunit protein of Ty1 virus-like particles. *Molecular microbiology* **1996**, *22*, 667-679, doi: 10.1046/j.1365-2958.1996.d01-1716.x.
39. Brookman, J.L.; Stott, A.J.; Cheeseman, P.J.; Burns, N.R.; Adams, S.E.; Kingsman, A.J. and Gull, K. An immunological analysis of Ty1 virus-like particle structure. *Virology* **1995**, *207*, 59-67, doi: 10.1006/viro.1995.1051.
40. Checkley, M.A.; Nagashima, K.; Lockett, S.J.; Nyswaner, K.M. and Garfinkel, D.J. P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. *Molecular and cellular biology* **2010**, *30*, 382-398, doi: 10.1128/MCB.00251-09.
41. Doh, J.H.; Lutz, S. and Curcio, M.J. Co-translational localization of an LTR-retrotransposon RNA to the endoplasmic reticulum nucleates virus-like particle assembly sites. *PLoS genetics* **2014**, *10*, e1004219, doi: 10.1371/journal.pgen.1004219.

42. Saraogi, I. and Shan, S.O. Molecular mechanism of co-translational protein targeting by the signal recognition particle. *Traffic* **2011**, *12*, 535-542, doi: 10.1111/j.1600-0854.2011.01171.x.
43. Olson, E.D. and Musier-Forsyth, K. Retroviral Gag protein-RNA interactions: Implications for specific genomic RNA packaging and virion assembly. *Seminars in cell & developmental biology* **2019**, *86*, 129-139, doi: 10.1016/j.semcdb.2018.03.015.
44. Rein, A. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA biology* **2010**, *7*, 700-705, doi: 10.4161/rna.7.6.13685.
45. Zuniga, S.; Sola, I.; Cruz, J.L. and Enjuanes, L. Role of RNA chaperones in virus replication. *Virus research* **2009**, *139*, 253-266, doi: 10.1016/j.virusres.2008.06.015.
46. Cristofari, G.; Ficheux, D. and Darlix, J.L. The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. *The Journal of biological chemistry* **2000**, *275*, 19210-19217, doi: 10.1074/jbc.M001371200.
47. Nishida, Y.; Pachulska-Wieczorek, K.; Blaszczyk, L.; Saha, A.; Gumna, J.; Garfinkel, D.J. and Purzycka, K.J. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic acids research* **2015**, *43*, 7414-7431, doi: 10.1093/nar/gkv695.
48. Cottee, M.A.; Beckwith, S.L.; Letham, S.C.; Kim, S.J.; Young, G.R.; Stoye, J.P.; Garfinkel, D.J. and Taylor, I.A. Structure of a Ty1 restriction factor reveals the molecular basis of transposition copy number control. *Nature communications* **2021**, *12*, 5590, doi: 10.1038/s41467-021-25849-0.
49. Rajkowitsch, L.; Chen, D.; Stampfl, S.; Semrad, K.; Waldsich, C.; Mayer, O.; Jantsch, M.F.; Konrat, R.; Blasi, U. and Schroeder, R. RNA chaperones, RNA annealers and RNA helicases. *RNA biology* **2007**, *4*, 118-130, doi: 10.4161/rna.4.3.5445.
50. Woodson, S.A.; Panja, S. and Santiago-Frangos, A. Proteins That Chaperone RNA Regulation. *Microbiology spectrum* **2018**, *6*, doi: 10.1128/microbiolspec.RWR-0026-2018.
51. Tucker, J.M.; Larango, M.E.; Wachsmuth, L.P.; Kannan, N. and Garfinkel, D.J. The Ty1 Retrotransposon Restriction Factor p22 Targets Gag. *PLoS genetics* **2015**, *11*, doi: 10.1371/journal.pgen.1005571.
52. Xu, H. and Boeke, J.D. Localization of sequences required *in cis* for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. *Molecular and cellular biology* **1990**, *10*, 2695-2702, doi: 10.1128/mcb.10.6.2695-2702.1990.
53. Chapman, K.B.; Bystrom, A.S. and Boeke, J.D. Initiator methionine tRNA is essential for Ty1 transposition in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 3236-3240, doi: 10.1073/pnas.89.8.3236.
54. Friant, S.; Heyman, T.; Bystrom, A.S.; Wilhelm, M. and Wilhelm, F.X. Interactions between Ty1 retrotransposon RNA and the T and D regions of the tRNA(iMet) primer are required for initiation of reverse transcription *in vivo*. *Molecular and cellular biology* **1998**, *18*, 799-806, doi: 10.1128/mcb.18.2.799.

55. Friant, S.; Heyman, T.; Wilhelm, M.L. and Wilhelm, F.X. Extended interactions between the primer tRNAⁱ(Met) and genomic RNA of the yeast Ty1 retrotransposon. *Nucleic acids research* **1996**, *24*, 441-449, doi: 10.1093/nar/24.3.441.
56. Friant, S.; Heyman, T.; Wilhelm, F.X. and Wilhelm, M. Role of RNA primers in initiation of minus-strand and plus-strand DNA synthesis of the yeast retrotransposon Ty1. *Biochimie* **1996**, *78*, 674-680, doi: 10.1016/s0300-9084(96)80013-7.
57. Laueremann, V. and Boeke, J.D. Plus-strand strong-stop DNA transfer in yeast Ty retrotransposons. *The EMBO journal* **1997**, *16*, 6603-6612, doi: 10.1093/emboj/16.21.6603.
58. Cristofari, G.; Bampi, C.; Wilhelm, M.; Wilhelm, F.X. and Darlix, J.L. A 5'-3' long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *The EMBO journal* **2002**, *21*, 4368-4379, doi: 10.1093/emboj/cdf436.
59. Bolton, E.C.; Coombes, C.; Eby, Y.; Cardell, M. and Boeke, J.D. Identification and characterization of critical *cis*-acting sequences within the yeast Ty1 retrotransposon. *Rna* **2005**, *11*, 308-322, doi: 10.1261/rna.7860605.
60. Andrzejewska, A.; Zawadzka, M.; Gumna, J.; Garfinkel, D.J. and Pachulska-Wieczorek, K. *In vivo* structure of the Ty1 retrotransposon RNA genome. *Nucleic acids research* **2021**, *49*, 2878-2893, doi: 10.1093/nar/gkab090.
61. Purzycka, K.J.; Legiewicz, M.; Matsuda, E.; Eizentstat, L.D.; Lusvardi, S.; Saha, A.; Le Grice, S.F. and Garfinkel, D.J. Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic acids research* **2013**, *41*, 463-473, doi: 10.1093/nar/gks983.
62. Huang, Q.; Purzycka, K.J.; Lusvardi, S.; Li, D.; Legrice, S.F. and Boeke, J.D. Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *Rna* **2013**, *19*, 320-332, doi: 10.1261/rna.035535.112.
63. Fu, W.; Gorelick, R.J. and Rein, A. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *Journal of virology* **1994**, *68*, 5013-5018, doi: 10.1128/JVI.68.8.5013-5018.1994.
64. Johnson, S.F. and Telesnitsky, A. Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS pathogens* **2010**, *6*, e1001007, doi: 10.1371/journal.ppat.1001007.
65. Comas-Garcia, M.; Davis, S.R. and Rein, A. On the Selective Packaging of Genomic RNA by HIV-1. *Viruses* **2016**, *8*, doi: 10.3390/v8090246.
66. D'Souza, V. and Summers, M.F. How retroviruses select their genomes. *Nature reviews. Microbiology* **2005**, *3*, 643-655, doi: 10.1038/nrmicro1210.
67. Dubois, N.; Marquet, R.; Paillart, J.C. and Bernacchi, S. Retroviral RNA Dimerization: From Structure to Functions. *Frontiers in microbiology* **2018**, *9*, 527, doi: 10.3389/fmicb.2018.00527.
68. Kharytonchyk, S.; Monti, S.; Smaldino, P.J.; Van, V.; Bolden, N.C.; Brown, J.D.; Russo, E.; Swanson, C.; Shuey, A.; Telesnitsky, A. *et al.* Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 13378-13383, doi: 10.1073/pnas.1616627113.

69. Boeras, I.; Seufzer, B.; Brady, S.; Rendahl, A.; Heng, X. and Boris-Lawrie, K. The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. *Scientific reports* **2017**, *7*, 6902, doi: 10.1038/s41598-017-06883-9.
70. Lu, K.; Heng, X. and Summers, M.F. Structural determinants and mechanism of HIV-1 genome packaging. *Journal of molecular biology* **2011**, *410*, 609-633, doi: 10.1016/j.jmb.2011.04.029.
71. Miyazaki, Y.; Miyake, A.; Nomaguchi, M. and Adachi, A. Structural dynamics of retroviral genome and the packaging. *Frontiers in microbiology* **2011**, *2*, 264, doi: 10.3389/fmicb.2011.00264.
72. Russell, R.S.; Liang, C. and Wainberg, M.A. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **2004**, *1*, 23, doi: 10.1186/1742-4690-1-23.
73. Mikkelsen, J.G. and Pedersen, F.S. Genetic reassortment and patch repair by recombination in retroviruses. *Journal of biomedical science* **2000**, *7*, 77-99, doi: 10.1007/BF02256615.
74. Mikkelsen, J.G.; Lund, A.H.; Duch, M. and Pedersen, F.S. Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination *in vivo*. *Journal of virology* **2000**, *74*, 600-610, doi: 10.1128/jvi.74.2.600-610.2000.
75. Roth, J.F.; Kingsman, S.M.; Kingsman, A.J. and Martin-Rendon, E. Possible regulatory function of the *Saccharomyces cerevisiae* Ty1 retrotransposon core protein. *Yeast* **2000**, *16*, 921-932, doi: 10.1002/1097-0061(200007)16:10<921::AID-YEA588>3.0.CO;2-#.
76. Feng, Y.X.; Moore, S.P.; Garfinkel, D.J. and Rein, A. The genomic RNA in Ty1 virus-like particles is dimeric. *Journal of virology* **2000**, *74*, 10819-10821, doi: 10.1128/jvi.74.22.10819-10821.2000.
77. Gamache, E.R.; Doh, J.H.; Ritz, J.; Laederach, A.; Bellaousov, S.; Mathews, D.H. and Curcio, M.J. Structure-Function Model for Kissing Loop Interactions That Initiate Dimerization of Ty1 RNA. *Viruses* **2017**, *9*, doi: 10.3390/v9050093.
78. Polge, E.; Darlix, J.L.; Paoletti, J. and Fosse, P. Characterization of loose and tight dimer forms of avian leukosis virus RNA. *Journal of molecular biology* **2000**, *300*, 41-56, doi: 10.1006/jmbi.2000.3832.
79. Purzycka, K.J.; Pachulska-Wieczorek, K. and Adamiak, R.W. The *in vitro* loose dimer structure and rearrangements of the HIV-2 leader RNA. *Nucleic acids research* **2011**, *39*, 7234-7248, doi: 10.1093/nar/gkr385.
80. Aktar, S.J.; Jabeen, A.; Ali, L.M.; Vivet-Boudou, V.; Marquet, R. and Rizvi, T.A. SHAPE analysis of the 5' end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. *Rna* **2013**, *19*, 1648-1658, doi: 10.1261/rna.040931.113.
81. Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E.L.; Kharytonchyk, S.; Dorjsuren, B.; Kurlandaivel, G.; Jones, S.; Hiremath, A. *et al.* NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **2011**, *334*, 242-245, doi: 10.1126/science.1210460.
82. Berkhout, B.; Ooms, M.; Beerens, N.; Huthoff, H.; Southern, E. and Verhoef, K. *In vitro* evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates

- multiple functions through conformational changes. *The Journal of biological chemistry* **2002**, 277, 19967-19975, doi: 10.1074/jbc.M200950200.
83. Tran, T.; Liu, Y.; Marchant, J.; Monti, S.; Seu, M.; Zaki, J.; Yang, A.L.; Bohn, J.; Ramakrishnan, V.; Singh, R. *et al.* Conserved determinants of lentiviral genome dimerization. *Retrovirology* **2015**, 12, 83, doi: 10.1186/s12977-015-0209-x.
84. Lusvarghi, S.; Sztuba-Solinska, J.; Purzycka, K.J.; Pauly, G.T.; Rausch, J.W. and Grice, S.F. The HIV-2 Rev-response element: determining secondary structure and defining folding intermediates. *Nucleic acids research* **2013**, 41, 6637-6649, doi: 10.1093/nar/gkt353.
85. Popena, M.; Szachniuk, M.; Antczak, M.; Purzycka, K.J.; Lukasiak, P.; Bartol, N.; Blazewicz, J. and Adamiak, R.W. Automated 3D structure composition for large RNAs. *Nucleic acids research* **2012**, 40, e112, doi: 10.1093/nar/gks339.
86. Antczak, M.; Popena, M.; Zok, T.; Sarzynska, J.; Ratajczak, T.; Tomczyk, K.; Adamiak, R.W. and Szachniuk, M. New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta biochimica Polonica* **2016**, 63, 737-744, doi: 10.18388/abp.2016_1329.
87. Antczak, M.; Zablocki, M.; Zok, T.; Rybarczyk, A.; Blazewicz, J. and Szachniuk, M. RNAvista: a webserver to assess RNA secondary structures with non-canonical base pairs. *Bioinformatics* **2019**, 35, 152-155, doi: 10.1093/bioinformatics/bty609.
88. Purzycka, K.J.; Garfinkel, D.J.; Boeke, J.D. and Le Grice, S.F. Influence of RNA structural elements on Ty1 retrotransposition. *Mobile genetic elements* **2013**, 3, e25060, doi: 10.4161/mge.25060.
89. Curcio, M.J. and Garfinkel, D.J. Heterogeneous functional Ty1 elements are abundant in the *Saccharomyces cerevisiae* genome. *Genetics* **1994**, 136, 1245-1259.
90. Pachulska-Wieczorek, K.; Le Grice, S.F. and Purzycka, K.J. Determinants of Genomic RNA Encapsulation in the *Saccharomyces cerevisiae* Long Terminal Repeat Retrotransposons Ty1 and Ty3. *Viruses* **2016**, 8, doi: 10.3390/v8070193.
91. Mailler, E.; Bernacchi, S.; Marquet, R.; Paillart, J.C.; Vivet-Boudou, V. and Smyth, R.P. The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses* **2016**, 8, doi: 10.3390/v8090248.
92. Blaszczyk, L.; Biesiada, M.; Saha, A.; Garfinkel, D.J. and Purzycka, K.J. Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression. *Viruses* **2017**, 9, doi: 10.3390/v9040074.
93. Jerabek-Willemsen, M.; Wienken, C.J.; Braun, D.; Baaske, P. and Duhr, S. Molecular interaction studies using microscale thermophoresis. *Assay and drug development technologies* **2011**, 9, 342-353, doi: 10.1089/adt.2011.0380.
94. Record, M.T., Jr.; Lohman, M.L. and De Haseth, P. Ion effects on ligand-nucleic acid interactions. *Journal of molecular biology* **1976**, 107, 145-158, doi: 10.1016/s0022-2836(76)80023-x.
95. Rouzina, I. and Bloomfield, V.A. Competitive electrostatic binding of charged ligands to polyelectrolytes: practical approach using the non-linear Poisson-Boltzmann equation. *Biophysical chemistry* **1997**, 64, 139-155, doi: 10.1016/s0301-4622(96)02231-4.

96. Comas-Garcia, M.; Datta, S.A.; Baker, L.; Varma, R.; Gudla, P.R. and Rein, A. Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. *eLife* **2017**, *6*, doi: 10.7554/eLife.27055.
97. Comas-Garcia, M.; Kroupa, T.; Datta, S.A.; Harvin, D.P.; Hu, W.S. and Rein, A. Efficient support of virus-like particle assembly by the HIV-1 packaging signal. *eLife* **2018**, *7*, doi: 10.7554/eLife.38438.
98. Rye-McCurdy, T.; Olson, E.D.; Liu, S.; Binkley, C.; Reyes, J.P.; Thompson, B.R.; Flanagan, J.M.; Parent, L.J. and Musier-Forsyth, K. Functional Equivalence of Retroviral MA Domains in Facilitating Psi RNA Binding Specificity by Gag. *Viruses* **2016**, *8*, doi: 10.3390/v8090256.
99. Webb, J.A.; Jones, C.P.; Parent, L.J.; Rouzina, I. and Musier-Forsyth, K. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: implications for viral genomic RNA packaging. *Rna* **2013**, *19*, 1078-1088, doi: 10.1261/rna.038869.113.
100. Muriaux, D.; Mirro, J.; Harvin, D. and Rein, A. RNA is a structural element in retrovirus particles. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 5246-5251, doi: 10.1073/pnas.091000398.
101. Jones, C.P.; Datta, S.A.; Rein, A.; Rouzina, I. and Musier-Forsyth, K. Matrix domain modulates HIV-1 Gag's nucleic acid chaperone activity via inositol phosphate binding. *Journal of virology* **2011**, *85*, 1594-1603, doi: 10.1128/JVI.01809-10.
102. Nilsen, T.W. Mapping RNA-protein interactions using hydroxyl-radical footprinting. *Cold Spring Harbor protocols* **2014**, *2014*, 1333-1336, doi: 10.1101/pdb.prot080952.
103. Szachniuk, M. RNAPolis: Computational Platform for RNA Structure Analysis. *Found Comput Decis S* **2019**, *44*, 241-257, doi: 10.2478/fcds-2019-0012.
104. Spitale, R.C.; Flynn, R.A.; Torre, E.A.; Kool, E.T. and Chang, H.Y. RNA structural analysis by evolving SHAPE chemistry. *Wiley interdisciplinary reviews. RNA* **2014**, *5*, 867-881, doi: 10.1002/wrna.1253.
105. Mitchell, D., 3rd; Assmann, S.M. and Bevilacqua, P.C. Probing RNA structure *in vivo*. *Current opinion in structural biology* **2019**, *59*, 151-158, doi: 10.1016/j.sbi.2019.07.008.
106. McGinnis, J.L.; Dunkle, J.A.; Cate, J.H. and Weeks, K.M. The mechanisms of RNA SHAPE chemistry. *Journal of the American Chemical Society* **2012**, *134*, 6617-6624, doi: 10.1021/ja2104075.
107. Merino, E.J.; Wilkinson, K.A.; Coughlan, J.L. and Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society* **2005**, *127*, 4223-4231, doi: 10.1021/ja043822v.
108. Busan, S.; Weidmann, C.A.; Sengupta, A. and Weeks, K.M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **2019**, *58*, 2655-2664, doi: 10.1021/acs.biochem.8b01218.
109. Andrade, J.M.; Dos Santos, R.F. and Arraiano, C.M. RNA Structure Analysis by Chemical Probing with DMS and CMCT. *Methods in molecular biology* **2020**, *2106*, 209-223, doi: 10.1007/978-1-0716-0231-7_13.

110. Siegfried, N.A.; Busan, S.; Rice, G.M.; Nelson, J.A. and Weeks, K.M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature methods* **2014**, *11*, 959-965, doi: 10.1038/nmeth.3029.
111. Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A.M.; Weissman, J.S. and Rouskin, S. DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nature methods* **2017**, *14*, 75-82, doi: 10.1038/nmeth.4057.
112. Vasa, S.M.; Guex, N.; Wilkinson, K.A.; Weeks, K.M. and Giddings, M.C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *Rna* **2008**, *14*, 1979-1990, doi: 10.1261/rna.1166808.
113. Karabiber, F.; McGinnis, J.L.; Favorov, O.V. and Weeks, K.M. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *Rna* **2013**, *19*, 63-73, doi: 10.1261/rna.036327.112.
114. Low, J.T. and Weeks, K.M. SHAPE-directed RNA secondary structure prediction. *Methods* **2010**, *52*, 150-158, doi: 10.1016/j.ymeth.2010.06.007.
115. Reuter, J.S. and Mathews, D.H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC bioinformatics* **2010**, *11*, 129, doi: 10.1186/1471-2105-11-129.
116. Wiegreffe, D.; Alexander, D.; Stadler, P.F. and Zeckzer, D. RNAPuzzler: efficient outerplanar drawing of RNA-secondary structures. *Bioinformatics* **2019**, *35*, 1342-1349, doi: 10.1093/bioinformatics/bty817.

Załącznik nr 1 do rozprawy doktorskiej „Identyfikacja i charakterystyka elementów strukturalnych genomowego RNA retrotranspozonu Ty1 oraz jego oddziaływań z białkiem Gag kluczowych dla dimeryzacji i pakowania genomu do cząstek wirusopodobnych”

ŻYCIORYS NAUKOWY

JULITA GUMNA

1. EDUKACJA

- 2015-2021 studia III stopnia, profil: biologiczny
Środowiskowe Studium Doktoranckie ICHB PAN
data otwarcia przewodu doktorskiego: 04.04.2019
- 2015 **Tytuł magistra biotechnologii**
Wydział Biologii, Uniwersytet im A. Mickiewicza w Poznaniu
Tytuł pracy dyplomowej: *Aspekty strukturalne oddziaływania RNA Ty1 z białkiem Gag: implikacje dla procesu retrotranspozycji*
Promotor: prof. dr hab. Ryszard W. Adamiak, dr Katarzyna J. Purzycka
Praca wykonana w Instytucie Chemii Bioorganicznej PAN w ramach projektu Homing Plus FNP „*How RNA structure governs Ty1 retrotransposition*”; kierownik projektu: dr Katarzyna J. Purzycka
- 2013-2015 studia II stopnia, kierunek: biotechnologia
Wydział Biologii; Uniwersytet im. A. Mickiewicza w Poznaniu
- 2013 **Tytuł licencjata biotechnologii**
Wydział Biologii, Uniwersytet im A. Mickiewicza w Poznaniu
Tytuł pracy dyplomowej: *Izolacja genu czynnika transkrypcyjnego HvZIP1 i charakterystyka jego mRNA ulegającego ekspresji w warunkach niedoboru wody w siewkach jęczmienia*
Promotor: dr Mateusz de Mezer
Praca wykonana w Instytucie Genetyki Roślin PAN w ramach projektu POLAPGEN-BD „*Narzędzia biotechnologiczne służące do otrzymywania zbóż o zwiększonej odporności na suszę*”; beneficjent: Instytut Genetyki Roślin PAN
- 2010-2013 studia I stopnia, kierunek: biotechnologia
Wydział Biologii; Uniwersytet im. A. Mickiewicza w Poznaniu

2. PUBLIKACJE

Julita Gumna, Angelika Andrzejewska-Romanowska, David J. Garfinkel and Katarzyna Pachulska-Wieczorek (2021) *RNA binding properties of the Ty1 LTR-retrotransposon Gag protein*; **International Journal of Molecular Sciences** 22(16):9103

Angelika Andrzejewska, Małgorzata Zawadzka, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek (2021) *In vivo structure of the Ty1 retrotransposon RNA genome*; **Nucleic Acids Research** 49(5):2878-2893

Julita Gumna, Tomasz Żok, Kacper Figurski, Katarzyna Pachulska-Wieczorek, Marta Szachniuk (2020) *RNAthor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis*; **PLoS One** 15(10):e0239287

Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel, Katarzyna Pachulska-Wieczorek (2019) *Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA*; **RNA Biology** 16: 1749-1763

Katarzyna Pachulska-Wieczorek, Leszek Błaszczuk, Julita Gumna, Yuri Nishida, Agniva Saha, Marcin Biesiada, David J. Garfinkel and Katarzyna J. Purzycka (2016) *Characterizing the functions of Ty1 Gag and the Gag-derived restriction factor p22/p18*; **Mobile Genetic Elements** 7;6(2):e1154637

Yuri Nishida, Katarzyna Pachulska-Wieczorek, Leszek Błaszczuk, Agniva Saha, Julita Gumna, David J. Garfinkel, Katarzyna J. Purzycka (2015) *Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions*; **Nucleic Acids Research** 43 (15): 7414-7431

3. DONIESIENIA KONFERENCYJNE

3.1. POSTER

The life cycle of retrovirus-like Ty1 transposon: the structure of genomic RNA in the cell and Gag function; A. Andrzejewska, M. Zawadzka, J. Gumna, D. J. Garfinkel, K. Pachulska-Wieczorek; RNA goes viral; Poznań 1-2.07.2021

Dissecting the mechanism of REGE-1 mediated mRNA silencing; D. Sobańska, A. Komur, P. Kumari, A. Chabowska-Kita, B. Juskowiak, J. Gumna, K. Pachulska-Wieczorek, K. J. Purzycka, R. Ciosk; RNA goes viral; Poznań 1-2.07.2021

Retrovirus-like transposons genome packaging: Gag function and structural determinants of Ty1 RNA dimerization; J. Gumna, K. J. Purzycka, H. W. Ahn, D. J. Garfinkel, K. Pachulska-Wieczorek, RNA 2019, Kraków 11-16.06.2019

From retrotransposon genome structure to function: Probing the structure of Ty1 genomic RNA in distinct biological states using SHAPE; A. Andrzejewska, M. Zawadzka, J. Gumna, D. J. Garfinkel, K. Pachulska-Wieczorek; RNA 2019, Kraków 11-16.06.2019

Regnase-1 and Roquin - partners in an evolutionarily conserved RNA decay mechanism; D. Sobańska, B. Juskowiak, A. Neagu, K. Pachulska-Wieczorek, K. J. Purzycka, J. Gumna, R. Ciosk, RNA 2019, Kraków 11-16.06.2019

Regulation of LTR-retrotransposons life cycle; J. Gumna, K. J. Purzycka, A. Saha, M. Zawadzka, D. J. Garfinkel, K. Pachulska-Wieczorek, 3rd Congress of Polish Biosciences BIO2018 “Through interdisciplinary approach into new solutions”, Gdańsk 18-21.09.2018

Structural determinants of Ty1 genomic RNA dimerization and packaging in the LTR-retrotransposon Ty1; J. Gumna, K. J. Purzycka, M. Zawadzka, D. J. Garfinkel, K. Pachulska-Wieczorek, 3rd Congress of Polish Biosciences BIO2018 “Through interdisciplinary approach into new solutions”, Gdańsk 18-21.09.2018

Determinants of genomic RNA dimerization and packaging in the Saccharomyces cerevisiae LTR-retrotransposon Ty1; J. Gumna, K. J. Purzycka, D. J. Garfinkel, K. Pachulska-Wieczorek, RNAtion Computational and Experimental RNA Biology Conference for Young Scientists, Poznań 18-19.09.2017.

Structural determinants of Ty1 genomic RNA dimerization: the role of Gag and p22 restriction factor; J. Gumna, K. J. Purzycka, A. Saha, D. J. Garfinkel and K. Pachulska-Wieczorek, Central European Congress of Life Science EUROBIOTECH, Kraków 11-14.09.2017.

Insight into the biological functions of retroelement's RNA through the predictions and analyses of their RNA 3D structures; M. Biesiada, K. Pachulska-Wieczorek, L. Błaszczyk, J. Gumna, R. W. Adamiak, K. J. Purzycka, RNA 2015, Madison (USA) 26-31.05.2015.

3.2. REFERAT

RNA dimerization in Ty1 LTR-retrotransposon; KNOW, Obrzycko 24-26.09.2018

Structural determinants of Ty1 genomic RNA dimerization; RNA&Computing, Biedrusko 17-18.05.2018.

4. KIEROWANIE PROJEKTAMI BADAWCZYMI

PRELUDIUM 18 NCN [2019/35/N/NZ1/01954]; „*Arc jako neuronalny Gag. Badanie wirusowych właściwości i funkcji głównego regulatora plastyczności synaptycznej*”

GRANT DLA MŁODYCH „*Wpływ dimeryzacji na stabilność RNA Ty1 w komórce i retrotranspozycję*” otrzymany w konkursie na finansowanie badań naukowych lub prac rozwojowych, służących rozwojowi młodych naukowców oraz uczestników studiów doktoranckich, ICHB PAN, 2018

GRANT DLA MŁODYCH „*Analiza funkcji sekwencji PAL w procesie dimeryzacji genomowego RNA retrotranspozonu Ty1*” otrzymany w konkursie na finansowanie badań naukowych lub prac rozwojowych, służących rozwojowi młodych naukowców oraz uczestników studiów doktoranckich, ICHB PAN, 2017

5. UDZIAŁ W PROJEKTACH BADAWCZYCH (WYKONAWCA)

OPUS 20 NCN [2020/39/B/NZ3/03020]; „*Analiza struktury RNA w komórce i jej kompartmentach w skali transkryptomowej oraz identyfikacja wpływu czynników komórkowych na strukturę RNA w S. cerevisiae*”; kierownik projektu: dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

SONATA 12 NCN [2016/23/D/NZ1/02565]; „*W jaki sposób struktura RNA kieruje funkcją macrycowego RNA? Aspekty strukturalne i funkcjonalne kluczowe dla syntezy białka p53 – głównego regulatora podstawowych procesów w komórce*”; kierownik projektu: dr Leszek Błaszczyk

OPUS 12 NCN [2016/23/B/ST6/03931]; „*RNApolis - metody i algorytmy do modelowania i analizy struktury RNA*”; kierownik projektu: prof. dr hab. inż. Marta Szachniuk

SONATA BIS 6 NCN [2016/22/E/NZ3/00426]; „*Struktura RNA i jego oddziaływania z białkiem Gag warunkujące funkcje RNA Ty1 na różnych etapach replikacji retrotranspozonu*”; kierownik projektu: dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

MAESTRO 3 NCN [2012/06/A/ST6/00384]; „*Automatyczne, wysokoprzepustowe modelowanie struktur przestrzennych RNA*”; kierownik projektu: prof. dr hab. Ryszard W. Adamiak

HOMING PLUS FNP [HOMING PLUS/2012-6/12]; „*How RNA structure governs Ty1 retrotransposition*”; kierownik projektu: dr Katarzyna J. Purzycka

6. SZKOLENIA

Szkolenie „Analiza i wizualizacja danych biologicznych w R”; Ideas4biology, Poznań; 15-16.02.2020

Nanotemper Workshop „Measure binding affinities and protein stability with NanoTemper Technologies”, ICHB PAN; Poznań; 25-26.06.2019

Szkolenie „NanoTemper Bioanalytics Meeting”; NanoTemper Technologies; ICHB PAN; Poznań; 20.04.2018

Szkolenie „GeXP Application Training: Introduction and Operation of GeXP Genetic Analysis System, Sequencing and Fragment Analysis Samples Preparation, Software training for analysis of Sequencing, Fragment and Gene Expression Data”; AB SCIEX; Darmstadt; Niemcy; 9-13.05.2016

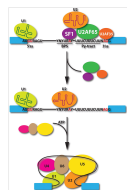
7. NAGRODY I STYPENDIA

Stypendium dla najlepszych doktorantów ICHB PAN 2020/2021

Stypendium dla najlepszych doktorantów ICHB PAN 2019/2020

Załącznik nr 2 do rozprawy doktorskiej „Identyfikacja i charakterystyka elementów strukturalnych genomowego RNA retrotranspozonu Ty1 oraz jego oddziaływań z białkiem Gag kluczowych dla dimeryzacji i pakowania genomu do cząstek wirusopodobnych”

PRACE NAUKOWE WCHODZĄCE W SKŁAD
ROZPRAWY DOKTORSKIEJ



Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA

Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel & Katarzyna Pachulska-Wieczorek

To cite this article: Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel & Katarzyna Pachulska-Wieczorek (2019): Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA, RNA Biology, DOI: [10.1080/15476286.2019.1657370](https://doi.org/10.1080/15476286.2019.1657370)

To link to this article: <https://doi.org/10.1080/15476286.2019.1657370>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



View supplementary material [↗](#)



Published online: 30 Aug 2019.



Submit your article to this journal [↗](#)



Article views: 360



View related articles [↗](#)







View Crossmark data [↗](#)

RESEARCH PAPER

 OPEN ACCESS 

Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA

Julita Gumna ^a, Katarzyna J. Purzycka ^a, Hyo Won Ahn^b, David J. Garfinkel ^b, and Katarzyna Pachulska-Wieczorek ^a

^aDepartment of Structure and Function of Retrotransposons, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland;

^bDepartment of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

ABSTRACT

During replication of long terminal repeat (LTR)-retrotransposons, their proteins and genome (g) RNA assemble into virus-like particles (VLPs) that are not infectious but functionally related to retroviral virions. Both virions and VLPs contain gRNA in a dimeric form, but contrary to retroviruses, little is known about how gRNA dimerization and packaging occurs in LTR-retrotransposons. The LTR-retrotransposon Ty1 from *Saccharomyces cerevisiae* is an informative model for studying LTR-retrotransposon and retrovirus replication. Using structural, mutational and functional analyses, we explored dimerization of Ty1 genomic RNA. We provide direct evidence that interactions of self-complementary PAL1 and PAL2 palindromic sequences localized within the 5'UTR are essential for Ty1 gRNA dimer formation. Mutations disrupting PAL1-PAL2 complementarity restricted RNA dimerization *in vitro* and Ty1 mobility *in vivo*. Although dimer formation and mobility of these mutants was inhibited, our work suggests that Ty1 RNA can dimerize via alternative contact points. In contrast to previous studies, we cannot confirm a role for PAL3, tRNA_i^{Met} as well as recently proposed initial kissing-loop interactions in dimer formation. Our data also supports the critical role of Ty1 Gag in RNA dimerization. Mature Ty1 Gag binds in the proximity of sequences involved in RNA dimerization and tRNA_i^{Met} annealing, but the 5' pseudoknot in Ty1 RNA may constitute a preferred Gag-binding site. Taken together, these results expand our understanding of genome dimerization and packaging strategies utilized by LTR-retroelements.

ARTICLE HISTORY

Received 16 May 2019
Revised 8 August 2019
Accepted 12 August 2019

KEYWORDS



LTR-retrotransposons; RNA dimerization; Ty1; Gag; RNA structure


Introduction

Retrotransposons are widespread in most eukaryotic genomes, and active elements are present in fungi, plants, insects and mammals, including humans. They influence the architecture and evolution of genomes, and can impact the expression of genes [1]. Furthermore, expression of retrotransposon RNAs and proteins, and in some cases the production of VLPs, accompanies various types of cancer and autoimmune diseases in humans [2,3], and enhances synaptic plasticity in *Drosophila* and mammals [4,5]. The packaging of gRNA into virions or VLPs is critical for propagation of infectious and endogenous LTR-retroelements. Extensive studies with retroviruses indicate that gRNA packaging is coupled with dimerization, which precludes gRNA encapsidation into virions and ensures the proper packaging specificity [6–9]. A variety of structure-based mapping strategies have been used to reveal the molecular mechanisms and structural determinants of gRNA dimerization in retroviruses [10–12]. In general, dimeric retroviral RNA is stabilized by base-pairing between *cis*-acting palindromic sequences located near the 5'-end of the gRNA [10,13]. The nucleocapsid (NC) domain of Gag is responsible for recognition of specific dimerization and packaging signals in viral RNA, and then Gag or mature NC nucleic acid chaperone activity promotes the structural

rearrangements of gRNA and interactions between RNA partners [8,10,14–16]. In addition, Gag participates in selective packaging of cellular tRNA into virions and facilitates its annealing with the complementary primer-binding site (PBS) in viral gRNA [17–19].

