

Instytut Chemii Bioorganicznej

Polskiej Akademii Nauk



Rozprawa doktorska pt.

**Charakterystyka dynamiki strukturalnej i funkcjonalnej
genomów RNA aktywnych retrotranspozonów LTR**

*Characterization of the structural and functional dynamics of active LTR
retrotransposon RNA genomes*

mgr inż. Angelika Andrzejewska-Romanowska

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

Zakład Struktury i Funkcji RNA

Poznań, 2024

*Chciałabym serdecznie podziękować wszystkim osobom,
które przyczyniły się do powstania tej rozprawy:*

*Mojej Promotor, **dr hab. Katarzynie Pachulskiej-Wieczorek, prof. ICHB PAN,**
za opiekę naukową, przekazaną wiedzę i doświadczenie,
a także za niesamowite wsparcie i poświęcony czas*

***Koleżankom i Kolegom** z Zakładu Struktury i Funkcji RNA, ICHB PAN, i nie tylko,
szczególnie *dr Julicie Gumnej, dr Małgorzacie Zawadzkiej, mgr inż. Ewie Tykwińskiej,
dr inż. Pawłowi Śledzińskiemu, dr Leszkowi Błaszczykowi i dr Grzegorzowi Framskiemu,*
za wspólne lata pracy, inspirujące rozmowy i codzienną życzliwość*

***Pani Małgorzacie Dąbkiewicz,**
za wszelką pomoc oraz wiele ciepłych słów*

***Moim Rodzicom, Bratu i Najbliższym,**
za wiarę we mnie i troskę*

*A w szczególności **Mojemu Mężowi, Piotrkowi,**
za nieocenione wsparcie i obecność każdego dnia, dodawanie mi sił,
oraz za wyrozumiałość.*

Z całego serca Dziękuję!

SPIS TREŚCI

1	FINANSOWANIE.....	4
2	WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ.....	5
3	STRESZCZENIE.....	6
4	ABSTRACT.....	7
5	WPROWADZENIE.....	8
5.1	Retrotranspozony LTR.....	8
5.1.1	Retrotranspozony Ty1/Copia i Ty3/Gypsy.....	8
5.1.2	Organizacja i cykl replikacyjny drożdżowych retrotranspozonów Ty1 i Ty3.....	9
5.1.3	Sekwencje regulatorowe w gRNA drożdżowych retrotranspozonów Ty1 i Ty3.....	11
5.2	Badanie struktury drugorzędowej RNA <i>in vivo</i>	12
5.2.1	Mapowanie chemiczne struktury drugorzędowej RNA <i>in vivo</i>	14
5.2.2	Detekcja modyfikacji SHAPE.....	16
5.2.3	Analiza bioinformatyczna danych z eksperymentów SHAPE.....	18
5.2.4	Analiza i interpretacja reaktywności SHAPE.....	19
5.2.5	Predykcja oraz interpretacja modeli struktury drugorzędowej RNA.....	20
5.2.6	Aktualne wyzwania w badaniach struktury drugorzędowej RNA.....	22
5.3	Struktura gRNA wirusów i retrotranspozonów oraz RNA drożdży <i>in vivo</i>	24
6	CEL NAUKOWY PRZEDSTAWIONEGO CYKLU PRAC.....	26
7	KRÓTKIE OMÓWIENIE PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ.....	27
7.1	On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective.....	27
7.2	<i>In vivo</i> structure of the Ty1 retrotransposon RNA genome.....	28
7.3	Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome.....	31
8	PODSUMOWANIE.....	35
9	WYKAZ SKRÓTÓW.....	36
10	BIBLIOGRAFIA.....	37

ZAŁĄCZNIKI:

Załącznik nr 1 ŻYCIORYS NAUKOWY

Załącznik nr 2 PRACE NAUKOWE WCHODZĄCE W SKŁAD ROZPRAWY
DOKTORSKIEJ

Załącznik nr 3 OŚWIADCZENIA OKREŚLAJĄCE WKŁAD KANDYDATKI W
POWSTANIE PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY
DOKTORSKIEJ

1 FINANSOWANIE

Niniejsza praca powstała przy wsparciu finansowym udzielonym przez Narodowe Centrum Nauki:

- Grant Sonata Bis (2016/22/E/NZ3/00426)
- Grant Opus (2020/39/B/NZ3/03020)
- Grant Preludium (2021/41/N/NZ3/04060)

Moja aktywność naukowa była również wspierana przez:

- Stypendium START Fundacji na Rzecz Nauki Polskiej (2022)
- Stypendium L'Oreal-UNESCO dla Kobiet i Nauki (2023)
- Stypendium Dyrektora ICHB PAN dla najlepszych doktorantów (lata 2019/2020, 2021/2022 oraz 2022/2023)

2 WYKAZ PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

Rozprawę doktorską stanowi cykl powiązanych tematycznie publikacji wymienionych poniżej. Wszystkie prace zostały opublikowane w czasopismach z listy JCR. Sumaryczny współczynnik wpływu (IF, ang. *impact factor*)¹ przedstawionego cyklu prac naukowych wynosi 39,98, a liczba punktów MNiSW² – 540.

1. Andrzejewska, A.*, Zawadzka, M.*, Pachulska-Wieczorek, K.
On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective.
International Journal Of Molecular Sciences, **2020**, 21(18), 6770.
(IF₂₀₂₀ – 5,92; MNiSW₂₀₂₄ – 140)
2. Andrzejewska, A., Zawadzka, M., Gumna, J., Garfinkel, D. J., Pachulska-Wieczorek, K.
In vivo structure of the Ty1 retrotransposon RNA genome.
Nucleic Acids Research, **2021**, 49(5), 2878–2893.
(IF₂₀₂₁ – 19,16; MNiSW₂₀₂₄ – 200)
3. Andrzejewska-Romanowska, A., Gumna, J., Tykwińska, E., Pachulska-Wieczorek, K.
Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome.
Nucleic Acids Research, **2024**, doi: 10.1093/nar/gkae494.
(IF₂₀₂₂ – 14,9; MNiSW₂₀₂₄ – 200)

* równorzędny pierwszy autor

¹ IF na podstawie danych w bazie Journal Citation Reports na platformie Web of Science

² pkt MNiSW na podstawie ujednoliconego wykazu czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego z dnia 5 stycznia 2024

3 STRESZCZENIE

Retrotranspozony LTR (ang. *long terminal repeat*) replikują za pośrednictwem genomowego RNA (gRNA), który służy jako matryca do syntezy białek i odwrotnej transkrypcji. Podobnie do wirusów RNA zarówno sekwencja, jak i struktura gRNA retrotranspozonów niesie instrukcje niezbędne do replikacji. Rozwój metod chemicznego mapowania struktury RNA pozwala pozyskać coraz lepszą wiedzę na temat natywnej struktury cząsteczek RNA w środowisku komórkowym. Jest ona niezbędna do uzyskania kompleksowego obrazu wpływu architektury RNA na jego funkcję. Mimo rosnącej liczby modeli strukturalnych cząsteczek RNA, dane dotyczące struktury gRNA retrotranspozonów *in vivo* nie były dostępne.

Najważniejszym celem moich badań było określenie po raz pierwszy struktury gRNA aktywnych retrotranspozonów LTR podczas replikacji w komórkach. Moimi modelami badawczymi były retrotranspozony Ty1 (*Pseudoviridae*) oraz Ty3 (*Metaviridae*), naturalnie występujące w genomie *Saccharomyces cerevisiae*. Badania nad biologią drożdżowych elementów Ty są istotnym źródłem informacji na temat mechanizmów retrotranspozycji. Moją rozprawę doktorską tworzy cykl trzech powiązanych tematycznie publikacji. W pracy **Andrzejewska i in., NAR, 2021** wykorzystałam metodę SHAPE-CE do zbadania struktury gRNA retrotranspozonu Ty1 (5,7 kz) *in vivo*. Dokonałam również analizy porównawczej uzyskanego modelu z modelami gRNA Ty1 *in vitro* oraz *in virio*. Scharakteryzowałam kontekst strukturalny sekwencji funkcjonalnych zaangażowanych w oddziaływanie RNA-RNA istotne dla retrotranspozycji. Wskazałam, które oddziaływania mogą zachodzić przed pakowaniem gRNA do cząstek wirusopodobnych. Dodatkowo przeprowadziłam eksperymenty w warunkach zahamowanej translacji, pokazując wpływ rybosomów na destabilizację i reorganizację struktury gRNA Ty1. W publikacji **Andrzejewska-Romanowska i in., NAR, 2024** zastosowałam strategię SHAPE-MaP do ustalenia struktury gRNA Ty3 (5,1 kz) w warunkach *in vivo* oraz *ex vivo*. Zidentyfikowałam motywy strukturalne specyficzne dla danego stanu eksperymentalnego, a także stabilnie zwinięty rdzeń tworzący się niezależnie od środowiska komórkowego. Opisałam kontekst strukturalny znanych sekwencji funkcjonalnych i zaproponowałam nową sekwencję istotną dla dimeryzacji gRNA Ty3. Na koniec scharakteryzowałam kluczowe cechy strukturalne gRNA wspólne dla reprezentatywnych retroelementów (Ty1, Ty3 i HIV-1). Cykl publikacji wchodzących w skład rozprawy doktorskiej rozpoczyna praca przeglądowa **Andrzejewska i in., IJMS, 2020**, podsumowująca dostępną wiedzę na temat struktury mRNA *in vivo*, wynikającą z najnowszych badań w komórkach różnych organizmów. Skupiliśmy się na czynnikach regulujących strukturę RNA wewnątrz komórek oraz na związkach pomiędzy strukturą cząsteczek RNA a ich funkcją.

4 ABSTRACT

LTR (long terminal repeat)-retrotransposons replicate via genomic RNA (gRNA), which serves as a template for protein synthesis and reverse transcription. Similar to RNA viruses, both the sequence and structure of gRNA retrotransposons carry the instructions necessary for replication. The advancement of RNA structure chemical mapping methods helps us learn more about the natural structure of RNA molecules within cells. This understanding is crucial for gaining a comprehensive picture of how RNA's architecture influences its function. Despite the growing number of RNA structural models, data on the *in vivo* structure of retrotransposon gRNA have not been available.

The primary objective of my studies was to determine the gRNA structure of active LTR-retrotransposons during replication in cells for the first time. My research models were retrotransposons Ty1 (*Pseudoviridae*) and Ty3 (*Metaviridae*), naturally occurring in the *Saccharomyces cerevisiae* genome. The studies on yeast Ty element biology are essential sources of information about the mechanisms of retrotransposition. My doctoral dissertation comprises three thematically related publications. In the work by **Andrzejewska et al., NAR, 2021**, I used the SHAPE-CE method to examine the gRNA structure of the Ty1 retrotransposon (5.7 kb) *in vivo*. I also conducted a comparative analysis of the obtained model with the *in vitro* and *in virio* Ty1 gRNA models. I characterized the structural properties of functional sequences involved in RNA-RNA interactions significant for retrotransposition, indicating those that may occur before gRNA packaging into virus-like particles. Additionally, I performed experiments in yeast inhibited for translation, showing the influence of ribosomes on the destabilization and remodeling of the Ty1 gRNA structure. In the publication by **Andrzejewska-Romanowska et al., NAR, 2024**, I used the SHAPE-MaP strategy to determine the structure of gRNA Ty3 (5.1 kb) *in vivo* and *ex vivo*. I identified the state-specific structural motifs and a well-folded core forming independently from the cellular environment. I described the structural context of known functional sequences and proposed a novel Ty3 gRNA dimerization sequence. Finally, I characterized gRNA key structural features shared between representative retroelements (Ty1, Ty3, and HIV-1). The series of publications included in the doctoral dissertation begins with a review by **Andrzejewska et al., IJMS, 2020**, summarizing the available knowledge about mRNA structure *in vivo* resulting from the most recent studies in different organism cells. We focused on factors regulating the structure of RNA inside cells and the relationships between the structure of RNA molecules and their function.

5 WPROWADZENIE

5.1 Retrotranspozony LTR

Retrotranspozony LTR to grupa ruchomych elementów genetycznych (ang. *mobile genetic elements*), które replikują za pośrednictwem genomowego RNA (gRNA). Wyróżnia je obecność długich końcowych powtórzeń sekwencji (LTRs, ang. *long terminal repeats*), które bezpośrednio otaczają wewnętrzny rejon kodujący. Retrotranspozony LTR są szeroko rozpowszechnione w genomach eukariotycznych, odgrywając istotną rolę w ich funkcjonowaniu (Havecker i in., 2004). Wstawienie kopii retroelementu może zmieniać ekspresję sąsiednich genów, prowadzić do mutacji i różnych chorób genetycznych (Mita & Boeke, 2016). Ponadto ich aktywność może być źródłem przegrupowań genomowych i niestabilności chromatyny. Z drugiej strony retrotranspozony LTR mogą być korzystne dla swoich gospodarzy, zapewniając regulacyjne elementy genetyczne i napędzając ewolucję genomów. Duża część retrotranspozonów utraciła zdolność do mobilności, ale jednocześnie zachowały one aktywną, otwartą ramkę odczytu (ORF, ang. *open reading frame*). Dobrze znane są przykłady kooptacji takich sekwencji w genomach różnych organizmów i aktualnie kodują one białka wykorzystywane w różnych szlakach biologicznych komórek gospodarza. W genomie ludzkim duża grupa zaadaptowanych sekwencji pochodzi z genu *GAG* retrotranspozonów *Ty3/Gypsy* (Kaneko-Ishino & Ishino, 2012; Volff i in., 2001). Na przykład neuronalny gen *ARC* pochodzący z *Ty3 GAG* jest niezbędny do regulacji plastyczności synaptycznej, kluczowej dla procesów pamięciowych i jest powiązany z wieloma zaburzeniami neurologicznymi (Pastuzyn i in., 2018). Ponadto sekwencje retrotranspozonów LTR bogate w sygnały wzmacniacza, promotora lub poliadenylacji są często wykorzystywane do regulacji ekspresji genów komórkowych, promując adaptację gospodarza do różnych warunków stresowych (Curcio, 2019; Zawadzka & Pachulska-Wieczorek, 2019).

5.1.1 Retrotranspozony *Ty1/Copia* i *Ty3/Gypsy*

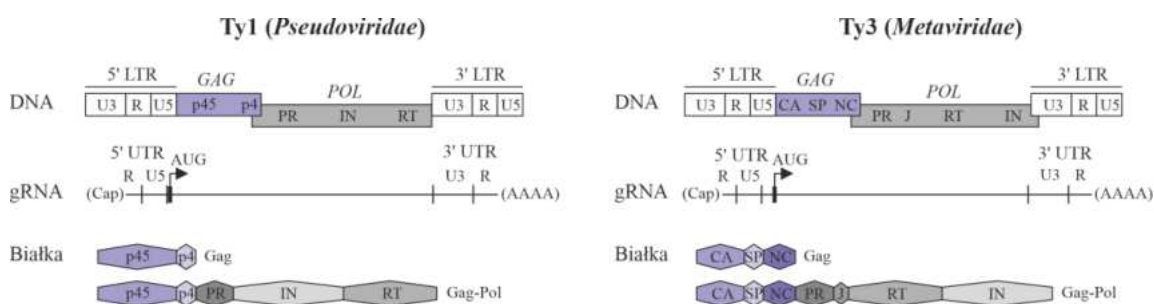
Ty1/Copia (*Pseudoviridae*) i *Ty3/Gypsy* (*Metaviridae*) to najważniejsze i najbardziej reprezentatywne rodziny retroelementów LTR u eukariotów. Należą one do rzędu *Ortervirales*, gromadzącego wirusy przeprowadzające odwrotną transkrypcję jednoniciowego RNA (ssRNA-RT), w tym członków rodziny *Retroviridae*, takich jak HIV-1 (Krupovic i in., 2018). *Ty3/Gypsy* i *Ty1/Copia* dzielą homologiczne białka oraz ważne cechy strukturalne i funkcjonalne z retrowirusami. Ponadto z uwagi na organizację genomu i podobieństwa w sekwencji białkowej, elementy *Ty3/Gypsy* uważane są za prekursorów retrowirusów (Llorens i in., 2008).

Znaczną reprezentację sekwencji retrotranspozonów *Ty3/Gypsy* i *Ty1/Copia* stwierdzono w genomach roślin, grzybów, zwierząt, alg i protistów. Obfitość sekwencji retrotranspozonów (ale nie ich typ) koreluje z rozmiarem genomu gospodarza. Około 85% sekwencji dużego genomu kukurydzy stanowią retrotranspozony, głównie typu LTR (Bonchev, 2016). W przypadku genomu

ludzkiego, retrotranspozony zajmują ~37% sekwencji i są to w większości elementy typu nie-LTR (LINE i SINE). Kompaktowy genom drożdży zawiera tylko niewielką ilość sekwencji retrotranspozonowych (1,3–3,4%), wszystkie są typu LTR (retrotranspozony Ty1–5). Ty1 jest najliczniejszym mobilnym elementem genetycznym w wielu szczepach *S. cerevisiae* (Curcio i in., 2015). Przykładowo referencyjny szczep *S. cerevisiae* BY4741 zawiera 32 kopie pełnej długości Ty1 na komórkę, ale tylko dwie pełne kopie elementu Ty3, z których tylko jedna, YGRWTy3-1, jest transpozycyjnie aktywna (S. Sandmeyer i in., 2015). Systemy drożdżowe Ty są szeroko wykorzystywane jako modele do zrozumienia, w jaki sposób retroelementy typu LTR replikują, kształtują genom gospodarza, oddziałują z czynnikami kodowanymi przez gospodarza oraz w jaki sposób ich aktywacja wpływa na ekspresję sąsiednich genów.

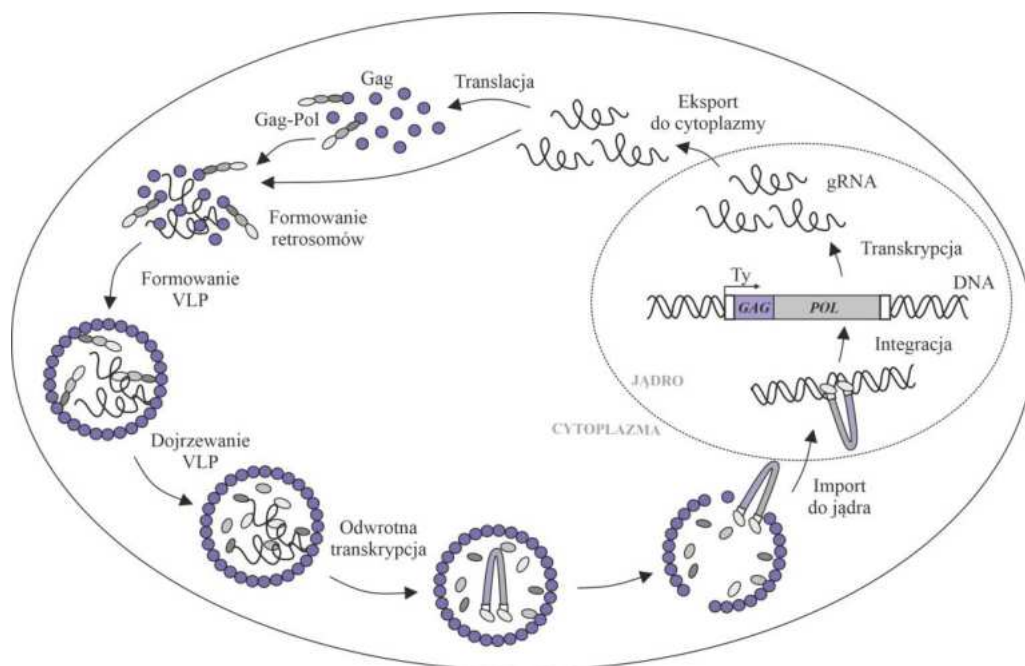
5.1.2 Organizacja i cykl replikacyjny drożdżowych retrotranspozonów Ty1 i Ty3

Pełna zintegrowana z genomem gospodarza kopia DNA Ty1 (~5,9 kbp) i Ty3 (~5,4 kbp) zawiera dwie nakładające się otwarte ramki odczytu, *GAG* i *POL*, flankowane długimi powtórzeniami końcowymi (Rysunek 1). ORF *GAG* koduje strukturalne białko Gag, funkcjonalnie homologiczne do poliproteiny Gag retrowirusów. Gag Ty3, w odróżnieniu od Gag Ty1, zawiera charakterystyczną dla retrowirusów domenę nukleokapsydową z motywem palca cynkowego (Orlinsky & Sandmeyer, 1994; S. B. Sandmeyer & Clemens, 2010). ORF *POL* koduje białka katalityczne wymagane do retrotranspozycji: proteazę (PR), odwrotną transkryptazę (RT) i integrazę (IN). ORF *POL* Ty3 jest zorganizowany podobnie do retrowirusów (PR-RT-IN), podczas gdy kolejność sekwencji w ORF *POL* Ty1 jest inna (PR-IN-RT). W przeciwieństwie do retrowirusów, retrotranspozony Ty przeważnie nie mają genów dla białka otoczki (Env), tym samym nie ma etapu pozakomórkowego w ich cyklu replikacji.



Rysunek 1. Organizacja DNA, RNA i białek drożdżowych elementów Ty1 oraz Ty3. Elementy Ty zawierają dwie nakładające się sekwencje ORF w rejonie centralnym: *GAG* i *POL*. Sekwencje LTR znajdują się na granicach i obejmują rejony U3, R i U5. gRNA zawiera rejony nieulegające translacji (UTR), czapkę (ang. *cap*) na końcu 5' oraz ogon poli(A) na końcu 3'. Produktami translacji są poliproteiny Gag i Gag-Pol. Gag Ty3 ulega cięciu potranslacyjnemu na białko kapsydu (CA), krótki peptyd SP i białko nukleokapsydu (NC). Gag Ty1 jest cięty na Gag-p45 i peptyd p4. Poliproteina Gag-Pol, będąca wynikiem przesunięcia ramki odczytu (ang. *frameshift*), jest cięta dodatkowo na proteazę (PR), odwrotną transkryptazę (RT) i integrazę (IN). W przypadku Ty3 powstaje także mały peptyd łącznikowy (J).

Podczas retrotranspozycji, zintegrowany z genomem komórki element Ty ulega transkrypcji od LTR do LTR przez polimerazę RNA II gospodarza, czego wynikiem jest gRNA o długości około 5,7 kb w przypadku Ty1 i 5,1 kb dla Ty3 (Rysunek 2). Po eksporcie z jądra do cytoplazmy pełni on rolę mRNA, stając się matrycą do translacji prekursorowych poliprotein Gag i Gag-Pol. Jednocześnie, część gRNA jest kierowana do cytoplazmatycznych ognisk, zwanych retrosomami, akumulujących białka i gRNA retroelementu, które następnie łączą się w cząstki wirusopodobne (VLPs, ang. *virus-like particles*) (Malagon & Jensen, 2011; S. B. Sandmeyer & Clemens, 2010). W VLPs białka prekursorowe dojrzewają poprzez cięcie proteolityczne. Dojrzałe VLPs zawierają białka strukturalne pochodzące z Gag, enzymy kodowane przez *POL*, gRNA w formie dimeru i niektóre czynniki komórkowe (np. $tRNA_i^{Met}$). Struktura VLP Ty3 wykazuje duże podobieństwo do dojrzałego kapsydu retrowirusa HIV-1 (Dodonova i in., 2019). Spakowany gRNA służy jako matryca do odwrotnej transkrypcji, która jest przeprowadzana przy użyciu odwrotnej transkryptazy elementu Ty i komórkowego $tRNA_i^{Met}$ jako startera (Keeney i in., 1995). Wytworzona zostaje w ten sposób dwuniciowa kopia DNA retrotranspozonu, która jest transportowana do jądra komórkowego i integruje się z genomem gospodarza. Miejscem integracji obu elementów jest rejon inicjacji transkrypcji genów transkrybowanych przez polimerazę RNA III (Bonnet & Lesage, 2021; Devine & Boeke, 1996; Qi i in., 2012).



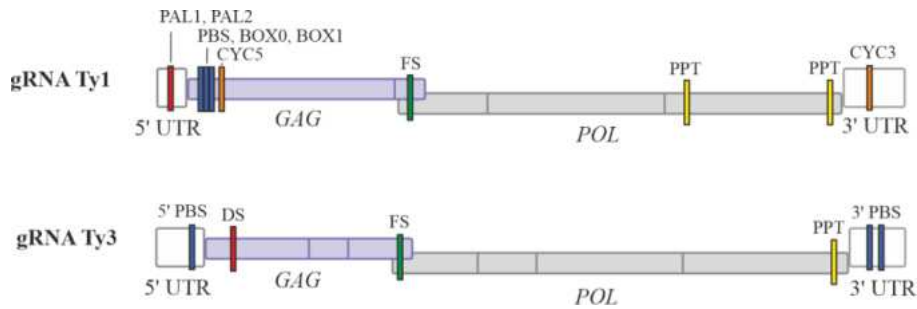
Rysunek 2. Cykl replikacyjny retrotranspozonów Ty.

Mimo że retrotranspozony Ty1 i Ty3 dzielą tego samego gospodarza i replikują w bardzo podobny sposób, to zajmują odrębne nisze podczas retrotranspozycji (Lesage & Todeschini, 2005). W komórkach haploidalnych obserwuje się wysoki poziom gRNA Ty1, a transkrypcja i retrotranspozycja Ty1 są aktywowane przez czynniki stresowe, takie jak warunki ograniczonego

wzrostu (niedobór glukozy, azotu lub węgla w pożywce). Poziom gRNA Ty3 jest ledwie wykrywalny w komórkach haploidalnych, ale jego transkrypcja i transpozycja są w dużym stopniu indukowane przez stymulację drożdży feromonem godowym (czynnik a/α , ang. *a/ α -factor*), wydzielanym przez komórki haploidalne podczas parowania (ang. *mating*). Feromony nie aktywują ekspresji Ty1 i silnie hamują jego retrotranspozycję na poziomie potranskrypcyjnym. Transkrypcja obu elementów jest natomiast hamowana w komórkach diploidalnych.

5.1.3 Sekwencje regulatorowe w gRNA drożdżowych retrotranspozonów Ty1 i Ty3

Wielofunkcyjne genomy RNA retrotranspozonów Ty1 i Ty3 wykazują wiele podobieństw, zarówno na poziomie organizacyjnym, jak i funkcjonalnym. Analogicznie do retrowirusów, końce 5' i 3' gRNA retrotranspozonów Ty1 i Ty3 są bogate w sekwencje regulatorowe bezpośrednio zaangażowane w odwrotną transkrypcję, cyklizację i dimeryzację gRNA (Rysunek 3). Obydwa elementy Ty zawierają kilkuczlönowe miejsca wiązania startera do odwrotnej transkrypcji (PBS, ang. *primer binding site*), jednak PBS Ty1 jest zlokalizowany w ORF *GAG*, podczas gdy części 5' i 3' PBS Ty3 znajdują się na przeciwległych końcach genomu (Gabus i in., 1998; Keeney i in., 1995). Ponadto zaproponowano, że w przypadku Ty3 przyłączanie $tRNA_i^{Met}$ do PBS pośredniczy w cyklizacji gRNA wymaganej do wydajnej inicjacji odwrotnej transkrypcji (Gabus i in., 1998). Natomiast w gRNA Ty1 znajdują się dodatkowe, komplementarne sekwencje cyklizacji, zlokalizowane na końcach 5' i 3' (CYC5 i CYC3) (Cristofari i in., 2002). Wczesne analizy sugerowały, że w odróżnieniu od retrowirusów, dimeryzacja gRNA elementów Ty nie zachodzi przez bezpośrednie oddziaływanie dwóch nici gRNA, lecz poprzez utworzenie dimeru $tRNA_i^{Met}$, który jednocześnie oddziałuje z rejonami PBS w dwóch cząsteczkach gRNA (Cristofari i in., 2000; Gabus i in., 1998). Nowsze badania wykazały jednak obecność sekwencji palindromowych zaangażowanych w bezpośrednią dimeryzację gRNA, zarówno dla Ty1, jak i Ty3 (Andrzejewska-Romanowska i in., 2024; Gumna i in., 2019; Purzycka i in., 2013). W gRNA elementów Ty znajduje się również specyficzna 7-nt sekwencja (FS, ang. *frameshift*), w której dochodzi do przesunięcia ramki odczytu na pozycję +1, koniecznego do translacji ORF *POL* (Belcourt & Farabaugh, 1990; Farabaugh i in., 1993). gRNA Ty1 i Ty3 zawierają także polipurynowe trakty (PPT), różniące się rozmiarem i sekwencją, które rozpoznają i wiążą RT, powodując inicjację syntezy drugiego łańcucha DNA (Dash i in., 2006; Lener i in., 2003). Struktura krystalograficzna RT Ty3 była pierwszą rozwiązana dla retrotranspozonu (Nowak i in., 2014). Pokazano, że choć ogólne konformacje RT Ty3 i HIV-1 są porównywalne, krytyczną różnicą jest, indukowane przyłączeniem substratu, utworzenie asymetrycznego homodimeru RT Ty3, w odróżnieniu od stabilnego niezależnie od substratu heterodimeru RT HIV-1.

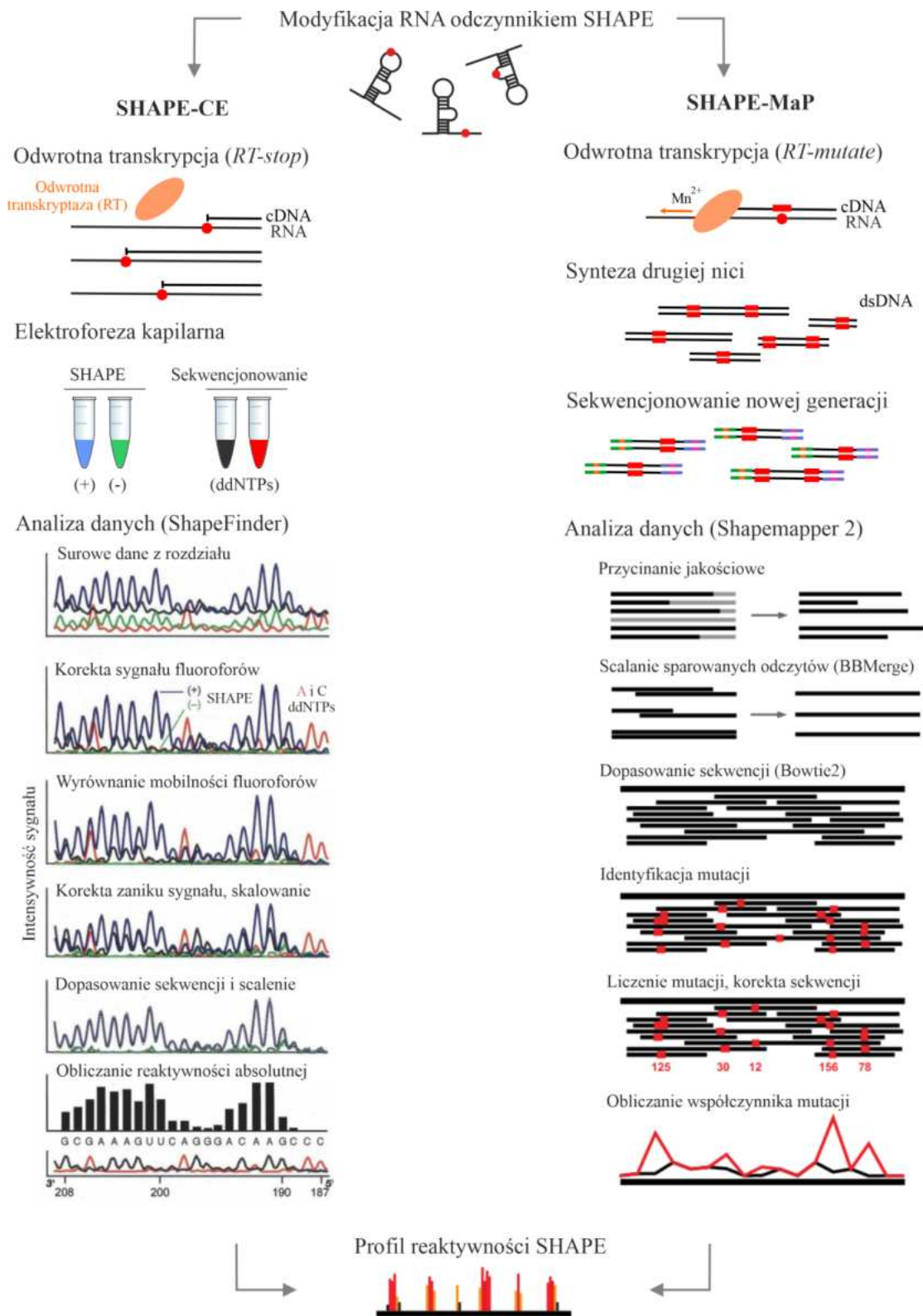


Rysunek 3. Sekwencje regulatorowe w gRNA elementów Ty1 i Ty3. PAL1, PAL2, DS – miejsca dimeryzacji gRNA; PBS, BOX0, BOX1 – miejsca wiązania tRNA^{Met}, jako startera do odwrotnej transkrypcji; CYC5, CYC3 – miejsca cyklizacji gRNA Ty1; FS – miejsce przesunięcia ramki odczytu; PPT – trakt polipurynowy.

5.2 Badanie struktury drugorzędowej RNA *in vivo*

Wiedza na temat natywnej struktury cząsteczek RNA w środowisku komórkowym jest niezbędna do kompleksowego zrozumienia roli architektury RNA w regulacji procesów biologicznych. Istnieje wiele metod badania struktury drugorzędowej RNA, które opierają się zarówno na podejściach eksperymentalnych, jak i na coraz bardziej zaawansowanych algorytmach do predykcji modeli strukturalnych i ich analizy bioinformatycznej (Strobel i in., 2018; J. Zhang i in., 2022). Przez długi czas techniki eksperymentalne pozwalały jedynie na badanie struktury RNA *in vitro*, a zrozumienie, w jaki sposób zachodzi zwijanie RNA *in vivo*, było ograniczone do badań w warunkach co najwyżej imitujących środowisko komórki (np. w lizatach komórkowych). Metodą, która umożliwiła badania struktury RNA w żywych komórkach, było chemiczne mapowanie RNA z użyciem małych elektrofilowych reagentów łatwo przenikających przez błony biologiczne, które na przestrzeni ostatniej dekady zrewolucjonizowało dziedzinę badań strukturalnych i funkcjonalnych RNA (Assmann i in., 2023; Kubota i in., 2015; Xu i in., 2022). Dane pochodzące z eksperymentów w komórkach lub *in vivo* dostarczają wielu przydatnych informacji na temat wpływu środowiska komórkowego na strukturę RNA. Tak kompleksowe zagadnienie najlepiej analizować w sposób porównawczy do danych pochodzących z RNA mapowanych w warunkach pozbawionych czynników komórkowych. Kompleksowa analiza dwóch różnych stanów eksperymentalnych, może wskazać motywy strukturalne RNA występujące zależnie lub niezależnie od środowiska komórki, wpływ maszyny translacyjnej i degradacyjnej, czy też potencjalne miejsca wiązania białek.

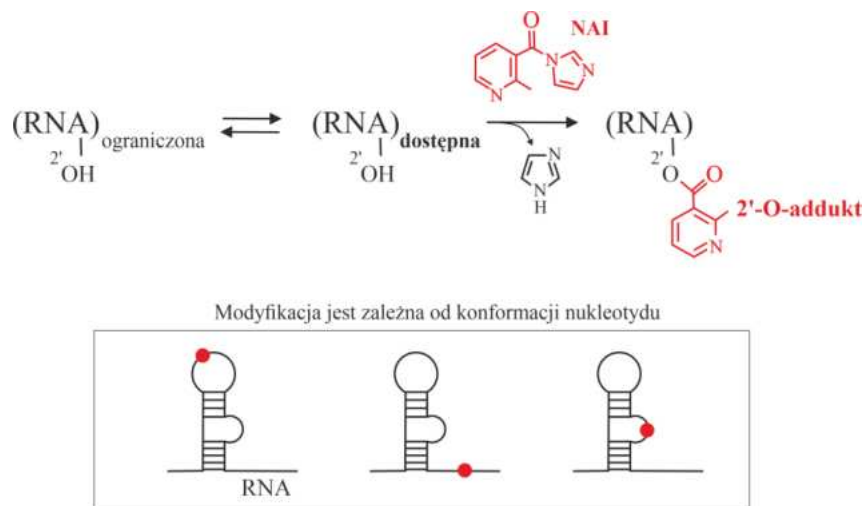
W pracach przedstawionych w rozprawie doktorskiej zastosowałam metody SHAPE-CE (ang. *selective 2' hydroxyl acylation analyzed by primer extension and capillary electrophoresis*) oraz SHAPE-MaP (ang. *SHAPE and mutational profiling*) do mapowania struktury RNA *in vivo*, *ex vivo* i *in vitro* (Rysunek 4), dlatego w kolejnych rozdziałach skupiłam się głównie na dokładniejszym wyjaśnieniu tych strategii oraz opisanie ich zalet i ograniczeń.



Rysunek 4. Schematyczna prezentacja zastosowanych strategii mapowania struktury drugorzędowej RNA – SHAPE-CE oraz SHAPE-MaP. Części schematu przedstawiające etapy analizy danych z wykorzystaniem programów ShapeFinder oraz Shapemapper 2 zostały zaczerpnięte z oryginalnych publikacji (Busan & Weeks, 2018; Vasa i in., 2008).

5.2.1 Mapowanie chemiczne struktury drugorzędowej RNA *in vivo*

Metoda mapowania chemicznego polega na wykorzystaniu małych specyficznych związków chemicznych, które selektywnie reagują z nukleotydami w cząsteczce RNA w zależności od ich konformacji. Aktualnie, jedną z najczęściej stosowanych i najlepiej zwalidowanych technik w tej grupie jest SHAPE (Merino i in., 2005). Technika ta wykorzystuje małe odczynniki elektrofilowe do pomiaru konformacji każdej reszty nukleotydowej w RNA poprzez tworzenie kowalencyjnych adduktów na dostępnej grupie 2'-OH rybozy, tym samym modyfikacja następuje niezależnie od typu zasady azotowej (Rysunek 5). Oznacza to, że nukleotydy z wysokim poziomem modyfikacji (wykazujące wysoki poziom reaktywności SHAPE) będą najprawdopodobniej niesparowane bądź strukturalnie dynamiczne w cząsteczce RNA, podczas gdy niski poziom modyfikacji (niska reaktywność SHAPE) wskaże nukleotydy, które są strukturalnie ograniczone lub zaangażowane w oddziaływania. Do tej pory opracowano szeroką gamę odczynników SHAPE, takich jak, np. bezwodnik kwasu 1-metylo-7-nitroizatowego (1M7), bezwodnik kwasu N-metyloizatowego (NMIA), imidazolid kwasu 2-metylonikotynowego (NAI), czy cyjanek benzoilu (BzCN), które różnią się rozpuszczalnością, czasem połowicznego rozpadu, jak również pewną swoistą preferencją dla poszczególnych nukleotydów (Busan i in., 2019). Wśród nich największą popularność zdobyły te ze zdolnością łatwego przenikania przez błony i ściany komórkowe, jak na przykład NAI, wydajnie modyfikujący cząsteczki RNA wewnątrz komórek (Spitale i in., 2013).



Rysunek 5. Schemat ilustrujący reakcję modyfikacji RNA odczynnikiem SHAPE.

Jednym z nowszych odczynników SHAPE jest imidazolid kwasu 2-aminopirydyno-3-karboksylowego (2A3), którego raportowana czułość jest nawet dwa razy większa niż NAI (Marinus i in., 2021). Jednak odczynnik ten nie zdobył jeszcze dużej popularności w badaniach *in vivo*, co może wynikać z faktu, że jego podwyższona aktywność została zaprezentowana tylko dla rybosomalnych RNA. Tym samym nie jest do końca jasne, czy 2A3 równie wydajnie modyfikuje transkrypty mniej liczne w komórce. Oprócz odczynników SHAPE, związkiem zdolnym do

modyfikacji RNA *in vivo*, jest siarczek dimetylu (DMS), który bezpośrednio oddziałuje z zasadą azotową rybonukleotydów, które nie są zaangażowane w kanoniczne pary zasad Watsona-Cricka (Tijerina i in., 2007). Jednak w przeciwieństwie do odczynników SHAPE, DMS dostarcza informacji o kanonicznych parach tylko reszt nukleotydowych A i C, gdyż wprowadza modyfikację na atomach N1 adenozyiny i N3 cytozyny. Ponadto DMS modyfikuje pozycję N7 guanozyny, jednak jej dostępność nie zależy od udziału w parowaniu zasad Watsona-Cricka. Istnieją doniesienia wskazujące, że w podwyższonym pH roztworu (~8), DMS jest w stanie modyfikować także pozycje N1 guanozyny i N3 uracylu w sposób zależny od ich zaangażowania w kanoniczne pary zasad, jednak z dużo mniejszą wydajnością (Mitchell i in., 2023; Mustoe i in., 2019). Najnowszym odczynnikiem zaproponowanym do mapowania struktury RNA *in vivo* jest jodometan 1-etylo-3-(3-dimetyloaminopropyl) karbodiimidu (ETC), który modyfikując niesparowane reszty G i U, dostarcza informacji uzupełniających dla modyfikacji DMS (Douds i in., 2024).

Mapowanie chemiczne RNA przeprowadza się zwyczajowo w komórkach w logarytmicznej fazie wzrostu. Do zawiesiny komórek dodaje się ściśle określoną ilość odczynnika mapującego, rozpuszczonego najczęściej w DMSO (bezwodny dimetylosulfotlenek) i przeprowadza się inkubację w określonym czasie. Dla odczynników SHAPE z dłuższym czasem półtrwania zakończenie reakcji wymaga etapu neutralizacji (ang. *quenching*) z użyciem ditiotretitolu (DTT) lub β -merkaptoetanolu. Równolegle przeprowadza się również kontrolne mapowanie RNA, stosując tylko rozpuszczalnik danego odczynnika mapującego. Wynik reakcji kontrolnej stanowi tło podczas analizy próby modyfikowanej. Po reakcji modyfikacji przeprowadzana jest izolacja RNA, osobno z próby modyfikowanej i niemodyfikowanej. Planując eksperymenty mapowania chemicznego RNA *in vivo* trzeba mieć na względzie typ badanych komórek. Ze względu na brak ściany, komórki zwierzęce cechuje większa przepuszczalność dla odczynnika modyfikującego. Należy jednak pamiętać, że komórki adherentne są trudniejszym celem niż komórki rosnące w zawieszynie (Busan i in., 2019; Smola & Weeks, 2018). W przypadku bakterii czy drożdży odczynnik musi spenetrować nie tylko półprzepuszczalną błonę, ale też ścianę komórkową, co może wymagać zwiększenia jego stężenia lub wydłużenia czasu modyfikacji. Dlatego przy wyborze odczynnika mapującego należy brać pod uwagę jego właściwości, w szczególności rozpuszczalność oraz czas półtrwania. Duże znaczenie ma również pH pożywki hodowlanej, gdyż optymalnym zakresem działania dla odczynników SHAPE jest 7,4–8,3 (Smola & Weeks, 2018). W niektórych przypadkach może być konieczna wymiana medium na takie o dostosowanym pH lub zawieszenie komórek w roztworze PBS bezpośrednio przed modyfikacją SHAPE. Ze względu na powyższe kwestie wymagane jest wykonanie dokładnych testów dla każdego badanego typu komórek przed rozpoczęciem właściwych eksperymentów. Zarówno zbyt intensywna, jak i niewystarczająca modyfikacja SHAPE jest efektem niepożądanym, rzutującym na dalsze etapy eksperymentu oraz finalny wynik. Zbyt słaba

modyfikacja może dostarczyć profil reaktywności gorszej rozdzielczości, a nawet nieodróżnialny od tła, co może być błędnie zinterpretowane jako wysoce ustrukturyzowany RNA. Zbyt intensywna modyfikacja może utrudnić dalsze etapy pracy, ponieważ częste addukty w RNA nadmiarowo zmniejszają procesywność odwrotnej transkryptazy w kolejnym kroku procedury SHAPE. Ponadto w przypadku niektórych technik detekcji modyfikacji (np. elektroforeza kapilarna), zbyt duża reaktywność nukleotydów może sztucznie zawyżać reaktywność nukleotydów sąsiednich.

5.2.2 Detekcja modyfikacji SHAPE

Kolejnym kluczowym krokiem jest dokładne wykrycie miejsc modyfikacji w cząsteczce RNA. W tym celu wykorzystuje się reakcję odwrotnej transkrypcji, podczas której enzym napotyka resztę nukleotydową z adduktem oddysocjowuje od RNA tworząc skrócony produkt cDNA (ang. *RT-stop*) lub wprowadza mutację w cDNA, którego synteza jest dalej kontynuowana (ang. *RT-mutate*) (Rysunek 4). W podejściu *RT-stop* wytworzona jest pula różnej długości produktów cDNA, których końce wskazują miejsca modyfikacji w cząsteczce RNA. Z uwagi na spadek procesywności odwrotnej transkryptazy w obecności modyfikacji w RNA, enzym nie jest w stanie wyprodukować fragmentów dłuższych niż 500 nt. W celu odczytania modyfikacji z dłuższych cząsteczek RNA (na przykład ponad 5000-nt gRNA retrotranspozonów Ty) wymagane jest przeprowadzenie nawet kilkunastu reakcji odwrotnej transkrypcji z różnymi starterami w celu wygenerowania produktów cDNA nakładających się wzdłuż całej cząsteczki RNA. Pierwotnie do analizy uzyskanej puli cDNA wykorzystywano klasyczną metodę rozdziału elektroforetycznego w żelu poliakrylamidowym, którą później zastąpiła elektroforeza kapilarna (SHAPE-CE) (Wilkinson i in., 2008). Podejście *RT-stop* jest także wykorzystywane w nowszych technikach SHAPE (np. SHAPE-seq), gdzie z puli cDNA przygotowana jest biblioteka dsDNA, która jest następnie analizowana z wykorzystaniem sekwencjonowania nowej generacji (NGS, ang. *next generation sequencing*) (Lucks i in., 2011).

Ograniczeniem technik elektroforetycznych jest konieczność użycia podczas odwrotnej transkrypcji starterów o specyficznej sekwencji, gdyż analizowana jest długość produktów cDNA, a więc początek sekwencji musi być stały. Ogranicza to ich stosowanie do detekcji modyfikacji w konkretnych RNA. Wprowadzenie do protokołu SHAPE technik NGS pozwoliło na stosowanie starterów o sekwencji losowej zarówno w podejściu *RT-stop*, jak i *RT-mutate* (opisanym poniżej), gdyż odczytywana jest cała sekwencja produktów. Umożliwia to badanie od jednej do wielu cząsteczek RNA w heterogennej puli, a nawet całych transkryptomów. Gdy celem badań jest RNA o niskiej frekwencji w puli, również stosuje się specyficzne startery w celu zwiększenia czułości detekcji.

W przypadku metody SHAPE-CE podczas odwrotnej transkrypcji stosuje się fluorescencyjnie znakowane startery, dzięki czemu każda cząsteczka cDNA wzbudza swoisty

sygnał podczas rozdziału, który jest zapisywany jako pik na elektroferogramie (Vasa i in., 2008). W ten sposób każdy koniec cząsteczki cDNA wskazujący miejsce oddysocjowania enzymu, a tym samym miejsce modyfikacji SHAPE, zostaje wykryty. W pojedynczej próbie poddawanej rozdzielaniu znajduje się próba modyfikowana, próba kontrolna oraz jedna lub dwie reakcje sekwencjonowania z zastosowaniem dideoksynukleotydów (np. ddCTP i ddATP), które są potrzebne do zidentyfikowania miejsc modyfikacji w sekwencji badanego RNA. Każda z reakcji zawiera inny znacznik fluorescencyjny, więc mogą być one jednocześnie odczytywane przez laser podczas jednego rozdziału. Wysoka przepustowość i czułość elektroforezy kapilarnej pozwala uzyskać informację o intensywności wykrywanego sygnału, która podczas analizy przeliczana jest na poziom modyfikacji nukleotydu.

Dużym przełomem był rozwój techniki SHAPE-MaP, która bazuje na podejściu *RT-mutate* (Siegfried i in., 2014). Główna zmiana polega na zastąpieniu jonów Mg^{2+} jonami Mn^{2+} w reakcji odwrotnej transkrypcji. Spowodowało to większą tolerancję enzymu na obecność modyfikacji w cząsteczce RNA, dzięki czemu napotykając addukt wprowadza on preferencyjnie punktową mutację w powstającym cDNA, rzadziej oddysocjując od matrycy RNA. Ta z pozoru niewielka zmiana umożliwiła otrzymanie informacji o więcej niż jednej modyfikacji w pojedynczej cząsteczce RNA, znacznie zwiększając czułość metody. Informacja o modyfikacjach współistniejących w jednej nici RNA umożliwia także badanie oddziaływań w przestrzeni oraz wykrywanie współistniejących konformacji RNA. Przykładami takich zastosowań są metody RING-MaP (ang. *RNA interaction groups by mutational profiling*) i DANCE-MaP (ang. *deconvolution and annotation of ribonucleic conformational ensembles by mutational profiling*) (Homan i in., 2014; Mustoe i in., 2023). Strategia MaP może być również wykorzystana do wykrywania miejsc niektórych modyfikacji epigenetycznych w RNA (Y. Zhang i in., 2022). Co ważne, w przypadku metody SHAPE-MaP informacja o miejscu modyfikacji reagentem SHAPE nie znajduje się na końcu cząsteczki cDNA, co sprawia, że nie wpływa na nią proces przygotowania bibliotek do sekwencjonowania (Smola, Rice, i in., 2015). W tym podejściu modyfikacje w RNA oraz obecność jonów Mn^{2+} również znacznie obniżają procesywność odwrotnej transkryptazy, jednak przy odpowiednio dopasowanych warunkach reakcji modyfikacji, możliwe jest uzyskanie cDNA o długości dochodzącej do 1 000 nt. Opracowano również warunki pozwalające na zwiększenie długości generowanych produktów nawet do 2 500 nt, dzięki zastosowaniu odwrotnej transkryptazy o większej procesywności – MarathonRT (Guo i in., 2020). To podejście zostało wykorzystane do zbadania struktury gRNA SARS-CoV-2, który liczy ponad 30 000 nt, z użyciem jedynie 16 nakładających się amplikonów (Huston i in., 2021). Przy każdej zmianie warunków reakcji ważne jest zachowanie wydajnego poziomu wykrywanych modyfikacji w RNA i, jak na razie, MarathonRT nie wyparła oryginalnie stosowanej odwrotnej transkryptazy SuperScript II (firmy Invitrogen). Następnym etapem po odwrotnej transkrypcji jest synteza drugiej nici DNA, która w przypadku analizy pojedynczych cząsteczek RNA może

odbywać się poprzez standardową reakcję łańcuchową polimerazy (PCR, ang. *polymerase chain reaction*) z zastosowaniem specyficznych starterów. Z powstałych w ten sposób produktów dsDNA (tzw. amplikonów) przygotowywane są fragmentowane biblioteki, które są sekwencjonowane techniką krótkich odczytów (technologia firmy Illumina) w celu zidentyfikowania miejsc mutacji, a tym samym modyfikacji w cząsteczkach RNA.

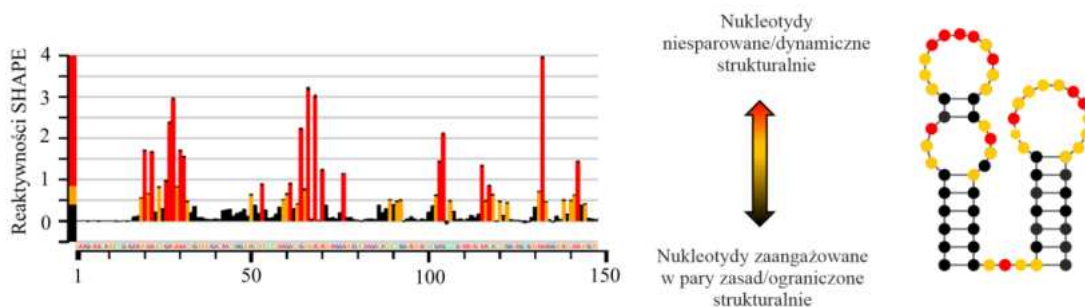
5.2.3 Analiza bioinformatyczna danych z eksperymentów SHAPE

Nieodłącznym elementem opisanych metod jest analiza bioinformatyczna uzyskanych danych w celu otrzymania finalnych reaktywności SHAPE dla poszczególnych reszt nukleotydowych w cząsteczce RNA. W przypadku metody SHAPE-CE, wyniki rozdziału elektroforetycznego analizowane są z zastosowaniem np. programu ShapeFinder (Vasa i in., 2008), w którym przeprowadza się dopasowywanie pików do sekwencji referencyjnej oraz kilkietapową normalizację (Rysunek 4). Program ma za zadanie obliczyć powierzchnię każdego z pików dla próby modyfikowanej, traktując sygnał z próby kontrolnej jako tło. Następnie dane liczbowe eksportowane są do programu Microsoft Excel, gdzie dane poddaje się dodatkowej normalizacji, a powierzchnie pików przelicza się na finalne reaktywności SHAPE. Można wykorzystać do tego celu także narzędzie RNathor, opracowane w naszym zakładzie we współpracy z Zakładem Bioinformatyki Strukturalnej ICHB PAN, które przeprowadza zautomatyzowaną normalizację, wizualizację oraz analizę statystyczną danych (Gumna i in., 2020).

W strategii SHAPE-MaP końcowym rezultatem części eksperymentalnej jest wynik sekwencjonowania NGS, osobny dla próby modyfikowanej i kontrolnej. Głównym celem analizy danych jest porównanie wszystkich odczytanych sekwencji do sekwencji referencyjnej badanego RNA i wskazanie miejsc oraz poziomu wprowadzonych mutacji. Poziom mutacji każdego nukleotydu skorelowany jest z poziomem jego modyfikacji przez odczynnik SHAPE. Do analizy uzyskanych danych NGS można wykorzystać dedykowane oprogramowanie Shapemapper 2 (Busan & Weeks, 2018). Do przeprowadzenia poszczególnych etapów analizy korzysta on z towarzyszących programów i kontroluje ich działanie, umożliwiając spójną pracę i przekazywanie danych pośrednich (Rysunek 4). Plikami wejściowymi są w tym przypadku pliki z odczytami sekwencji (format .fastq) oraz sekwencją referencyjną (format .fasta). Etapy analizy obejmują przycinanie odczytów według jakości, scalanie sparowanych odczytów, dopasowywanie do sekwencji referencyjnej, a następnie kilkietapowe obliczanie współczynnika mutacji dla każdej pozycji w DNA. Finalnym wynikiem jest profil reaktywności SHAPE, któremu towarzyszą dane analityczne, takie jak głębokość sekwencjonowania, współczynnik mutacji, odchylenie standardowe dla każdej pozycji, a także bardziej szczegółowe dane liczbowe.