Much less is known about gRNA dimerization and packaging during propagation of endogenous retroelements. The Ty1 retrotransposon of *S. cerevisiae* is a well-studied member of the widely disseminated Ty1/Copia family of the Pseudoviridae [20]. Ty1 shares many critical features with endogenous and infectious LTR-retroelements [21]. For example, Ty1 contains partially overlapping GAG and POL genes flanked by LTRs at the 5' and 3' termini. GAG encodes the capsid protein of VLPs, and POL encodes enzymes required for replication (reverse transcriptase, protease, integrase). Analogous to other retroelements, Ty1 replicates via an RNA intermediate and full-length Ty1 gRNA plays dual role in replication [22]. It is translated into Ty1 proteins and utilized as a genome that is packaged into VLPs, and reverse transcribed to form a DNA copy that is integrated into the host genome [21,23,24]. Assembly of VLPs from Ty1 or the yeast Metavirus Ty3 occurs in specific cytoplasmic foci (termed retrosomes or T-bodies) where element-encoded proteins and gRNA colocalize during an early stage of retrotransposition [25–30]. Ty1 VLPs are mainly composed of

CONTACT Katarzyna Pachulska-Wieczorek  kasiapw@ibch.poznan.pl  Department of Structure and Funktion of Retrotransposons, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61–704, Poznan, Poland

 Supplemental data for this article can be accessed [here](#).

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Gag and contain Ty1-encoded enzymes and gRNA in a dimeric form [24]. During or soon after VLP assembly, the precursor Ty1 Gag polypeptide (Gag-p49) undergoes a single C-terminal cleavage that results in mature Gag-p45 (Fig. 1b) [31–34]. Despite the lack of sequence homology, Ty1 Gag is functionally related to retroviral Gag proteins as it comprises the capsid of VLPs and mediates critical RNA transactions during the retrotransposon replication cycle [21,28]. Direct interactions between Gag and gRNA are required for effective Ty1 retrotransposition and mutations in GAG can disrupt retroviral nucleation and VLP assembly [26,30,35–37]. Based on *in vitro* studies of truncated derivatives of Ty1 Gag, the RNA binding and nucleic acid chaperone region has been mapped to the C-terminus and contains three clusters of basic amino acid residues (Fig. 1b) [38,39].

Mutational and functional analyses identified a ~ 380 nt region at the 5'-terminus of the gRNA that is required for Ty1 retrotransposition [40,41]. The 5'-terminus of Ty1 gRNA contains *cis*-acting sequences required for packaging, reverse transcription, translation initiation and genome cyclization [42–45]. In VLPs, this region forms a long-range pseudoknot that is needed for efficient retrotransposition [46,47]. However, the sequences essential for Ty1 gRNA dimerization and packaging have not been precisely defined. An early study, based on *in vitro* dimerization assays with a synthetic peptide from a predicted chaperone domain of Ty1 Gag, suggested a role for tRNA_i^{Met} in Ty1 RNA dimerization and proposed a model for dimer formation mediated by interaction of two tRNA_i^{Met} molecules bound to the PBS sequences from two gRNAs [38]. More recently, SHAPE (selective 2'-hydroxyl acylation analysed by primer extension) [48] analysis of Ty1 gRNA *in vitro* and *in vivo* identified the changes in the flexibility of nucleotides contiguous to the three palindromic PAL sequences from the 5' terminus [46]. By analogy to retroviruses and in agreement with SHAPE reactivity profiles, PAL1, PAL2 and PAL3 sequences were suggested as potential contact sites for dimer formation. Two palindromes (PAL1 and PAL2) are located in 5'UTR, while PAL3 is located in the Gag open reading frame (Fig. 1a). However, the role of PALs in Ty1 RNA dimerization was not analysed. Recent work also suggests that two SL1a – SL3a kissing-loop interactions (SL1 – SL4 according to our notation, Fig. 1a) participate in Ty1

RNA dimerization [49]. Mutations disturbing the complementarity between SL1 and SL4 decreased Ty1 RNA stability *in vivo* and significantly inhibited Ty1 mobility, but the direct contact between these hairpins in dimerized RNA was not shown. The SL1 – SL4 kissing-loop complex is mutually exclusive with the SHAPE-based model of dimeric Ty1 RNA, as SL1 is formed by an intramolecular PAL1 and PAL2 interaction [46].

Here, we have taken a multifaceted approach to characterize the nucleotide sequences involved in Ty1 RNA dimerization and define features of monomeric and dimeric states of Ty1 RNA. The importance of selected sequences was assessed by their impact on RNA dimerization and structure, and Ty1 mobility. Our results indicate that similar to retroviruses dimeric Ty1 gRNA is maintained by interactions of self-complementary palindromic sequences within the 5'UTR. We show that intermolecular PAL1 – PAL2 interactions are important for *in vitro* dimerization of Ty1 RNA and Ty1 mobility *in vivo*. However mutations that disrupt PAL1 – PAL2 complementarity partially inhibit Ty1 RNA functionality and Ty1 RNA apparently dimerizes via alternative contact points. Our data support the critical role of Ty1 Gag in RNA dimerization. Interestingly, dimerization induced by Gag does not require initiator tRNA_i^{Met} annealing. Although mature Ty1 Gag binds in the proximity of sequences involved in RNA dimerization and tRNA_i^{Met} annealing, the 5' pseudoknot may serve as a major Gag-binding site. We propose that this interaction could be essential for Ty1 RNA stability during the process of retrotransposition.

Materials and methods

Genetic techniques, media, strains, and plasmids

Standard yeast genetic and microbiological procedures were used in this work [50]. Ty1 mobility was determined in the Ty1-less *S. paradoxus* strain DG3582 (*MATα his3-Δ200hisG trp1 ura3*) [39]. The helper-Ty1 and mini-Ty1*his3-AI* plasmids were kindly provided by M. Joan Curcio. The Ty1 helper plasmid (pEIB, 2μ *TRP1*) contains nucleotides 241–5561 of Ty1-H3 [42]. The mini-Ty1*his3AI* plasmid (pJC994, 2μ *URA3*) was constructed by

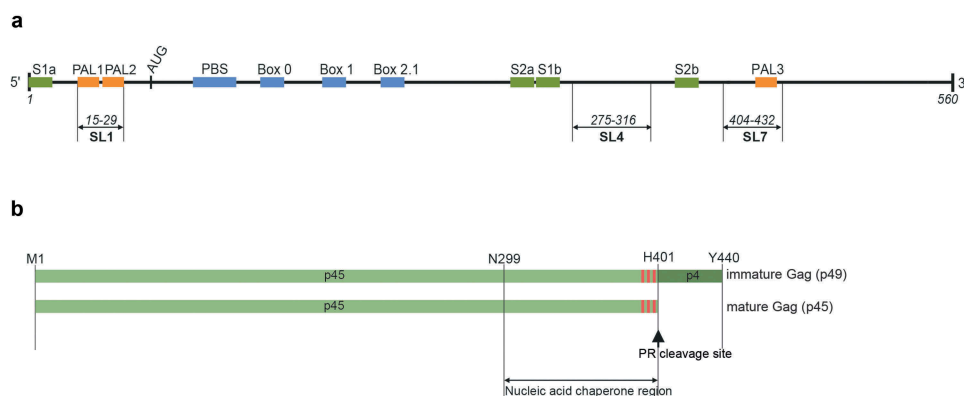


Figure 1. Organization of mini Ty1 RNA and Ty1 Gag. (a) Mini Ty1 RNA (+1–560) contains the primer binding site (PBS), three short boxes (Box 0, Box 1, Box 2.1; blue) that anneal with cellular tRNA_i^{Met}, and the palindromic sequences PAL1, PAL2, PAL3 (orange). The strands forming the stems of the pseudoknot (S1 and S2) are marked with green. (b). Schematic representation of Ty1 Gag: p49 (immature) and p45 (mature) with amino acid coordinates. The red boxes represent stretches of basic amino acids present in the nucleic acid chaperone region.

deleting the HpaI-SnaBI fragment of pGTy1*his3AI-Δ1* (818–5463 nt of Ty1-H3 DNA) [51]. Plasmids PAL1, PAL2, PAL3, and IL3 were derived from pJC994 by site-directed mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) following the manufacturer's protocol (for specific oligonucleotide sequences see Supplementary Table S1). Plasmid DNA was purified using PureLink Quick Plasmid Miniprep Kit (Invitrogen) and mutations were verified by DNA sequencing. The multicopy (2 μ) *TRP1*-based pEIB plasmid was introduced into DG3582 by transformation to generate DG4148. Multicopy (2 μ) *URA3*-based plasmids pJC994, PAL1, PAL2, PAL3, and IL3 were transformed individually into DG4148 to generate strains DG4151–DG4155, respectively, containing the helper and mini-Ty1*his3-AI* plasmids. Both Ty1 helper and mini-Ty1*his3-AI* element transcription is under the control of the *GAL1* promoter.

Ty1*his3-AI* mobility

Ty1 movement was monitored using the selectable indicator gene *his3-AI* [52]. A single colony of each strain was grown in SC-Ura-Trp 2% raffinose broth for 16 hrs at 30°C on a tube roller. Cultures diluted 25-fold into SC-Ura-Trp 2% galactose broth in quadruplicate were grown at 22°C for 2 days to induce Ty1 expression from the *GAL1* promoter. Cells were centrifuged, suspended in water and dilutions plated on SC-Ura-Trp 2% glucose and SC-Ura-Trp-His 2% glucose plates. Following incubation for 5 days at 30°C, the frequency of Ty1*his3-AI* mobility was calculated by dividing the number of Ura⁺ Trp⁺ His⁺ colonies by the number of Ura⁺ Trp⁺ colonies. Mobility analysis for wild type and the mutants were performed together, repeated at least twice, and representative data are shown. Note that Ty1*HIS3* insertions usually occur by retrotransposition following splicing of the artificial intron. Since His⁺ cells can also result from recombination of Ty1*HIS3* cDNA with Ty1 elements or solo-LTRs [53,54], the term Ty1 'mobility' is used to describe both types of insertion.

Cloning, expression and purification of Ty1 Gag-p45

Recombinant Gag-p45 was expressed from pGEX-p45 and purified as described previously [55]. Briefly, the Gag-p45-GST fusion protein was purified from *Escherichia coli* BL21 (DE3)pLysS strain (Invitrogen) by affinity chromatography using Glutathione Sepharose (GE Healthcare). The GST tag was removed by thrombin cleavage and Gag-p45 was eluted using a high – salt wash buffer containing 1 M NaCl, concentrated by centrifugal filtration and stored at –80°C.

DNA and RNA substrates

The template for transcription of unmodified yeast tRNA_i^{Met} was generated by PCR, and RNA was synthesized using T7-MEGashortscript (Invitrogen). RNA was purified by denaturing gel electrophoresis (8 M urea) in 1 x TBE buffer, eluted from the gel matrix and concentrated by ethanol precipitation. The template for *in vitro* transcription of mini Ty1 RNA (560 nt) was generated by PCR

amplification of sequences corresponding to the RNA nt +1–560 from pBDG433 using a forward primer F-miniRNA containing an SP6 promoter sequence followed by 5' Ty1 RNA sequence and a reverse primer R-miniRNA (Supplementary Table S2). DNA templates for PAL1, PAL2 and Δ S1a mini Ty1 RNA mutants were obtained using forward primers introducing desired mutations and primer R-mini RNA. Templates for the PAL3, SL4 and IL3 mutants were generated by PCR-driven overlap extension. All transcripts were synthesized using SP6-MEGascript (Invitrogen) and purified using Direct-zol RNA MiniPrep Kit (Zymo Research). The quality of transcripts was monitored by high-resolution agarose-gel electrophoresis in the presence of formaldehyde. Mini Ty1RNA and Ty1 RNA mutants were 3'-end labeled with [α -³²P] pCp using T4 RNA ligase (Fermentas) and purified on NucAway Spin Columns (Invitrogen Ambion). Purified RNAs were stored at –20°C.

Ty1 RNA dimerization

³²P-labeled Ty1 RNA (1 pmol) was refolded in buffer containing 40 mM Tris-HCl pH 8.0, 130 mM KCl, 0.1 mM EDTA by heating at 95°C for 3 minutes, slowly cooled to 60°C, placed on ice for 1 minute, and then incubated at 37°C for 30 minutes following the addition of MgCl₂ to 2 mM. For the dimerization assays with tRNA_i^{Met}, tRNA_i^{Met} was folded separately in equivalent conditions and Ty1 RNA was combined with unlabeled tRNA_i^{Met} at 1:2 molar ratio before addition of Ty1 Gag-p45. RNAs were incubated with increasing concentrations of protein at 37°C for 30 minutes. All reactions were quenched by incubation with 1% (w/v) SDS at room temperature for 5 minutes. The samples were phenol/chloroform extracted and 10 μ l of aqueous phase was mixed with 2 μ l of 50% glycerol. RNA was resolved on a 1% agarose gel in 0.5 x TB at room temperature or TBM (TB with 0.2 mM MgCl₂) running buffer at 4°C. Gels were autoradiographed and quantified by phosphorimaging using FLA-5100 phosphorimager with MultiGaugeV 3.0 software (FujiFilm). The obtained data were analysed using Origin software (OriginLab). In all cases, at least three independent experiments were performed and the data presented are representative of the whole.

Ty1 RNA structure probing

RNAs (8 pmol) were refolded in 20 μ l of renaturation buffer (10 mM Tris-HCl [pH 8.0], 100 mM KCl, and 0.1 mM EDTA) by heating for 3 minutes at 95°C, slow cooling to 4°C, then adding 100 μ l of water and 30 μ l of 5 \times folding buffer (final concentration: 40 mM Tris-HCl [pH 8.0], 130 mM KCl, 0.2 mM EDTA, and 4 mM MgCl₂), followed by incubation for 25 minutes at 37°C. The RNA sample was divided into two tubes and treated with 8 μ l of NMIA in DMSO ([+], final concentration 1 mM) or DMSO alone (–), and the modification reaction was carried for 50 minutes at 37°C. RNA was recovered by ethanol precipitation and resuspended in 10 μ l of water.

Detecting Gag-induced changes in RNA structure

Ty1 RNA (8 pmol) and tRNA_i^{Met} (16 pmols) were folded as described for dimerization assays. Subsequently, 240 pmols of Gag-p45 (8 µl in the protein buffer containing 50 mM Tris-HCl, pH 8.0, 1 M NaCl, 10 mM β-mercaptoethanol, 2.5 mM DTT) or equal volume of protein buffer were added to 72 µl of RNA mixture. The reaction was incubated at 37°C for 30 minutes and quenched by incubation with 1% SDS at room temperature for 5 minutes followed by phenol/chloroform extraction. Each reaction was divided into two separate tubes and treated with 8 µl of NMIA in DMSO [(+), 1.5mM NMIA, final concentration] or DMSO alone (-). The modification reactions were carried out for 50 minutes at 24°C. RNA was recovered by ethanol precipitation and was resuspended in 10 µl of water. Detection of 2'-O-adducts and data processing were performed as described below. The contribution of Ty1 RNA dimer to the ensemble SHAPE reactivity profile(s) was calculated according to:

$$R_E = R_M p_M + R_D p_D$$

Where R_E is the overall reactivity obtained experimentally, R_M , R_D are the reactivities for monomer and dimer and p_M , p_D are the per cent contribution from monomer/dimer to total RNA. To reduce bias produced by extreme values, calculated reactivity values greater than 1 were set equal to 1 and negative values were set to 0 [56].

Hydroxyl radical footprinting

RNA samples (5 pmol) were refolded by heating at 95°C for 1 minute followed by incubation for 25 minutes at 37°C in folding buffer (40 mM Tris-HCl, pH 8.0, 130 mM KCl, 0.5 mM EDTA and 5 mM MgCl₂). Folded mini Ty1 RNA samples were diluted 20-fold with 20 mM Tris-HCl, pH 8.0 prior to adding Ty1 Gag-p45. Subsequently, 50 or 100 pmol of Gag-p45 in a total volume of 6 µl was added to a 70 µl reaction. As a control for non-specific cleavage, buffer was added instead of protein. RNA/protein complexes were formed by incubating 20 minutes at 0°C. To initiate footprinting reactions, 1.5 µl of 2.5 mM (NH₄)Fe(SO₄)₂, 50 mM sodium ascorbate, 1.5% H₂O₂ and 2.75 mM EDTA were applied on the wall of the tube followed by centrifugation [57]. Reactions were incubated for 15 seconds at 24°C and quenched by addition of thiourea and EDTA to final concentrations of 20 mM and 40 mM, respectively. RNA was purified using Direct-zol RNA MiniPrep Kit (Zymo Research).

Primer extension reactions and data processing

A total of 10 pmols of fluorescently labeled primer PR3 [5'-TCAGGTGATGGAGTGCTCAG-3'] was added to 4 pmols of RNA [Cy5 (+) and Cy5.5 (-)], and 12 µl of primer-template was incubated at 95°C for 3 minutes, 37°C for 10 minutes and 55°C for 2 minutes, then reverse transcribed as described previously [46]. Sequencing ladders were prepared using primers labeled with WellRed D2 (ddA) or LicorIRD-800 (ddT) and a Thermo Sequenase Cycle Sequencing kit (Affymetrix) according to the manufacturer's protocol. Samples and sequencing ladders were purified using ZR DNA Sequencing Clean-up Kit (Zymo

Research) and analysed on a GenomeLab GeXP Analysis System (Beckman-Coulter). Electropherograms were processed using SHAPEfinder software [58], normalized as described previously [18] and converted into nucleotide reactivity tables [59]. All reactivity data used in analysis were averaged from at least three independent experiments. Tables with HR footprinting and SHAPE data are provided in Supplementary Data set 2.

RNA structure prediction

Secondary structure models were generated using RNAstructure software with default parameters [60]. RNAComposer was used for prediction of RNA 3D structural models (<http://rnacomposer.cs.put.poznan.pl/>). This is a fully automated RNA structure modeling server that predicts RNA 3D structures based on sequence and secondary structure [61–63].

Results

Requirements for *in vitro* dimerization of mini Ty1 RNA

Similar to retroviral RNAs, full-length Ty1 gRNA (5900 nt) dimerizes inefficiently *in vitro*, even in the presence of recombinant Gag-p45 (Fig. S1). To analyse Ty1 RNA dimerization and further define nucleotide sequences and structural motifs required for this process, we used a mini Ty1 RNA that encompasses +1–560 nucleotides of Ty1 gRNA and contains sequences suggested previously as candidates for mediating Ty1 gRNA dimerization, packaging into VLPs, and undergoing reverse transcription and integration (Fig. 1a) [22,38,41,46,49]. To determine if mature Ty1 Gag-p45 displays properties similar to retroviral Gag polyproteins, we purified recombinant Gag-p45 from *E. coli* and examined whether Gag-p45 enhanced mini Ty1 RNA dimerization *in vitro*. ³²P-labeled mini Ty1 RNA was incubated with the increasing amounts of the Gag-p45 and dimeric RNA formation was assayed by electrophoresis under TB and TBM (includes MgCl₂) running conditions after protein removal. This experimental approach allows one to distinguish 'loose dimers', which are mediated by kissing loops from 'tight dimers', which may be held together by extended intermolecular duplexes [64–67]. Kissing loop complexes require Mg²⁺ for stabilization and are only detectable by electrophoresis using running buffers containing Mg²⁺ that prevent electrophoresis-dependent dissociation. In contrast, tight dimers are stable under mild denaturing conditions (TB or TBE running buffers) in the absence of Mg²⁺.

Mini Ty1 RNA formed dimers in the presence of Gag-p45 (Fig. 2), whereas the RNA migrated as a single band corresponding to the monomeric form under TBM and TB electrophoresis conditions in the absence of Gag-p45. Incubation of mini Ty1 RNA with increasing amounts of Gag-p45 resulted in significant conversion toward the dimeric state. At a 3 µM concentration of Gag-p45 (1:23 protein to nt molar ratio), ~45% of mini Ty1 RNA was present in a dimeric state. Dimerization assays were performed in the presence (Fig. 2a) or absence of tRNA_i^{Met} (Fig. 2b) to directly assess whether tRNA_i^{Met}–tRNA_i^{Met} interactions act as a 'bridge' in Ty1 RNA dimerization, as suggested by earlier work [38]. The inclusion

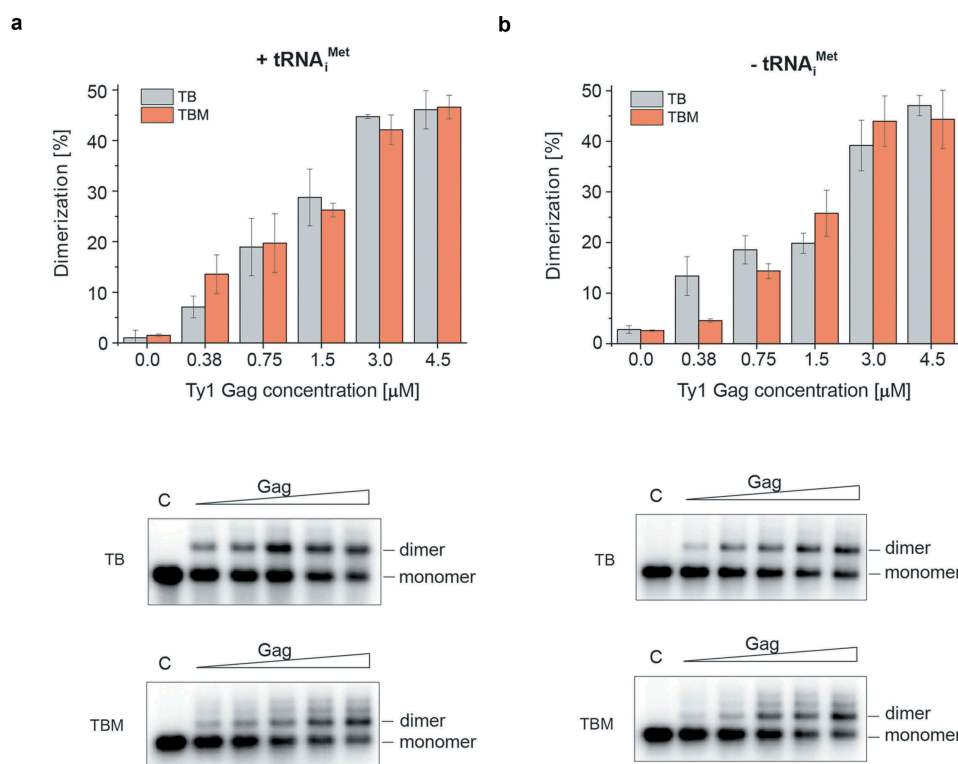


Figure 2. Gag-mediated dimerization of mini Ty1 RNA in the presence (a) and absence (b) of tRNA_i^{Met} analysed in the presence or absence of Mg²⁺. The graphs represent dimerized RNA (%) with increasing concentrations of Gag-p45 (0; 0.375; 0.75; 1.5; 3; 4.5 μM). Samples were analysed by agarose gel electrophoresis in TBM (2 mM MgCl₂, orange) and TB (no MgCl₂, grey). Representative analyses of mini Ty1 RNA dimerization are presented at the bottom. Lanes denoted C represent control samples that lack protein.

of tRNA_i^{Met} did not alter the monomer – dimer ratio, indicating that similar to retroviruses, dimeric Ty1 RNA is stabilized by direct base pairing of complementary *cis*-acting sequences. A similar efficiency of dimerization was obtained under TB or TBM running conditions, showing that mature Gag promotes formation of tight mini Ty1 RNA dimers. However, two discrete bands migrating slower than the major dimer were visible in the presence of Mg²⁺ (Fig. 2, bottom panel TBM gels). These bands may contain the higher-order multimers of mini Ty1 RNA or transient forms of dimeric RNA that are unstable during electrophoresis when the running buffer lacks Mg²⁺.

Gag binding sites on mini Ty1 RNA

Hydroxyl radical (HR) footprinting was used to map Gag-p45 binding sites within mini Ty1 RNA. This method offers an advantage over SHAPE analysis for detecting sequences protected by protein binding because HR cleavage of the RNA backbone occurs independently of RNA secondary structure [68]. Ribonucleoprotein complexes were formed on ice to minimize Gag-induced structural changes of mini Ty1 RNA and maintain the monomeric state of the RNA (Fig. S2). Overall reactivity profiles for mini Ty1 RNA in the presence or absence of recombinant Gag-p45 were compared to reveal protected nucleotide sequences (Fig. 3a). HR footprinting data were correlated with the SHAPE-based secondary structure model of the mini Ty1 RNA monomer (Figs. 3b & S3). The strongest and statistically significant decrease in HR cleavage

was observed in sequences forming the Ty1 pseudoknot, raising the possibility that this is a preferred region for Gag-p45 binding. We also observed smaller but statistically significant protection in three regions located downstream of PBS and in the apical part of the SL1 hairpin that is formed by intramolecular base pairing between PAL1 and PAL2 palindromes (Figs. 3b & S3). Analysis of the composition of protected sites revealed that the content of C, A and U residues in the protected stretches of nucleotides were similar, while the number of G residues was significantly lower (Fig. 3c). These results suggest a lack of Ty1 Gag preference for particular nucleotide base(s), whereas retroviral Gag binding sites contain a high purine content (A and/or G) [18,69].

SHAPE analysis of wild type mini Ty1 RNA in monomeric and dimeric states

To further explore Ty1 RNA dimerization, we analysed the structure of the Gag-induced mini Ty1 RNA dimers by SHAPE [48,70]. In addition to mini Ty1 RNA, tRNA_i^{Met} was also present in the reaction mixtures as dimerization and primer binding are two critical RNA-RNA hybridization processes occurring during Ty1 VLPs assembly. When Gag-p45 was not added to the reaction, dimers were not formed and tRNA_i^{Met} was not annealed. Therefore, this reaction was used to monitor the monomeric state prior tRNA_i^{Met} annealing (thereafter referred to as monomer). Addition of Gag-p45 induced both dimerization of mini Ty1 RNA and tRNA_i^{Met} annealing (thereafter referred to as dimer). Prior to SHAPE analysis, protein was

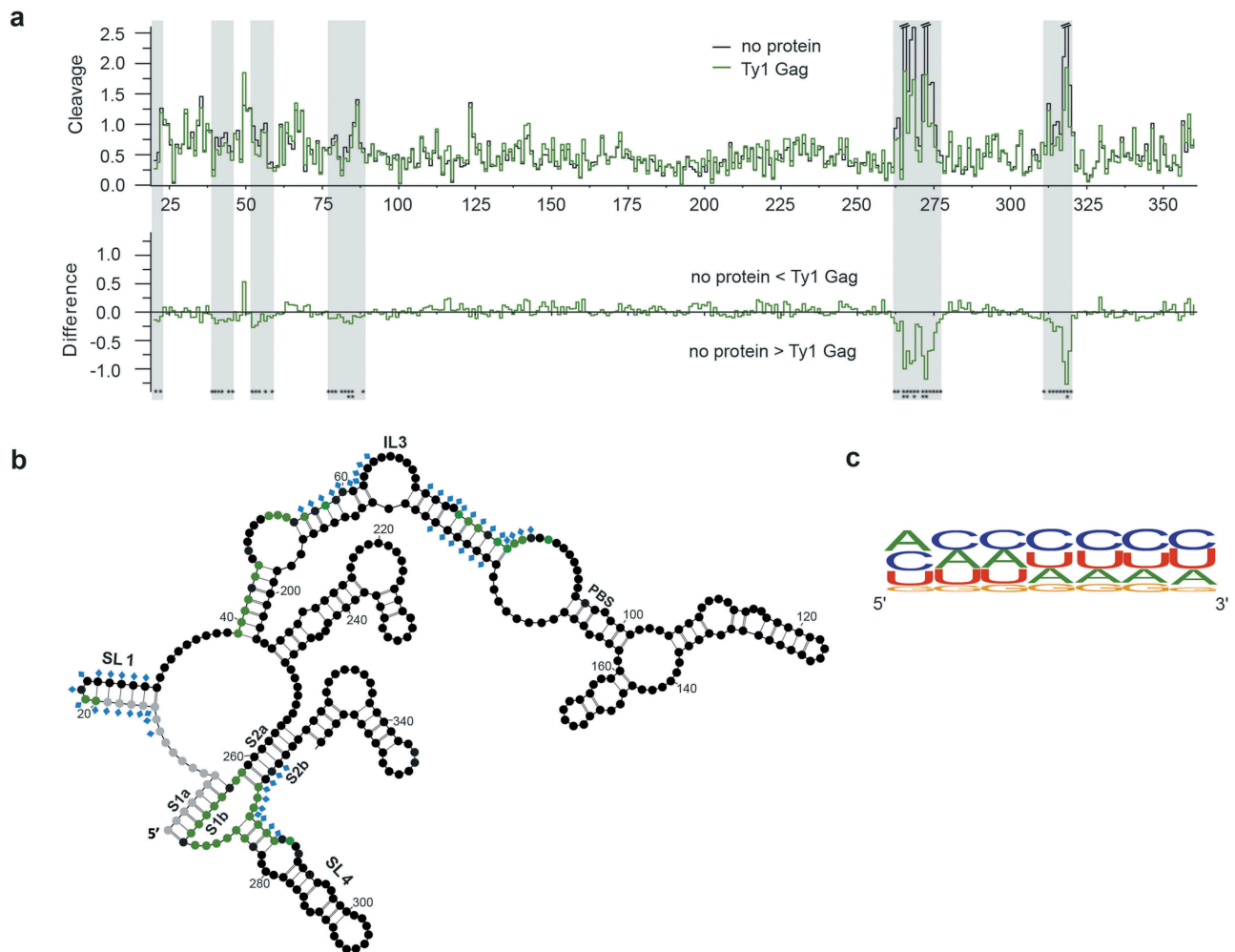


Figure 3. Gag-p45 binding sites in mini Ty1 RNA. (A) Hydroxyl radical (HR) cleavage and difference plot analyses of protein free mini Ty1 RNA in comparison with RNA probed in the presence of Gag-p45. Regions showing statistically significant decrease HR reactivity over several nucleotides are indicated by gray stripes (absolute cleavage decrease of 0.1 or greater and a p-value <0.05). Sites of decreased HR cleavage upon Gag-p45 binding are indicated by negative peaks. Asterisks below the plot correspond to statistical significance (Student's t-test). (B) 2D structure model of +1–358 region of mini Ty1 RNA with the positions protected from HR cleavage in the presence of Gag-p45 coloured with green (absolute cleavage decrease of 0.1 or greater and a p-value <0.05). Not analysed nucleotides are marked with gray. Nucleotide positions which most likely correspond to protein binding sites within the Ty1 genomic RNA in VLPs are marked with blue diamonds [46]. Tables with HR footprinting data are provided in Supplementary Data set 2. (C) Frequency of nucleotide occurrence within Gag-p45 binding sites and their vicinities represented as a logo (<https://weblogo.berkeley.edu/>).

gently removed by phenol-chloroform extraction. Since the maximal efficiency of dimerization was ~ 50%, the mini Ty1 RNA population was a mixture of monomers and dimers after incubation with Gag-p45. To address this problem, we applied a mathematical method that allows one to determine the secondary structures of individual RNAs in a mixture of the conformers. This method was used successfully to analyse RNA folding intermediates of the HIV-2 Rev-responsive element [56]. Accordingly, the reactivity value of each nucleotide obtained for the mixture of two RNA states is a sum of reactivity values of both states, weighted according to the fractional contribution of each to the total RNA population. Structural differences were detected both by comparing monomer and dimer SHAPE reactivity profiles and by subtracting individual nucleotide reactivities for the monomer from values of dimer to yield a difference plot (Fig. 4a).

The SHAPE profiles for mini Ty1 RNA monomer and dimer are similar, indicating that Gag-induced transition from the monomer to dimer state does not cause extensive refolding of the RNA (Fig. 4a). Despite this overall similarity, we observed statistically significant reactivity differences that reveal important local changes in RNA conformation. NMIA reactivity alterations in the region +82–176 can be attributed to the formation of a mini Ty1 RNA/tRNA_i^{Met} complex in the presence of Gag-p45. These changes correspond well with NMIA reactivity profiles characteristic for Ty1 RNA/tRNA_i^{Met} complex described earlier [46]. The tRNA_i^{Met} binding sequences constitute a specific border in SHAPE reactivity (Fig. 4a). In the downstream region, stretches of small positive peaks are revealed by difference plots, while upstream from the PBS, stretches of both positive and negative peaks are detected. Most of positive peaks correspond to increases in local nucleotide reactivity resulting from a structure destabiliz-

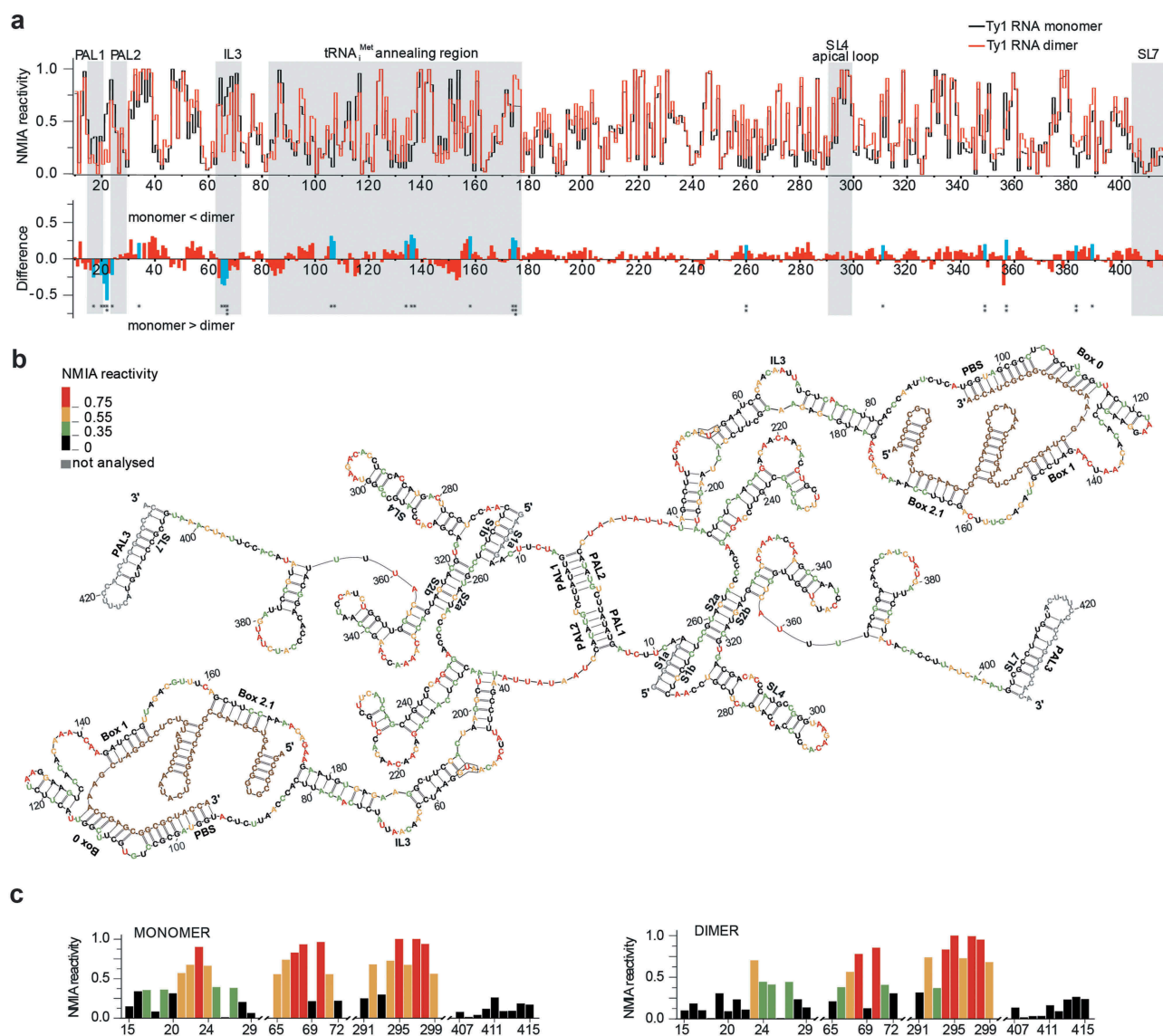


Figure 4. SHAPE analysis of mini Ty1 RNA. (a) Comparison of SHAPE reactivities for mini Ty1 RNA in monomeric and dimeric states. At the top is the step plot of NMIA reactivity for monomeric (black) and dimeric RNA (red). At the bottom is the difference plot calculated by subtracting the monomer intensities from those of the dimer. Negative values indicate nucleotides that are less flexible in the dimer. Regions of the interest are indicated by gray stripes. Columns corresponding to nucleotides that exhibit statistically significant differential reactivity (10% of highest SHAPE reactivity differences and a p -value < 0.05 using the Student's t -test) are coloured in blue. Asterisks below the plot correspond to statistical significance. (b) +1–433 region of mini Ty1 RNA in the dimeric form. Nucleotide residue reactivity to NMIA are scaled by colour (see insert on the left). The $tRNA_i^{Met}$ is shown in brown. (c) Colour-coded histograms of NMIA reactivity for mini Ty1 RNA in monomeric and dimeric states. Tables with SHAPE data are provided in Supplementary Data set 2.

ing activity of Gag-p45 that is similar to the nucleic acid chaperone activity of retroviral Gag polyproteins [28,38,39]. The stretches of negative peaks are attributable to intermolecular interactions critical for dimer formation. The strongest and statistically significant reactivity decrease was observed in the 21UCU23 sequence linking PAL1 and PAL2. In the monomer, nucleotides 21UCU23 were highly reactive since they comprise apical loop of the SL1 hairpin formed by intramolecular base-pairing of PAL1 and PAL2 (Figs. 4c & S2). PAL1 and PAL2 palindromic sequences were unreactive in both monomer and dimer states, therefore, it is challenging to prove that they constitute contact sites in the dimer. Importantly, the observed reactivity decrease in 21UCU23 is independent from $tRNA_i^{Met}$ annealing (Fig. S4) and supports a Gag-induced structural transition that may correspond to the SL1 hairpins forming an

intermolecular extended duplex with a UCU/UCU distortion (Fig. 4b). Alternatively, reactivity changes in SL1 apical part may arise from intermolecular kissing loop interactions between SL1 and SL4 that were recently proposed to help initiate Ty1 RNA dimerization [49]. Since the SL4 apical loop was highly reactive in both monomeric and dimeric states (Figs. 4c & S2), these kissing loop interactions are unlikely to be present in dimeric mini Ty1 RNA. Similar to PAL1 and PAL2, PAL3 was also unreactive in the monomeric and dimeric states. This palindrome is present in the stem of SL7 (Figs. 4c & S2) and its engagement in intermolecular base-pairing is anticipated to disturb the structure of SL7 and increase reactivity in the complementary strand [46]. However, we did not observe SHAPE reactivity changes in the SL7 stem (Fig. 4c) that would support PAL3-PAL3 interactions in the Ty1 RNA dimer.

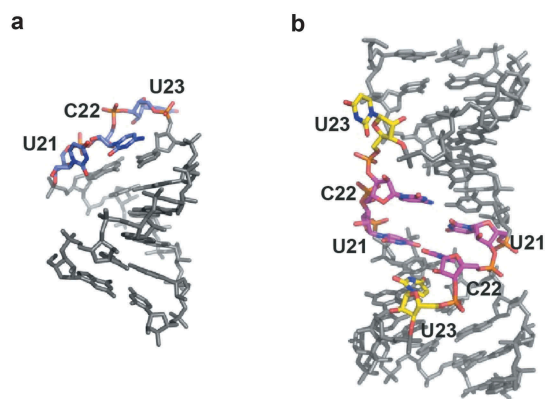


Figure 5. 3D structure models of SL1 (a) and the PAL1-PAL2 intermolecular duplex (b), characteristic for monomeric and dimeric RNA, respectively. In SL1, nucleotides 21UCU23 are marked with blue. In duplex U21 and C22 are marked with magenta, U23 are marked with yellow.

In addition to the change of 21UCU23 reactivity in dimeric mini Ty1 RNA, a reproducible statistically relevant decrease in reactivity was detected for nucleotides in the third internal loop – IL3 (nt 63–71) (Fig. 4a). This asymmetric internal loop contains a short 67AAUU70 palindromic sequence that could serve as an additional dimerization site (Fig. 4b). However, despite the reactivity decrease in IL3, nucleotides A67, A68 and U70 in the palindrome remained reactive to NMIA in the dimer (Figs. 4b, c), suggesting this interaction is less stable or transient in nature. In support of the validity of our approach, a similar change in the IL3 reactivity pattern was detected between full-length monomeric Ty1 RNA probed *in vitro* and in dimeric Ty1 RNA probed within VLPs [46].

3D models of PAL1-PAL2 interactions in monomer and dimer

Our data suggest the SHAPE reactivity changes in the 21UCU23 sequence linking PAL1 and PAL2 result from a Gag-induced monomer to dimer transition. To extend and validate these findings, we generated 3D structure models of SL1 and the PAL1-PAL2 extended duplex (Fig. 5). Models were generated using RNAComposer, which is an advanced tool for modeling atomic-resolution 3D models of RNAs based on their sequence and secondary structures [61,62]. Predicted structural models correspond well with the NMIA reactivity changes observed in the linking sequence. In monomeric Ty1 RNA, nucleotides 21UCU23 constitute the apical loop of SL1 and were highly reactive to NMIA. The 3D

structure model of PAL1-PAL2 extended duplex suggests formation of non-canonical C-U intermolecular base-pairs involving U21 and C22 residues. Non-canonical C-U base pairs were identified in several biologically important RNAs and RNA duplexes containing the C-U base pairs can be considered similar to typical A-RNA helices [71,72]. The preservation of continuous stacking contributes significantly to the stability of RNA folds and minimizes the exposure of hydrophobic base surfaces to polar solvent. This can explain the lack of U21 and C22 NMIA modification in dimeric mini Ty1 RNA. Moreover, U23 residues were predicted to be extruded from the duplex with the base pointing towards the solvent. Such looped out nucleotides would allow the overall duplex geometry to remain close to the regular A-form. In agreement with the 3D structure model, U23 residues were readily modified by NMIA in dimeric mini Ty1 RNA.

Mutational analysis of cis-acting sequences involved in mini Ty1 RNA dimerization *in vitro*

To further define the role of candidate sequences in Ty1 RNA dimerization, we analysed *in vitro* dimerization of mini Ty1 RNA mutants with nucleotide substitutions in the postulated contact sites (Table 1). Our data reveal that substitution of PAL1 or PAL2 by non-palindromic sequences (PAL1 mutant and PAL2 mutant, respectively) results in a two-fold decrease in dimerization when compared to wild type mini Ty1 RNA (Figs. 6a & S5). Investigation of a mini Ty1 RNA mutant with a single nucleotide substitution (U16C) in PAL1 additionally confirmed the role of this palindrome in Ty1 RNA dimerization (Fig. S6). The U16C RNA mutant dimerized less efficiently than wild type but as expected, the inhibition was weaker than observed for the PAL1 non-palindromic mutant. U16C disturbs the U16-A28 base pair and thereby might destabilize the PAL1-PAL2 intermolecular interaction. However, mutations in SL1 hairpin could also impact the potential SL1-SL4 interactions [49] and thus, the observed inhibitory effects might result from disturbance of an initial kissing loop complex. Despite the lack of support for this possibility from our SHAPE analysis, we also assayed dimerization of a mini Ty1 RNA mutant with substitutions in the SL4 apical loop (SL4 mutant) that disturb SL1-SL4 complementarity and impairs Ty1 retrotransposition [49]. We observed that mutations in SL4 altered monomer to dimer ratio only slightly, indicating that the SL1-SL4 kissing complex is not required for dimerization of mini Ty1 RNA *in vitro* (Figs. 6b & S5).

Table 1. Ty1 RNA mutants used in this study.

Mutant	Description of introduced mutations	Sequence of wild type mini Ty1 RNA and mutants
U16C mutant	Single nucleotide substitution in PAL1 sequence	15GUAAU20 → GCAUUA
PAL1 mutant	Substitution of PAL1 sequence with non palindromic sequence	13UAGUAUA19 → CGCCCGC
PAL2 mutant	Substitution of PAL2 sequence with non palindromic sequence	24GUUAUC29 → CGCCCG
PAL3 mutant	Substitution of PAL3 sequence with non palindromic sequence	423CCUGGG428 → AAGAGG
IL3 mutant	Substitution of internal loop 3 sequence	65ACAAUUUAU72 → CCAACCAU
SL4 mutant	Substitution of SL4 apical loop sequence	291CCACAGAAU299 → CCUCUCUAA
ΔS1a mutant	Deletion of 5 nt within S1a pseudoknot stem	1GAGGA5 → - - - - -

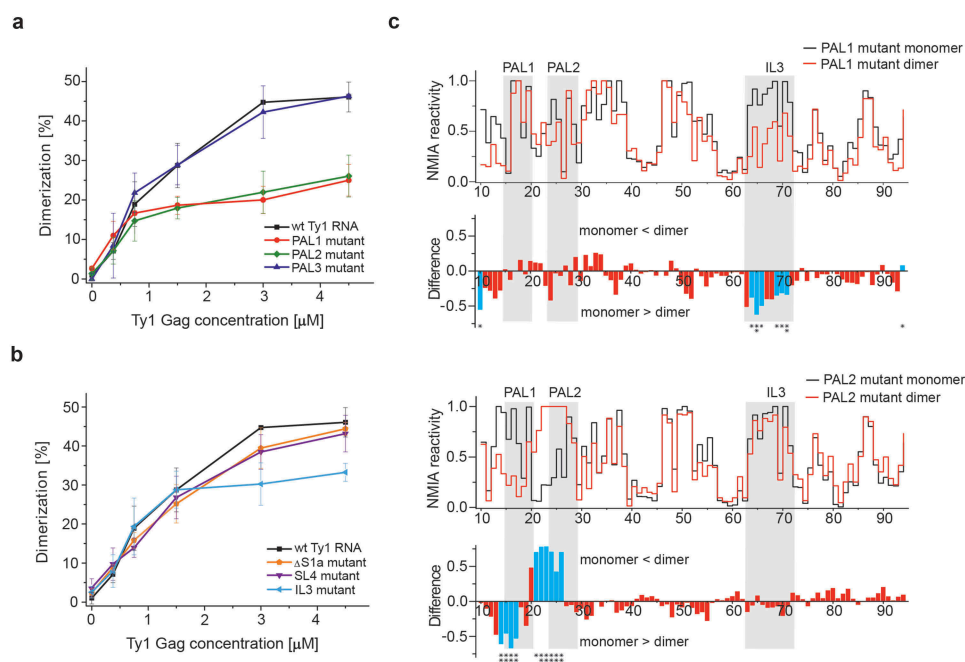


Figure 6. Identification of *cis*-acting sequences important for mini Ty1 RNA dimerization. (a) Dimerization assays for wild type mini Ty1 RNA and PAL1, PAL2, PAL3 mutants. (b) Dimerization assays for wild type mini Ty1 RNA and Δ S1a, SL4 and IL3 mutants. The graphs represent averaged measurements from three independent experiments. (c) SHAPE analysis of PAL1 and PAL2 RNA mutants in monomeric and dimeric states. For clarity, the plots show the +10–90 region of mini Ty1 RNA mutants. At the top is the step plot of NMA reactivity for monomeric (black) and dimeric (red) RNA. At the bottom is the difference plot calculated by subtracting the monomer intensities from those of the dimer. Negative values indicate nucleotides that are less flexible in the dimer. Columns corresponding to nucleotides that exhibit statistically significant differential reactivity (10% of highest SHAPE reactivity differences and a p-value <0.05 using the Student's t-test) are coloured in blue. Asterisks below the plot correspond to statistical significance. Tables with SHAPE data are provided in Supplementary Data set 2.

Taken together, our results suggest that PAL1-PAL2 intermolecular interactions play an important role in Ty1 RNA dimerization. However, dimerization of RNA containing PAL1 or PAL2 mutations was reduced but not abolished completely, suggesting another sequence could facilitate dimerization in addition to PAL1 and PAL2. Although SHAPE analysis of Ty1 RNA in VLPs raises the possibility that a PAL3–PAL3 interaction influences dimer formation [46], a non-palindromic PAL3 mutant displayed a dimerization efficiency comparable to that of wild type mini Ty1 RNA (Fig. 6a), suggesting that PAL3 is dispensable for dimerization *in vitro*. We also analysed dimerization of mini Ty1 RNA with substitutions in IL3 (IL3 mutant) because NMA reactivity of dimeric RNA decreased in the IL3 internal loop. The IL3 mutant dimerized similar to wild type mini Ty1 RNA at lower Gag-p45 concentration, but less efficiently at higher Gag-p45 concentration (Figs. 6b & S5). However, the decrease in IL3 RNA mutant dimerization was weaker than that observed for the PAL1 or PAL2 non-palindromic RNA mutants.

SHAPE analysis of monomeric and dimeric mini Ty1 RNA containing non-palindromic mutations

To exclude mutation-induced RNA misfolding that could indirectly affect dimerization, we performed SHAPE analysis of non-palindromic RNA mutants in monomer form. The obtained reactivity profiles of PAL1, PAL2 and PAL3 non-palindromic RNA mutants resembled closely that of monomeric wild type mini Ty1 RNA and the majority of statistically significant reactivity changes were found only for the mutated sequences and

adjacent nucleotides (Fig. 7a–c). These results confirm that the observed perturbations in dimerization are not caused by extensive alterations in the secondary structure of the mutated RNAs extending beyond the altered sequence. Interestingly, the IL3 RNA mutant displayed decreased NMA reactivity in residues of the SL1 apical loop (Fig. 7d). Therefore, the IL3 mutation may influence the architecture of the 5' terminus of Ty1 RNA and indirect effect on mini Ty1 RNA dimerization *in vitro* is also possible.

We also characterized PAL1 and PAL2 non-palindromic RNA mutants in the dimeric form by SHAPE as described above (Fig. 6c). Similar to wild type mini Ty1 RNA, differences in SHAPE reactivity between monomeric and dimeric RNAs were limited to the regions (i) +82–176 (indicating tRNA annealing) and (ii) upstream from the PBS. For the PAL1 mutant, a stretch of statistically relevant high negative peaks was observed in the region corresponding to IL3. This loss of reactivity in the dimeric state suggests that the PAL1 mutant dimerizes via an interaction between palindromic sequences located in the IL3. The decrease in SHAPE reactivity in this region was also observed for wild type mini Ty1 RNA dimer but was less evident when PAL1-PAL2 intermolecular interactions were present (Fig. 4a). This effect was not detected for the PAL2 mutant, where a surprisingly strong and statistically significant reactivity decrease was identified in PAL1 sequence, suggesting intermolecular PAL1-PAL1 interaction in dimeric PAL2 RNA mutant. Taken together, our observations suggest that alternate dimerization contacts are utilized when the wild-type dimerization sites are mutated.

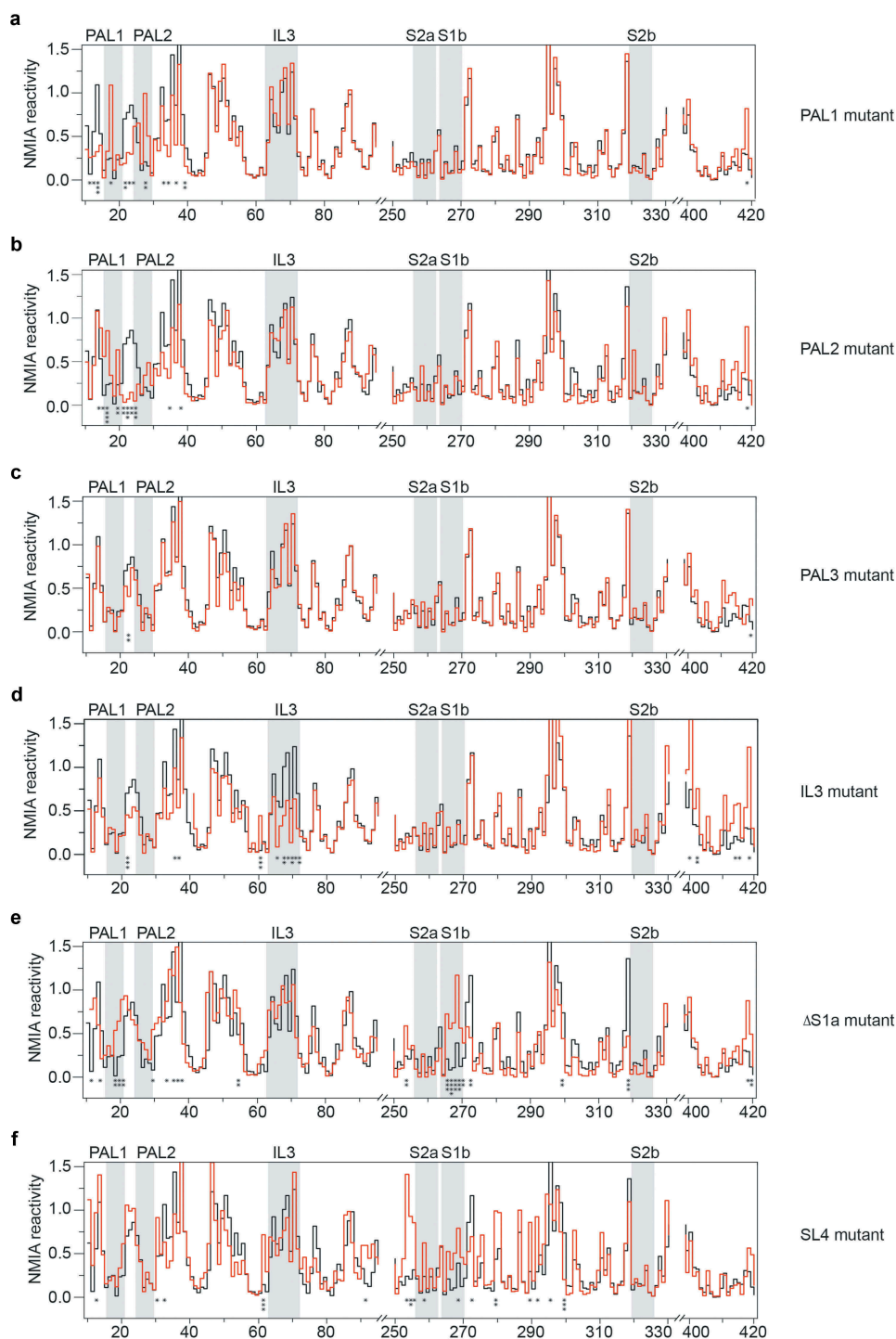


Figure 7. SHAPE analysis of the secondary structure of mini Ty1 RNA mutants in monomeric form. Step plots of NMIA reactivity of PAL1 (a), PAL2 (b), PAL3 (c), IL3 (d), Δ S1a (e) and SL4 (f) mutants. For clarity, the plots show the selected regions of mini Ty1 RNA mutants. Each mutant (red line) is plotted against the wild type mini Ty1 RNA (black line) for comparison. Nucleotides within the PAL1, PAL2, IL3 and stems (S1b, S2a, S2b) of the pseudoknot are indicated by gray stripes. Nucleotides that exhibit statistically significant differential reactivity (10% of highest SHAPE reactivity differences and a p-value < 0.05 using the Student's t-test) are marked with asterisks below the plot. Tables with SHAPE data are provided in Supplementary Data set 2.

Role of the 5'-pseudoknot in Ty1 RNA dimerization

The 5'-terminus of Ty1 RNA in VLPs is stabilized by a long-range pseudoknot structure with two helical regions: S1 and S2 [46]. This tertiary motif provides an important biological function as mutations disrupting its conformation interfere with Ty1 retrotransposition [41,47,49]. This pseudoknot is not present in the full-length Ty1 RNA *in vitro*, but is stable in the

mini Ty1 RNA monomer [47]. Our structural analysis reveals that the dimerization does not change the reactivity pattern in the sequence forming the S1 and S2 stems of the pseudoknot (Fig. 4b), and this critical motif is also present in dimeric mini Ty1 RNA. To explore the relationship between pseudoknot formation and dimerization, we analysed the structure and dimerization of a mutant with a defect in S1 stem of the

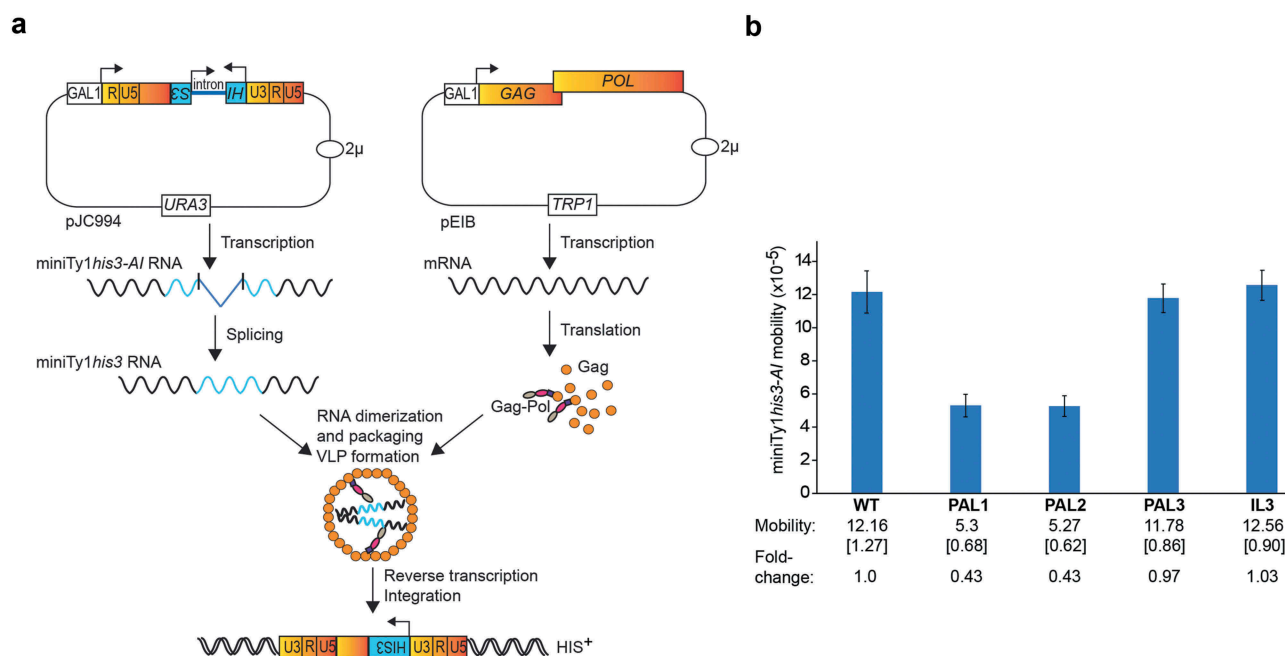


Figure 8. Involvement of dimerization sequences in Ty1 mobility. (a) Mini Ty1*his3-AI* mobility assay using two-plasmid system. Expression of the mini Ty1*his3-AI* element (pJC994) and Ty1 helper (pEIB), is driven from the *GAL1* promoter in a Ty1-less *S. paradoxus* strain. The Ty1 helper element contains *GAG* and *POL* ORFs. The absence of 3' LTR and silent substitutions in PBS preclude this element from being used as a template for reverse transcription. The mini Ty1*his3-AI* element contains the first 575 nucleotides, as well as the last 357 nucleotides of Ty1 gRNA and the *his3-AI* retrotransposition indicator gene. After splicing, miniTy1*his3* RNA is reverse transcribed and cDNA is integrated into the host genome. Cells containing spliced miniTy1/*HIS3* element give rise to His⁺ colonies. (b) Effect of PAL1, PAL2, PAL3 and IL3 mutations on mini Ty1*his3-AI* mobility. Cells expressing the Ty1 helper plasmid pEIB and mini Ty1*his3-AI* plasmids containing wild type sequences (WT) or PAL1, PAL2, PAL3 or IL3 mutations were analysed for mini Ty1*his3-AI* mobility as described in the Materials and Methods. The average frequency of mini Ty1*his3-AI* mobility is shown with standard deviations in brackets, and fold-change in mobility relative to WT.

pseudoknot (Δ S1a mutant) (Figs. 6b & 7e). Deletion of 5 nts from the S1a strand caused a statistically relevant increase in NMIA reactivity in the S1b strand, but not in the S2 stem, confirming that two helical regions of pseudoknot fold independently (Fig. 7e) [47]. Interestingly, statistically significant increased NMIA reactivity was also detected for residues A19, U20 and U21, suggesting destabilization of the upper part of SL1. This SL1 destabilization did not greatly influence dimerization as the Δ S1a mutant when compared with wild type mini Ty1 RNA (Fig. 6b).

The pseudoknot stems are linked by the sequence constituting SL4 hairpin. Although our findings do not support a role for SL4 in dimerization, a strong retrotransposition defect was previously reported for a mini Ty1 RNA mutant with identical substitutions in SL4 apical loop [49]. To further explain the role of SL4 in Ty1 retrotransposition, we performed SHAPE analysis of the SL4 RNA mutant in the monomer state (Fig. 7f). It revealed that substitutions in the SL4 apical loop disrupt the pseudoknot as increased NMIA reactivity was detected for residues in S1b and in direct proximity of S2a. Similar to the Δ S1a mutant, dimerization of the SL4 mutant was moderately weaker than that of wild type mini Ty1 RNA (Fig. 6b), additionally confirming a lack of direct correlation between the disruption of pseudoknot and Ty1 RNA dimerization. However, our data show the importance of SL4 for pseudoknot architecture, which might explain the retrotransposition defect induced by mutations in SL4 apical loop [49].

Involvement of dimerization sequences in Ty1 retrotransposition

To determine if sequences involved in dimerization of Ty1 RNA affect retrotransposition, we utilized a mini-Ty1*his3-AI* element containing the same sequences as those in our analyses *in vitro*, and that can be mobilized by Ty1 proteins supplied by a helper element (Fig. 8a) [41,49,54]. The helper Ty1 RNA expressed from pEIB lacks the PBS and 3' U3 and R sequences, which should block its involvement in reverse transcription. Unlike previous studies, Ty1 mobility assays were performed in a yeast strain that lacks Ty1 elements [73,74] (C. Bergman and D. J. Garfinkel, personal communication), thereby minimizing the effects of Ty1 products such as the p22 restriction factor that may complicate interpretation of the results [75]. Here, the PAL1 and PAL2 mutations decreased Ty1 mobility about 2-fold whereas the PAL3 and IL3 mutations did not affect mobility (Fig. 8b). The Ty1 mobility assays correspond well with results obtained *in vitro* and support our proposition of Ty1 RNA dimerization via alternative RNA-RNA contact points when PAL1-PAL2 complementarity is disrupted. Unchanged mobility of the IL3 mutant suggests that the IL3 palindrome is not required when PAL1-PAL2 can interact. Like retroviruses [76,77], alternative dimerization sites appear to be biologically active. Therefore, comparing *in vivo* replication and *in vitro* RNA dimerization confirms that *in vitro* assays reflect the biological relevance of dimerization sequences.

Discussion

Understanding how RNA dimerization occurs and its critical role in propagation has been well-studied in retroviruses [8,10], but has not been extensively analysed for other retroelements. Here, we used a combination of mutational and structural analyses to develop a model for RNA dimerization of the yeast LTR-retrotransposon Ty1 that builds on earlier work where we characterized the structure of Ty1 RNA in three distinct biological states (*in virio*, *ex virio* and *in vitro*) using SHAPE [46]. This analysis suggested PAL1, PAL2, and PAL3 might be involved in forming dimeric RNA if Ty1 followed a retroviral paradigm that involves base-pairing between palindromic sequences from two RNAs [46]. Full-length Ty1 RNA exists as monomer *in vitro* while it forms a dimer in VLPs [24,46]. In monomeric Ty1 RNA PAL1 and PAL2 form the stem of SL1, while PAL3 is a part of SL7 stem [46]. Like other classical RNA probing methods, SHAPE cannot discriminate intra- from intermolecular base-pairing, therefore, PAL sequences were unreactive in all analysed biological states and PAL's engagement in dimerization was based on changes in reactivity of nucleotides adjacent to PAL sequences. Our current model was refined by analysing the effects of mutations in PAL sequences and additional potential contact sites on Ty1 RNA structure, Gag-induced dimerization *in vitro* and Ty1 mobility. By demonstrating that mutations in PAL1 and PAL2, designed to impair dimerization, do not cause major RNA secondary structure refolding while diminishing the ability of RNA to form dimers (Figs. 6a & 7a, b), we provide evidence for PAL1 and PAL2 involvement in Ty1 RNA intermolecular interactions. Furthermore, comparative SHAPE analyses of wild type mini Ty1 RNA monomeric and dimeric states (Fig. 4) together with 3D models of PAL1 and PAL2 intra- and inter-molecular interactions (Fig. 5) have allowed us to assign reactivity alterations in the sequence linking PAL1 and PAL2 to dimer formation. These alterations are also observed in VLPs [46], which validates and extends our understanding of how Ty1 RNA dimers form. Also, the present work is the first to directly implicate PAL1 and PAL2 sequences in Ty1 RNA dimerization, and the proposed model clearly explains the inhibitory effect of mutations in PAL1 or PAL2 on Ty1 mobility in yeast (Fig. 8). The biological relevance of the proposed dimerization sequences is supported by analysis of conserved nucleotides within the Ty1 RNA 5' terminus indicating high preservation of PAL1 and PAL2 palindromes among *Saccharomyces* Ty1 elements [49].

Interestingly, we do not detect reactivity changes in the SL7 hairpin harboring PAL3 that were proposed to result from PAL3-PAL3 intermolecular base pairing in VLPs [46]. Although our analyses support the idea that PAL3 is not required for Ty1 RNA dimerization, we cannot exclude the possibility that PAL3-PAL3 interactions occur under specific conditions in VLPs. However, even if PAL3-PAL3 interactions occur in VLPs, they are not critical for Ty1 replication as the PAL3 mutations do not affect mobility (Fig. 8). Our data initially suggested that the IL3 palindrome might be a contact site that stabilizes the dimer in addition to PAL1-PAL2 duplex formation. Additional dimerization sequences have been identified in retroviral RNAs, and they can support stability of the RNA

dimer or serve as alternative contact sites when primary dimerization sites are altered [10]. For example, wild type MMTV RNA dimerizes via PAL II as a major point of contact, while deletion of PAL II induces dimerization via the palindromic sequence located within the PBS (PBS-PAL), which constitutes additional point of contact in wild type MMTV RNA dimer [78]. Repression of the HIV-2 RNA dimerization initiation site (DIS) by antisense oligonucleotides also promotes tight dimer formation by interaction between 10-nt long PAL sequence located upstream from DIS [64,79]. Ty1 RNA with a mutated PAL1 dimerizes via IL3-IL3 interaction, and IL3 mutations negatively affect dimerization *in vitro* (Fig. 6). However, despite the reactivity decrease in IL3, the IL3 palindrome is still reactive toward NMIA in the wild type mini Ty1 RNA dimer (Fig. 4) and in full-length Ty1 RNA probed *in virio* [46]. Also, IL3 mutations do not affect Ty1 mobility (Fig. 8b). Taken together, our results indicate that IL3 serves as alternative contact point in Ty1 RNA dimer when PAL1-PAL2 intermolecular interaction is disrupted. Although, we cannot exclude that under some conditions IL3-IL3 interaction additionally stabilizes the dimer of wild type Ty1 RNA, this interaction is transient and not critical for Ty1 replication.

Ty1 RNA dimers resemble retroviral dimers in that they undergo stabilization during maturation of the VLPs, resulting from proteolytic processing of Gag-p49 to Gag-p45 as well as the Gag-Pol precursor [24]. A recent study proposed that initial Ty1 RNA dimers are maintained by two intermolecular kissing loop interactions between the SL1 and SL4 hairpins [49]. These kissing loop interactions are not detected in full-length Ty1 RNA in VLPs [46] or in the mini Ty1 RNA dimer induced by Gag-p45 (Fig. 4). The RNA dimers in VLPs and induced *in vitro* by Gag-p45 represent the mature dimer stabilized by the intermolecular PAL1-PAL2 extended duplex that is mutually exclusive with the SL1-SL4 kissing loop. Retroviral transcripts can form loose dimers stabilized by kissing loop interactions even in the absence of proteins [64,66,80]. Therefore, spontaneously-formed kissing loop dimers of Ty1 RNA could be selected for packaging by the Gag-p49 precursor and then converted to mature dimer stabilized by PAL1-PAL2 interactions. However, Ty1 RNA dimers are not detected in the absence of protein (Fig. 2) and mutational disruption of the SL1-SL4 kissing loop interactions impacts mini Ty1 RNA dimerization *in vitro* only modestly, if any (Fig. 6b). Our results do not support SL1-SL4 intermolecular interactions, but they confirm the significance of SL4 in Ty1 retrotransposition. We propose that while SL4 is not involved in dimerization, it is critical for global architecture of Ty1 RNA including folding of the biologically important pseudoknot.