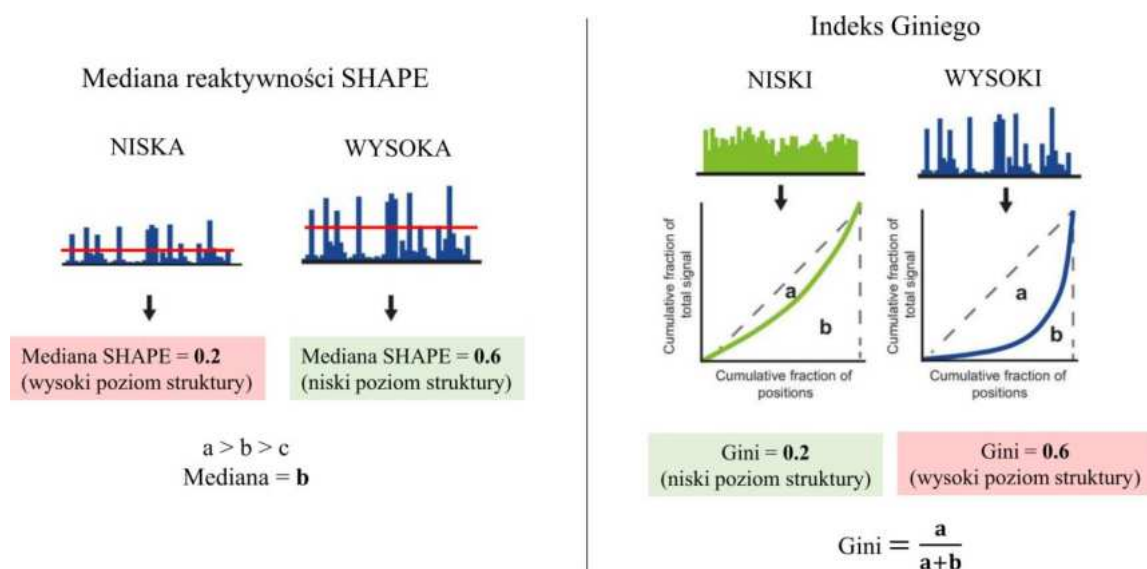
5.2.4 Analiza i interpretacja reaktywności SHAPE

Z założenia otrzymany profil reaktywności SHAPE zawiera zbiór informacji o poziomie modyfikacji dla każdej reszty nukleotydowej, co odpowiada jej dostępności w strukturze danego RNA (Rysunek 6). Podczas analizy reaktywności SHAPE przyjmuje się skalę kolorystyczną, gdzie czarne pozycje wskazują najniższe reaktywności (0–0,4), żółte – reaktywności pośrednie (0,4–0,85), a czerwone – najwyższe wartości (>0,85). W ten sposób wizualna analiza otrzymanego profilu jest w stanie wskazać rejony sparowane i jednoniciowe.



Rysunek 6. Przykładowy profil reaktywności SHAPE oraz model struktury drugorzędowej.

Przydatnymi parametrami podczas wstępnej analizy właściwości strukturalnych RNA są mediana reaktywności SHAPE i indeks (współczynnik) Giniego (Rysunek 7). Mediana, jako powszechna miara statystyczna, wskazuje środkową wartość reaktywności SHAPE w zbiorze wszystkich reaktywności występujących w danej cząsteczce, uporządkowanych od wartości najmniejszej do największej. Porównanie median reaktywności SHAPE pozwala szybko ocenić, czy dwie cząsteczki RNA mają podobny ogólny poziom struktury. Indeks Giniego analizuje dysproporcje w profilach reaktywności SHAPE. Wysoce ustrukturyzowane rejony RNA mają nierównomierny rozkład wysokich i niskich reaktywności, a tym samym wykazują wysoki indeks Giniego. Rejony RNA o niskim poziomie struktury mają większość zasad dostępnych dla modyfikacji odczynikiem SHAPE i tym samym wykazują niski indeks Giniego. Przykładowo, profil reaktywności dla cząsteczki rybosomalnego RNA o bardzo wysokim poziomie ustrukturyzowania (Giannetti i in., 2019; Petrov i in., 2014), będzie odznaczał się niską medianą SHAPE oraz wysokim indeksem Giniego. Obydwie miary dostarczają informacji, które pomagają wstępnie zaklasyfikować daną cząsteczkę RNA jako wysoce lub nisko ustrukturyzowaną w danych warunkach eksperymentalnych. Ponadto, takie metryki są szczególnie przydatne w przypadku analizy całych transkryptomów, gdy mamy do czynienia z tysiącami cząsteczek RNA.



Rysunek 7. Schematyczny opis i interpretacja mediany reaktywności oraz indeksu Giniego w analizie danych SHAPE (na podstawie publikacji (Burkhardt i in., 2017)).

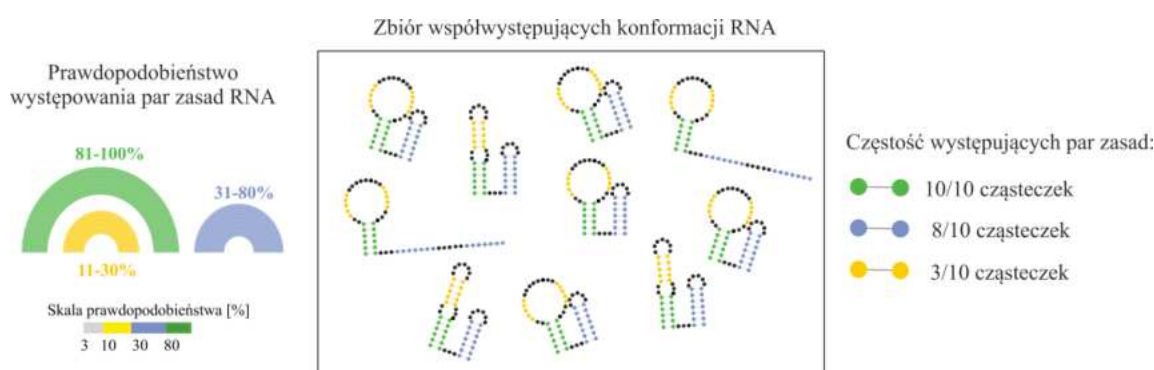
W celu porównania profili reaktywności SHAPE, zarówno dla dwóch sekwencji RNA tej samej długości, jak i dla tej samej cząsteczki RNA badanej w różnych warunkach, stosuje się również współczynnik korelacji (r), który informuje o stopniu podobieństwa między zestawami danych, zależnym od pozycji. Wysoka wartość korelacji oznacza duże podobieństwo wzoru reaktywności, natomiast niska korelacja oznacza różnice. Oprócz ogólnej wartości korelacji pomiędzy profilami, warto obliczyć korelację w oknie przesuwającym się wzdłuż sekwencji RNA (ang. *sliding window*), która ma na celu dokładniejsze wskazanie lokalnych wzrostów i spadków podobieństwa w reaktywności SHAPE.

5.2.5 Predykcja oraz interpretacja modeli struktury drugorzędowej RNA

Zdolność RNA do tworzenia par zasad, w tym kanonicznych par zasad Watsona-Cricka i niekanonicznych par zasad (na przykład G:U), powoduje tworzenie drugorzędowych struktur RNA, w których nukleotydy tworzą motywy strukturalne, takie jak trzony, pętle, wybrzuszenia, czy też pseudowęzły. Te interakcje tworzą następnie podstawę do oddziaływań przestrzennych, tworząc struktury trzeciorzędowe RNA. Rozwój algorytmów do predykcji modeli strukturalnych stał się ważnym elementem badania struktur RNA (J. Zhang i in., 2022). Wśród nich znalazł się między innymi powszechnie stosowany model obliczeniowy, który poszukuje struktur drugorzędowych RNA o minimalnej energii swobodnej (MFE, ang. *minimum free energy*), czyli teoretycznie najbardziej stabilnych. Obliczenia dokonywane są na podstawie sekwencji i w oparciu o zestaw eksperymentalnie określonych parametrów termodynamicznych, znanych jako model najbliższego sąsiada (NN, ang. *nearest neighbor*) lub reguły Turnera (Xia i in., 1998). Algorytm ten jest wykorzystywany między innymi przez program RNAstructure (Reuter & Mathews, 2010). Choć opiera się on na założeniach dotyczących obliczeń energii w roztworze, co

nie do końca odzwierciedla sposób związania się cząsteczek RNA w komórkach, opracowano wydajne metody włączania do modelowania eksperymentalnych danych strukturalnych w celu poprawy dokładności predykcji (Cordero i in., 2012; Deigan i in., 2009; Hajdin i in., 2013; Mathews, 2004).

W mojej pracy do modelowania struktury drugorzędowej cząsteczek RNA wykorzystałam algorytmy zawarte w programie RNAstructure, a otrzymane dane SHAPE zastosowałam jako liczbowe ograniczenia pseudoenergetyczne do wsparcia obliczeń. Analizy wykonałam za pomocą oprogramowania SuperFold, które wykorzystuje funkcje zaimplementowane w programie RNAstructure, wykonując obliczenia w przesuującym się oknie, co usprawnia predykcję długich cząsteczek RNA (Siegfried i in., 2014; Smola, Rice, i in., 2015). Pierwszym etapem jest obliczenie prawdopodobieństwa tworzenia się par zasad i entropii Shannona dla poszczególnych nukleotydów RNA (Mathews, 2004). Prawdopodobieństwo par zasad pokazuje, z jaką częstością występują one w puli możliwych koegzystujących konformacji RNA (Rysunek 8). Natomiast wartość entropii Shannona podsumowuje liczbę możliwych konformacji dla każdego nukleotydu, określając poziom uporządkowania każdej pozycji w cząsteczce RNA w danych warunkach eksperymentalnych. Niska wartość entropii informuje o wysokim poziomie uporządkowania, natomiast dla rejonów dynamicznych strukturalnie jej wartość będzie wysoka.



Rysunek 8. Schemat ilustrujący prawdopodobieństwo występowania par zasad w cząsteczce RNA. Przedstawiony po lewej stronie łukowy model wraz z kolorystyczną skalą prezentuje częstość występowania poszczególnych par zasad w zbiorze współwystępujących konformacji RNA zaprezentowanym obok.

Kolejnym etapem analizy jest predykcja modelu struktury drugorzędowej MFE, podczas której oprócz danych eksperymentalnych, oprogramowanie SuperFold wykorzystuje informację z poprzedniego etapu o parach zasad tworzonych z prawdopodobieństwem powyżej 99%, aby dodatkowo wzmocnić ich predykcję w ostatecznym modelu. Pary zasad przewidywane są osobno w każdym z dziesiątek analizowanych okien, a te występujące w większości okien są włączane do finalnego modelu strukturalnego cząsteczki RNA. Takie podejście ogranicza wyniki predykcji tylko do jednego konsensusowego modelu, w porównaniu do kilkunastu modeli o podobnej energii swobodnej, generowanych przez program RNAstructure. Ostatnim etapem analizy

SuperFold jest przeskanowanie uzyskanego modelu w celu znalezienia rejonów o stabilnej i dobrze zdefiniowanej strukturze. Następuje to poprzez wyselekcjonowanie rejonów o długości co najmniej 40 reszt nukleotydowych, charakteryzujących się lokalną reaktywnością SHAPE oraz entropią Shannona niższą niż globalne wartości mediany dla tych parametrów – strategia *lowSS* (ang. *low-SHAPE*, *low-Shannon*) (Siegfried i in., 2014).

Analiza porównawcza modeli struktury RNA otrzymanych w różnych stanach eksperymentalnych może dostarczyć wiele informacji na temat dynamiki strukturalnej danej cząsteczki. Do określenia stopnia podobieństwa dwóch modeli strukturalnych służą metryki wykorzystywane również do oceny poprawności predykcji, gdy struktura referencyjna jest znana, takie jak czułość (ang. *sensitivity / true positive rate*, TPR) i PPV (ang. *predictive positive value*). Wskazują one procentową liczbę par zasad wspólnych dla porównywanych modeli. Czułość wskazuje liczbę wspólnych par zasad w stosunku do struktury referencyjnej, a PPV w stosunku do struktury przewidzianej. Nawet przy wysokiej wartości tych parametrów, nie należy pomijać faktu, że poszczególne pary mogą tworzyć się z nierównym prawdopodobieństwem. Ponadto należy pamiętać, że powszechnie stosowany do prezentacji drugorzędowej struktury RNA pojedynczy model MFE reprezentuje tylko jedną konformację ze zbioru możliwych o podobnych energiach swobodnych. Dlatego modelowi MFE powinny towarzyszyć opisane powyżej analizy, biorące pod uwagę dynamikę strukturalną RNA, która wraz z porównaniem modeli otrzymanych w różnych stanach eksperymentalnych stanowi kompleksowe podejście podczas interpretacji struktury drugorzędowej RNA.

5.2.6 Aktualne wyzwania w badaniach struktury drugorzędowej RNA

Połączenie chemicznego mapowania RNA z sekwencjonowaniem nowej generacji i zaawansowanymi narzędziami bioinformatycznymi zwiększyło przepustowość badań struktury RNA, pozwalając na badanie cząsteczek o długości nawet kilkudziesięciu tysięcy nukleotydów, dając wynik wysokiej rozdzielczości dla każdego rybonukleotydu w złożonym środowisku komórkowym. Badania coraz większej ilości cząsteczek RNA rzuciły nowe światło na niedocenianą wcześniej złożoność i dynamikę ich strukturalnej organizacji (Vicens & Kieft, 2022; Weeks, 2021).

Jednym z aktualnych wyzwań w dziedzinie badań strukturalnych RNA jest próba rekonstrukcji współistniejących alternatywnych konformacji RNA (Bose i in., 2024; Spitale & Incarnato, 2023). Zarówno wewnątrz komórki, jak i w roztworze, wiele kopii tego samego RNA może przyjmować różne konformacje, które dodatkowo mogą zmieniać się w czasie. Taka zdolność cząsteczek RNA może regulować różne procesy komórkowe, co ma miejsce między innymi dla genomów RNA wirusów (Lan i in., 2022; Larman i in., 2017; Tomczko i in., 2020). Aby sprostać temu zagadnieniu, opracowywane są metody eksperymentalne, takie jak chemiczne mapowanie produktów pośrednich transkrypcji, które tworzą krótkotrwałe motywy. Do takich

metod należą przykładowo SHAPE-seq w czasie transkrypcji (ang. *cotranscriptional SHAPE-seq*) (Watters i in., 2016) i SPET-seq (ang. *structural probing of elongating transcripts*) (Incarnato i in., 2017). Rozwijane są również metody obliczeniowe oparte zarówno na termodynamicznej selekcji struktur RNA najlepiej pasujących do danych eksperymentalnych, jak i algorytmy próbujące bezpośrednio przypisać odczyty sekwencjonowania do odrębnych klastrów, reprezentujących współistniejące konformacje RNA (Aviran & Incarnato, 2022).

Innym kierunkiem podejmowanych wyzwań jest także wykorzystanie technologii sekwencjonowania długich odczytów, czy też bezpośredniego sekwencjonowania RNA. W porównaniu do sekwencjonowania technologią krótkich odczytów, długie odczyty mogą dostarczyć informacji o całej cząsteczce RNA w jednym odczycie, co umożliwia detekcję współistniejących modyfikacji na długich odcinkach sekwencji, ułatwiając analizę współistniejących konformerów. Pojawiły się już pierwsze próby zastosowania sekwencjonowania RNA do zbadania jego struktury przy użyciu platformy Oxford Nanopore (Aw i in., 2021; Bizuayehu i in., 2022; Bohn i in., 2023; Stephenson i in., 2022). W jednym z ostatnich badań wyselekcjonowano acetyloimidazol (AcIm), jako najbardziej obiecujący odczynnik SHAPE do tego zastosowania, z uwagi na mały rozmiar generowanych adduktów (Stephenson i in., 2022). W pracy tej pokazano również, że technologia Nanopore jest w stanie wykryć jednocześnie endogenne, jak i egzogenne modyfikacje RNA.

Atrakcyjną alternatywą dla eksperymentalnego badania struktury drugorzędowej RNA jest predykcja struktury oparta na samej sekwencji. Niezwykły sukces w rozwoju technik uczenia maszynowego i ich zastosowanie do przewidywania struktury białek (algorytmy AlphaFold — (Abramson i in., 2024; Jumper i in., 2021)) stanowi ogromną motywację do stworzenia podobnego podejścia do przewidywania struktury RNA. Choć przewidywania modeli struktury drugorzędowej RNA oparte na termodynamice osiągnęły znaczący postęp, uważa się, że ich dokładność jest wciąż niezadowalająca przez ograniczoną liczbę eksperymentalnie wyznaczonych parametrów termodynamicznych, przez co w dalszym ciągu wyzwaniem jest prawidłowa predykcja pseudowęzłów, czy niekanonicznych par zasad (S. Zhang i in., 2024). Dodatkowo istnieje duża luka między liczbą ustalonych eksperymentalnie struktur a liczbą znanych sekwencji RNA. Pozyskanie danych eksperymentalnych, które mogą znacząco poprawić dokładność predykcji, wymaga czasu. W ostatnich latach pojawiło się kilka obiecujących podejść wykorzystujących techniki uczenia maszynowego do przewidywania struktur RNA, które są intensywnie rozwijane i testowane, jak na przykład SPOT-RNA, CONTRAfold, ATTFold, czy ostatnio udostępniony sincFold (Bugnon i in., 2024; Justyna i in., 2023; K. E. Wu i in., 2023; Yu i in., 2022). Choć przeprowadzają one coraz celniejsze predykcje modeli RNA, na podstawie których były uczone, to nadal jednym z głównych problemów jest nadmierne dopasowanie (tzw. przeuczenie, ang. *overfitting*), co oznacza zbyt duże wyspecjalizowanie na dane treningowe i tym samym błędne analizowanie nowych danych. Głównym powodem tego zjawiska jest zbyt mały i

jednorodny zestaw danych treningowych, co w przypadku struktur drugorzędowych RNA jest nadal problemem, gdyż zbiór wiarygodnych i eksperymentalnie potwierdzonych modeli RNA jest dużo mniejszy niż dla białek. Dlatego do dopracowania tych metod kluczowe jest powiększenie bazy eksperymentalnych zestawów danych o strukturze RNA w różnych organizmach i warunkach.

5.3 Struktura gRNA wirusów i retrotranspozonów oraz RNA drożdży *in vivo*

W odróżnieniu od licznych modeli *in vitro*, dotychczas ustalono struktury genomu RNA *in vivo* lub *in cellulo* tylko dla kilku wirusów – Dengue, Zika i HCV (*Flaviviridae*), HIV-1 (*Retroviridae*), RVA (*Reoviridae*), CHIKV (*Togaviridae*), a także SARS-CoV-2 (*Coronaviridae*) (Coria i in., 2022; Huber i in., 2019; Huston i in., 2021; Lan i in., 2022; Li i in., 2018; Madden i in., 2020; Manfredonia i in., 2020; Sun i in., 2021; Tomezsko i in., 2020; Wan i in., 2022; Y. Zhang i in., 2021). Otrzymane dane dostarczyły dodatkowych informacji na temat roli struktury gRNA w procesie replikacji wirusów. Co ważne, zwróciły też uwagę na istotne różnice w zwijaniu się wirusowych gRNA *in vivo* i *in vitro* oraz wykazały, że funkcjonalnie ważne sekwencje RNA mają tendencję do lokalizowania się w rejonach o stabilnej strukturze drugorzędowej. Jednocześnie badania te wskazały rejony gRNA, które mogą służyć jako potencjalne cele dla nakierowanych na RNA leków przeciwwirusowych.

W momencie rozpoczęcia moich badań nie było dostępnych danych na temat struktury genomów RNA retrotranspozonów wewnątrz komórki. Otwarte pozostawało więc pytanie, czy gRNA retrotranspozonów będą odzwierciedlać cechy strukturalne wirusów infekcyjnych, czy też raczej będą podobne do transkryptów swojego gospodarza. Modele struktury drugorzędowej gRNA drożdżowego retrotranspozonu Ty1 (uzyskane metodą SHAPE-CE) zostały zaproponowane jedynie dla pierwszych 1482 nt (26%) w warunkach *ex virio*, *in virio* i *in vitro* (Purzycka i in., 2013) oraz dla stanów monomerycznych i dimerycznych *in vitro* transkryptu obejmującego rejon +1–560 nt gRNA Ty1 (Gamache i in., 2017; Gumna i in., 2019; Huang i in., 2013). Z kolei dla retrotranspozonu Ty3 brakowało jakichkolwiek danych eksperymentalnych dotyczących struktury drugorzędowej genomu RNA, nawet w warunkach *in vitro*.

W porównaniu do *E. coli*, roślin lub ssaków, znacznie mniej wiadomo o zwijaniu mRNA w drożdżach, które są gospodarzem badanych przeze mnie retrotranspozonów. Co ciekawe, drożdże *S. cerevisiae* były pierwszym organizmem, w którym przeprowadzono mapowanie struktury transkryptomu *in vivo* (Rouskin i in., 2014). Do tego celu zastosowano strategię DMS-seq, a uzyskane wyniki zasugerowały, że mRNA w drożdżach są w dużej mierze rozwinięte w porównaniu do stanu *in vitro*. Nie znaleziono również globalnej różnicy w poziomie struktury drugorzędowej pomiędzy CDS (ang. *coding DNA sequence*) i UTR transkryptów *in vivo*. Było to sprzeczne z poprzednimi danymi *in vitro* pochodzącymi z eksperymentów PARS (ang. *parallel analysis of RNA structure*) wskazującymi, że mRNA drożdży wykazują wyższy poziom struktury

w CDS (Kertesz i in., 2010). Analiza DMS-seq *in vivo* nie wykryła również korelacji między strukturą CDS a jego statusem translacyjnym (Rouskin i in., 2014). Późniejsza analiza mapowania oddziaływań RNA-RNA w transkryptomie drożdży *in vivo* przy użyciu biotynylowanego psoralenu i ligacji zbliżeniowej – SPLASH (ang. *sequencing of psoralen crosslinked, ligated, and selected hybrids*) pokazała, że stabilne struktury w 5' UTR zmniejszają wydajność translacji i okres półtrwania mRNA, natomiast ich obecność w 3' UTR zwiększa stabilność mRNA, najprawdopodobniej poprzez hamowanie kompleksu egzozomu podczas degradacji RNA (Aw i in., 2016). Pozytywny wpływ struktury 3' UTR na stabilność mRNA drożdży został potwierdzony *in vivo* także przy użyciu zmodyfikowanych wersji DMS-seq (DREADS, ang. *DMS region extraction and deep sequencing*) i SHAPE-seq (SHREADS, ang. *SHAPE region extraction and deep sequencing*), które są przeznaczone do określania struktury drugorzędowej poliadenylowanych końców 3' RNA (Moqtaderi i in., 2018). Analogiczne wyniki uzyskano w kilku badaniach obejmujących cały transkryptom w komórkach ssaków, co sugeruje, że regulacja funkcji mRNA za pośrednictwem struktur w UTR jest ewolucyjnie konserwowana (Sun i in., 2019; X. Wu & Bartel, 2017). Podczas gdy rola drugorzędowej struktury UTR w regulacji ekspresji genów była szeroko badana, zakres, w jakim struktura CDS wpływa na wydajność translacji i stabilność drożdżowych mRNA *in vivo*, jest znacznie mniej poznany. Globalna analiza struktury RNA na komórkach ssaków połączona z zastosowaniem substytucji modyfikowanymi nukleotydami i obliczeniowym projektowaniem sekwencji wykazała, że wysoki poziom struktury drugorzędowej w CDS koreluje dodatnio z syntezą białka i okresem półtrwania mRNA (Mauger i in., 2019). Korelacja między strukturą CDS a wydajnością translacji i stabilnością mRNA została zaobserwowana u niższych kręgowców (*Danio rerio*), bakterii (*Escherichia coli*), a także u roślin (*Oryza sativa* i *Arabidopsis thaliana*) (Beaudoin i in., 2018; Burkhardt i in., 2017; Deng i in., 2018; Ding i in., 2014; Mustoe i in., 2018).

6 CEL NAUKOWY PRZEDSTAWIONEGO CYKLU PRAC

Nadrzędnym celem przedstawionego cyklu prac było ustalenie struktury drugorzędowej cząsteczek genomowego RNA retrotranspozonów LTR w drożdżach. Badania obejmowały również analizę struktury gRNA poza komórką w celu identyfikacji zmian strukturalnych spowodowanych środowiskiem komórki, a także analizy korelacji pomiędzy strukturą gRNA a funkcją pełnioną w czasie cyklu replikacyjnego retrotranspozonu. Finalnie, wyniki moich badań miały na celu poszerzenie wiedzy na temat związania się długich cząsteczek RNA *in vivo*. Cele szczegółowe moich badań pokrywają się z tematyką trzech publikacji wchodzących w skład cyklu tworzącego rozprawę doktorską:

1. Praca przeglądowa **Andrzejewska, Zawadzka i in., IJMS, 2020** miała na celu zebranie i usystematyzowanie dostępnej wiedzy na temat wpływu środowiska komórki na strukturę cząsteczek mRNA oraz wzajemnej relacji między strukturą mRNA a jego funkcją w procesach komórkowych.
2. Badania przedstawione w pracy **Andrzejewska i in., NAR, 2021** były ukierunkowane na dostarczenie pierwszego modelu struktury genomu RNA aktywnego retrotranspozonu Ty1 *in vivo* oraz porównanie go z modelami struktury otrzymanymi w warunkach *in vitro* oraz *in virio*. Istotna była również charakterystyka kontekstu strukturalnego znanych sekwencji funkcjonalnych zaangażowanych w oddziaływanie RNA-RNA istotne dla retrotranspozycji. Moim dodatkowym celem była analiza wpływu aktywnych rybosomów na strukturę gRNA.
3. Celem pracy **Andrzejewska-Romanowska i in., NAR, 2024** było ustalenie struktury genomu RNA retrotranspozonu Ty3 w warunkach *in vivo*, jak i poza komórką, oraz charakterystyka porównawcza dynamiki strukturalnej gRNA Ty3 w różnych stanach. Moim zadaniem była także szczegółowa analiza struktury rejonów ważnych funkcjonalnie oraz wytypowanie najbardziej stabilnych motywów strukturalnych w gRNA Ty3. Ideą pracy było również dostarczenie strukturalno-funkcjonalnego porównania gRNA reprezentantów trzech ewolucyjnie spokrewnionych rodzin retroelementów – retrotranspozonów Ty1 oraz Ty3, a także wirusa HIV-1.

7 KRÓTKIE OMÓWIENIE PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

7.1 On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective

Angelika Andrzejewska*, Małgorzata Zawadzka*, Katarzyna Pachulska-Wieczorek

International Journal Of Molecular Sciences, 2020, 21(18), 6770

* równorzędny pierwszy autor

RNA uczestniczy w niemal każdym procesie zachodzącym w komórce i działa jako centralny nośnik wymiany informacji. Informacje te są zawarte zarówno w sekwencji nukleotydowej RNA, jak i jego strukturze. Powszechna dostępność technik sekwencjonowania nowej generacji zrewolucjonizowała wiele obszarów badań RNA, wliczając badania strukturalne. Umożliwiło to przesunięcie perspektywy z pojedynczych transkryptów na tysiące heterogennych RNA, w tym całe transkryptomy w złożonym środowisku komórkowym. Rosnąca liczba danych strukturalnych dla transkryptomów różnych organizmów, w tym ssaków, drożdży, roślin i bakterii oraz płynące z nich czasami sprzeczne wnioski stworzyły potrzebę usystematyzowania zgromadzonej wiedzy. Podjęliśmy się tego w niniejszej pracy przeglądowej, której przygotowanie pozwoliło mi na zdobycie niezbędnej wiedzy w dziedzinie badań strukturalnych RNA.

W pierwszej części zebraliśmy, przeprowadzone do momentu przygotowania pracy, badania struktury drugorzędowej transkryptomów oraz zastosowane w nich metody. Wszystkie najważniejsze informacje z tego rozdziału zostały zebrane w Tabeli 1 w publikacji. Następnie, podjęliśmy się uporządkowania wniosków płynących z zebranych badań. W pierwszej kolejności podsumowaliśmy obserwacje dotyczące wzajemnego oddziaływania pomiędzy strukturą mRNA a procesem translacji podkreślając, że kierunek tej relacji nadal stanowi kwestię sporną. Dokonaliśmy również podziału zebranych wyników w zależności od tego, czy obserwacje dotyczyły rejonów kodujących, czy nieulegających translacji. W dalszej części pracy zestawiliśmy najnowsze obserwacje dotyczące zależności między strukturą mRNA a jego stabilnością i procesem degradacji, zwracając szczególną uwagę na liczne wnioski dotyczące znaczenia struktury 3' UTR. Kolejny rozdział skupia się na badaniach dotyczących aktywności białek wiążących RNA, zarówno w sposób specyficzny jak i niezależny od konkretnych sekwencji bądź motywów RNA. Badanie tych oddziaływań jest wciąż dużym wyzwaniem w skali całego proteomu i transkryptomu *in vivo*. Ostatni rozdział poświęciliśmy podsumowaniu badań dotyczących obecności modyfikacji potranskrypcyjnych w RNA i ich wpływu na strukturę drugorzędową.

7.2 *In vivo* structure of the Ty1 retrotransposon RNA genome

Angelika Andrzejewska, Małgorzata Zawadzka, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek

Nucleic Acids Research, 2021, 49(5), 2878–2893

Wieloletnie badania dotyczące biologii Ty1 podkreśliły kluczową rolę funkcjonalnych sekwencji w jego gRNA dla oddziaływań RNA-RNA i RNA-białko niezbędnych do procesu retrotranspozycji (Curcio i in., 2015). Mimo to, dane strukturalne dla gRNA Ty1 były ograniczone tylko do 5'-końcowego fragmentu w warunkach *in vitro* i *in virio* (Gamache i in., 2017; Gumna i in., 2019; Purzycka i in., 2013). Odpowiedź na pytanie, jak architektura genomu RNA wpływa na replikację Ty1 oraz które interakcje mogą zachodzić przed utworzeniem cząstek wirusopodobnych, wymagała zbadania struktury niespakowanego gRNA Ty1 w warunkach *in vivo*.

W niniejszej pracy nasz zespół podjął się scharakteryzowania struktury drugorzędowej pełnej długości gRNA retrotranspozonu Ty1 (5652 nt) *in vivo* oraz odpowiadającego mu transkryptu *in vitro*, zwijanego w ściśle ustalonych warunkach (Rysunek 1A w publikacji). Wykorzystując metodę SHAPE-CE w drożdżach oraz narzędzia bioinformatyczne dedykowane analizie długich RNA, dostarczyłam informacji na temat struktury gRNA Ty1 w jego natywnym stanie *in vivo*. Do mapowania chemicznego wykorzystałam odczynnik NMIA z uwagi na to, że został on zastosowany we wcześniejszych badaniach tego RNA *in vitro* i *in virio*. Istotnym etapem pracy były eksperymenty porównujące zdolność różnych odczynników SHAPE do mapowania struktury RNA w drożdżach (Rysunki 1B, C w publikacji). Udowodniłam, wbrew wcześniejszym danym (Spitale i in., 2013), że NMIA może być z powodzeniem stosowany do badań struktury RNA *in vivo*. Z wykorzystaniem analogicznego podejścia dr Małgorzata Zawadzka wykonała mapowanie chemiczne transkryptu *in vitro*, co pozwoliło na przeprowadzenie szczegółowej analizy struktury gRNA Ty1 w dwóch stanach eksperymentalnych.

Przeprowadzone przeze mnie eksperymenty dostarczyły wysokiej jakości danych strukturalnych dla 98% reszt nukleotydowych gRNA Ty1, z czego aż 53% nukleotydów wykazywało wysokie (powyżej 0,85) lub pośrednie (0,4–0,85) reaktywności SHAPE. Taka proporcja sugerowała wysoką ilość strukturalnie dynamicznych, jednoniciowych rejonów gRNA. Dodatkowo, zaobserwowałam, że wiele rejonów gRNA Ty1 *in vivo* odznaczało się wyższymi reaktywnościami SHAPE niż *in vitro*, co skutkowało znacząco wyższą ogólną medianą reaktywności (0,42 vs 0,35) (Rysunki 2A, C, D, F w publikacji). Wykorzystując oprogramowanie Superfold oraz otrzymane dane eksperymentalne, jako liczbowe ograniczenia pseudoenergetyczne, obliczyłam prawdopodobieństwa tworzenia par zasad i entropii Shannona

dla poszczególnych nukleotydów. Mediany obu parametrów wskazały na niższy poziom uporządkowania gRNA Ty1 *in vivo* w porównaniu do stanu *in vitro* (Rysunki 2G, H w publikacji).

W następnym etapie przeprowadziłam modelowanie konsensusowej struktury MFE gRNA Ty1 (Rysunek 3C w publikacji). Analiza otrzymanego modelu sugerowała, że 44% wszystkich nukleotydów w gRNA jest sparowane *in vivo* (Rysunki 4B, E w publikacji). Podobny wynik uzyskano dla modelu *in vitro*. Spośród przewidzianych par zasad jedynie 55% było obecnych także *in vitro*, tym samym duża część sparowań była specyficzna dla stanu *in vivo* (Rysunki 4A, B w publikacji). Co ciekawe, model *in vivo* posiadał dwa razy mniej wysoce prawdopodobnych par zasad (prawdopodobieństwo >80%) i stanowiły one jedynie 37,5% wszystkich obecnych w modelu MFE (Rysunki 3C, 4E, F, G w publikacji). Oznaczało to, że pozostałe sparowania mogą występować przejściowo lub być obecne tylko w części współistniejących konformacji. Uzyskane wyniki ujawniły silny wpływ środowiska komórkowego na zwijanie gRNA retrotranspozonu, wskazując jednocześnie na wyższą dynamikę i niższy poziom struktury *in vivo*.

Dodatkowo, zbadalam obecność potencjalnych motywów pseudowężła w gRNA Ty1 wykorzystując otrzymane *in vivo* reaktywności SHAPE jako dane wejściowe do predykcji struktury programem ShapeKnots (Hajdin i in., 2013). Funkcjonalnie ważny pseudowęzeł został wcześniej zidentyfikowany w rejonie 5' końca gRNA Ty1 mapowanego w cząstkach wirusopodobnych (Huang i in., 2013; Purzycka i in., 2013). Pozostawało jednak niejasne, czy motyw ten tworzy się tylko w VLP. Wykonane przeze mnie analizy wskazały na obecność tego motywu w badanej puli gRNA Ty1, która pochodzi głównie z niespakowanej frakcji cytoplazmatycznej i retrosomów, wspierając formowanie się pseudowężła przed pakowaniem gRNA do VLP (Rysunki 3C, D w publikacji).

Mimo dynamicznego charakteru struktury gRNA Ty1 wewnątrz komórki, zidentyfikowałam 11 rejonów o stabilnej i dobrze zdefiniowanej strukturze, wykorzystując do tego strategię *lowSS* (Rysunki 3B, D w publikacji). Wskazało to na mozaikowy charakter architektury gRNA Ty1, zauważony wcześniej dla innych długich wirusowych i komórkowych RNA (Dethoff i in., 2018; Mustoe i in., 2018; Siegfried i in., 2014; Smola i in., 2016). Pokazałam również, że struktura tych rejonów jest zachowana pomiędzy stanami *in vivo* i *in vitro* w większym stopniu niż dla reszty gRNA Ty1 (Rysunek 4C w publikacji). Co ważne, rejony stabilne znajdujące się na 5' i 3' końcach gRNA zawierały większość znanych sekwencji funkcjonalnych, w tym sekwencje palindromowe PAL1 i PAL2 pośredniczące w dimeryzacji gRNA, sekwencje CYC5 i CYC3 biorące udział w cyklizacji gRNA, a także PBS, BOX0 i BOX1 będące miejscem przyłączenia tRNA^{Met} (Rysunek 3D w publikacji). Znaczenie pozostałych zidentyfikowanych rejonów stabilnych pozostaje na razie niejasne, i stanowią one atrakcyjne cele do dalszych badań funkcjonalnych.

Dla ustalenia, czy kluczowe dla retrotranspozycji oddziaływania RNA-RNA, takie jak dimeryzacja i cyklizacja gRNA oraz przyłączanie tRNA_i^{Met}, mogą zachodzić przed pakowaniem gRNA do VLP, przeprowadziliśmy szczegółową analizę zmian reaktywności SHAPE w rejonie końca 5' gRNA Ty1. W tym celu porównaliśmy dane *in vivo* z danymi uzyskanymi dla stanu *in vitro*, gdzie wszystkie powyższe interakcje są obecne (Purzycka i in., 2013) oraz stanu *in vitro* pozbawionego tych interakcji, z uwagi na brak białka Gag i tRNA (Rysunek 5A w publikacji). Na tej podstawie wykryliśmy charakterystyczne zmiany w reaktywności nukleotydów, wskazujące na obecność frakcji dimerycznej w puli gRNA *in vivo*, jak również międzycząsteczkowej interakcji z tRNA_i^{Met} (Rysunki 5A, B w publikacji). Obserwowane rozkłady reaktywności SHAPE wspierają również występowanie interakcji CYC5-CYC3 w gRNA Ty1 przed spakowaniem do VLP.

W ostatnim etapie zbadałam relację między strukturą gRNA Ty1 a procesem translacji, gdyż przed spakowaniem gRNA stanowi matrycę do produkcji białek retrotranspozonu. Dane, które uzyskałam z eksperymentów SHAPE-CE w komórkach w stanie zahamowanej translacji, wskazywały na znaczącą stabilizację struktury gRNA Ty1 w rejonie kodującym (Rysunki 6B, C w publikacji) oraz reorganizację par zasad w porównaniu do kontrolnego stanu *in vivo* (Rysunki 6D, E, F w publikacji). Sugeruje to, że aktywne rybosomy mogą uczestniczyć zarówno w rozwijaniu, jak i remodelowaniu struktury gRNA retrotranspozonu.

7.3 Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome

Angelika Andrzejewska-Romanowska, Julita Gumna, Ewa Tykwińska, Katarzyna Pachulska-Wieczorek

Nucleic Acids Research, 2024, doi: 10.1093/nar/gkae494

W niniejszej pracy wykorzystałam nowszą metodę mapowania chemicznego – SHAPE-MaP – wraz z zaawansowanymi narzędziami bioinformatycznymi do ustalenia struktury drugorzędowej gRNA drożdżowego retrotranspozonu Ty3 (5052 nt) w natywnym stanie *in vivo* oraz w warunkach pozakomórkowych (Rysunek 1B w publikacji). Dla uzyskania wysokiej jakości danych strukturalnych do mapowania chemicznego wykorzystałam odczynnik NAI, który cechuje się poziomem wydajności modyfikacji RNA *in vivo* optymalnym dla zapewnienia efektywnego współczynnika mutacji w podejściu *RT-mutate* (Busan i in., 2019). Optymalizacja warunków modyfikacji RNA oraz długości analizowanych amplikonów umożliwiła uzyskanie powtarzalnych danych o wysokiej rozdzielczości, co finalnie przełożyło się na efektywną reaktywność SHAPE dla 99,1% nukleotydów w gRNA Ty3 (Rysunki 1C, D w publikacji).

W pierwszej kolejności przeprowadziłam analizę porównawczą reaktywności SHAPE otrzymanych w warunkach *in vivo* i *ex vivo*. Obydwa zestawy danych SHAPE wykazały bardzo silną korelację globalną, natomiast różnica w wartościach mediany reaktywności zasugerowała nieco bardziej rozwiniętą strukturę gRNA Ty3 *in vivo* (Rysunki 2A, B w publikacji). Aby uchwycić lokalne cechy strukturalne obu stanów gRNA Ty3, przeprowadziłam obliczenia mediany reaktywności, indeksu Giniego oraz korelacji w przesuwających się oknach, a także zastosowałam algorytm statystyczny Δ SHAPE (Smola, Calabrese, i in., 2015). To kompleksowe podejście pozwoliło mi zidentyfikować statystycznie istotne różnice w wielu rejonach gRNA, połączone z miejscowymi spadkami korelacji oraz potwierdziło niższy poziom struktury gRNA w stanie *in vivo* (Rysunki 2C, D w publikacji).

Następnie otrzymane reaktywności SHAPE wykorzystałam do modelowania struktury drugorzędowej badanego gRNA z wykorzystaniem oprogramowania SuperFold (Siegfried i in., 2014). Konsensusowy model MFE struktury drugorzędowej gRNA Ty3 *in vivo* zawierał nieco mniej par zasad, w tym par G-C niż model struktury *ex vivo* (Rysunki 3A, B, oraz Tabela 1 w publikacji). Dodatkowo 31% par zasad było specyficznych tylko dla modelu *in vivo*, a 40% dla modelu *ex vivo* wskazując, że struktura gRNA w analizowanych stanach różni się bardziej niż sugerowała globalna korelacja reaktywności SHAPE. Co ciekawe, w obu stanach porównywalna liczba nukleotydów (~24%) była zaangażowana w wysoce prawdopodobne pary zasad, a ich duża część jest obecna niezależnie od środowiska (Rysunek 3D w publikacji).

Dalsza analiza miała na celu identyfikację rejonów gRNA Ty3 z najbardziej stabilnymi motywami strukturalnymi. W tym celu posłużyłam się strategią *lowSS*, która pozwoliła mi na

znalezienie 14 takich rejonów dla stanu *in vivo* oraz 12 dla stanu *ex vivo* (Rysunki 3A, B w publikacji). Zawierały one większość funkcjonalnych sekwencji Ty3, w tym 5' PBS (Gabus i in., 1998), FS (Farabaugh i in., 1993) oraz miejsce dimeryzacji (DS), zidentyfikowane w niniejszej pracy. Zauważyłam, że aż 9 rejonów stabilnych nakładało się pomiędzy stanami eksperymentalnymi, a 7 spośród nich zawierało te same motywy strukturalne, bogate z wysoce prawdopodobne pary zasad (Rysunek 3C w publikacji). Zaproponowałam, że motywy te stanowią stabilnie zwinięty rdzeń (ang. *well-folded core*), tworzący się niezależnie od środowiska komórkowego. Rdzeń obejmuje 19,4% sekwencji gRNA Ty3 i, co ciekawe, duża część tworzących go motywów gromadzi się w obrębie lub w pobliżu rejonu kodującego odwrotną transkryptazę, najbardziej ewolucyjnie zachowawczą domenę białkową wśród retroelementów (Xiong & Eickbush, 1990).

Aby lepiej zrozumieć międzycząsteczkowe interakcje RNA-RNA, które zachodzą podczas replikacji Ty3, razem z dr Julią Gumną zastosowałyśmy uproszczony system *in vitro*, który opierał się na wykorzystaniu dwóch krótkich transkryptów obejmujących końce 5' lub 3' elementu Ty3 (Ty3 5' RNA i Ty3 3' RNA) oraz białka nukleokapsydu Ty3 (NCp9) (Rysunek 4A w publikacji). Eksperymenty dr Julity Gumnej wykazały, że Ty3 3' RNA wiąże tRNA_i^{Met} dużo wydajniej niż Ty3 5' RNA (Rysunek 4B w publikacji). Przeprowadzone przeze mnie analizy SHAPE-MaP krótkich transkryptów wskazały dokładnie miejsce interakcji z tRNA_i^{Met} *in vitro* (Rysunek 4C w publikacji). Przypadało ono w obrębie 3' PBS, wyznaczonego wcześniej na podstawie funkcjonalnych analiz mutacyjnych (Gabus i in., 1998). Rozwiązanie problemu wiązania tRNA_i^{Met} w rejonie 5' końca było trudne ze względu na niską reaktywność 5' PBS we wszystkich stanach eksperymentalnych i jego tendencję do alternatywnych oddziaływań wewnątrzcząsteczkowych (Rysunki 6A, B, C i S7 w publikacji). Niemniej wyniki SHAPE uzyskane dla 3' PBS sugerują, że tylko niewielka część gRNA Ty3 oddziałuje z tRNA_i^{Met} *in vivo* i *ex vivo*.

Choć wcześniejsze badania proponowały, że związany tRNA_i^{Met} pośredniczy jednocześnie w dimeryzacji gRNA Ty3, eksperymenty przeprowadzone przez dr Julię Gumną pokazały, że Ty3 5' RNA ulega indukowanej NCp9 dimeryzacji bez pośrednictwa tRNA (Rysunek 4B w publikacji). Wynik ten zachęcił nas do poszukania sekwencji Ty3, która bierze bezpośredni udział w tworzeniu dimeru. Skupiłyśmy się na analizie sekwencji palindromowych, gdyż ich udział w bezpośrednim oddziaływaniu dwóch nici gRNA był już wcześniej pokazany dla innych retroelementów (Dubois i in., 2018; Gumna i in., 2019). Porównawcza analiza profili reaktywności z eksperymentów SHAPE-MaP dla monomeru i dimeru Ty3 5' RNA wskazała miejsca spadków reaktywności w dimerze, które mogą wskazywać na tworzenie interakcji międzycząsteczkowych (Figura 5A i S8 w publikacji). Spośród sześciu zidentyfikowanych sekwencji palindromowych, sekwencja PAL6 wskazywała najbardziej istotny spadek reaktywności. Dalsze eksperymenty pokazały, że usunięcie sekwencji PAL6 całkowicie hamuje

dimeryzację Ty3 5' RNA *in vitro*. Analizy SHAPE-MaP dla tego skróconego transkryptu nie wykazały globalnych różnic w jego strukturze wskazując, że delecja ma efekt lokalny i bezpośrednio wpływa na zahamowanie dimeryzacji (Rysunki 5B, C w publikacji). Istotnie statystycznie spadki reaktywności w obrębie PAL6 zaobserwowałam również w przypadku gRNA Ty3 *in vivo* w porównaniu do stanu *ex vivo*, którego warunki uniemożliwiały formowanie dimeru z powodu braku indukującego ten proces białka (Rysunki 6A B, C i S9 w publikacji). Wynik ten wskazuje na obecność frakcji dimerycznej *in vivo* i wzmacnia, sugerowane przez nas na podstawie danych *in vitro*, funkcjonalne znaczenie PAL6 jako sekwencji dimeryzacyjnej (DS).

Na podstawie analizy porównawczej otrzymanych danych scharakteryzowałam również kontekst strukturalny pozostałych znanych sekwencji funkcjonalnych gRNA Ty3 w stanie natywnym (Rysunki 6A, B, C w publikacji). Pokazałam, że 7-nt sekwencja FS jest niereaktywna *in vivo* i zlokalizowana w trzonie wysoce prawdopodobnego motywu spinki do włosów. Następująca po niej regulacyjna sekwencja „kontekstu” jest wysoce reaktywna i w większości niesparowana *in vivo* i *ex vivo*. Sekwencja PPT była umiarkowanie reaktywna i zaangażowana w pary zasad o niższym prawdopodobieństwie, co może przyczyniać się do jej większej dostępności dla RNazy H. Uzyskane dane strukturalne wyjaśniają pewne obserwacje z badań funkcjonalnych, jednak dokładna rola ustalonych struktur sekwencji regulatorowych w przebiegu cyklu replikacyjnego Ty3 wymaga bardziej zaawansowanych analiz funkcjonalno-strukturalnych.

Przeprowadziłam również analizę gRNA Ty3 pod kątem obecności motywów typu pseudowęzeł oraz G-kwadrupleks (Rysunek 7 w publikacji). Do tego celu zastosowałam programy ShapeKnots oraz QGRS Mapper (Hajdin i in., 2013; Kikin i in., 2006). Predykcja przeprowadzona z wykorzystaniem danych SHAPE *in vivo* wskazała na potencjalną obecność dwóch pseudowęzłów – w ORF GAG (^{GAG}PK) i ORF POL (^{POL}PK). ^{GAG}PK został również przewidziany w oparciu o dane *ex vivo*. Analiza przy użyciu QGRS Mapper wskazała jedno miejsce w gRNA Ty3 z sekwencją bogatą w guanozyny zdolną do utworzenia G-kwadrupleksu, zlokalizowane 12 nt po kodonie AUG. Wyniki obu tych analiz były zgodne z otrzymanym wzorem reaktywności SHAPE uzyskanym *in vivo* oraz *ex vivo*.

W ostatniej części pracy przeprowadziłam kompleksową analizę wspólnych i specyficznych cech strukturalnych dla gRNA trzech modelowych reprezentantów retroelementów – retrotranspozonów Ty1 (*Pseudoviridae*) i Ty3 (*Metaviridae*) oraz retrowirusa HIV-1 (*Retroviridae*). Zestawiłam, między innymi, korelację pomiędzy wzbogaceniem w guanozyny a lokalizacją sekwencji funkcjonalnych oraz rozmieszczenie rejonów wysoko ustrukturyzowanych (Rysunki 8A, C w publikacji). Zestawienie to pokazało, że charakterystyczna dla HIV-1 korelacja między składem sekwencji a lokalizacją elementów funkcjonalnych, jest częściowo zachowana dla Ty3, ale nie dla Ty1. Dla otrzymanych przeze mnie modeli struktury gRNA Ty1 i Ty3 przeprowadziłam dodatkową analizę porównawczą liczby i dystansu tworzonych par zasad (Rysunek 8B w publikacji). Potwierdziła ona, że struktura gRNA obu elementów jest bardziej

rozluźniona *in vivo*, ale mniej różni się od stanu *ex vivo* (Ty3) niż *in vitro* gRNA (Ty1). W odróżnieniu od HIV-1, dla obu retrotranspozonów nie zaobserwowałam tendencji do lokalnego wzrostu poziomu struktury gRNA na granicach kodowanych domen białkowych (Rysunek S10 w publikacji).

8 PODSUMOWANIE

Dwie prace oryginalne współtworzące niniejszą rozprawę doktorską (Andrzejewska i in., 2021; Andrzejewska-Romanowska i in., 2024) dostarczają pierwszej charakterystyki struktury drugorzędowej genomów RNA retrotranspozonów LTR w warunkach *in vivo*. Przeprowadzona przeze mnie kompleksowa analiza porównawcza danych strukturalnych otrzymanych w różnych stanach eksperymentalnych dla reprezentantów dwóch rodzin retrotranspozonów drożdżowych – Ty1 (*Pseudoviridae*) i Ty3 (*Metaviridae*), znacząco wzbogaca wiedzę na temat różnorodności strukturalnej genomów RNA i ich relacji ze środowiskiem gospodarza. Uważam, że stanowi to solidną podstawę do dalszego badania zależności między strukturą genomów RNA a ich funkcją w replikacji retroelementów. Wyniki moich badań są również wartościowe dla zrozumienia problemu zwijania się długich, wielofunkcyjnych RNA w żywych komórkach. Prace eksperymentalne poprzedza praca przeglądowa (Andrzejewska i in., 2020), która stanowi wnikliwą analizę i usystematyzowanie ówczesnej wiedzy na temat wpływu środowiska komórkowego na zwijanie cząsteczek mRNA oraz istniejących zależności między strukturą mRNA a jego funkcją w komórkach różnych organizmów.

Najważniejszymi rezultatami badań eksperymentalnych i analiz bioinformatycznych przedstawionych w niniejszej rozprawie doktorskiej są:

- dostarczenie pierwszych, wspartych eksperymentalnie, modeli struktury drugorzędowej genomów RNA drożdżowych retrotranspozonów Ty1 *in vivo* oraz Ty3 *in vivo* i *ex vivo*,
- identyfikacja cech strukturalnych gRNA Ty1 i Ty3 zależnych i niezależnych od środowiska komórki, a także najbardziej stabilnych motywów RNA,
- wskazanie znaczącego udziału aktywnych rybosomów w rozwijaniu i remodelowaniu struktury gRNA Ty1,
- szczegółowa charakterystyka kontekstu strukturalnego sekwencji funkcjonalnych w gRNA Ty1 i Ty3 w warunkach *in vivo*,
- zaproponowanie nowego mechanizmu dimeryzacji gRNA Ty3 (we współpracy z dr Julią Gumną),
- identyfikacja cech strukturalnych wspólnych lub odrębnych dla genomów RNA reprezentantów trzech ewolucyjnie powiązanych rodzin retroelementów,
- udowodnienie, że NMIA może być z powodzeniem stosowany do badań struktury RNA w drożdżach i bakteriach,
- wprowadzenie eksperymentów i analizy danych SHAPE-MaP do wachlarza metod badania struktury RNA w Zakładzie Struktury i Funkcji RNA ICHB PAN.

9 WYKAZ SKRÓTÓW

Wykaz obejmuje skróty nieobjaśnione w tekście pracy

CHIKV	wirus Chikungunya (ang. <i>Chikungunya virus</i>)
ddATP	trifosforan 2',3'-dideoksyadenozyny
ddCTP	trifosforan 2',3'-dideoksyctozyny
ddNTP	dideoksynukleotydy
DNA	kwas deoksyrybonukleinowy (ang. <i>deoxyribonucleic acid</i>)
dsDNA	dwuniciowy DNA (ang. <i>double-stranded</i>)
HCV	wirus zapalenia wątroby typu C (ang. <i>hepatitis C virus</i>)
HIV-1	ludzki wirus niedoboru odporności typu 1 (ang. <i>human immunodeficiency virus type 1</i>)
LINE	długie rozproszone elementy jądrowe (ang. <i>long interspersed nuclear elements</i>)
mRNA	informacyjny RNA (ang. <i>messenger RNA</i>)
R	rejon powtórzonej sekwencji w LTR lub UTR retroelementów (ang. <i>repetitive</i>)
RNA	kwas rybonukleinowy (ang. <i>ribonucleic acid</i>)
RVA	rotawirus grupy A (ang. <i>group A rotavirus</i>)
SARS-CoV-2	drugi koronawirus ciężkiego ostrego zespołu oddechowego (ang. <i>severe acute respiratory syndrome coronavirus 2</i>)
SINE	krótkie rozproszone elementy jądrowe (ang. <i>short interspersed nuclear elements</i>)
tRNA	transportujący RNA (ang. <i>transfer RNA</i>)
U3, U5	rejony unikalnej sekwencji w LTR lub UTR retroelementów (ang. <i>unique</i>)
UTR	rejon nieulegający translacji (ang. <i>untranslated region</i>)

10 BIBLIOGRAFIA

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
- Andrzejewska, A., Zawadzka, M., Gumna, J., Garfinkel, D. J., & Pachulska-Wieczorek, K. (2021). In vivo structure of the Ty1 retrotransposon RNA genome. *Nucleic Acids Research*, 49(5), 2878–2893. <https://doi.org/10.1093/nar/gkab090>
- Andrzejewska, A., Zawadzka, M., & Pachulska-Wieczorek, K. (2020). On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective. *International Journal of Molecular Sciences*, 21(18), 6770. <https://doi.org/10.3390/ijms21186770>
- Andrzejewska-Romanowska, A., Gumna, J., Tykwińska, E., & Pachulska-Wieczorek, K. (2024). Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome. *Nucleic Acids Research*, gkae494. <https://doi.org/10.1093/nar/gkae494>
- Assmann, S. M., Chou, H.-L., & Bevilacqua, P. C. (2023). Rock, scissors, paper: How RNA structure informs function. *The Plant Cell*, 35(6), 1671–1707. <https://doi.org/10.1093/plcell/koad026>
- Aviran, S., & Incarnato, D. (2022). Computational approaches for RNA structure ensemble deconvolution from structure probing data. *Journal of Molecular Biology*, 434(18), 167635. <https://doi.org/10.1016/j.jmb.2022.167635>
- Aw, J. G. A., Lim, S. W., Wang, J. X., Lambert, F. R. P., Tan, W. T., Shen, Y., Zhang, Y., Kaewsapsak, P., Li, C., Ng, S. B., Vardy, L. A., Tan, M. H., Nagarajan, N., & Wan, Y. (2021). Determination of isoform-specific RNA structure with nanopore long reads. *Nature Biotechnology*, 39(3), 336–346. <https://doi.org/10.1038/s41587-020-0712-z>
- Aw, J. G. A., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K.-L., Tapsin, S., Chan, Y.-S., Tan, C.-P., Sim, A. Y. L., Zhang, T., Susanto, T. T., Fu, Z., Nagarajan, N., & Wan, Y. (2016). In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Molecular Cell*, 62(4), 603–617. <https://doi.org/10.1016/j.molcel.2016.04.028>
- Beaudoin, J.-D., Novoa, E. M., Vejnar, C. E., Yartseva, V., Takacs, C. M., Kellis, M., & Giraldez, A. J. (2018). Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nature Structural & Molecular Biology*, 25(8), 677–686. <https://doi.org/10.1038/s41594-018-0091-z>
- Belcourt, M. F., & Farabaugh, P. J. (1990). Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell*, 62(2), 339–352. [https://doi.org/10.1016/0092-8674\(90\)90371-K](https://doi.org/10.1016/0092-8674(90)90371-K)
- Bizuayehu, T. T., Labun, K., Jakubec, M., Jefimov, K., Niazi, A. M., & Valen, E. (2022). Long-read single-molecule RNA structure sequencing using nanopore. *Nucleic Acids Research*, gkac775. <https://doi.org/10.1093/nar/gkac775>
- Bohn, P., Gribling-Burrer, A.-S., Ambi, U. B., & Smyth, R. P. (2023). Nano-DMS-MaP allows isoform-specific RNA structure determination. *Nature Methods*, 20(6), 849–859. <https://doi.org/10.1038/s41592-023-01862-7>

- Bonchev, G. N. (2016). Useful parasites: The evolutionary biology and biotechnology applications of transposable elements. *Journal of Genetics*, 95(4), 1039–1052. <https://doi.org/10.1007/s12041-016-0702-6>
- Bonnet, A., & Lesage, P. (2021). Light and shadow on the mechanisms of integration site selection in yeast Ty retrotransposon families. *Current Genetics*, 67(3), 347–357. <https://doi.org/10.1007/s00294-021-01154-7>
- Bose, R., Saleem, I., & Mustoe, A. M. (2024). Causes, functions, and therapeutic possibilities of RNA secondary structure ensembles and alternative states. *Cell Chemical Biology*. <https://doi.org/10.1016/j.chembiol.2023.12.010>
- Bugnon, L. A., Di Persia, L., Gerard, M., Raad, J., Prochetto, S., Fenoy, E., Chorostecki, U., Ariel, F., Stegmayer, G., & Milone, D. H. (2024). sincFold: End-to-end learning of short- and long-range interactions in RNA secondary structure. *Briefings in Bioinformatics*, 25(4), bbae271. <https://doi.org/10.1093/bib/bbae271>
- Burkhardt, D. H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J. S., & Gross, C. A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *eLife*, 6, e22037. <https://doi.org/10.7554/eLife.22037>
- Busan, S., & Weeks, K. M. (2018). Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, 24(2), 143–148. <https://doi.org/10.1261/rna.061945.117>
- Busan, S., Weidmann, C. A., Sengupta, A., & Weeks, K. M. (2019). Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry*, 58(23), 2655–2664. <https://doi.org/10.1021/acs.biochem.8b01218>
- Cordero, P., Kladwang, W., VanLang, C. C., & Das, R. (2012). Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry*, 51(36), 7037–7039. <https://doi.org/10.1021/bi3008802>
- Coria, A., Wienecke, A., Knight, M. L., Desirò, D., Laederach, A., & Borodavka, A. (2022). Rotavirus RNA chaperone mediates global transcriptome-wide increase in RNA backbone flexibility. *Nucleic Acids Research*, gkac738. <https://doi.org/10.1093/nar/gkac738>
- Cristofari, G., Bampi, C., Wilhelm, M., Wilhelm, F.-X., & Darlix, J.-L. (2002). A 5'-3' long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *The EMBO Journal*, 21(16), 4368–4379. <https://doi.org/10.1093/emboj/cdf436>
- Cristofari, G., Ficheux, D., & Darlix, J.-L. (2000). The Gag-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. *Journal of Biological Chemistry*, 275(25), 19210–19217. <https://doi.org/10.1074/jbc.M001371200>
- Curcio, M. J. (2019). Border collies of the genome: Domestication of an autonomous retrovirus-like transposon. *Current Genetics*, 65(1), 71–78. <https://doi.org/10.1007/s00294-018-0857-1>
- Curcio, M. J., Lutz, S., & Lesage, P. (2015). The Ty1 LTR-retrotransposon of budding yeast, *Saccharomyces cerevisiae*. *Microbiology spectrum*, 3(2), 1–35. <https://doi.org/10.1128/microbiolspec.MDNA3-0053-2014>
- Dash, C., Marino, J. P., & Le Grice, S. F. J. (2006). Examining Ty3 polypurine tract structure and function by nucleoside analog interference. *Journal of Biological Chemistry*, 281(5), 2773–2783. <https://doi.org/10.1074/jbc.M510369200>

- Deigan, K. E., Li, T. W., Mathews, D. H., & Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences*, 106(1), 97–102. <https://doi.org/10.1073/pnas.0806929106>
- Deng, H., Cheema, J., Zhang, H., Woolfenden, H., Norris, M., Liu, Z., Liu, Q., Yang, X., Yang, M., Deng, X., Cao, X., & Ding, Y. (2018). Rice in vivo RNA structurome reveals RNA secondary structure conservation and divergence in plants. *Molecular Plant*, 11(4), 607–622. <https://doi.org/10.1016/j.molp.2018.01.008>
- Dethoff, E. A., Boerneke, M. A., Gokhale, N. S., Muhire, B. M., Martin, D. P., Sacco, M. T., McFadden, M. J., Weinstein, J. B., Messer, W. B., Horner, S. M., & Weeks, K. M. (2018). Pervasive tertiary structure in the dengue virus RNA genome. *Proceedings of the National Academy of Sciences*, 115(45), 11513–11518. <https://doi.org/10.1073/pnas.1716689115>
- Devine, S. E., & Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes & Development*, 10(5), 620–633. <https://doi.org/10.1101/gad.10.5.620>
- Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, 505(7485), 696–700. <https://doi.org/10.1038/nature12756>
- Dodonova, S. O., Prinz, S., Bilanchone, V., Sandmeyer, S., & Briggs, J. A. G. (2019). Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proceedings of the National Academy of Sciences*, 116(20), 10048–10057. <https://doi.org/10.1073/pnas.1900931116>
- Douds, C. A., Babitzke, P., & Bevilacqua, P. (2024). A new reagent for in vivo structure probing of RNA G and U residues that improves RNA structure prediction alone and combined with DMS. *RNA*, rna.079974.124. <https://doi.org/10.1261/rna.079974.124>
- Dubois, N., Marquet, R., Paillart, J.-C., & Bernacchi, S. (2018). Retroviral RNA dimerization: from structure to functions. *Frontiers in Microbiology*, 9, 527. <https://doi.org/10.3389/fmicb.2018.00527>
- Farabaugh, P. J., Zhao, H., & Vimaladithan, A. (1993). A novel programmed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: Frameshifting without tRNA slippage. *Cell*, 74(1), 93–103. [https://doi.org/10.1016/0092-8674\(93\)90297-4](https://doi.org/10.1016/0092-8674(93)90297-4)
- Gabus, C., Ficheux, D., Rau, M., Keith, G., Sandmeyer, S., & Darlix, J. L. (1998). The yeast Ty3 retrotransposon contains a 5'-3' bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. *The EMBO Journal*, 17(16), 4873–4880. <https://doi.org/10.1093/emboj/17.16.4873>
- Gamache, E., Doh, J., Ritz, J., Laederach, A., Bellaousov, S., Mathews, D., & Curcio, M. (2017). Structure-function model for kissing loop interactions that initiate dimerization of Ty1 RNA. *Viruses*, 9(5), 93. <https://doi.org/10.3390/v9050093>
- Giannetti, C. A., Busan, S., Weidmann, C. A., & Weeks, K. M. (2019). SHAPE probing reveals human rRNAs are largely unfolded in solution. *Biochemistry*, 58(31), 3377–3385. <https://doi.org/10.1021/acs.biochem.9b00076>
- Gumna, J., Purzycka, K. J., Ahn, H. W., Garfinkel, D. J., & Pachulska-Wieczorek, K. (2019). Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biology*, 16(12), 1749–1763. <https://doi.org/10.1080/15476286.2019.1657370>

- Gumna, J., Zok, T., Figurski, K., Pachulska-Wieczorek, K., & Szachniuk, M. (2020). RNATHOR – 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>
- Guo, L.-T., Adams, R. L., Wan, H., Huston, N. C., Potapova, O., Olson, S., Gallardo, C. M., Graveley, B. R., Torbett, B. E., & Pyle, A. M. (2020). Sequencing and structure probing of long RNAs using MarathonRT: a next-generation reverse transcriptase. *Journal of Molecular Biology*, 432(10), 3338–3352. <https://doi.org/10.1016/j.jmb.2020.03.022>
- Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., & Weeks, K. M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences*, 110(14), 5498–5503. <https://doi.org/10.1073/pnas.1219988110>
- Havecker, E. R., Gao, X., & Voytas, D. F. (2004). The diversity of LTR retrotransposons. *Genome Biology*, 5, 225. <https://doi.org/10.1186/gb-2004-5-6-225>
- Homan, P. J., Favorov, O. V., Lavender, C. A., Kursun, O., Ge, X., Busan, S., Dokholyan, N. V., & Weeks, K. M. (2014). Single-molecule correlated chemical probing of RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 111(38), 13858–13863. <https://doi.org/10.1073/pnas.1407306111>
- Huang, Q., Purzycka, K. J., Lusvardi, S., Li, D., LeGrice, S. F. J., & Boeke, J. D. (2013). Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA*, 19(3), 320–332. <https://doi.org/10.1261/rna.035535.112>
- Huber, R. G., Lim, X. N., Ng, W. C., Sim, A. Y. L., Poh, H. X., Shen, Y., Lim, S. Y., Sundstrom, K. B., Sun, X., Aw, J. G., Too, H. K., Boey, P. H., Wilm, A., Chawla, T., Choy, M. M., Jiang, L., de Sessions, P. F., Loh, X. J., Alonso, S., ... Wan, Y. (2019). Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nature Communications*, 10(1), Article 1. <https://doi.org/10.1038/s41467-019-09391-8>
- Huston, N. C., Wan, H., Strine, M. S., de Cesaris Araujo Tavares, R., Wilen, C. B., & Pyle, A. M. (2021). Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Molecular Cell*, 81(3), 584-598.e5. <https://doi.org/10.1016/j.molcel.2020.12.041>
- Incarnato, D., Morandi, E., Anselmi, F., Simon, L. M., Basile, G., & Oliviero, S. (2017). In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding. *Nucleic Acids Research*, 45(16), 9716–9725. <https://doi.org/10.1093/nar/gkx617>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Justyna, M., Antczak, M., & Szachniuk, M. (2023). Machine learning for RNA 2D structure prediction benchmarked on experimental data. *Briefings in Bioinformatics*, 24(3), bbad153. <https://doi.org/10.1093/bib/bbad153>
- Kaneko-Ishino, T., & Ishino, F. (2012). The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Frontiers in Microbiology*, 3, 262. <https://doi.org/10.3389/fmicb.2012.00262>

- Keeney, J. B., Chapman, K. B., Lauermaun, V., Voytas, D. F., Aström, S. U., von Pawel-Rammingen, U., Byström, A., & Boeke, J. D. (1995). Multiple molecular determinants for retrotransposition in a primer tRNA. *Molecular and Cellular Biology*, 15(1), 217–226. <https://doi.org/10.1128/MCB.15.1.217>
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., & Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 467(7311), 103–107. <https://doi.org/10.1038/nature09322>
- Kikin, O., D'Antonio, L., & Bagga, P. S. (2006). QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research*, 34(Web Server issue), W676–682. <https://doi.org/10.1093/nar/gkl253>
- Krupovic, M., Blomberg, J., Coffin, J. M., Dasgupta, I., Fan, H., Geering, A. D., Gifford, R., Harrach, B., Hull, R., Johnson, W., Kreuze, J. F., Lindemann, D., Llorens, C., Lockhart, B., Mayer, J., Muller, E., Olszewski, N. E., Pappu, H. R., Pooggin, M. M., ... Kuhn, J. H. (2018). Ortervirales: new virus order unifying five families of reverse-transcribing viruses. *Journal of Virology*, 92(12), e00515–18. <https://doi.org/10.1128/JVI.00515-18>
- Kubota, M., Tran, C., & Spitale, R. C. (2015). Progress and challenges for chemical probing of RNA structure inside living cells. *Nature Chemical Biology*, 11(12), 933–941. <https://doi.org/10.1038/nchembio.1958>
- Lan, T. C. T., Allan, M. F., Malsick, L. E., Woo, J. Z., Zhu, C., Zhang, F., Khandwala, S., Nyeo, S. S. Y., Sun, Y., Guo, J. U., Bathe, M., Näär, A., Griffiths, A., & Rouskin, S. (2022). Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nature Communications*, 13(1), 1128. <https://doi.org/10.1038/s41467-022-28603-2>
- Larman, B. C., Dethoff, E. A., & Weeks, K. M. (2017). Packaged and free satellite Tobacco Mosaic Virus (STMV) RNA genomes adopt distinct conformational states. *Biochemistry*, 56(16), 2175–2183. <https://doi.org/10.1021/acs.biochem.6b01166>
- Lener, D., Kvaratskhelia, M., & Le Grice, S. F. J. (2003). Nonpolar thymine isosteres in the Ty3 polypurine tract DNA template modulate processing and provide a model for its recognition by Ty3 reverse transcriptase. *Journal of Biological Chemistry*, 278(29), 26526–26532. <https://doi.org/10.1074/jbc.M302374200>
- Lesage, P., & Todeschini, A. L. (2005). Happy together: The life and times of Ty retrotransposons and their hosts. *Cytogenetic and Genome Research*, 110(1–4), 70–90. <https://doi.org/10.1159/000084940>
- Li, P., Wei, Y., Mei, M., Tang, L., Sun, L., Huang, W., Zhou, J., Zou, C., Zhang, S., Qin, C.-F., Jiang, T., Dai, J., Tan, X., & Zhang, Q. C. (2018). Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host & Microbe*, 24(6), 875–886.e5. <https://doi.org/10.1016/j.chom.2018.10.011>
- Llorens, C., Fares, M. A., & Moya, A. (2008). Relationships of Gag-pol diversity between Ty3/Gypsy and Retroviridae LTR retroelements and the three kings hypothesis. *BMC Evolutionary Biology*, 8(1), 276. <https://doi.org/10.1186/1471-2148-8-276>
- Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., Pachter, L., Doudna, J. A., & Arkin, A. P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proceedings of the National Academy of Sciences of the United States of America*, 108(27), 11063–11068. <https://doi.org/10.1073/pnas.1106501108>

- Madden, E. A., Plante, K. S., Morrison, C. R., Kutchko, K. M., Sanders, W., Long, K. M., Taft-Benz, S., Cruz Cisneros, M. C., White, A. M., Sarkar, S., Reynolds, G., Vincent, H. A., Laederach, A., Moorman, N. J., & Heise, M. T. (2020). Using SHAPE-MaP to model RNA secondary structure and identify 3'UTR variation in Chikungunya virus. *Journal of Virology*, 94(24), 10.1128/jvi.00701-20. <https://doi.org/10.1128/jvi.00701-20>
- Malagon, F., & Jensen, T. H. (2011). T-body formation precedes virus-like particle maturation in *S. cerevisiae*. *RNA Biology*, 8(2), 184–189. <https://doi.org/10.4161/rna.8.2.14822>
- Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T. K., Marinus, T., Ogando, N. S., Snijder, E. J., van Hemert, M. J., Bujnicki, J. M., & Incarnato, D. (2020). Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Research*, 48(22), 12436–12452. <https://doi.org/10.1093/nar/gkaa1053>
- Marinus, T., Fessler, A. B., Ogle, C. A., & Incarnato, D. (2021). A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy. *Nucleic Acids Research*, 49(6), e34. <https://doi.org/10.1093/nar/gkaa1255>
- Mathews, D. H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, 10(8), 1178–1190. <https://doi.org/10.1261/rna.7650904>
- Mauger, D. M., Cabral, B. J., Presnyak, V., Su, S. V., Reid, D. W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M. J., & McFadyen, I. J. (2019). mRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences*, 116(48), 24075–24083. <https://doi.org/10.1073/pnas.1908052116>
- Merino, E. J., Wilkinson, K. A., Coughlan, J. L., & Weeks, K. M. (2005). RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society*, 127(12), 4223–4231. <https://doi.org/10.1021/ja043822v>
- Mita, P., & Boeke, J. D. (2016). How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development*, 37, 90–100. <https://doi.org/10.1016/j.gde.2016.01.001>
- Mitchell, D., III, Cotter, J., Saleem, I., & Mustoe, A. M. (2023). Mutation signature filtering enables high-fidelity RNA structure probing at all four nucleobases with DMS. *Nucleic Acids Research*, 51(16), 8744–8757. <https://doi.org/10.1093/nar/gkad522>
- Moqtaderi, Z., Geisberg, J. V., & Struhl, K. (2018). Extensive structural differences of closely related 3' mRNA isoforms: links to Pab1 binding and mRNA stability. *Molecular Cell*, 72(5), 849–861.e6. <https://doi.org/10.1016/j.molcel.2018.08.044>
- Mustoe, A. M., Busan, S., Rice, G. M., Hajdin, C. E., Peterson, B. K., Ruda, V. M., Kubica, N., Nutiu, R., Baryza, J. L., & Weeks, K. M. (2018). Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell*, 173(1), 181–195.e18. <https://doi.org/10.1016/j.cell.2018.02.034>
- Mustoe, A. M., Lama, N. N., Irving, P. S., Olson, S. W., & Weeks, K. M. (2019). RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proceedings of the National Academy of Sciences*, 116(49), 24574–24582. <https://doi.org/10.1073/pnas.1905491116>
- Mustoe, A. M., Weidmann, C. A., & Weeks, K. M. (2023). Single-molecule correlated chemical probing: A revolution in RNA structure analysis. *Accounts of chemical research*, 56(7), 763–775. <https://doi.org/10.1021/acs.accounts.2c00782>

- Nowak, E., Miller, J. T., Bona, M. K., Studnicka, J., Szczepanowski, R. H., Jurkowski, J., Le Grice, S. F. J., & Nowotny, M. (2014). Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nature structural & molecular biology*, 21(4), 389–396. <https://doi.org/10.1038/nsmb.2785>
- Orlinsky, K. J., & Sandmeyer, S. B. (1994). The Cys-His motif of Ty3 NC can be contributed by Gag3 or Gag3-Pol3 polyproteins. *Journal of Virology*, 68(7), 4152–4166. <https://doi.org/10.1128/jvi.68.7.4152-4166.1994>
- Pastuzyn, E. D., Day, C. E., Kearns, R. B., Kyrke-Smith, M., Taibi, A. V., McCormick, J., Yoder, N., Belnap, D. M., Erlendsson, S., Morado, D. R., Briggs, J. A. G., Feschotte, C., & Shepherd, J. D. (2018). The neuronal gene ARC encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. *Cell*, 172(1–2), 275–288.e18. <https://doi.org/10.1016/j.cell.2017.12.024>
- Petrov, A. S., Bernier, C. R., Hsiao, C., Norris, A. M., Kovacs, N. A., Waterbury, C. C., Stepanov, V. G., Harvey, S. C., Fox, G. E., Wartell, R. M., Hud, N. V., & Williams, L. D. (2014). Evolution of the ribosome at atomic resolution. *Proceedings of the National Academy of Sciences*, 111(28), 10251–10256. <https://doi.org/10.1073/pnas.1407205111>
- Purzycka, K. J., Legiewicz, M., Matsuda, E., Eizentstat, L. D., Lusvarghi, S., Saha, A., Grice, S. F. J. L., & Garfinkel, D. J. (2013). Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Research*, 41(1), 463–473. <https://doi.org/10.1093/nar/gks983>
- Qi, X., Daily, K., Nguyen, K., Wang, H., Mayhew, D., Rigor, P., Forouzan, S., Johnston, M., Mitra, R. D., Baldi, P., & Sandmeyer, S. (2012). Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Research*, 22(4), 681–692. <https://doi.org/10.1101/gr.131219.111>
- Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11(1), 129. <https://doi.org/10.1186/1471-2105-11-129>
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., & Weissman, J. S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505(7485), 701–705. <https://doi.org/10.1038/nature12894>
- Sandmeyer, S. B., & Clemens, K. A. (2010). Function of a retrotransposon nucleocapsid protein. *RNA Biology*, 7(6), 642–654. <https://doi.org/10.4161/rna.7.6.14117>
- Sandmeyer, S., Patterson, K., & Bilanchone, V. (2015). Ty3, a position-specific retrotransposon in budding yeast. *Microbiology Spectrum*, 3(2), MDNA3-0057–2014. <https://doi.org/10.1128/microbiolspec.MDNA3-0057-2014>
- Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E., & Weeks, K. M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature Methods*, 11(9), 959–965. <https://doi.org/10.1038/nmeth.3029>
- Smola, M. J., Calabrese, J. M., & Weeks, K. M. (2015). Detection of RNA–protein interactions in living cells with SHAPE. *Biochemistry*, 54(46), 6867–6875. <https://doi.org/10.1021/acs.biochem.5b00977>
- Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., Keene, J. D., Lee, D. M., Calabrese, J. M., & Weeks, K. M. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proceedings of the National Academy of Sciences*, 113(37), 10322–10327. <https://doi.org/10.1073/pnas.1600008113>

- Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., & Weeks, K. M. (2015). Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols*, 10(11), 1643–1669. <https://doi.org/10.1038/nprot.2015.103>
- Smola, M. J., & Weeks, K. M. (2018). In-cell RNA structure probing with SHAPE-MaP. *Nature Protocols*, 13(6), 1181–1195. <https://doi.org/10.1038/nprot.2018.010>
- Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., & Chang, H. Y. (2013). RNA SHAPE analysis in living cells. *Nature Chemical Biology*, 9(1), 18–20. <https://doi.org/10.1038/nchembio.1131>
- Spitale, R. C., & Incarnato, D. (2023). Probing the dynamic RNA structurome and its functions. *Nature Reviews Genetics*, 24(3), 178–196. <https://doi.org/10.1038/s41576-022-00546-w>
- Stephenson, W., Razaghi, R., Busan, S., Weeks, K. M., Timp, W., & Smibert, P. (2022). Direct detection of RNA modifications and structure using single-molecule nanopore sequencing. *Cell Genomics*, 2(2), 100097. <https://doi.org/10.1016/j.xgen.2022.100097>
- Strobel, E. J., Yu, A. M., & Lucks, J. B. (2018). High-throughput determination of RNA structures. *Nature Reviews Genetics*, 19(10), 615–634. <https://doi.org/10.1038/s41576-018-0034-x>
- Sun, L., Fazal, F. M., Li, P., Broughton, J. P., Lee, B., Tang, L., Huang, W., Kool, E. T., Chang, H. Y., & Zhang, Q. C. (2019). RNA structure maps across mammalian cellular compartments. *Nature Structural & Molecular Biology*, 26(4), 322–330. <https://doi.org/10.1038/s41594-019-0200-7>
- Sun, L., Li, P., Ju, X., Rao, J., Huang, W., Ren, L., Zhang, S., Xiong, T., Xu, K., Zhou, X., Gong, M., Miska, E., Ding, Q., Wang, J., & Zhang, Q. C. (2021). In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell*, 184(7), 1865–1883.e20. <https://doi.org/10.1016/j.cell.2021.02.008>
- Tijerina, P., Mohr, S., & Russell, R. (2007). DMS footprinting of structured RNAs and RNA-protein complexes. *Nature Protocols*, 2(10), 2608–2623. <https://doi.org/10.1038/nprot.2007.380>
- Tomezsko, P. J., Corbin, V. D. A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M. D., Mcintosh, L., Papenfuss, A. T., Emery, A., Swanstrom, R., Zang, T., Lan, T. C. T., Bieniasz, P., Kuritzkes, D. R., Tsbiris, A., & Rouskin, S. (2020). Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature*, 582(7812), 438–442. <https://doi.org/10.1038/s41586-020-2253-5>
- Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M., & Giddings, M. C. (2008). ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA (New York, N.Y.)*, 14(10), 1979–1990. <https://doi.org/10.1261/rna.1166808>
- Vicens, Q., & Kieft, J. S. (2022). Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences*, 119(17), e2112677119. <https://doi.org/10.1073/pnas.2112677119>
- Volff, J., Körting, C., & Scharl, M. (2001). Ty3/Gypsy retrotransposon fossils in mammalian genomes: Did they evolve into new cellular functions? *Molecular Biology and Evolution*, 18(2), 266–270. <https://doi.org/10.1093/oxfordjournals.molbev.a003801>
- Wan, H., Adams, R. L., Lindenbach, B. D., & Pyle, A. M. (2022). The in vivo and in vitro architecture of the Hepatitis C virus RNA genome uncovers functional RNA secondary and tertiary structures. *Journal of Virology*, e01946-21. <https://doi.org/10.1128/jvi.01946-21>

- Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T., & Lucks, J. B. (2016). Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nature Structural & Molecular Biology*, 23(12), 1124–1131. <https://doi.org/10.1038/nsmb.3316>
- Weeks, K. M. (2021). SHAPE directed discovery of new functions in large RNAs. *Accounts of Chemical Research*, 54(10), 2502–2517. <https://doi.org/10.1021/acs.accounts.1c00118>
- Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., & Weeks, K. M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*, 6(4), e96. <https://doi.org/10.1371/journal.pbio.0060096>
- Wu, K. E., Zou, J. Y., & Chang, H. (2023). Machine learning modeling of RNA structures: Methods, challenges and future perspectives. *Briefings in Bioinformatics*, 24(4), bbad210. <https://doi.org/10.1093/bib/bbad210>
- Wu, X., & Bartel, D. P. (2017). Widespread influence of 3'-end structures on mammalian mRNA processing and stability. *Cell*, 169(5), 905-917.e11. <https://doi.org/10.1016/j.cell.2017.04.036>
- Xia, T., SantaLucia, J. Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., & Turner, D. H. (1998). Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, 37(42), 14719–14735. <https://doi.org/10.1021/bi9809425>
- Xiong, Y., & Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO Journal*, 9(10), 3353–3362. <https://doi.org/10.1002/j.1460-2075.1990.tb07536.x>
- Xu, B., Zhu, Y., Cao, C., Chen, H., Jin, Q., Li, G., Ma, J., Yang, S. L., Zhao, J., Zhu, J., Ding, Y., Fang, X., Jin, Y., Kwok, C. K., Ren, A., Wan, Y., Wang, Z., Xue, Y., Zhang, H., ... Zhou, Y. (2022). Recent advances in RNA structurome. *Science China Life Sciences*, 65(7), 1285–1324. <https://doi.org/10.1007/s11427-021-2116-2>
- Yu, H., Qi, Y., & Ding, Y. (2022). Deep learning in RNA structure studies. *Frontiers in Molecular Biosciences*, 9. <https://doi.org/10.3389/fmolb.2022.869601>
- Zawadzka, M., & Pachulska-Wieczorek, K. (2019). Na dobre i na złe: Rola endogennych retroelementów u człowieka. *Postępy Biochemii*, 65(3), 217–223. https://doi.org/10.18388/pb.2019_272
- Zhang, J., Fei, Y., Sun, L., & Zhang, Q. C. (2022). Advances and opportunities in RNA structure experimental determination and computational modeling. *Nature Methods*, 19(10), 1193–1207. <https://doi.org/10.1038/s41592-022-01623-y>
- Zhang, S., Li, J., & Chen, S.-J. (2024). Machine learning in RNA structure prediction: Advances and challenges. *Biophysical Journal*. <https://doi.org/10.1016/j.bpj.2024.01.026>
- Zhang, Y., Huang, K., Xie, D., Lau, J. Y., Shen, W., Li, P., Wang, D., Zou, Z., Shi, S., Ren, H., Wang, Y., Mao, Y., Jin, M., Kudla, G., & Zhao, Z. (2021). In vivo structure and dynamics of the SARS-CoV-2 RNA genome. *Nature Communications*, 12(1), 5695. <https://doi.org/10.1038/s41467-021-25999-1>
- Zhang, Y., Lu, L., & Li, X. (2022). Detection technologies for RNA modifications. *Experimental & Molecular Medicine*, 54(10), 1601–1616. <https://doi.org/10.1038/s12276-022-00821-0>

Załącznik nr 1 do rozprawy doktorskiej pt.

„Charakterystyka dynamiki strukturalnej i funkcjonalnej genomów RNA aktywnych retrotranspozonów LTR”

ŻYCIORYS NAUKOWY
ANGELIKA ANDRZEJEWSKA-ROMANOWSKA

1. EDUKACJA I DOŚWIADCZENIE ZAWODOWE

07.2018 - obecnie **Studia III stopnia** (doktoranckie), Środowiskowe Studium Doktoranckie Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk, Poznań, Polska, **profil biologiczny**.

Tytuł rozprawy doktorskiej: *Charakterystyka dynamiki strukturalnej i funkcjonalnej genomów RNA aktywnych retrotranspozonów LTR*

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

Praca realizowana w Zakładzie Struktury i Funkcji RNA (poprzednio Zakład Struktury i Funkcji Retrotranspozonów), Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, Poznań, Polska.

01.2017 - 05.2018 **Studia II stopnia** (magisterskie), Uniwersytet Przyrodniczy w Poznaniu, Poznań, Polska, **biotechnologia**, specjalność - diagnostyka genetyczna.

Tytuł pracy magisterskiej: *Analiza częstości występowania wariantu c.3020insC w genie NOD2 u chorych z nieswoistymi zapalnymi chorobami jelit i w populacji polskiej*

Promotor: dr Marta Kaczmarek-Ryś

Praca realizowana w Zakładzie Funkcji Kwasów Nukleinowych, Instytut Genetyki Człowieka Polskiej Akademii Nauk, Poznań, Polska.

10-2013 - 01.2017 **Studia I stopnia** (inżynierskie), Uniwersytet Przyrodniczy w Poznaniu, Poznań, Polska, **biotechnologia**.

Tytuł pracy inżynierskiej: *Analiza wybranych polimorfizmów genu receptora kannabinoidowego CNR1 w populacji polskiej*

Promotor: dr inż. Szymon Hryhorowicz

Praca realizowana w Zakładzie Funkcji Kwasów Nukleinowych, Instytut Genetyki Człowieka Polskiej Akademii Nauk, Poznań, Polska).

2. KRAJOWE I ZAGRANICZNE STAŻE NAUKOWE

2018 Laboratorium Replikacji DNA i Integralności Genomu (prof. Rodrigo Bermejo Moreno), Centrum Badań Biologicznych (Centro de Investigaciones Biológicas – CIB, CSIC), Madryt, Hiszpania (2 miesiące).

2017 Zakład Fenomiki Zbóż (prof. Anetta Kuczyńska), Instytut Genetyki Roślin Polskiej Akademii Nauk, Poznań, Polska (1 miesiąc).

2015 Zakład Biologii Molekularnej i Systemowej (prof. Maciej Figiel), Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, Poznań, Polska (praktyki studenckie, 1 miesiąc).

3. PROJEKTY BADAWCZE

- 03.2024 – 04.2024 Wykonawca, NCN MAESTRO 8 [2016/22/A/NZ2/00184], “*Analysis of somatic mutations in miRNA genes and miRNA biogenesis genes to identify new therapy targets and biomarkers of cancer diseases*”, PI – Piotr Kozłowski.
- 10.2023 – obecnie Wykonawca, EIC PATHFINDER, “*Yeast cell factory for mRNA bioproduction*”, projekt realizowany w konsorcjum międzynarodowym: PI – Chantal Pichon, Centre National de la Recherche Scientifique, Francja, Kierownik projektu w ICHB PAN – Katarzyna Pachulska-Wieczorek.
- 03.2022 - obecnie Kierownik projektu i wykonawca, NCN PRELUDIUM 20 [2021/41/N/NZ3/04060], “*Exploring Ty3 genomic RNA structure during retrotransposition in yeast*”, PI – Angelika Andrzejewska-Romanowska.
- 10.2021 - obecnie Wykonawca (stypendium NCN), NCN OPUS 20 [2020/39/B/NZ3/0302], “*Transcriptome-wide analysis of RNA structure in the cell and across cellular compartments, and identification of the impact of cellular factors on RNA structure in S. cerevisiae*”, PI – Katarzyna Pachulska-Wieczorek.
- 07.2022 – 01.2023 Wykonawca, NCN SONATA 12 [2016/23/D/NZ1/02565], “*How does RNA structure guide messenger RNA function? Structural and functional aspects crucial for the synthesis of p53 protein - the main regulator of basic processes in the cell*”, PI – Leszek Błaszczak.
- 07.2018 – 09.2021 Wykonawca (stypendium NCN), NCN SONATA BIS 6 [2016/22/E/NZ3/00426], “*RNA structures and interactions with Gag protein that specify Ty1 RNA functions during retrotransposon replication*”, PI – Katarzyna Pachulska-Wieczorek.

4. WYRÓŻNIENIA I NAGRODY

- 11.2023 **Stypendium L'Oréal-UNESCO dla Kobiet i Nauki**, kategoria doktorancka.
- 06.2022 – 06.2023 **Stypendium START 2022** Fundacji na Rzecz Nauki Polskiej (FNP).
- 12.2022 – 06.2023 **Stypendium** dla najlepszych doktorantów, ICHB PAN.
- 11.2021 – 06.2022 **Stypendium** dla najlepszych doktorantów, ICHB PAN
- 12.2021 **Nagroda w konkursie RNA Salon** na najciekawszą publikację w latach 2020-2021, której pierwszym lub drugim autorem jest doktorant (*Andrzejewska, A. i in., 2021, In vivo structure of the Ty1 retrotransposon RNA genome. NAR, 49(5), 2878-2893*).
- 11.2019 – 06.2020 **Stypendium** dla najlepszych doktorantów, ICHB PAN.
- 10.2017 – 06.2018 **Stypendium** dla najlepszych studentów, Uniwersytet Przyrodniczy w Poznaniu.
- 01-03.12.2017 **Wyróżnienie posteru**, “*Molecular diagnostics of selected diseases of the muscular and nervous system in horses (Equus caballus)*”, The International Conference of Natural and Medical Sciences: Young Scientists, PhD Students and Students, Lublin, Polska.

5. PUBLIKACJE

1. Andrzejewska-Romanowska, A., Gumna, J., Tykwińska, E., Pachulska-Wieczorek, K. *Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome. Nucleic Acids Research*, **2024**, doi: 10.1093/nar/gkae494.
(IF₂₀₂₂ – 14.9; MNiSW₂₀₂₄ – 200)
2. Zawadzka, M., Andrzejewska-Romanowska, A., Gumna, J., Garfinkel, D. J., Pachulska-Wieczorek, K. *Cell compartment-specific folding of Ty1 long terminal repeat retrotransposon RNA genome. Viruses*, **2022**, 14(9) 2007.
(IF₂₀₂₂ – 4.7; MNiSW₂₀₂₄ – 100)
3. Andrzejewska, A., Zawadzka, M., Gumna, J., Garfinkel, D. J., Pachulska-Wieczorek, K. *In vivo structure of the Ty1 retrotransposon RNA genome. Nucleic Acids Research*, **2021**, 49(5), 2878-2893.
(IF₂₀₂₁ – 19,16; MNiSW₂₀₂₄ – 200)
4. Gumna, J., Andrzejewska-Romanowska, A., Garfinkel, D. J., Pachulska-Wieczorek, K. *RNA binding properties of the Ty1 LTR-retrotransposon Gag protein. International Journal Of Molecular Sciences*, **2021**, 22(16), 9103.
(IF₂₀₂₁ – 6.21; MNiSW₂₀₂₄ – 140)
5. Andrzejewska, A.*, Zawadzka, M.*, Pachulska-Wieczorek, K. *On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective. International Journal Of Molecular Sciences*, **2020**, 21(18), 6770. * równorzędny pierwszy autor
(IF₂₀₂₀ – 5.92; MNiSW₂₀₂₄ – 140)
6. Hryhorowicz, S., Kaczmarek-Ryś, M., Andrzejewska, A., Staszak, K., Hryhorowicz, M., Korcz, A., Słomski, R. *Allosteric modulation of cannabinoid receptor 1 – Current challenges and future opportunities. International Journal Of Molecular Sciences*, **2019**, 20(23), 5874.
(IF₂₀₁₉ – 4.56; MNiSW₂₀₂₄ – 140)
7. Andrzejewska, A., Staszak, K., Kaczmarek-Ryś, M., Słomski, R., Hryhorowicz, S. *Understanding cannabinoid receptors: structure and function. Acta Universitatis Lodzianis. Folia Biologica Et Oecologica*, **2018**, 14, 1-13.
(MNiSW₂₀₂₄ – 20)
8. Andrzejewska, A., Staszak, K., Lisiak-Teodorczyk, K., Bociąg, P., Cholewiński, G., Wojciechowicz, J. *Diagnostyka molekularna wybranych chorób układu mięśniowego i nerwowego u koni. Życie Weterynaryjne*, **2018**, 93(8), 554-557.
9. Staszak, K., Andrzejewska, A., Lisiak-Teodorczyk, K., Bociąg, P., Cholewiński, G., Wojciechowicz, J. *Diagnostyka molekularna wybranych chorób narządu wzroku u psów. Życie Weterynaryjne*, **2018**, 93(10), 691-695.
10. Andrzejewska, A., Staszak, K., Lisiak-Teodorczyk, K., Bociąg, P., Cholewiński, G., Wojciechowicz, J. *Diagnostyka molekularna wybranych letalnych wad genetycznych u koni. Życie Weterynaryjne*, **2018**, 93(11), 785-788.
11. Monografia pokonferencyjna - Staszak, K., Andrzejewska, A., Słomski, R., Hryhorowicz, S. *Modulacja allosteryczna receptora CBI. Nauki Przyrodnicze i Medyczne: Najnowsze doniesienia dotyczące nauk medycznych i biotechnologicznych*, **2018**, 106-115, 2018, ISBN: 978-83-950719-1-1.

6. DONIESIENIA KONFERENCYJNE

- 29.05-02.06.2024 Poster, “*Exploring structural and functional features of retrotransposon RNA genomes and Gag-derived proteins*”, Andrzejewska-Romanowska A, Gumna J, Tykwińska E, Pachulska-Wieczorek K, 29th Annual Meeting of the RNA Society (RNA 2024), Edynburg, Szkocja.
- Poster, “*Yeast-Based Factory for Therapeutic mRNA Bioproduction*”, Gumna J, Błaszczuk L, Andrzejewska-Romanowska A, Černigoj U, Daboussi F, Freire Maria G, Pachulska-Wieczorek K, Perkovic M, Sekirnik R, Sousa F, Pichon C, 29th Annual Meeting of the RNA Society (RNA 2024), Edynburg, Szkocja.
- 30.10-02.11.2023 Poster, “*Yeast-based factory for therapeutic mRNA bioproduction*”, Gumna J, Błaszczuk L, Zawadzka M, Andrzejewska-Romanowska A, Černigoj U, Daboussi F, Freire MG, Pachulska-Wieczorek K, Perkovic M, Sekirnik R, Sousa F, Pichon C, 11th International mRNA Health Conference, Berlin, Niemcy.
- 28-30.09.2023 Poster, “*Discovering structural features of the active Ty3 retrotransposon RNA genome*”, Andrzejewska-Romanowska A, Gumna J, Tykwińska E, Pachulska-Wieczorek K, Polish RNA Biology Meeting, Warszawa, Polska.
- Poster, “*Transcriptome-wide analysis of yeast RNA secondary structures*”, Tykwińska E, Andrzejewska-Romanowska A, Śledziński P, Gumna J, Pachulska-Wieczorek K, Polish RNA Biology Meeting, Warszawa, Polska.
- 19-21.11.2021 Prezentacja ustna, “*Not entirely mysterious folding of an active retrotransposon RNA genome in the cell*”, Andrzejewska-Romanowska A, Zawadzka M, Gumna J, Garfinkel DJ, Pachulska-Wieczorek K, International Conference of the Centenary of Natural Sciences Club of Adam Mickiewicz University, Poznań, Polska.
- 01-02.07.2021 Poster, “*The life cycle of retrovirus-like Ty1 transposon: the structure of genomic RNA in the cell and Gag function*”, Andrzejewska-Romanowska A, Zawadzka M, Gumna J, Garfinkel DJ, Pachulska-Wieczorek K, RNA Goes Viral Conference, Poznań, Polska.
- 11-16.06.2019 Poster, “*From retrotransposon genome structure to function: Probing the structure of Ty1 genomic RNA in distinct biological states using SHAPE*”, Andrzejewska A, Zawadzka M, Gumna J, Garfinkel DJ, Pachulska-Wieczorek K, 24th Annual Meeting of the RNA Society (RNA 2019), Kraków, Polska.
- 20-22.04.2018 Poster, “*Non-invasive prenatal diagnosis of chromosomal aberrations conditioning Down, Edwards and Patau syndromes using free-circulating fetal DNA and NGS technology*”, Andrzejewska A, Lisiak-Teodorczyk K, Bociąg P, Cholewiński G, Wojciechowicz J, IV National Genetic Conference “Genomica”, Kraków, Polska.
- Poster, “*MTHFR gene mutation analysis in a pregnant woman*”, Staszak K, Andrzejewska A, Lisiak-Teodorczyk K, Bociąg P, Cholewiński G, Wojciechowicz J, IV National Genetic Conference “Genomica”, Kraków, Polska.
- 01-03.12.2017 Poster, “*Molecular diagnostics of selected diseases of the muscular and nervous system in horses (Equus caballus in Latin)*”, Andrzejewska A, Lisiak-Teodorczyk K, Bociąg P, Cholewiński G, Wojciechowicz J. The International Conference of

Natural and Medical Sciences: Young Scientists, PhD Students and Students, Lublin, Polska.

Poster, “*Allosteric modulation of cannabinoid receptor type 1 (CB1) – a new strategy in pharmacotherapy?*”, Staszak K, Andrzejewska A, Słomski R, Hryhorowicz S, The International Conference of Natural and Medical Sciences: Young Scientists, PhD Students and Students, Lublin, Polska.

20.11.2017 Poster, “*Analysis of the frequency of c.802C>T and c.3020incC variants of the NOD2 gene in patients with inflammatory bowel diseases and in the Polish population*”, Andrzejewska A, Staszak K, Pławski A, Kaczmarek-Ryś M, Hryhorowicz S, Boruń P, Hoppe-Gołębiowska J, Słomski R, VI National Scientific Conference Young Scientists in Poland - Research and Development, Poznań, Polska.

Poster, “*Molecular diagnostics of eye diseases in dogs (Canis lupus familiaris)*”, Staszak K, Andrzejewska A, Lisiak-Teodorczyk K, Bociąg P, Cholewiński G, Wojciechowicz J, VI National Scientific Conference Young Scientists in Poland - Research and Development, Poznań, Polska.

11-12.05.2017 Poster, “*The importance of cannabinoid receptors: analysis of the 1359 C/T polymorphism (p.Thr453Thr; rs1049353) of the CNR1 cannabinoid receptor gene in patients with inflammatory bowel disease (IBD) and in the Polish population*”, Andrzejewska A, Staszak K, Hryhorowicz S, Kaczmarek-Ryś M, Krela-Kaźmierczak I, Dobrowolska A, Słomski R, III National Conference of PhD Students in Life Sciences BIOOPEN 2017, Łódź, Polska.

16-17.04.2016 Poster, “*Analysis of the 1359 C/T polymorphism (p.Thr453Thr; rs1049353) of the CNR1 cannabinoid receptor gene in a group of patients with inflammatory bowel disease and in the Polish population*”, Andrzejewska A, Staszak K, Hryhorowicz S, Kaczmarek-Ryś M, Krela-Kaźmierczak I, Dobrowolska A, Słomski R, 5th Wrocław Conference of Students of Science and Technology Puzzel 2016, Wrocław, Polska.

20-22.11.2015 Poster, “*The effect of Interleukin 13 gene on immunoglobulin IgE concentration in children under 3 years of age*”, Andrzejewska A, Rutecki S, XVII National Academic Seminar of Biotechnology Students & VII International Conference of Biotechnology Students, Poznań, Polska.

20-22.03.2015 Poster, “*Mistletoe as a source of anticancer substances*”, Andrzejewska A, Staszak K, 5th Copernican Symposium of Students of Natural Sciences, Toruń, Polska.

7. DODATKOWE DOŚWIADCZENIE I AKTYWNOŚCI

01.2020 Kurs “Wprowadzenie do obróbki i analizy danych NGS”, Ideas4Biology, Poznań.

02.2020 Kurs “Analiza i wizualizacja danych biologicznych w języku R”, Ideas4biology, Poznań.

2014-2017 Czynne uczestnictwo w Kole Naukowym Studentów Biotechnologii “OPERON”, Uniwersytet Przyrodniczy w Poznaniu.

- 20.04.2016* Czynne uczestnictwo w XIX edycji Poznańskiego Festiwalu Nauki i Sztuki.
- 2015-2016* Reprezentant studentów na Radzie Wydziału Rolnictwa i Bioinżynierii Uniwersytetu Przyrodniczego w Poznaniu.
- 25.09.2015* Czynne uczestnictwo w VII edycji Nocy Naukowców, Poznań.
- 06.2015* Kurs "Eucaryotic Molecular Genetics" (prof. Eva Czarnecka-Verner z University of Florida) (1 miesiąc).
- 11.2014-05.2015* Cykl kursów dot. rozwoju umiejętności miękkich (mgr Agnieszka Harmak - Santander), Poznań.
- 16.04.2015* Czynne uczestnictwo w XVIII edycji Poznańskiego Festiwalu Nauki i Sztuki.
- 26.09.2014* Czynne uczestnictwo w VI edycji Nocy Naukowców, Poznań.

Załącznik nr 2 do rozprawy doktorskiej pt.

„Charakterystyka dynamiki strukturalnej i funkcjonalnej genomów RNA aktywnych retrotranspozonów LTR”

PRACE NAUKOWE WCHODZĄCE W SKŁAD
ROZPRAWY DOKTORSKIEJ



Review

On the Way to Understanding the Interplay between the RNA Structure and Functions in Cells: A Genome-Wide Perspective

Angelika Andrzejewska [†] , Małgorzata Zawadzka [†] and Katarzyna Pachulska-Wieczorek ^{*}

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

* Correspondence: kasiapw@ibch.poznan.pl; Tel.: +48-618528503

† These authors contributed equally to this work.

Received: 6 August 2020; Accepted: 11 September 2020; Published: 15 September 2020



Abstract: RNAs adopt specific structures in order to perform their biological activities. The structure of RNA is an important layer of gene expression regulation, and can impact a plethora of cellular processes, starting with transcription, RNA processing, and translation, and ending with RNA turnover. The development of high-throughput technologies has enabled a deeper insight into the sophisticated interplay between the structure of the cellular transcriptome and the living cells environment. In this review, we present the current view on the RNA structure in vivo resulting from the most recent transcriptome-wide studies in different organisms, including mammals, yeast, plants, and bacteria. We focus on the relationship between the mRNA structure and translation, mRNA stability and degradation, protein binding, and RNA posttranscriptional modifications.

Keywords: transcriptome-wide studies; RNA structure in cells; structure-function relationship; translation; RNA stability and degradation; RNA binding proteome; RNA modifications

1. Introduction

RNA molecules fulfill a number of important functions. They take part in not only gene expression, as the intermediate between the DNA and protein, but also in many other cellular processes. It has emerged that the RNA structure, which single stranded RNA can fold into [1], adds another layer of information needed for the regulation of mechanisms such as the transcription, post-transcriptional processing, translation and protein folding, cellular localization, stability, and decay of RNA [2–6]. Although in vitro studies have enabled determination of the secondary structure of many RNAs, and the majority of the RNA secondary structure information is encoded within RNA sequence [7], plenty of results have shown that the RNA structures determined in vitro are not fully consistent with those in the cellular environment [4,8–11]. Thus, studying RNA architecture and its interactions with other RNA or protein partners in living cells has become fundamental to understanding the biochemical pathways of RNA acting.

In the last decade, the development of new, sophisticated methods for the measurement of RNA structures in living cells has revolutionized the field of structural and functional RNA studies. Coupling RNA chemical probing in vivo with next generation sequencing and advanced bioinformatics tools has allowed for shifting the perspective from low-throughput studies to thousands of heterogeneous RNAs, and even whole transcriptomes, in complex cellular environments (reviewed in [12–15]). First genome-wide studies have been published for plant [8] and yeast [9] transcriptomes, and have suggested that RNAs are highly unfolded in vivo. Nevertheless, with the increasing number

of *in vivo* studies, a view of the cell RNA structure as “wet spaghetti” has been displaced [16]. A growing amount of evidence has indicated that the RNA structure can be affected by various cellular factors, but RNA still contains well-defined structural motifs that often display important regulatory functions (Figure 1). The interplay between the cellular environment and RNA structure has become an important question that needs to be resolved.

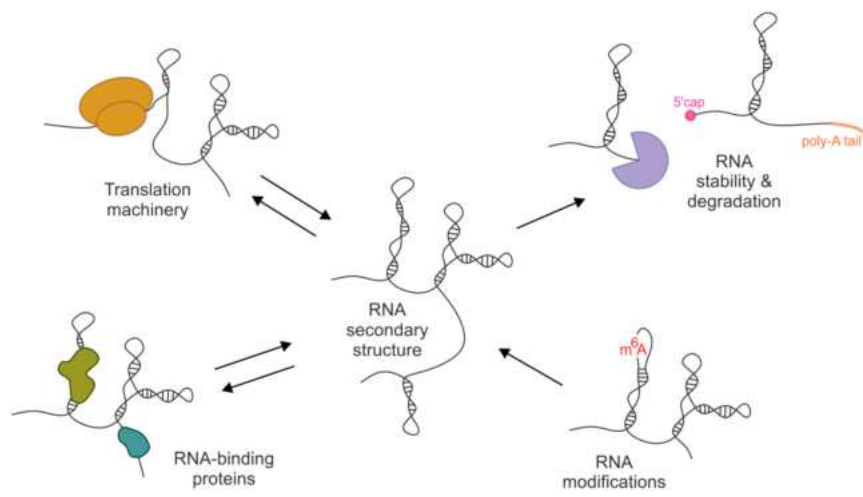


Figure 1. Schematic view on the relationship between the RNA secondary structure and cellular environment elements.

Here, we present the current view on the RNA secondary structure *in vivo* resulting from the most recent transcriptome-wide studies. We summarize and discuss new findings about the impact of cellular factors on the folding of RNA and the relationship between RNA structure and its functioning in the cell. We focus on the correlation between RNA structure and translation, protein binding, RNA stability, and degradation and the role of RNA modifications in the folding of RNA molecules in the cell. However, this review also shows that despite the growing number of transcriptome-wide RNA structural studies *in vivo*, there are still large gaps in the understanding of RNA folding in living cells.

2. Approaches for RNA Structure Determination

For a long time, our knowledge on RNA secondary structure was based mainly on *in silico* methods that calculate the most thermodynamically favorable states of RNA or predict the consensus structures conserved in multiple homologous RNA sequences [17–20]. Although the computational predictions are still developed, they are often not effective for long RNAs with complex structural motifs and usually do not take into account physiological conditions and cellular factors that can impact RNA folding [21]. However, the advancement of techniques for enzymatic and chemical probing of RNA structure and incorporation of experimental-based data into the folding algorithms allowed to significantly improve the accuracy of computational RNA structure predictions [22–24]. Nowadays, RNA probing techniques rely mostly on chemical modifications of RNA in a base- or sugar-specific way, which are detected by the reverse transcriptase enzyme (RT), which either truncates synthesized cDNA one nucleotide before or inserts the mutation at the site of modifications on the RNA strand [13,21,25]. Coupling of RNA chemical probing with next generation sequencing and advanced bioinformatics tools expands the number of available protocols for the high-throughput effective studying of RNA structure *in vitro* as well as *in vivo* (Table 1) [10,26–29]. For *in cell* RNA structural studies, the ability of chemicals to penetrate the cell membrane is crucial. To date, there are only few reagents that can be used for RNA secondary structure determination *in vivo* [30]. Among them, DMS (dimethyl sulfate) and reagents developed for the SHAPE method (selective 2'-hydroxyl acylation analyzed by primer extension) are most widely used in studies *in vivo*. DMS modifies unpaired adenine and cytosine, but uracil and guanine remain without structural information [3,31,32]. SHAPE reagents act in a base-independent

manner and provide structural data for all four nucleotide residues [25,33,34]. Moreover, the methods for incorporation of SHAPE and DMS probing data into RNA structure prediction algorithms are well established [22,35,36].

Table 1. High-Throughput RNA Structure Probing Methods Used in Studies Described in this Review.

Method	Application Described in This Review	Used Probe	Modification Readout	Condition
icSHAPE	Mouse [10,37], human [37], zebrafish [38]	NAI-N ₃	RT-stop	In vivo and in vitro
DMS-seq	Yeast [9], zebrafish [39], <i>E. coli</i> [40]	DMS	RT-stop	In vivo and in vitro
SHAPE-Structure-seq	<i>A. thaliana</i> [41]	NAI	RT-stop	In vivo and in vitro
Structure-seq	<i>A. thaliana</i> [8], rice [42]	DMS	RT-stop	In vivo
SHAPE-MaP	<i>E. coli</i> [43]	1M7	RT-mutate	In vivo and in vitro
SPLASH	Human and yeast [44]	Biotinylated psoralen	Mapping of ligated junctions	In vivo
PARS	<i>E. coli</i> [45]	RNase V1 (dsRNA) and S1 (ssRNA)	Fragments analysis	In vitro
CIRS-seq	Mouse [46]	DMS and CMCT	RT-stop	In vitro

1M7, 1-methyl-7-nitroisatoic anhydride; CIRS, chemical inference of RNA structures; CMCT, N-cyclohexyl-N-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate; DMS, dimethyl sulfate; icSHAPE, in vivo click SHAPE; NAI, 2-methylnicotinic acid imidazolide; NAI-N₃, 2-(azidomethyl)nicotinic acid acyl imidazole; PARS, parallel analysis of RNA structure, RT, reverse transcriptase; SHAPE-MaP, selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling; SPLASH, sequencing of psoralen crosslinked, ligated, and selected hybrids.

3. The Correlation between mRNA Structure and Translation

In addition to their protein-coding functions, mRNAs contain cis-acting sequences with a specific secondary structure that may regulate the translation efficiency (TE). The interplay between the translation and RNA structure constitutes the subject of discussion, and it remains an open question whether the RNA structure guides translation or whether translation guides the RNA structure in cells (Table 2 and Figure 2). In general, stable mRNA structural elements tend to reduce the rate of TE, probably by hampering ribosome binding or slowing down its movement [6,40]; however, on the other hand, the ribosome possesses a helicase activity [47], and has been proposed as a major remodeler of the mRNA structure in cells [39,43].