Our work re-examines an early study where it was proposed that unlike retroviruses, Ty1 dimeric RNA results from base pairing between two tRNA^{Met} molecules bound to PBS sequences that initiate reverse transcription [38]. This model is based on dimerization assays with a 103-residue synthetic peptide from the C-terminal region of Ty1 Gag that possesses RNA chaperone activity (Fig. 1b). Here we show that Ty1 RNA dimerization induced by mature Gag-p45 does not require tRNA^{Met} sequences (Fig. 2), and that the retrotransposon RNA dimers resemble those of retroviruses in that they are

maintained via *cis*-acting RNA sequences as described above (Fig. 4b). The lack of tRNA involvement in Ty RNA dimer formation is also supported by data showing that dimerization of Ty3 retrotransposon gRNA is not reduced in mutants that fail to package tRNA_i^{Met} and deletion of the PBS does not alter gRNA packaging into Ty3 VLPs [81–83].

Assembly of Ty1 VLPs and effective retrotransposition requires direct interactions between Gag and gRNA [26,28,30], but specific nucleotide sequences recognized by Ty1 Gag have not been defined. Recently, we identified sites within mini Ty1 RNA that are recognized *in vitro* by a protein comprising the C-terminal half of Ty1 Gag called the CTR [39]. The N-terminal part of Gag does not bind RNA, but is known to play an important role in VLP structure as residues from this region face the outside of the particle [37]. Gag-p45 binding with Ty1 RNA is significantly less salt-dependent than that of CTR [55], raising the possibility that Gag-p45 has a higher non-electrostatic RNA-binding component than the CTR, and therefore, may be more specific. We used hydroxyl radical footprinting to show that Gag-p45 recognizes the same nucleotide sequences in Ty1 RNA that we previously identified as CTR binding sites [39]. This observation indicates that although the N-terminal part of Gag-p45 influences the salt-dependent protein binding to Ty1 RNA it does not change the selective recognition of nucleotide sequences in Ty1 RNA *in vitro*.

A majority of RNA binding sites for Gag-p45 identified *in vitro* may also bind Gag or other proteins in VLPs [46]. In support of this idea, Gag-p45 binds close to RNA sequences critical for Ty1 dimerization as well as to sequences involved in cellular tRNA_i^{Met} packaging and initiation of reverse transcription (Fig. 3). Interestingly, the most prominent Gag-p45 binding site *in vitro* contains sequences forming the pseudoknot. This motif is functionally important as mutations that destabilize the S1 stem inhibit reverse transcription [47], and mutations in S2 stem decrease RNA stability [49]. Surprisingly, the pseudoknot is not required for packaging since mutations in S1 or S2 do not influence the level of Ty1 RNA in VLPs [47]. Although it is not known whether LTR-retrotransposon gRNA dimerization is a precondition for packaging or occurs in VLPs, we show that pseudoknot disruption by deletions in the S1 stem or by substitutions in SL4 have only moderate effects on Ty1 RNA dimerization, suggesting that the link between the pseudoknot and dimerization may be indirect. Indeed, pseudoknot destabilization might impact the overall Ty1 RNA structure and interactions with Gag at other steps during retrotransposition. Checkley et al. showed that Gag is required for Ty1 RNA stability, efficient nuclear export and localization into cytoplasmic foci [30]. Although, it remains to be determined if Ty1 Gag recognizes pseudoknot architecture or is involved in its folding process *in vivo*, we propose that sequences involved in pseudoknot formation constitute a critical Gag binding site and these interactions are essential for Ty1 RNA stability and trafficking in the cell. Further work will be required to determine whether Ty1 RNA contains high affinity binding sites for Gag-p45 that are involved in selective packaging and support VLPs assembly, as has been demonstrated for several retroviral Gag proteins [8,16].

Acknowledgments

We thank M. J. Curcio for plasmids pEIB and pJC994 and A. Kilizek for help preparing Fig. 5. We also thank members of K.P.W. and D.J. G. laboratories for fruitful discussion and critical reading of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was sponsored by National Science Centre Poland [2016/22/E/NZ3/00426 to K.P.W. and 2016/23/B/ST6/03931] and the National Institutes of Health [GM095622 and GM124216, to D.J.G.]. This publication was also supported by the Polish Ministry of Science and Higher Education, under the KNOW programme, and the grants for young researches and Ph.D. students [to J.G.]. Funding for open access charge was provided by the National Science Centre Poland [2016/22/E/NZ3/00426 to K.P.W.].

ORCID

Julita Gumna  <http://orcid.org/0000-0002-1045-2823>
 Katarzyna J. Purzycka  <http://orcid.org/0000-0002-0271-9161>
 David J. Garfinkel  <http://orcid.org/0000-0001-6234-2426>
 Katarzyna Pachulska-Wieczorek  <http://orcid.org/0000-0002-5723-6204>

References

- [1] Huang CR, Burns KH, Boeke JD. Active transposition in genomes. *Annu Rev Genet.* 2012;46:651–675.
- [2] Contreras-Galindo R, Kaplan MH, Dube D, et al. Human endogenous retrovirus type K (HERV-K) particles package and transmit HERV-K-related sequences. *J Virol.* 2015;89:7187–7201.
- [3] Mager DL, Stoye JP. Mammalian endogenous retroviruses. *Microbiol Spectr.* 2015;3:1–20.
- [4] Ashley J, Cordy B, Lucia D, et al. Retrovirus-like Gag protein Arc1 binds RNA and traffics across synaptic boutons. *Cell.* 2018;172:262–274 e211.
- [5] Pastuzyn ED, Day CE, Kearns RB, et al. The neuronal gene arc encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. *Cell.* 2018;172:275–288 e218.
- [6] Johnson SF, Telesnitsky A. Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog.* 2010;6:e1001007.
- [7] Kuzembayeva M, Dilley K, Sardo L, et al. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology.* 2014;454–455:362–370.
- [8] Comas-Garcia M, Davis SR, Rein A. On the selective packaging of genomic RNA by HIV-1. *Viruses-Basel.* 2016;8:246.
- [9] D'Souza V, Summers MF. How retroviruses select their genomes. *Nature Rev Microbiol.* 2005;3:643–655.
- [10] Dubois N, Marquet R, Paillart JC, et al. Retroviral RNA dimerization: from structure to functions. *Front Microbiol.* 2018;9.
- [11] Jayaraman D, Kenyon JC. New windows into retroviral RNA structures. *Retrovirology.* 2018;15:11.
- [12] Keane SC, Summers MF. NMR studies of the structure and function of the HIV-1 5-leader. *Viruses-Basel.* 2016;8.
- [13] Lu K, Heng X, Summers MF. Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol.* 2011;410:609–633.
- [14] Abd El-Wahab EW, Smyth RP, Mailler E, et al. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat Commun.* 2014;5:4304.

- [15] Mailler E, Bernacchi S, Marquet R, et al. The life-cycle of the HIV-1 Gag-RNA complex. *Viruses-Basel*. 2016;8.
- [16] Olson ED, Musier-Forsyth K. Retroviral Gag protein-RNA interactions: implications for specific genomic RNA packaging and virion assembly. *Semin Cell Dev Biol*. 2018;8:129–139
- [17] Guo F, Saadatmand J, Niu MJ, et al. Roles of Gag and NCp7 in facilitating tRNA(3)(Lys) annealing to viral RNA in human immunodeficiency virus type 1. *J Virol*. 2009;83:8099–8107.
- [18] Pachulska-Wieczorek K, Blaszczyk L, Biesiada M, et al. The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag. *Retrovirology*. 2016;13:18.
- [19] Kleiman L, Jones CP, Musier-Forsyth K. Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett*. 2010;584:359–365.
- [20] Peterson-Burch BD, Voytas DF. Genes of the pseudoviridae (Ty1/copia retrotransposons). *Mol Biol Evol*. 2002;19:1832–1845.
- [21] Curcio MJ, Lutz S, Lesage P. The Ty1 LTR-retrotransposon of budding yeast. *Microbiol Spectr*. 2015;3:1–35.
- [22] Boeke JD, Garfinkel DJ, Styles CA, et al. Ty elements transpose through an RNA intermediate. *Cell*. 1985;40:491–500.
- [23] Devine SE, Boeke JD. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev*. 1996;10:620–633.
- [24] Feng YX, Moore SP, Garfinkel DJ, et al. The genomic RNA in Ty1 virus-like particles is dimeric. *J Virol*. 2000;74:10819–10821.
- [25] Dutko JA, Kenny AE, Gamache ER, et al. 5' to 3' mRNA decay factors colocalize with Ty1 gag and human APOBEC3G and promote Ty1 retrotransposition. *J Virol*. 2010;84:5052–5066.
- [26] Malagon F, Jensen TH. T-body formation precedes virus-like particle maturation in *S. cerevisiae*. *RNA Biol*. 2011;8:184–189.
- [27] Beliakova-Bethell N, Beckham C, Giddings TH Jr., et al. Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. *RNA*. 2006;12:94–101.
- [28] Pachulska-Wieczorek K, Le Grice SF, Purzycka KJ. Determinants of genomic RNA encapsidation in the *Saccharomyces cerevisiae* long terminal repeat retrotransposons Ty1 and Ty3. *Viruses*. 2016;8:193.
- [29] Sandmeyer S, Patterson K, Bilanchone V. Ty3, a position-specific retrotransposon in budding yeast. *Microbiol Spectr*. 2015;3:1–29
- [30] Checkley MA, Mitchell JA, Eizenstat LD, et al. Ty1 gag enhances the stability and nuclear export of Ty1 mRNA. *Traffic*. 2013;14:57–69.
- [31] Garfinkel DJ, Hedge AM, Youngren SD, et al. Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1. *J Virol*. 1991;65:4573–4581.
- [32] Garfinkel DJ, Mastrangelo MF, Sanders NJ, et al. Transposon tagging using Ty elements in yeast. *Genetics*. 1988;120:95–108.
- [33] Kingsman AJ, Mellor J, Adams S, et al. The genetic organization of the yeast Ty element. *J Cell Sci Suppl*. 1987;7:155–167.
- [34] Muller F, Bruhl KH, Freidel K, et al. Processing of TY1 proteins and formation of Ty1 virus-like particles in *Saccharomyces cerevisiae*. *Mol Gen Genet*. 1987;207:421–429.
- [35] Tucker JM, Larango ME, Wachsmuth LP, et al. The Ty1 retrotransposon restriction factor p22 targets Gag. *PLoS Genet*. 2015;11:e1005571.
- [36] Martin-Rendon E, Hurd DW, Marfany G, et al. Identification of proteolytic cleavage sites within the gag-analogue protein of Ty1 virus-like particles. *Mol Microbiol*. 1996;22:1035–1043.
- [37] Brookman JL, Stott AJ, Cheeseman PJ, et al. Analysis of Tya protein regions necessary for formation of the Ty1 virus-like particle structure. *Virology*. 1995;212:69–76.
- [38] Cristofari G, Ficheux D, Darlix JL. The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. *J Biol Chem*. 2000;275:19210–19217.
- [39] Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, et al. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic Acids Res*. 2015;43:7414–7431.
- [40] Xu H, Boeke JD. Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. *Mol Cell Biol*. 1990;10:2695–2702.
- [41] Bolton EC, Coombes C, Eby Y, et al. Identification and characterization of critical cis-acting sequences within the yeast Ty1 retrotransposon. *RNA*. 2005;11:308–322.
- [42] Boeke JD, Eichinger D, Castrillon D, et al. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. *Mol Cell Biol*. 1988;8:1432–1442.
- [43] Chapman KB, Bystrom AS, Boeke JD. Initiator methionine tRNA is essential for Ty1 transposition in yeast. *Proc Natl Acad Sci U S A*. 1992;89:3236–3240.
- [44] Friant S, Heyman T, Bystrom AS, et al. Interactions between Ty1 retrotransposon RNA and the T and D regions of the tRNA(iMet) primer are required for initiation of reverse transcription in vivo. *Mol Cell Biol*. 1998;18:799–806.
- [45] Cristofari G, Bampi C, Wilhelm M, et al. A 5'-3' long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *Embo J*. 2002;21:4368–4379.
- [46] Purzycka KJ, Legiewicz M, Matsuda E, et al. Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res*. 2013;41:463–473.
- [47] Huang Q, Purzycka KJ, Lusvardi S, et al. Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA*. 2013;19:320–332.
- [48] Merino EJ, Wilkinson KA, Coughlan JL, et al. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*. 2005;127:4223–4231.
- [49] Gamache ER, Doh JH, Ritz J, et al. Structure-function model for kissing loop interactions that initiate dimerization of Ty1 RNA. *Viruses*. 2017;9:93.
- [50] Guthrie CF, Fink GR. Guide to yeast genetics and molecular biology. *Meth Enzymol*. 1991;194:1–863.
- [51] Scholes DT, Banerjee M, Bowen B, et al. Multiple regulators of Ty1 transposition in *Saccharomyces cerevisiae* have conserved roles in genome maintenance. *Genetics*. 2001;159:1449–1465.
- [52] Curcio MJ, Garfinkel DJ. Single-step selection for Ty1 element retrotransposition. *Proc Natl Acad Sci U S A*. 1991;88:936–940.
- [53] Melamed C, Nevo Y, Kupiec M. Involvement of cDNA in homologous recombination between Ty elements in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1992;12:1613–1620.
- [54] Sharon G, Burkett TJ, Garfinkel DJ. Efficient homologous recombination of Ty1 element cDNA when integration is blocked. *Mol Cell Biol*. 1994;14:6540–6551.
- [55] Blaszczyk L, Biesiada M, Saha A, et al. Structure of Ty1 internally initiated RNA influences restriction factor expression. *Viruses*. 2017;9:74.
- [56] Lusvardi S, Sztuba-Solinska J, Purzycka KJ, et al. The HIV-2 Rev-response element: determining secondary structure and defining folding intermediates. *Nucleic Acids Res*. 2013;41:6637–6649.
- [57] Somarowthu S, Legiewicz M, Keating KS, et al. Visualizing the ai5gamma group IIB intron. *Nucleic Acids Res*. 2014;42:1947–1958.
- [58] Vasa SM, Guex N, Wilkinson KA, et al. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *Rna*. 2008;14:1979–1990.
- [59] Low JT, Weeks KM. SHAPE-directed RNA secondary structure prediction. *Methods*. 2010;52:150–158.
- [60] Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *Bmc Bioinformatics*. 2010;11.
- [61] Antczak M, Popenda M, Zok T, et al. New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim Pol*. 2016;63:737–744.
- [62] Popenda M, Szachniuk M, Antczak M, et al. Automated 3D structure composition for large RNAs. *Nucleic Acids Res*. 2012;40:e112.

- [63] Purzycka KJ, Popena M, Szachniuk M, et al. Automated 3D RNA structure prediction using the RNAComposer method for riboswitches. *Methods Enzymol.* **2015**;553:3–34.
- [64] Purzycka KJ, Pachulska-Wieczorek K, Adamiak RW. The in vitro loose dimer structure and rearrangements of the HIV-2 leader RNA. *Nucleic Acids Res.* **2011**;39:7234–7248.
- [65] Berkhout B, Ooms M, Beerens N, et al. In vitro evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes. *J Biol Chem.* **2002**;277:19967–19975.
- [66] Tran T, Liu Y, Marchant J, et al. Conserved determinants of lentiviral genome dimerization. *Retrovirology.* **2015**;12:83.
- [67] Laughrea M, Jette L. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248–271 are dispensable for dimer formation. *Biochemistry.* **1996**;35:1589–1598.
- [68] Nilsen TW. Mapping RNA-protein interactions using hydroxyl-radical footprinting. *Cold Spring Harb Protoc.* **2014**;2014:1333–1336.
- [69] Kutluay SB, Zang T, Blanco-Melo D, et al. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell.* **2014**;159:1096–1109.
- [70] Wilkinson KA, Gorelick RJ, Vasa SM, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **2008**;6:e96.
- [71] Rypniewski W, Banaszak K, Kulinski T, et al. Watson-Crick-like pairs in CCUG repeats: evidence for tautomeric shifts or protonation. *RNA.* **2016**;22:22–31.
- [72] Blaszczyk L, Rypniewski W, Kiliszek A. Structures of RNA repeats associated with neurological diseases. *Wiley Interdiscip Rev RNA.* **2017**;8:e1412.
- [73] Moore SP, Liti G, Stefanisko KM, et al. Analysis of a Ty1-less variant of *Saccharomyces paradoxus*: the gain and loss of Ty1 elements. *Yeast.* **2004**;21:649–660.
- [74] Garfinkel DJ, Nyswaner K, Wang J, et al. Post-transcriptional cosuppression of Ty1 retrotransposition. *Genetics.* **2003**;165:83–99.
- [75] Saha A, Mitchell JA, Nishida Y, et al. A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. *J Virol.* **2015**;89:3922–3938.
- [76] Laughrea M, Shen N, Jette L, et al. Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry.* **1999**;38:226–234.
- [77] Paillart JC, Berthoux L, Ottmann M, et al. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol.* **1996**;70:8348–8354.
- [78] Aktar SJ, Vivet-Boudou V, Ali LM, et al. Structural basis of genomic RNA (gRNA) dimerization and packaging determinants of mouse mammary tumor virus (MMTV). *Retrovirology.* **2014**;11:96.
- [79] Lanchy JM, Ivanovitch JD, Lodmell JS. A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA.* **2003**;9:1007–1018.
- [80] Huthoff H, Berkhout B. Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry.* **2002**;41:10439–10445.
- [81] Clemens K, Bilanchone V, Beliakova-Bethell N, et al. Sequence requirements for localization and packaging of Ty3 retroelement RNA. *Virus Res.* **2013**;171:319–331.
- [82] Nymark-McMahon MH, Beliakova-Bethell NS, Darlix JL, et al. Ty3 integrase is required for initiation of reverse transcription. *J Virol.* **2002**;76:2804–2816.
- [83] Larsen LS, Zhang M, Beliakova-Bethell N, et al. Ty3 capsid mutations reveal early and late functions of the amino-terminal domain. *J Virol.* **2007**;81:6957–6972.

SUPPLEMENTARY DATA 1

Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA

Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel and Katarzyna Pachulska-Wieczorek

Table S1. Primers used for pJC994 site-directed mutagenesis.

PRIMER	SEQUENCE
PAL1mutant_PF	CTCTACCGCCTCGAGGAGAACTTCCGCCCGCTTCTGTATACCTAATATTATAGCC
PAL1mutant_PR	GGCTATAATATTAGGTATACAGAAGCGGGCGGAAGTTCTCCTCGAGGCGGTAGAG
PAL2mutant_PF	CTACCGCCTCGAGGAGAACTTCTAGTATATTCTCGCCCGCTAATATTATAGCCTTTATCA ACAATGGAATCC
PAL2mutant_PR	GGATTCCATTGTTGATAAAGGCTATAATATTAGCGGGCGAGAATATACTAGAAGTTCTCC TCGAGGCGGTAG
PAL3mutant_PF	CAAATGTCGCCTATGTACTTTCCAAAGAGGCCACAATCACAGTTTCCGC
PAL3mutant_PR	GCGGAAACTGTGATTGTGGCCTCTTTGGAAAGTACATAGGCGCATTG
IL3mutant_PF	GCCTTTATCAACAATGGAATCCCACCAACCATCTCAACATTCACCAATTCTC
IL3mutant_PR	GAGAATTGGGTGAATGTTGAGATGGTTGGTGGGATTCCATTGTTGATAAAGGC

Table S2. Primers used for DNA templates construction.

PRIMER	SEQUENCE
F-miniRNA	GATTTAGGTGACACTATAGAGGAGAACTTCTAGT
R-miniRNA	ACATTGGTGGTGGTCTGAC
ΔS1amutant_PF	GATTTAGGTGACACTATAGAACTTCTAGTATATT
PAL1mutant_PF	GATTTAGGTGACACTATAGAGGAGAACTTCCGCCCGCTTCTGTATACCTAA
PAL2mutant_PF	GATTTAGGTGACACTATAGAGGAGAACTTCTAGTATATTCTCACACGATAATATTATAGC
PAL3mutant_PF	TTCCAAAGAGGCCACAAT
PAL3mutant_PR	ACTGTGATTGTGGCCTCTTTGGAAAGTACATAG
SL4mutant_PF	GGGTACGGCCCTTAGAGAGGAGGTGGTACTGAAGC
SL4mutant_PR	CCUCUCUAAGGGCCGUACCCACAGCAGUG
IL3mutant_PF	GAATCCCACCAACCATCTCAACATTCAC
IL3mutant_PR	TGAATGTTGAGATGGTTGGTGGGATTCC

Figure S1. Dimerization of full length Ty1 gRNA *in vitro*. Lane M: mass marker; lane 1: RNA denatured in formamide, lane 2: RNA in water; lane 3: RNA dimerization in the absence of Gag-p45; line 4: RNA dimerization in the presence of Gag-p45.

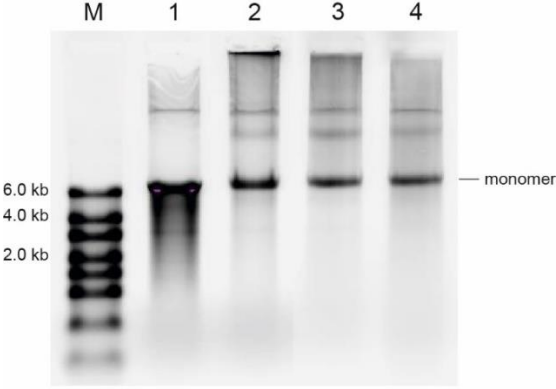


Figure S2. Gag-dependent dimerization of wild type mini Ty1 RNA detected at 0°C in the presence of 0; 0.375; 0.75; 1.5; 3; and 4.5 μM Gag-p45. Lane denoted C represents control sample that lacks protein.

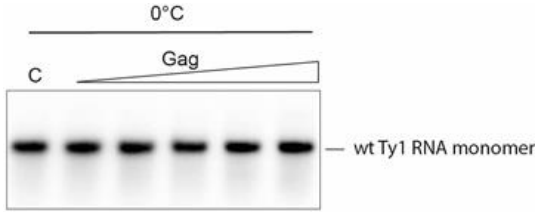


Figure S3. SHAPE analysis of mini Ty1 RNA in the monomeric state. +1-433 region of mini Ty1 RNA in the monomer form. Nucleotide residues accessibilities to NMIA are color-coded according to scale.

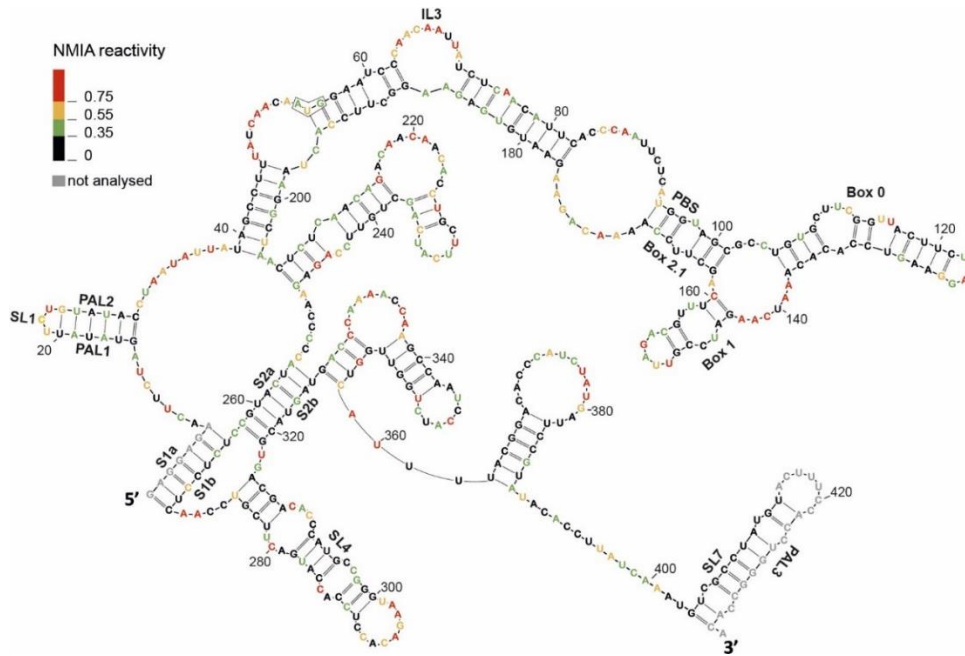


Figure S4. SHAPE analysis of mini Ty1 RNA in monomeric and dimeric states in the absence of tRNA^{Met}. For clarity, the plots show the +10 – 85 region of mini Ty1 RNA. At the top is the step plot of NMIA reactivity for monomeric (black) and dimeric (red) RNA. At the bottom is the difference plot calculated by subtracting the monomer intensities from those of the dimer. Negative values indicate nucleotides that are less flexible in the dimer. Columns corresponding to nucleotides that exhibit statistically significant differential reactivity (10% of highest SHAPE reactivity differences and a p-value <0.05 using the Student's t-test) are colored in blue. Asterisks below the plot correspond to statistical significance.

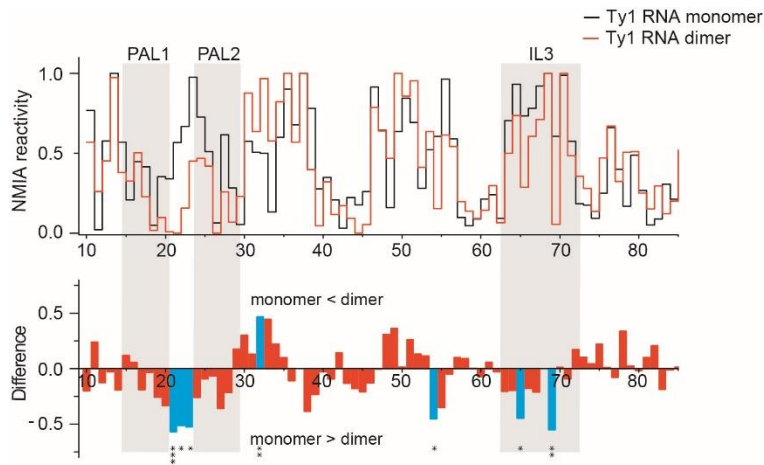


Figure S5. Representative Gag-dependent dimerization assays of wild type and mutant mini Ty1 RNA. 0; 0.375; 0.75; 1.5; 3; 4.5 μM Gag-p45 concentrations were used. Experiments were performed in the presence of $\text{tRNA}_i^{\text{Met}}$. Lanes denoted C represent control samples that lack protein.

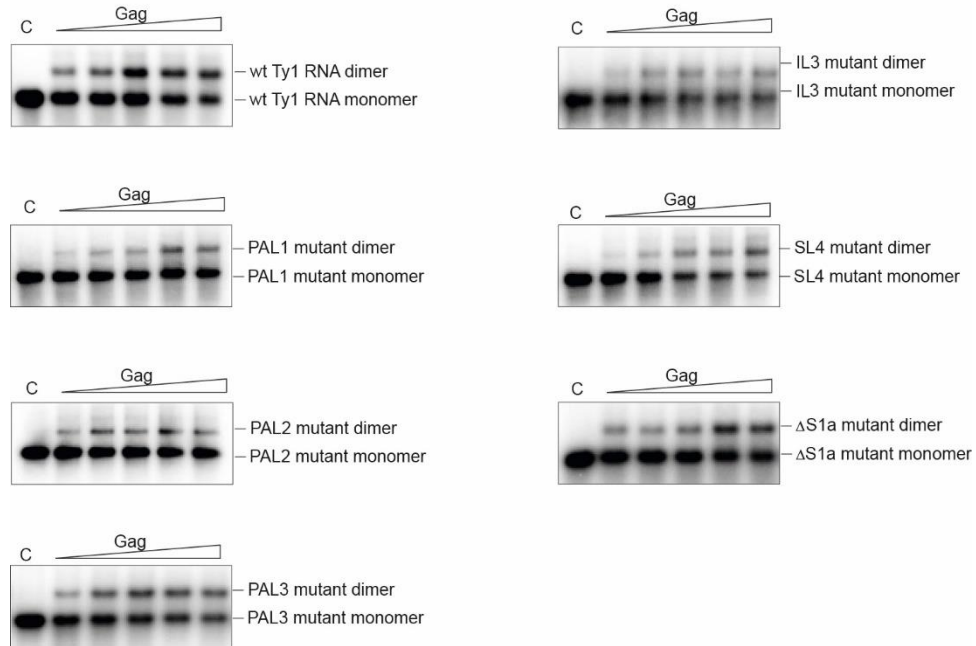
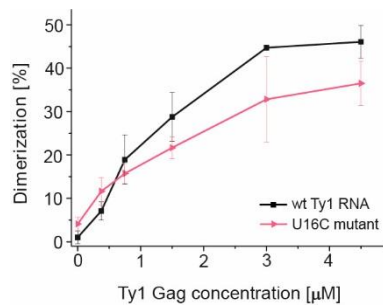


Figure S6. Dimerization of wild type mini Ty1 and U16C RNA mutant. The graph represents averaged results from three independent experiments.



Załącznik nr 2 do rozprawy doktorskiej „Identyfikacja i charakterystyka elementów strukturalnych genomowego RNA retrotranspozonu Ty1 oraz jego oddziaływań z białkiem Gag kluczowych dla dimeryzacji i pakowania genomu do cząstek wirusopodobnych”

Dodatek *Supplementary data 2* do pracy naukowej J. Gumna, K. J. Purzycka, H. W. Ahn, D. J. Garfinkel, K. Pachulska-Wieczorek; *Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA*; RNA Biology 2019, zawierający średnie wartości reaktywności dla poszczególnych nukleotydów z eksperymentów próbkowania strukturalnego RNA oraz wyniki analiz statystycznych znajduje się pod linkiem:

<https://www.tandfonline.com/doi/suppl/10.1080/15476286.2019.1657370?scroll=top>



Article

RNA Binding Properties of the Ty1 LTR-Retrotransposon Gag Protein

Julita Gumna ¹, Angelika Andrzejewska-Romanowska ¹, David J. Garfinkel ²
and Katarzyna Pachulska-Wieczorek ^{1,*}

¹ Department of Structure and Function of Retrotransposons, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland; jgumna@ibch.poznan.pl (J.G.); aandrzejewska@ibch.poznan.pl (A.A.-R.)

² Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA; djgarf@uga.edu

* Correspondence: kasiapw@ibch.poznan.pl; Tel.: +48-61-852-85-03; Fax: +48-61-852-05-32

Abstract: A universal feature of retroelement propagation is the formation of distinct nucleoprotein complexes mediated by the Gag capsid protein. The Ty1 retrotransposon Gag protein from *Saccharomyces cerevisiae* lacks sequence homology with retroviral Gag, but is functionally related. In addition to capsid assembly functions, Ty1 Gag promotes Ty1 RNA dimerization and cyclization and initiation of reverse transcription. Direct interactions between Gag and retrotransposon genomic RNA (gRNA) are needed for Ty1 replication, and mutations in the RNA-binding domain disrupt nucleation of retrosomes and assembly of functional virus-like particles (VLPs). Unlike retroviral Gag, the specificity of Ty1 Gag-RNA interactions remain poorly understood. Here we use microscale thermophoresis (MST) and electrophoretic mobility shift assays (EMSA) to analyze interactions of immature and mature Ty1 Gag with RNAs. The salt-dependent experiments showed that Ty1 Gag binds with high and similar affinity to different RNAs. However, we observed a preferential interaction between Ty1 Gag and Ty1 RNA containing a packaging signal (Psi) in RNA competition analyses. We also uncover a relationship between Ty1 RNA structure and Gag binding involving the pseudoknot present on Ty1 gRNA. In all likelihood, the differences in Gag binding affinity detected in vitro only partially explain selective Ty1 RNA packaging into VLPs in vivo.

Keywords: Ty1 retrotransposon; Gag; protein-RNA interactions; binding affinity; RNA packaging; MST



Citation: Gumna, J.; Andrzejewska-Romanowska, A.; Garfinkel, D.J.; Pachulska-Wieczorek, K. RNA Binding Properties of the Ty1 LTR-Retrotransposon Gag Protein. *Int. J. Mol. Sci.* **2021**, *22*, 9103. <https://doi.org/10.3390/ijms22169103>

Academic Editors: Tatyana Karabencheva-Christova and Christo Z. Christov

Received: 3 August 2021

Accepted: 22 August 2021

Published: 23 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Ty1/Copia (*Pseudoviridae*) family contains well-studied LTR retrotransposons present in many eukaryotic genomes [1]. Although yeast Ty1 elements replicate intracellularly, they package genomic RNA (gRNA) into virus-like particles (VLPs) that are comparable to retroviral virions. Both types of particles form by the assembly of Gag precursors to protect retroelement gRNA from nucleases and concentrate factors necessary for reverse transcription of gRNA into double-stranded DNA that can integrate into the host genome [2–5]. Retroviral virion assembly has been extensively studied using a variety of approaches that highlight the critical role of the Gag nucleocapsid (NC) domain, containing one or two zinc finger (ZF) motifs, in the selection, dimerization, and packaging of gRNA [6–9].

Ty1 Gag appears to be different based on its functional organization and low sequence homology when compared with retroviral Gag, but carries out similar processes [10]. The Ty1 Gag precursor is a 441-amino acid protein (Gag-p49) which, unlike a retroviral Gag polyprotein, does not contain recognizable matrix (MA), capsid (CA), and nucleocapsid (NC) domains. During VLP maturation, Gag-p49 undergoes one C-terminal cleavage by Ty1 protease to form mature Gag (Gag-p45, 401 aa) and a small peptide (p4, 40 aa) with

unknown function [11,12]. The N-terminal and central regions of Ty1 Gag are essential for protein-protein interactions to form the VLP shell, and the C-terminal region is required for RNA binding [13,14]. The RNA-binding region of Ty1 Gag is rich in basic amino acids and does not contain a recognizable ZF motif [14]. Ty1 Gag plays an important role in the nuclear export and stability of the Ty1 gRNA and its trafficking to specific cytoplasmic foci termed retrosomes where Ty1 RNA and proteins accumulate prior to assembly into VLPs [15–18]. Mutations in the C-terminal RNA binding domain of Gag disrupt retrosome nucleation and VLP assembly [15,17]. Like retroviral Gag, Ty1 Gag also displays nucleic acid chaperone (NAC) activity. In vitro assays showed that Ty1 Gag or its shorter forms containing C-terminal domain promote annealing of tRNA_i^{Met}, initiation of reverse transcription, Ty1 gRNA dimerization, and cyclization [14,19–21].