Table 2. The Critical Conclusions from Reviewed Transcriptome-Wide Studies.

Work	Organism	Main Conclusions
Del Campo et al., 2015 [45]	<i>E. coli</i>	The unstructured sequence upstream of the start codon is a general feature of <i>E. coli</i> genes and is positively correlated with gene expression.
Mustoe et al., 2018 [43]	<i>E. coli</i>	Translation is the main source of mRNA structural destabilization in cells. The structure in RBS is a strong determinant of TE. CDS structure is not critical for TE.
Burkhardt et al., 2017 [40]	<i>E. coli</i>	The structure in RBS does not determine TE. The intrinsic CDS structure plays the critical role in TE tuning.
Beaudoin et al., 2018 [39]	Zebrafish	Translation guides RNA structure rather than structure guiding translation. The ribosome is a major remodeler of RNA structure. Structural elements in the 3' UTR are major regulators of transcript stability during the MZT.
Shi et al., 2020 [38]	Zebrafish	TE is correlated with RNA unfolding.
Rouskin et al., 2014 [9]	Yeast	ATP-dependent processes strongly contribute to the unfolded state of mRNAs inside cells.
Geisberg et al., 2014 [48]	Yeast	The double-stranded structures at the 3'-ends, involving or not involving poly(A) tails, are a critical determinant of mRNA stability.
Moqtaderi et al., 2018 [49]	Yeast	The single-strandedness in the proximity of 3'-end, double-strandedness of the poly(A) tail, together with low Pab1 binding, are linked with mRNA stability and are evolutionarily conserved.
Aw et al., 2016 [44]	Human and yeast	The structure of 5'-UTRs is negatively correlated with mRNA stability, whereas the secondary structure in 3' UTRs is associated with longer mRNA half-life.
Wu et al., 2017 [50]	Human	In cells, 3'-ends are generally more folded than are other mRNA regions and their structure regulates mRNA metabolic stability. Specific structure of the 3'-end can facilitate cleavage and polyadenylation of mRNAs.
Roost et al., 2015 [51]	Human	Ex vivo studies of human transcriptome confirmed the structural RNA changes at the m ⁶ A modification sites, with a strong tendency for unwinding RNA secondary structure.
Sun et al., 2019 [37]	Mouse and human	The intrinsic RNA structure plays a central role in connecting transcription, translation, and RNA degradation. The majority of the transcripts preserve their structure as they transfer from chromatin to the nucleoplasm and cytoplasm. RBPs and RNA modifications account for local RNA structure changes between cellular compartments. CDS structure and TE are only weakly correlated. More-structured RNAs tended to have shorter half-lives. RNA degradation is not RNA-region specific.
Spitale et al., 2015 [10]	Mouse	m ⁶ A modifications impact RNA structure in vivo, favoring the transition from paired to unpaired RNA.

Table 2. Cont.

Work	Organism	Main Conclusions
Deng et al., 2018 [42]	Rice	Higher m ⁶ A modification tends to have less RNA structure in the 3' UTR in plants.
Su et al., 2018 [52]	Rice	Transcripts are subjected to degradation by a mechanism involving secondary structure unfolding in 5' and 3' UTRs.
Ding et al., 2014 [8]	<i>A. thaliana</i>	Less structured regions immediately upstream the start codon region facilitate ribosome binding and increase TE.
Liu et al., 2019 [41]	<i>A. thaliana</i>	Nuclear mRNAs fold differently from cytosolic mRNAs.
Sanchez De Groot et al., 2019 [53]	Various	Highly structured RNAs bind a large amount of proteins.

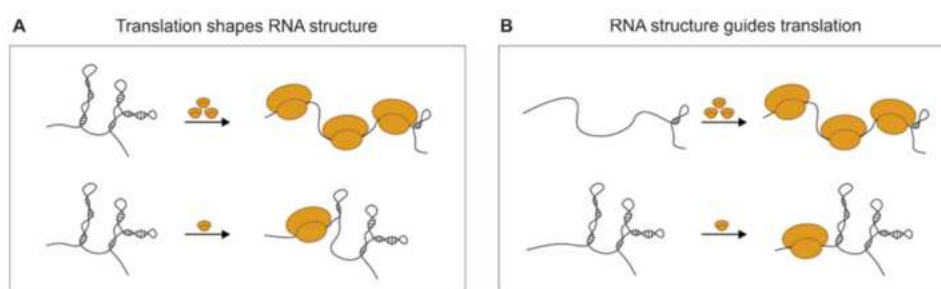


Figure 2. Schematic representation of the interplay between the translation and RNA structure. (A) Translation shapes the RNA structure in the cells by ribosome helicase activity and (B) the RNA structure guides translation by modulating ribosome binding.

3.1. The 5' and 3' UTRs

Numerous studies have shown that the stable structures in the 5' UTR and near the start codon of mRNA can repress the effective initiation of translation in bacteria and eukaryotes [54,55]. The first parallel analysis of RNA structure (PARS; which combines digestion of RNA fragments in vitro by structure-specific enzymes with deep sequencing) on *E. coli* transcriptome complemented with ribosome profiling has indicated that the unstructured sequence upstream of the start codon is a general feature of *E. coli* genes and the secondary structure in this region is negatively correlated with gene expression [45]. Consistent findings were found in recent SHAPE-MaP experiments in living *E. coli* cells that showed that the translation efficiency is regulated by unfolding kinetics of structures overlapping the ribosome binding site (RBS) [43]. According to the RBS kinetic unfolding model, genes with unstructured RBSs have high TE, whereas low TE was observed for those with highly structured RBS. In contrast, DMS-seq analysis of the *E. coli* transcriptome in vivo suggests that TE is only weakly correlated with local RBS structure and is rather regulated by structure of the entire coding sequence (CDS) [40] (see below).

A negative correlation between the high structure of cytoplasmic 5' UTRs and TE was observed in the mouse and human mRNAs studies using icSHAPE [10,37]. In mouse mRNAs, AUG codons are preceded by a 5-nt sequence with increased accessibility, both in vitro and in vivo, suggesting that the structures around translational start sites are programmed by RNA sequences [10]. Consistent findings have been reported in structural studies of the *Arabidopsis thaliana* transcriptome [8]. By applying structure-seq protocol, they showed that decreases in the average mRNA structure near the start codon facilitate ribosome binding and the start of the translation machinery. The structures that 5' UTRs fold into may regulate both the cap-dependent and cap-independent initiation of translation [54]. A variety of higher-order RNA structures in the cap region, including pseudoknots, hairpins, and RNA G-quadruplexes, tend to inhibit translation [56,57]. However, cap-independent regulatory RNA

structures, including IRES (internal ribosome binding site) or eIF3-binding stem-loop structures, can stimulate translation by promoting loading the translation machinery on the mRNA [58–60].

The direct influence of the structural elements within 3' UTRs on translation remains incompletely discovered, and these structures are usually explored in the context of RNA stability (see below). Nevertheless, the RNA stability and translation efficiency are undeniably inseparable and very often, factors, including 3' UTR binding proteins that control transcript stability and decay, are also engaged in TE regulation [61–63].

3.2. The Coding Sequence (CDS)

In general, mRNAs have a lower structure in cells than in vitro, but different RNA structural patterns between species have been found. Across *E. coli* transcripts, especially the coding regions of mRNA tend to be less structured in vivo [40,43]. A strong destabilization of the CDS was also observed in zebrafish in vivo [39]. In addition, global profiling of deproteinized mouse transcripts showed lower degree of structure in CDS compared with UTRs [46]. In contrast, in *Arabidopsis thaliana* and yeast, the average in vivo structure of CDSs was higher or not distinguishable from the UTRs, respectively [8,9]. Additionally, in rice (*Oryza sativa*), the structure of the CDS regions is higher than 3' UTR, but lower than 5' UTR [39,42].

The correlation between CDS structure and ribosome density was not found in the PARS study of *E. coli* transcriptome in vitro [45], but the transcriptome-wide analyses in vivo based on DMS-seq or SHAPE-MaP and the ribosome profiling in *E. coli* showed a strong negative correlation between the mRNA structure and TE [40,43]. In cells, the highly translated ORFs appear to have a lower RNA structure and, inversely, those poorly translated ORFs exhibit a high level of RNA structure. In addition, these studies have shown that inhibition of translation (by kasugamycin treatment) leads to stabilization of CDS structure and the greatest structural difference tends to appear precisely in highly translated genes. These observations thus suggest that the ribosome-induced unfolding contributes to mRNA structure destabilization in cells. Mustoe et al. have also found that the correlation between TE and the structure of CDS in vivo is much weaker than those between RBS structure and TE and proposed that TE is rather unaffected by downstream CDS structure (see also Section 3.1). In contrast, the Burkhardt et al. study pointed to the critical role of the intrinsic CDS structure in TE tuning, whereas other features, such as secondary structure and the strength of the RBS, or the codon usage, are only weakly correlated with mRNA translatability [40]. Furthermore, ORF mRNAs have been shown to have modular structures, and the structures and TE of adjacent ORFs inside the one operon can differ significantly.

A strong correlation between the RNA structure and TE was also observed in transcriptome-wide DMS-seq analysis of zebrafish embryos during maternal to zygotic transition (MZT) [39]. A decreased TE of the maternal transcripts has been found to be connected to reduced DMS reactivity, suggesting an increase in the RNA structure and vice versa; transcripts with a higher TE rate were less structured. In addition, in highly translated mRNAs, the CDS regions are much more accessible than 5' UTRs. Moreover, in embryos treated with an inhibitor of translation initiation (pateamine A), both TE and CDS accessibility to DMS significantly decreased. However, in embryos treated with an inhibitor of translation elongation (cycloheximide), such an effect was not observed. These findings suggest that the initiation of translation and ribosomal entry are crucial for unwinding mRNA into less stable, more open structures. A recent study (icSHAPE) further supports that translation efficiency is correlated with mRNA unfolding in zebrafish embryos [38]. Together, existing data indicate that, at least in some cases, ribosomes have a profound effect on the RNA architecture, and the inherent structure of the coding region appears to have little effect on the translation efficiency.

In contrast to the above studies, the early genome-wide DMS-seq RNA structure analysis in yeast suggested the global unfolding of RNA in vivo, which is not correlated with translation machinery [9]. In this study, the average in vivo RNA structure of CDS does not differ from the structure of UTRs, and a decrease in the RNA structure is not linked to ribosome occupancy, thus suggesting that the

translation and RNA structure are not correlated in yeast. Based on the observation of the RNA structure stabilization under ATP-depleted conditions, the authors proposed the strong contribution of energy-dependent factors, such as ATP-dependent helicases in the unfolding of yeast mRNA. Nevertheless, translation is one of the main energy-consuming processes, and ATP-depletion can also inhibit the action of the ribosomes [64,65].

Furthermore, a recent study of mammalian RNA structures across subcellular compartments has shown that the majority of the transcripts preserve their structure as they transfer from chromatin to the nucleoplasm and cytoplasm [37]. This suggests that the intrinsic RNA structure plays a central role in connecting transcription, translation, and RNA degradation. Although a general trend was observed, where RNA with a lower TE rate tends to be more structured, this study challenges the link between CDS structure and translation efficiency, and highlights the role of RNA binding proteins and RNA posttranscriptional modifications for local RNA structural changes. In contrast, structural differences between nuclear and cytosolic mRNAs were found in the genome-wide analysis of the *A. thaliana* transcriptome [8,41].

An intriguing feature of the coding region detected in diverse organisms *in vivo*, including mouse, *Arabidopsis thaliana*, and rice (*Oryza sativa*), is a three-nucleotide periodicity in the mRNA secondary structure [8,10,42,66]. These studies have found that this periodic repeat pattern is significantly associated with ribosome density *in vivo*, and thus could facilitate translation. However, a three-nucleotide periodicity was not observed in the recent *E. coli* transcriptome-wide study [43].

4. RNA Structure in Relation to RNA Stability and Degradation

The stability and fate of mRNA molecules in the cell are under strict control, and at the end of their life cycle, RNAs undergo a carefully regulated degradation. The cap at the 5'-end and the poly(A) tail at the 3'-end are considered as major determinants of mRNA stability, but it can be also regulated by the other intrinsic features of transcripts, such as the sequence, length, and structure of UTRs and RNA modifications in UTRs [3,48,49,67,68]. Transcriptome-wide studies *in vivo* provide a growing number of evidence that RNA structural features can govern the RNA lifespan (Table 2). The sequences located in UTRs constitute binding sites for diverse trans-acting cellular factors, including RNA-binding proteins (RBPs) and microRNAs [69,70], and the alterations in structure of UTRs might impact their binding and stability of RNA in either positive or negative way [71,72]. The length of mRNA is negatively correlated with its stability in cell and longer transcripts have shorter half-lives in many species [48,73].

In human and yeast, the structure of 5' UTRs is negatively correlated with mRNA stability, whereas the secondary structure in 3' UTRs is associated with longer half-lives and higher abundance of mRNAs [44,50,68]. The stable structural elements in the 3' UTR may block the exosome complex that mainly account for RNA degradation in eukaryotes [74]. In human cells, mRNA 3'-ends tend to be less unfolded than other mRNA regions and specific structure of the 3'-end can facilitate cleavage and polyadenylation of human mRNAs by juxtaposing poly(A) signals (PASs) and cleavage sites that are otherwise too far apart [50]. Interestingly, another study found a negative correlation between RNA structure and RNA half-life in human and mouse cells: more-structured RNAs tended to have shorter half-lives in both the nucleus and cytoplasm [37]. They also proposed that the RNA degradation is not RNA-region specific since the same trends were observed in 5' UTRs, CDS, and 3' UTRs.

In contrast to human mRNAs, in yeast, RNA structure content in 3'-ends is similar to other mRNA regions [9]. However, a global analysis of the clusters of yeast mRNA isoforms with different half-lives showed that the double-stranded structures at the 3'-ends, involving or not involving poly(A) tails, are a critical determinant of mRNA stability in yeast [48]. It was also found that the formation of a stable polyU–poly(A) stem-loop can inhibit the association of poly(A)-binding proteins (e.g., Pab1, Ski2, and Xrn1) and lead to increased mRNA stability. Similar findings were obtained in a more recent study of yeast transcriptome, which confirmed the correlation between structure of the 3'-end and mRNA isoform stability, and showed that even closely related mRNA isoforms can form radically

different structures in the 3' UTR in vivo, and they can occur far from the poly(A) site. This study also showed that single-strandedness in the proximity of 3'-end, double-strandedness of the poly(A) tail, together with low Pab1 binding, are linked with mRNA stability and are evolutionarily conserved [49].

An important role of the 3'-end structure for mRNA stability was also observed in zebrafish in vivo. The 3' UTRs of zebrafish mRNAs are structurally dynamic and changes in their structure can regulate the stability of mRNAs during MZT by modulating miRNA activity [39]. In addition, the 3' UTRs of zebrafish mRNAs are enriched in cis-acting regulatory elements that control mRNA decay during MZT and have been characterized in detail using a high-throughput RNA-element selection assay (RESA), which enables to identify sequences regulating RNA stability with near nucleotide resolution [75]. An in vivo study of heat-regulated RNA structures in rice showed that mRNA unfolding is correlated with RNA half-life and supports the importance of both UTRs for mRNA stability in plants [52]. It was found that heat-induced structural change is greater in UTRs than in other RNA regions and unfolding of 5'- and 3'- ends facilitates access to the RNA degradation machinery.

Moreover, genome-wide RNA decay analysis indicates that codon optimality is also critical for mRNA stability in many organisms, including *E. coli* [76], yeast [77], zebrafish [78], *D. melanogaster* [79], and human [80]. The stable mRNAs are enriched in optimal codons, whereas less stable mRNAs contain predominately non-optimal codons and possess shorter poly(A)-tails.

5. The Relationship between RNA Structure and Proteins Binding

Modern technologies combining high-throughput sequencing with in vivo UV crosslinking and RNA immunoprecipitation (e.g., CLIP-seq, RIP-seq, and RIP-Chip), together with new computational approaches provided new insights into the landscape of the RNA binding proteome [81]. To date, 1753 proteins have been identified in the human RNA-binding proteome, including 978 proteins interacting with poly(A) RNA and 775 proteins that bind non-poly(A) RNA, highlighting the complexity of RNA–protein interactions in vivo [82]. RNA-binding proteins (RBPs) coordinate all of the essential cellular processes, and are an indispensable element in the co- and post-transcriptional regulation of mRNAs and ncRNAs. Various RBPs can dynamically bind RNAs across cellular compartments and during different steps of the lifecycle, including RNA transcription, post-transcriptional processing, translation, stability, and decay [70,83]. A comparative analysis of the human and yeast RNA-binding proteome showed that the RNA-binding activity in vivo and the structural features of many RBPs are strongly conserved in Eukaryotes [84].

Many RBPs need to recognize a specific RNA sequence or structure for their function in a cell [85–87], but other proteins can bind RNA in nonspecific manner, including diverse cellular and viral RNA chaperones that can remodel the RNA structure and facilitate interactions with other partners, as has been shown in studies in vitro [88,89]. However, the regulatory networks between the cellular transcriptome and proteome are only beginning to be understood, and little is known about the target RNA sequences and structural preferences of RBPs in vivo. Genome-wide studies of the sophisticated interplay in the RNA interactome in vivo are still challenging, as RNA–protein complexes tend to be dynamic and change during the RNA lifetime. Advanced global studies correlate the large-scale studies of the RNA-binding proteome with structural information from the transcriptome-wide RNA structure probing in vivo. They confirm the binding of RBPs to specific sequences in the 5' and 3' UTRs, and the critical role of these interactions for the initiation of translation, RNA processing, and its stability in cells (Table 2, and see above). These studies also show that a significant fraction of RNA-binding motifs are present in the coding region and introns [90,91]. The association of individual RNA structures within *S. cerevisiae* transcripts with their interacting proteins revealed that many RBPs recognize evolutionary conserved RNA structures in CDS, possibly formed by the degenerated codons [90]. These interactions have been proposed to regulate post-transcriptional processes, such as tRNA binding and ribosomal biosynthesis, or that these RBPs may act as metabolic enzymes or kinases.

A recent investigation of human transcriptome focused on the interplay between the structure of RNA and its ability to facilitate protein binding revealed a relationship called the RNA structure-driven

protein interactivity, which has an important functional role [53]. According to this theory, the structural content in RNA molecules regulates the number of protein bindings. RBPs interact more with the highly structured RNAs that are rich in double-stranded regions, whereas an opposite trend has been found for poorly structured transcripts. Furthermore, highly structured transcripts preferentially bind polypeptides and encode the regulatory proteins involved in a large number of cellular networks. These findings indicate functional differences between highly and poorly structured RNAs, and suggest the existence of a new, sophisticated layer of post-transcriptional regulation of genes expression. Although this relationship needs to be more closely investigated, a recent comparative analysis of 114 *in vivo* RBP interaction maps from multiple PAR-CLIP experiments performed in HEK293 cells identified the modules of RBPs that are constituted by subsets of proteins that preferentially bind to specific sets of RNAs and targeted regions, and possibly play role in posttranscriptional regulation [91].

In contrast, other genome-wide studies in mammalian cells point to the contribution of RBPs to the RNA structural rearrangements that are distinct from ribosome—or ATP-helicase induced RNA unwinding [10,37]. As diverse subsets of RBPs can bind RNA in specific cellular compartments, they have been proposed to account for local RNA structural changes observed between chromatin, nucleoplasm, and cytoplasm [37]. For example, the heterogeneous nuclear ribonucleoprotein C (HNRNPC) splicing factor, which preferentially recognizes single-stranded uridine tracts [92], is directly involved in the destabilization of the RNA structure around its binding site that, together with m⁶A modification, facilitates HNRNPC-binding in the chromatin RNA fraction [37]. In contrast, the Staufen homolog I, which is a double-stranded-binding RBP involved in the transport and localization of mRNAs to different subcellular compartments [93], seems to participate in stabilizing the RNA structure after RNA release from the chromatin [37].

6. Impact of RNA Modifications on RNA Structure In Vivo

As the RNA, in its life cycle, undergoes numerous post-transcriptional modifications, adding various chemical groups to their bases can significantly influence the RNA folding, stability, and interactions with cellular factors [94–96]. The aberration of RNA modification patterns has been associated with various diseases in human, such as cancerogenesis [97]. Among the hundreds of possible RNA chemical modifications, the most abundant across the mRNA is N⁶-methyladenosine (m⁶A). Thus, it is not surprising that m⁶A modification is of particular interest in research. Development of an m⁶A RNA immunoprecipitation approach followed by high-throughput sequencing (MeRIP-seq) allowed to study m⁶A modification landscape in a transcriptome-wide manner [98,99]. These studies identify over 12,000 m⁶A sites, mainly in the context of the sequence GGm⁶ACU, in more than 7000 human transcripts. The sites of m⁶A modification are highly conserved between humans and mice and preferentially appear in internal long exons, around the stop codons, and in the 3' UTRs. However, the exact functions and, in particular, the effects on RNA folding are still not completely understood. The regulatory function of m⁶A mRNA modification has been shown in transcription, splicing, mRNA export and stability, and translation [37,99–102]. It can also impact various physiological processes, such as the clearance of maternal mRNAs during zebrafish MZT [103], mammalian cortical neurogenesis [104], and plays regulatory role in human cancer [105,106].

Thermodynamic study has shown that m⁶A influences the RNA structure because of the rotation of the methylamino group, from *syn* to *anti* conformation, with a higher energy, thus destabilizing the RNA duplexes by 0.5–1.7 kcal/mol [51]. The opposite effect was observed in the single stranded region of RNA, where m⁶A can contribute to increasing the stability of the RNA molecule, probably by base stacking. *Ex vivo* studies of human transcriptome confirmed the structural RNA changes at the m⁶A modification sites, with a strong tendency for unwinding RNA secondary structure [51]. Therefore, the appearance of m⁶A in RNA has been proposed to work as a “molecular switch” for the RNA structure [107]. Spitale et al. first comprised the *in vivo* SHAPE reactivities of m⁶A-modified vs. unmodified transcripts, and showed that in the cell RNA regions, both surrounding and including the modified A residues, tend to be unpaired [10]. To check whether decreased base-pairing at the

modification sites is caused by the m⁶A destabilizing effect on RNA duplexes or the modification machinery preferentially methylate adenosine at single-stranded sites of RNA, they performed genetic knockout of N⁶-adenosine-methyltransferase (*Mettl3*) in mouse ES cells. The depletion of *Mettl3* led to a transcriptome-wide reduction of the SHAPE signal at the m⁶A modification sites, confirming the RNA destabilizing properties of m⁶A modification in vivo.

Recent transcriptome-wide studies across the three different cellular compartments in mammalian cells further supported that m⁶A enriched regions in transcripts are far less structured than the same, but unmethylated, sites of RNAs, and the patterns of m⁶A-induced structural destabilization are similar in chromatin, nucleoplasm, and cytoplasm [37]. However, the RNA structure is more open after RNA release from the chromatin to nucleoplasm, consistent with *METTL3-METTL14* complex's localization to specific nuclear loci [102,108]. Furthermore, across all of the analyzed compartments, RNA modifications significantly overlap with both structural-change sites and RBP bindings, suggesting that many RBPs require induced by the m⁶A local destabilization of RNA for their binding. For example, in the chromatin fraction, m⁶A modification can facilitate the binding of HNRNPC by the disruption of the local RNA secondary structure in close proximity to the binding sites [37,107]. In addition, m⁶A can work as an "RNA molecular switch" in plants [42]. However, the significant correlation between the structural changes and m⁶A in rice has been observed only in 3' UTR and not in the CDS or 5' UTR regions.

7. Conclusions and Future Perspectives

Recent transcriptome-wide studies have significantly increased the knowledge about the interplay between the RNA secondary structure and RNA functions in cells. They allow for considering the influence of various cellular factors on RNA folding in vivo, and vice versa, as well as the impact of the RNA structure on critical biological processes. However, there are still many unanswered questions and challenges. For example, it remains unclear whether the RNA folding pattern is species-specific or observed differences result from application of various RNA structure probing methods. An important area for advancement is further development of methods for more accurate studying of the coexisting RNA conformers, RNA co-transcriptional folding, and differentiation of intramolecular RNA interactions and intermolecular RNA–protein or RNA–RNA bindings. Together, with continuous experimental/technical development, there is also a need to advance the computational tools not only for high-throughput data analysis, but also for the experimentally supported accurate modeling of the RNA structure in the native in vivo form.

Author Contributions: K.P.-W. conceptualized the work, A.A. and M.Z. wrote the manuscript, K.P.-W. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Polish National Science Centre (2016/22/E/NZ3/00426 to K.P.-W.). Funding for open access charge was provided by the Polish National Science Centre (2016/22/E/NZ3/00426 to K.P.-W.).

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

References

1. Hendrix, D.K.; Brenner, S.E.; Holbrook, S.R. RNA structural motifs: Building blocks of a modular biomolecule. *Q. Rev. Biophys.* **2005**, *38*, 221–243. [[CrossRef](#)] [[PubMed](#)]
2. Kwok, C.K. Dawn of the in vivo RNA structurome and interactome. *Biochem. Soc. Trans.* **2016**, *44*, 1395–1410. [[CrossRef](#)] [[PubMed](#)]
3. Piao, M.; Sun, L.; Zhang, Q.C. RNA Regulations and Functions Decoded by Transcriptome-wide RNA Structure Probing. *Genom. Proteom. Bioinform.* **2017**, *15*, 267–278. [[CrossRef](#)] [[PubMed](#)]
4. Mortimer, S.A.; Kidwell, M.A.; Doudna, J.A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **2014**, *15*, 469–479. [[CrossRef](#)]
5. Ignatova, Z.; Narberhaus, F. Systematic probing of the bacterial RNA structurome to reveal new functions. *Curr. Opin. Microbiol.* **2017**, *36*, 14–19. [[CrossRef](#)] [[PubMed](#)]

6. Faure, G.; Ogurtsov, A.Y.; Shabalina, S.A.; Koonin, E.V. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res.* **2016**, *44*, 10898–10911. [[CrossRef](#)]
7. Wan, Y.; Qu, K.; Zhang, Q.C.; Flynn, R.A.; Manor, O.; Ouyang, Z.; Zhang, J.; Spitale, R.C.; Snyder, M.P.; Segal, E.; et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **2014**, *505*, 706–709. [[CrossRef](#)] [[PubMed](#)]
8. Ding, Y.; Tang, Y.; Kwok, C.K.; Zhang, Y.; Bevilacqua, P.C.; Assmann, S.M. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **2013**, *505*, 696–700. [[CrossRef](#)]
9. Rouskin, S.; Zubradt, M.; Washietl, S.; Kellis, M.; Weissman, J.S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **2013**, *505*, 701–705. [[CrossRef](#)]
10. Spitale, R.C.; Flynn, R.A.; Zhang, Q.C.; Crisalli, P.; Lee, B.; Jung, J.-W.; Kuchelmeister, H.Y.; Batista, P.J.; Torre, E.A.; Kool, E.T.; et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **2015**, *519*, 486–490. [[CrossRef](#)]
11. Smola, M.J.; Christy, T.W.; Inoue, K.; Nicholson, C.O.; Friedersdorf, M.; Keene, J.D.; Lee, D.M.; Calabrese, J.M.; Weeks, K.M. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10322–10327. [[CrossRef](#)]
12. Bevilacqua, P.C.; Ritchey, L.E.; Su, Z.; Assmann, S.M. Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **2016**, *50*, 235–266. [[CrossRef](#)]
13. Strobel, E.J.; Yu, A.M.; Lucks, J.B. High-throughput determination of RNA structures. *Nat. Rev. Genet.* **2018**, *19*, 615–634. [[CrossRef](#)]
14. Kubota, M.; Tran, C.; Spitale, R. Progress and challenges for chemical probing of RNA structure inside living cells. *Nat. Methods* **2015**, *11*, 933–941. [[CrossRef](#)] [[PubMed](#)]
15. Weeks, K.M. Review toward all RNA structures, concisely. *Biopolymers* **2015**, *103*, 438–448. [[CrossRef](#)] [[PubMed](#)]
16. Pyle, A.M. Rediscovering RNA. *RNA* **2015**, *21*, 714–715. [[CrossRef](#)] [[PubMed](#)]
17. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)] [[PubMed](#)]
18. Lorenz, R.; Bernhart, S.H.; Höner Zu Siederdisen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [[CrossRef](#)]
19. Reuter, J.S.; Mathews, D.H. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinform.* **2010**, *11*, 129. [[CrossRef](#)]
20. Xu, Z.; Mathews, D.H. Prediction of Secondary Structures Conserved in Multiple RNA Sequences. *Adv. Struct. Saf. Stud.* **2016**, *1490*, 35–50. [[CrossRef](#)]
21. Mailler, E.; Paillart, J.-C.; Marquet, R.; Smyth, R.P.; Vivet-Boudou, V. The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip. Rev. RNA* **2018**, *10*, e1518. [[CrossRef](#)] [[PubMed](#)]
22. Mathews, D.H.; Disney, M.D.; Childs, J.L.; Schroeder, S.J.; Zuker, M.; Turner, D.H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7287–7292. [[CrossRef](#)]
23. Deigan, K.E.; Li, T.W.; Mathews, D.H.; Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* **2008**, *106*, 97–102. [[CrossRef](#)]
24. Wu, Y.; Shi, B.; Ding, X.; Liu, T.; Hu, X.; Yip, K.; Yang, Z.R.; Mathews, D.H.; Lu, Z.J. Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.* **2015**, *43*, 7247–7259. [[CrossRef](#)] [[PubMed](#)]
25. Busan, S.; Weidmann, C.A.; Sengupta, A.; Weeks, K.M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **2019**, *58*, 2655–2664. [[CrossRef](#)] [[PubMed](#)]
26. Ritchey, L.E.; Su, Z.; Tang, Y.; Tack, D.C.; Assmann, S.M.; Bevilacqua, P.C. Structure-seq2: Sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Res.* **2017**, *45*, e135. [[CrossRef](#)]
27. Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A.M.; Weissman, J.S.; Rouskin, S. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **2016**, *14*, 75–82. [[CrossRef](#)]
28. Siegfried, N.A.; Busan, S.; Rice, G.M.; Nelson, J.A.; Weeks, K.M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **2014**, *11*, 959–965. [[CrossRef](#)]

29. Kertesz, M.; Wan, Y.; Mazor, E.; Rinn, J.L.; Nutter, R.C.; Chang, H.Y.; Segal, E. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **2010**, *467*, 103–107. [[CrossRef](#)]
30. Bevilacqua, P.C.; Assmann, S.M. Technique Development for Probing RNA Structure In Vivo and Genome-Wide. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a032250. [[CrossRef](#)]
31. Lawley, P.D.; Brookes, P.; Swann, P.F.; Magee, P.N.; Orr, D.J.; Jarman, M. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.* **1963**, *89*, 127–138. [[CrossRef](#)]
32. Wells, S.E.; Hughes, J.M.; Igel, A.H.; Ares, M. Use of dimethyl sulfate to probe RNA structure in vivo. *Enzym. Eng. Evol. Gen. Methods* **2000**, *318*, 479–493. [[CrossRef](#)]
33. Merino, E.J.; Wilkinson, K.A.; Coughlan, J.L.; Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231. [[CrossRef](#)]
34. Spitale, R.C.; Flynn, R.A.; Torre, E.A.; Kool, E.T.; Chang, H.Y. RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip. Rev. RNA* **2014**, *5*, 867–881. [[CrossRef](#)]
35. Lorenz, R.; Wolfinger, M.T.; Tanzer, A.; Hofacker, I.L. Predicting RNA secondary structures from sequence and probing data. *Methods* **2016**, *103*, 86–98. [[CrossRef](#)]
36. Sloma, M.F.; Mathews, D.H. Improving RNA Secondary Structure Prediction with Structure Mapping Data. *Enzym. Eng. Evol. Gen. Methods* **2015**, *553*, 91–114. [[CrossRef](#)]
37. Sun, L.; Fazal, F.M.; Li, P.; Broughton, J.P.; Lee, B.; Tang, L.; Huang, W.; Kool, E.T.; Chang, H.Y.; Zhang, Q.C. RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **2019**, *26*, 322–330. [[CrossRef](#)] [[PubMed](#)]
38. Shi, B.; Zhang, J.; Heng, J.; Gong, J.; Zhang, T.; Li, P.; Sun, B.-F.; Yang, Y.; Zhao, Y.-L.; Wang, H.-L.; et al. RNA structural dynamics regulate early embryogenesis through controlling transcriptome fate and function. *Genome Biol.* **2020**, *21*, 1–27. [[CrossRef](#)]
39. Beaudoin, J.-D.; Novoa, E.M.; Vejnar, C.E.; Yartseva, V.; Takacs, C.M.; Kellis, M.; Giraldez, A.J. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* **2018**, *25*, 677–686. [[CrossRef](#)]
40. Burkhardt, D.H.; Rouskin, S.; Zhang, Y.; Li, G.-W.; Weissman, J.S.; Gross, C.A. Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *eLife* **2017**, *6*, 811. [[CrossRef](#)]
41. Liu, Z.; Liu, Q.; Yang, X.; Zhang, Y.; Norris, M.; Chen, X.; Cheema, J.; Ding, Y. In vivo nuclear RNA structurome reveals RNA-structure regulation of mRNA processing in plants. *bioRxiv* **2019**, 839506. preprint. [[CrossRef](#)]
42. Deng, H.; Cheema, J.; Zhang, H.; Woolfenden, H.C.; Norris, M.; Liu, Z.; Liu, Q.; Yang, X.; Yang, M.; Deng, X.; et al. Rice In Vivo RNA Structurome Reveals RNA Secondary Structure Conservation and Divergence in Plants. *Mol. Plant* **2018**, *11*, 607–622. [[CrossRef](#)] [[PubMed](#)]
43. Mustoe, A.M.; Busan, S.; Rice, G.M.; Hajdin, C.E.; Peterson, B.K.; Ruda, V.M.; Kubica, N.; Nutiu, R.; Baryza, J.L.; Weeks, K.M. Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **2018**, *173*, 181–195.e18. [[CrossRef](#)] [[PubMed](#)]
44. Aw, J.G.A.; Shen, Y.; Wilm, A.; Sun, M.; Ni Lim, X.; Boon, K.-L.; Tapsin, S.; Chan, Y.-S.; Tan, C.-P.; Sim, A.Y.; et al. In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol. Cell* **2016**, *62*, 603–617. [[CrossRef](#)]
45. Del Campo, C.; Bartholomäus, A.; Fedyunin, I.; Ignatova, Z. Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet.* **2015**, *11*, e1005613. [[CrossRef](#)]
46. Incarnato, D.; Neri, F.; Anselmi, F.; Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* **2014**, *15*, 491. [[CrossRef](#)]
47. Takyar, S.; Hickerson, R.P.; Noller, H.F. mRNA Helicase Activity of the Ribosome. *Cell* **2005**, *120*, 49–58. [[CrossRef](#)]
48. Geisberg, J.V.; Moqtaderi, Z.; Fan, X.; Oszolak, F.; Struhl, K. Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell* **2014**, *156*, 812–824. [[CrossRef](#)]
49. Moqtaderi, Z.; Geisberg, J.V.; Struhl, K. Extensive Structural Differences of Closely Related 3' mRNA Isoforms: Links to Pab1 Binding and mRNA Stability. *Mol. Cell.* **2018**, *72*, 849–861. [[CrossRef](#)]
50. Wu, X.; Bartel, D.P. Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability. *Cell* **2017**, *169*, 905–917. [[CrossRef](#)]

51. Roost, C.; Lynch, S.R.; Batista, P.J.; Qu, K.; Chang, H.Y.; Kool, E.T. Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J. Am. Chem. Soc.* **2015**, *137*, 2107–2115. [[CrossRef](#)]
52. Su, Z.; Tang, Y.; Ritchey, L.E.; Tack, D.C.; Zhu, M.; Bevilacqua, P.C.; Assmann, S.M. Genome-wide RNA structurome reprogramming by acute heat shock globally regulates mRNA abundance. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12170–12175. [[CrossRef](#)] [[PubMed](#)]
53. De Groot, N.S.; Armaos, A.; Graña-Montes, R.; Alriquet, M.; Calloni, G.; Vabulas, R.M.; Tartaglia, G.G. RNA structure drives interaction with proteins. *Nat. Commun.* **2019**, *10*, 3246. [[CrossRef](#)]
54. Leppek, K.; Das, R.; Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell. Biol.* **2018**, *19*, 158–174. [[CrossRef](#)]
55. Chiaruttini, C.; Guillier, M. On the role of mRNA secondary structure in bacterial translation. *Wiley Interdiscip. Rev. RNA* **2020**, *11*, e1579. [[CrossRef](#)]
56. Bolduc, F.; Garant, J.-M.; Allard, F.; Perreault, J.-P. Irregular G-quadruplexes Found in the Untranslated Regions of Human mRNAs Influence Translation. *J. Biol. Chem.* **2016**, *291*, 21751–21760. [[CrossRef](#)]
57. Cohen-Chalamish, S.; Hasson, A.; Weinberg, D.; Namer, L.S.; Banai, Y.; Osman, F.; Kaempfer, R. Dynamic refolding of IFN-gamma mRNA enables it to function as PKR activator and translation template. *Nat. Chem. Biol.* **2009**, *5*, 896–903. [[CrossRef](#)]
58. Baird, S.D.; Lewis, S.M.; Turcotte, M.; Holcik, M. A search for structurally similar cellular internal ribosome entry sites. *Nucleic Acids Res.* **2007**, *35*, 4664–4677. [[CrossRef](#)]
59. Lee, A.S.Y.; Kranzusch, P.J.; Cate, J.H. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **2015**, *522*, 111–114. [[CrossRef](#)]
60. Weingarten-Gabbay, S.; Elias-Kirma, S.; Nir, R.; Gritsenko, A.A.; Stern-Ginossar, N.; Yakhini, Z.; Weinberger, A.; Segal, E. Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **2016**, *351*. [[CrossRef](#)]
61. Radhakrishnan, A.; Green, R. Connections Underlying Translation and mRNA Stability. *J. Mol. Biol.* **2016**, *428*, 3558–3564. [[CrossRef](#)] [[PubMed](#)]
62. Tuck, A.C.; Rankova, A.; Arpat, A.B.; Liechti, L.A.; Hess, D.; Iesmantavicius, V.; Castelo-Szekely, V.; Gatfield, D.; Bühler, M. Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. *Mol. Cell* **2020**, *77*, 1222–1236. [[CrossRef](#)] [[PubMed](#)]
63. Otsuka, H.; Fukao, A.; Funakami, Y.; Duncan, K.E.; Fujiwara, T. Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. *Front. Genet.* **2019**, *10*, 332. [[CrossRef](#)] [[PubMed](#)]
64. Buttgereit, F.; Brand, M.D. A hierarchy of ATP-consuming processes in mammalian cells. *Biochem. J.* **1995**, *312*, 163–167. [[CrossRef](#)]
65. Ashe, M.P.; De Long, S.K.; Sachs, A.B. Glucose Depletion Rapidly Inhibits Translation Initiation in Yeast. *Mol. Biol. Cell* **2000**, *11*, 833–848. [[CrossRef](#)]
66. Li, F.; Zheng, Q.; Vandivier, L.E.; Willmann, M.R.; Chen, Y.; Gregory, B.D. Regulatory Impact of RNA Secondary Structure across the Arabidopsis Transcriptome. *Plant. Cell* **2012**, *24*, 4346–4359. [[CrossRef](#)]
67. Gagliardi, D.; Dziembowski, A. 5' and 3' modifications controlling RNA degradation: From safeguards to executioners. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2018**, *373*. [[CrossRef](#)]
68. Ringner, M.; Krogh, M. Folding free energies of 5'-UTRs impact post-transcriptional regulation on a genomic scale in yeast. *PLoS Comput. Biol.* **2005**, *1*, e72.
69. Fabian, M.R.; Sonenberg, N.; Filipowicz, W. Regulation of mRNA Translation and Stability by microRNAs. *Annu. Rev. Biochem.* **2010**, *79*, 351–379. [[CrossRef](#)]
70. Hentze, M.W.; Castellóo, A.; Schwarzl, T.; Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 327–341. [[CrossRef](#)]
71. Mayya, V.K.; Duchaine, T.F. Ciphers and Executioners: How 3'-Untranslated Regions Determine the Fate of Messenger RNAs. *Front. Genet.* **2019**, *10*, 6. [[CrossRef](#)] [[PubMed](#)]
72. Mayr, C. Regulation by 3'-Untranslated Regions. *Annu. Rev. Genet.* **2017**, *51*, 171–194. [[CrossRef](#)] [[PubMed](#)]
73. Feng, L.; Niu, D.-K. Relationship Between mRNA Stability and Length: An Old Question with a New Twist. *Biochem. Genet.* **2007**, *45*, 131–137. [[CrossRef](#)]
74. Decker, C.J.; Parker, R. mRNA decay enzymes: Decappers conserved between yeast and mammals. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12512–12514. [[CrossRef](#)]

75. Vejnar, C.E.; Abdelmessih, M.; Takacs, C.; Yartseva, V.; Oikonomou, P.; Christiano, R.; Stoeckius, M.; Lau, S.; Lee, M.; Beaudoin, J.-D. Genome wide analysis of 3' UTR sequence elements and proteins regulating mRNA stability during maternal-to-zygotic transition in zebrafish. *Genome Res.* **2019**, *29*, 1100–1114. [[CrossRef](#)]
76. Boel, G.; Letso, R.; Neely, H.; Price, W.N.; Wong, K.-H.; Su, M.; Luff, J.D.; Valecha, M.; Everett, J.K.; Acton, T.B.; et al. Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature* **2016**, *529*, 358–363. [[CrossRef](#)]
77. Presnyak, V.; Alhusaini, N.; Chen, Y.-H.; Martin, S.; Morris, N.; Kline, N.; Olson, S.; Weinberg, D.; Baker, K.E.; Graveley, B.R.; et al. Codon optimality is a major determinant of mRNA stability. *Cell* **2015**, *160*, 1111–1124. [[CrossRef](#)]
78. Bazzini, A.A.; Viso, F.; Moreno-Mateos, M.A.; Johnstone, T.; Vejnar, C.E.; Qin, Y.; Yao, J.; Khokha, M.K.; Giraldez, A.J. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **2016**, *35*, 2087–2103. [[CrossRef](#)]
79. Burrow, D.A.; Martin, S.; Quail, J.F.; Alhusaini, N.; Coller, J.; Cleary, M.D. Attenuated Codon Optimality Contributes to Neural-Specific mRNA Decay in Drosophila. *Cell Rep.* **2018**, *24*, 1704–1712. [[CrossRef](#)]
80. Wu, Q.; Medina, S.G.; Kushawah, G.; Devore, M.L.; Castellano, L.; Hand, J.M.; Wright, M.; Bazzini, A.A. Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife* **2019**, *8*, 8. [[CrossRef](#)]
81. Wheeler, E.C.; Van Nostrand, E.L.; Yeo, G.W. Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. *Wiley Interdiscip. Rev. RNA* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
82. Trendel, J.; Schwarzl, T.; Horos, R.; Prakash, A.; Bateman, A.; Hentze, M.W.; Krijgsveld, J. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **2019**, *176*, 391–403. [[CrossRef](#)] [[PubMed](#)]
83. Dominguez, D.; Freese, P.; Alexis, M.S.; Su, A.; Hochman, M.; Palden, T.; Bazile, C.; Lambert, N.J.; Van Nostrand, E.L.; Pratt, G.A.; et al. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **2018**, *70*, 854–867. [[CrossRef](#)] [[PubMed](#)]
84. Beckmann, B.M.; Horos, R.; Fischer, B.; Castelló, A.; Eichelbaum, K.; Alleaume, A.-M.; Schwarzl, T.; Curk, T.; Foehr, S.; Huber, W.; et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **2015**, *6*, 10127. [[CrossRef](#)] [[PubMed](#)]
85. Jankowsky, E.; Harris, M.E. Specificity and nonspecificity in RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 533–544. [[CrossRef](#)]
86. Ray, D.; Kazan, H.; Cook, K.; Weirauch, M.T.; Najafabadi, H.S.; Li, X.; Gueroussov, S.; Albu, M.; Zheng, H.; Yang, A.; et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **2013**, *499*, 172–177. [[CrossRef](#)]
87. Kumari, P.; Aeschmann, F.; Gaidatzis, D.; Keusch, J.J.; Ghosh, P.; Neagu, A.; Pachulska-Wieczorek, K.; Bujnicki, J.; Gut, H.; Großhans, H.; et al. Evolutionary plasticity of the NHL domain underlies distinct solutions to RNA recognition. *Nat. Commun.* **2018**, *9*, 1549. [[CrossRef](#)]
88. 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*, 18. [[CrossRef](#)]
89. Woodson, S.A.; Panja, S.; Santiago-Frangos, A. Proteins That Chaperone RNA Regulation. *Regul. RNA Bact. Archaea* **2018**, *6*, 383–397. [[CrossRef](#)]
90. Casas-Vila, N.; Sayols, S.; Pérez-Martínez, L.; Scheibe, M.; Butter, F. The RNA fold interactome of evolutionary conserved RNA structures in S. cerevisiae. *Nat. Commun.* **2020**, *11*, 1–12. [[CrossRef](#)]
91. Mukherjee, N.; Wessels, H.-H.; Lebedeva, S.; Sajek, M.; Ghanbari, M.; Garzia, A.; Munteanu, A.; Yusuf, D.; Farazi, T.; Hoell, J.I.; et al. Deciphering human ribonucleoprotein regulatory networks. *Nucleic Acids Res.* **2018**, *47*, 570–581. [[CrossRef](#)] [[PubMed](#)]
92. Konig, J.; Zarnack, K.; Rot, G.; Curk, T.; Kayikci, M.; Zupan, B.; Turner, D.J.; Luscombe, N.M.; Ule, J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **2010**, *17*, 909–915. [[CrossRef](#)]
93. Furic, L.; Maher-Laporte, M.; DesGroseillers, L. A genome-wide approach identifies distinct but overlapping subsets of cellular mRNAs associated with Staufen1- and Staufen2-containing ribonucleoprotein complexes. *RNA* **2007**, *14*, 324–335. [[CrossRef](#)] [[PubMed](#)]
94. Incarnato, D.; Oliviero, S. The RNA Epistruome: Uncovering RNA Function by Studying Structure and Post-Transcriptional Modifications. *Trends Biotechnol.* **2017**, *35*, 318–333. [[CrossRef](#)]

95. Boo, S.H.; Kim, Y.K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp. Mol. Med.* **2020**, *52*, 400–408. [[CrossRef](#)] [[PubMed](#)]
96. Harcourt, E.M.; Kietrys, A.M.; Kool, E.T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **2017**, *541*, 339–346. [[CrossRef](#)]
97. Delaunay, S.; Frye, M. RNA modifications regulating cell fate in cancer. *Nature* **2019**, *21*, 552–559. [[CrossRef](#)]
98. Meyer, K.D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C.E.; Jaffrey, S.R. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **2012**, *149*, 1635–1646. [[CrossRef](#)]
99. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **2012**, *485*, 201–206. [[CrossRef](#)]
100. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S.R.; Qian, S.-B. Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* **2015**, *526*, 591–594. [[CrossRef](#)]
101. Yue, Y.; Liu, J.; He, C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev.* **2015**, *29*, 1343–1355. [[CrossRef](#)] [[PubMed](#)]
102. Roundtree, I.A.; Evans, M.E.; Pan, T.; He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017**, *169*, 1187–1200. [[CrossRef](#)] [[PubMed](#)]
103. Zhao, B.S.; Wang, X.; Beadell, A.V.; Lu, Z.; Shi, H.; Kuuspalu, A.; Ho, R.K.; He, C. m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **2017**, *542*, 475–478. [[CrossRef](#)] [[PubMed](#)]
104. Yoon, K.-J.; Ringeling, F.R.; Vissers, C.; Jacob, F.; Pokrass, M.; Jimenez-Cyrus, D.; Su, Y.; Kim, N.-S.; Zhu, Y.; Zheng, L.; et al. Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. *Cell* **2017**, *171*, 877–889. [[CrossRef](#)]
105. Chen, X.-Y.; Zhang, J.; Zhu, J. The role of m6A RNA methylation in human cancer. *Mol. Cancer* **2019**, *18*, 103. [[CrossRef](#)]
106. Jaffrey, S.R.; Kharas, M.G. Emerging links between m6A and misregulated mRNA methylation in cancer. *Genome Med.* **2017**, *9*, 2. [[CrossRef](#)]
107. Liu, N.; Dai, Q.; Zheng, G.; He, C.; Parisien, M.; Pan, T. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* **2015**, *518*, 560–564. [[CrossRef](#)]
108. Roundtree, I.A.; He, C. RNA epigenetics—Chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem. Biol.* **2016**, *30*, 46–51. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).

In vivo structure of the Ty1 retrotransposon RNA genome

Angelika Andrzejewska¹, Małgorzata Zawadzka¹, Julita Gumna¹, David J. Garfinkel²
and Katarzyna Pachulska-Wieczorek^{1,*}

¹Department of Structure and Function of Retrotransposons, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland and ²Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

Received November 09, 2020; Revised January 28, 2021; Editorial Decision January 29, 2021; Accepted February 02, 2021

ABSTRACT

Long terminal repeat (LTR)-retrotransposons constitute a significant part of eukaryotic genomes and influence their function and evolution. Like other RNA viruses, LTR-retrotransposons efficiently utilize their RNA genome to interact with host cell machinery during replication. Here, we provide the first genome-wide RNA secondary structure model for a LTR-retrotransposon in living cells. Using SHAPE probing, we explore the secondary structure of the yeast Ty1 retrotransposon RNA genome in its native *in vivo* state and under defined *in vitro* conditions. Comparative analyses reveal the strong impact of the cellular environment on folding of Ty1 RNA. *In vivo*, Ty1 genome RNA is significantly less structured and more dynamic but retains specific well-structured regions harboring functional *cis*-acting sequences. Ribosomes participate in the unfolding and remodeling of Ty1 RNA, and inhibition of translation initiation stabilizes Ty1 RNA structure. Together, our findings support the dual role of Ty1 genomic RNA as a template for protein synthesis and reverse transcription. This study also contributes to understanding how a complex multifunctional RNA genome folds *in vivo*, and strengthens the need for studying RNA structure in its natural cellular context.

INTRODUCTION

LTR-retrotransposons and RNA viruses possess a compact RNA genome that encodes information required for replication and encapsidation in nucleotide sequence and secondary and tertiary structures. Genome-wide *in vitro*, *in virio* and *ex virio* RNA structure models were presented for several infectious RNA viruses (1). However, since these RNA structure models may not reflect conditions *in vivo*,

elucidating the native structure of viral RNA genomes exposed to the cellular environment is necessary to reach a comprehensive view for how genome architecture influences viral replication. *In vivo* RNA structure models are available for Zika and Dengue viruses (*Flaviviridae*), and the coronavirus SARS-CoV-2 (2–5). There is also a transcriptome-wide *in vivo* secondary structure model of influenza A virus (IAV) mRNAs (*Orthomyxoviridae*) (6). These studies reveal important differences in the *in vivo* folding of viral RNAs, but it remains unclear what factors account for remodeling of the structure of viral RNAs in cells.

Reverse-transcribing single strand RNA viruses (ssRNA-RT viruses) are widespread in diverse eukaryotes. The most extensively studied are *Pseudoviridae* and *Metaviridae* families, better known as Ty1/Copia and Ty3/Gypsy long terminal repeat (LTR)-retrotransposons, respectively, and the *Retroviridae* family, which contains prominent human pathogens such as HIV-1 (7). These families share homologous proteins and show mechanistic similarities in genome replication, polyprotein processing and virus particle formation (8–11). As critical steps of their replication cycle, these retroelements encapsidate two copies of single-stranded RNA genome (gRNA), synthesize DNA copy of genome utilizing host tRNA as primers and integrate into host DNA (12,13).

Saccharomyces cerevisiae Ty1-5 retrotransposons continue to provide fundamental insights into the mechanism of retrotransposition and the impact of retroelements on eukaryotic cells (8,9,14). Compared to other eukaryotes, the compact yeast genome contains a relatively small fraction of transposon sequences (1.3–3.4%) (15). Except for Ty3, yeast Ty elements belong to Ty1/Copia family, and Ty1 is the most abundant mobile genetic element in many *S. cerevisiae* strains (16). Ty1 resembles simple retroviruses, but lacks an *ENV* (envelope) gene required for infectivity (17). Ty1 gRNA is 5.6-kb long and contains two partially overlapping *GAG* and *POL* ORFs, flanked by a 53-nt 5'UTR and 318-nt 3'UTR (18). Ty1 transcripts are not spliced and

*To whom correspondence should be addressed. Tel: +48 61 852 85 03; Email: kasiapw@ibch.poznan.pl

about 15% of total Ty1 gRNA is polyadenylated. Unlike cellular mRNAs, however, Ty1 transcripts lacking a poly(A) tail do not undergo rapid degradation (19,20). Ty1 gRNAs' unusual stability (21) contributes to its high abundance, comprising about 0.8% of total cellular RNA (22). Cytoplasmic Ty1 gRNA is utilized as mRNA for translation of Gag and Gag-Pol polyprotein or is directed to cytoplasmic foci, termed retrosomes (19), where full-length Ty1 gRNA and proteins assemble into virus like particles (VLPs) (23). Within the VLP, Ty1 gRNA is reverse transcribed to a linear dsDNA copy that integrates predominantly upstream of genes transcribed by RNA polymerase III (23–27). Numerous functional studies show that analogous to retroviruses, the 5' and 3' termini of Ty1 gRNA are rich in regulatory sequences directly involved in encapsidation, dimerization, cyclization and reverse transcription (25,28–32).

Secondary structure models of Ty1 gRNA derived by SHAPE (Selective 2' Hydroxyl Acylation analyzed by Primer Extension) probing have been proposed for the first 1482 nts (26%) of *ex vivo*, *in vivo* and *in vitro* Ty1 gRNA (33), and for monomeric and dimeric states of the first 560 nts of the transcript *in vitro* (30–32). However, the *in vivo* structure of Ty1 gRNA or other LTR-retrotransposon transcripts has not been determined. Moreover, as compared to *Escherichia coli*, plant or mammalian cells, much less is known about the folding of mRNA in yeast (34). The first transcriptome-wide measurement of yeast mRNA structure was performed *in vitro* using an enzymatic PARS method (Parallel Analysis of RNA Structure), which suggested that mRNA transcripts with similar biological functions or cellular localization have common structural features (35). PARS analysis also suggests that yeast mRNAs contain more secondary structure in the coding sequences (CDSs) compared to the untranslated regions (UTRs), and unstructured sequences adjacent to the start codon is positively correlated with translation efficiency (TE). In contrast, transcriptome-wide DMS-seq mapping *in vivo* suggests that yeast mRNAs are largely unfolded, and secondary structures are similar in the CDS and the UTRs (36). Also, the correlation between the CDS structure and ribosome occupancy is not detected. Therefore, reconciling RNA structures determined *in vitro* with those determined *in vivo* is required to understand how structure relates to function.