For retroviruses, it is generally accepted that gRNA packaging is mediated by specific binding between Gag and the packaging signal (Psi) located near the 5'-terminus of viral gRNA [22,23]. In vitro, at physiological salt conditions, HIV-1 Gag binds HIV-1 Psi RNA and non-Psi RNAs with similar affinities but the specificity for Psi RNA is evident when binding affinity is measured in high ionic strength buffers or in the presence of an excess of competing heterologous RNA (e.g., tRNA) [24,25]. The mechanism by which Ty1 Gag identifies and selects Ty1 gRNA for packaging into VLPs remains unclear. Psi sequences for the Ty1 element have not been precisely defined, but functional analyses suggest that *cis*-acting sequences critical for gRNA packaging encompass nucleotides +230–580 [26–28]. Biochemical and genetic results suggest *cis*-acting sequences within the first 580 nucleotides are critical for Ty1 retrotransposition. This region includes palindromes (PAL1 and PAL2) involved in Ty1 RNA dimerization, PBS, Box0, and Box1 that bind primer tRNA_i^{Met}, and CYC5 sequence required for RNA genome cyclization (Figure 1) [20,21,26,27,29,30]. Additionally, the Ty1 gRNA 5'-end is stabilized by a functionally important pseudoknot structure that may also be a preferred binding site for mature Ty1 Gag [20,31].

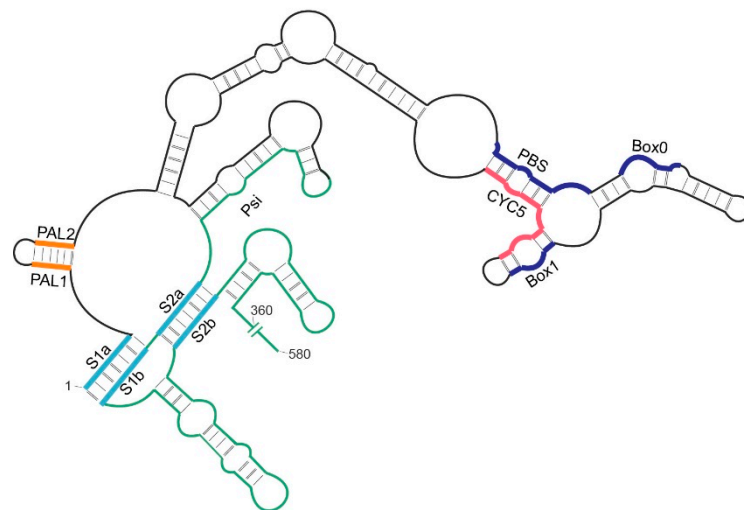


Figure 1. Schematic representation of the 5'-end of Ty1 gRNA containing *cis*-acting sequences: palindromes (PAL1 and PAL2; orange) involved in RNA dimerization, primer binding site (PBS) and short boxes (Box0, Box1; dark blue) that anneal cellular tRNA_i^{Met} and 5' cyclization site (CYC5; pink). The region containing the proposed packaging signal (+230–580) is shown in green. The strands forming the stems (S1 and S2) of the pseudoknot motif are shown in blue.

In this work, we use microscale thermophoresis (MST) and electrophoretic mobility shift assays (EMSA) to investigate RNA-binding properties of recombinant Ty1 Gag-p49 and -p45. We assess Gag binding to relevant segments of Ty1 gRNA with, or without, a packaging signal and to heterologous RNA. Our results suggest that Ty1 Gag binds diverse RNAs with high affinity but low specificity. In contrast to retroviral Gag, the affinity of Ty1 Gag for various RNAs is not significantly different even at high ionic strength. However,

competition experiments reveal preferential binding of Gag to Ty1 RNA containing Psi, and correct folding of the 5' pseudoknot structure facilitates selective recognition of Ty1 RNA by Gag.

2. Results

2.1. RNA Binding Affinity and Specificity of Immature and Mature Ty1 Gag

To characterize RNA binding properties of Gag-p49 and Gag-p45, we used microscale thermophoresis, which provides a fast and sensitive way to dissect intermolecular interactions through changes in fluorescence [32]. We analyzed the binding properties of recombinant Gag-p49 and Gag-p45 to three RNA molecules of similar length: mini Ty1 RNA (mTy1 RNA), non-Psi Ty1 RNA, and *S. cerevisiae* 18S rRNA. mTy1 RNA encompasses the first 576 nucleotides of Ty1 gRNA and contains *cis*-acting sequences crucial for the retrotransposition, including the packaging region (Figure 1) [20,21,26,27,30,31]. The non-Psi Ty1 RNA encompasses a region of the Ty1 coding sequence (+1000–1616). This RNA corresponds to an internally-initiated transcript (Ty1i), which encodes the Gag-like p22 restriction factor but is not detected in VLPs [33,34]. A yeast 18S rRNA fragment (+1–576) was used as an additional control for Ty1 Gag binding specificity. Experiments were performed in a buffer similar to that used in binding reactions for HIV-1 Gag [25,35,36]. To determine Ty1 Gag binding specificity, Gag titration experiments were performed at different NaCl concentrations (150–500 mM). At high NaCl concentrations, non-specific (electrostatic) interactions are increasingly masked by the salt. Specific binding to the RNA should be more salt-resistant due to the higher contribution of hydrogen bonds or aromatic residue stacking from protein-RNA interactions [37,38]. Our results were fitted to the Hill equation model to estimate the average K_D values for protein-RNA complexes and to provide further information on the binding mechanism.

At physiological ionic strength (150 mM NaCl), Gag-p49 binds both mTy1 RNA and non-Psi Ty1 RNA with high affinity, and only a small difference in K_D values (K_D 176.8 nM and 231.9 nM, respectively) was detected by MST (Figure 2A,B; Figure S1; Table 1). The baseline-corrected normalized fluorescence (ΔF_{norm}) increased with Gag-p49 concentration, reaching a plateau at ~560 nM for mTy1 RNA (1:37 RNA: protein ratio) and ~790 nM for non-Psi Ty1 RNA (1:53 RNA: protein ratio). Salt concentrations of 300 mM and 400 mM caused a ~2.4-fold and ~9-fold rise in the K_D value of Gag-p49 binding to mTy1 RNA, respectively. For non-Psi Ty1 RNA, we detected more substantial increases in K_D value, with a ~3.8-fold increase at 300 mM NaCl and a 15-fold increase at 400 mM NaCl. At 500 mM NaCl, binding of Gag-p49 to both RNAs was significantly reduced, as shown by a 50- and 67-fold increase in K_D for mTy1 RNA and non-Psi Ty1 RNA, respectively. Also, a plateau was not reached at 500 mM salt, thus, a measurement error is greater than for other salt concentrations. For heterologous 18S rRNA, higher ionic strength of the binding buffer (300 mM to 500 mM NaCl) caused an increase in K_D comparable to that observed for interactions with mTy1 RNA (Figure 2C; Table 1). Therefore, Gag-p49 binding to RNA containing the Ty1 packaging signal is not more salt-resistant when compared with the heterologous RNA. Nevertheless, at each salt concentration, Ty1 Gag-p49 displays a slightly higher affinity to mTy1 RNA than to non-Psi Ty1 RNA or 18S rRNA.

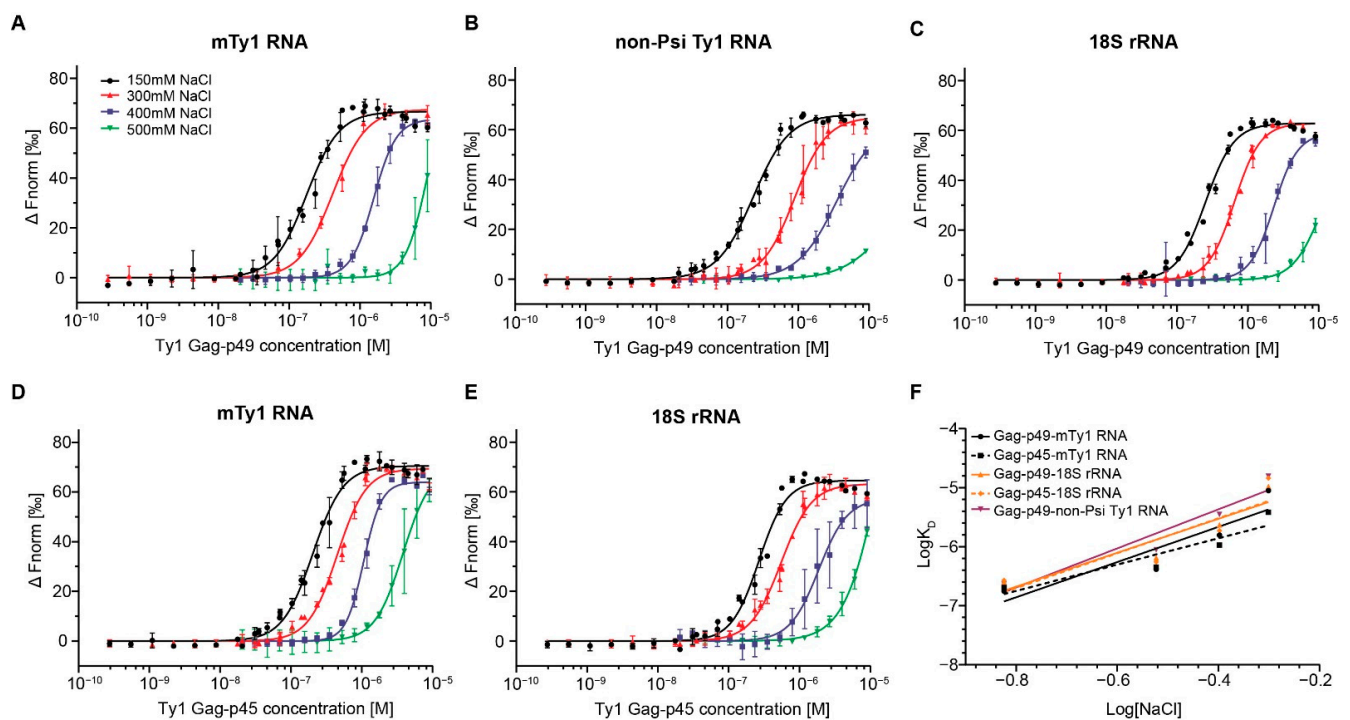


Figure 2. MST analysis of Ty1 Gag-RNA interactions at increasing ionic strength. Dose–response binding curves of Ty1 Gag-p49 to mTy1 RNA (A), non-Psi Ty1 RNA (B), 18S rRNA (C) and of Ty1 Gag-p45 to mTy1 RNA (D) and 18S rRNA (E) at 150, 300, 400 and 500 mM NaCl. Lines represent fits of the data points using the Hill equation (EC50). (F) The log-log plot of the binding affinity as a function of NaCl concentration for Ty1 Gag-p49 (solid lines) and Ty1 Gag-p45 (dashed lines).

Table 1. Impact of ionic strength on Ty1 Gag-RNA interactions. Binding parameters were based on binding curves from the salt-dependent experiments that were fitted to the Hill equation. Mean \pm SD of K_D from at least three independent MST experiments. n.d.: not determined.

	K_D [nM]	n_H	K_D [nM]	n_H	K_D [nM]	n_H	K_D [nM]	n_H	$K_{D(1M)}$ [M]	Z_{eff}
	150 mM NaCl		300 mM NaCl		400 mM NaCl		500 mM NaCl			
mTy1 RNA-Gag-p49	176.8 \pm 11.4	1.9	414.7 \pm 21.7	1.8	1566.2 \pm 36	2.7	8935 \pm 3544.5	2.8	$(3.4 \pm 5.2) \times 10^{-5}$	3.0 \pm 0.9
18S rRNA-Gag-p49	253.9 \pm 11.9	2.2	647.3 \pm 17.2	2.4	2337.4 \pm 127.7	2.6	10466 \pm 4083.1	2.2	$(4.2 \pm 4.4) \times 10^{-5}$	2.9 \pm 0.8
non-Psi Ty1 RNA-Gag-p49	231.9 \pm 10.5	1.7	886.4 \pm 37.7	1.9	3520 \pm 258.4	1.6	15635 \pm 19989	1.4	$(9 \pm 3.9) \times 10^{-5}$	3.3 \pm 0.7
mTy1 RNA-Gag-p45	204.3 \pm 10.8	2	451.1 \pm 21.9	2.1	1064.7 \pm 32.6	3.4	3823.2 \pm 148.8	2.2	$(1.1 \pm 3.4) \times 10^{-5}$	2.2 \pm 0.6
18S rRNA-Gag-p45	267.8 \pm 15	2.2	552.4 \pm 23.4	2	1871.3 \pm 139.4	2.1	14525 \pm 6330.2	1.7	$(4.6 \pm 6.1) \times 10^{-5}$	3.0 \pm 1.1
Δ S1a RNA-Gag-p49	214.7 \pm 9.4	2.1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

To compare RNA binding features of immature and mature forms of Ty1 Gag, we performed salt-dependent experiments using Gag-p45 and mTy1 RNA or 18S rRNA (Figure 2D,E; Table 1). At physiological salt concentration, the difference between K_D values obtained for the Gag-p45 complexes with mTy1 RNA and 18S rRNA was also small (204.3 nM and 267.8 nM, respectively), and a plateau in binding was reached at the same protein concentrations as for immature Ty1 Gag. At 400 mM NaCl, we observed a \sim 5.2-fold increase of K_D value for Gag-p45 complexes formed with mTy1 and a \sim 7-fold increase for complexes with 18S rRNA. Furthermore, at 500 mM NaCl, there was a smaller increase in the K_D value for Gag-p45 interactions with mTy1 RNA than with 18S rRNA (19- and

54-fold, respectively). The results suggest that Gag-p45 may display greater discrimination between Ty1 and other RNAs.

To estimate the relative contribution of electrostatic and non-electrostatic interactions in Ty1 Gag-RNA binding, we prepared a log-log plot of the K_D as a function of NaCl concentration (Figure 2F). The slope of this plot can be interpreted as the effective charge of the protein-nucleic acid interaction (Z_{eff}), and the Y-intercept represents $K_{D(1M)}$ which reports the non-electrostatic component strength of binding [24,25,37,38]. mTy1 RNA was bound by Gag-p49 with a $K_{D(1M)}$ of 3.4×10^{-5} , which is similar to that obtained for 18S rRNA binding ($K_{D(1M)}$ 4.2×10^{-5}) (Table 1). The Gag-p49-non-Psi Ty1 RNA complexes were characterized by the weakest non-electrostatic binding component ($K_{D(1M)}$ of 9×10^{-5}). The difference in non-electrostatic binding component for interactions of Gag-p45 with mTy1 RNA and 18S rRNA was greater than that observed for Gag-p49 ($K_{D(1M)}$ 1.1×10^{-5} and 4.6×10^{-5} , respectively). The number of electrostatic interactions was similar for all Gag-p49-RNA complexes ($Z_{\text{eff}} \sim 3$) (Table 1).

The Hill coefficient (n_H) ranged between 1.7 and 3.4 at each salt concentration for both Ty1 Gag proteins in interactions with all RNAs (Table 1). A value greater than 1 suggests positive cooperativity in which binding of ligand to one site facilitates binding of subsequent ligands at other sites. Therefore, RNA binding by Gag-p49 or Gag-p45 in vitro is a cooperative process that is independent of the RNA substrates analyzed here.

The results suggest that both Ty1 Gag proteins bind diverse RNAs with high affinity but low specificity. Given that trafficking of Ty1 RNA to retrosomes and packaging into VLPs occurs before maturation of Gag-p49 to p45, we focused on the RNA binding properties of Ty1 Gag-p49 in further analyses.

2.2. Gag Interactions in the Presence of Competitor RNA

To further investigate Gag-p49 binding specificity, we performed MST assays with diverse RNAs added in excess. In the competition reactions, fluorescently-labeled RNA was added to a mixture of Gag-p49 and competitor RNA at physiological salt concentration. The competition efficiency was determined by the competitor's relative ability to inhibit the formation of protein-labeled RNA complexes, as monitored by a change in K_D .

Initially, we performed Gag-p49 titration experiments with labeled RNA and total tRNA from *E. coli* in 50-fold molar excess. Adding excess tRNA only modestly decreased Gag-p49 binding to labeled mTy1 RNA or 18S rRNA, as shown by a ~2.5- or ~1.7-fold increase in K_D , respectively (Figure 3A, Table 2). Increasing the molar excess of tRNA to 133-fold resulted in similar increases in K_D for both mTy1 RNA and 18S rRNA (Figure S2). Thus, competition with excess tRNA failed to reveal specific Gag-p49 binding to RNA containing packaging signal, which is in contrast to results obtained with HIV-1 Gag [25,39].

Table 2. Impact of RNA competition on Ty1 Gag-p49 binding to mTy1 RNA and 18S rRNA. Binding parameters for Ty1 Gag-RNA interactions were obtained based on the Hill equation. Mean \pm SD of K_D from at least three independent MST experiments. n.d.: not determined.

Competitor \ Labeled RNA	mTy1 RNA		18S rRNA	
	K_D [nM]	n_H	K_D [nM]	n_H
competitor	176.8 \pm 11.4	1.9	253.9 \pm 11.9	2.2
tRNA <i>E.coli</i>	434.6 \pm 15.8	1.8	430.2 \pm 20.5	1.9
mTy1 RNA	1491.6 \pm 28.7	2.3	1224.9 \pm 42.3	1.7
18S rRNA	679.2 \pm 30.1	1.9	973.9 \pm 37	1.8
non-Psi Ty1 RNA	405.2 \pm 7.5	1.8	n.d.	n.d.
Δ S1a RNA	1150.2 \pm 36.5	2.6	n.d.	n.d.

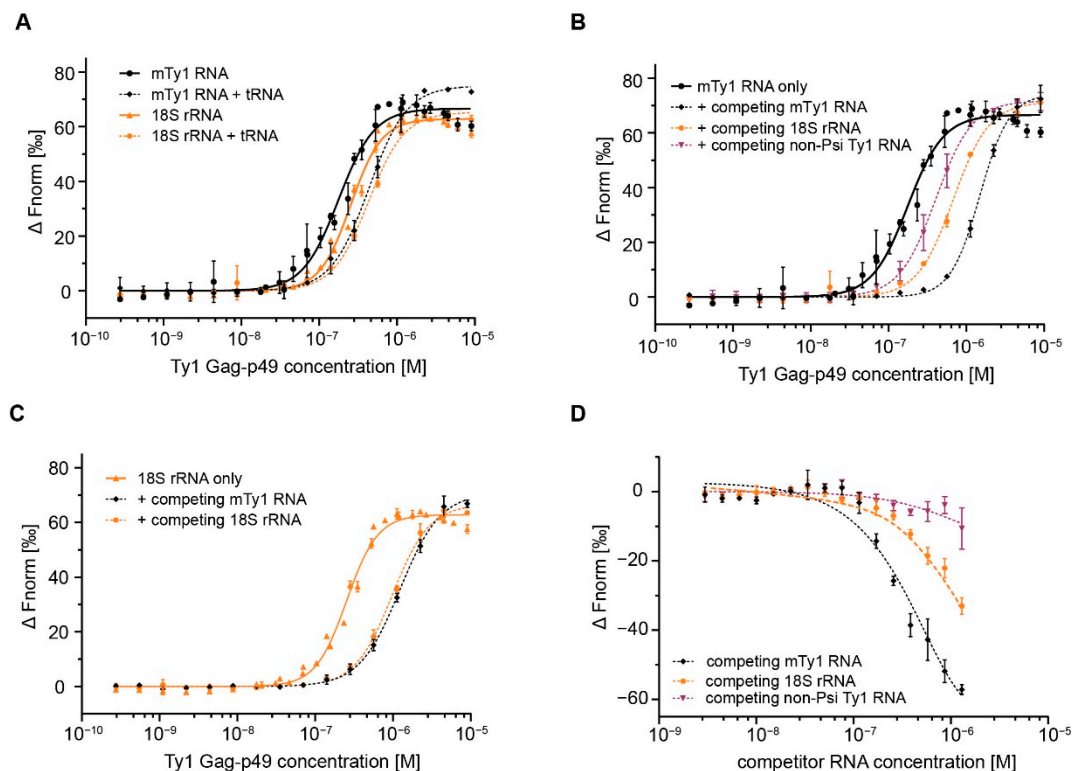


Figure 3. MST analysis of Ty1 Gag-p49 interactions with RNA in the presence of competitor RNA. (A) Dose–response binding curves of Ty1 Gag-p49 to mTy1 RNA (black) and 18S rRNA (orange) in the presence (dashed lines) or absence (solid lines) of total *E. coli* tRNA in 50-fold molar excess. (B) Dose–response binding curves of Ty1 Gag-p49 to mTy1 RNA in the presence of competing mTy1 RNA (black, dashed line), non-Psi Ty1 RNA (purple, dashed line) or 18S rRNA (orange, dashed line) in constant 10-fold molar excess. mTy1 RNA binding by Ty1 Gag-p49 in 150 mM NaCl lacking competitor RNA is used as a control (black, solid line) (C) Dose–response binding curves of Ty1 Gag-p49 to 18S rRNA in the presence of competing mTy1 RNA (black, dashed line) or 18S rRNA (orange, dashed line) in 10-fold molar excess. 18S rRNA binding by Ty1 Gag-p49 in 150 mM NaCl lacking competitor is used as a control (orange, solid line) (D) Dose-response binding curves of Ty1 Gag-p49 to mTy1 RNA in the presence of an increasing concentration of competing mTy1 RNA (black), non-Psi Ty1 RNA (purple) or 18S rRNA (orange).

Next, we used non-Psi Ty1 RNA and 18S rRNA as competitors in 10-fold molar excess to labeled mTy1 RNA. To establish the maximum level of competition, a control MST experiment was performed with labeled mTy1 RNA and unlabeled mTy1 RNA. Adding unlabeled mTy1 RNA in excess strongly inhibited the formation of Gag-p49 complexes with labeled mTy1 RNA, as reflected in an 8-fold increase in K_D (Figure 3B, Table 2). When unlabeled non-Psi Ty1 RNA or 18S rRNA competed with mTy1 RNA for the binding to Gag-p49, we observed a much smaller increase in K_D of 2.3-fold and 3.8-fold, respectively. We also tested the impact of the competitor on Gag-p49 interactions with 18S rRNA (Figure 3C, Table 2). Excess of 18S rRNA or mTy1 RNA affected the Gag-p49-18S rRNA complexes formation similarly with a ~3.8- and ~4.8-fold increase in K_D , respectively.

To obtain more information about the specificity of the complexes formed between Gag-p49 and mTy1 RNA, we performed competitive RNA binding assays with increasing amounts of RNA competitor at constant concentrations of labeled mTy1 RNA and Gag-p49. We observed that mTy1 RNA effectively self-competed for interactions with Gag-p49 (Figure 3D). At a 20-fold molar excess, competitor mTy1 RNA reduced complexes by 50%, and at 100-fold molar excess, essentially all labeled RNA was displaced from the complexes. When 18S rRNA or non-Psi RNA were used as competitors, the inhibitory effects were much weaker. About 80% and 40% of Gag-p49-mTy1 RNA complexes were still observed in the reactions at 100-fold molar excess of non-Psi Ty1 RNA or 18S rRNA, respectively.

Together, our results suggest that a slightly higher Gag-p49 binding affinity to mTy1 RNA is sufficient to make 18S rRNA and non-Psi Ty1 RNA less-effective competitors with mTy1 RNA for binding to Gag-p49. The reduced effect of non-Psi Ty1 RNA suggests this RNA may contain specific properties that may be investigated in future studies.

2.3. Comparing MST and EMSA to Detect Gag-RNA Complexes

To extend the MST analyses, we detected Ty1 Gag-RNA interactions using EMSA. Initially, we analyzed the binding affinity of Gag-p49 and Gag-p45 to mTy1 RNA and 18S rRNA at various salt concentrations. At physiological salt concentration, Gag-p49 and Gag-p45 bound mTy1 RNA with high affinity, as reflected by K_D values of 88.5 nM and 179 nM, respectively (Figure 4A, Table S1). Also, 18S rRNA was bound efficiently by both forms of Gag, and the difference between K_D values between Ty1 Gag complexes with mTy1 RNA and 18S rRNA was greater than that detected with MST. Increasing the salt concentration increased the K_D for Gag-p49 and Gag-p45 interactions with both mTy1 and 18S RNAs, but observed K_D changes were much smaller than that detected using MST. The binding of mTy1 RNA and 18S rRNA was relatively strong even at 800 mM NaCl ($K_D < 1 \mu\text{M}$) (Table S1). Similar to MST, we observed a slightly higher affinity of Ty1 Gag proteins to mTy1 RNA than to 18S rRNA at each salt concentration.

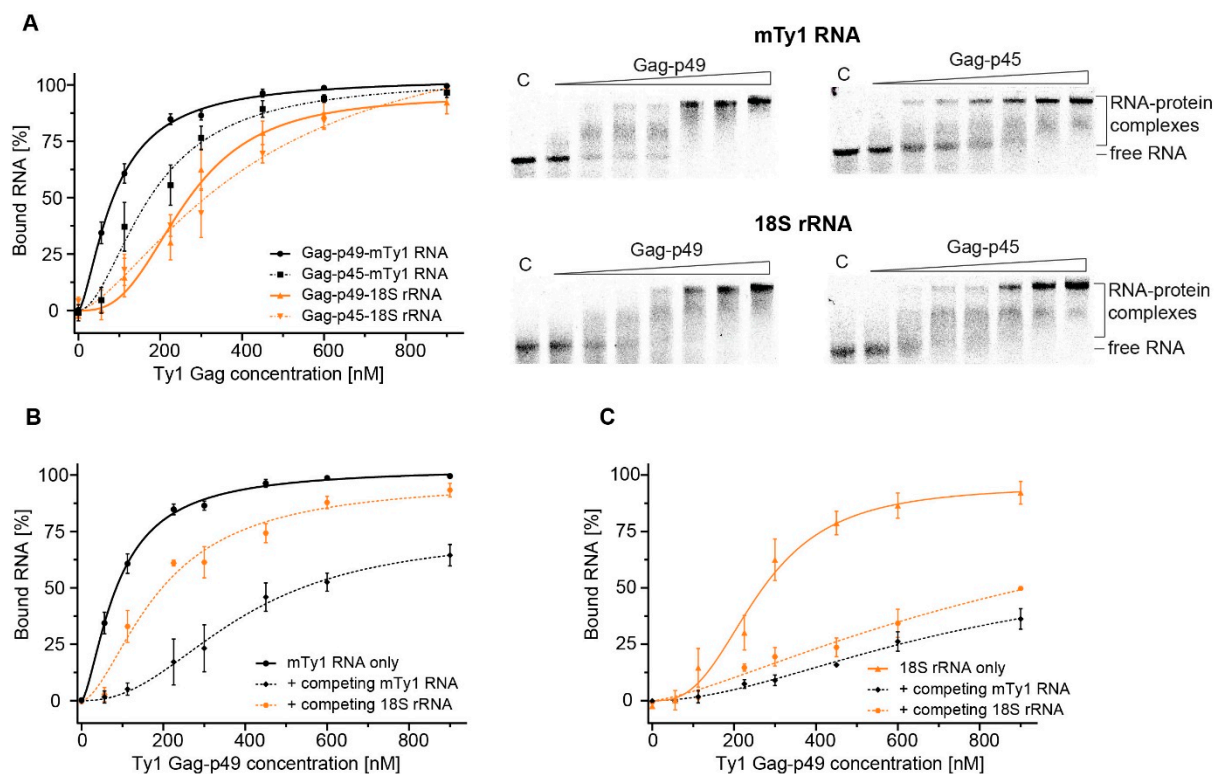


Figure 4. EMSA analysis of Ty1 Gag-RNA interactions. (A) Plots represent the fraction of bound mTy1 RNA (black) or 18S rRNA (orange) with increasing concentration of Gag-p49 (solid line) or Gag-p45 (dashed line) at 150 mM NaCl. Representative agarose gels are presented alongside the plots. Lanes denoted C lack protein. Effects of addition of competing mTy1 RNA (black, dashed line) or 18S rRNA (orange, dashed line) in 10-fold molar excess on Ty1 Gag-p49 interactions with mTy1 RNA (B) or 18S rRNA (C). The measurement of RNA binding by Ty1 Gag-p49 in 150 mM NaCl lacking competitor RNA is used as a control (solid lines).

EMSA was also used to monitor competition between mTy1 RNA and 18S rRNA for binding with Gag-p49. Similar to MST, unlabeled mTy1 RNA or 18S rRNA was added in 10-fold molar excess relative to labeled mTy1 RNA. Inhibition of the Gag-p49-mTy1 RNA interaction by 18S rRNA was 50% weaker than self-competition with mTy1 RNA (Figure 4B, Table S2), which is comparable to the results obtained using MST. Similar results

to those obtained by MST were also observed with labeled 18S rRNA. Gag-p49-18S rRNA interactions were comparable when competed with mTy1 and 18S rRNA, as reflected in a ~3- and ~4-fold increase of K_D , respectively (Figure 4C, Table S2).

Although the absolute binding parameters are different between MST and EMSA, this can be explained by differences in the accuracy and sensitivity of these two methods [40]. Importantly, the same trends were observed in the salt-dependent and competition experiments using EMSA and MST methods.

2.4. Role of the 5' Pseudoknot in Ty1 RNA Binding

In vitro and in virio RNA structure mapping experiments suggest that the stems of the long-range pseudoknot near the 5'-end of Ty1 gRNA and adjacent regions constitute primary binding sites for Gag-p45 [19,20,41]. However, whether the pseudoknot facilitates Ty1 Gag binding or Ty1 RNA interactions with Gag promote the formation of this structural motif remains unexplored. To explain the relationship between the pseudoknot motif and Gag, we studied Gag-p49 interactions with mTy1 RNA mutant containing a 5-nt deletion at the 5'-end (Δ S1a RNA) that is destabilized for pseudoknot formation. Prior SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) analyses show that this mutation disrupts the pseudoknot, but does not extensively alter the secondary structure in other regions of mTy1 RNA [20].

We analyzed the binding affinity of Gag-p49 for Δ S1a RNA at 150 mM NaCl. The results suggest that the 5-nt deletion moderately impairs protein binding, resulting in a K_D of 214.7 nM (Figure 5A; Figure S1; Table 1). Next, Δ S1a RNA was used to compete with wild-type mTy1 RNA for the binding to Gag-p49. A constant 10-fold molar excess of Δ S1a RNA caused a ~6.5-fold rise in K_D for Gag-p49-mTy1 RNA complexes formation (Figure 5B, Table 2), which was weaker than that obtained for self-competing mTy1 RNA. Moreover, Δ S1a RNA did not completely disrupt Gag-p49 interactions with mTy1 RNA when present at higher concentration (Figure 5C). At a 1:100 ratio of wild type to mutated mTy1 RNA, 15% of Gag-p49-mTy1 RNA complexes still remained in solution. Together, our data suggest that disrupting the pseudoknot affects the interactions between Gag-p49 and mTy1 RNA.

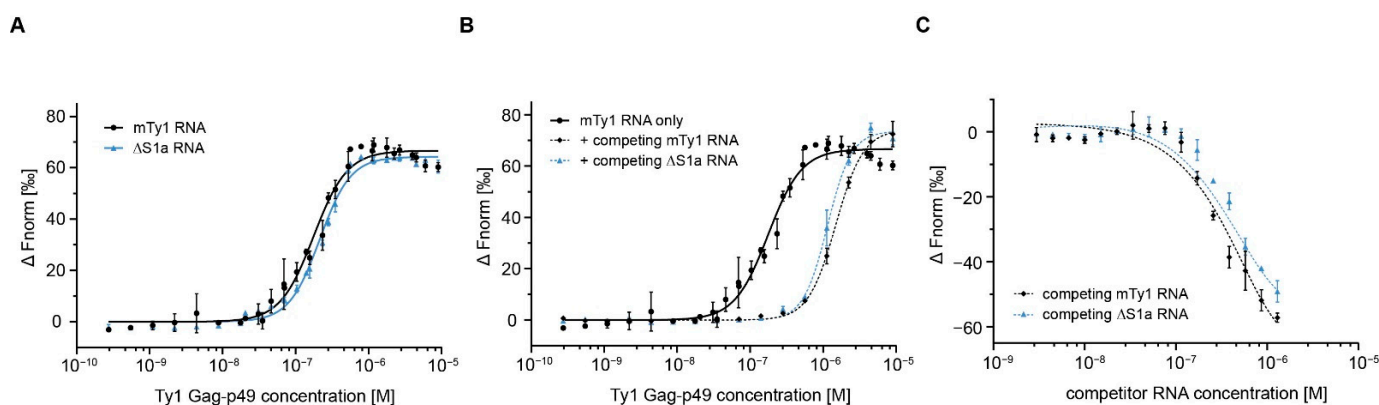


Figure 5. Role of the 5' pseudoknot in Gag-Ty1 RNA interactions. (A) Dose–response binding curves of Ty1 Gag-p49 with mTy1 RNA (black) and Δ S1a (blue) at 150 mM NaCl. (B) Dose–response binding curves Gag-p49-mTy1 RNA interactions in the presence of competing mTy1 RNA (black, dashed line) or Δ S1a RNA (blue, dashed line) in 10-fold molar excess. mTy1 RNA binding by Ty1 Gag-p49 in 150 mM NaCl lacking competitor is used as a control (black, solid line). (C) Dose–response binding curves of Ty1 Gag-p49 to mTy1 RNA in the presence of an increasing concentration of competing mTy1 RNA (black) or Δ S1a RNA (blue).

To further investigate differences in interactions between Gag-p49 and mTy1 or Δ S1a RNA, we used hydroxyl radical (HR) footprinting to map Gag-p49 binding sites in these two RNAs. The advantage of HR is that cleavage of the RNA backbone occurs independently of RNA secondary structure [42]. We compared HR cleavage profiles for mTy1 RNA

and Δ S1a RNA in the protein-free state. The most significant changes for mutated RNA were present near the pseudoknot stems, confirming RNA structural changes in the pseudoknot region (Figure S3). Protein-RNA complexes were formed under the same conditions as for the MST analyses. The regions protected by Gag-p49 against \bullet OH were identified by comparing the cleavage profiles of RNA in the presence and absence of Gag-p49. For mTy1 RNA, the most significant protection occurred near nucleotides forming the pseudoknot stems (Figure 6A). We also observed a strong Gag-p49 binding effect in PAL dimerization sequences and adjacent regions (nt 7–13 and 34–40). These data correspond well with results obtained using mature Ty1 Gag-p45 [20]. In contrast, the Δ S1a RNA cleavage profile in the presence of Gag-p49 was very similar to that obtained in the protein-free RNA state (Figure 6B). Strong protection at the pseudoknot region was not detected, but instead, multiple small cleavage decreases were observed along most of the RNA. We observed slightly stronger protection against cleavage only for nucleotides +30–42 and +65–68, but this was much weaker than for wild-type mTy1 RNA-Gag-p49 complexes. Thus, despite a high affinity of Gag-p49 for Δ S1a RNA, preferential protein binding sites were not detected in the absence of a stable pseudoknot. Taken together, our results provide evidence that the pseudoknot may facilitate Ty1 Gag binding. This novel function of the pseudoknot may also be important during the process of Ty1 retrotransposition.

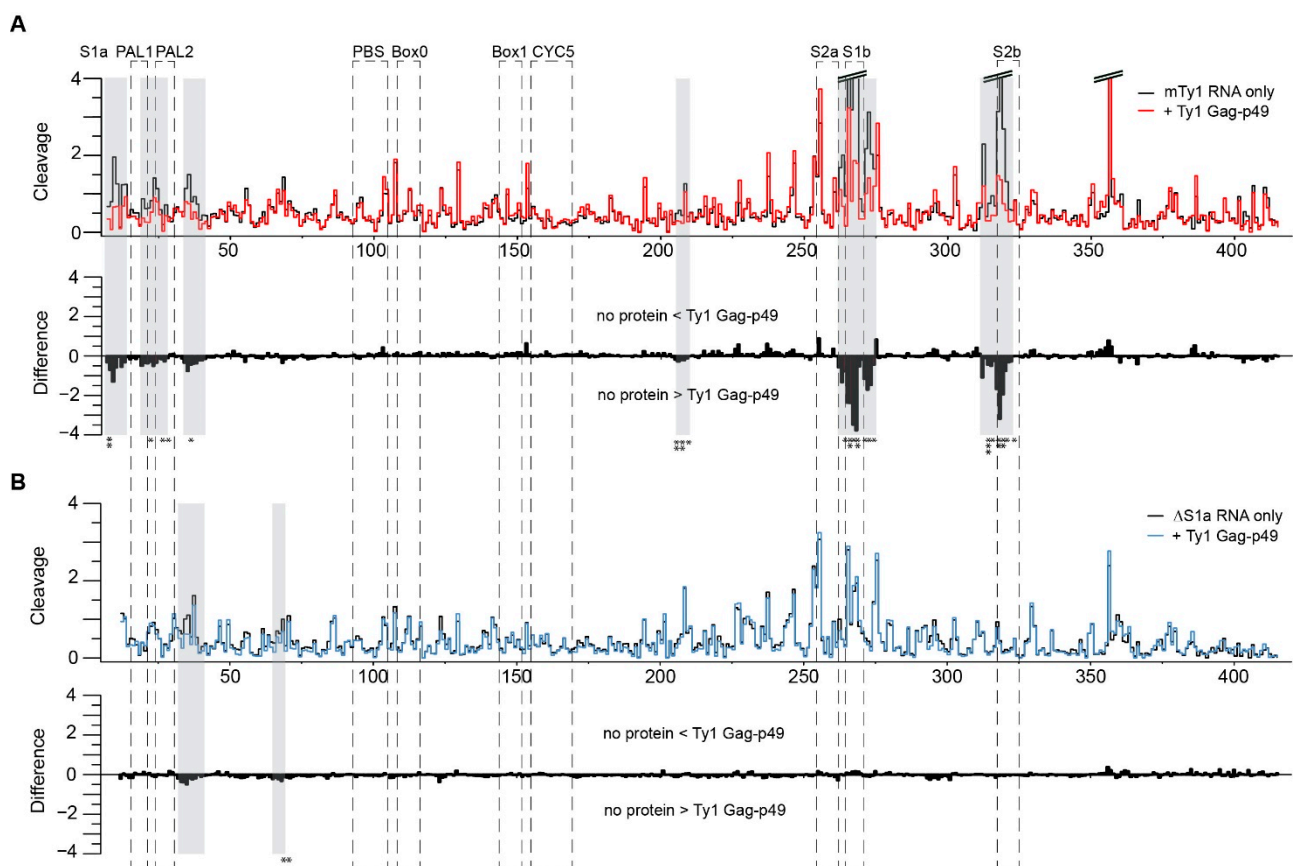


Figure 6. Ty1 Gag-p49 binding sites on mTy1 (A) and Δ S1a RNA (B). Hydroxyl radical (HR) cleavage profiles and difference plots of protein free RNA (black) in comparison with RNA probed in the presence of Gag-p49 (red for mTy1 RNA and blue for Δ S1a RNA). On the difference plots, sites of decreased HR cleavage upon Gag-p49 binding are indicated by negative peaks. Regions showing significant decrease HR reactivity over several nucleotides are indicated by gray stripes (absolute cleavage decrease of >0.2 and/or a p -value <0.05). Asterisks below the plot correspond to statistical significance (Student's t -test). Regions important for Ty1 retrotransposition are boxed.

3. Discussion

Genomic RNA of retroelements is selectively packaged into virions or VLPs despite the presence of a large excess of diverse cellular RNAs. This process has been extensively studied for retroviruses, but much less is known about how gRNA packaging occurs during the replication of endogenous retroelements. Here, we characterize *S. cerevisiae* Ty1 Gag interactions with RNAs with, or without, a packaging sequence. We performed a thorough analysis of RNA binding properties of recombinant immature and mature Ty1 Gag using two different techniques, MST and EMSA. No major differences in overall RNA binding between Gag-p49 and Gag-p45 were detected with either technique. Moreover, Gag-p49 recognizes mTy1 RNA regions that are very similar to that detected for Gag-p45 [20]. These findings are confirmed and extend prior work showing that both precursor and mature Ty1 Gag promote retosome formation, Ty1 RNA packaging, and assembling into VLPs in vivo [15,17,43]. Our analyses also reveal that both Ty1 Gag proteins bind RNA through a cooperative mechanism independent of the RNA molecule, which is similar to the general mechanism of action of retroviral Gag proteins [25]. Similar to retroviral virion assembly [44], any RNA might serve as a scaffold for Ty1 VLP formation and drive interactions of multiple Gag molecules with each other, as well as with RNA. However, testing this hypothesis requires further experiments.

Our study reveals that like retroviral Gag proteins [24,25,35,45,46], Ty1 Gag-p49 and -p45 bind diverse RNAs with high but similar affinities at physiological salt conditions. Gag-binding to Ty1 RNA containing Psi is less sensitive to an increase of NaCl concentration than binding to non-Psi Ty1 RNA. Although this result suggests a more specific Ty1 Gag interaction with Psi RNA, K_D changes caused by increasing ionic strength are similar for mTy1 RNA and 18S rRNA. In contrast to retroviral Gag [24,45], recognition of RNA containing Psi by Ty1 Gag-p49 or -p45 does not involve a significantly greater non-electrostatic binding component than interactions with control RNAs. Notably, the $K_{D(1M)}$ values for all tested Ty1 Gag-RNA interactions are comparable to those obtained for HIV-1 and RSV Gag complexes with Psi RNAs [24,45]. The differences between retroviral Gag and Ty1 Gag may result from the low sequence homology of RNA binding domains. Specific RNA binding by retroviral Gag proteins is mediated primarily by the NC domain, containing one or two well-structured zinc finger motifs [6–8]. Additionally, retroviral Gag polyproteins can interact with RNA by basic amino acid residues of NC and positively charged MA domain, but MA interactions are not specific and limited mainly to tRNA [47]. In contrast, Ty1 Gag contains a disordered region rich in basic amino acids arginine and lysine (25.53%) that binds RNA, but lacks a ZF motif [14]. We hypothesize that, due to the interaction of Ty1 Gag with RNA through stretches of basic amino acids, specific binding to mTy1 RNA is not detected in the presence of an excess of tRNA or at high ionic strength. For HIV-1 Gag, the addition of excess tRNA blocks the positively charged amino acids of MA and NC domains from interacting with non-specific RNA and enhances specific interaction by ZF motifs [25,48,49]. Interestingly, when non-Psi Ty1 RNA or 18S rRNA competes with mTy1 RNA for binding to Gag-p49, we observe Psi-dependent interactions with Ty1 RNA. Surprisingly, the non-Psi Ty1 RNA is less effective in competition analyses than 18S rRNA. The non-Psi Ty1 RNA used in our work is not only a part of the coding region but is present in the Ty1i transcript – a template for restriction factor p22 that is used to modulate the level of Ty1 retrotransposition in vivo [33,34]. Our findings suggest that the slight difference in Ty1 Gag affinity to diverse RNAs might help discriminate between Ty1 gRNA and other RNAs during packaging into VLPs. Nevertheless, it seems unlikely that differences in Ty1 Gag binding affinity to diverse RNAs in vitro fully account for selective Ty1 RNA packaging. Our results also raise the possibility that cellular or additional Ty1 factors help identify Ty1 gRNA as the preferred RNA for packaging into VLPs.

An analysis of the nucleotide composition of Gag-p49 binding sites in mTy1 RNA shows a lack of preference for particular nucleotide base(s) (Figure S4), which extends earlier work using the mature Gag-p45 [20]. These results are consistent with the generally accepted rule for RNA-binding proteins with intrinsically disordered regions rich in basic

residues [50,51]. Our data suggest that Gag-p49 can recognize specific structural motifs in Ty1 RNA. For various retroviruses, mutational analyses reveal that destabilization of particular stem-loops at the 5'-terminus negatively affects RNA packaging [52–55]. Here, we provide data suggesting that the stems of the 5'-long-range pseudoknot motif and adjacent nucleotides constitute primary binding sites of Gag-p49 in vitro. Destabilizing the pseudoknot structure results in the loss of preferential binding of Gag-p49 in this region of Ty1 RNA. In addition, we show that Gag-p49 can distinguish wild-type mTy1 RNA from RNA with an unstable pseudoknot ($\Delta S1a$) as evidenced by RNA competition analyses. Therefore, Ty1 Gag does not promote pseudoknot formation, rather the pseudoknot structure may facilitate Ty1 RNA binding by Gag. Since the pseudoknot is also present in Ty1 gRNA in vivo [56], this structural motif may be present in Ty1 gRNA before packaging into VLPs. However, the role of the pseudoknot motif in Ty1 RNA packaging is not completely understood. Point mutations in the pseudoknot stems do not affect the Ty1 RNA packaging [31], but defects resulting from complete disruption of base-pairing in S1 and S2 stems remain unexplored. Gag interactions with nucleotides forming the pseudoknot stems may also be important for other stages of the Ty1 life cycle.

Our study suggests that Ty1 Gag is less effective in specific gRNA recognition than retroviral Gag polyproteins. Indeed, cellular mRNAs lacking any obvious sequence homology with Ty1 gRNA associate with Ty1 VLPs [26,57,58]. On the other hand, only several, but simultaneously abundant at cytoplasm cellular mRNAs have been identified in VLPs, suggesting that Ty1 gRNA packaging is not a completely random process and encapsidation of other mRNAs is based on their abundance. It is likely that the recognition of Ty1 gRNA is facilitated by specific structural motifs present in the transcript. The correct Ty1 gRNA structure might promote the optimal presentation of Psi for interaction with Gag and efficient RNA packaging. Moreover, there are also specific in vivo mechanisms supporting selective packaging of Ty1 RNA. VLPs assembly occurs in retrosomes where Ty1 RNA and Gag are concentrated, which increases the likelihood of efficient Ty1 gRNA packaging. Besides, retrosomes partially co-localize with eukaryotic mRNA processing bodies (P-bodies) and two P-body-associated 5' to 3' mRNA decay pathways enhance Ty1 retrotransposition [18,59]. It raises the possibility that these 5' to 3' decay pathways degrade cellular mRNAs competing with Ty1 gRNA for binding to Gag and non-specific packaging into VLPs [59]. However, further studies are required to clarify these issues.

4. Materials and Methods

4.1. Expression and Purification of Ty1 Gag Proteins

The Ty1 Gag-p45-GST and Ty1 Gag-p49-GST fusion proteins were expressed in *Escherichia coli* BL21(DE3)pLysS strain (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA). The starter cultures were grown overnight at 37 °C in Luria-Bertani (LB) medium supplemented with ampicillin and chloramphenicol. These cultures (47mL) were used to inoculate large-scale 6 L cultures of LB medium. Cells were grown at 28 °C at 180 rpm to OD₆₀₀ of 0.6–0.7. Following the addition of IPTG (0.8 mM), the cultures were grown in an orbital shaker (180 rpm) at 18 °C for 18 h. Cells were pelleted by centrifugation at 4000× g for 10 min at 4 °C and resuspended in lysis buffer (50 mM HEPES pH 8.0, 1 M NaCl, 5 mM β-mercaptoethanol, 5 mM DTT, 1% Tween 20, 0.5 mg/mL lysozyme and protease inhibitor cocktail (Roche, Basel, Switzerland)). The cell suspension was sonicated 50 × 2 s pulse with 28 s pause after each pulse. Debris was removed by centrifugation at 22,000× g for 30 min at 4 °C. Nucleic acids were precipitated from supernatant by dropwise addition of 5% poly(ethyleneimine) solution (pH 7.9) to a final concentration of 0.45%, followed by incubation at 4 °C for 30 min and pelleting by centrifugation at 25,500× g for 30 min at 4 °C (this step was performed twice). The cleared supernatants were passed through a 0.45 μm filter and loaded onto a gravity flow column with pre-equilibrated Glutathione Sepharose 4 Fast Flow (Cytiva, Marlborough, MA, USA). The sepharose column was washed with 5 volumes of wash buffer (50 mM HEPES pH 8.0, 1 M NaCl, 5 mM β-mercaptoethanol, 5 mM DTT, 1% Tween 20) and 5 column volumes of

wash buffer without detergent. The GST tag was removed by thrombin cleavage on-resin overnight at 4 °C, and Ty1 Gag was eluted with protein storage buffer (50 mM HEPES pH 6.0, 1 M NaCl, 5mM β -mercaptoethanol, 5 mM DTT), concentrated by centrifugal filtration, aliquoted and stored at –80 °C.

4.2. DNA and RNA Substrates

DNA templates for in vitro transcription of Ty1 RNAs (mTy1 RNA, Δ S1a RNA and non-Psi Ty1 RNA) were obtained by PCR amplification from plasmid pBDG433 containing the Ty1-H3 element (Accession M18706.1). The DNA template for *S. cerevisiae* 18S rRNA was amplified from cDNA synthesized from total RNA. All primers are listed in Supplementary Table S3. Transcripts were synthesized using SP6- or T7-MEGAscript transcription kits (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA) according to the manufacturer's protocols and purified using a Direct-zol RNA MiniPrep Kit (Zymo Research, Irvine, CA, USA). Transcript integrity was monitored by agarose gel electrophoresis under denaturing conditions. 3'-end labeling of RNA was carried out overnight at 4 °C in an 18 μ L reaction containing T4 RNA ligase (Thermo Fisher Scientific Inc., Waltham, MA, USA), 1 \times T4 RNA Ligase Buffer, 20 μ M ATP, 20 μ M pCp-Cy5 (Jena Bioscience, Jena, Germany), and 30 pmols of RNA. Labeled RNA was purified using MEGAclean Transcription Clean-Up Kit (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA) as recommended by the manufacturer. Purified transcripts were stored at –20 °C.

4.3. Microscale Thermophoresis

Cy5-labeled RNA was denatured in water by heating at 90 °C for 2 min, placed on ice for 3 min, then adjusted to 30 nM with binding buffer (30 mM HEPES, pH 6.0, 10 mM DTT, 2.5 mM MgCl₂, 0.1% PF127), and incubated at 37 °C for 20 min. Ty1 Gag was dissolved in binding buffer and a series of 16 1:1 or 2:1 dilutions were prepared using the same buffer. Each protein dilution was mixed with one volume of labeled RNA, which led to a final concentration of RNA of 15 nM and final protein concentrations ranging from 0.000275 to 9 μ M. In salt-dependent experiments, the binding buffer was supplemented with the NaCl. For competition experiments, unlabeled RNA competitor (final concentration 750 nM or 2000 nM for tRNA and 150 nM for other competing RNAs) was combined with Gag before adding labeled RNA. For competitor titration experiments, increasing concentrations of unlabeled RNA competitor (final concentration from 2.97 to 1300 nM) were used, and binding was performed with a constant concentration of Gag-p49 (2000 nM) and labeled mTy1 RNA (13 nM). After incubation in the dark at 4 °C for 7.5 h, the samples were loaded into standard Monolith NT.115 Capillaries (NanoTemper Technologies, München, Germany) according to the manufacturer's instructions. Various temperatures (37 °C, 30 °C, 22 °C, 4 °C) and reaction times (from 5 min to 18 h) were tested in pilot experiments. Finally, a low temperature (4 °C) was used in the reactions to reduce aggregate formation, and a time of 7.5 h was necessary to reach a plateau in these conditions. MST was measured using a Monolith NT.115 instrument (NanoTemper Technologies, München, Germany) set to 22 °C. Instrument parameters were adjusted to 80–100% LED power and medium MST power. The data obtained from at least three independent measurements were analyzed with MO. Affinity Analysis software (version 2.3, NanoTemper Technologies, München, Germany) using the signal from an MST-on time of 5 s. The reproducibility of the experiments was assessed by standard deviation.

4.4. Electrophoretic Mobility Shift Assays

Cy5-labeled RNA (0.1 pmol) was refolded in water by heating at 90 °C for 2 min, placing on ice for 3 min and incubation at 37 °C for 20 min following the addition of binding buffer (30 mM HEPES, pH 6.0, 10 mM DTT, 2.5 mM MgCl₂, 0.1% PF127). To induce complex formation, RNA was incubated with increasing concentrations of protein (from 0 to 900 nM) for 15 min on ice to minimize the formation of aggregates. In salt-dependent experiments, the binding buffer was supplemented with NaCl. In the competition experiments, the

unlabeled RNA competitor (100 nM) was incubated for 10 min with Gag before the labeled RNA was added. Samples (10 μ L) were mixed with 2 μ L of 25% Ficoll 400, and RNA-protein complexes were resolved on a 1.1% agarose gel in $0.5 \times$ TB running buffer at 4 $^{\circ}$ C. Gels were quantified by imaging using Fujifilm FLA-5100 imaging system with MultiGaugeV 3.0 software (Fujifilm Life Science, Stamford, CT, USA). Data were then analyzed using a non-linear regression function (option *Specific binding with Hill Slope*; GraphPad Prism 8, GraphPad Software, San Diego, CA, USA). In all cases, at least three independent experiments were performed and the data presented were representative of the whole. The reproducibility of the experiments was assessed by standard deviation.

4.5. Hydroxyl Radical Footprinting, Primer Extension Reactions, and Data Processing

RNA samples (5 pmol) were denatured by heating at 90 $^{\circ}$ C for 2 min in water followed by incubation for 3 min on ice. Next, binding buffer (30 mM HEPES, pH 6.0, 10 mM DTT, 2.5 mM MgCl₂, 0.1% PF127, 75 mM NaCl) was added and RNA was incubated for 20 min at 37 $^{\circ}$ C. Subsequently, 200 pmol of Ty1 Gag-p49 in a total volume of 6 μ L was added to a 70 μ L reaction, and samples were incubated on ice for 15 min. As a control for non-specific cleavage, protein storage buffer was added instead of protein. To initiate the production of hydroxyl radicals, 1 μ L of 2.5 mM (NH₄)Fe(SO₄)₂, 50 mM sodium ascorbate, 1.5% H₂O₂, and 2.75 mM EDTA were applied on the wall of the tube followed by centrifugation. Reactions were incubated for 15 s at 24 $^{\circ}$ C and quenched by the addition of 20 μ L of stop solution (100 mM thiourea, 200 mM EDTA). RNA was purified using Direct-zol RNA MiniPrep Kit (Zymo Research, Irvine, CA, USA). For detection of cleavage sites, samples containing 4–5 pmols of RNA and 10 pmols of fluorescently labeled [Cy5 (+) and Cy5.5 (–)] primer PR3 [5'-TCAGGTGATGGAGTGCTCAG-3'], 0.1 mM EDTA, were incubated at 95 $^{\circ}$ C for 3 min, at 37 $^{\circ}$ C for 10 min and at 55 $^{\circ}$ C for 2 min, next reverse transcribed at 50 $^{\circ}$ C for 50 min using Superscript III Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA). Sequencing ladders were carried out using plasmid pBDG433 as a template, primer PR3 labeled with WellRed D2 (ddA) or IRD-800 (ddT), and a Thermo Sequenase Cycle Sequencing kit (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham, MA, USA) as recommended by the manufacturer. The following cycling parameters were used: 96 $^{\circ}$ C/10 s; 24 cycles: 96 $^{\circ}$ C/20 s, 55 $^{\circ}$ C/20 s, 72 $^{\circ}$ C/20 s; 72 $^{\circ}$ C/1 min. cDNA samples and sequencing ladders were purified using ZR DNA Sequencing Clean-up Kit (Zymo Research, Irvine, CA, USA) and analyzed on a GenomeLab GeXP Analysis System (Beckman-Coulter, Brea, CA, USA). Raw data from capillary electrophoresis were processed using SHAPEfinder software [60], and then normalized and converted into nucleotide reactivity tables using RNATHOR [61]. All reactivity data used in the analysis were averaged from at least two independent experiments. To assess the reproducibility of experiments, we performed statistical analysis using RNATHOR. To determine statistically significant changes in HR profiles of RNAs probed in the presence and absence of protein, we used Student's t-test, and a *p*-value was calculated separately for each nucleotide across the samples (*p*-values < 0.05 indicate statistically significant difference).

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22169103/s1>.

Author Contributions: K.P.-W. supervised the project; J.G. designed experiments; J.G. performed MST and footprinting experiments; J.G. and A.A.-R. performed EMSA experiments; J.G. performed data analysis and, together with K.P.-W. and D.J.G. interpreted the results. J.G., K.P.-W., and D.J.G. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Polish National Science Centre [2019/35/N/NZ1/01954 to J.G. and 2016/22/E/NZ3/00426 to K.P.W.]; National Institutes of Health [GM124216 to D.J.G.]. Funding for open access charge: Polish National Science Centre [2016/22/E/NZ3/00426 to K.P.W.].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Most of the data are provided in this work and in Supplementary Data 1. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We thank members of K.P.-W. group for exceptional support and fruitful discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eickbush, T.H.; Malik, H.S. Origin and Evolution of retrotransposons. In *Mobile DNA II*; Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M., Eds.; ASM Press: Washington, DC, USA, 2002; pp. 1111–1144.
2. Coffin, J.M.; Hughes, S.H.; Varmus, H.E. (Eds.) *Retroviruses*; Cold Spring Harbor Laboratory Press: New York, NY, USA, 1997.
3. Garfinkel, D.J.; Boeke, J.D.; Fink, G.R. Ty element transposition: Reverse transcriptase and virus-like particles. *Cell* **1985**, *42*, 507–517. [[CrossRef](#)]
4. Shiba, T.; Saigo, K. Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. *Nature* **1983**, *302*, 119–124. [[CrossRef](#)]
5. Zhang, W.; Mendonça, L.M.; Mansky, L.M. The Retrovirus Capsid Core. In *Virus Protein and Nucleoprotein Complexes*; Harris, J.R., Bhella, D., Eds.; Springer Nature Singapore Pte Ltd.: Singapore, 2018; pp. 169–187.
6. Olson, E.D.; Musier-Forsyth, K. Retroviral Gag protein–RNA interactions: Implications for specific genomic RNA packaging and virion assembly. *Semin. Cell Dev. Biol.* **2019**, *86*, 129–139. [[CrossRef](#)] [[PubMed](#)]
7. Rein, A.; Datta, S.; Jones, C.P.; Musier-Forsyth, K. Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* **2011**, *36*, 373–380. [[CrossRef](#)] [[PubMed](#)]
8. Muriaux, D.; Darlix, J.-L. Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol.* **2010**, *7*, 744–753. [[CrossRef](#)] [[PubMed](#)]
9. Mailler, E.; Bernacchi, S.; Marquet, R.; Paillart, J.-C.; Vivet-Boudou, V.; Smyth, R.P. The Life-Cycle of the HIV-1 Gag–RNA Complex. *Viruses* **2016**, *8*, 248. [[CrossRef](#)] [[PubMed](#)]
10. Pachulska-Wieczorek, K.; Le Grice, S.F.; Purzycka, K.J. Determinants of Genomic RNA Encapsidation in the *Saccharomyces cerevisiae* Long Terminal Repeat Retrotransposons Ty1 and Ty3. *Viruses* **2016**, *8*, 193. [[CrossRef](#)]
11. Al-Khayat, A.H.; Bhella, D.; Kenney, J.M.; Roth, J.-F.; Kingsman, A.J.; Martin-Rendon, E.; Saibil, H.R. Yeast Ty retrotransposons assemble into virus-like particles whose T-numbers depend on the C-terminal length of the capsid protein. *J. Mol. Biol.* **1999**, *292*, 65–73. [[CrossRef](#)]
12. Merkulov, G.V.; Swiderek, K.M.; Brachmann, C.B.; Boeke, J.D. A critical proteolytic cleavage site near the C terminus of the yeast retrotransposon Ty1 Gag protein. *J. Virol.* **1996**, *70*, 5548–5556. [[CrossRef](#)]
13. Tucker, J.M.; Larango, M.E.; Wachsmuth, L.; Kannan, N.; Garfinkel, D.J. The Ty1 Retrotransposon Restriction Factor p22 Targets Gag. *PLoS Genet.* **2015**, *11*, e1005571. [[CrossRef](#)]
14. Cristofari, G.; Ficheux, D.; Darlix, J.-L. The Gag-like Protein of the Yeast Ty1 Retrotransposon Contains a Nucleic Acid Chaperone Domain Analogous to Retroviral Nucleocapsid Proteins. *J. Biol. Chem.* **2000**, *275*, 19210–19217. [[CrossRef](#)]
15. Checkley, M.A.; Mitchell, J.A.; Eizenstat, L.D.; Lockett, S.J.; Garfinkel, D.J. Ty1 Gag Enhances the Stability and Nuclear Export of Ty1 mRNA. *Traffic* **2012**, *14*, 57–69. [[CrossRef](#)] [[PubMed](#)]
16. Malagon, F.; Jensen, T.H. The T Body, a New Cytoplasmic RNA Granule in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **2008**, *28*, 6022–6032. [[CrossRef](#)] [[PubMed](#)]
17. Malagon, F.; Jensen, T.H. T-body formation precedes virus-like particle maturation in *S. cerevisiae*. *RNA Biol.* **2011**, *8*, 184–189. [[CrossRef](#)] [[PubMed](#)]
18. Checkley, M.A.; Nagashima, K.; Lockett, S.J.; Nyswaner, K.M.; Garfinkel, D.J. P-Body Components Are Required for Ty1 Retrotransposition during Assembly of Retrotransposition-Competent Virus-Like Particles. *Mol. Cell. Biol.* **2010**, *30*, 382–398. [[CrossRef](#)]
19. Nishida, Y.; Pachulska-Wieczorek, K.; Blaszczyk, L.; Saha, A.; Gumna, J.; Garfinkel, D.J.; Purzycka, K.J. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic Acids Res.* **2015**, *43*, 7414–7431. [[CrossRef](#)]
20. Gumna, J.; Purzycka, K.J.; Ahn, H.W.; Garfinkel, D.J.; Pachulska-Wieczorek, K. Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biol.* **2019**, *16*, 1749–1763. [[CrossRef](#)]
21. Cristofari, G.; Bampi, C.; Wilhelm, M.; Wilhelm, F.; Darlix, J. A 5′-3′ long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *EMBO J.* **2002**, *21*, 4368–4379. [[CrossRef](#)]
22. Kuzembayeva, M.; Dilley, K.; Sardo, L.; Hu, W.-S. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* **2014**, *454–455*, 362–370. [[CrossRef](#)]
23. D’Souza, V.; Summers, M.F. How retroviruses select their genomes. *Nat. Rev. Genet.* **2005**, *3*, 643–655. [[CrossRef](#)]
24. Webb, J.A.; Jones, C.P.; Parent, L.; Rouzina, I.; Musier-Forsyth, K. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging. *RNA* **2013**, *19*, 1078–1088. [[CrossRef](#)]
25. Comas-Garcia, M.; Datta, S.A.; Baker, L.; Varma, R.; Gudla, P.R.; Rein, A. Dissection of specific binding of HIV-1 Gag to the ‘packaging signal’ in viral RNA. *eLife* **2017**, *6*, e27055. [[CrossRef](#)]

26. Xu, H.; Boeke, J.D. Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: Analysis of mini-Ty1 elements. *Mol. Cell. Biol.* **1990**, *10*, 2695–2702. [[CrossRef](#)]
27. Bolton, E.C.; Coombes, C.; Eby, Y.; Cardell, M.; Boeke, J.D. Identification and characterization of critical cis-acting sequences within the yeast Ty1 retrotransposon. *RNA* **2005**, *11*, 308–322. [[CrossRef](#)] [[PubMed](#)]
28. Luschnig, C.; Bachmair, A. RNA Packaging of Yeast Retrotransposon Ty1 in the Heterologous Host, *Escherichia coli*. *Biol. Chem.* **1997**, *378*, 39–46. [[CrossRef](#)]
29. Friant, S.; Heyman, T.; Wilhelm, M.L.; Wilhelm, F.X. Extended Interactions Between the Primer tRNAⁱMet and Genomic RNA of the Yeast Ty1 Retrotransposon. *Nucleic Acids Res.* **1996**, *24*, 441–449. [[CrossRef](#)]
30. Friant, S.; Heyman, T.; Byström, A.S.; Wilhelm, M.; Wilhelm, F.X. Interactions between Ty1 Retrotransposon RNA and the T and D Regions of the tRNAⁱMet Primer Are Required for Initiation of Reverse Transcription In Vivo. *Mol. Cell. Biol.* **1998**, *18*, 799–806. [[CrossRef](#)] [[PubMed](#)]
31. Huang, Q.; Purzycka, K.J.; Lusvarghi, S.; Li, D.; LeGrice, S.F.; Boeke, J.D. Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA* **2013**, *19*, 320–332. [[CrossRef](#)] [[PubMed](#)]
32. Jerabek-Willemsen, M.; Wienken, C.J.; Braun, D.; Baaske, P.; Dühr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *ASSAY Drug Dev. Technol.* **2011**, *9*, 342–353. [[CrossRef](#)] [[PubMed](#)]
33. Saha, A.; Mitchell, J.A.; Nishida, Y.; Hildreth, J.E.; Ariberre, J.A.; Gilbert, W.V.; Garfinkel, D.J. A trans-Dominant Form of Gag Restricts Ty1 Retrotransposition and Mediates Copy Number Control. *J. Virol.* **2015**, *89*, 3922–3938. [[CrossRef](#)] [[PubMed](#)]
34. Błaszczuk, L.; Biesiada, M.; Saha, A.; Garfinkel, D.J.; Purzycka, K.J. Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression. *Viruses* **2017**, *9*, 74. [[CrossRef](#)]
35. Comas-Garcia, M.; Kroupa, T.; Datta, S.A.; Harvin, D.P.; Hu, W.-S.; Rein, A. Efficient support of virus-like particle assembly by the HIV-1 packaging signal. *eLife* **2018**, *7*. [[CrossRef](#)]
36. El-Wahab, E.A.; Smyth, R.; Mailler, E.; Bernacchi, S.; Vivet-Boudou, V.; Hijnen, M.; Jossinet, F.; Mak, J.; Paillart, J.-C.; Marquet, R. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* **2014**, *5*, 4304. [[CrossRef](#)]
37. Record, M.T.; Lohman, T.M.; de Haseth, P. Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* **1976**, *107*, 145–158. [[CrossRef](#)]
38. Rouzina, I.; Bloomfield, V.A. Competitive electrostatic binding of charged ligands to polyelectrolytes: Practical approach using the non-linear Poisson-Boltzmann equation. *Biophys. Chem.* **1997**, *64*, 139–155. [[CrossRef](#)]
39. Jones, C.P.; Datta, S.; Rein, A.; Rouzina, I.; Musier-Forsyth, K. Matrix Domain Modulates HIV-1 Gag's Nucleic Acid Chaperone Activity via Inositol Phosphate Binding. *J. Virol.* **2010**, *85*, 1594–1603. [[CrossRef](#)] [[PubMed](#)]
40. König, F.; Schubert, T.; Längst, G. The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLoS ONE* **2017**, *12*, e0178875. [[CrossRef](#)] [[PubMed](#)]
41. Purzycka, K.J.; Legiewicz, M.; Matsuda, E.; Eizentstat, L.D.; Lusvarghi, S.; Saha, A.; Le Grice, S.F.J.; Garfinkel, D.J. Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res.* **2012**, *41*, 463–473. [[CrossRef](#)] [[PubMed](#)]
42. Nilsen, T.W. Mapping RNA-Protein Interactions Using Hydroxyl-Radical Footprinting. *Cold Spring Harb. Protoc.* **2014**, *2014*. [[CrossRef](#)]
43. Kingsman, S.M.; Kingsman, A.J.; Martin-Rendon, E. Possible regulatory function of the *Saccharomyces cerevisiae* Ty1 retrotransposon core protein. *Yeast* **2000**, *16*, 921–932. [[CrossRef](#)]
44. Muriaux, D.; Mirro, J.; Harvin, D.; Rein, A. RNA is a structural element in retrovirus particles. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5246–5251. [[CrossRef](#)] [[PubMed](#)]
45. Rye-McCurdy, T.; Olson, E.D.; Liu, S.; Binkley, C.; Reyes, J.-P.; Thompson, B.; Flanagan, J.M.; Parent, L.J.; Musier-Forsyth, K. Functional Equivalence of Retroviral MA Domains in Facilitating Psi RNA Binding Specificity by Gag. *Viruses* **2016**, *8*, 256. [[CrossRef](#)]
46. Pachulska-Wieczorek, K.; Błaszczuk, L.; Biesiada, M.; Adamiak, R.W.; Purzycka, K.J. The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag. *Retrovirology* **2016**, *13*, 1–15. [[CrossRef](#)] [[PubMed](#)]
47. Kutluay, S.B.; Zang, T.; Blanco-Melo, D.; Powell, C.; Jannain, D.; Errando, M.; Bieniasz, P.D. Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis. *Cell* **2014**, *159*, 1096–1109. [[CrossRef](#)] [[PubMed](#)]
48. Berkowitz, R.D.; Luban, J.; Goff, S.P. Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* **1993**, *67*, 7190–7200. [[CrossRef](#)] [[PubMed](#)]
49. Comas-Garcia, M.; Davis, S.R.; Rein, A. On the Selective Packaging of Genomic RNA by HIV-1. *Viruses* **2016**, *8*, 246. [[CrossRef](#)]
50. Corley, M.; Burns, M.C.; Yeo, G.W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol. Cell* **2020**, *78*, 9–29. [[CrossRef](#)]
51. Balcerak, A.; Trebńska-Stryjewska, A.; Konopiński, R.; Wakuła, M.; Grzybowska, E.A. RNA-protein interactions: Disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biol.* **2019**, *9*, 190096. [[CrossRef](#)] [[PubMed](#)]
52. Russell, R.S.; Hu, J.; Bériault, V.; Moulard, A.J.; Kleiman, L.; Wainberg, M.A.; Liang, C. Sequences Downstream of the 5' Splice Donor Site Are Required for both Packaging and Dimerization of Human Immunodeficiency Virus Type 1 RNA. *J. Virol.* **2003**, *77*, 84–96. [[CrossRef](#)]
53. Doria-Rose, N.A.; Vogt, V.M. In Vivo Selection of *Rous Sarcoma Virus* Mutants with Randomized Sequences in the Packaging Signal. *J. Virol.* **1998**, *72*, 8073–8082. [[CrossRef](#)] [[PubMed](#)]

54. Gherghe, C.; Lombo, T.; Leonard, C.W.; Datta, S.; Bess, J.W.; Gorelick, R.J.; Rein, A.; Weeks, K.M. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19248–19253. [[CrossRef](#)] [[PubMed](#)]
55. Mougél, M.; Barklis, E. A role for two hairpin structures as a core RNA encapsidation signal in murine leukemia virus virions. *J. Virol.* **1997**, *71*, 8061–8065. [[CrossRef](#)] [[PubMed](#)]
56. Andrzejewska, A.; Zawadzka, M.; Gumna, J.; Garfinkel, D.J.; Pachulska-Wieczorek, K. *In vivo* structure of the Ty1 retrotransposon RNA genome. *Nucleic Acids Res.* **2021**, *49*, 2878–2893. [[CrossRef](#)] [[PubMed](#)]
57. Maxwell, P.H.; Coombes, C.; Kenny, A.E.; Lawler, J.F.; Boeke, J.; Curcio, M.J. Ty1 Mobilizes Subtelomeric Y' Elements in Telomerase-Negative *Saccharomyces cerevisiae* Survivors. *Mol. Cell. Biol.* **2004**, *24*, 9887–9898. [[CrossRef](#)]
58. Maxwell, P.H.; Curcio, M.J. Retrosequence formation restructures the yeast genome. *Genes Dev.* **2007**, *21*, 3308–3318. [[CrossRef](#)]
59. Dutko, J.A.; Kenny, A.E.; Gamache, E.R.; Curcio, M.J. 5' to 3' mRNA Decay Factors Colocalize with Ty1 Gag and Human APOBEC3G and Promote Ty1 Retrotransposition. *J. Virol.* **2010**, *84*, 5052–5066. [[CrossRef](#)]
60. Vasa, S.M.; Guex, N.; Wilkinson, K.A.; Weeks, K.M.; Giddings, M.C. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **2008**, *14*, 1979–1990. [[CrossRef](#)]
61. Gumna, J.; Zok, T.; Figurski, K.; Pachulska-Wieczorek, K.; Szachniuk, M. RNATHor—Fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis. *PLoS ONE* **2020**, *15*, e0239287. [[CrossRef](#)]

SUPPLEMENTARY DATA 1

RNA binding properties of the Ty1 LTR-retrotransposon Gag protein

Julita Gumna, Angelika Andrzejewska-Romanowska, David J. Garfinkel and Katarzyna Pachulska-Wieczorek

Figure S1. Raw MST traces for Ty1 Gag-p49 interactions with mTy1 RNA (A), non-Psi Ty1 RNA (B), 18S rRNA (C), and Δ S1a RNA (D) at 150mM NaCl.