Here, we provide the first genome-wide *in vivo* RNA secondary structure model for the endogenous yeast ssRNA-RT virus Ty1. Using SHAPE probing, we establish the secondary structure of the entire Ty1 RNA genome in its native *in vivo* state as well as under defined *in vitro* conditions. We find that *in vivo* Ty1 gRNA undergoes significant remodeling, is more dynamic and strongly destabilized, in large part by translating ribosomes. Nevertheless, Ty1 gRNA still contains well-structured regions harboring functional *cis*-sequences. The present work increases our understanding of viral RNA folding in living cells, and helps explain the interplay between translation and RNA structure. Furthermore, we show that commonly used SHAPE reagents, NMIA, NAI and 1M7 robustly modify RNA in yeast, as evidenced by strong correlations between position-dependent reactivities *in vivo* for all three reagents.

MATERIALS AND METHODS

Yeast strain, media and growth conditions

Strain DG3412 used in this study is Ty1-less *Saccharomyces paradoxus* strain (DG1768) (*MAT α his3- Δ 200hisG ura3 gal3 Spo⁻*) containing pBDG202 (pGTy1-H3CLA/URA3/2 μ), a multicopy plasmid with Ty1 under the control of the *GALI* promoter (37,38). Strain DG3412 was grown in SC-Ura 2% raffinose broth at 30°C with constant shaking at 250 rpm. Saturated cultures were diluted to OD_{600 nm} of approximately 0.2 with SC-Ura 2% galactose broth to induce Ty1 expression from the *GALI* promoter. Cultures were grown to a final OD_{600 nm} of 1.0 at 22°C with constant shaking at 250 rpm. For antibiotic treatment, 10 mg/ml kasugamycin solution (Enzo Life Sciences) was added to final concentration of 1 mg/ml when the culture OD_{600 nm} reached 1.0, and cells were incubated with shaking for 20 min. The reference strain BY4742 (*MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0*) was used for NMIA testing in *S. cerevisiae* and the growth conditions were as above except that SC medium was used instead of SC-Ura.

E. coli strain, media and growth conditions

Escherichia coli DH5 α cells (Invitrogen) were grown in LB medium without antibiotics at 37°C overnight with constant shaking at 225 rpm.

In vivo RNA modification

A 100 ml culture of exponentially growing yeast cells or 1 ml overnight culture of *E. coli* cells were centrifuged, the cell pellet was washed with PBS and resuspended in 720 μ l of PBS or PBS with 1 mg/ml kasugamycin. Each sample was divided into two equal amounts and treated with 40 μ l SHAPE reagent in DMSO (10 mM NMIA, 10 mM 1M7 or 20 mM NAI, final concentration) or 40 μ l DMSO alone. For each reagent, different concentrations were tested and the concentration yielding repeatable, high quality results, without over-modification signals, were chosen for final experiments. The modification reactions were carried at 37°C for 15 min (NMIA and NAI) or for 5 min (1M7). For NMIA, longer reaction time (45 min) was also tested and significant differences in the reactivity patterns were not observed. RNA modification reaction with NAI was quenched by adding 400 μ l of 1 M DTT. Cells were collected at 8000 \times g (10°C) for 7 min followed by total RNA isolation.

Isolation of total RNA from cells

Yeast pellets were resuspended in 1 ml of lysis buffer (10 mM Tris-HCl, pH 8.5, 5 mM EDTA, 2% SDS, 2% 2-mercaptoethanol) and incubated at 83°C for 20 min with constant shaking at 450 rpm. The tubes were centrifuged at 12 000 \times g for 5 min. The supernatants were transferred to a fresh tubes and RNA was extracted twice with phenol (pH 8.0), followed by two extractions with phenol: chloroform (pH 4.5). RNA was recovered by LiCl precipitation overnight at -20°C. RNA pellets were washed twice in 70% ethanol and resuspended in 30 μ l of water. Purified RNA

samples were stored at -20°C . RNA from *E. coli* cells was isolated using Direct-zol RNA MiniPrep Kit (Zymo Research) according to the manufacturer's protocol, and additionally purified by ethanol precipitation and resuspended in 10 μl of water.

***In vitro* transcription**

pBDG433, which contains full-length Ty1 DNA under the control of the SP6 promoter, was digested with BamHI restriction enzyme (Thermo Scientific) to generate a linear sense-strand template for transcription *in vitro*. DNA was phenol–chloroform extracted, ethanol precipitated and resuspended in 20 μl of water. Ty1 RNA was synthesized using MEGAscript™ SP6 Transcription Kit (Invitrogen) and recovered by LiCl precipitation overnight at -20°C , and then washed twice in 70% ethanol. Purified RNA was resuspended in 30 μl of water. The integrity of the RNA was determined by high-resolution gel electrophoresis (1.5% agarose gel with formaldehyde). RNA was stored at -20°C .

***In vitro* NMIA modification**

Ty1 *in vitro* transcript (3 pmol) was incubated in 100 μl of refolding buffer (10 mM Tris–HCl, pH 8.0, 100 mM KCl, 0.1 mM EDTA) at 95°C for 3 min, then cooled by placing on ice. RNA was incubated at 37°C for 30 min following the addition 50 μl of $3\times$ folding buffer (120 mM Tris–HCl, pH 8.0, 600 mM KCl, 1.5 mM EDTA, 15 mM MgCl_2). The RNA sample was divided into two separate tubes and treated with 8 μl of NMIA in DMSO (1mM NMIA, final concentration) or DMSO alone. The reaction was performed as described above for modification of RNA *in vivo*. RNA was recovered by ethanol precipitation and resuspended in 10 μl of water.

Primer extension and data processing

The optimal amounts of modified (+) and control (–) RNAs were mixed with 1 μl of 10 μM fluorescently labeled primer [Cy5 (+) or Cy5.5 (–)] and 1 μl 2 mM EDTA (for *in vivo* additionally with 3.84 μl of 5M betaine) to a final volume 12 μl . Primer–template solutions were incubated at 95°C for 3 min, 37°C for 10 min and 55°C for 2 min, then 8 μl of reverse transcriptase mix (SuperScript III, Invitrogen) was added to each tube and RNA was reverse transcribed at 50°C for 45 min (*in vitro*) or 90 min (*in vivo*) to provide optimal reaction efficiency. Sequencing ladders were prepared using primers labeled with WellRed D2 (ddC) or LicorIRD-800 (ddA) and a Thermo Sequenase Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's protocol. Samples and sequencing ladders were purified using ZR DNA Sequencing Clean-up Kit (Zymo Research) and analyzed on a GenomeLab GeXP Analysis System (Beckman-Coulter). Electropherograms were processed using ShapeFinder software (39), normalized as described previously (40). Exemplary raw electropherograms are presented on Supplementary Figure S1. Reactivity information was obtained from eighteen overlapping reads of ~ 350 nts each and at least two independent experiments were obtained for each read. Primers are listed in Supplementary Table S1.

Signal-to-background ratio and correlation calculation

Signal-to-background (S/B) ratio (Figure 1B) was calculated based on the output files from ShapeFinder. Medians for signal and background were estimated from the averaged values of signal or background peak areas for each nucleotide. S/B is the ratio of the obtained signal and background medians. Spearman's correlations and linear regressions were computed using the GraphPad Prism 8 software.

RNA secondary structure modeling and analysis

SuperFold software (41) was used for minimum free energy (MFE) secondary structure prediction, calculation of base pairing probabilities for all possible canonical base pairs and identification of regions with well-defined structures. SuperFold uses Partition and Fold functions implemented in RNAstructure (42) and the SHAPE reactivities were used as a pseudo-energy constraints. For all calculations slope and intercept folding parameters were set to 1.8 and -0.6 kcal mol $^{-1}$, respectively.

Prediction of RNA pseudoknots. ShapeKnots (implemented in RNAstructure) was used to identify pseudoknots in the *in vivo* and *in vitro* Ty1 RNAs. Folding was performed in 600-nt sliding windows with 100-nt increments. Four additional folds were computed at the 5' and 3' ends to increase sampling of terminal regions. Pseudoknots occurring in $>50\%$ of the windows were included in the MFE structure.

Partition function calculation. Partition was run in 1200-nt sliding windows, incremented by 100 nts, with four additional calculations performed on the 5' and 3' ends of the Ty1 RNA. Next, 300 nts were trimmed from the 5' and 3' ends of each window to eliminate end effects. The base-pairing distance was limited to 600 nts. Base-pair probabilities from the multiple partition function windows were combined and pseudoknot nucleotides were involved. Shannon entropies were calculated from individual base-pairing probabilities and combined into a single profile in 55-nt sliding windows.

Minimum free energy structure prediction. The MFE structure models were predicted using the Fold function with pseudo-energy constraints, and base pairs with $>99\%$ probability were used as hard pairing constraints. Fold was performed in 3000-nt sliding windows, incremented by 300 nts to predict individual MFE structures over each window, with four additional fold calculations on the 5' and 3' ends of the genome. Next, 300 nts were trimmed from the 5' and 3' ends of each window to eliminate end effects. These multiple folds were then combined into a final MFE structure by requiring that base pairs appeared in a majority of the windowed folds. The pseudoknotted helices were added to the final RNA structure. The maximum pairing distance was set to 600 nts. Additional prediction of Ty1 MFE structure was performed with tRNA $_{i}^{\text{Met}}$ binding, dimerization and cyclization sequences forced to be single-stranded. The MFE structure model of the *in vivo* Ty1 RNA + 1–1482 region was generated based on previously published SHAPE data (33), using the RNAstructure.

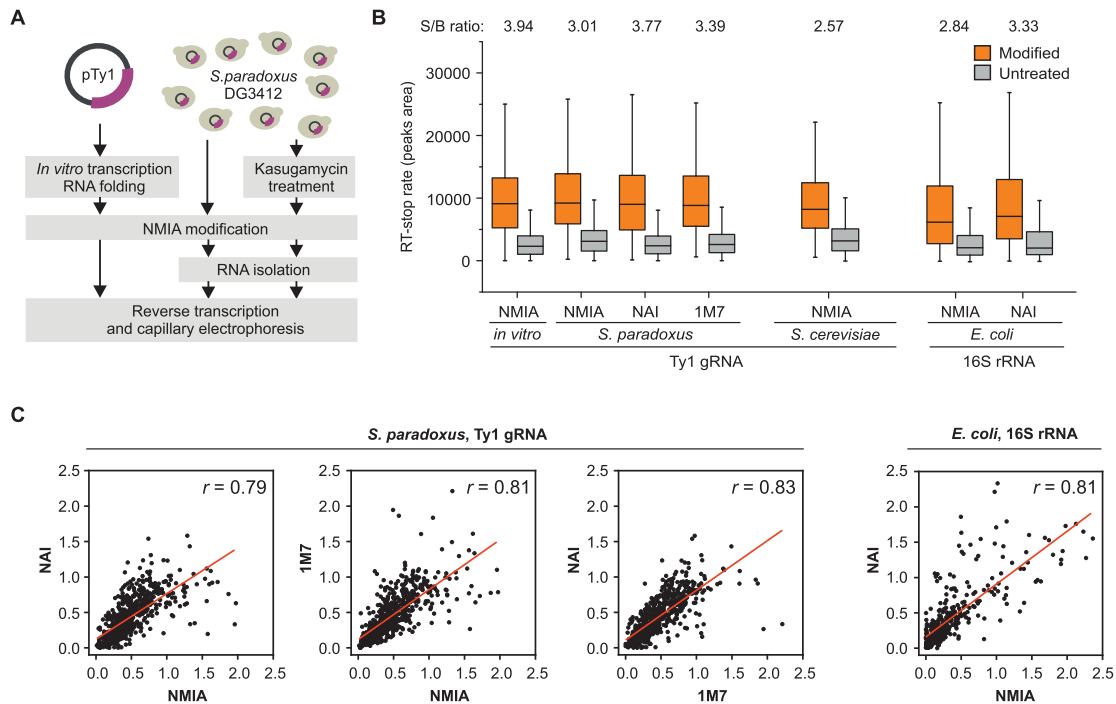


Figure 1. Experimental workflow and SHAPE reagent comparison. (A) Experimental workflow. (B) Box plot analysis of RT-stop rate (CE peaks area) measurement with medians for signal and background from capillary electrophoresis experiments for different SHAPE reagents and different cells. Plots present data for approximately 600 nts of Ty1 gRNA *in vitro* and in *S. paradoxus* and approximately 400 nts of Ty1 gRNA in *S. cerevisiae* and 400 nts of 16S rRNA in *E. coli*. (C) Correlation of position-dependent NMIA, NAI and 1M7 reactivities from *in vivo* probing of Ty1 gRNA in *S. paradoxus* and 16S rRNA in *E. coli*.

Identification of regions with well-defined structures. Well-defined regions (lowSS) were identified by selecting regions of at least 40 nts with both median SHAPE reactivities and median Shannon entropies below the global medians, as described previously (43). Some lowSS regions were expanded or combined to preserve the entirety of helices predicted in the MFE structure. The pseudoknot helix was ignored in lowSS regions identification, because its presence caused the lowSS region 1 extension to high entropy and low base-pair probability regions. RNA structures were visualized with VARNA (<http://varna-gui.software.informer.com/>) (44).

Sensitivity and positive prediction value (PPV) calculations. The sensitivity and PPV for the obtained MFE models were calculated using Scorer function (implemented in RNAs-structure). If the region of interest contained only half the predicted helix, calculations were performed with the manual addition of nucleotides from disrupted helix. For the entire Ty1 RNA and identified lowSS regions, the *in vitro* structure was used as the reference state while in other cases the reference structure was indicated in the figures. Sensitivity is defined as the percentage of base pairs in the reference MFE structure that are also present in the compared MFE structure, and the PPV represents the percentage of base pairs in the MFE structure that are also present in the reference model (45).

RESULTS

High-resolution probing of Ty1 genome under *in vivo* and *in vitro* conditions

We used SHAPE followed by capillary electrophoresis to explore the structure of gRNA of an active Ty1 retrotransposon at single-nucleotide resolution. We probed Ty1 gRNA structure under three experimental conditions: in *S. paradoxus* cells expressing an active Ty1 element, *in vitro* in the absence of cellular components, and in cells partially inhibited for translation following treatment with kasugamycin (Figure 1A). The parental yeast strain DG1768 lacks full-length chromosomal Ty1 sequences [(37) and Garfinkel and Bergman, unpublished results], and expression of Ty1 occurs exclusively from a multicopy pGTy1 plasmid driven by the inducible *GALI* promoter (23,46). Galactose induction produces a high level of correctly initiated Ty1 RNA that participates in all steps of the replication cycle, including translation, co-localization into retrosomes, encapsidation into VLPs, and reverse transcription (38). The resulting cDNA copy integrates into host DNA, and a high frequency of Ty1 retromobility can be measured. Based on the available data, we assume that the major fraction of Ty1 RNA is present in the cytoplasm or associated with retrosomes, and a much smaller amount is packaged in VLPs (38). As this approach allowed us to obtain SHAPE reactivity data from a homogenous Ty1 gRNA population, a more reliable RNA secondary structure model was gener-

ated. In contrast, the widely used *S. cerevisiae* S288c laboratory strain background contains 32 full length Ty1 elements comprised of three subfamilies: Ty1, Ty1/Ty2 hybrids, and the ancestral Ty1' element (15,47,48). Ty1 elements are expressed at different levels that vary up to 50-fold, with 75% of total Ty1 expression coming from eleven highly expressed elements, and eight of these are Ty1/Ty2 hybrids (49). Additionally, some full length Ty1 copies contain deletions and mutations that may affect transposition, but they can be still expressed.

SHAPE exploits 2'-hydroxyl-selective reagents to interrogate local backbone flexibility in RNA by forming covalent 2'-*O*-adducts at flexible nucleotides (50). Generally, SHAPE reactivities are inversely proportional to local nucleotide conformations, such that flexible or unpaired nucleotides are reactive, while structurally constrained or base-paired nucleotides tend to be less reactive (51). Diverse cell types have different permeabilities to SHAPE reagents, and in general, those growing in a suspension culture are more permeable than adherent cell lines (52). Surprisingly little is known about yeast permeability to SHAPE reagents and only NAI (2-methylnicotinic acid imidazolide) has been used in SHAPE analyses in yeast (53–55). To the best of our knowledge, other well-validated SHAPE reagents, such as NMIA (N-methylisatoic anhydride) or closely related 1M7 (1-methyl-7-nitroisatoic anhydride) have not been used for RNA structure mapping in yeast, however, their ability to modify RNA occurs in other cell types (52,56–59). In contrast, one group reports that NMIA reactivity is not detected in an adherent mouse cell line (mESCs) (53), and a very low 1M7 signal is detected in mESCs cells and *E. coli* (60).

SHAPE reagents have local-structure and nucleobase biases that can impact reactivity profiles (52,61,62). All prior studies of Ty1 RNA structure have been performed using NMIA as the probing reagent (30,32,33,63). We reasoned that the application of the same SHAPE reagent for Ty1 RNA structure mapping *in vivo* would allow for direct and unbiased comparison of reactivity patterns with those presented for Ty1 RNA 5' terminal region *in vitro* and *ex vivo* (33). This approach also facilitates detection of NMIA reactivity changes *in vivo* that result from Gag-induced RNA dimerization (30), pseudoknot formation (32,33), and primer tRNA annealing (30,33). Therefore, we initially tested the ability of NMIA to modify the 5' end of Ty1 gRNA in yeast, and compared this data with that obtained from NAI and 1M7 probing. We clearly detected high reactivity signals well above background for all three reagents using capillary electrophoresis for analysis of adduct-induced cDNA truncations (Figure 1B). Since the signal-to-background ratio is the raw measurement of structure-dependent RNA modification, these results show that NMIA, NAI and 1M7 robustly modify RNA in *S. paradoxus*. Additionally, *E. coli* and *S. cerevisiae* treated with NMIA also displayed high reactivity above the background (Figure 1B). Our data indicate that NAI is slightly more efficient than NMIA in both yeast and *E. coli*, but a recent study suggests that NAI has stronger nucleotide biases and less effectively differentiates paired and unpaired C and G residues (52). Despite that, we observed strong correlations between position-dependent NMIA, 1M7 and NAI

reactivities *in vivo*, but some local single-nucleotide reactivity differences were also detected (Figure 1C). Therefore, the reactivity patterns obtained with the same SHAPE reagents are more informative for detailed analysis of RNA structure alterations specific for diverse biological or experimental states. Using NMIA, we obtained high-quality SHAPE data for 98% of the nucleotides in the Ty1 gRNA *in vivo* and under defined conditions *in vitro*. SHAPE data derived from *in vivo* and *in vitro* experiments were used for structural modeling of Ty1 gRNA using SuperFold pipeline, which incorporates a windowing approach along with the *Partition* and *Fold* functions implemented in RNAstructure (41,42).

Impact of cell environment on SHAPE reactivities, base-pairing probabilities and Shannon entropies

To determine the impact of cell environment on Ty1 gRNA folding, we compared the *in vivo* and *in vitro* SHAPE reactivities and base-pairing probabilities combined with the Shannon entropy calculation. The difference in overall median SHAPE reactivities between *in vivo* and *in vitro* Ty1 gRNAs was smaller than expected (0.07) for such distinct environmental conditions (Figure 2A). However, the correlation between datasets from these two states was moderate ($r = 0.49$), which indicates a significant change in the pattern of SHAPE modifications *in vivo* (Figure 2B). In contrast to experiments under defined *in vitro* conditions, SHAPE reactivities *in vivo* are dependent not only on local nucleotide flexibility, but also on interactions with proteins or other RNA molecules (59,64). In general, increased *in vivo* SHAPE reactivities, relative to the *in vitro* state, usually reflect RNA conformational changes induced by the cellular environment, whereas decreased reactivity *in vivo* tends to reflect the binding of proteins and other biomolecules (65). The comparison of median SHAPE reactivity profiles across the Ty1 RNA genome showed that many regions were significantly more reactive *in vivo* than *in vitro*, some were less modified *in vivo*, and several had quite similar reactivity in both states (Figure 2C, D). Overall, the differences in SHAPE reactivity between both states (Δ SHAPE, Figure 2D) suggest there is a strong destabilization of Ty1 gRNA in the cellular environment, and the few specific regions protected from NMIA modification *in vivo* may serve as protein binding sites. However, all protected regions are probably not detected due to strong destabilization of RNA structure *in vivo* that reduces the impact of protein binding on the total SHAPE reactivity measurement.

The difference in SHAPE reactivity between *in vivo* and *in vitro* measurements was the most evident for adenosine residues and the smallest difference was calculated for cytidines (Figure 2E). Since Ty1 gRNA is particularly rich in A-residues (35.5%) while the C-content is low (21.9%), specific Ty1 gRNA nucleotide content likely contributes to differences between *in vivo* and *in vitro* SHAPE reactivity profiles.

Ty1 gRNA contained almost 50% less unreactive nucleotides *in vivo*, but was richer in nucleotides with intermediate SHAPE reactivity ($0.4 > 0.85$) (Figure 2F). The SHAPE reactivity at any given nucleotide position does not provide information about a single RNA structure but represents the average reactivity over the ensemble of

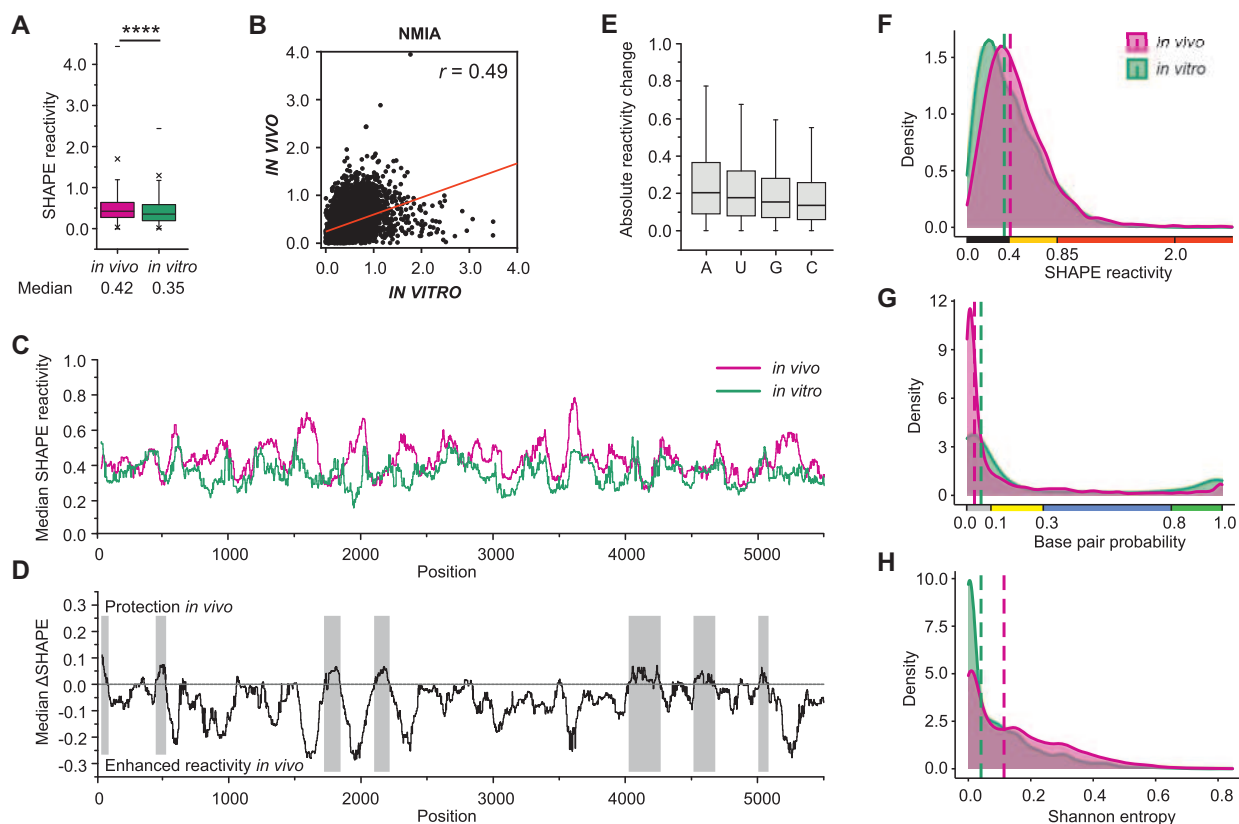


Figure 2. SHAPE reactivities, pairing probabilities and Shannon entropies for *in vivo* and *in vitro* Ty1 gRNAs. (A) Box plot analysis of SHAPE reactivity distributions with medians. Significance was computed by unpaired two-tailed Mann–Whitney test; **** $P < 0.0001$. (B) Correlation of position-dependent NMIA reactivities for *in vivo* and *in vitro* Ty1 gRNA states. (C) The median SHAPE reactivity profiles smoothed with a 75-nt window. (D) The median Δ SHAPE profile calculated by subtracting the *in vivo* reactivities from those of the *in vitro* RNA, smoothed with a 75-nt window. Gray shadings indicate regions that may serve as protein binding sites. (E) Distribution of the absolute difference in reactivity between *in vitro* and *in vivo* SHAPE measurements, calculated for each nucleotide type. (F) NMIA reactivity distributions with medians (dashed lines). (G) Base pair probability distributions with medians. (H) Shannon entropy distributions with medians.

all coexisting RNA structures (66,67). Nucleotides with low SHAPE reactivity (<0.4) are conformationally constrained and likely to be double-stranded in most conformers, whereas those with intermediate reactivity may be involved in diverse coexisting RNA conformations (68). To consider the entire structural ensemble, we calculated a probability for each base pair across all SHAPE-directed structures predicted for the Ty1 RNA *in vitro* and *in vivo* (45). A significant increase in the number of low-probability base pairs and a concurrent decrease of high-probability pairings was detected *in vivo* (Figures 2G and 3C, F). Consistent with these findings, Shannon (base-pairing) entropy distributions indicated that 2-fold fewer nucleotides form well-defined structures (with low entropy) *in vivo* and the median entropy was significantly higher than *in vitro* (0.11 and 0.04, respectively) (Figures 2H and 3B, E). Taken together, these results strongly suggest that Ty1 gRNA structure is destabilized in yeast cells and is more heterogeneous than under *in vitro* conditions.

We queried the RASP database (69) to determine whether structural information for Ty1 transcripts can be retrieved from published transcriptome-wide studies in yeast. The only available data are from DMS-MaPseq experiments *in vitro* and *in vivo* (70). However, in all Ty1 transcripts, only

about 30% of nucleotides are mapped and all transcripts contain many long 100–200 nts sequences without structural data. In addition, we observe a low correlation between SHAPE and DMS reactivity patterns for fragments with higher DMS data coverage. This may result from different probing reagents and methods of adduct detection.

Identification of Ty1 RNA regions with well-determined secondary structure

Long RNA molecules often adopt a mosaic structure, such that conformationally dynamic regions are interspaced with regions with well-determined secondary structure (43,66,71,72). Importantly, RNA regions with persistent stable structural motifs tend to be functionally important and contain *cis*-acting regulatory sequences (73–75). Based on low SHAPE and low Shannon entropy metrics (lowSS regions), we identified 11 lowSS regions within *in vivo* and *in vitro* Ty1 RNAs (Figure 3B,E; numbered dark violet strips). They were separated by conformationally dynamic regions that exhibited higher Shannon entropy even if SHAPE values remained below the global median reactivity. The 5' and 3' terminal lowSS regions (regions 1 and 11) detected *in vivo* correspond well with sequences required

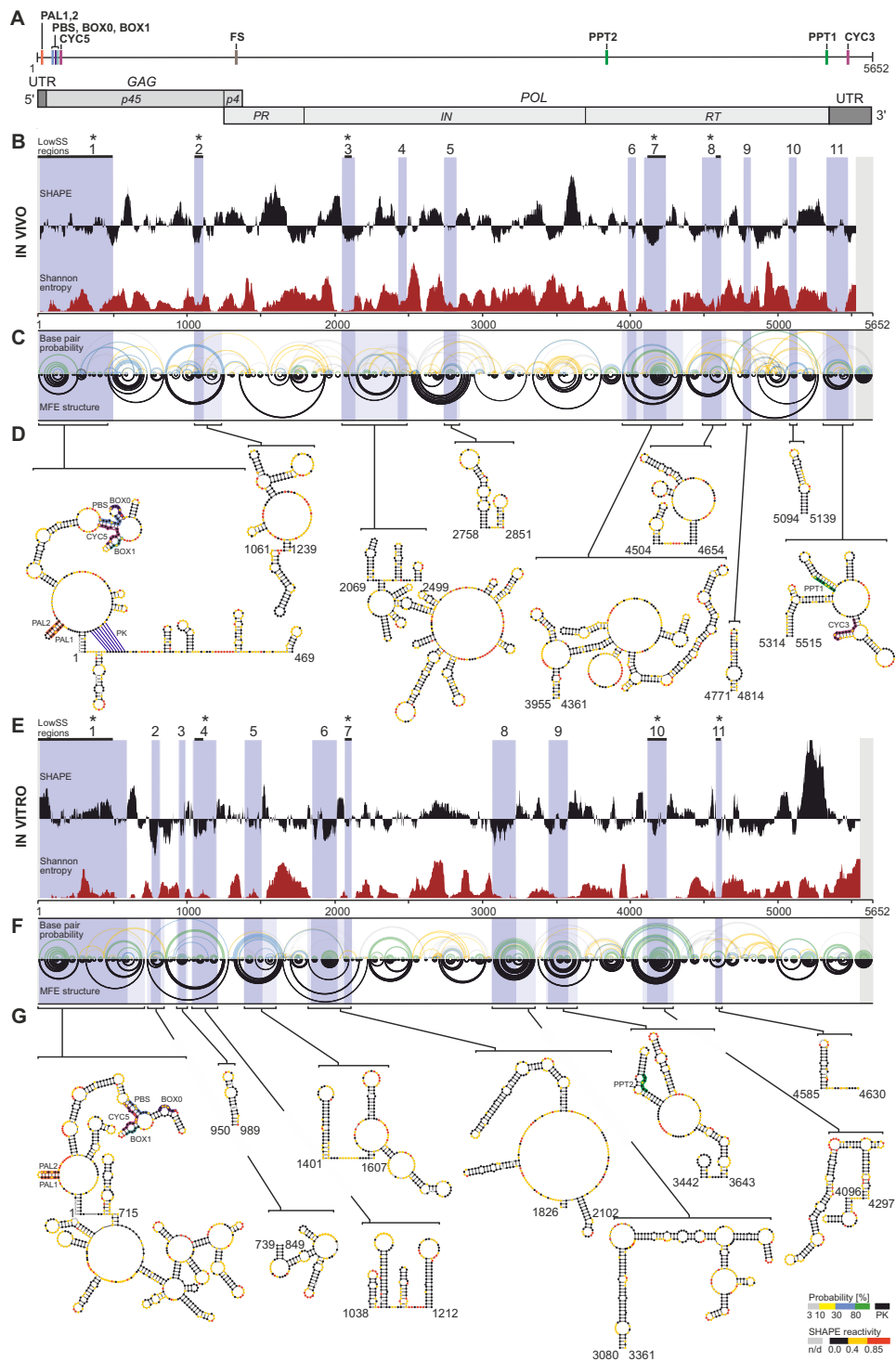


Figure 3. SHAPE-directed structure models and structurally well-defined regions for *in vivo* and *in vitro* Ty1 gRNAs. (A) Functional organization of Ty1 gRNA. The known *cis*-acting sequences are indicated: dimerization sequences (PAL1, PAL2), primer-binding sites (PBS, BOX0, BOX1), cyclization sites (CYC5, CYC3), frameshift site (FS) and polypurine tracts (PPT1, PPT2). (B) *In vivo* median SHAPE reactivity and Shannon entropy profiles smoothed with a 55-nt window. LowSS regions are marked by dark violet shadings with light violet shadings extended to encompass entire intersecting helices from MFE structures. Grey shadings indicate regions without experimental data. Stars mark regions that overlap in both Ty1 gRNAs and black horizontal lines indicate overlapping fragments. (C) Arc plots showing the base-pairing probabilities and the predicted MFE structure for *in vivo* Ty1 gRNA. At the top base pairs are colored with respect to their probabilities (see scale), black arcs indicate a predicted pseudoknot (PK). (D) Secondary structure models for *in vivo* lowSS regions. Nucleotides are colored by SHAPE reactivity (see scale). Panels (E), (F) and (G) show the analogous analysis performed for *in vitro* Ty1 gRNA. The MFE structure models were predicted with the maximum pairing distance of 600 nts, and they do not include CYC5/CYC3 long-range interaction. High-resolution structures of lowSS regions and entire Ty1 gRNAs are provided in Supplementary Figures S2 and S3.

for transposition of a minimal Ty1 element when Gag and Gag-Pol are provided *in trans* (28,76). The 5' and 3' ends of Ty1 RNA contain *cis*-acting sequences directly involved in retromobility, such as palindromes (PAL) required for Ty1 RNA dimerization, PBS, BOX0 and BOX1 that anneal with primer tRNA_i^{Met}, CYC5 and CYC3 required for genome cyclization, and the polypurine tract, which is required for plus-strand synthesis (23,25,28–30,32) (Figure 3A). These important RNA sequences were embedded in low SS regions identified *in vivo*. Thus, our data highlight the correlation between well-determined RNA conformation and biological function. An interesting but expected exception is the sequence mediating Ty1 translational frameshifting. Although the frameshift heptamer is not nested within a well-structured region, Ty1 frameshifting depends on the low abundance of a specific isoacceptor tRNA and does not require a specific RNA structure (77).

Only 5 lowSS regions detected *in vivo* overlapped with those identified *in vitro* (Figure 3B,E; strips marked with asterisks), while others were unique to the *in vivo* or *in vitro* state. Interestingly, the 5' terminus containing the majority of known *cis*-acting sequences was well-structured both *in vivo* and *in vitro*. However, the 3' terminal region with CYC3 and PPT1 was well-structured only in the *in vivo* state. Thus, our data suggest that detection of regions meeting low SHAPE and low entropy criteria *in vivo* might enable more precise identification of the positions of unknown regulatory RNA sequences. Here, we identified eight lowSS regions in the POL ORF of *in vivo* Ty1 RNA. Although their functions remain to be determined, four of them (regions 3, 6, 7 and 8) overlapped well with the regions protected from NMIA modification *in vivo*, perhaps resulting from protein binding (Figure 2D).

A prior study of the entire HIV-1 RNA genome structure *in vitro* shows decreases in SHAPE at RNA regions that encode protein domain junctions (78). These more structured RNA regions are proposed to induce ribosome pausing that facilitates proper protein folding. However, we do not detect decreases in SHAPE reactivity in regions of Ty1 gRNA that encode protein domain boundaries *in vivo* or *in vitro*. Our data indicate that such regions in Ty1 gRNA exhibit equal or higher reactivity values than the median reactivity in each state. A decrease in SHAPE reactivity at boundary RNA regions is also not detected for Zika and Dengue gRNA *in vivo*, suggesting different strategies of ribosome elongation regulation (3). Indeed, RNA structure is one of the mechanisms promoting ribosome pausing, but there are also other contributors to this process, such as mRNA sequence, charged tRNA and translation factor availability, and the nascent protein chain itself (79).

Comparison of minimum free energy structure models of the *in vivo* and *in vitro* Ty1 RNA genome

The consensus minimum free energy (MFE) secondary structure of Ty1 gRNA analyzed *in vivo* and *in vitro* was predicted using the SuperFold pipeline, which is dedicated to modeling structures of large RNAs (Figure 3C, F and Supplementary Figure S3) (41). SuperFold takes a windowing approach, combines base pairs predicted for overlapping windows, incorporates base pairs with the highest probabili-

ties (>99%), and the final structure is predicted by requiring base pairs that occur in more than one-half of the windows. Comparing the resulting MFE structures in terms of sensitivity and positive predictive value (PPV) showed a significant difference in the folding of Ty1 RNAs. Both parameters were in the range of 49–55%, which indicated that only about half of the base pairs were shared between *in vitro* and *in vivo* MFE structures, while the rest of the pairings remained unique for each model (Figure 4A, B). Despite these differences, roughly 50% of the nucleotides were predicted to be base-paired in both MFE structures, suggesting similar levels of Ty1 gRNA structure *in vitro* and *in vivo* (Figure 4E). In contrast, analysis of SHAPE reactivities, base-pairing probabilities and Shannon entropies strongly suggest that Ty1 gRNA structure is destabilized in yeast cells. A single MFE model, even predicted as a consensus structure, represents one conformer from an ensemble of possible RNA conformers with similar free energies (ΔG) (42,45). Thus, focusing on a single MFE structure can lead to an incomplete or biased understanding of RNA folding, especially for structurally dynamic long RNAs.

We reasoned that alterations occurring in well-structured regions (lowSS) may better reflect changes in Ty1 gRNA folding under different experimental conditions. Therefore, we compared the MFE structures of Ty1 RNAs across all identified *in vivo* and *in vitro* lowSS regions by calculating a sensitivity value and PPV for these selected RNA segments. Most of the base pairs predicted in the *in vivo* lowSS regions were retained in the *in vitro* MFE structure, even if they were not in lowSS regions of the *in vitro* RNA (median sensitivity and PPV – 92% and 76%) (Figure 4C). This trend was not observed when we compared base pairs from *in vitro* lowSS regions with the *in vivo* MFE structure (median sensitivity and PPV – 30.5% and 38.5%) (Figure 4D). However, higher structural similarity was found for lowSS regions that partially overlapped in *in vitro* and *in vivo* Ty1 gRNA. Lower sensitivity and PPV metrics for overlapping region 1 can be caused by *cis*-acting sequences that mediate inter- and intramolecular RNA interactions that may occur *in vivo* but not *in vitro* (see below). Together, these analyses suggest that a large fraction of well-defined structural motifs detected *in vivo* are preserved *in vitro*, but not vice versa.

Analysis of base-pairing probabilities supports destabilization of Ty1 RNA *in vivo*

To better understand Ty1 gRNA folding, we repeated the comparative analysis of the *in vivo* and *in vitro* Ty1 gRNA structures considering only highly-probable base pairs (HP bps, pairing probability > 80%). HP bps are predicted with the highest confidence and are more likely to be present in the gRNA structure (45). Although a similar number of MFE base pairs was predicted for both Ty1 gRNA states, the contribution of high-probability base pairs in the *in vivo* MFE structure was much lower. Accordingly, 17% of Ty1 gRNA nucleotides participated in HP bps *in vivo*, representing 37.5% of the base pairs in the MFE structure, whereas 34.1% of nucleotides were involved in HP bps *in vitro*, and they represented 68% of the MFE base pairs (Figure 4E, F). We found that only 39% of *in vitro* HP bps were also predicted to occur in the *in vivo* MFE structure, while

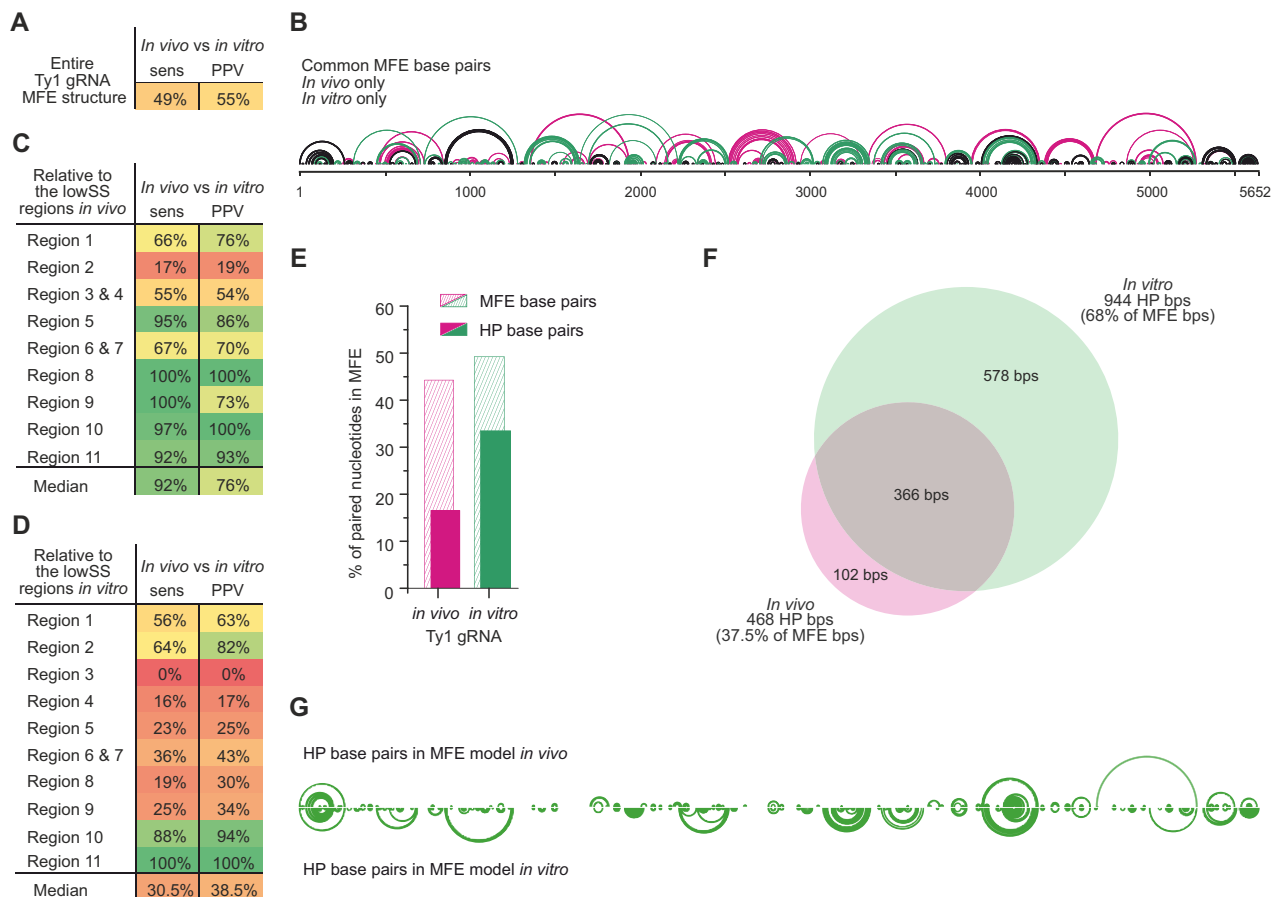


Figure 4. Comparison of SHAPE-directed MFE structures predicted for *in vivo* and *in vitro* Ty1 gRNAs. (A) Calculation of sensitivity (sens) and PPV parameters for the MFE structures of *in vivo* and *in vitro* Ty1 RNAs. *In vitro* Ty1 gRNA MFE structure was used as a reference structure. Values are colored from low (red) to high (green). (B) Combined model of MFE structures of *in vivo* and *in vitro* Ty1 gRNAs. Common base pairs are indicated by black arcs, base pairs present only *in vivo* by pink arcs and those unique for *in vitro* gRNA by dark green arcs. Calculation of sensitivity and PPV parameters for lowSS regions identified *in vivo* (C) and *in vitro* (D). (E) Percentages of paired nucleotides in the predicted MFE structures, including high probability base pairs (HP bps, pairing probability > 80%). (F) Venn diagram showing overlap of highly probable base pairs between *in vivo* and *in vitro* MFE structures. (G) Comparison of high probability base pairs in the *in vivo* and *in vitro* MFE structure models of Ty1 gRNA.

almost 80% of *in vivo* HP bps were shared with *in vitro* MFE structure (Figure 4F). This disparity arises from the lower number of HP bps predicted for the *in vivo* model of Ty1 gRNA. We observed that among HP bps the short-range interactions representing locally stable secondary RNA structures were preferentially shared between both Ty1 gRNA states (Figure 4G). In contrast, a majority of highly probable longer-range base pairings *in vitro* were disrupted or became less probable *in vivo*. These analyses confirm that Ty1 gRNA possesses a significantly lower and more dynamic structure *in vivo*, and show that only part of stable (highly probable) base pairs *in vitro* are retained *in vivo*, consistent with the findings from comparison of the MFE structures across lowSS regions.

Ty1 RNA dimerization, cyclization and tRNA_i^{Met} annealing *in vivo*

We analyzed SHAPE reactivity changes in the 5'-end of Ty1 gRNA in more detail to establish whether critical RNA-RNA interactions could occur *in vivo* prior to packaging

into VLPs. Ty1 gRNA dimerization, cyclization or primer tRNA annealing do not occur under the *in vitro* conditions employed here because Ty1 proteins such as Gag and tRNA are absent in the analysis. RNA sequences involved in tRNA_i^{Met} binding (PBS, BOX0, and BOX1) and genome dimerization (PAL1 and PAL2) were engaged in intramolecular base-pairing *in vitro*, and we did not detect changes in reactivity patterns that could directly support these processes occurring *in vivo* (Figure 5A). This problem arises from a limitation of the SHAPE method that, like other classical RNA structure mapping techniques, cannot discriminate intra- from intermolecular base-pairing. However, changes in reactivity of nucleotides separating PALs were detected *in vivo*, analogous to those detected in VLPs and in Gag-induced Ty1 RNA dimerization *in vitro*, and supporting PALs involvement in the intermolecular PAL1-PAL2 interactions (30,33). We also detected SHAPE reactivity increases for nucleotides between the PBS and BOX0 (105–107) that occur in Gag- or temperature-induced tRNA_i^{Met}/Ty1 RNA *in vitro* complexes and that are also observed in VLPs (30,33). Highly significant reactivity in-

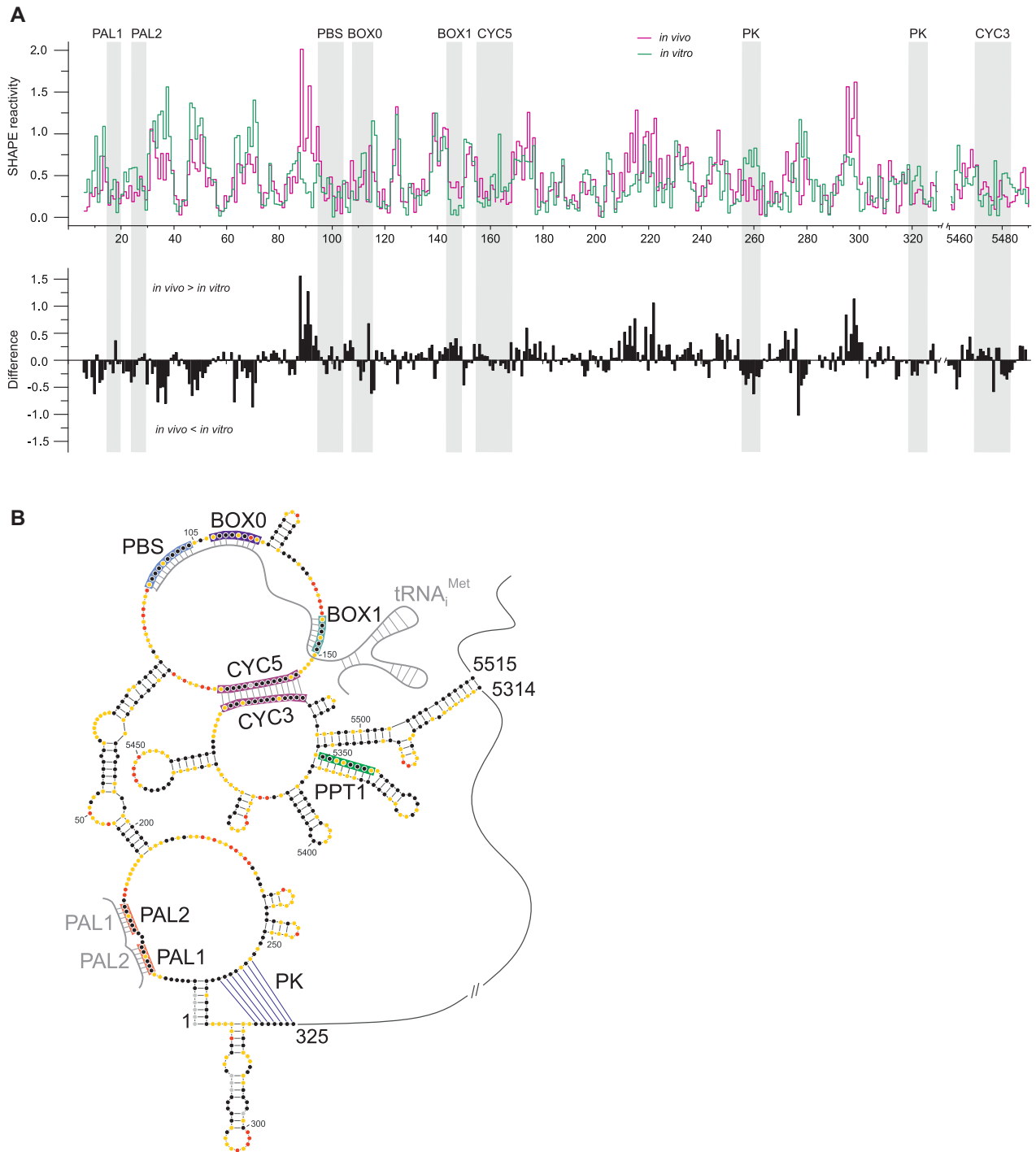


Figure 5. SHAPE-based analysis of Ty1 gRNA dimerization, cyclization and tRNA^{Met} annealing. **(A)** The step plot (top) of NMIA reactivity for *in vivo* and *in vitro* gRNAs and the difference plot (bottom) calculated by subtracting the *in vitro* reactivities from those of the *in vivo* gRNA. Negative values indicate nucleotides that are less reactive *in vivo*. Locations of *cis*-acting sequences are indicated by gray stripes. **(B)** Ty1 gRNA in its dimeric and circular form with annealed tRNA^{Met}. *Cis*-acting sequences are marked with colored boxes, nucleotides are colored by SHAPE reactivity according to the scale presented in Figure 2. For simplicity, only 5' and 3' ends are shown. The present RNA structure was predicted with PAL, CYC and tRNA^{Met} binding sequences forced to be single-stranded. The intermolecular base-pairs and long-range CYC5/CYC3 interaction were introduced manually based on differences in SHAPE reactivity patterns *in vitro* and *in vivo* and published data (30,32,33).

creases were also observed *in vivo* in the sequence (88–94) preceding the PBS (compare Figure 3D and G, leftmost structures). The majority of this sequence was base-paired and unreactive *in vitro*. *In vivo* SHAPE reactivity increases likely reflect disruption of the helical region preceding the PBS caused by tRNA_i^{Met} hybridization (Figure 5A, B). This interpretation is further supported by reactivity increases in the opposite strand (173–176) of the helix. In addition, destabilization of this helix may be enhanced by CYC5 interaction with CYC3. In the absence of tRNA_i^{Met}, CYC5 base-paired with the PBS and BOX1 *in vitro*, thus the reactivity increases in CYC5 were expected as a consequence of tRNA_i^{Met} hybridization *in vivo*. Since CYC5 remains unreactive *in vivo*, this region may be involved in long-range base-pairing with CYC3 at the 3' end of Ty1 gRNA (Figure 5B), as suggested by genetic studies (29). Although SHAPE can detect RNA structure changes resulting from long-range interactions, the prediction of such interactions is challenging because the accuracy of the predicted RNA structure model decreases with distance between base-pairs (80).

Importantly, the SHAPE data supported the presence of the functionally important pseudoknot at the 5' terminus of Ty1 gRNA *in vivo* (Figures 3D and 5A, B). The 5' pseudoknot was identified *in vitro*, and mutations disrupting its structure inhibit reverse transcription (32,33). Additionally, our current findings support previous work (33) indicating that the pseudoknot is not formed in the full-length Ty1 transcript *in vitro*.

Together, our data raise the possibility that functionally important RNA intramolecular interactions such as cyclization and pseudoknot formation occur in retrosomes or the cytoplasm prior to packaging of Ty1 gRNA into VLPs. The primer for reverse transcription may be annealed prior to packaging, which also occurs with primer tRNA^{Lys} annealing during HIV-1 propagation (81). Like retroviral RNA genomes, Ty1 gRNA may undergo dimerization prior to being packaged into VLPs. However, further confirmation of these observations is required.

Ty1 RNA structure changes during translation

Ty1 gRNA is both the template for reverse transcription and the mRNA for translation of Gag and the Gag-Pol polyprotein, thus translation may contribute to differences between Ty1 RNA structures *in vivo* and *in vitro*. The correlation between translation and *in vivo* mRNA structure is observed in transcriptome-wide structure probing experiments in *Escherichia coli* and zebrafish (73,82). However, interplay between ongoing translation and mRNA structure *in vivo* is not detected in global yeast transcriptome mapping, and the average *in vivo* structure of coding regions is not distinguishable from that of the UTRs (36). Interestingly, our results suggest Ty1 gRNA may not fit this pattern. The *in vivo* median SHAPE reactivities of GAG and POL ORFs were higher than that of UTRs, indicating that the Ty1 RNA coding region was less structured in cells than non-coding sequences, while this correlation was not detected *in vitro* (Figure 6A).

To analyze the relationship between translation and RNA structure, we probed Ty1 gRNA structure *in vivo* fol-

lowing addition of the aminoglycoside kasugamycin, an antibiotic that specifically inhibits the initial step of protein synthesis by preventing formation of the translation initiation complex (83). This experimental approach was used recently to show the impact of translating ribosomes on mRNA structure in *E. coli* (73). Cells were treated with kasugamycin for 20 min to allow ongoing translation elongation cycles to finish (84) followed by the addition of NMIA. Comparison of median SHAPE reactivity profiles revealed that in cells treated with kasugamycin Ty1 gRNA was less reactive toward NMIA, and thus is likely more structured than under native *in vivo* conditions (Figure 6B). As expected for mRNA undergoing translation, statistically relevant decreases in SHAPE reactivity was observed for ORFs ($P < 0.0001$) but not for UTRs (Figure 6C). To better understand the effect of translating ribosomes on RNA conformation, MFE structures of a +1–2500 fragment of Ty1 RNA were compared in each probing condition. Ty1 gRNA from kasugamycin-treated cells contains more MFE and HP base pairs than RNA *in vivo*, but less than RNA *in vitro* (Figure 6D). The structural model of +1–2500 Ty1 gRNA from kasugamycin-treated cells shared roughly 60% of MFE base pairs with the *in vivo* or *in vitro* RNA model (Figure 6E). Nevertheless, the Ty1 gRNA in kasugamycin-treated cells appeared more similar to the *in vitro* state, since the gRNAs shared more highly probable base pairs with the *in vitro* than *in vivo* structure (70% and 50%, respectively) (Figure 6F). Together, our results show that ribosomes participate in the unfolding and remodeling of Ty1 gRNA *in vivo*, and support the idea that inhibition of translation initiation partially stabilizes Ty1 gRNA structure.

DISCUSSION

We took advantage of the well-developed Ty1 experimental system to define the first genome-wide RNA secondary structure model for a retrovirus-like retrotransposon *in vivo*. To reveal the effect of the cellular environment on folding of the Ty1 RNA genome, we directly compare the SHAPE-derived *in vivo* structure with that obtained under defined *in vitro* conditions. Furthermore, we analyze the role of active ribosomes in RNA structural remodeling by probing Ty1 gRNA structure in yeast treated with a translation initiation inhibitor. We find a strong impact of the cellular environment on folding Ty1 gRNA and show that ribosomes participate in the unfolding and remodeling of Ty1 gRNA.