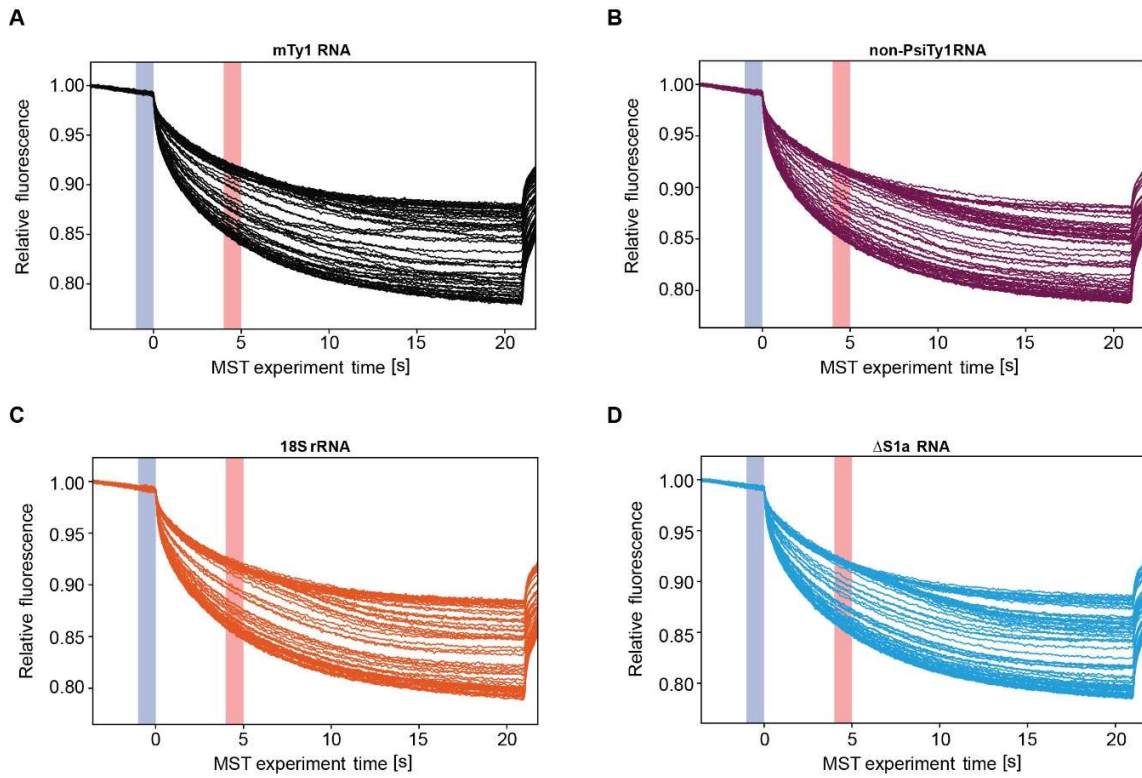


Figure S2. Impact of a large excess of tRNA on Ty1 Gag-RNA binding. Dose–response binding curves of Ty1 Gag-p49 to mTy1 RNA (black) and 18S rRNA (orange) in the presence (dashed lines) or absence (solid lines) of total *E.coli* tRNA in ~133-fold molar excess. Lines represent fits of the data points using the Hill equation. The excess of tRNA resulted in 5.2-fold and 5.7-fold increase in K_D value for mTy1 RNA binding and for 18S rRNA, respectively (K_D 928.7nM and 1445.7nM).

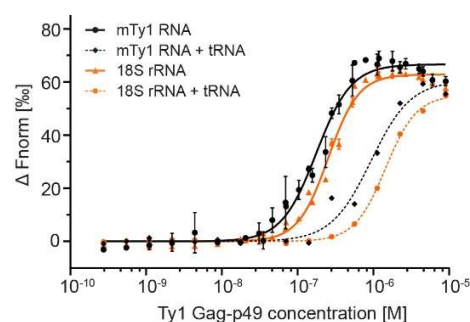


Table S1. Impact of buffer ionic strength on Ty1 Gag-RNA interactions analyzed by EMSA. Binding parameters were obtained based on quantitative analysis of gel images. Mean \pm SD of K_D from at least 3 independent EMSA experiments.

	K_D [nM]	n_H	K_D [nM]	n_H	K_D [nM]	n_H	K_D [nM]	n_H
	150 mM NaCl		300 mM NaCl		500 mM NaCl		800 mM NaCl	
mTy1 RNA-Gag-p49	88.5 \pm 3.6	1.5	136.7 \pm 5.6	2.2	212.4 \pm 27.4	1.5	454.0 \pm 133.5	1.5
18S rRNA-Gag-p49	261.5 \pm 14.2	2.8	551.8 \pm 98	1.9	711.7 \pm 307.9	1.7	634.1 \pm 133.5	2.5
mTy1 RNA-Gag-p45	179 \pm 20.9	1.9	227.9 \pm 20	2	363.1 \pm 56.2	1.9	469.9 \pm 29.1	4.6
18S rRNA-Gag-p45	425 \pm 95.1	1.5	514.7 \pm 56.7	2.5	936 \pm 239.7	2.6	891 \pm 309.8	2.6

Table S2. Impact of RNA competitor on Ty1 Gag-p49 binding to mTy1 RNA and 18S rRNA analyzed by EMSA. Binding parameters for Ty1 Gag-RNA interactions were obtained based on quantitative analysis of gel images. Mean \pm SD of K_D from at least 3 independent EMSA experiments.

Competitor \ Labeled RNA	mTy1 RNA		18S rRNA	
	K_D [nM]	n_H	K_D [nM]	n_H
- competitor	88.5 \pm 3.6	1.5	261.5 \pm 20.7	2.8
mTy1 RNA	390.1 \pm 34.7	2.3	752.7 \pm 217.6	1.9
18S rRNA	191.4 \pm 33.1	1.7	1043 \pm 124	1.3

Figure S3. Comparative analysis of the 3D structure of wild type mTy1 RNA and Δ S1a RNA. Hydroxyl radical (HR) cleavage profiles and difference plot of protein free mTy1 RNA (black) in comparison with protein free Δ S1a RNA (blue). On the difference plot, sites of decreased HR cleavage for Δ S1a RNA are indicated by positive peaks. Regions important for Ty1 retrotransposition are boxed.

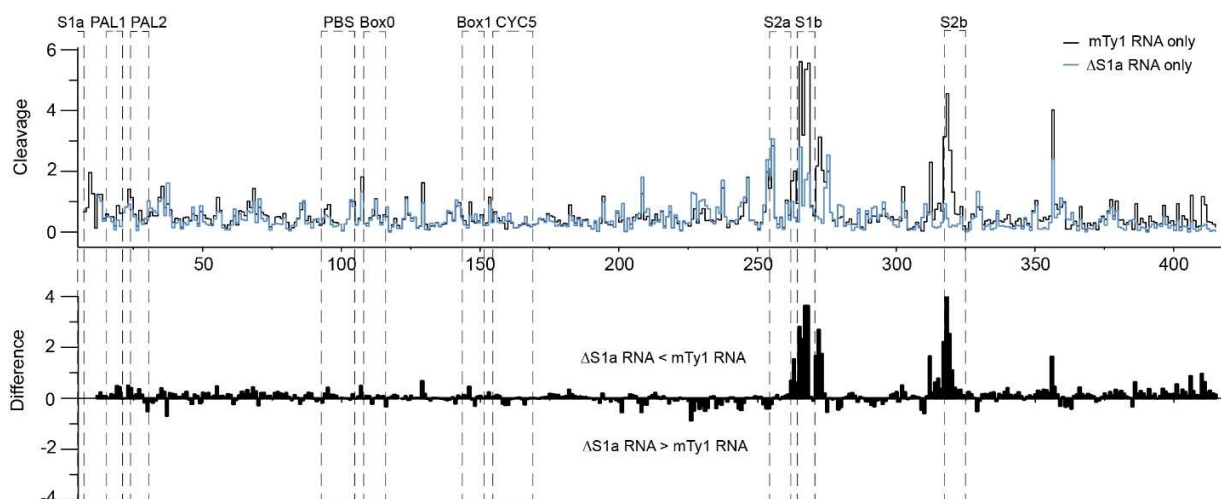


Figure S4. Frequency of nucleotide occurrence within Ty1 Gag-p49 binding sites and their vicinities represented as a logo (<http://weblogo.berkeley.edu/>).



Table S3. Primers used for DNA templates construction.

Primer	Sequence
F-miniRNA	GATTTAGGTGACACTATAGAGGAGAACTTCTAGT
R-miniRNA	ACATTGGTGGTGGTCTGAC
Δ S1a_RNA_PF	GATTTAGGTGACACTATAGAACTTCTAGTATATT
non-Psi Ty1_RNA_PF	TAATACGACTCACTATAGGGTCAAAGACATCCTATCC
non-Psi Ty1_RNA_PR	TTTACTGTAGATTTCAGTAAGTTTCTGG
18S_rRNA_PF	GATTTAGGTGACACTATAGTATCTGGTTGATCCT
18S_rRNA_PR	CGCGGCTGCTGGCACCAGAC

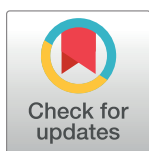
RESEARCH ARTICLE

RNAuthor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis

Julita Gumna¹, Tomasz Zok², Kacper Figurski², Katarzyna Pachulska-Wieczorek^{1*}, Marta Szachniuk^{1,2*}

1 Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, **2** Institute of Computing Science, Poznan University of Technology, Poznan, Poland

* kasiapw@ibch.poznan.pl (KPW); mszachniuk@cs.put.poznan.pl (MS)



OPEN ACCESS

Citation: Gumna J, Zok T, Figurski K, Pachulska-Wieczorek K, Szachniuk M (2020) RNAuthor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis. PLoS ONE 15(10): e0239287. <https://doi.org/10.1371/journal.pone.0239287>

Editor: Danny Barash, Ben-Gurion University, ISRAEL

Received: April 7, 2020

Accepted: September 3, 2020

Published: October 1, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0239287>

Copyright: © 2020 Gumna et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Abstract

RNAs adopt specific structures to perform their functions, which are critical to fundamental cellular processes. For decades, these structures have been determined and modeled with strong support from computational methods. Still, the accuracy of the latter ones depends on the availability of experimental data, for example, chemical probing information that can define pseudo-energy constraints for RNA folding algorithms. At the same time, diverse computational tools have been developed to facilitate analysis and visualization of data from RNA structure probing experiments followed by capillary electrophoresis or next-generation sequencing. RNAuthor, a new software tool for the fully automated normalization of SHAPE and DMS probing data resolved by capillary electrophoresis, has recently joined this collection. RNAuthor automatically identifies unreliable probing data. It normalizes the reactivity information to a uniform scale and uses it in the RNA secondary structure prediction. Our web server also provides tools for fast and easy RNA probing data visualization and statistical analysis that facilitates the comparison of multiple data sets. RNAuthor is freely available at <http://rnathor.cs.put.poznan.pl/>.

Introduction

Structural features are of importance for the biological functions of RNA molecules. Specific RNA structures are recognized by RNA binding proteins, ligands, and other RNAs—these interactions impact almost every aspect of cell life or viral replication. Therefore, there is a great interest in developing novel approaches for proper and rapid RNA structure modeling. The computational methods enable the obtaining of good quality models of short RNAs based on sequence only, but the accuracy of structure prediction decreases with the length of RNA molecules [1–5]. The inclusion of RNA structure probing data as pseudo-energy constraints into the thermodynamic folding algorithms significantly improves the accuracy of RNA structure prediction [6, 7]. Among chemical and enzymatic methods, SHAPE (selective 2'-hydroxyl

Funding: This work was funded by the National Science Centre Poland (<https://ncn.gov.pl/?language=en>) in the form of grants awarded to MS (2016/23/B/ST6/03931, 2019/35/B/ST6/03074) and KPW (2016/22/E/NZ3/00426). Funding for the open access charge was provided by the statutory funds of Poznan University of Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

acylation analyzed by primer extension) [8] and DMS (dimethyl sulfate) mapping [9] are the best validated and most widely used techniques of RNA structure probing *in vitro* and *in vivo* [10, 11]. Besides, the pipelines of SHAPE and DMS probing data incorporation into RNA structure prediction software are well established [6, 12]. DMS modifies the Watson–Crick edge of unpaired adenosines or cytosines, whereas SHAPE reagents create covalent adducts at the 2'-OH group on the RNA sugar ring in a flexibility-sensitive manner [8, 9, 13]. Several SHAPE reagents that differ in their half-life and solubility have been developed until now [14–16]. They act independently from nitrogen base and, consequently, one probing reagent can be used instead of a combination of base-specific chemicals.

Effective detection and quantitative measurement of modification sites are critical for all RNA probing experiments. Typically, RNA chemical modification is followed by a reverse transcription to cDNA that is truncated or mutated at the adducts position [8, 17]. The sites of RT stops in the cDNA can be read-out using the capillary electrophoresis (CE) or next-generation sequencing (NGS) but only the second technique can be used for the detection of adduct-induced mutations [17, 18]. The NGS-based techniques allow genome-wide and transcriptome-wide profiling of RNA structure. The CE is widely used for resolving reactivity data from medium- and low-throughput RNA probing experiments. There are many examples of SHAPE-CE usage for analysis of the structure of many important RNAs, including ribosomal RNAs [19, 20], long noncoding RNAs [21–23], viral RNAs [24–30], and retrotransposon RNAs [31–33]. Besides, CE can also be used for the analysis of RNA probing experiments utilizing other chemical reagents such as CMCT, kethoxal, hydroxyl radicals, and RNases [34–37].

The extraction of quantitative data from CE electropherograms is challenging and requires complicated, multistep analysis of fluorescence signals. Several computational tools can process electropherograms from SHAPE-CE experiments [38–42]. Among them, ShapeFinder [41] and QuShape [42] are the most widely used and yield high-quality SHAPE reactivity data for 300–600 nucleotides in one experiment. Before the incorporation of probing information into the thermodynamic RNA folding algorithms, the reactivity values must be normalized to a uniform scale that is valid for diverse RNAs. Additionally, visual inspection of nonspecific RT strong-stops (non-induced by adduct formation) is required.

Normalization and other quality control steps are very important aspects of structure probing data analysis. Therefore, we developed RNAtbor, a user-friendly tool for fast, automatic normalization, and analysis of the CE-based RNA probing data (Fig 1). Features of our tool include (i) normalization of data from several experiments in the box-plot scheme at once, (ii) automatic detection of strong-stops of reverse transcriptase, (iii) reactivity data visualization, (iv) statistical analysis of the results to compare multiple data sets, and (v) RNA secondary structure prediction based on reactivity data.

Materials and methods

RNAtbor workflow

In the RNAtbor workflow, we distinguish five general stages: validation of the input data (ShapeFinder or QuShape file(s) and optionally RNA sequence), exclusion of unreliable data, normalization of probing data, prediction of the secondary structure (optional), and statistical analysis of the normalized data (optional) (Fig 1).

Validation of the input data. Initially, the user-uploaded files, resulting from ShapeFinder or QuShape, are parsed, and the basic validation of their format is executed. If, additionally, a sequence is entered, RNAtbor checks whether it is RNA and whether it is at least as long as the sequence in the input file(s). A positive validation results in the next step of the

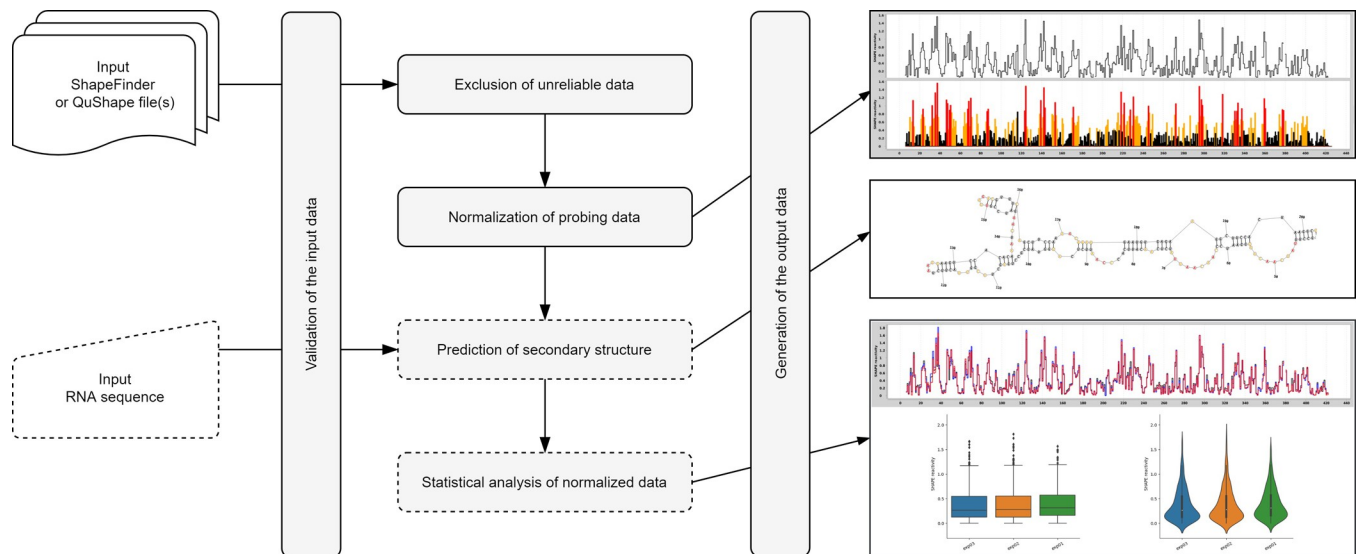


Fig 1. Workflow in the RNAthor system.

<https://doi.org/10.1371/journal.pone.0239287.g001>

computational process. Otherwise, the user receives an error message and is asked to provide correct data.

Exclusion of unreliable data. Unreliable data usually correspond to premature terminations of primer extension reaction due to reasons other than the formation of the adduct (e.g., preexisting cleavage or modification in RNA). These nucleotide positions are called RT strong-stops. RNAthor offers two ways of detecting such data and excluding them from further processing: a fully automated algorithm and an interactive procedure requiring manual selections. The automated procedure (Fig 2) was implemented based on our experience with analyzing data from RNA chemical probing experiments *in vitro* and *in vivo*. It was optimized for the analysis of SHAPE and DMS probing data. It eliminates the data, which meet one of the following criteria: the absolute reactivity value is negative; the background peak area is at least five times larger than the average background peak area; the difference in peak areas between background and reaction is less than 35% of the average background peak area, and the background peak area in this position is equal to or greater than this average. The alternative is a manual procedure, recommended especially for processing the data from RNA probing experiments other than SHAPE or DMS-probing. In this approach, users can identify unreliable data according to their own experience. They define the negative reactivity threshold, and indicate how to treat the negative reactivity values—they can be left as negative values, changed to 0, or marked as no data. RNAthor displays the histogram with peak areas for modification reaction and background for each nucleotide. Based on this view, users manually select RT strong-stops' positions. All identified RT strong-stops are next excluded from the normalization step.

Data normalization. In this step, the data are brought into proportion with one another, and outliers are removed, to provide users with easy to interpret reactivity data on a uniform scale. RNAthor applies the standard box-plot scheme, recommended to normalize the SHAPE-CE data [43, 44]. The normalization process involves: identifying outliers, determining the effective maximum reactivity, and calculating the normalized reactivity values. The initial task is to determine the first (Q1) and the third (Q3) quartile, the interquartile range (IQR), and compute the upper extreme: $UP = Q3 + 1.5(IQR)$. Reactivities greater than UP are considered outliers and not taken into account in subsequent calculations following the principle: for

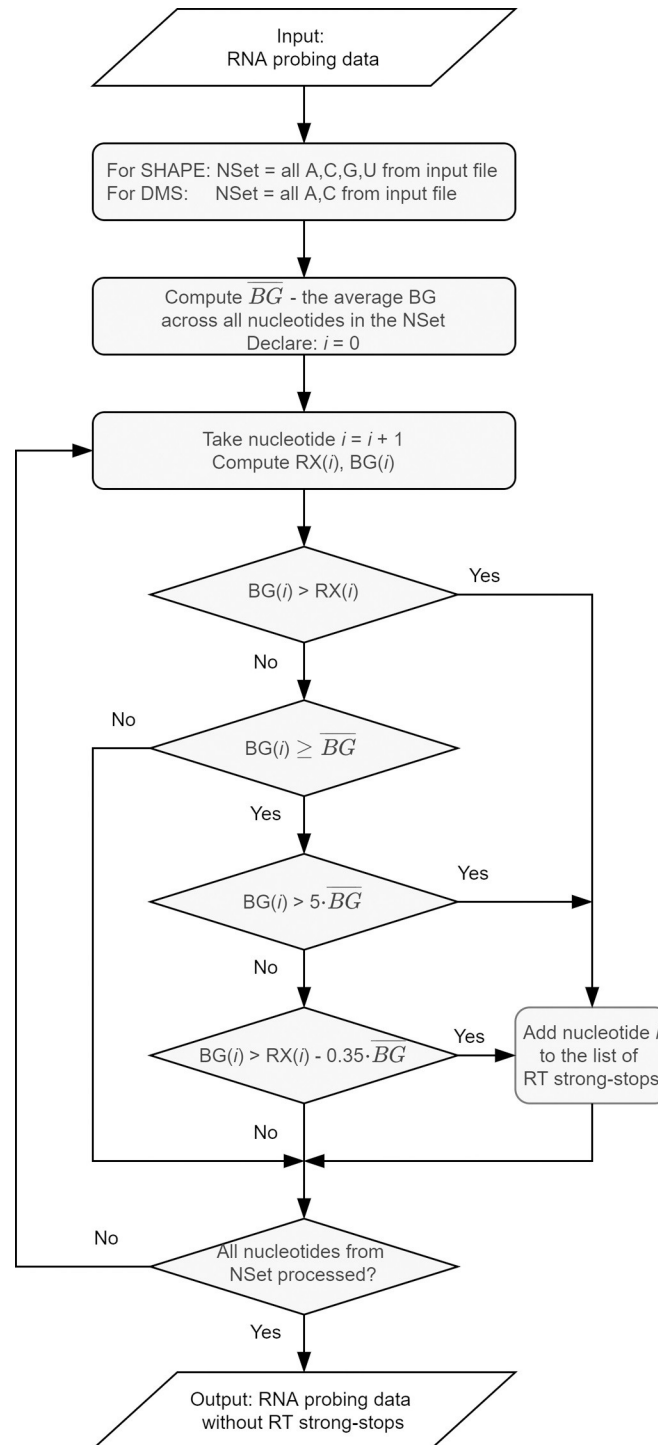


Fig 2. Scheme of the RNAthor algorithm for unreliable data exclusion.

<https://doi.org/10.1371/journal.pone.0239287.g002>

RNAs longer than 100 nucleotides, no more than 10% of the data are identified as outliers; for shorter RNAs, maximum 5% of data are removed. The remaining values are used to compute the effective maximum reactivity, i.e., the average of the top 8% of reactivity values. Finally, all absolute reactivity values are divided by the effective maximum reactivity. It results in

obtaining the normalized reactivity data on a uniform scale. Values close to 0 indicate no reactivity (and highly constrained nucleotides), while values greater than 0.85 correspond to high reactivity (and flexible nucleotides).

Secondary structure prediction. Optionally, users can obtain the secondary structure predicted for the RNA sequence provided at the input. If the sequence is given, RNAtbor automatically executes the incorporated RNAstructure algorithm [45] that supports SHAPE / DMS data-driven prediction. It takes the RNA sequence and the normalized probing data and generates the respective secondary structure. The graphical diagram of the structure is colored according to the color scheme defined for the default reactivity ranges. The output structure is also encoded in the dot-bracket notation.

Statistical analysis of the normalized data. Logged users can perform additional statistical analysis of the normalized probing data. The analysis includes 2–5 experiments selected by the user. It consists of running the Shapiro-Wilk test for normal data distribution, Bartlett test of variance homogeneity, non-parametric Mann-Whitney test (if the user selected 2 experiments), and Kruskal-Wallis rank-sum test (if the user selected 3–5 experiments). Two latter tests are performed if the probing data departure from the normal distribution. As a result of the analysis, users receive numerical, textual, and graphical data—among others, the comparative step plot, the box-and-whisker plot, and the violin plot.

Experimental setup

RNA probing data for the RNAtbor validation were obtained from SHAPE-CE and DMS-CE experiments performed in our laboratory for Ty1 RNA (+1–560). The results of SHAPE-based manual analysis were already published in [32]. The DMS experiment was performed especially for this work; its details are presented below. Electropherograms from SHAPE and DMS probing were processed using ShapeFinder software according to the authors' instructions [41].

For RNA probing with DMS, RNA (8 pmol) was refolded in 30 μ l of renaturation buffer (10 mM Tris-HCl pH 8.0, 100 mM KCl and 0.1 mM EDTA) by heating for 3 minutes at 95°C, slow cooling to 4°C, then adding 90 μ l of water and 30 μ l of 5x folding buffer (final concentration: 40 mM Tris-HCl pH 8.0, 130 mM KCl, 0.2 mM EDTA, 5 mM MgCl₂), followed by incubation for 20 minutes at 37°C. The RNA sample was divided into two tubes and treated with DMS dissolved in ethanol (+) or ethanol alone (-), and incubated at RT for 1 minute. The reaction was quenched by the addition of 14.7 M β -mercaptoethanol. RNA was recovered by ethanol precipitation and resuspended in 10 μ l of water. Primer extension reactions were performed using fluorescently labeled primer [Cy5 (+) and Cy5.5 (-)] as described previously [32]. Sequencing ladders were prepared using primers labeled with WellRed D2 (ddA) and LicorIRD-800 (ddT) and a Thermo Sequenase Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's protocol. Samples were analyzed on a GenomeLab GeXP Analysis System (Beckman-Coulter).

Web application

RNAtbor, implemented as a publicly available web server, has a simple and intuitive interface. It runs on all major web browsers and is accessible at <http://rnatbor.cs.put.poznan.pl/>. The web service is hosted and maintained by the Institute of Computing Science, Poznan University of Technology, Poland.

Implementation details

The architecture of RNAtbor comprises two components: the computational engine (backend layer) and the web application (frontend layer). The backend layer, implemented in Java

OpenJDK 8.0, applies selected modules of the Spring Framework: Spring Boot 2.1.6 enables fast configuration of the application; Spring Security ensures user authentication and basic security; Spring MVC allows compatibility with the Model View Controller and Apache Tomcat server; Spring Test, via Junit and Mockito libraries, enables unit tests and integration; Spring Data allows for comprehensive database services, including transaction management. The user interface (frontend layer) is implemented in Angular technology. User data and basic information about the experiments are collected in the PostgreSQL relational database, the input and output data are saved on the server's hard drive. The tool uses the Apache License 2.0.

Input and output description

At the input, RNAtbor accepts ShapeFinder or QuShape output files in a tab-delimited text format. Users upload their data via the *New experiment* page by selecting 1–15 files from the local folder. All files in the multiple-input should come from several repetitions of the RNA probing experiment performed for the same RNA. Repetitions increase the reliability of structural data for RNA secondary structure prediction. RNAtbor processes all the input files in a single run. It starts after data uploading and setting additional parameters for the normalization process (algorithm for RT strong-stops detection, probing reagent, color settings). Additionally, users can provide an RNA sequence that is used to predict the RNA secondary structure.

RNAtbor generates a selection of output data. First of all, users obtain the output file in the SHAPE format (*.shape) that is compatible with the RNAstructure software [45]. The file comprises two columns with nucleotide positions and normalized SHAPE / DMS-CE reactivity data. For the multiple-input, the generated SHAPE file contains averaged reactivities from all normalized data. Nucleotides for which there is no reactivity data are assigned -999 values as recommended in [43]. If the user uploaded the sequence of the analyzed RNA molecule, RNAtbor provides the RNA secondary structure in dot-bracket notation and the graphical diagram. Additionally, RNAtbor generates files that can facilitate the analysis of RNA probing experiments. One of them is the MS Excel file with spreadsheets containing the input data, the normalized reactivity values, and averaged normalized reactivity data with standard deviation (the average and standard deviation are calculated separately for each nucleotide across samples). Each spreadsheet with the input data contains a histogram, identical to this created during manual removal of RT strong-stops. Rows with normalized reactivity values are colored depending on the user's settings. In the processing of large RNAs, this file can help to combine probing data from overlapping reads (with a different set of primers). RNAtbor also prepares a graphical output: step plot and bar plot presenting a reactivity profile for one experiment or averaged data from several repetitions. The bar plot is colored depending on the settings: black for reactivities in [0, 0.4), orange for reactivities in [0.4, 0.85), and red for reactivities in [0.85, ∞) by default. Logged users that run statistical analysis of experimental data also obtain comparative step plot, box-and-whisker plot, violin plot, and summary of test results. The latter one, available for download in .txt file, informs whether the uploaded data come from a normal distribution, whether they have equal variances, what statistical test was performed, and what is the p-value. The comparative step plot shows reactivity profiles of all compared experiments in one chart. The box-and-whisker plot displays the distribution of data based on the position measures, such as quartiles, minimum, and maximum. The violin plot presents the shape of the distribution and probability density of normalized reactivity values. All generated plots can be saved in PNG or EPS format. Users download the output files separately or in a single zipped archive. They can also obtain them as an email attachment—if the email was provided at the input. Additionally, the email contains a unique link to the result page. The results are

stored in the system for 3 days (for guest users) or 3 months (for logged users). Logged users can extend the storage time by an additional month.

Results

RNAtbor allows for efficient, automated processing and analysis of RNA probing data from SHAPE-CE and DMS-CE experiments and their use in data-driven RNA secondary structure prediction. It was tested on multiple datasets, containing data from SHAPE and DMS probing experiments resolved by capillary electrophoresis. The tests confirmed the reliability of the results and showed the utility of the tool. Here, we describe the experiments performed to compare the results of RNA probing data analysis carried out manually by an expert, and automatically by RNAtbor. For the experiments, we chose SHAPE-CE and DMS-CE probing data obtained for RNA of yeast retrotransposon Ty1. The structure of the 5'-end of Ty1 RNA was extensively studied and determined under different experimental conditions and biological states [31, 32, 46].