In support of previous observations made for viral RNAs (2,6) and cellular transcripts (36,53,82), we detect a moderate correlation between the *in vivo* and *in vitro* structure of Ty1 gRNA. About half of the MFE and <40% of HP base pairs are shared between *in vivo* and *in vitro* structure models, indicating significant remodeling of Ty1 gRNA occurs *in vivo*. Base-pairing probabilities and Shannon entropy calculations suggest that Ty1 gRNA adopts a significantly lower and more dynamic structure *in vivo* when compared to the structure *in vitro*. In the *in vivo* state, Ty1 gRNA contains 50% less HP base pairs, and is more likely to form alternative less probable pairings. The observed increase in structural heterogeneity *in vivo* may result from the parallel occurrence of various Ty1 gRNA conformers or from structural transitions through different stages of the replication

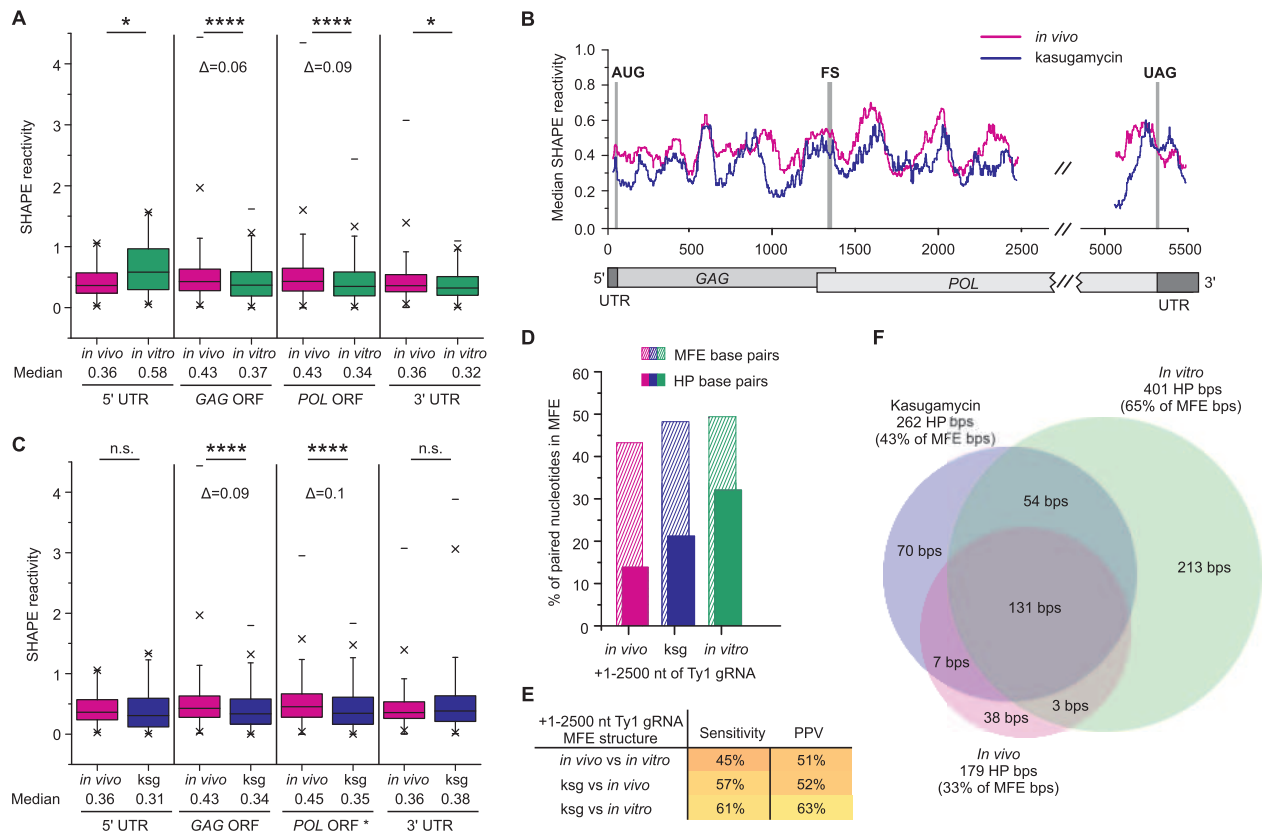


Figure 6. Correlating Ty1 RNA structure and translation. (A) Box plot analysis of SHAPE reactivity distributions with medians for coding and untranslated regions of *in vivo* and *in vitro* Ty1 gRNAs. (B) Median SHAPE reactivity profiles (smoothed with a 75-nt window) for Ty1 gRNA from cells treated with kasugamycin or not. The start codon (AUG), frameshift site (FS) and stop codon (UAG) were marked by grey strips. (C) Box plot analysis of SHAPE reactivity distributions with medians for coding and untranslated regions of Ty1 gRNA from cells treated with kasugamycin-treated (ksg) or not. The *POL ORF** calculation was performed for fragments presented in panel (B). Significance was computed by unpaired two-tailed Mann–Whitney test; * $P < 0.05$, **** $P < 0.0001$, n.s., not significant. (D) Percentages of paired nucleotides in the predicted MFE structures of +1–2500 region of Ty1 gRNAs, including high probability base pairs (HP bps). (E) Sensitivity and PPV parameters calculated for the MFE structure models of +1–2500 Ty1 gRNA from cells with and without kasugamycin treatment or synthesized *in vitro* (colored as in Figure 3). (F) Venn diagram showing overlap of highly probable base pairs between MFE structures of +1–2500 region of Ty1 gRNAs.

cycle. Our results are in agreement with recent studies investigating *in vivo* and *in vitro* structures of viral RNAs, which suggest a more open structure for ZIKV gRNA (2) and a large unfolding of IAV mRNAs in infected cells (6). ZIKV and DENV gRNAs are also less structured in infected cells than *in virio* (3). Consistent with these findings, comparison of *in vivo* SHAPE reactivity data with those obtained previously *in virio* and *ex virio* for the first 1482 nts of Ty1 RNA (33) show that Ty1 gRNA is also less structured in cells than in VLPs (Supplementary Figure S4A). Additionally, the overall similarity between *in virio*, *ex virio* and *in vitro* MFE structures of Ty1 gRNA is quite high despite the fact that *in virio* RNA is probed in a protein-bound state (Supplementary Figure S4B). Whereas the *in vivo* structure differs significantly from those observed in other experimental states (Supplementary Figure S4B).

Not only are viral RNAs less structured in infected cells, transcriptome-wide structure probing analyses indicate a global decrease in the *in vivo* structure of eukaryotic and prokaryotic mRNAs (36,73,82,85). However, the reasons for RNA structural alteration *in vivo* are poorly understood. Whereas some studies show that ribosomes account

for RNA remodeling *in vivo* (73,82), others suggest a more significant role for RNA binding proteins, ATP-dependent helicases and covalent RNA modifications (36,85,86). To date, the mechanisms underlying viral RNA destabilization *in vivo* have not been thoroughly studied. Our observations indicate that ribosomes play an important role in Ty1 gRNA structure remodeling in yeast. First, the coding region of *in vivo* Ty1 gRNA is less structured than the UTRs. Secondly, kasugamycin treatment leads to stabilization of Ty1 gRNA structure. An analogous effect of translation initiation inhibitors is also observed for *E. coli* and zebrafish mRNAs (73,82). In yeast, a marked increase in mRNA structure is detected under ATP-depleted conditions, and the strong contribution of energy-dependent processes other than translation to mRNA remodeling *in vivo* is proposed (36). However translation is one of the major energy-consuming cellular processes, and upon ATP depletion yeast shut global protein synthesis (87,88). Thus, we conclude that translation is one of several energy-dependent processes contributing to mRNA structure destabilization in yeast. In addition, other cellular and viral factors may be involved in Ty1 gRNA structure remodeling *in vivo*.

Ty1 replication requires direct interactions between the RNA genome and Ty1 Gag (19,38), which displays nucleic acid chaperone activity and likely promotes structural rearrangements of genomic RNA and interactions between RNA partners, analogous to retroviral Gag polyproteins (40,89).

Although Ty1 gRNA is generally less structured *in vivo*, it retains well-structured 5' and 3' terminal regions that contain functional *cis*-acting sequences directly involved in dimerization, cyclization and packaging of the Ty1 genome, primer tRNA annealing, and reverse transcription (8,11). However, our results suggest that Ty1 gRNA may contain additional functional *cis*-acting sequences in *POL* where we identify eight well-structured regions. Recent studies show that low-SHAPE and low-entropy analyses can identify previously unknown functional motifs in viral RNA and cellular transcripts (43,74,75,90). Perhaps the stable RNA motifs in *POL* serve as specific binding sites for cellular proteins, as multiple cellular factors influence Ty1 replication (8). Alternatively, these well-structured motifs might contribute to the stability of Ty1 gRNA *in vivo* since *cis*-elements within the coding region help stabilize certain yeast transcripts in addition to the poly(A) tail (91). Although a small fraction of Ty1 transcripts are polyadenylated, Ty1 gRNA half-life is longer than the average half-life of many cellular transcripts (19–21), suggesting the presence of additional stabilizing elements. The correlation between the CDS structure *in vivo* and yeast RNA half-life remains unclear, and codon optimality is proposed as a major determinant of mRNA stability in yeast (92). A significant increase of yeast mRNA half-life is detected when non-optimal codons are converted to optimal codons in a manner minimizing changes in the GC content and the predicted RNA secondary structure (92). Interestingly codon-optimized Ty1 gRNA (CO-Ty1 gRNA) is about as stable as Ty1 gRNA (93). The secondary structure of CO-Ty1 RNA has not been analyzed, but the lack of an increase in stability suggests that the positive effect of codon optimization may be masked by recoding-induced changes in other Ty1 transcript features, such as RNA secondary structure in the CDS. Furthermore, Ty1 gRNA stability is enhanced by interactions with Ty1 Gag (38). Thus, the 1209 base changes in CO-Ty1 gRNA may also induce sequence or structure changes that alter Gag or cellular factor binding. Although extensive sequence recoding in the CO-Ty1 element does not alter the level of retrotransposition or protein expression, there are examples of mutations that strongly inhibit Ty1 mobility located in coding sequences and the UTRs (30,31,76). Indeed, a Rap1 binding site fortuitously introduced into CO-Ty1 during recoding dramatically alters Ty1 gRNA expression (93). Therefore, it stands to reason that RNA structural elements mimicking functions of natural Ty1 gRNA could also have been introduced in CO-Ty1.

Ty1 gRNA also shares structural properties with yeast mRNAs that contribute to stability. There is growing evidence that the low structure (high ΔG) of 5'UTRs and the high structure (low ΔG) of 3'UTRs are associated with longer half-lives and a higher abundance of yeast transcripts (94,95). The 5'UTR of Ty1 RNA has a low GC-content (30%) and is very short (53 nts), consequently, its folding free energy is relatively high. Our data also sup-

port the formation of a long stem-loop structure within the 3'UTR of Ty1 RNA *in vivo* and *in vitro* (Figure 3C, F and Supplementary Figure S3). Previous work indicates that stable stem-loop motifs in the 3'UTR but outside the poly(A) tail increase the stability of yeast mRNAs (96). Additional analyses will be required to address the function of well-structured regions in *POL* and the 3' stem-loop motif in the Ty1 life cycle. Future work on RNA folding in the powerful yeast model will also be enhanced by our discovery that diverse SHAPE reagents can be used successfully in growing yeast cultures.

DATA AVAILABILITY

The SHAPE reactivity values are provided in Supplemental Dataset. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Katarzyna J. Purzycka for help developing the first project idea and Piotr Romanowski for providing computer and informatics support.

Author contributions: K.P.W. supervised the project. A.A. performed SHAPE experiments in yeast, A.A. and M.Z. performed SHAPE experiments in *E. coli*, M.Z. and J.G. performed SHAPE experiments *in vitro*, D.J.G. helped develop the project, provided yeast strains, and guidance on cell culturing conditions, A.A. and M.Z. performed data analysis and, together with K.P.W., J.G. and D.J.G. interpreted the results. K.P.W., A.A., M.Z. and D.J.G. wrote the manuscript.

FUNDING

Polish National Science Centre [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.].

Conflict of interest statement. None declared.

REFERENCES

- Boerneke, M.A., Ehrhardt, J.E. and Weeks, K.M. (2019) Physical and functional analysis of viral RNA genomes by SHAPE. *Annu Rev Virol*, **6**, 93–117.
- Li, P., Wei, Y., Mei, M., Tang, L., Sun, L., Huang, W., Zhou, J., Zou, C., Zhang, S., Qin, C.F. *et al.* (2018) Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe*, **24**, 875–886.
- Huber, R.G., Lim, X.N., Ng, W.C., Sim, A.Y.L., Poh, H.X., Shen, Y., Lim, S.Y., Sundstrom, K.B., Sun, X., Aw, J.G. *et al.* (2019) Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.*, **10**, 1408.
- Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T.K., Marinus, T., Ogando, N.S., Snijder, E.J., van Hemert, M.J., Bujnicki, J.M. *et al.* (2020) Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.*, **48**, 12436–12452.

5. Huston, N.C., Wan, H., de Cesaris Araujo Tavares, R., Wilen, C. and Pyle, A.M. (2020) Comprehensive in-vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. doi:10.1101/2020.07.10.197079.
6. Simon, L.M., Morandi, E., Lukanini, A., Gribaudo, G., Martinez-Sobrido, L., Turner, D.H., Oliviero, S. and Incarnato, D. (2019) In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs. *Nucleic Acids Res.*, **47**, 7003–7017.
7. Krupovic, M., Blomberg, J., Coffin, J.M., Dasgupta, I., Fan, H., Geering, A.D., Gifford, R., Harrach, B., Hull, R., Johnson, W. *et al.* (2018) Ortervirales: new virus order unifying five families of reverse-transcribing viruses. *J. Virol.*, **92**, e00515-18.
8. Curcio, M.J., Lutz, S. and Lesage, P. (2015) The Ty1 LTR-retrotransposon of budding yeast, *Saccharomyces cerevisiae*. *Microbiol. Spectr.*, **3**, MDNA3-0053-2014.
9. Sandmeyer, S., Patterson, K. and Bilanchone, V. (2015) Ty3, a position-specific retrotransposon in budding yeast. *Microbiol. Spectr.*, **3**, MDNA3-0057-2014.
10. Dodonova, S.O., Prinz, S., Bilanchone, V., Sandmeyer, S. and Briggs, J.A.G. (2019) Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 10048–10057.
11. Pachulska-Wieczorek, K., Le Grice, S.F. and Purzycka, K.J. (2016) Determinants of genomic RNA encapsidation in the *Saccharomyces cerevisiae* long terminal repeat retrotransposons Ty1 and Ty3. *Viruses*, **8**, 193.
12. Dubois, N., Marquet, R., Paillart, J.C. and Bernacchi, S. (2018) Retroviral RNA Dimerization: from structure to functions. *Front. Microbiol.*, **9**, 527.
13. Menendez-Arias, L., Sebastian-Martin, A. and Alvarez, M. (2017) Viral reverse transcriptases. *Virus Res.*, **234**, 153–176.
14. Sultana, T., Zamborlini, A., Cristofari, G. and Lesage, P. (2017) Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.*, **18**, 292–308.
15. Carr, M., Bensasson, D. and Bergman, C.M. (2012) Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. *PLoS One*, **7**, e50978.
16. Curcio, M.J. and Garfinkel, D.J. (1994) Heterogeneous functional Ty1 elements are abundant in the *Saccharomyces cerevisiae* genome. *Genetics*, **136**, 1245–1259.
17. Malik, H.S., Henikoff, S. and Eickbush, T.H. (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.*, **10**, 1307–1318.
18. Elder, R.T., Loh, E.Y. and Davis, R.W. (1983) RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 2432–2436.
19. Malagon, F. and Jensen, T.H. (2008) The T body, a new cytoplasmic RNA granule in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **28**, 6022–6032.
20. Voytas, D., Boeke, J., Craig, N., Craigie, R., Gellert, M. and Lambowitz, A. (2002) Mobile DNA II.
21. Nonet, M., Scafe, C., Sexton, J. and Young, R. (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.*, **7**, 1602–1611.
22. Curcio, M.J., Hedge, A.M., Boeke, J.D. and Garfinkel, D.J. (1990) Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **220**, 213–221.
23. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) Ty elements transpose through an RNA intermediate. *Cell*, **40**, 491–500.
24. Devine, S.E. and Boeke, J.D. (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev.*, **10**, 620–633.
25. Friant, S., Heyman, T., Bystrom, A.S., Wilhelm, M. and Wilhelm, F.X. (1998) 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.*, **18**, 799–806.
26. Garfinkel, D.J., Boeke, J.D. and Fink, G.R. (1985) Ty element transposition: reverse transcriptase and virus-like particles. *Cell*, **42**, 507–517.
27. Asif-Laidin, A., Conesa, C., Bonnet, A., Grison, C., Adhya, I., Menouni, R., Fayol, H., Palmic, N., Acker, J. and Lesage, P. (2020) A small targeting domain in Ty1 integrase is sufficient to direct retrotransposon integration upstream of tRNA genes. *EMBO J.*, **39**, e104337.
28. Xu, H. and Boeke, J.D. (1990) Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. *Mol. Cell. Biol.*, **10**, 2695–2702.
29. Cristofari, G., Bampi, C., Wilhelm, M., Wilhelm, F.X. and Darlix, J.L. (2002) A 5'-3' long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *EMBO J.*, **21**, 4368–4379.
30. Gumna, J., Purzycka, K.J., Ahn, H.W., Garfinkel, D.J. and Pachulska-Wieczorek, K. (2019) Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biol.*, **16**, 1749–1763.
31. Gamache, E.R., Doh, J.H., Ritz, J., Laederach, A., Bellaousov, S., Mathews, D.H. and Curcio, M.J. (2017) Structure-function model for kissing loop interactions that initiate dimerization of Ty1 RNA. *Viruses*, **9**, 93.
32. Huang, Q., Purzycka, K.J., Lusvardi, S., Li, D., Legrice, S.F. and Boeke, J.D. (2013) Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA*, **19**, 320–332.
33. Purzycka, K.J., Legiewicz, M., Matsuda, E., Eizenstat, L.D., Lusvardi, S., Saha, A., Le Grice, S.F. and Garfinkel, D.J. (2013) Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res.*, **41**, 463–473.
34. Andrzejewska, A., Zawadzka, M. and Pachulska-Wieczorek, K. (2020) On the way to understanding the interplay between the RNA structure and functions in cells: A genome-wide perspective. *Int. J. Mol. Sci.*, **21**, 6770.
35. Kertesz, M., Wan, Y., Mazar, E., Rinn, J.L., Nutter, R.C., Chang, H.Y. and Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
36. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. and Weissman, J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.
37. Moore, S.P., Liti, G., Stefanisko, K.M., Nyswaner, K.M., Chang, C., Louis, E.J. and Garfinkel, D.J. (2004) Analysis of a Ty1-less variant of *Saccharomyces paradoxus*: the gain and loss of Ty1 elements. *Yeast*, **21**, 649–660.
38. Checkley, M.A., Mitchell, J.A., Eizenstat, L.D., Lockett, S.J. and Garfinkel, D.J. (2013) Ty1 gag enhances the stability and nuclear export of Ty1 mRNA. *Traffic*, **14**, 57–69.
39. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. and Giddings, M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.
40. Pachulska-Wieczorek, K., Blaszczyk, L., Biesiada, M., Adamiak, R.W. and Purzycka, K.J. (2016) The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag. *Retrovirology*, **13**, 18.
41. Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A. and Weeks, K.M. (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.
42. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
43. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A. and Weeks, K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, **11**, 959–965.
44. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
45. Mathews, D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**, 1178–1190.
46. Garfinkel, D.J., Mastrangelo, M.F., Sanders, N.J., Shafer, B.K. and Strathern, J.N. (1988) Transposon tagging using Ty elements in yeast. *Genetics*, **120**, 95–108.
47. Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A. and Voytas, D.F. (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.*, **8**, 464–478.

48. Czaja, W., Bensasson, D., Ahn, H.W., Garfinkel, D.J. and Bergman, C.M. (2020) Evolution of Ty1 copy number control in yeast by horizontal transfer and recombination. *PLoS Genet.*, **16**, e1008632.
49. Morillon, A., Benard, L., Springer, M. and Lesage, P. (2002) Differential effects of chromatin and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons. *Mol. Cell Biol.*, **22**, 2078–2088.
50. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
51. Low, J.T. and Weeks, K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.
52. Busan, S., Weidmann, C.A., Sengupta, A. and Weeks, K.M. (2019) Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry*, **58**, 2655–2664.
53. Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T. and Chang, H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.
54. Guo, J.U. and Bartel, D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353**, aaf5371.
55. Moqtaderi, Z., Geisberg, J.V. and Struhl, K. (2018) Extensive structural differences of closely related 3' mRNA isoforms: links to Pab1 binding and mRNA stability. *Mol. Cell*, **72**, 849–861.
56. Watters, K.E., Abbott, T.R. and Lucks, J.B. (2016) Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.*, **44**, e12.
57. Sztuba-Solinska, J., Rausch, J.W., Smith, R., Miller, J.T., Whitby, D. and Le Grice, S.F.J. (2017) Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA: a structural scaffold for nuclear, cytoplasmic and viral proteins. *Nucleic Acids Res.*, **45**, 6805–6821.
58. Sherpa, C., Rausch, J.W. and Le Grice, S.F. (2018) Structural characterization of maternally expressed gene 3 RNA reveals conserved motifs and potential sites of interaction with polycomb repressive complex 2. *Nucleic Acids Res.*, **46**, 10432–10447.
59. Smola, M.J. and Weeks, K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.*, **13**, 1181–1195.
60. Lee, B., Flynn, R.A., Kadina, A., Guo, J.K., Kool, E.T. and Chang, H.Y. (2017) Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA*, **23**, 169–174.
61. Rice, G.M., Leonard, C.W. and Weeks, K.M. (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA*, **20**, 846–854.
62. Steen, K.A., Rice, G.M. and Weeks, K.M. (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. *J. Am. Chem. Soc.*, **134**, 13160–13163.
63. Blaszczyk, L., Biesiada, M., Saha, A., Garfinkel, D.J. and Purzycka, K.J. (2017) Structure of Ty1 internally initiated RNA influences restriction factor expression. *Viruses*, **9**, 74.
64. Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. and Chang, H.Y. (2011) Understanding the transcriptome through RNA structure. *Nature reviews. Genetics*, **12**, 641–655.
65. Smola, M.J., Calabrese, J.M. and Weeks, K.M. (2015) Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry*, **54**, 6867–6875.
66. Kutchko, K.M. and Laederach, A. (2017) Transcending the prediction paradigm: novel applications of SHAPE to RNA function and evolution. *Wiley Interdiscipl. Rev. RNA*, **8**, e1374.
67. Deigan, K.E., Li, T.W., Mathews, D.H. and Weeks, K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
68. Weeks, K.M. and Mauger, D.M. (2011) Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.*, **44**, 1280–1291.
69. Li, P., Zhou, X., Xu, K. and Zhang, Q.C. (2021) RASP: an atlas of transcriptome-wide RNA secondary structure probing data. *Nucleic Acids Res.*, **49**, D183–D191.
70. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S. and Rouskin, S. (2017) DMS-MaPSeq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*, **14**, 75–82.
71. Giannetti, C.A., Busan, S., Weidmann, C.A. and Weeks, K.M. (2019) SHAPE probing reveals human rRNAs are largely unfolded in solution. *Biochemistry*, **58**, 3377–3385.
72. Smola, M.J., Christy, T.W., Inoue, K., Nicholson, C.O., Friedersdorf, M., Keene, J.D., Lee, D.M., Calabrese, J.M. and Weeks, K.M. (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 10322–10327.
73. Mustoe, A.M., Busan, S., Rice, G.M., Hajdin, C.E., Peterson, B.K., Ruda, V.M., Kubica, N., Nutiu, R., Baryza, J.L. and Weeks, K.M. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell*, **173**, 181–195.
74. Dethoff, E.A., Boerneke, M.A., Gokhale, N.S., Muhire, B.M., Martin, D.P., Sacco, M.T., McFadden, M.J., Weinstein, J.B., Messer, W.B., Horner, S.M. et al. (2018) Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 11513–11518.
75. Mauger, D.M., Golden, M., Yamane, D., Williford, S., Lemon, S.M., Martin, D.P. and Weeks, K.M. (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3692–3697.
76. Bolton, E.C., Coombes, C., Eby, Y., Cardell, M. and Boeke, J.D. (2005) Identification and characterization of critical cis-acting sequences within the yeast Ty1 retrotransposon. *RNA*, **11**, 308–322.
77. Belcourt, M.F. and Farabaugh, P.J. (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell*, **62**, 339–352.
78. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W. Jr, Swanstrom, R., Burch, C.L. and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
79. Collart, M.A. and Weiss, B. (2020) Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Res.*, **48**, 1043–1055.
80. Lange, S.J., Maticzka, D., Mohl, M., Gagnon, J.N., Brown, C.M. and Backofen, R. (2012) Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res.*, **40**, 5215–5226.
81. Kleiman, L., Jones, C.P. and Musier-Forsyth, K. (2010) Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett.*, **584**, 359–365.
82. Beaudoin, J.D., Novoa, E.M., Vejnar, C.E., Yartseva, V., Takacs, C.M., Kellis, M. and Giraldez, A.J. (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.*, **25**, 677–686.
83. Schuwirth, B.S., Day, J.M., Hau, C.W., Janssen, G.R., Dahlberg, A.E., Cate, J.H. and Vila-Sanjurjo, A. (2006) Structural analysis of kasugamycin inhibition of translation. *Nat. Struct. Mol. Biol.*, **13**, 879–886.
84. Riba, A., Di Nanni, N., Mittal, N., Arhne, E., Schmidt, A. and Zavolan, M. (2019) Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 15023–15032.
85. Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T. et al. (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
86. Sun, L., Fazal, F.M., Li, P., Broughton, J.P., Lee, B., Tang, L., Huang, W., Kool, E.T., Chang, H.Y. and Zhang, Q.C. (2019) RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.*, **26**, 322–330.
87. Stouthamer, A.H. (1973) A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek*, **39**, 545–565.
88. Ashe, M.P., De Long, S.K. and Sachs, A.B. (2000) Glucose depletion rapidly inhibits translation initiation in yeast. *Mol. Biol. Cell*, **11**, 833–848.
89. Rein, A. (2010) Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol*, **7**, 700–705.
90. Dethoff, E.A. and Weeks, K.M. (2019) Effects of refolding on large-scale RNA structure. *Biochemistry*, **58**, 3069–3077.
91. Vemula, M., Kandasamy, P., Oh, C.S., Chellappa, R., Gonzalez, C.I. and Martin, C.E. (2003) Maintenance and regulation of mRNA stability of the *Saccharomyces cerevisiae* OLE1 gene requires multiple elements within the transcript that act through translation-independent mechanisms. *J. Biol. Chem.*, **278**, 45269–45279.

92. Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R. *et al.* (2015) Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111–1124.
93. Yarrington, R.M., Richardson, S.M., Lisa Huang, C.R. and Boeke, J.D. (2012) Novel transcript truncating function of Rap1p revealed by synthetic codon-optimized Ty1 retrotransposon. *Genetics*, **190**, 523–535.
94. Ringner, M. and Krogh, M. (2005) Folding free energies of 5'-UTRs impact post-transcriptional regulation on a genomic scale in yeast. *PLoS Comput. Biol.*, **1**, e72.
95. Aw, J.G., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.L., Tapsin, S., Chan, Y.S., Tan, C.P., Sim, A.Y. *et al.* (2016) In vivo mapping of Eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell*, **62**, 603–617.
96. Geisberg, J.V., Moqtaderi, Z., Fan, X., Oszolak, F. and Struhl, K. (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell*, **156**, 812–824.

SUPPLEMENTARY TABLES AND FIGURES

In vivo structure of the Ty1 retrotransposon RNA genome

Table S1 . Reverse transcription primers.

PRIMER	SEQUENCE (5'-3')
Ty1 PR_1	TCAGGTGATGGAGTGCTCAG
Ty1 PR_2	TCGTTTGCCTCTTGGGTATC
Ty1 PR_3	CATCACTCGGATTTCTCCTG
Ty1 PR_4	TTGGTGTGTCCTGGAAGTGA
Ty1 PR_5	TCGAATTGTCTGTGCATTGG
Ty1 PR_6	AGACACTGGCCTGAAACTGG
Ty1 PR_7	GCCACGAGGATGTATTTTGG
Ty1 PR_8	GGAGGTGTGGAATCGGTTGG
Ty1 PR_9	GGTACGTTTGTATGATTAGTCTCA
Ty1 PR_10	GTCTCGTGATACCTTAATTTTCAG
Ty1 PR_11	TGCTGAATATCACCTCTTGC
Ty1 PR_12	CAAATTGTCACTTGACTG
Ty1 PR_13	TGCATTTTCATGTACCTTCTC
Ty1 PR_14	GCTTCCGTAGTTGAAGTACAT
Ty1 PR_15	TCTATTCCAACATACCACCC
Ty1 PR_16	CTCCTCAAGGATTTAGGAATCC
Ty1 PR_17	CATTGTTGATAAAGGCTATA
16S rRNA PR_559	CTTTACGCCAGTAATT

Supplementary Figure Legends

Figure S1. Exemplary preprocessed SHAPE data from capillary electrophoresis. For clarity, (+) reagent channels (blue) are offset from (-) control channels (green). Sequencing ladders are not shown.

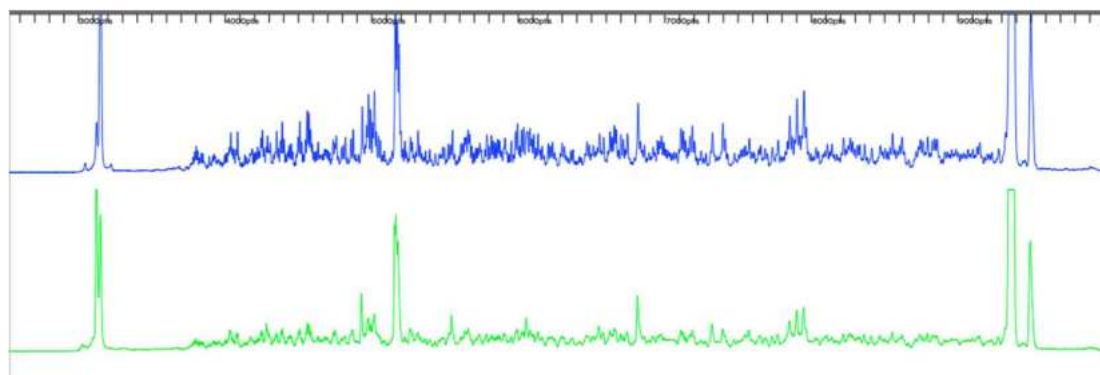
Figure S2. High-resolution secondary structures of lowSS regions identified for *in vivo* (**Panel 1**) and *in vitro* (**Panel 2**) Ty1 gRNAs. Structures are numbered and labelled with nucleotide identity and are colored by SHAPE reactivity (see scale). Pseudoknot (PK) is shown as blue lines. Positions of known functional RNA elements are indicated with colored boxes.

Figure S3. MFE structure models of entire Ty1 RNA genome in the *in vivo* and *in vitro* states. Nucleotides are colored by SHAPE reactivity (see scale) and pseudoknot is shown as blue lines. Positions of known functional RNA elements are indicated with colored boxes.

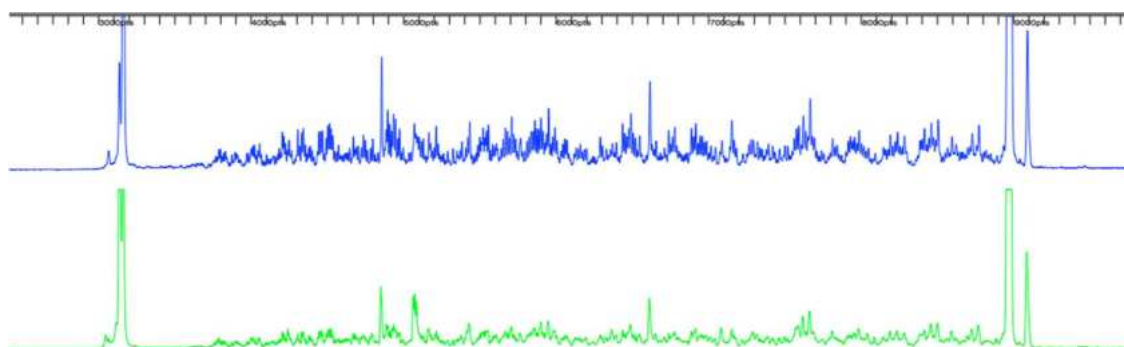
Figure S4. Comparison of Ty1 gRNA structure in four experimental states. Analysis was limited to the region with available *in virio* and *ex virio* SHAPE data. The *ex virio* structure of +1–1482 region and SHAPE reactivity data for *in virio* MFE structure prediction (this study) were taken from 33. **(A)** The median SHAPE reactivity profiles for +1–1482 of Ty1 gRNA smoothed with a 75-nt window (left). Box plot analysis of SHAPE reactivity distributions in +1–1482 region, with medians (right). Median *ex virio* reactivity was taken from 33, box plot is not shown since *ex virio* reactivity values were not provided. **(B)** Sensitivity and PPV parameters calculated for the MFE structure models of +1–1482 of Ty1 gRNA *in vivo*, *in vitro*, *in virio* and *ex virio*. Values are colored from low (red) to high (green).

Figure S1.

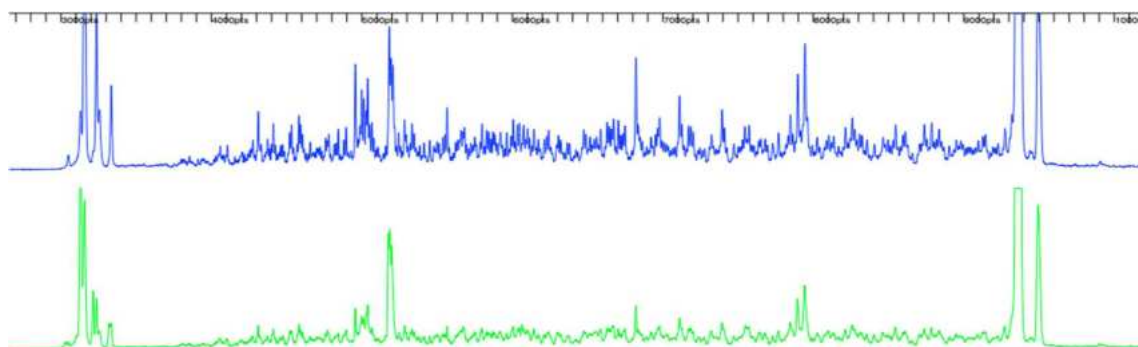
NMIA *in vivo*, Ty1 PR_1



NAI *in vivo*, Ty1 PR_1



1M7 *in vivo*, Ty1 PR_1



NMIA *in vitro*, Ty1 PR_1

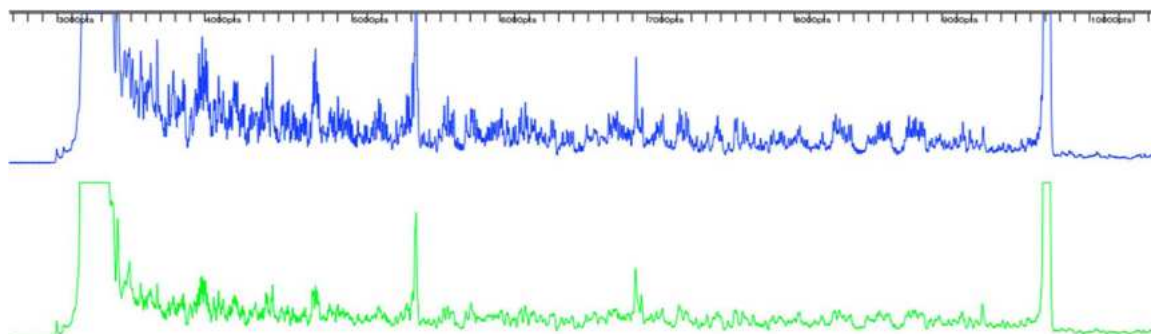


Figure S2. Panel 1

in vivo lowSS regions

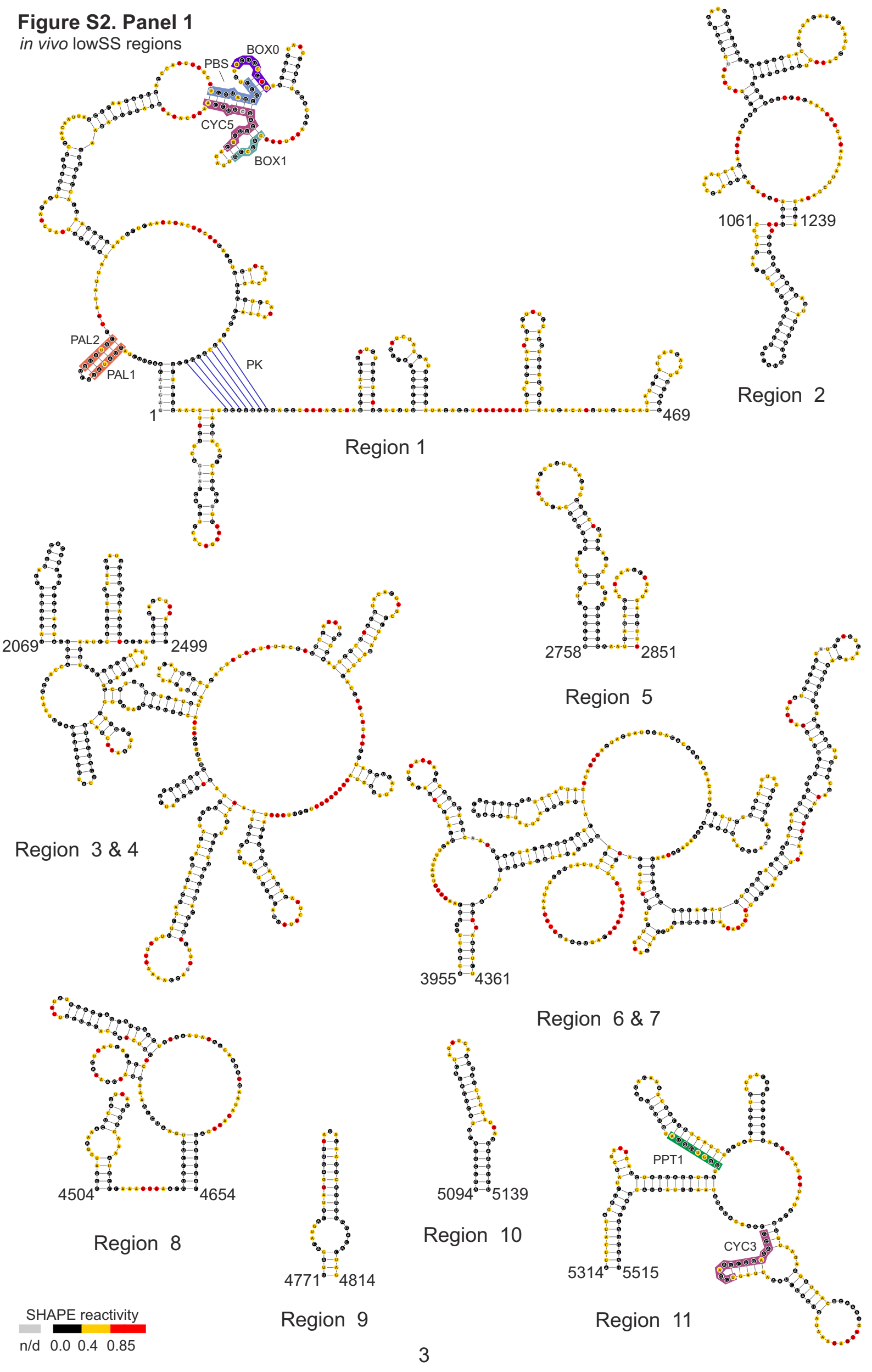


Figure S2. Panel 2
in vitro lowSS regions

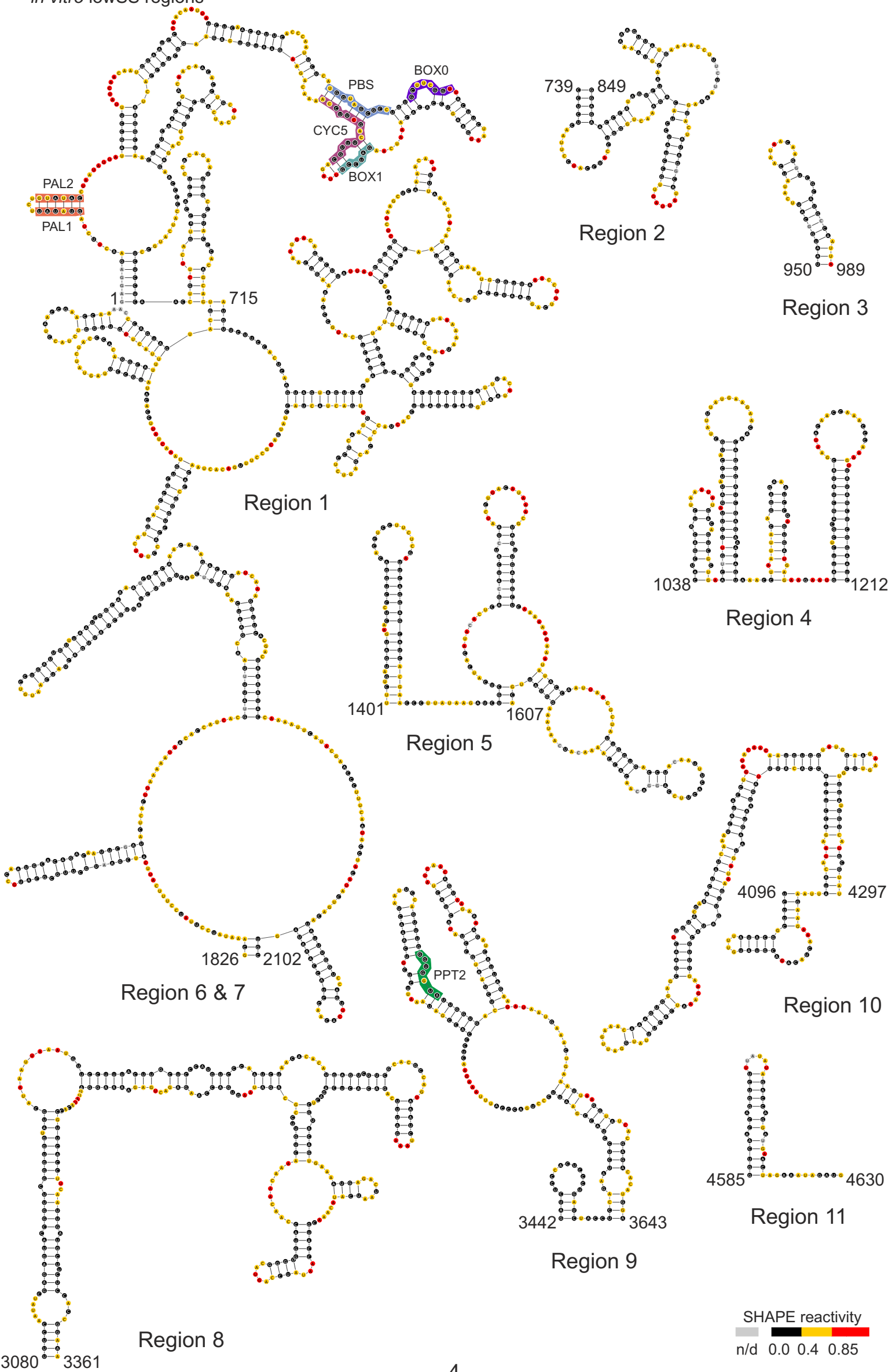
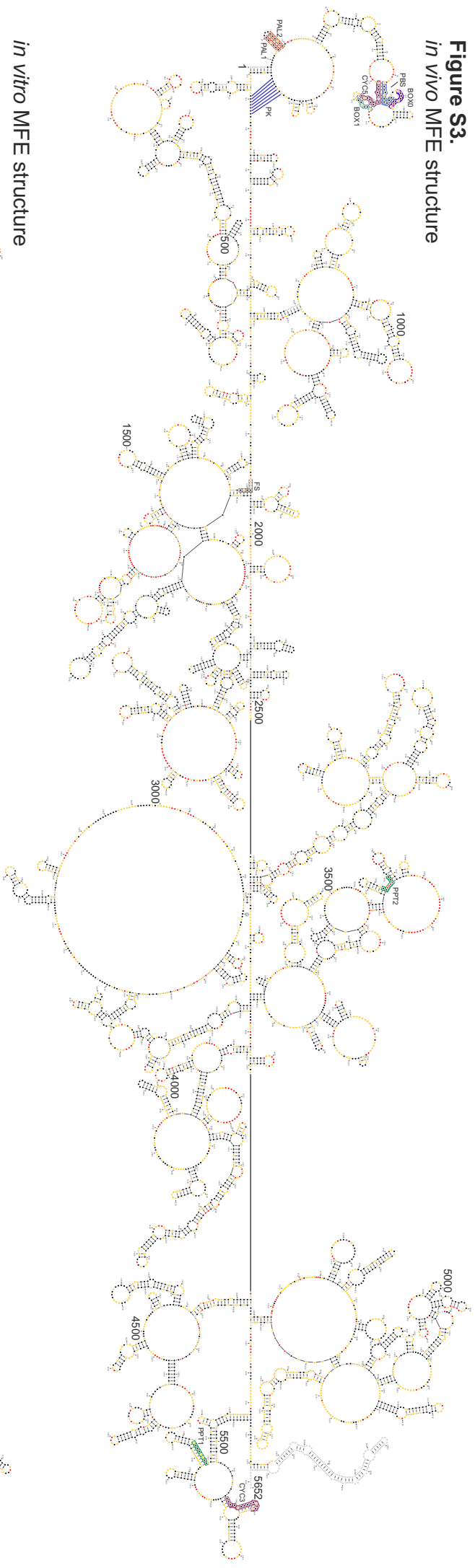
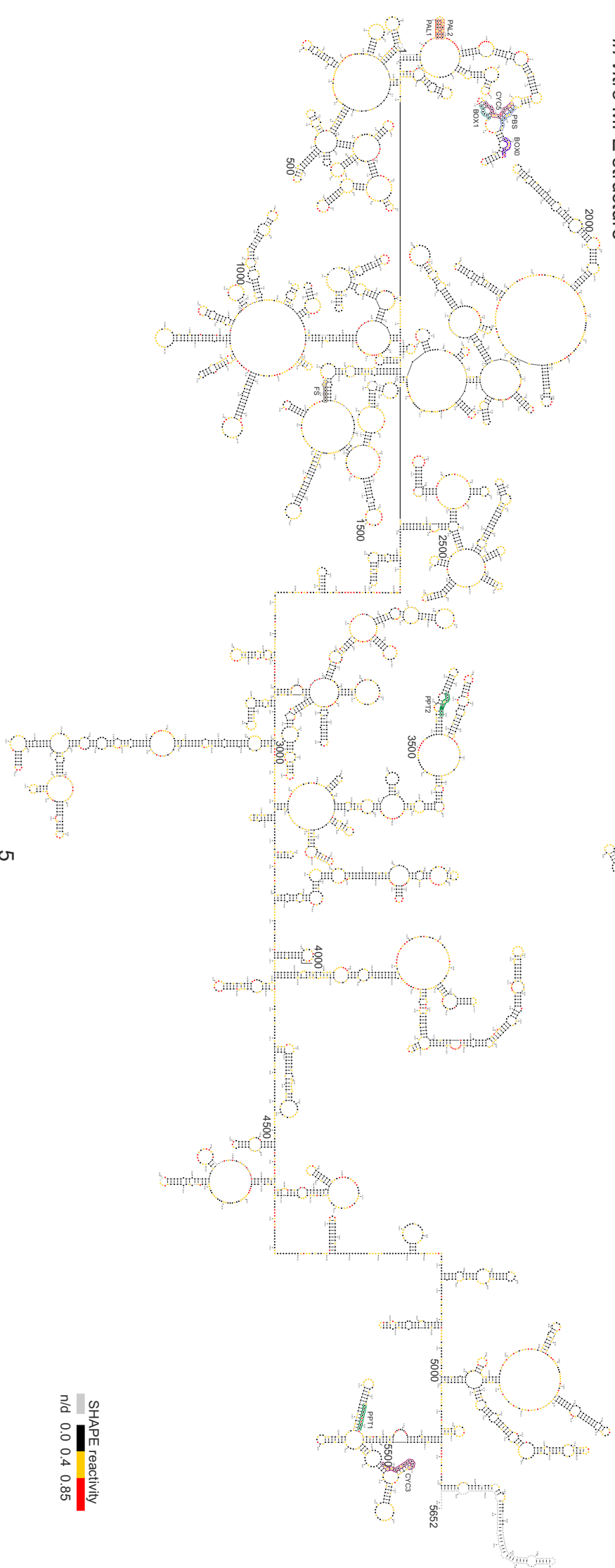


Figure S3.
in vivo MFE structure



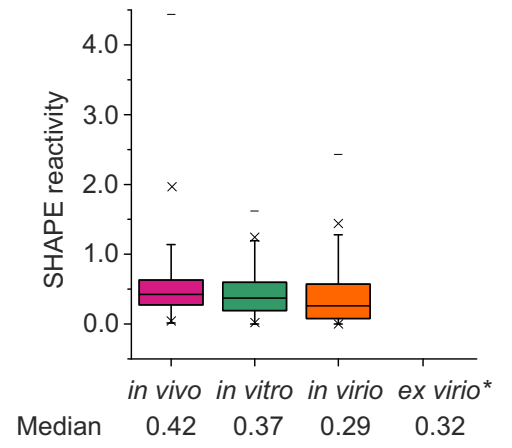
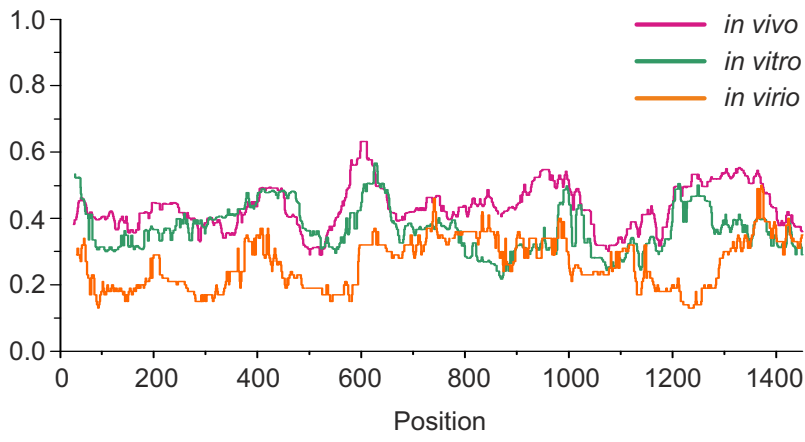
in vitro MFE structure



SHAPE reactivity
 n/d 0.0 0.4 0.85

Figure S4.

A



B

+1-1482 nt Ty1 gRNA MFE structure	sens	PPV
<i>in vivo</i> vs <i>in vitro</i>	43%	51%
<i>in vivo</i> vs <i>ex virio</i>	43%	43%
<i>in vivo</i> vs <i>in virio</i>	43%	48%
<i>in vitro</i> vs <i>ex virio</i>	86%	74%
<i>in vitro</i> vs <i>in virio</i>	87%	84%
<i>in virio</i> vs <i>ex virio</i>	90%	81%

Załącznik nr 2 do rozprawy doktorskiej pt.

„Charakterystyka dynamiki strukturalnej i funkcjonalnej genomów RNA aktywnych retrotranspozonów LTR”

Dodatek *Supplementary Dataset* do pracy naukowej A. Andrzejewska, M. Zawadzka, J. Gumna, D. J. Garfinkel, K. Pachulska-Wieczorek. *In vivo structure of the Ty1 retrotransposon RNA genome*. *Nucleic Acids Research*, 2021, zawierający średnie wartości reaktywności dla poszczególnych nukleotydów z eksperymentów próbkowania strukturalnego RNA oraz wartości odchylenia standardowego, znajduje się pod linkiem: <https://academic.oup.com/nar/article/49/5/2878/6148176>

Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome

Angelika Andrzejewska-Romanowska¹, Julita Gumna¹, Ewa Tykwińska¹ and Katarzyna Pachulska-Wieczorek^{1*}

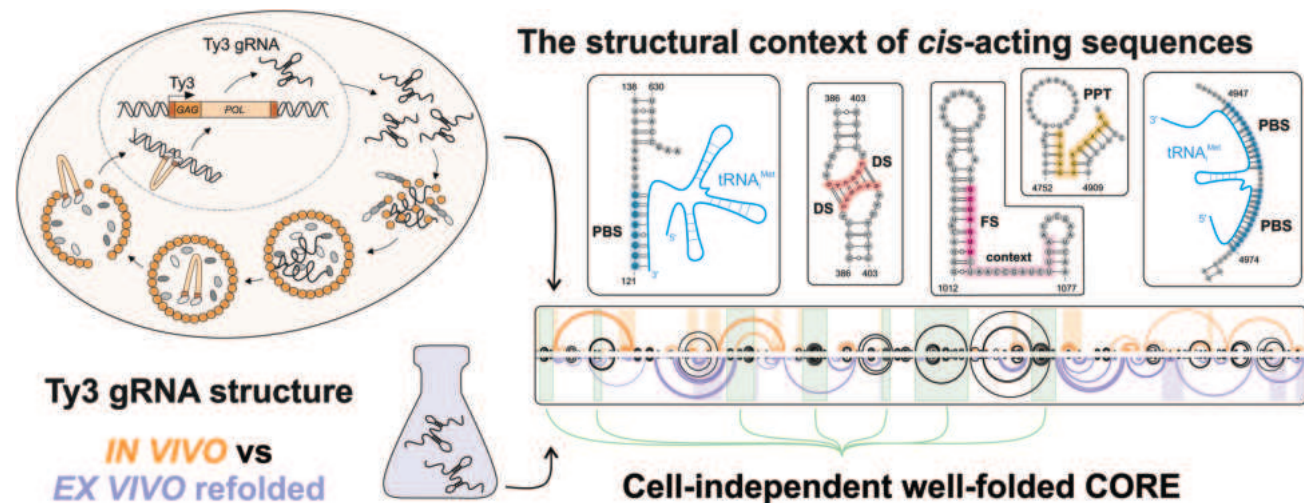
Department of RNA Structure and Function, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

*To whom correspondence should be addressed. Tel: +48 618528503; Fax: +48 618520532; Email: kasiapw@ibch.poznan.pl

Abstract

Long terminal repeat (LTR)-retrotransposons are significant contributors to the evolution and diversity of eukaryotic genomes. Their RNA genomes (gRNA) serve as a template for protein synthesis and reverse transcription to a DNA copy, which can integrate into the host genome. Here, we used the SHAPE-MaP strategy to explore Ty3 retrotransposon gRNA structure in yeast and under cell-free conditions. Our study reveals the structural dynamics of Ty3 gRNA and the well-folded core, formed independently of the cellular environment. Based on the detailed map of Ty3 gRNA structure, we characterized the structural context of *cis*-acting sequences involved in reverse transcription and frameshifting. We also identified a novel functional sequence as a potential initiator for Ty3 gRNA dimerization. Our data indicate that the dimer is maintained by direct interaction between short palindromic sequences at the 5' ends of the two Ty3 gRNAs, resembling the model characteristic for other retroelements like HIV-1 and Ty1. This work points out a range of cell-dependent and -independent Ty3 gRNA structural changes that provide a solid background for studies on RNA structure-function relationships important for retroelement biology.

Graphical abstract



Introduction

LTR-retrotransposons are endogenous reverse-transcribing single-stranded RNA viruses widely distributed in plants, fungi, and animals (1,2). Their RNA genome, akin to infectious RNA viruses, carries all essential instructions for replication and interactions with host cell machinery. In addition to genome function, it also serves as mRNA for element-encoded proteins. LTR-retrotransposons use a 'copy-and-paste' mechanism to propagate and spread within the host genome. New insertions of their copies lead to host genome remodeling and

instability but may also be beneficial by providing regulatory sequences (3,4).

Among mobile genetic elements, the Ty3/*Gypsy* retrotransposons (*Metaviridae* family) are a prominent and evolutionarily important group, considered progenitors of retroviruses (5). Recently, their potential involvement in the process of sex chromosome differentiation was suggested (6). Also, one of the neuronal proteins, Arc, crucial for memory processes and synaptic plasticity, was indicated to be derived from Ty3/*Gypsy*, and like Ty3 Gag protein, can self-assemble into

Received: February 9, 2024. Revised: April 30, 2024. Editorial Decision: May 19, 2024. Accepted: May 29, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

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 reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

virus-like particles (VLPs) that encapsulate Arc mRNA (7,8). The biology of the Ty3/*Gypsy* class has been extensively studied. The Ty3 retrotransposon is their sole representative in the *Saccharomyces cerevisiae*. This element is one of the most characterized and broadly utilized models to understand how LTR-retroelements mobilize, interact with the cell machinery, and shape the host genome.

The *S. cerevisiae* Ty3 gRNA is over 5 kb in length and contains two overlapping open reading frames (ORFs), GAG and POL, encoding structural and catalytic proteins, flanked by 5' and 3'UTRs (193 and 214 nts, respectively) (5) (Figure 1A). The first step of Ty3 replication involves element transcription from the DNA copy integrated with the host genome. After export from the nucleus to the cytoplasm, Ty3 gRNA is utilized as mRNA for translation of Gag and Gag-Pol or is directed to cytoplasmic foci, termed retrosomes (9,10), where a full-length Ty3 gRNA and proteins assemble into VLPs. Finally, mature VLPs contain Gag-derived structural proteins, enzymes encoded by POL, at least two copies of Ty3 gRNA in the dimeric form, and some cellular factors (e.g. tRNA_i^{Met}) (11–13). Ty3 capsid morphology and structure display striking similarities to mature HIV-1 capsids (14). In mature VLPs, gRNA is reverse transcribed to cDNA using self-encoded reverse transcriptase and cellular tRNA_i^{Met} as a primer. The final step of the replication includes integrating Ty3 DNA copy into the transcription start site of genes transcribed by RNA polymerase III (15).

Analogous to retroviruses, the 5' and 3' termini of Ty3 gRNA are rich in regulatory sequences directly involved in replication. The minimal Ty3 element capable of retrotransposition, when Ty3 proteins are provided in *trans*, encompasses the entire 5' UTR and part of both GAG ORF and 3' UTR (miniTy3: +1–426 nt and + 4980–5052 nt) (16). Ty3 contains a bipartite primer binding site (PBS) with the 5' and 3' portions located at opposite ends of the gRNA (17) (Figure 1A). This specific tRNA_i^{Met} binding is proposed to mediate gRNA cyclization and dimerization. Replication of Ty3 requires direct interaction between retrotransposon gRNA and Gag3 (18). Moreover, it was shown *in vitro* that the nucleic acid chaperone activity of Gag3-derived nucleocapsid protein (NCp9) facilitates tRNA_i^{Met} annealing, cyclization and Ty3 gRNA dimerization via tRNA_i^{Met}–tRNA_i^{Met} bridge (17,19). However, the *in vivo* studies do not support PBS role in dimerization and packaging of Ty3 gRNA (13,20). *Cis*-acting sequences essential for Ty3 gRNA localization in retrosomes and packaging to VLPs are not precisely defined, but both UTRs and POL are needed for the proper functioning of these processes (20,21).

Recent studies have highlighted the importance of investigating the relationship between RNA structure and function to better understand RNA-mediated cellular processes and viral RNA acting (22,23). Hence, the knowledge of the native structure of retrotransposon RNA genomes exposed to the cellular environment is necessary to understand comprehensively how genome architecture influences the retrotransposition process. However, there is still very little known about the folding of retrotransposon RNA genomes, and the only gRNA *in vivo* structure model is currently available for yeast Ty1 LTR-retrotransposon (*Pseudoviridae* family) (24,25). Until now, the structure of the Ty3 RNA genome remains unknown.

Here, we report the complete secondary structure model of the Ty3 retrotransposon RNA genome in living cells and

under cell-free conditions. Using the SHAPE-MaP (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension and Mutational Profiling) strategy, we characterized the state-specific structural folding of the retrotransposon gRNA. In Ty3 gRNA, we have identified a well-folded core that forms independently from the cellular environment and encompasses almost 20% of the sequence. We show the structural context of the known Ty3 *cis*-acting primer binding site (PBS), polypurine tract (PPT), and frameshift (FS) sequence. We also point out stable gRNA motifs, which may possess a functional role during retrotransposition. Finally, we identify a novel functional sequence critical for Ty3 gRNA dimer formation *in vitro*—dimerization site (DS). We also characterize gRNA key structural features shared between representative retroelements.

Materials and methods

Yeast strain and growth conditions

S. cerevisiae strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) (Horizon Discovery) were transformed with pDLC201 (Addgene plasmid #113388; formerly pEGTy3-1, 2μ, URA3, galactose inducible Ty3-1) (11). Yeast transformation was performed by the lithium acetate procedure (26,27). *Escherichia coli* strain DH5α (F[−] ϕ80lacZΔM15 Δ(*lacZYA-argF*)U169 *recA1 endA1 hsdR17*(r_K[−], m_K⁺) *phoA supE44 λ[−] thi-1 gyrA96 relA1*) (Invitrogen) was used for DNA cloning and plasmid preparation. *E. coli* cells were grown in LB broth + 0.1 mg/ml Ampicillin (or plates). For non-selective growth, yeast cells were grown on a YPDA medium (BioShop). Plasmid selection used a synthetic yeast growth medium with a drop-out mix (SC-Ura). The presence of plasmid in the transformed yeast strain was confirmed by colony PCR. Used primer sequences are listed in Supplementary Table S1.

For SHAPE-MaP experiments, BY4741 containing pDLC201 plasmid was grown in SC-Ura 2% raffinose broth at 30°C with constant shaking at 200 rpm. Saturated cultures were diluted to OD_{600 nm} of approximately 0.1 with SC-Ura 2% galactose broth to induce Ty3 expression from the GAL promoter. Cultures were grown to a final OD_{600 nm} of 0.7–0.8 at 30°C with constant shaking at 200 rpm.

DNA, RNA and protein substrates

The template for transcription of unmodified yeast tRNA_i^{Met} was generated by PCR, and RNA was synthesized using MEGAscript T7 Transcription Kit (Invitrogen). RNA was purified by denaturing gel electrophoresis (8 M urea) in 1× TBE buffer, eluted from the gel matrix, and concentrated by ethanol precipitation. The DNA templates for *in vitro* transcription of Ty3 5' RNA (+1–429 nt), 3' RNA (+4624–5052 nt), and ΔPAL6 5' RNA (+1–390 nt) were obtained by PCR amplification from plasmid pDLC201. All primers are listed in Supplementary Table S1. Transcripts were synthesized using MAXIscript T7 Transcription Kit (Invitrogen) according to the manufacturer's protocols and purified using a Monarch RNA Cleanup Kit (New England BioLabs). Transcripts integrity was monitored by agarose gel electrophoresis under denaturing conditions. RNA 3'-end labeling with fluorescent dye was carried out overnight at 4°C in an 18 μL reaction containing 20U of T4 RNA Ligase (Thermo Fisher Scientific Inc.), 1× T4 RNA Ligase Buffer, 20 μM ATP, 20 μM pCp-Cy5 or pCp-Cy3 (Jena Bioscience), and 30 pmol of RNA. Labeled

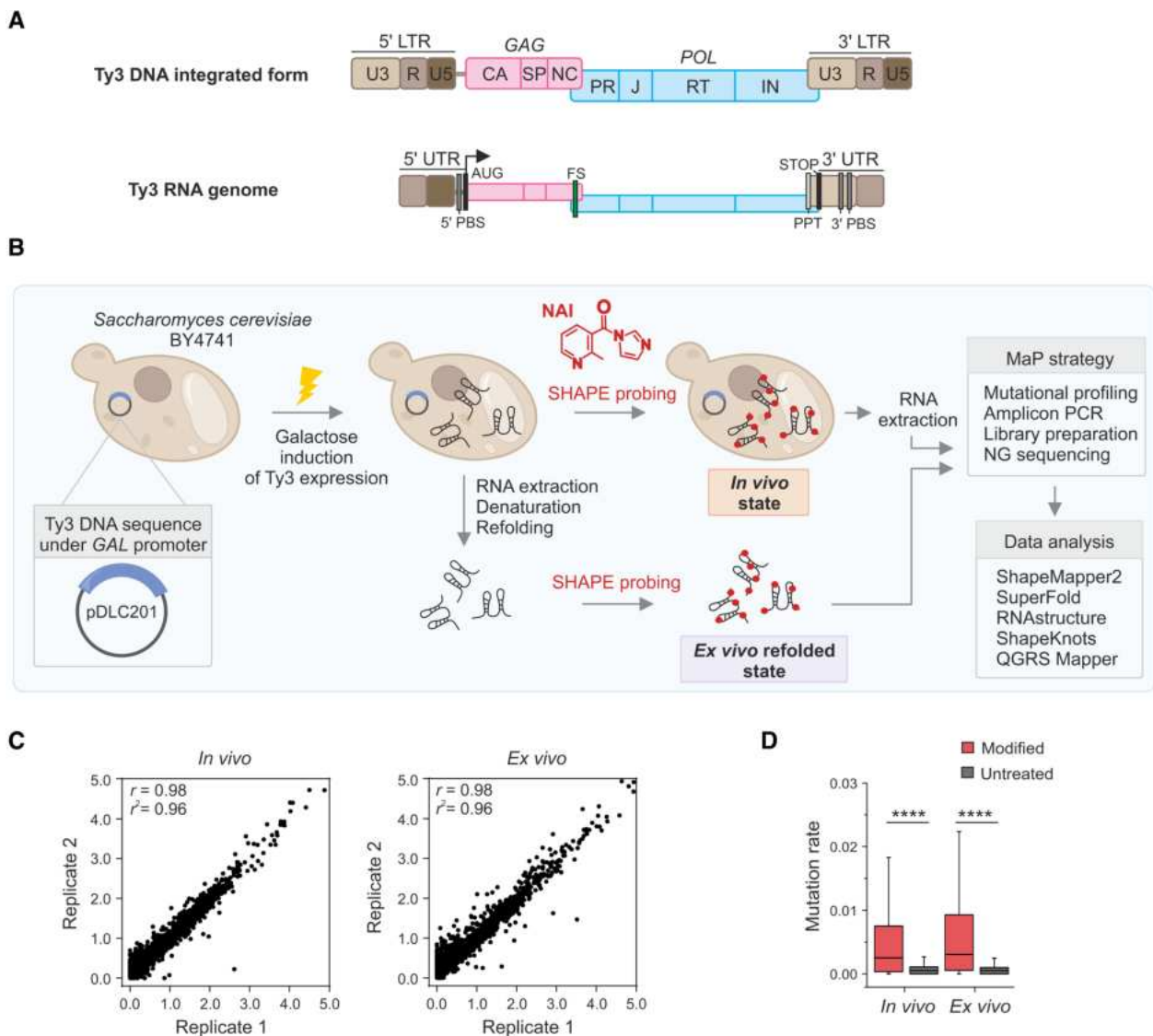


Figure 1. Ty3 DNA and gRNA organization, experimental strategy, and quality assessment of SHAPE-MaP data. **(A)** Organization of the Ty3 DNA and RNA genome. LTR – long terminal repeat, UTR - untranslated region, PBS - primer binding site, PPT - polypurine tract, FS – frameshift, CA - capsid domain, SP - short spacer, NC – nucleocapsid, PR – protease, J - junction domain, RT - reverse transcriptase, and IN - integrase. **(B)** Schematic presentation of the applied strategy. **(C)** Pearson correlation (r) plot of normalized SHAPE reactivities for the *in vivo* and *ex vivo* SHAPE datasets, (r^2) coefficient of determination. **(D)** Box plot analysis of mutation rates for modified and untreated samples from Ty3 gRNA probing under *in vivo* and *ex vivo* conditions. The boxes represent the interquartile range; a line indicates the median. Significance was computed by the Wilcoxon rank-sum test; **** $P < 0.0001$. The analysis for each nucleotide type separately is presented in [Supplementary Figure S2](#).

RNA was purified using a Monarch RNA Cleanup Kit (New England BioLabs). Purified transcripts were stored at -20°C .

The Ty3 NCp9 protein (57 amino acids: TVRTRRSYNKPMNSNHRNRRNNNPSREECIKNRLCFY-CKKEGHLNECRARKASSNRS) was prepared by chemical synthesis and purified by high-performance liquid chromatography (HPLC) (GenScript). NCp9 stocks were reconstituted at 0.8 mg/ml in a buffer containing 20 mM HEPES pH 6.5, 30 mM NaCl, 5 mM dithiothreitol, 0.15 mM ZnCl_2 and 10% glycerol.

Ty3 RNA dimerization and $\text{tRNA}_i^{\text{Met}}$ annealing assays

Cy3-labeled Ty3 RNA (0.5 pmol) was refolded in buffer containing 40 mM Tris-HCl pH 8.0 and 130 mM KCl by heating

at 95°C for 3 minutes, slowly cooled to 60°C , placed on ice for 2 minutes, and then incubated at 37°C for 30 min following the addition of MgCl_2 to 4 mM. For assays in the presence of $\text{tRNA}_i^{\text{Met}}$, Cy5-labeled tRNA was folded separately in equivalent conditions. Ty3 RNA was combined with $\text{tRNA}_i^{\text{Met}}$ at a 1:1 molar ratio before adding NCp9. RNAs were incubated with increasing protein concentrations at 37°C for 30 min. The reactions were stopped by incubation with a quenching solution (1% SDS, 5 mM EDTA) at room temperature for 5 min. The samples were phenol/chloroform extracted, and 15 μl of aqueous phase was mixed with 3 μl of 25% Ficoll 400. RNA was resolved on a 1.3% agarose gel in $0.5\times$ TB at room temperature. The gels were quantified by imaging using Amersham Typhoon 5 Biomolecular Imager with ImageQuantTL v10.1 software (Cytiva). The obtained data were analyzed using GraphPad Prism 8 (GraphPad Software). In all

cases, at least four independent experiments were performed, and the data presented are representative of the whole. The reproducibility of the experiments was assessed by standard deviation.

In vivo SHAPE modification

A 150 ml culture of exponentially growing yeast cells was centrifuged, and the cell pellet was washed once with cold PBS and then resuspended in the desired amount of PBS. Each sample was divided into two equal amounts, and 1 M NAI in DMSO (Merck) was added at a final concentration of 100 mM. A corresponding amount of DMSO was added to a control sample. The modification reactions were conducted at 30°C for 20 min, followed by quenching of NAI by adding DTT at a final concentration of 500 mM. Cells were collected at $8000 \times g$ (10°C) for 5 min, followed by total RNA extraction.

For both experimental states (*in vivo* and *ex vivo* refolded states), different NAI concentrations and modification times were tested, and conditions yielding the efficient NAI-induced mutation rate were chosen for final experiments.

Extraction of total RNA

RNA isolation was performed as previously described (24). Briefly, yeast pellets were resuspended in 1 ml of lysis buffer (10 mM Tris-HCl, pH 8.5, 5 mM EDTA, 2% SDS, 2% 2-mercaptoethanol) and incubated at 83°C for 20 min with constant shaking at 450 rpm. The tubes were centrifuged at $12\,000 \times g$ for 5 min. The supernatants were transferred to fresh tubes, and RNA was extracted twice with phenol (pH 8.0), followed by two extractions with phenol : chloroform (pH 4.5). RNA was recovered by LiCl precipitation overnight at -20°C. RNA pellets were washed twice in 70% ethanol and resuspended in 30 μ l of water. After *in vivo* SHAPE modification, samples were treated with DNase I (Ambion) and purified on an affinity column (RNeasy Mini Kit; QIAGEN). The integrity of the total RNA was evaluated using Qubit™ RNA IQ Assay (Invitrogen), and RNA IQ numbers were >8.5 for all samples. Purified RNA samples were stored at -20°C.

Ex vivo RNA folding and SHAPE modification

For RNA folding, ~40 μ g of total RNA in a volume of 120 μ l was denatured at 95°C for 2 min, then transferred immediately to ice and incubated for 2 min. Then, 60 μ l of 3.3 \times RNA folding buffer (333 mM HEPES, pH 8.0; 333 mM NaCl; 20 mM MgCl₂) was added. RNA was then incubated for 20 min at 30°C to allow secondary structure formation. For *ex vivo* SHAPE modification, each sample of refolded RNA was divided into two equal amounts, and 1 M NAI in DMSO was added at a final concentration of 100 mM. A corresponding amount of DMSO was added to a control sample. The modification reactions were carried at 30°C for 15 min, followed by quenching with DTT. RNA was recovered by ethanol precipitation overnight at -20°C. RNA pellets were washed twice in 70% ethanol and resuspended in 30 μ l of water. Samples were then treated with DNase I (Ambion) and purified on an affinity column (RNeasy Mini Kit; QIAGEN).

In vitro SHAPE modification

Ty3 RNA (3 pmol) in the presence or absence of tRNA^{Met} (3 pmol) was folded as described for dimerization/annealing assays. Subsequently, 600 pmol of NCp9 or an equal volume of protein buffer were added to the RNA mixture. The reaction was incubated at 37°C for 30 min and stopped by incubation with a quenching solution at room temperature for 5 min, followed by phenol/chloroform extraction. Each reaction was divided into two tubes and treated with NAI in DMSO at a final concentration of 100 mM or DMSO alone. The modification reactions were carried out for 15 min at 24°C, due to Ty3 5' RNA dimer instability at 30°C after protein removal (Supplementary Figure S4). The same modification conditions were used for Ty3 3' RNA. SHAPE modification was followed by quenching with DTT. To verify the efficiency of the reaction, one-third of the RNA was resolved on a 1.3% agarose gel in 0.5 \times TB. The rest of the RNA was recovered by ethanol precipitation overnight at -20°C. RNA pellets were washed twice in 70% ethanol and resuspended in 30 μ l of water. Detection of 2'-O-adducts and data processing were performed as described below. The contribution of RNA dimer to the ensemble SHAPE reactivity profile was calculated as previously described (28,29). Briefly, the SHAPE reactivity value of each nucleotide obtained for the mixture of RNA monomer and dimer is a sum of reactivity values of both states, weighted according to the fractional contribution of each to the total RNA population.

Amplicon SHAPE-MaP

To apply the SHAPE-MaP strategy to the Ty3 gRNA under *in vivo* and *ex vivo* states, eight pairs of primers for overlapping amplicons, tiling the entire genome length, were designed. For *in vitro* experiments, two additional shortened amplicons for 5' and 3' ends were also designed. Amplicon-specific reverse transcription was performed as described previously (30). In brief, 1 μ g of total RNA or 1 pmol of *in vitro* transcribed RNA was mixed with 1 μ l of the corresponding 2 μ M reverse primer. Primers were annealed at 65°C for 5 min and then cooled to 4°C, followed by the addition of 8 μ l of 2.5 \times MaP buffer (125 mM Tris, pH 8.0; 187.5 mM KCl; 15 mM MnCl₂; 25 mM DTT and 1.25 mM dNTPs) and incubation at 42°C for 2 min. After the addition of 1 μ l of SuperScript II reverse transcriptase (Invitrogen), the reaction in 20 μ l of total was incubated at 42°C for 3 h, and then RT was heat-inactivated at 70°C for 15 min. The generated cDNA was purified using ZR DNA Sequencing Clean-Up Kit (Zymo Research). 2 μ l 2 M NaOH was added to each reaction and incubated at 95°C for 5 min to degrade the RNA, followed by cooling on ice and the addition of 2 μ l 2 M HCl to neutralize the reaction. cDNA was again purified using the ZR DNA Sequencing Clean-Up Kit. dsDNA amplicons, tiling the Ty3 genome, were generated using NEBNext Ultra II Q5 Master Mix (New England BioLabs), amplicon-specific forward and reverse primers, and 1/10 of purified cDNA. PCR products were visualized on 1.2% agarose gel to confirm the production of correct-sized amplicons, and the desired amplicon products were extracted using Monarch DNA Gel Extraction Kit (New England BioLabs). Accurate concentration of dsDNA amplicons was measured by Qubit dsDNA High Sensitivity assay (Thermo Fisher Scientific Inc.). All primers are listed in Supplementary Table S1.

Illumina sequencing

The dsDNA amplicons were equimolarly pooled to yield the final odd and even sets. Obtained pools were fragmented and proceeded toward downstream library preparation by Genomed (Poland) or Novogene (UK). Sequencing was performed on Illumina MiSeq and NovaSeq instruments, outputting 2×300 or 2×250 paired-end data sets.

MaP analysis

Read quality assessment was performed using the FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). All SHAPE-MaP data were analyzed using the ShapeMapper 2 pipeline (42) and aligned to the sequence of the Ty3 RNA genome (YGRWTy3-1, SGD database). Due to repeated sequence (R) at both ends of the Ty3 genome, odd and even libraries were analyzed separately. The read-depth threshold setting of $5000\times$ was used as a quality control benchmark. All libraries passed the three quality-control checks of ShapeMapper 2.

RNA structure modeling

To Ty3 RNA genome MFE structure prediction, pairing probabilities calculation and lowSS regions identification, the SuperFold software (30) using functions implemented in RNAs-structure (31) was applied. Obtained SHAPE reactivities for replicates were averaged and used as pseudo-energy constraints. A default value of maximum pairing distance of 600 nt was imposed to force the prediction of local base pairs. Slope and intercept folding parameters were set to 1.8 and $-0.6 \text{ kcal mol}^{-1}$, respectively. RNA structures were visualized with StructureEditor, a tool in the RNAs-structure package, and VARNA (31,32).

To *de novo* identify pseudoknots in the *in vivo* and *ex vivo* Ty3 genome, the ShapeKnots software with experimental SHAPE constraints was used (33). Folding was performed in a 600-nt sliding window with 100-nt increments, which minimized biases resulting from sequence shortening and enabled the prediction of multiple pseudoknots in one RNA molecule. Two additional folds in different window sizes (800-nt and 1000-nt) were computed at the 5' and 3' ends to increase the sampling of terminal regions. To *in silico* prediction of potential RNA G-quadruplexes, QGRS Mapper was applied (34).

Sensitivity and positive predictive value (PPV)

The sensitivity and PPV for the obtained MFE models were calculated using the Scorer function (implemented in RNAs-structure). Sensitivity defines the percentage of base pairs in the reference MFE structure that are also present in the compared MFE structure, and the PPV describes the percentage of base pairs in the MFE structure that are present in the reference model. If the region contained only part of the predicted helix, calculations were performed with the manual addition of nucleotides from the disrupted helix.

Gini index

The Gini index was calculated using the R package (R Core Team, 2021). Based on the assessment of SHAPE reactivity profile inequality, this metric defines the level of RNA structure. A relatively even distribution of SHAPE modifications is reflected by a Gini index close to zero, and occurs when RNA is unfolded or RNA structure is highly heterogeneous.

A Gini index close to one occurs when a subset of residues is strongly protected from SHAPE, and indicates a highly stable RNA structure.

Local MFE and HP bps content

Local bps content was computed by calculating the percentage of nucleotides engaged in MFE or HP base pairs within a 75-nt sliding window across the Ty3 genome (in steps of 1 nt). Local bps content values were plotted in a heatmap using the R package (R Core Team, 2021).

Statistical analysis

Statistical analysis of SHAPE-MaP data, including Pearson's correlation, Student's *t*-test, and Wilcoxon rank sum test, was computed using statistical functions in Microsoft Excel and the R package (R Core Team, 2021). The Δ SHAPE framework was applied to detect statistically significant RNA structural differences between states (35). Data visualization was performed using the OriginLab and R package.

Results

In vivo and *ex vivo* mapping of the Ty3 gRNA secondary structure

To explore the secondary structure of the Ty3 retrotransposon gRNA, we utilized a high-throughput SHAPE-MaP strategy (36) (Figure 1B). In general, SHAPE uses small electrophilic reagents to measure the conformation of each nucleotide by forming covalent adducts at the flexible RNA ribose 2'-OH group (37). Thus, high SHAPE reactivities (a high level of SHAPE modification) indicate flexible or unpaired nucleotides, while low SHAPE reactivities point out structurally constrained or base-paired nucleotides. To probe RNA molecules, we used NAI (2-methylnicotinic acid imidazolidine), the SHAPE reagent known to effectively penetrate the cell wall and membranes (38). In the applied strategy, modifications are read out by mutational profiling (MaP). Under appropriate experimental conditions, the modified nucleotides induce the mutations in the cDNA sequence during reverse transcription rather than a stop. Combining this experimental approach with NGS allows the readout of multiple modifications from a single sequencing read, making the method very sensitive and accurate in determining the structure of RNA and enabling the study of the secondary structure, even of rare or less stable RNAs (30).

The reference *S. cerevisiae* strain BY4741 genome contains only two full-length Ty3 elements, one of which, YGRWTy3-1, is transpositionally active (5,11). However, multiple incomplete Ty3 sequences, including solo LTRs, are also present. In nature, transcription and replication of the Ty3 are induced by pheromone stimulation in mating yeast, but retrotransposition is a relatively rare event (39). To obtain structural data for an active and more homogenous pool of Ty3, we used the yeast strain BY4741 with Ty3-1 expressed from the galactose-inducible promoter on a high-copy plasmid (Figure 1B). This system enables synchronous Ty3 expression and retrotransposition regardless of the cell cycle (11). Ty3 gRNA participates in all replication steps, and its major pool is present in the cytoplasm or retosomes, while about 25% is packaged to VLPs (40,41). To determine the impact of cell environment on Ty3 gRNA architecture, we performed RNA structure mapping in living cells (*in vivo* state) or after extraction and refolding of

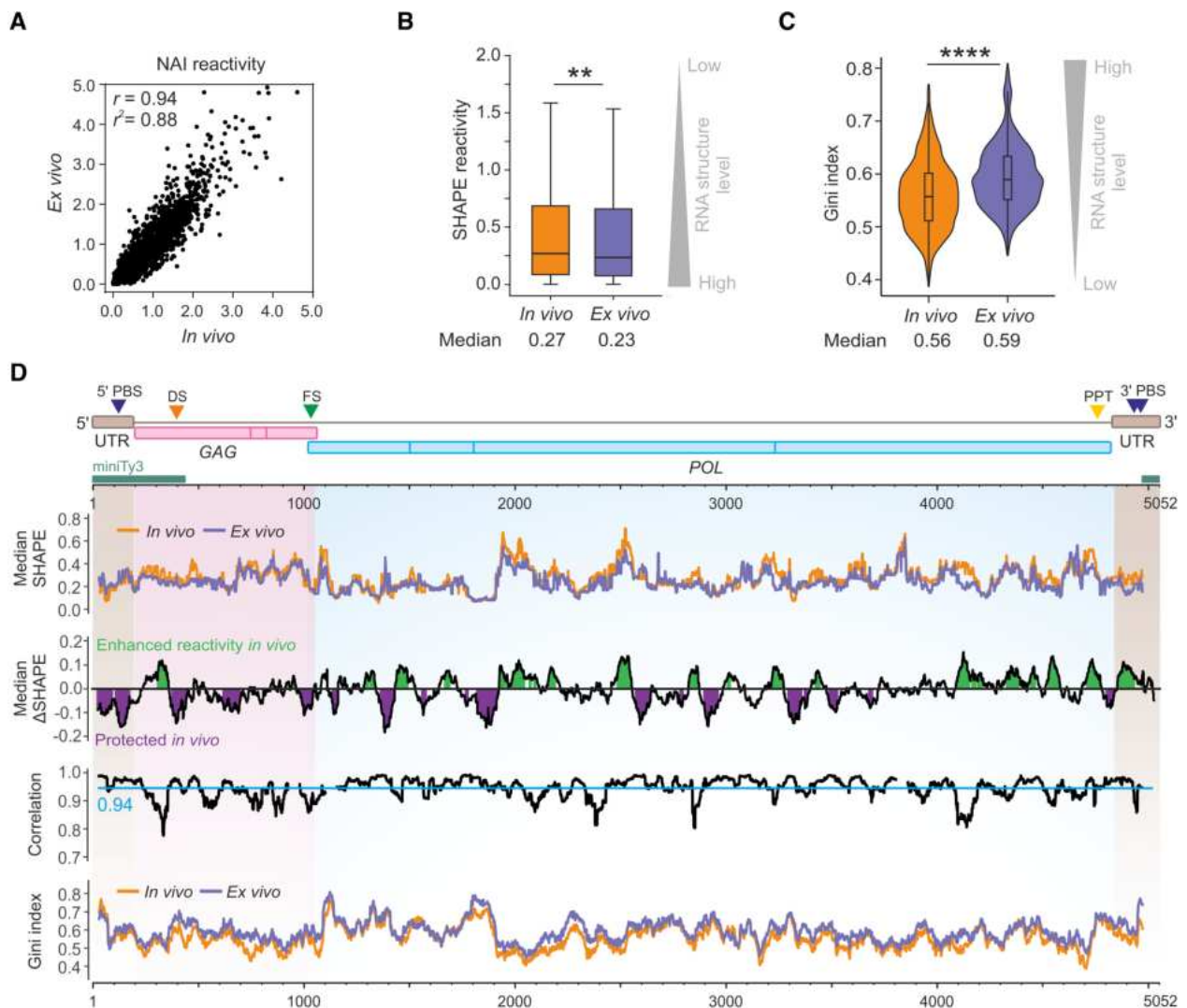


Figure 2. SHAPE-based comparison between *in vivo* and *ex vivo* structure of the Ty3 gRNA. **(A)** Pearson correlation (r) plot of normalized SHAPE reactivities between *in vivo* and *ex vivo* datasets. **(B)** Box plot analysis of SHAPE reactivity distributions with medians. **(C)** Violin plot analysis of Gini index distributions calculated in a 51-nt sliding window with medians. Significance was computed by the Wilcoxon rank-sum test; **** $P < 0.0001$; ** $P < 0.01$. **(D)** Profiles of the median SHAPE (upper plot), Δ SHAPE (*in vivo* – *ex vivo*, upper middle plot), Pearson correlation (lower middle plot), and Gini indexes distributions (lower plot), smoothed with a 51-nt sliding window. A more detailed presentation of Δ SHAPE analysis was presented in [Supplementary Figure S3](#). The *cis*-acting sequences are indicated. The dimerization site (DS) was identified in this work. Localization of miniTy3 sequences was marked as dark green boxes. The nucleotide numbering in the text and all figures refers to the Ty3 gRNA sequence.

total yeast RNA (refolded *ex vivo* state). Ty3 gRNA isolated from cells is devoid of all protein factors that can impact RNA folding but bears the same post-transcriptional RNA modifications as probed under *in vivo* conditions.

To analyze the complete retrotransposon gRNA, which spans 5052 nucleotide residues, we created eight overlapping amplicons, each covering approximately 730 nucleotides of the Ty3 gRNA sequence. During the optimization step, various NAI concentrations and SHAPE reaction times were tested ([Supplementary Figure S1](#)). We produced several biological and technical replicates, resulting in highly reproducible data. Finally, two independent biological replicates of SHAPE-MaP experiments in yeast and under *ex vivo* conditions were conducted using NAI at a final concentration of 100 mM for 20 and 15 min, respectively. All data were analyzed using the ShapeMapper2 pipeline (42). The resulting mutation profiles indicated a strong correlation between the replicates (*in vivo* replicates $r = 0.98$; *ex vivo* replicates

$r = 0.98$; [Figure 1C](#)). Importantly, we observed significantly elevated mutation rates in NAI-treated samples, compared to untreated samples, proving a high efficiency of SHAPE modification in both experimental states ($P < 0.0001$; [Figures 1D](#) and [Supplementary Figure S2](#)). Finally, the SHAPE-MaP data with a median effective depth $> 50\,000$ reads per site in each replicate provided effective reactivity information for 99.1% of nucleotides in the Ty3 gRNA. That guaranteed the correctness of our strategy, and we used these datasets for further analyses in this work.

Model-free analysis of the structural propensity in the Ty3 RNA genome

First, we conducted the model-free analysis of the SHAPE-MaP data obtained for Ty3 gRNA. The *in vivo* and *ex vivo* SHAPE reactivity datasets exhibited a very strong global correlation ($r = 0.94$, [Figure 2A](#)). Concurrently, the global me-

dian SHAPE reactivity was higher in the cell (Figure 2B), suggesting a slightly more unfolded RNA structure. As single-value results may not accurately reflect local characteristics of long RNA, we used a sliding window algorithm to capture unique structural features of Ty3 gRNA states. Based on the analysis of median SHAPE and Δ SHAPE (*in vivo* – *ex vivo*) reactivity profiles, we identified statistically important differences in many gRNA regions (Figure 2D and Supplementary Figure S3) (35). These differences were evident independently of the applied window size (Supplementary Figure S3). Also, the correlation profile dropped below the global value in many regions, indicating sites of decreased structural similarity (Figure 2D). Furthermore, the distribution of Gini indexes—a metric used to measure inequality—had higher *ex vivo* values than the *in vivo* state (median Gini - *in vivo*: 0.56; *ex vivo*: 0.59; Figure 2C and D). This additionally implied a lower level of the Ty3 gRNA structural organization *in vivo*. Together, model-free analyses have revealed both environment-specific features and regions of exceptionally high similarity within Ty3 gRNA.

SHAPE-directed structure prediction revealed a well-folded core of Ty3 gRNA

Next, using obtained SHAPE-MaP data as constraints, we modeled the secondary structure of the Ty3 retrotransposon gRNA with the SuperFold pipeline (Figure 3) (30). The consensus minimum-free energy (MFE) secondary structure model of Ty3 gRNA predicted with *in vivo* data contained slightly fewer base-paired nucleotides, lower GC bp content, and accordingly, exhibited a weaker folding strength, than the *ex vivo* model (Table 1 and Figure 3A, B, and D). The positive predictive value (PPV) and the sensitivity—metrics of RNA structure similarity—were 68.6% and 59.6%, respectively (Figure 3A and B, black arcs). Consequently, the predicted MFE structure models differed more than the global SHAPE reactivity correlation suggested. These comparisons indicate that minor and local SHAPE reactivity changes, which did not impact correlation value, can significantly affect RNA structure prediction.

To account for the structural dynamics of the Ty3 gRNA in both states, we conducted a thorough analysis of the probability of forming each base pair. This approach allowed us to characterize the structural heterogeneity across the RNA molecule and differentiate between RNA regions with non-mutually exclusive base pairs that fold into one stable conformation versus highly dynamic regions. Our calculations revealed that in both states, a comparable number of nucleotides (~24%) were involved in highly probable base pairs (HP bps, defined as having over 80% probability) (Figure 3A, B and D). Over 76% of the predicted HP bps were common for the *in vivo* and *ex vivo* Ty3 gRNA states. These results show that there is a significant group of HP bps that are formed regardless of the RNA folding environment. Consistent with the model-free analysis, in the predicted MFE structure, the 5' region of Ty3 gRNA is much more stabilized *in vivo* than *ex vivo* (higher content of HP bps, low level of alternative pairings) (Figure 3A).

To search for more stable and structured regions, we identified RNA positions with both low SHAPE reactivity and low Shannon entropy (lowSS regions) (30). For Ty3 gRNA, we identified 14 and 12 well-folded regions under *in vivo* and *ex vivo* states, respectively (Figure 3A and B, light orange and vi-

olet shadings). The 1st and 2nd lowSS regions were contained within the miniTy3 sequence (16). We found that most of the functional *cis*-acting sequences of Ty3, including the 5' PBS (17), FS (43) and DS (this work) were localized within lowSS regions identified *in vivo*. Furthermore, we dissected 9 lowSS regions that overlapped between states. Among them, seven shared the same secondary structure motifs enriched in the HP bps (Figure 3C). Therefore, we propose that these lowSS regions form a state-independent well-folded gRNA core. The core covered about 19.4% of the Ty3 gRNA sequence and included the most common HP base pairs (Figure 3D). Much of the core sequence accumulated within or in the vicinity of the region coding reverse transcriptase, which is known to be the most conserved domain among retroelements (44). Even outside of the identified well-folded core, the RT-coding region exhibits a very high level of structural similarity across the states (PPV – 90.5%, sensitivity – 82.8%).

In vitro analysis of RNA–RNA interactions essential for Ty3 replication

Efficient retroelement replication relies on intermolecular RNA–RNA interactions, such as forming a gRNA dimer or a gRNA-tRNA complex. To better comprehend these processes for Ty3, we utilized a simplified *in vitro* system analogous to the one employed for characterizing these interactions in Ty1 (28,45). We prepared two short transcripts covering the 5' or 3' part of the mini Ty3 element—5' RNA: +1–429 nt and 3' RNA: +4624–5052 nt. We then incubated fluorescently labeled transcripts and tRNA_i^{Met} (Cy3 and Cy5, respectively) with increasing amounts of NCp9 protein, and RNA complex formation was assayed by gel electrophoresis (Figure 4A). Consistent with prior study (17), we observed efficient tRNA_i^{Met} annealing to the 3' RNA (~80% of bound tRNA) and to a significantly lesser extent to 5' RNA (~3% of bound tRNA) (Figure 4B, Supplementary Figure S5). It may be caused by the shorter sequence complementary to tRNA_i^{Met} than in the 3' RNA (8 versus 24 nt) and its lower GC content (Table 1). It results in a stronger folding energy of 3' PBS-tRNA_i^{Met} (summed parts 1 and 2) than 5' PBS-tRNA_i^{Met}. Moreover, the folding energy of 3' PBS-tRNA_i^{Met} is comparable to tRNA_i^{Met}. We also detected specific tRNA_i^{Met} dimers formed via the interaction of palindromic sequences at their 5' ends. The tRNA_i^{Met} dimer co-linked to the 5' and 3' PBS sequences in two gRNA strands is thought to maintain the Ty3 gRNA dimeric state (17). Indeed, our *in vitro* assays showed the 3' RNA dimers and the higher order multimers formed only when tRNA_i^{Met} was present in the mixture (Figure 4B). Interestingly, in contrast to others, we found a more significant conversion toward the dimeric state for the 5' RNA (up to 40%), independent of tRNA_i^{Met} addition. This observation suggested that the Ty3 gRNA can dimerize via direct base-pairing of sequences located within its 5' end, similar to Ty1 and retroviruses (28,46).

The SHAPE-MaP data for 5' and 3' transcripts *in vitro* showed a strong correlation with those obtained for the corresponding regions in the Ty3 gRNA *in vivo* (Supplementary Figure S6). The correlations suggested their similar folding to the full-length gRNA, with only a few differences resulting from the sequence shortening. Further comparative analysis of SHAPE-MaP reactivity profiles for 3' RNA mapped in the absence or presence of tRNA_i^{Met} revealed local changes strictly in the 3' PBS region, proving the exact site of tRNA_i^{Met}

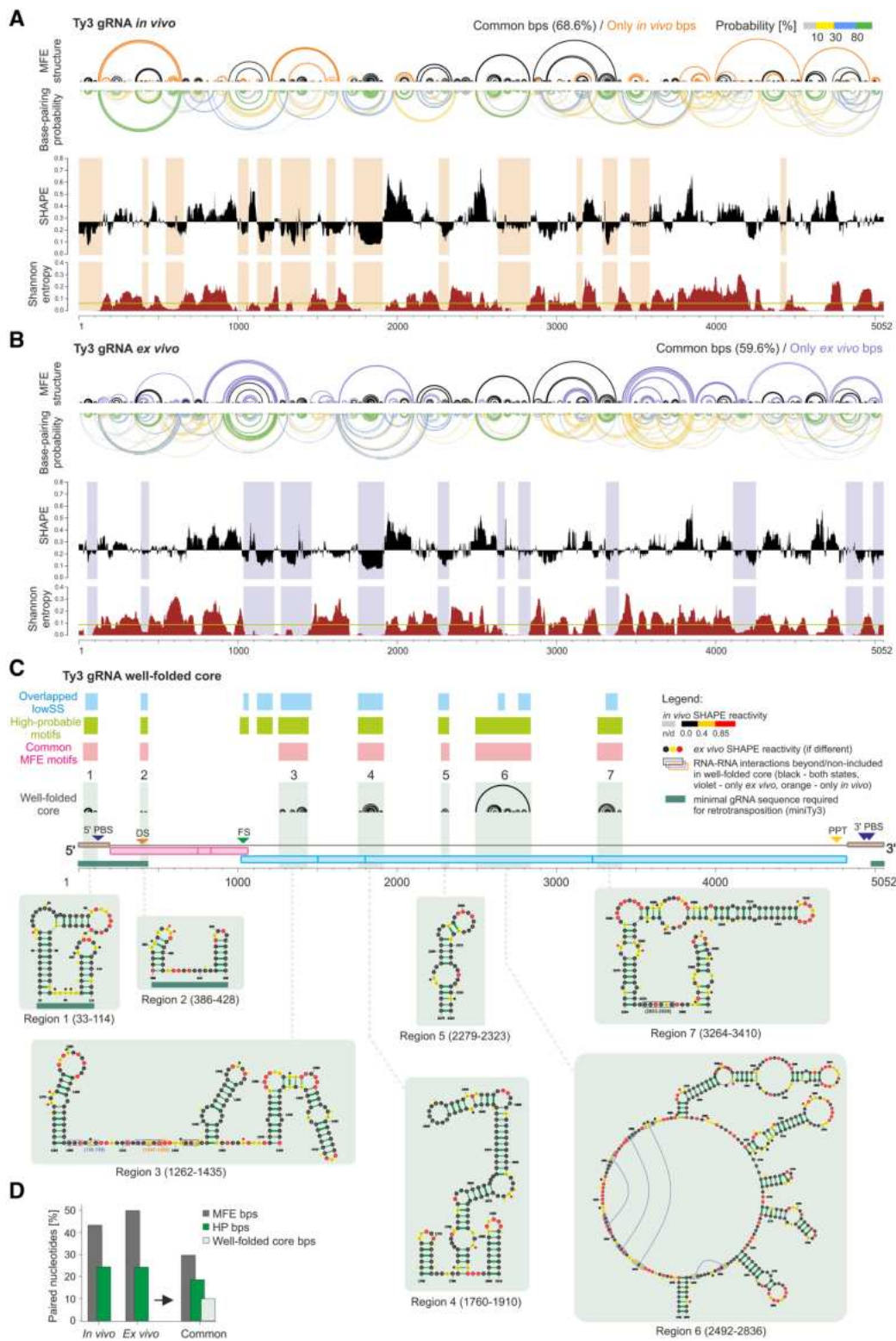


Figure 3. SHAPE – based prediction of the Ty3 gRNA structure and well-folded core. **(A)** Structure map of the Ty3 gRNA *in vivo* depicting: arc plots of predicted MFE structure and base-pairing probabilities (see scale); median SHAPE reactivity profile with respect to the global median reactivity and Shannon entropy profile (both in 51-nt sliding window). Shadings mark lowSS regions. **(B)** shows the analogous analysis performed for *ex vivo* Ty3 gRNA. **(C)** Well-folded core identification. Black arcs represent the MFE structure of the core regions. Detailed secondary structure models are presented below. Nucleotides are colored by SHAPE reactivity (see scale), HP bps are marked in green. **(D)** Percentages of paired nucleotides in the predicted MFE structures, including HP bps and well-folded core.

Table 1. ΔG energy and base pair content calculations for predicted RNA MFE models and RNA–RNA interactions

	ΔG energy (kcal/mol)	Total bp	GC bp
RNA			
Ty3 gRNA <i>in vivo</i>	−769.2	1095	481
Ty3 gRNA <i>ex vivo</i>	−844.3	1260	546
tRNA _i ^{Met}	−33.9	21	18
RNA–RNA interactions			
Ty3 5′ PBS-tRNA _i ^{Met}	−12.2	8	5
Ty3 3′ PBS-tRNA _i ^{Met} (part 1)	−14.8	12	6
Ty3 3′ PBS-tRNA _i ^{Met} (part 2)	−20.5	11	8

binding (Figure 4C). The highest SHAPE reactivity changes were observed for the second part of 3′ PBS, which agrees with its higher GC bp content and stronger folding energy (Table 1). However, we could not characterize the structural properties of the 5′ RNA bound by tRNA_i^{Met} due to the low efficacy of this complex formation *in vitro*. In a tRNA-free state, 5′ PBS was unreactive, most probably due to engagement in intramolecular base-pairing with nt + 172–191 (Supplementary Figure S7). Similar to 5′ PBS, this sequence was also protected toward the SHAPE reagent independently of tRNA_i^{Met} presence or absence in the mixture.

Identification of a candidate Ty3 gRNA dimerization site

To identify the dimerization site, we initially searched for the palindromic sequences in the 5′ end of Ty3 gRNA. This was based on the assumption that Ty3 could follow a retroviral dimerization paradigm, where two gRNA strands interact via short palindromic sequences from their 5′ ends (46). We identified six palindromic sequences, three in 5′ UTR and three in GAG ORF (Figure 5A). To determine palindromes that may be functional, we performed a structural analysis of the NCp9-induced Ty3 5′ RNA monomer and dimer using SHAPE-MaP. Since the maximal efficiency of dimerization was ~40%, we applied a mathematical data deconvolution that allows one to determine the secondary structures of individual RNAs in a mixture of the conformers (28,29). We detected a very strong correlation between the SHAPE reactivity datasets obtained for monomer and dimer *in vitro* ($r = 0.98$). However, the in-depth analysis of SHAPE profiles revealed statistically significant reactivity decreases for the palindrome 6 in the dimeric state (Figures 5A and Supplementary Figure S8). Concurrently, we did not find statistically relevant reactivity drops in other palindromic sequences that could support their involvement in the Ty3 5′ RNA dimerization.

These observations prompted us to further verify the involvement of selected palindromes in Ty3 gRNA dimerization. We found that deletion of the palindrome 6 (PAL6) results in complete inhibition of dimerization *in vitro*, even though the Δ PAL6 5′ RNA still contained the other five palindromes (Figure 5B). SHAPE-MaP analysis revealed that the 5′ RNA shortening has no significant impact on its folding ($r = 0.98$, Figure 5C), indicating that dimerization inhibition is not due to the overall 5′ RNA structure rearrangement. The PAL6 (391CUGCAG396) is located within the Ty3 gRNA lowSS region, further supporting its potential functional importance (Figure 3A, B). Taken together, the GC-rich PAL6 is required to form the Ty3 5′ RNA dimer, and we propose that it might serve as the primary dimerization site (DS) *in vitro*.

Structural analysis of Ty3 gRNA *cis*-acting elements *in vivo*

As a result of the parallel occurrence of Ty3 at different stages of the replication in yeast, the *in vivo* SHAPE data represented the average score from the nucleus, cytoplasm, retrosomes, and VLPs. Moreover, it is unclear how much of the *in vivo* Ty3 gRNA pool is in dimerized or tRNA_i^{Met}-bound form. In total yeast RNA probed *ex vivo*, tRNA_i^{Met} is still present and can bind to Ty3 3′ RNA spontaneously even without the support of proteins (Figure 4B). In Ty3 gRNA, the 5′ PBS exhibited a low level of SHAPE reactivity *in vivo* and *ex vivo*, resembling the SHAPE pattern obtained for *in vitro* probed Ty3 5′ RNA (Figures 6A and Supplementary Figure S7). However, it was challenging to dissect whether the 5′ PBS is occupied by tRNA_i^{Met} or intramolecularly constrained in Ty3 gRNA. SHAPE-based structure predictions suggested highly probable base-pairing of PBS with nt + 641–647 (Figure 6B, C). Indeed, nt + 641–647 were mainly unreactive *in vivo* and *ex vivo* (Figure 6A). This interaction cannot occur in the 5′ RNA (+1–429), where an alternative 5′ PBS base-pairing with nt + 172–191 is created (Supplementary Figure S6). Nevertheless, these data showed that apart from tRNA_i^{Met} binding, the 5′ PBS has a high tendency for intramolecular base-pairing. In contrast, the 3′ PBS segments of Ty3 gRNA were mainly reactive *in vivo* and *ex vivo* (Figure 6A). It suggests their weaker propensity to internal base-pairing but does not exclude tRNA_i^{Met} binding to the 3′ end of gRNA. It is highly probable that the unbound Ty3 gRNA fraction dominates over that with tRNA_i^{Met} annealed and camouflages the effect of intermolecular tRNA_i^{Met}-Ty3 gRNA interaction in SHAPE reactivity profiles. We found evidence to support our hypothesis through *in vitro* experiments on Ty3 3′ RNA (Figure 4C). As a result of very efficient tRNA_i^{Met} binding, we observed a strong SHAPE reactivity drop in the 3′ PBS but not complete protection. This is because some unbound Ty3 3′ RNA was still present in the mixture.

In contrast to the tRNA_i^{Met}-Ty3 gRNA complex, the reconstitution of the Ty3 gRNA dimers in the protein-free *ex vivo* conditions is highly unlikely. These dimers are heat-sensitive (13) and thus disrupted during the applied RNA isolation procedure. Moreover, Ty3 protein is needed for efficient Ty3 RNA dimerization (Figure 4B). For Ty3 gRNA *in vivo*, we detected statistically significant drops in SHAPE reactivity for the newly identified DS/PAL6 sequence, compared to the *ex vivo* state (Figures 6A and Supplementary Figure S9). This effect was not observed for other palindromic sequences. This result additionally supports DS involvement in Ty3 gRNA dimerization. The DS is located *in vivo* in the apical part of the highly probable short hairpin (Figure 6B, C), which suggests that the kissing-loop dimer is first formed, similar to the dimerization model proposed for retroviruses (46,47).

Like in other retroelements, Ty3 POL ORF undergoes translation as a result of +1 frameshifting (FS). The Ty3 FS event is mediated by a 7-nt sequence (1046GCGAGUU1052) and stimulated by the low availability of the tRNA specific for the hungry serine codon (AGU) (43). Additionally, a 12-nt stimulatory ‘context’ sequence nearby prolongs translational pause and increases the frameshifting about 7.5-fold (48). The exact mechanism underlying the stimulation in Ty3 frameshifting remains unknown, but it was postulated that the primary sequence of the ‘context’, not its secondary structure, stimulates frameshifting. We found the ‘context’ sequence highly reactive and mainly unpaired in both gRNA states. Inversely, the

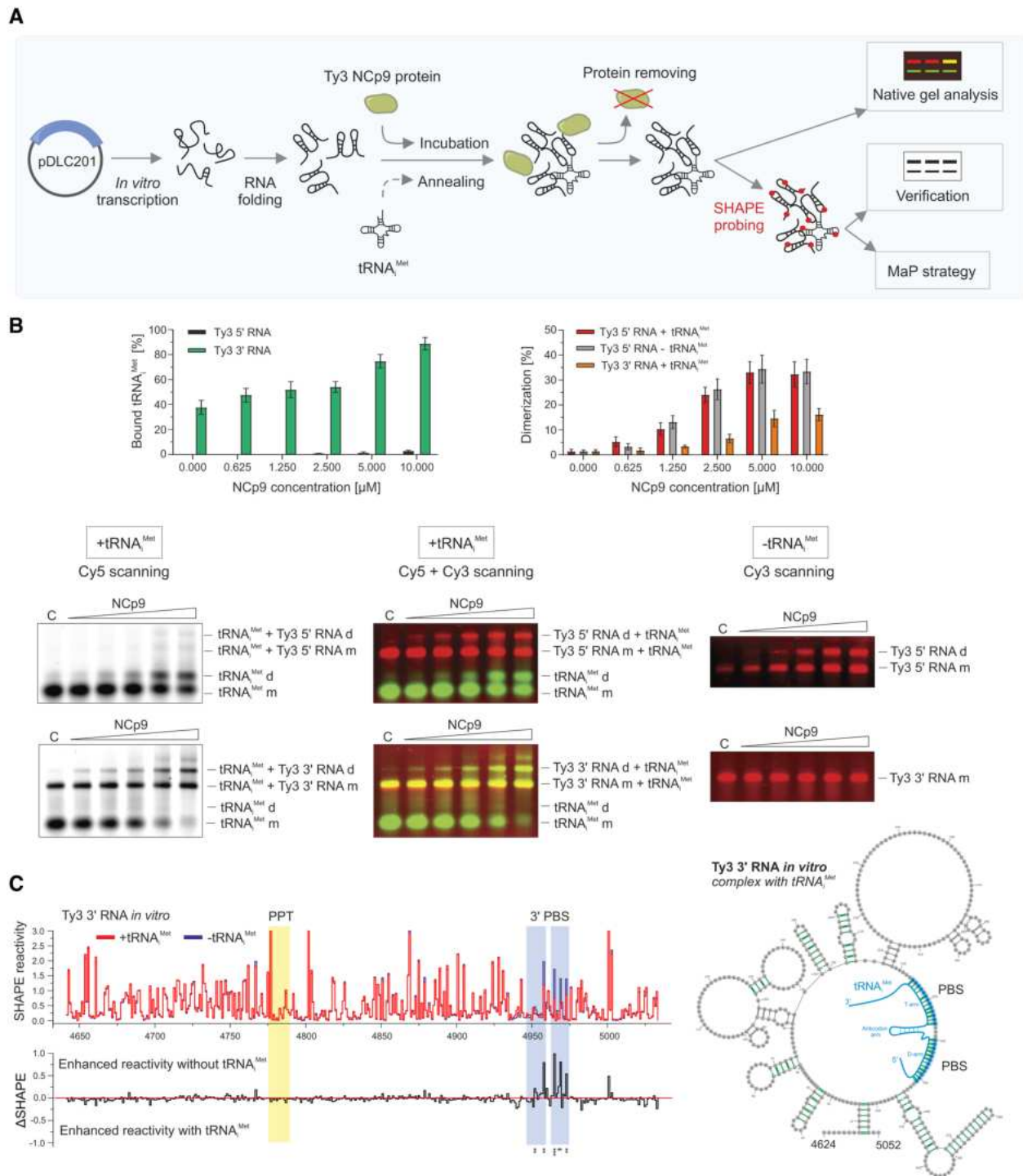


Figure 4. Analysis of NCp9-induced RNA–RNA interactions. **(A)** Schematic presentation of the applied strategy. **(B)** NCp9-mediated Ty3 5′–5′ and 3′–3′ RNA dimerization and tRNA_{Met} annealing assays. The graphs represent the percentage of bound tRNA_{Met} (left) and dimerized Ty3 RNA (right) with increasing concentrations of NCp9. Representative agarose gels for experiments with the addition or without tRNA_{Met} are presented below. Ty3 RNAs were labeled with Cy3, tRNA_{Met} with Cy5. Lanes C represent protein-free control reactions. Cy3 scan of gels are shown in [Supplementary Figure S5](#). **(C)** The step plot (top) and difference plot (bottom) of NAI reactivities for Ty3 3′ RNA in the presence and absence of tRNA_{Met}. Cis-acting sequences are marked with colored boxes. Nucleotide positions with statistically significant differential reactivity were marked by asterisks (SHAPE reactivity drop in tRNA_{Met} presence >0.15 and a *P*-value <0.05, using the Student's *t*-test). The MFE model for the Ty3 3′ RNA *in vitro* with the tRNA_{Met} annealed is presented on the right. HP bps are marked in green.