In the first test, we executed RNAtbor for the ShapeFinder-generated files containing the probing data obtained from three independent replicates of SHAPE experiment (raw data used in this experiment are provided in the [S1 File](#)). We ran RNAtbor with the default settings and the automated algorithm for the identification of RT strong-stops. The generated normalized reactivity data were next compared to the corresponding data published in [32], resulting from manual analysis of the same input. We aligned the obtained bar plots ([Fig 3A](#)), and we computed the correlation between normalized reactivity values ([Fig 3B](#)). In the second test, we repeated the same procedure for data obtained from the DMS experiment (unpublished data; the experiment was carried out for this work especially). “Blind” human experimentalist analyzed the DMS data preprocessed using ShapeFinder, normalized reactivity values, manually identified unreliable data and applied OriginPro to generate the bar plot presenting the reactivity profile. The results of this manual processing were compared to the output generated by RNAtbor that was executed with DMS reagent selected and automated identification of RT strong-stops ([Fig 3C and 3D](#)).

From these experiments, we observe that all RT strong-stops identified manually by the expert are also selected for exclusion by the automatic algorithm implemented in RNAtbor. On the other hand, few data assigned as RT strong-stops by RNAtbor can be considered reliable in the human-dependent analysis. This is due to the rigid criteria for determining RT strong-stops adopted in the algorithm. [Table 1](#) presents the results of the detailed analysis we did by comparing manual, expert-driven, and automatic, RNAtbor-performed detection of unreliable data. We computed basic measures used to evaluate the quality of binary classification: true positives (TP)—data classified as reliable by both expert and RNAtbor, true negatives (TN)—data classified as unreliable by both expert and RNAtbor, false positives (FP)—data indicated as unreliable by the expert but classified as reliable by RNAtbor, false negatives (FN)—data indicated as reliable by the expert but classified as unreliable by RNAtbor. Using these measures, we calculated the accuracy (ACC), sensitivity (TPR, true positive rate), specificity (TNR, true negative rate), and precision (PPV, positive predictive value) of the automatic algorithm implemented in RNAtbor. All these measures were determined for three datasets: SHAPE probing data separately, DMS probing data separately, and data from both sets together. They prove the high quality of the tested algorithm for all datasets. Accuracy and sensitivity equal 0.99, where accuracy, $ACC = (TP+TN)/(TP+TN+FP+FN)$, represents the ratio of correct classifications to the total number of input data, and sensitivity, $TPR = TP/(TP+FN)$, indicates what part of the actual reliable data has been correctly classified by RNAtbor. Specificity and precision are both equal to 1, which is because of $FP = 0$. Specificity, $TNR = TN/(TN$

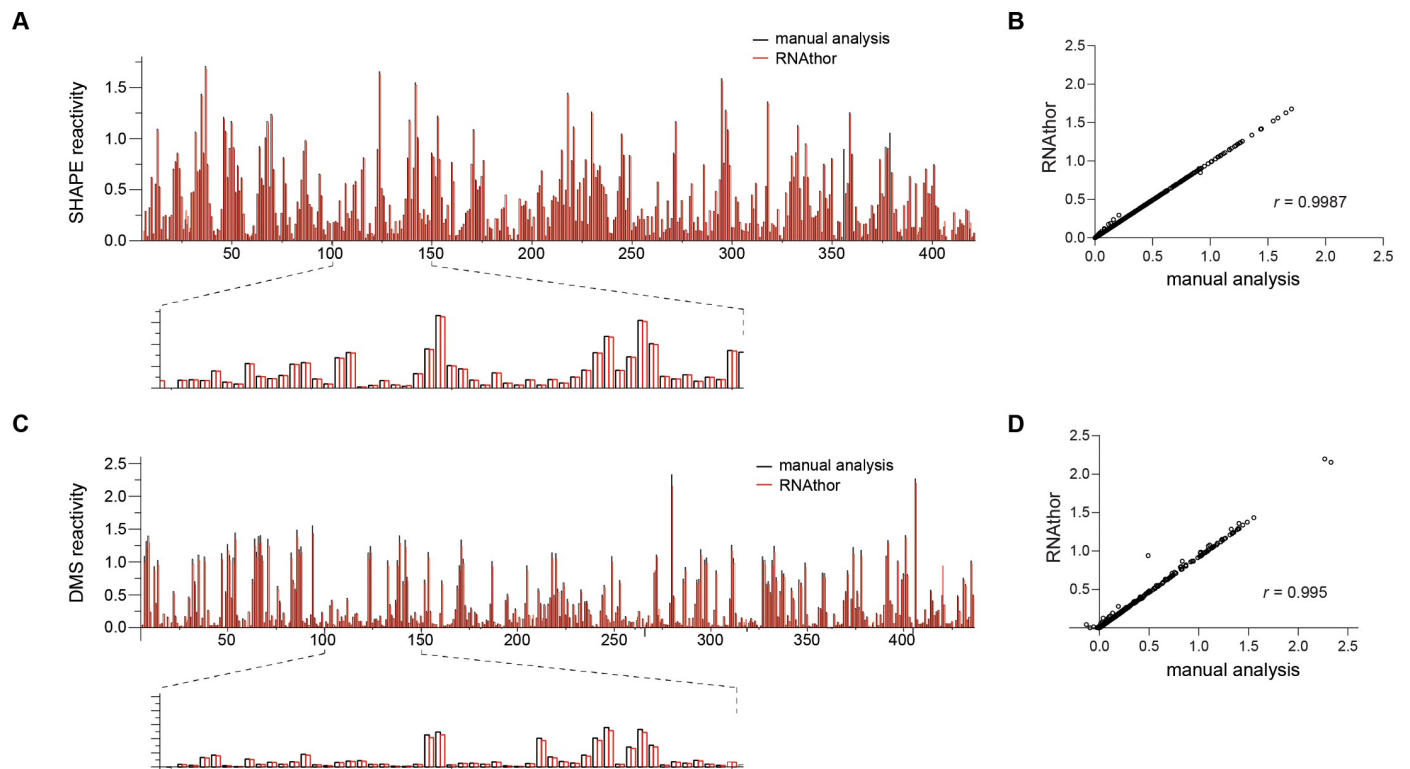


Fig 3. Automatic and manual normalization of RNA probing data. SHAPE (A) and DMS (C) reactivity profiles calculated by RNAtor (red) and manually (black). Correlation between RNAtor and manual analysis per nucleotide reactivity estimated for SHAPE (B) and DMS (D).

<https://doi.org/10.1371/journal.pone.0239287.g003>

+FP) is a fraction of correctly classified unreliable data, while the precision, $PPV = TP / (TP + FP)$, informs about the fraction of unreliable data classified as reliable. Finally, the experiments show that—despite some differences between expert- and RNAtor-driven analysis—the normalized RNA probing reactivity values obtained in both approaches are highly similar. The comparison of the reactivity profiles indicates the conformity of manual and automatic procedures. The averaged results from three independent probing experiments yield a Spearman correlation coefficient equal to 0.9987 for SHAPE and 0.995 for DMS-based analysis (Fig 3).

In the testing phase, we also executed statistical analysis to verify the repeatability of obtained results for each nucleotide between the replicates, and compare the reactivity profiles. Fig 4 shows an example of such verification for selected DMS-CE experiments previously performed by RNAtor (raw data used in these experiments are provided in the S1 File). Experiments 1 and 2 (denoted as DMSexp1 and DMSexp2 in Fig 4) were performed under identical experimental conditions, while the higher concentration of DMS was used in experiment 3 (denoted as DMSexp3). We observed a high similarity between reactivity profiles generated for experiments 1 and 2 (Fig 4A), whereas a significant difference was visible for experiment 3

Table 1. The results of validation of RNAtor algorithm for unreliable data identification.

dataset	TP	TN	FP	FN	ACC	PPV	TPR	TNR
SHAPE	2462	38	0	32	0.99	1	0.99	1
DMS	2431	35	0	25	0.99	1	0.99	1
ALL	4893	73	0	57	0.99	1	0.99	1

<https://doi.org/10.1371/journal.pone.0239287.t001>

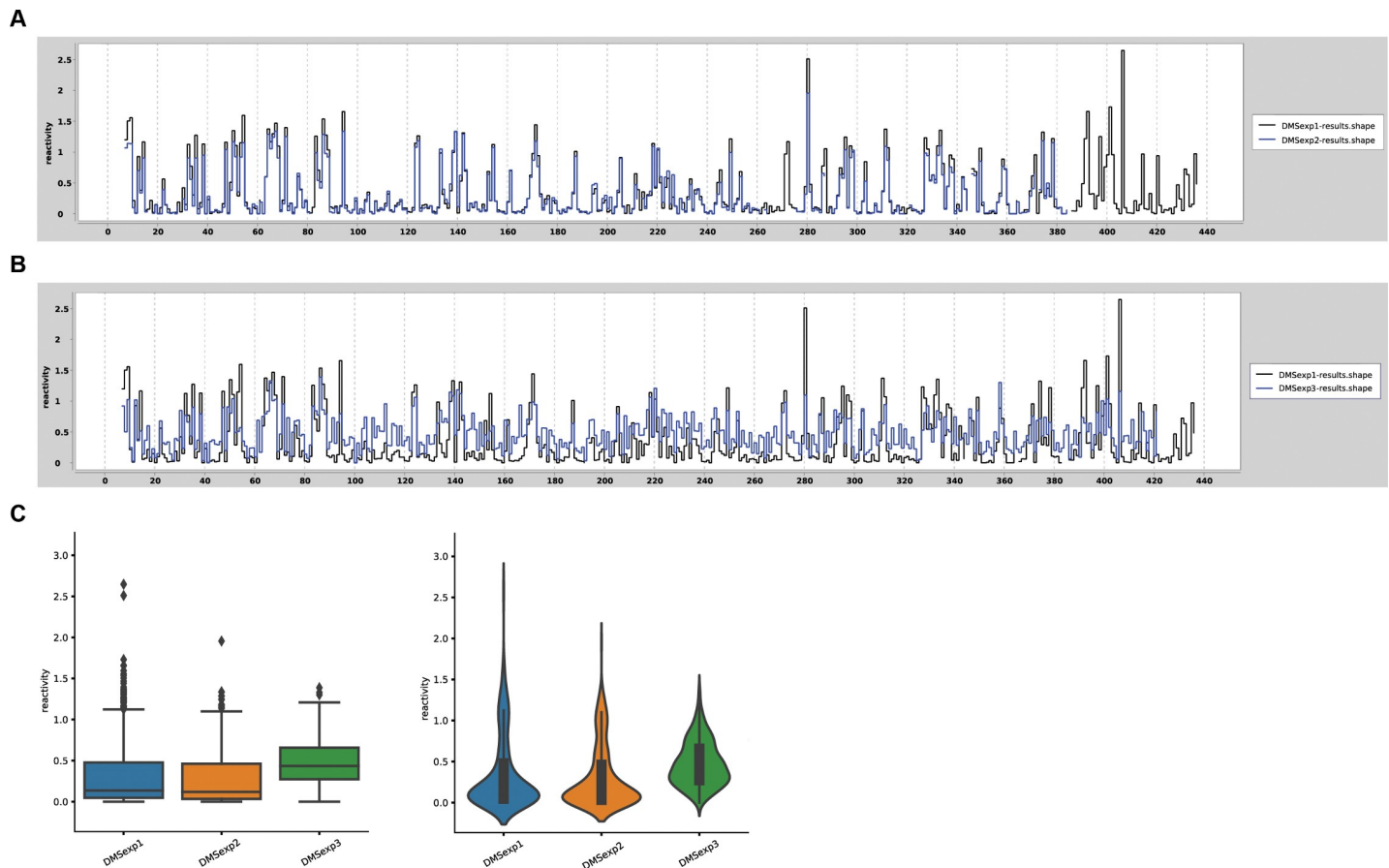


Fig 4. Example verification of the repeatability of RNA chemical probing experiments. Comparative step plot for repeatable (A) and non-repeatable (B) replicates of Ty1 RNA probing with DMS. (C) The box-and-whisker plot and violin plot presenting the differences in reactivity data distribution obtained for repeatable and non-repeatable experiments.

<https://doi.org/10.1371/journal.pone.0239287.g004>

(Fig 4B). As expected, the box plot and violin plot present the comparable DMS data distributions for experiments 1 and 2 (Fig 4C). The statistical plots for experiment 3 clearly show a significant increase in the number of more reactive nucleotides, and a concurrent decrease of unreactive nucleotides, consequently, the overall median reactivity is higher (Fig 4C). From these examples, we can see that additional options of RNAtor can be used for fast and easy comparative and statistical analysis for RNA chemical probing experiments.

Conclusions

In this work, we presented RNAtor, the new computational tool dedicated to the study of RNA structures that enriched the set of web-interfaced bioinformatics systems available within the RNapolis project [47]. RNAtor was designed for a fully automatic, quick normalization, and analysis of SHAPE / DMS-CE data. Although several programs can process the results of CE-based RNA probing, so far, no automatic procedure could identify unreliable data, and this step of the analysis was usually done manually. RNAtor incorporates the algorithm for the automatic exclusion of RT strong-stops to minimize user involvement in the probing data analysis. The tool can be applied to analyze data from other RNA probing methods if capillary electrophoresis and ShapeFinder or QuShape were used for data collection. RNAtor also visualizes the results of RNA probing data normalization, runs the data-driven prediction of

RNA secondary structure, and performs the statistical tests. The latter option facilitates the comparative study of multiple probing experiments, allows to assess the compatibility between experiments, and compare whole data sets of RNAs probed in different experimental conditions (e.g., *in vitro*, *in vivo*, *ex vivo*, *in virio*, *ex virio*), or in the absence or presence of protein/ligand. Compared to manual or semi-automated data processing, RNAtbor significantly reduces the time needed for data analysis; thus, it can highly improve the study and interpretation of data obtained from RNA chemical probing experiments.

In the future, we plan to extend the functionality of RNAtbor by implementing procedures combining RNA probing data from overlapping CE reads to facilitate the structural analysis of large RNAs.

Supporting information

S1 File.
(PDF)

Author Contributions

Conceptualization: Julita Gumna.

Data curation: Julita Gumna, Katarzyna Pachulska-Wieczorek.

Funding acquisition: Marta Szachniuk.

Methodology: Tomasz Zok.

Project administration: Katarzyna Pachulska-Wieczorek, Marta Szachniuk.

Software: Tomasz Zok, Kacper Figurski.

Supervision: Marta Szachniuk.

Validation: Julita Gumna, Tomasz Zok, Katarzyna Pachulska-Wieczorek, Marta Szachniuk.

Visualization: Kacper Figurski.

Writing – original draft: Julita Gumna.

Writing – review & editing: Julita Gumna, Katarzyna Pachulska-Wieczorek, Marta Szachniuk.

References

1. Waterman MS, Smith TF (1978) Rna Secondary Structure—Complete Mathematical-Analysis. *Mathematical Biosciences* 42: 257–266.
2. Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* 5: 105. <https://doi.org/10.1186/1471-2105-5-105> PMID: 15296519
3. Popena M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, et al. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40: e112. <https://doi.org/10.1093/nar/gks339> PMID: 22539264
4. Biesiada M, Pachulska-Wieczorek K, Adamiak RW, Purzycka KJ (2016) RNAComposer and RNA 3D structure prediction for nanotechnology. *Methods* 103: 120–127. <https://doi.org/10.1016/j.ymeth.2016.03.010> PMID: 27016145
5. Lukasiak P, Antczak M, Ratajczak T, Szachniuk M, Popena M, et al. (2015) RNAssess—a web server for quality assessment of RNA 3D structures. *Nucleic Acids Res* 43: W502–506. <https://doi.org/10.1093/nar/gkv557> PMID: 26068469

6. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, et al. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* 101: 7287–7292. <https://doi.org/10.1073/pnas.0401799101> PMID: 15123812
7. Ge P, Zhang S (2015) Computational analysis of RNA structures with chemical probing data. *Methods* 79–80: 60–66.
8. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127: 4223–4231. <https://doi.org/10.1021/ja043822v> PMID: 15783204
9. Inoue T, Cech TR (1985) Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci U S A* 82: 648–652. <https://doi.org/10.1073/pnas.82.3.648> PMID: 2579378
10. Spitale RC, Flynn RA, Torre EA, Kool ET, Chang HY (2014) RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip Rev RNA* 5: 867–881. <https://doi.org/10.1002/wrna.1253> PMID: 25132067
11. Mitchell D 3rd, Assmann SM, Bevilacqua PC (2019) Probing RNA structure in vivo. *Curr Opin Struct Biol* 59: 151–158. <https://doi.org/10.1016/j.sbi.2019.07.008> PMID: 31521910
12. Sloma MF, Mathews DH (2015) Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol* 553: 91–114. <https://doi.org/10.1016/bs.mie.2014.10.053> PMID: 25726462
13. McGinnis JL, Dunkle JA, Cate JH, Weeks KM (2012) The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 134: 6617–6624. <https://doi.org/10.1021/ja2104075> PMID: 22475022
14. Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129: 4144–4145. <https://doi.org/10.1021/ja0704028> PMID: 17367143
15. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, et al. (2013) RNA SHAPE analysis in living cells. *Nat Chem Biol* 9: 18–20. <https://doi.org/10.1038/nchembio.1131> PMID: 23178934
16. Busan S, Weidmann CA, Sengupta A, Weeks KM (2019) Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* 58: 2655–2664. <https://doi.org/10.1021/acs.biochem.8b01218> PMID: 31117385
17. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11: 959–965. <https://doi.org/10.1038/nmeth.3029> PMID: 25028896
18. Mailler E, Paillart JC, Marquet R, Smyth RP, Vivet-Boudou V (2019) The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip Rev RNA* 10: e1518. <https://doi.org/10.1002/wrna.1518> PMID: 30485688
19. Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* 106: 97–102. <https://doi.org/10.1073/pnas.0806929106> PMID: 19109441
20. Abeysirigunawardena SC, Kim H, Lai J, Rangunathan K, Rappe MC, et al. (2017) Evolution of protein-coupled RNA dynamics during hierarchical assembly of ribosomal complexes. *Nat Commun* 8: 492. <https://doi.org/10.1038/s41467-017-00536-1> PMID: 28887451
21. Garcia GR, Goodale BC, Wiley MW, La Du JK, Hendrix DA, et al. (2017) In Vivo Characterization of an AHR-Dependent Long Noncoding RNA Required for Proper Sox9b Expression. *Mol Pharmacol* 91: 609–619. <https://doi.org/10.1124/mol.117.108233> PMID: 28385905
22. Liu F, Somarowthu S, Pyle AM (2017) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat Chem Biol* 13: 282–289. <https://doi.org/10.1038/nchembio.2272> PMID: 28068310
23. Owens MC, Clark SC, Yankey A, Somarowthu S (2019) Identifying Structural Domains and Conserved Regions in the Long Non-Coding RNA lncTCF7. *Int J Mol Sci* 20.
24. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6: e96. <https://doi.org/10.1371/journal.pbio.0060096> PMID: 18447581
25. Soszynska-Jozwiak M, Michalak P, Moss WN, Kierzek R, Keszy J, et al. (2017) Influenza virus segment 5 (+)RNA—secondary structure and new targets for antiviral strategies. *Sci Rep* 7: 15041. <https://doi.org/10.1038/s41598-017-15317-5> PMID: 29118447
26. Michalak P, Soszynska-Jozwiak M, Biala E, Moss WN, Keszy J, et al. (2019) Secondary structure of the segment 5 genomic RNA of influenza A virus and its application for designing antisense oligonucleotides. *Sci Rep* 9: 3801. <https://doi.org/10.1038/s41598-019-40443-7> PMID: 30846846
27. Mahmud B, Horn CM, Tappich WE (2019) Structure of the 5' Untranslated Region of Enteroviral Genomic RNA. *J Virol* 93.

28. Newburn LR, White KA (2020) A trans-activator-like structure in RCNMV RNA1 evokes the origin of the trans-activator in RNA2. *PLoS Pathog* 16: e1008271. <https://doi.org/10.1371/journal.ppat.1008271> PMID: 31905231
29. Khoury G, Mackenzie C, Ayadi L, Lewin SR, Branlant C, et al. (2020) Tat IRES modulator of tat mRNA (TIM-TAM): a conserved RNA structure that controls Tat expression and acts as a switch for HIV productive and latent infection. *Nucleic Acids Res* 48: 2643–2660. <https://doi.org/10.1093/nar/gkz1181> PMID: 31875221
30. Kendall C, Khalid H, Muller M, Banda DH, Kohl A, et al. (2019) Structural and phenotypic analysis of Chikungunya virus RNA replication elements. *Nucleic Acids Res* 47: 9296–9312. <https://doi.org/10.1093/nar/gkz640> PMID: 31350895
31. Purzycka KJ, Legiewicz M, Matsuda E, Eizenstat LD, Lusvarghi S, et al. (2013) Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res* 41: 463–473. <https://doi.org/10.1093/nar/gks983> PMID: 23093595
32. Gumna J, Purzycka KJ, Ahn HW, Garfinkel DJ, Pachulska-Wieczorek K (2019) Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biol* 16: 1749–1763. <https://doi.org/10.1080/15476286.2019.1657370> PMID: 31469343
33. Blaszczyk L, Biesiada M, Saha A, Garfinkel DJ, Purzycka KJ (2017) Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression. *Viruses* 9.
34. McGinnis JL, Duncan CD, Weeks KM (2009) High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol* 468: 67–89. [https://doi.org/10.1016/S0076-6879\(09\)68004-6](https://doi.org/10.1016/S0076-6879(09)68004-6) PMID: 20946765
35. Novikova IV, Hennelly SP, Sanbonmatsu KY (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* 40: 5034–5051. <https://doi.org/10.1093/nar/gks071> PMID: 22362738
36. Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, Saha A, Gumna J, et al. (2015) Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic Acids Res* 43: 7414–7431. <https://doi.org/10.1093/nar/gkv695> PMID: 26160887
37. Pachulska-Wieczorek K, Blaszczyk L, Biesiada M, Adamiak RW, Purzycka KJ (2016) The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag. *Retrovirology* 13: 18. <https://doi.org/10.1186/s12977-016-0245-1> PMID: 26987314
38. Pang PS, Elazar M, Pham EA, Glenn JS (2011) Simplified RNA secondary structure mapping by automation of SHAPE data analysis. *Nucleic Acids Res* 39: e151. <https://doi.org/10.1093/nar/gkr773> PMID: 21965531
39. Yoon S, Kim J, Hum J, Kim H, Park S, et al. (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* 27: 1798–1805. <https://doi.org/10.1093/bioinformatics/btr277> PMID: 21561922
40. Cantara WA, Hatterschide J, Wu W, Musier-Forsyth K (2017) RiboCAT: a new capillary electrophoresis data analysis tool for nucleic acid probing. *RNA* 23: 240–249. <https://doi.org/10.1261/ma.058404.116> PMID: 27821510
41. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* 14: 1979–1990. <https://doi.org/10.1261/ma.1166808> PMID: 18772246
42. Karabiber F, McGinnis JL, Favorov OV, Weeks KM (2013) QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* 19: 63–73. <https://doi.org/10.1261/ma.036327.112> PMID: 23188808
43. Low JT, Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods* 52: 150–158. <https://doi.org/10.1016/j.ymeth.2010.06.007> PMID: 20554050
44. Lusvarghi S, Sztuba-Solinska J, Purzycka KJ, Rausch JW, Le Grice SF (2013) RNA secondary structure prediction using high-throughput SHAPE. *J Vis Exp*: e50243. <https://doi.org/10.3791/50243> PMID: 23748604
45. Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11: 129. <https://doi.org/10.1186/1471-2105-11-129> PMID: 20230624
46. Huang Q, Purzycka KJ, Lusvarghi S, Li D, LeGrice SF, et al. (2013) Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA* 19: 320–332. <https://doi.org/10.1261/ma.035535.112> PMID: 23329695
47. Szachniuk M (2019) RNApolis: Computational Platform for RNA Structure Analysis. *Foundations of Computing and Decision Sciences* 44: 241–257.

Załącznik nr 2 do rozprawy doktorskiej „Identyfikacja i charakterystyka elementów strukturalnych genomowego RNA retrotranspozonu Ty1 oraz jego oddziaływań z białkiem Gag kluczowych dla dimeryzacji i pakowania genomu do cząstek wirusopodobnych”

Dodatek S1 do pracy naukowej J. Gumna, T. Żok, K. Figurski, K. Pachulska-Wieczorek, M. Szachniuk; *RNAthor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis*; PLoS One 2020, zawierający dane z eksperymentów próbkowania strukturalnego RNA mTy1 metodą SHAPE-CE oraz DMS-CE i przeanalizowane w programie ShapeFinder znajduje się pod linkiem:

<https://doi.org/10.1371/journal.pone.0239287.s001>

Załącznik nr 3 do rozprawy doktorskiej „Identyfikacja i charakterystyka elementów strukturalnych genomowego RNA retrotranspozonu Ty1 oraz jego oddziaływań z białkiem Gag kluczowych dla dimeryzacji i pakowania genomu do cząstek wirusopodobnych”

OŚWIADCZENIA OKREŚLAJĄCE WKŁAD KANDYDATKI
W POWSTANIE PRAC NAUKOWYCH WCHODZĄCYCH
W SKŁAD ROZPRAWY DOKTORSKIEJ

Poznań, dnia 30.09.2021r.

mgr Julita Gumna

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE KANDYDATA DO NADANIA
STOPNIA NAUKOWEGO DOKTORA

Tytuł pracy naukowej: Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA

Autorzy: Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel, Katarzyna Pachulska-Wieczorek[§]


Czasopismo: RNA Biology

Rok wydania: 2019

Oświadczam, że mój wkład w powstanie wyżej wymienionej pracy naukowej polegał na:

- zaplanowaniu wraz z dr hab. Katarzyną Pachulską-Wieczorek wszystkich eksperymentów *in vitro*, a następnie samodzielnym ich wykonaniu,
- zaplanowaniu wraz z dr hab. Katarzyną Pachulską-Wieczorek oraz prof. David J. Garfinkel (University of Georgia, USA) testów retrotranspozycji Ty1 *in vivo*, samodzielnym przygotowaniu potrzebnych do tego konstruktów genetycznych oraz wykonaniu eksperymentów pilotażowych,
- otrzymaniu rekombinowanego białka Gag Ty1,
- analizie surowych danych otrzymanych z wykonanych przeze mnie eksperymentów,
- wykonaniu analiz statystycznych,
- udziale w analizie i interpretacji wyników badań,
- przygotowaniu rysunków 1-4, 6-8, tabeli 1 oraz *Supplementary data 1 i 2*,
- przygotowaniu pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek rozdziału *Materials and Methods*, opisu otrzymanych wyników, pierwszej wersji rozdziału *Introduction* oraz udziale w późniejszej edycji manuskryptu.

Rozdział „*Gag binding sites on mini Ty1 RNA*” został napisany na podstawie wyników eksperymentów, które wykonałam w ramach przygotowania mojej pracy magisterskiej pt.: „*Aspekty strukturalne oddziaływania RNA Ty1 z białkiem Gag: implikacje dla procesu retrotranspozycji*”.


Julita Gumna

[§] autor do korespondencji

Poznań, dnia 30.09.2021r.

dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE AUTORA DO KORESPONDENCJI
PRACY NAUKOWEJ ZAWARTEJ W ROZPRAWIE DOKTORSKIEJ

Tytuł pracy naukowej: Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA

Autorzy: Julita Gumna, Katarzyna J. Purzycka, Hyo Won Ahn, David J. Garfinkel, Katarzyna Pachulska-Wieczorek[§]


Czasopismo: RNA Biology

Rok wydania: 2019

Oświadczam, że wkład mgr Julity Gumnej w powstanie wyżej wymienionej pracy naukowej polegał na:

- współudziale w zaplanowaniu wszystkich eksperymentów *in vitro* i samodzielnym ich wykonaniu,
- udziale w planowaniu testów retrotranspozycji Ty1 *in vivo*, a następnie samodzielnym przygotowaniu potrzebnych do tego konstruktów genetycznych oraz wykonaniu eksperymentów pilotażowych,
- otrzymaniu rekombinowanego białka Gag Ty1,
- analizie surowych danych eksperymentalnych,
- wykonaniu analiz statystycznych,
- udziale w analizie i interpretacji wyników badań,
- przygotowaniu rysunków 1-4, 6-8, tabeli 1 oraz *Supplementary data 1 i 2*,
- przygotowaniu pod moim kierunkiem rozdziału *Materials and Methods*, opisu otrzymanych wyników, pierwszej wersji rozdziału *Introduction* oraz udziale w późniejszej edycji manuskryptu.

Rozdział „*Gag binding sites on mini Ty1 RNA*” został opracowany na podstawie wyników eksperymentów, które mgr Julita Gumna wykonała w ramach przygotowania pracy magisterskiej pt.: „*Aspekty strukturalne oddziaływania RNA Ty1 z białkiem Gag: implikacje dla procesu retrotranspozycji*”.


Katarzyna Pachulska-Wieczorek

[§] autor do korespondencji

Poznań, dnia 30.09.2021r.

mgr Julita Gumna

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE KANDYDATA DO NADANIA
STOPNIA NAUKOWEGO DOKTORA

Tytuł pracy naukowej: RNA binding properties of the Ty1 LTR-retrotransposon Gag protein

Autorzy: Julita Gumna, Angelika Andrzejewska-Romanowska, David J. Garfinkel and Katarzyna Pachulska-Wieczorek[§]

Czasopismo: International Journal of Molecular Sciences

Rok wydania: 2021

Oświadczam, że mój wkład w powstanie wyżej wymienionej pracy naukowej polegał na:

- opracowaniu planu badań pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek,
- otrzymaniu rekombinowanych białek Gag Ty1,
- wykonaniu eksperymentów MST i próbkowania kompleksów RNA/białko z wykorzystaniem rodników hydroksylowych,
- wykonaniu wraz z mgr inż. Angeliką Andrzejewską-Romanowską eksperymentów EMSA,
- analizie surowych danych otrzymanych ze wszystkich eksperymentów,
- wykonaniu analizy statystycznej,
- analizie i interpretacji wyników badań pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek oraz prof. David J. Garfinkel (University of Georgia, USA),
- przygotowaniu wszystkich rysunków, tabel oraz *Supplementary data 1*,
- przygotowaniu pierwszej wersji manuskryptu oraz jego wersji finalnej we współpracy z dr hab. Katarzyną Pachulską-Wieczorek i prof. David J. Garfinkel.


Julita Gumna

[§] autor do korespondencji

Poznań, dnia 30.09.2021r.

dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE AUTORA DO KORESPONDENCJI
PRACY NAUKOWEJ ZAWARTEJ W ROZPRAWIE DOKTORSKIEJ

Tytuł pracy naukowej: RNA binding properties of the Ty1 LTR-retrotransposon Gag protein

Autorzy: Julita Gumna, Angelika Andrzejewska-Romanowska, David J. Garfinkel and Katarzyna Pachulska-Wieczorek[§]

Czasopismo: International Journal of Molecular Sciences

Rok wydania: 2021

Oświadczam, że wkład mgr Julity Gumnej w powstanie wyżej wymienionej pracy naukowej polegał na:

- opracowaniu planu badań pod moim kierunkiem,
- otrzymaniu rekombinowanych białek Gag Ty1,
- wykonaniu eksperymentów MST i próbkowania kompleksów RNA/białko z wykorzystaniem rodników hydroksylowych,
- wykonaniu wraz z mgr inż. Angeliką Andrzejewską-Romanowską eksperymentów EMSA,
- analizie surowych danych otrzymanych ze wszystkich eksperymentów oraz wykonaniu analizy statystycznej,
- analizie i interpretacji wyników badań pod moim kierunkiem i prof. D.J. Garfinkel (University of Georgia, USA),
- przygotowaniu wszystkich rysunków, tabel oraz *Supplementary data 1*,
- przygotowaniu pierwszej wersji manuskryptu oraz jego wersji finalnej pod kierunkiem moim i prof. D.J. Garfinkel.



Katarzyna Pachulska-Wieczorek

[§] autor do korespondencji

Poznań, dnia 30.09.2021r.

mgr Julita Gumna
Zakład Struktury i Funkcji Retrotranspozonów
Instytut Chemii Bioorganicznej PAN
ul. Noskowskiego 12/14
61-704 Poznań

OŚWIADCZENIE KANDYDATA DO NADANIA
STOPNIA NAUKOWEGO DOKTORA

Tytuł pracy naukowej: RNAtThor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis

Autorzy: Julita Gumna, Tomasz Żok, Kacper Figurski, Katarzyna Pachulska-Wieczorek[§], Marta Szachniuk[§]

Czasopismo: PLoS One

Rok wydania: 2020

Oświadczam, że mój wkład w powstanie wyżej wymienionej pracy naukowej polegał na:

- zainicjowaniu projektu utworzenia narzędzia RNAtThor,
- zaprojektowaniu metody identyfikacji danych niewiarygodnych,
- sporządzeniu wraz z dr hab. Katarzyną Pachulską-Wieczorek specyfikacji dla narzędzia RNAtThor,
- przygotowaniu danych eksperymentalnych wykorzystanych do walidacji,
- testowaniu narzędzia RNAtThor,
- przygotowaniu rysunków 3 i 4 oraz pierwszej wersji rysunku 2,
- przygotowaniu pierwszej wersji manuskryptu oraz udziale w jego późniejszej edycji.


Julita Gumna

Poznań, dnia 30.09.2021r.

dr hab. Katarzyna Pachulska-Wieczorek, prof. ICHB PAN

Zakład Struktury i Funkcji Retrotranspozonów

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

prof. dr hab. inż. Marta Szachniuk

Zakład Bioinformatyki Strukturalnej

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE AUTORÓW DO KORESPONDENCJI
PRACY NAUKOWEJ ZAWARTEJ W ROZPRAWIE DOKTORSKIEJ

Tytuł pracy naukowej: RNAuthor – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis

Autorzy: Julita Gumna, Tomasz Żok, Kacper Figurski, Katarzyna Pachulska-Wieczorek[§], Marta Szachniuk[§]

Czasopismo: PLoS One

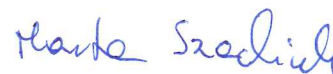
Rok wydania: 2020

Oświadczamy, że wkład mgr Julity Gumnej w powstanie wyżej wymienionej pracy naukowej polegał na:

- zainicjowaniu projektu oraz opracowaniu procedury identyfikacji danych niewiarygodnych,
- opracowaniu wraz z dr hab. Katarzyną Pachulską-Wieczorek specyfikacji dla narzędzia RNAuthor,
- przygotowaniu przykładowych danych eksperymentalnych i testowaniu narzędzia RNAuthor,
- przygotowaniu rysunków 3 i 4 oraz wstępnej wersji rysunku 2,
- przygotowaniu pierwszej wersji manuskryptu oraz udziale w jego edycji.



Katarzyna Pachulska-Wieczorek



Marta Szachniuk