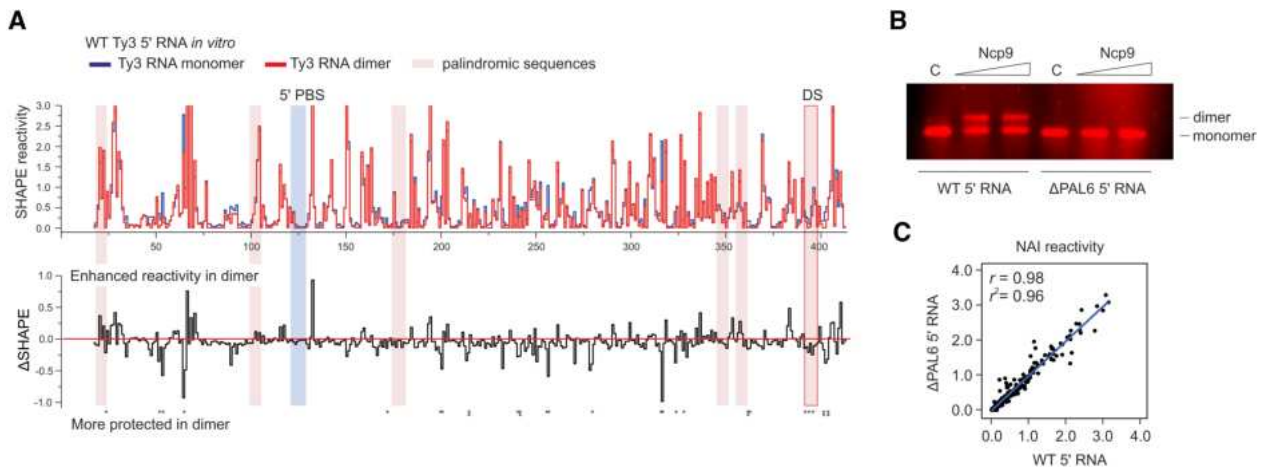


Figure 5. Identification of *cis*-acting sequence important for Ty3 5' RNA dimerization. **(A)** The step plot (top) and difference plot (bottom) of NAI reactivities for the Ty3 5' RNA *in vitro* in the monomeric and Ncp9-induced dimeric state. PBS and palindromic sequences (₁₈AUUAU₂₃, ₉₉CUCGAG₁₀₄, ₁₇₄UAGCUA₁₇₉, ₃₄₅ACAUGU₃₅₀, ₃₅₅GAUAUC₃₆₀, ₃₉₁CUGCAG₃₉₆) are marked with colored boxes. Nucleotide positions with statistically significant differential reactivity were marked by asterisks (SHAPE reactivity drop in dimer >0.15 and a *P*-value <0.05, using the Student's *t*-test). **(B)** Protein-mediated dimerization assay for WT and Δ PAL6 5' RNAs; concentrations of Ncp9: 5 and 10 μ M. Lanes C represent protein-free control reactions. **(C)** Pearson correlation (r) plot of normalized SHAPE reactivities between WT and Δ PAL6 5' RNAs *in vitro* datasets.

FS sequence remained unreactive toward SHAPE reagent both *in vivo* and *ex vivo* and is localized in the stem of the highly probable hairpin (Figure 6A–C).

We also found that the polypurine tract (PPT, ₄₇₇GAGAGAGAGGAAGA₄₇₈), which serves as a primer for Ty3 plus-strand synthesis, is located in a base-paired region in both gRNA states (Figure 6A–C). However, these pairings were predicted with lower probability, and boundary nucleotides exhibited high or moderate SHAPE modification levels. Studies indicate that the 5' and 3' ends of PPT contribute to RNase H recognition and are more sensitive to substitutions than internal positions, which may suggest the need for their higher accessibility (49,50).

Ty3 gRNA possesses potential pseudoknots and G-quadruplex motifs

Many RNA viruses utilize pseudoknot structures to control replication, translation, and the switching between these processes (51–53). Ty1 and HIV-1 gRNA contain experimentally validated pseudoknots, which are functionally important, as their disruption results in an inhibition of replication (45,54–56). Also, RNA G-quadruplexes regulate the life cycle of multiple viruses, including HIV-1 and SARS-CoV-2, by modulating the efficiency of reverse transcription or translation (57,58). *De novo* prediction and experimental validation of G-quadruplexes and pseudoknots, especially *in vivo*, is challenging. MFE structure algorithms are not able to identify such RNA motifs. Therefore, to identify *de novo* potential RNA pseudoknots in Ty3 gRNA, we used the SHAPE-MaP data as constraints for the dedicated ShapeKnots software (33). This analysis revealed two potential pseudoknots *in vivo* – in the GAG ORF (^{GAG}PK) and the POL ORF (^{POL}PK) in the proximity of the FS (Figure 7). In contrast to ^{POL}PK, the ^{GAG}PK was also predicted to be formed *ex vivo*. The ShapeKnots results were sustained by SHAPE-MaP data since nucleotides engaged in both pseudoknots were mainly unreactive in SHAPE-MaP experiments. In addition, we performed *in silico* predictions of potential RNA G-quadruplexes using QGRS Mapper

(34). We found only one potential G-quadruplex in Ty3 gRNA at the very beginning of the GAG ORF (12 nt after the AUG codon) (Figure 7). In this case, SHAPE data also supported the computational prediction result since guanosines engaged in potential G-quadruplex motif formation remained unreactive in Ty3 gRNA under both experimental conditions. Nevertheless, these RNA motifs' existence in the cells needs further experimental validation.

Identification of common structural features of retroelement RNA genomes

We searched for common gRNA secondary structure characteristics across diverse retroelements as there are deep evolutionary relationships between retroelements (59). For this purpose, we compared the SHAPE data obtained for Ty3 gRNA with those determined for yeast Ty1 retrotransposon and the HIV-1 retrovirus (24,25,36,60). These retroelements share important similarities in genome organization and replication cycle, but the HIV-1 gRNA is almost twice as long (9.2 kb) and encodes additional regulatory proteins and the Env, responsible for infectivity.

The gRNA of Ty3 is highly enriched in adenosines (~36%), similar to Ty1 and HIV-1 (Figure 8A). The overall base composition of the viral genome was shown to be related to local RNA structure stability (61,62). A-rich structures may facilitate RNA unfolding and refolding, resulting in less stable secondary structures prone to the formation of alternative conformers, while regions rich in Gs may stabilize functionally critical structures. In HIV-1 gRNA, accumulation of Gs was found in the 5' and 3' UTRs and the functional *cis*-acting sequences across the coding sequence—FS and the Rev response element (RRE) (Figure 8A, C). In contrast to HIV-1, we observed guanosine accumulation neither in 5' nor in 3' UTRs of both Ty elements (Figure 8A). However, we found Ty3 FS and DS are enriched in Gs, even though there is a lack of G domination over A (Figure 8C). This analysis indicates that the Ty3 retrotransposon, as evolutionarily more related to HIV-1, displays a better correlation between the sequence composition

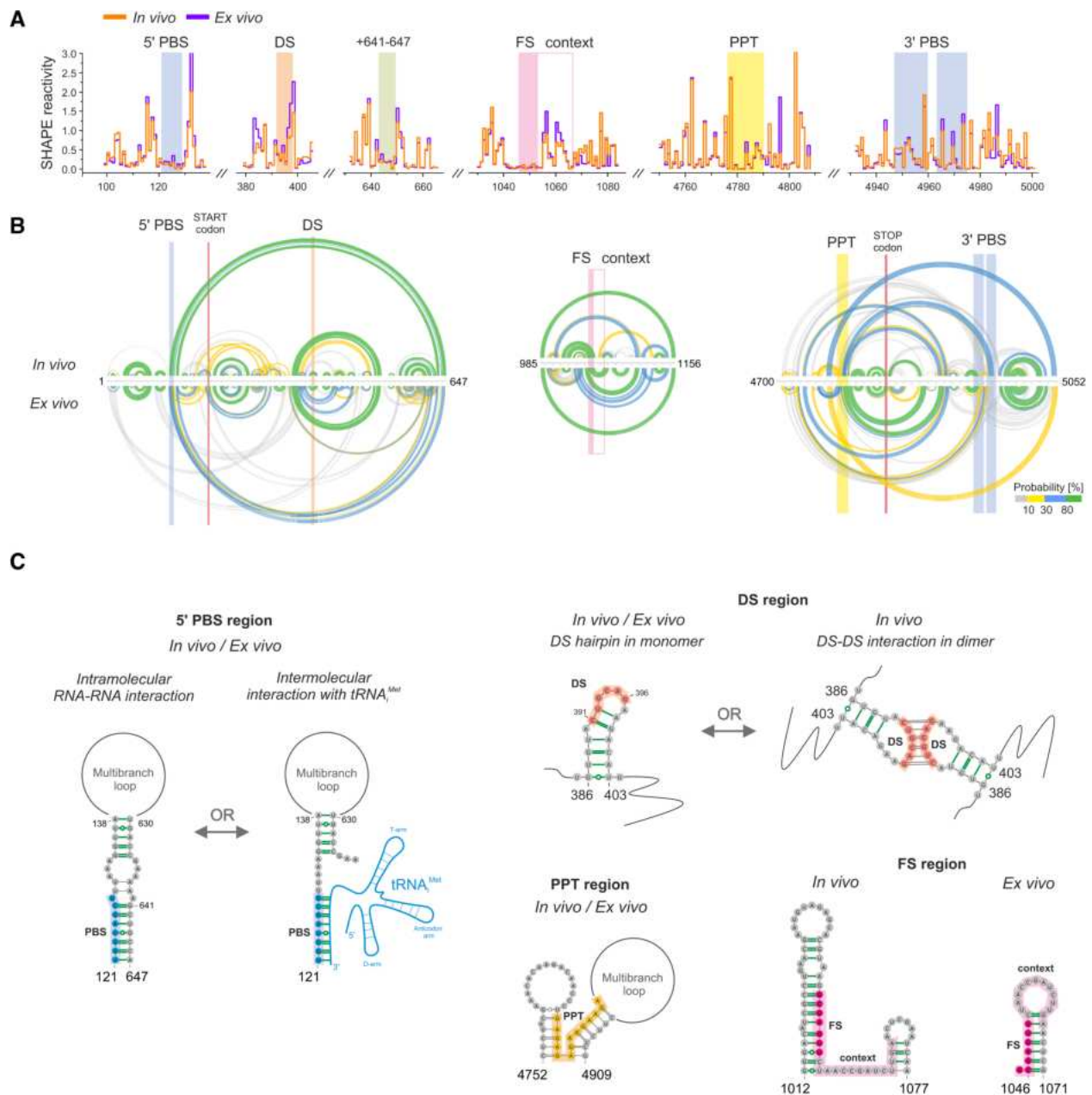


Figure 6. SHAPE-MaP-based analysis of Ty3 gRNA *cis*-acting elements. **(A)** The step plot of *in vivo* and *ex vivo* SHAPE reactivities for the selected regions of gRNA. *Cis*-acting sequences and 5' PBS interaction partner are marked with colored boxes. **(B)** Comparative analysis of base-pairing probabilities, represented by arc plots for regions containing *cis*-acting sequences. *Cis*-acting sequences and start or stop codons are marked with colored boxes. **(C)** Structural context of Ty3 gRNA *cis*-acting sequences. *In vivo* HP bps are marked in green.

and the location of *cis*-elements, in contrast to Ty1, where this correlation was very low.

SHAPE-directed structure models of Ty1 and Ty3 gRNA *in vivo* were characterized by a lower number and distance of MFE pairings, compared to cell-free state (Figure 8B). The more detailed heatmap of the local MFE bps content showed that highly structured regions are dispersed across these genomes, and their location is extensively shared between experimental states (Figure 8C). Interestingly, the analysis focused on *in vivo* HP bps revealed that the Ty3 gRNA possesses more regions of HP bps accumulation than Ty1 gRNA (Figure 8B, C). Simultaneously, we found higher similarity in

HP bps content, distance, and distribution profiles between the Ty3 *in vivo* and *ex vivo* states, compared to *in vivo* state and *in vitro* transcribed Ty1 gRNAs (Figure 8B, C). A common characteristic for the *in vivo* state of both Ty elements was a high accumulation of HP bps within the RT domain.

Furthermore, our studies showed that stable and highly structured (lowSS) regions constitute up to 30% of Ty1 and Ty3 gRNAs (Figure 8C) (24,25). For HIV-1 gRNA, identification of lowSS regions was performed only for cell-free state and they encompass about 40% of the sequence (36) (Figure 8C). More recently, a high structural heterogeneity across the entire HIV-1 gRNA was also shown in cells using the DREEM

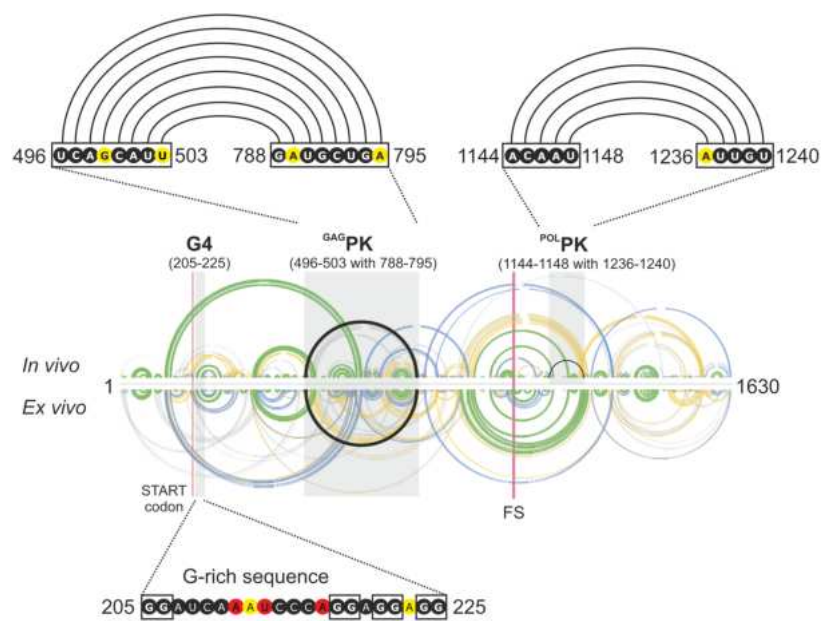


Figure 7. Potential pseudoknots and G-quadruplex predicted in Ty3 gRNA. The ^{GAG}PK and ^{POL}PK interactions are presented as black arcs. Black frames indicate nts involved in PK or G-quadruplexes, with their SHAPE reactivity. Base-pairing probabilities and SHAPE reactivities are colored with respect to scale in Figure 3.

algorithm (60). Altogether, these sequence and structural comparisons demonstrate that gRNAs from diverse retroelements possess a mosaic secondary structure with regions of higher structural stability interspersed with dynamic ones. Importantly, Ty1, Ty3, and HIV-1 have well-characterized and functionally analogous *cis*-acting sequences in their gRNAs. The common feature is a localization of sequences involved in tRNA annealing, gRNA dimerization, cyclization and packaging in the well-structured 5' end of gRNA (Figure 8C). In contrast, the FS is located in a structurally stable region of gRNA for Ty3 and HIV-1 but not for Ty1.

We also looked for the correlation between the location of retrotransposon protein domain boundaries and the local level of gRNA structure, as it was previously clearly observed for HIV-1 (63). For Ty3, we observed an evident increase in RNA structure level in the region linking the J and RT domains in *POL* ORF, and a little increase at the beginning of the CA in *GAG* ORF (Supplementary Figure S10). Thus, the correlation is much less evident than observed for HIV-1. For Ty1 retrotransposon, most protein domain boundaries do not exhibit a higher level of RNA structure, excluding the correlation between RNA structure and protein folding.

Discussion

The yeast Ty3 LTR-retrotransposon is an informative model system for studying retrotransposition mechanisms and the impact of transposable elements on host genomes (5). In this work, we present the complete structural map for the yeast Ty3 retrotransposon RNA genome using the SHAPE-MaP strategy both in living cells and *ex vivo*. A comparative analysis of Ty3 gRNA structure in different experimental states facilitates a more correct interpretation of the relationship between RNA structure, function, and cellular factors. Only several studies provide such comparisons for viral RNAs

(24,64–69). Consistent with these studies, we observed a slight shift in Ty3 gRNA structure towards a more unfolded state *in vivo*. Nevertheless, over half of the predicted MFE bps and three-fourths of the HP bps are shared between the *in vivo* and *ex vivo* Ty3 gRNA states. This structural similarity indicates that many interactions are independent of the folding environment Ty3 gRNA encounters *in vivo*. Like other single-stranded RNA viruses, Ty3 gRNA has a complex secondary structure enriched in stable, specific structural motifs. We identified a well-folded core formed by more stable regions without the support of cellular factors, and it encompasses about 20% of the Ty3 gRNA sequence. Among identified structurally stable regions with low SHAPE reactivity and Shannon entropy (lowSS regions), half is included in the well-folded core, but the others are specific for each experimental state. The stable motifs were shown to be functionally necessary for the replication of RNA viruses (36,67,70–72). Accordingly, Ty3 lowSS regions encompass *cis*-acting sequences, like 5' PBS, FS, and DS, but the role of other gRNA stable motifs remains to be explored.

Analogous to other retroelements, Ty3 gRNA is present in the dimeric form in VLP (13). The sequences involved in Ty3 gRNA dimerization were not clearly identified, and it remains unknown whether, like in retroviruses, dimerization is essential for Ty3 RNA packaging into VLP. An early study suggests that the Ty3 gRNA dimer *in vitro* is maintained by the interaction between two tRNA_i^{Met} molecules co-linked to the 5' and 3' PBS sequences (17). However, further study demonstrates that deletion of 5' and 3' PBS does not prevent Ty3 gRNA packaging (20). Therefore, either tRNA_i^{Met} is unnecessary for Ty3 gRNA dimerization or dimerization is not a pre-requisite for this RNA packaging. Alternatively, different mechanisms of dimerization are also possible. We showed that the transcript encompassing Ty3 gRNA 5' end dimerizes efficiently without tRNA_i^{Met} mediation. This finding indicates that Ty3 retrotransposon follows the common mechanism of

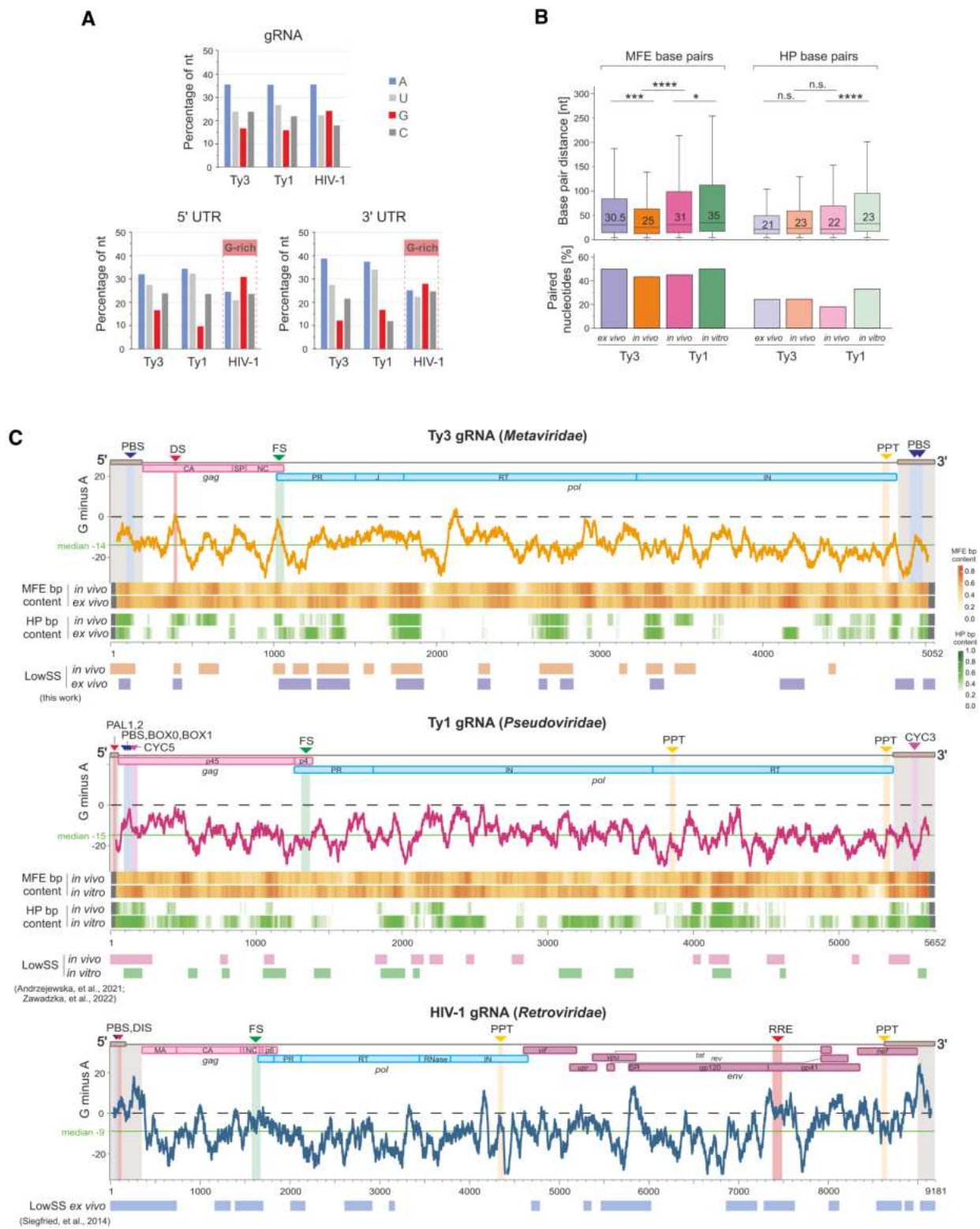


Figure 8. Comparative analysis of Ty3, Ty1, and HIV-1 RNA genomes. **(A)** Bar plot analysis of nucleotide composition in CDS, 5' UTR, and 3' UTR. **(B)** Distribution of MFE and HP bp distances with medians and the percentage of nucleotides engaged in these pairings. **(C)** Profiles of G minus A comparison for gRNA sequences. The number of As was subtracted from the number of Gs (G minus A) in a 75-nt sliding window. The green solid line indicates the median, and the black dashed line indicates where the difference between the number of Gs and As is zero. Values above the dashed line indicate a higher G concentration than A. Heatmaps representing Ty3 or Ty1 bp content (MFE and HP) are shown below the G minus A profiles. The lowSS regions are shown on the bottom. Their locations for Ty3 gRNA were taken from this work, for Ty1 gRNA from (24,25); and for HIV-1 gRNA from (36). PBS, BOX0, BOX1 – primer binding sites, DS, PAL1, PAL2, and DIS – dimerization sequences, CYC5, CYC3 – cyclization sequences, FS – frameshift sequence, PPT – polypurine tract, RRE – Rev response element.

retroelement gRNA dimerization, which occurs through the direct interaction of *cis*-acting palindromic sequences from the 5' end. And importantly, we propose a novel functional Ty3 RNA sequence – DS/PAL6, that directly drives the dimerization process. We also showed that DS/PAL6 is less reactive *in vivo* than *ex vivo*, suggesting its involvement in Ty3 gRNA dimerization in yeast. Further studies are needed to explain the role of dimerization via DS/PAL6 for Ty3 gRNA packaging into VLPs.

The primary function of tRNA_i^{Met} in Ty3 replication is the initiation of gRNA reverse transcription, and its annealing to both 5' and 3' PBS sequences is necessary for this process (17). Since SHAPE does not discriminate intra- and intermolecular RNA interactions, resolving the issue of tRNA_i^{Met} annealing to 5' PBS is challenging due to its low reactivity in all experimental states and tendency to intramolecular interactions. Nevertheless, SHAPE results obtained for 3' PBS suggest that only a tiny fraction of Ty3 gRNA interacts with tRNA_i^{Met} *in vivo* and *ex vivo*.

Studies for other RNA viruses highlight the role of -1 FS structural context in modulating frameshifting efficiency (53). In contrast, the +1 frameshift mechanism in Ty3 is postulated to rely mainly on the specific peptidyl-tRNA availability but can be additionally induced by the 'context' sequence (48). Our results provide a structural basis to further understand how the stable secondary structure of the Ty3 FS element along with the single-stranded 'context' sequence regulate frameshifting. However, further experimental validation is required to confirm this hypothesis, such as studying the frameshift rate when a well-folded hairpin with FS sequence is disrupted, or some stable motifs are incorporated within the 'context' sequence.

In the cytoplasm, Ty3 gRNA co-localizes with Ty3 proteins in P-body-associated foci (termed retrosomes), where VLP assembly occurs (9,16). Apart from the UTRs, the *POL* region plays an independent role in the localization of Ty3 gRNA in the cytoplasmic foci (20). A level of ribosome occupancy of *POL* was shown to be inversely correlated with Ty3 gRNA localization in cytoplasmic foci. It was proposed that ribosomes may impede access to located in *POL* sequences important for binding host factors involved in Ty3 localization. However, these sequences are not distinguished explicitly so far. We identified several stable lowSS regions across Ty3 *POL*, which represent attractive candidates for further Ty3 gRNA localization studies.

To date, Ty3 is the first endogenous retroelement in the *Metaviridae* family, along with the yeast Ty1 retrotransposon in the *Pseudoviridae* family, whose gRNA structure has been determined in a native *in vivo* state (24,25). RNA genomes of both Ty retrotransposons are significantly enriched in A nucleobases and exhibit some general structural similarities, including a mosaic RNA structure, which is more open *in vivo* than under cell-free conditions. Nevertheless, due to higher HP bp content, Ty3 gRNA structural motifs may be more stable *in vivo* than Ty1 gRNA. These HP bps, together with the same RNA posttranscriptional processing, may contribute to better structural similarity between Ty3 gRNA *in vivo* and *ex vivo* states. Moreover, only Ty3 gRNA is enriched for G nucleobases in functionally essential sequences, similar to the HIV-1 gRNA (61). However, there is no evident correlation between RNA structure level and protein domain boundaries as was found in HIV-1 (63).

Overall, the comprehensive characterization of Ty3 gRNA secondary structure, along with a comparative analysis with diverse retroelements, broadens our understanding of the relationship between gRNA architecture and the replication process. This study also provides the basis for understanding the function of retrotransposon RNA transactions at a molecular level.

Data availability

Raw sequencing data can be obtained from <https://www.ncbi.nlm.nih.gov/bioproject/> and can be accessed under accession code PRJNA1072546. All further data generated or analyzed during this study are included in this published article and its supplementary file.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

A.A.-R. was supported by L'Oreal-UNESCO for Women in Science and the START Scholarship from the Foundation for Polish Science. A.A.-R. and E.T. were Polish National Science Centre scholarship holders (2020/39/B/NZ3/03020). The scientific work was created using the infrastructure of the Poznan Supercomputing and Networking Center (PSNC).

We thank David J. Garfinkel for the helpful discussions and all the suggestions that improved this manuscript. We also thank Tomasz Żok for providing informatics support with the PSNC infrastructure and applying the ShapeKnots in sliding windows.

Author contributions: A.A.-R.: Conceptualization, Investigation, Methodology, Validation, Data curation, Formal analysis, Visualization, Writing - original draft, review & editing, Funding acquisition. J.G.: Investigation, Methodology, Validation, Data curation, Formal analysis, Writing - review & editing, Funding acquisition. E.T.: Investigation. K.P.-W.: Supervision, Conceptualization, Formal analysis, Writing - review & editing, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Polish National Science Centre [2021/41/N/NZ3/04060 to A.A.-R., 2019/35/N/NZ1/01954 to J.G., 2020/39/B/NZ3/03020 to K.P.-W.]. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. Funding for open access charge: Polish National Science Centre [2021/41/N/NZ3/04060 to A. A.-R., 2019/35/N/NZ1/01954 to J. G.], and the Institute of Bioorganic Chemistry PAS.

Conflict of interest statement

None declared.

References

- Krupovic, M., Blomberg, J., Coffin, J.M., Dasgupta, I., Fan, H., Geering, A.D., Gifford, R., Harrach, B., Hull, R., Johnson, W., et al. (2018) Ortervirales: new virus order unifying five families of reverse-transcribing viruses. *J. Virol.*, **92**, e00515-18.
- Neumann, P., Novák, P., Hošťáková, N. and Macas, J. (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mob. DNA*, **10**, 1.
- Lesage, P. and Todeschini, A.L. (2005) Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenet. Genome Res.*, **110**, 70–90.
- Mita, P. and Boeke, J.D. (2016) How retrotransposons shape genome regulation. *Curr. Opin. Genet. Dev.*, **37**, 90–100.
- Sandmeyer, S., Patterson, K. and Bilanchone, V. (2015) Ty3, a position-specific retrotransposon in Budding Yeast. *Microbiol. Spectr.*, **3**, MDNA3-A0057-2014.
- Srikulnath, K., Ahmad, S.F., Singchat, W. and Panthum, T. (2022) Do Ty3/gypsy transposable elements play preferential roles in sex chromosome differentiation? *Life*, **12**, 522.
- Ashley, J., Cordy, B., Lucia, D., Fradkin, L.G., Budnik, V. and Thomson, T. (2018) Retrovirus-like gag protein Arc1 binds RNA and traffics across synaptic boutons. *Cell*, **172**, 262–274.
- Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A.V., McCormick, J., Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R., et al. (2018) The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell*, **172**, 275–288.
- Beliakova-Bethell, N., Beckham, C., Giddings, T.H., Winey, M., Parker, R. and Sandmeyer, S. (2006) Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. *RNA*, **12**, 94–101.
- Malagon, F. and Jensen, T.H. (2011) T-body formation precedes virus-like particle maturation in *S. cerevisiae*. *RNA Biol*, **8**, 184–189.
- Hansen, L.J., Chalker, D.L. and Sandmeyer, S.B. (1988) Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Microbiol. Spectr.*, **8**, 12.
- Keeney, J.B., Chapman, K.B., Lauer, M., Voytas, D.F., Aström, S.U., von Pawel-Rammingen, U., Byström, A. and Boeke, J.D. (1995) Multiple molecular determinants for retrotransposition in a primer tRNA. *Mol. Cell. Biol.*, **15**, 217–226.
- Nymark-McMahon, M.H., Beliakova-Bethell, N.S., Darlix, J.-L., Le Grice, S.F.J. and Sandmeyer, S.B. (2002) Ty3 Integrase is required for initiation of reverse transcription. *J. Virol.*, **76**, 2804–2816.
- Dodonova, S.O., Prinz, S., Bilanchone, V., Sandmeyer, S. and Briggs, J.A.G. (2019) Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 10048–10057.
- Chalker, D.L. and Sandmeyer, S.B. (1992) Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev.*, **6**, 117–128.
- Kirchner, J., Sandmeyer, S.B. and Forrest, D.B. (1992) Transposition of a Ty3 GAG3-POL3 fusion mutant is limited by availability of capsid protein. *J. Virol.*, **66**, 6081–6092.
- Gabus, C., Ficheux, D., Rau, M., Keith, G., Sandmeyer, S. and Darlix, J.L. (1998) The yeast Ty3 retrotransposon contains a 5'-3' bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. *EMBO J.*, **17**, 4873–4880.
- Larsen, L.S.Z., Beliakova-Bethell, N., Bilanchone, V., Zhang, M., Lamsa, A., DaSilva, R., Hatfield, G.W., Nagashima, K. and Sandmeyer, S. (2008) Ty3 Nucleocapsid controls localization of particle assembly. *J. Virol.*, **82**, 2501–2514.
- Chaurasiya, K.R., Geertsema, H., Cristofari, G., Darlix, J.-L. and Williams, M.C. (2012) A single zinc finger optimizes the DNA interactions of the nucleocapsid protein of the yeast retrotransposon Ty3. *Nucleic Acids Res.*, **40**, 751–760.
- Clemens, K., Bilanchone, V., Beliakova-Bethell, N., Larsen, L.S.Z., Nguyen, K. and Sandmeyer, S. (2013) Sequence requirements for localization and packaging of Ty3 retroelement RNA. *Virus Res.*, **171**, 319–331.
- Pachulski-Wieczorek, K., Le Grice, S. and Purzycka, K. (2016) Determinants of genomic RNA encapsidation in the *Saccharomyces cerevisiae* long terminal repeat retrotransposons Ty1 and Ty3. *Viruses*, **8**, 193.
- Andrzejewska, A., Zawadzka, M. and Pachulski-Wieczorek, K. (2020) On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective. *Int. J. Mol. Sci.*, **21**, 6770.
- Boerneke, M.A., Ehrhardt, J.E. and Weeks, K.M. (2019) Physical and functional analysis of viral RNA genomes by SHAPE. *Annu. Rev. Virol.*, **6**, 93–117.
- Andrzejewska, A., Zawadzka, M., Gumna, J., Garfinkel, D.J. and Pachulski-Wieczorek, K. (2021) *In vivo* structure of the Ty1 retrotransposon RNA genome. *Nucleic Acids Res.*, **49**, 2878–2893.
- Zawadzka, M., Andrzejewska-Romanowska, A., Gumna, J., Garfinkel, D.J. and Pachulski-Wieczorek, K. (2022) Cell compartment-specific folding of Ty1 long terminal repeat retrotransposon RNA genome. *Viruses*, **14**, 2007.
- Schiestl, R.H. and Gietz, R.D. (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.*, **16**, 339–346.
- Tripp, J.D., Lilley, J.L., Wood, W.N. and Lewis, L.K. (2013) Enhancement of plasmid DNA transformation efficiencies in early stationary-phase yeast cell cultures. *Yeast*, **30**, 191–200.
- Gumna, J., Purzycka, K.J., Ahn, H.W., Garfinkel, D.J. and Pachulski-Wieczorek, K. (2019) Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biol*, **16**, 1749–1763.
- Lusvardi, S., Sztuba-Solinska, J., Purzycka, K.J., Pauly, G.T., Rausch, J.W. and Grice, S.F.J.L. (2013) The HIV-2 rev-response element: determining secondary structure and defining folding intermediates. *Nucleic Acids Res.*, **41**, 6637–6649.
- Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A. and Weeks, K.M. (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.
- Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinf.*, **11**, 129.
- Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinforma. Oxf. Engl.*, **25**, 1974–1975.
- Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H. and Weeks, K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5498–5503.
- Kikin, O., D'Antonio, L. and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.
- Smola, M.J., Calabrese, J.M. and Weeks, K.M. (2015) Detection of RNA-Protein interactions in living cells with SHAPE. *Biochemistry*, **54**, 6867–6875.
- Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E. and Weeks, K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, **11**, 959–965.
- Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
- Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T. and Chang, H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.

39. Kinsey,P.T. and Sandmeyer,S.B. (1995) Ty3 transposes in mating populations of yeast: a novel transposition assay for Ty3. *Genetics*, **139**, 81–94.
40. Larsen,L.S.Z., Zhang,M., Beliakova-Bethell,N., Bilanchone,V., Lamsa,A., Nagashima,K., Najdi,R., Kosaka,K., Kovacevic,V., Cheng,J., *et al.* (2007) Ty3 Capsid mutations reveal early and late functions of the amino-terminal domain. *J. Virol.*, **81**, 6957–6972.
41. Clemens,K., Larsen,L., Zhang,M., Kuznetsov,Y., Bilanchone,V., Randall,A., Harned,A., DaSilva,R., Nagashima,K., McPherson,A., *et al.* (2011) The TY3 Gag3 spacer controls intracellular condensation and uncoating. *J. Virol.*, **85**, 3055–3066.
42. Busan,S. and Weeks,K.M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, **24**, 143–148.
43. Farabaugh,P.J., Zhao,H. and Vimaladithan,A. (1993) A novel programmed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: frameshifting without tRNA slippage. *Cell*, **74**, 93–103.
44. Xiong,Y. and Eickbush,T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
45. Purzycka,K.J., Legiewicz,M., Matsuda,E., Eizenstat,L.D., Lusvardi,S., Saha,A., Grice,S.F.J.L. and Garfinkel,D.J. (2013) Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res.*, **41**, 463–473.
46. Dubois,N., Khoo,K.K., Ghossein,S., Seisser,T., Wolff,P., McKinstry,W.J., Mak,J., Paillart,J.-C., Marquet,R. and Bernacchi,S. (2018) The C-terminal p6 domain of the HIV-1 Pr55^{Gag} precursor is required for specific binding to the genomic RNA. *RNA Biol*, **15**, 923–936.
47. Purzycka,K.J., Pachulska-Wieczorek,K. and Adamiak,R.W. (2011) The in vitro loose dimer structure and rearrangements of the HIV-2 leader RNA. *Nucleic Acids Res.*, **39**, 7234–7248.
48. Farabaugh,P. (1996) Programmed Translational frameshifting. *Annu. Rev. Genet.*, **30**, 507–528.
49. Dash,C., Marino,J.P. and Le Grice,S.F.J. (2006) Examining Ty3 polypurine tract structure and function by nucleoside analog interference. *J. Biol. Chem.*, **281**, 2773–2783.
50. Lener,D., Kvaratskhelia,M. and Le Grice,S.F.J. (2003) Nonpolar thymine isosteres in the Ty3 polypurine tract DNA template modulate processing and provide a model for its recognition by Ty3 reverse transcriptase. *J. Biol. Chem.*, **278**, 26526–26532.
51. Brierley,I., Pennell,S. and Gilbert,R.J.C. (2007) Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat. Rev. Microbiol.*, **5**, 598–610.
52. Olsthoorn,R.C.L. (2022) Replication of alphaviruses requires a pseudoknot that involves the poly(A) tail. *RNA*, **28**, 1348–1358.
53. Penn,W.D., Harrington,H.R., Schleich,J.P. and Mukhopadhyay,S. (2020) Regulators of viral frameshifting: more than RNA influences translation events. *Annu. Rev. Virol.*, **7**, 219–238.
54. Huang,Q., Purzycka,K.J., Lusvardi,S., Li,D., LeGrice,S.F.J. and Boeke,J.D. (2013) Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA*, **19**, 320–332.
55. Paillart,J.-C., Skripkin,E., Ehresmann,B., Ehresmann,C. and Marquet,R. (2002) In vitro evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA*. *J. Biol. Chem.*, **277**, 5995–6004.
56. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
57. Amrane,S., Jaubert,C., Bedrat,A., Rundstadler,T., Recordon-Pinson,P., Akinin,C., Guédin,A., De Rache,A., Bartolucci,L., Diene,I., *et al.* (2022) Deciphering RNA G-quadruplex function during the early steps of HIV-1 infection. *Nucleic Acids Res.*, **50**, 12328–12343.
58. Qin,G., Zhao,C., Liu,Y., Zhang,C., Yang,G., Yang,J., Wang,Z., Wang,C., Tu,C., Guo,Z., *et al.* (2022) RNA G-quadruplex formed in SARS-CoV-2 used for COVID-19 treatment in animal models. *Cell Discov.*, **8**, 86.
59. Boeke,J.D. and Stoye,J.P. (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin,J.M., Hughes,S.H. and Varmus,H.E. (eds.) *Retroviruses*. Cold Spring Harbor Laboratory Press, NY.
60. Tomezsko,P.J., Corbin,V.D.A., Gupta,P., Swaminathan,H., Glasgow,M., Persad,S., Edwards,M.D., Mcintosh,L., Papenfuss,A.T., Emery,A., *et al.* (2020) Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature*, **582**, 438–442.
61. Pollom,E., Dang,K.K., Potter,E.L., Gorelick,R.J., Burch,C.L., Weeks,K.M. and Swanstrom,R. (2013) Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. *PLoS Pathog.*, **9**, e1003294.
62. van Hemert,F.J., van der Kuyl,A.C. and Berkhout,B. (2013) The A-nucleotide preference of HIV-1 in the context of its structured RNA genome. *RNA Biol*, **10**, 211–215.
63. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
64. Dadonaite,B., Gilbertson,B., Knight,M.L., Trifkovic,S., Rockman,S., Laederach,A., Brown,L.E., Fodor,E. and Bauer,D.L.V. (2019) The structure of the influenza A virus genome. *Nat. Microbiol.*, **4**, 1781–1789.
65. Huber,R.G., Lim,X.N., Ng,W.C., Sim,A.Y.L., Poh,H.X., Shen,Y., Lim,S.Y., Sundstrom,K.B., Sun,X., Aw,J.G., *et al.* (2019) Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.*, **10**, 1408.
66. Li,P., Wei,Y., Mei,M., Tang,L., Sun,L., Huang,W., Zhou,J., Zou,C., Zhang,S., Qin,C.-F., *et al.* (2018) Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe*, **24**, 875–886.
67. Madden,E.A., Plante,K.S., Morrison,C.R., Kutchko,K.M., Sanders,W., Long,K.M., Taft-Benz,S., Cruz Cisneros,M.C., White,A.M., Sarkar,S., *et al.* (2020) Using SHAPE-MaP to model RNA secondary structure and identify 3'UTR variation in Chikungunya virus. *J. Virol.*, **94**, e00701-20.
68. Manfredonia,I., Nithin,C., Ponce-Salvatierra,A., Ghosh,P., Wirecki,T.K., Marinus,T., Ogando,N.S., Snijder,E.J., van Hemert,M.J., Bujnicki,J.M., *et al.* (2020) Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.*, **48**, 12436–12452.
69. Simon,L.M., Morandi,E., Luginani,A., Gribaudo,G., Martinez-Sobrido,L., Turner,D.H., Oliviero,S. and Incarnato,D. (2019) In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs. *Nucleic Acids Res.*, **47**, 7003–7017.
70. Dethoff,E.A., Boerneke,M.A., Gokhale,N.S., Muhire,B.M., Martin,D.P., Sacco,M.T., McFadden,M.J., Weinstein,J.B., Messer,W.B., Horner,S.M., *et al.* (2018) Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 11513–11518.
71. Huston,N.C., Wan,H., Strine,M.S., de Cesaris Araujo Tavares,R., Wilen,C.B. and Pyle,A.M. (2021) Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell*, **81**, 584–598.
72. Wan,H., Adams,R.L., Lindenbach,B.D. and Pyle,A.M. (2022) The *In vivo* and *In vitro* architecture of the Hepatitis C Virus RNA genome uncovers functional RNA secondary and tertiary structures. *J. Virol.*, **96**, e01946-21.

SUPPLEMENTARY TABLES AND FIGURES

Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome

Table S1. Primers used in the study.

PRIMER	SEQUENCE (5'-3')
Colony PCR	
PF_Gal_Ty3_part	GCTTCTAATCCGTA CTTC AATATAG
PR_Ty3_part	CGTCTTTTAATGTCTCTGGTTTG
PF_AMP_part	CCTATCTCAGCGATCTGTCTAT
PR_AMP_part	GAGTATTCAACATTTCCGTGTCTG
Reverse transcription and PCR	
PF_amp1	TAAGTAACATTCCGTG
PR_amp1	CGTCTTTTAATGTCTC
PF_amp2	GACCGAAAAAGCTGCCATAAT
PR_amp2	CTAAAGTTATATGCAGGTCATTGAT
PF_amp3	TCAAATACGAAATCTACGAGACC
PR_amp3	GCCAGGTTTAATTTCAATATCATG
PF_amp4	ACAACAGAAGTATAGAGAGATCATAAC
PR_amp4	TAGTGGAGCTATTTTCTGGATTCC
PF_amp5	CATTGGAAACATTTAGACACGG
PR_amp5	CATCTGCGACAACGTTCTTG
PF_amp6	TATCATTACAAAACAAGAACGAACC
PR_amp6	GTTTTCCCTTG TAGCTATGAAGTG
PF_amp7	TATCAATGGATTTTGTGACAGGATT
PR_amp7	GAAATGGTCCGACGTATATTTG
PF_amp8	TTATTGTTAAACATAGGAGATCACG
PR_amp8	GTCAAAACAGTTTATCAGATTAATTCAC
PF_invitro_5end	TAAGTAACATTCCGTG
PR_invitro_5end_16	GCCCACTGAGCAGCGG
PR_invitro_5end_12	GCCCACTGAGCA
PR_invitro_5end_ΔPAL6	TACACAATACGACTGGCATCATT
PF_invitro_3end	CTACACACAAAACCTTACT
PR_invitro_3end	GTCAAAACAGTTTATCAG
PCR for <i>in vitro</i> transcription templates	
PF_5'_Ty3	TAATACGACTCACTATAGTAAGTAACATTCCGT
PR_5'_Ty3	GCCCACTGAGCAGCGGGGTT
PR_5'_Ty3ΔPAL6	TACACAATACGACTGGCATCATT
PF_3'_Ty3	TAATACGACTCACTATAGAGGGGCCAGC
PR_3'_Ty3	GTCAAAACAGTTTATCAGATTAATTCACGGAATG

Supplementary Figure Legends

Figure S1. Pearson correlation (r) plots of normalized SHAPE reactivities between different NAI concentrations and modification times *in vivo* and *ex vivo*. Tests were performed for the first amplicon.

Figure S2. Box plot analysis of mutation rates for each nucleotide identity for modified and untreated samples from Ty3 gRNA probing under *in vivo* and *ex vivo* conditions. The boxes represent the interquartile range; a line indicates the median. Significance was computed by the Wilcoxon rank-sum test; **** $P < 0.0001$.

Figure S3. Identification of significant SHAPE reactivity differences between *in vivo* and *ex vivo* Ty3 gRNA states using the Δ SHAPE framework. To identify local and global differences, SHAPE profiles were smoothed with 3-nt and 51-nt sliding windows, respectively. Regions that meet the standard score and Z-factor criteria are marked with green and violet shadings.

Figure S4. Thermal stability of NCp9-induced Ty3 5'-5' and 3'-3' RNA dimers after protein removal. For 3' RNA, tRNA_i^{Met} was present in the reaction. Ty3 RNAs were labeled with Cy3, tRNA_i^{Met} with Cy5.

Figure S5. Cy3 scan (signal for Ty3 5' or 3' RNA) of representative agarose gels for NCp9-mediated 5'-5' and 3'-3' Ty3 RNA dimerization assays in the presence of tRNA_i^{Met}. Lanes marked C represent protein-free samples.

Figure S6. Base-pairing probabilities for Ty3 5' RNA and Ty3 3' RNAs *in vitro* compared to Ty3 gRNA *in vivo*. *Cis*-acting sequences and start or stop codons are marked with colored boxes. Pearson correlation (r) plot of normalized SHAPE reactivities between *in vivo* and *in vitro* datasets.

Figure S7. The step plot of NAI reactivities for Ty3 5' RNA *in vitro* in the presence and absence of tRNA_i^{Met}. PBS and predicted interaction partner are marked with colored boxes. The arc plot of this interaction is presented below.

Figure S8. Identification of significant SHAPE reactivity differences between monomeric and dimeric states of Ty3 5' RNA *in vitro* using the Δ SHAPE framework. Before Δ SHAPE calculation, SHAPE profiles were smoothed with a 3-nt sliding window. Regions that meet the standard score and Z-factor criteria are marked with green and violet shadings. PBS and palindromic sequences are marked with colored boxes.

Figure S9. The step plot (top) and difference plot (bottom) of NAI reactivities for the 5' end of Ty3 gRNA *in vivo* and *ex vivo*. Nucleotide positions that exhibit statistically significant differential reactivity (SHAPE reactivity drop *in vivo* > 0.15 and a P -value < 0.05 , using the

Student's t-test) were marked by asterisks according to statistical significance. PBS and palindromic sequences are marked with colored boxes.

Figure S10. Comparative analysis of the gRNA structure level (median SHAPE and Gini index profiles) and the protein domain boundaries for Ty3 and Ty1 genomes.

Figure S1.

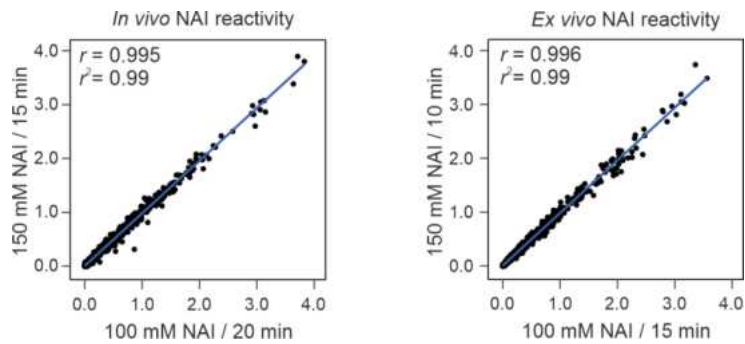


Figure S2.

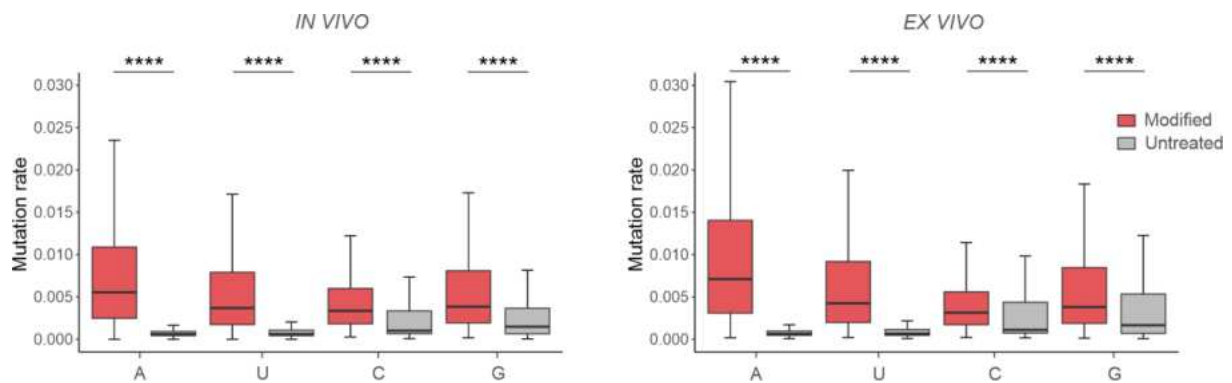


Figure S3.

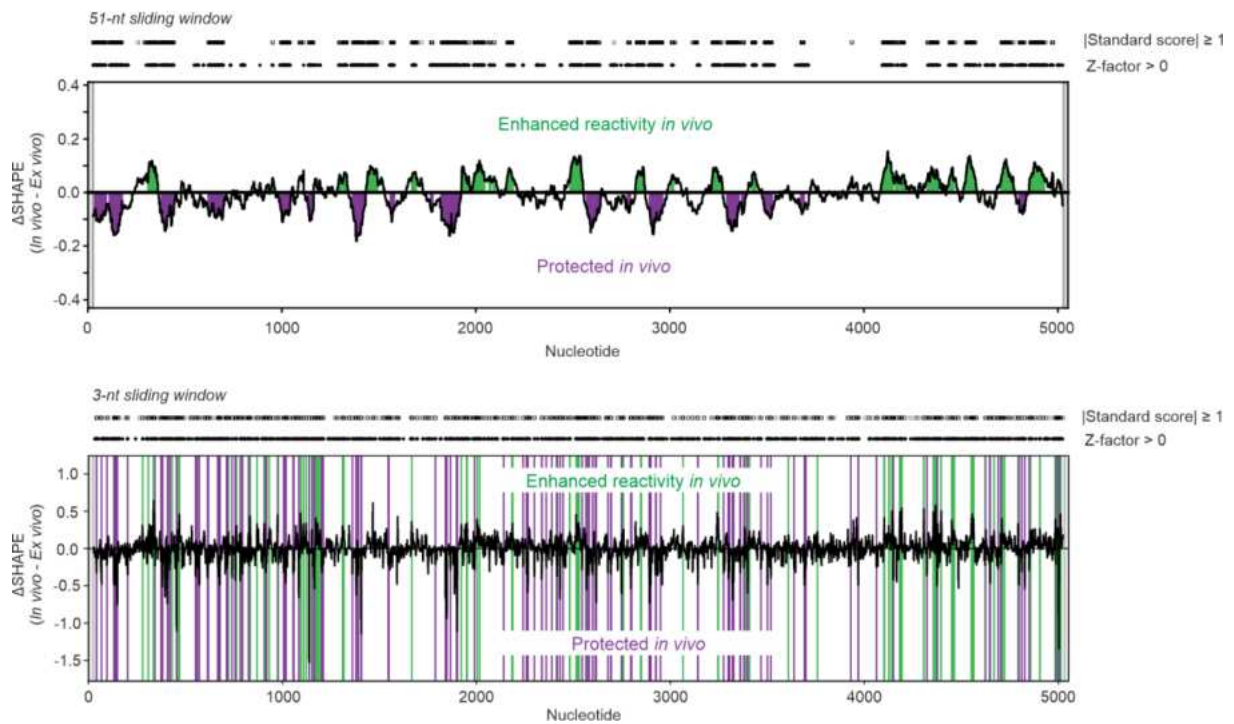


Figure S4.

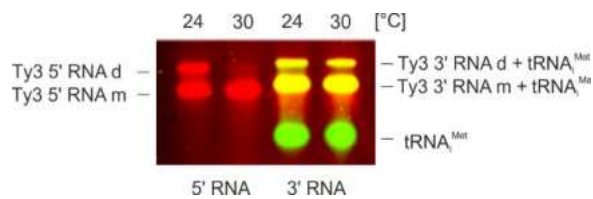


Figure S5.

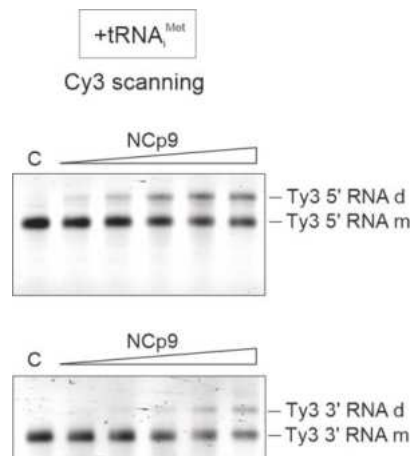


Figure S6.

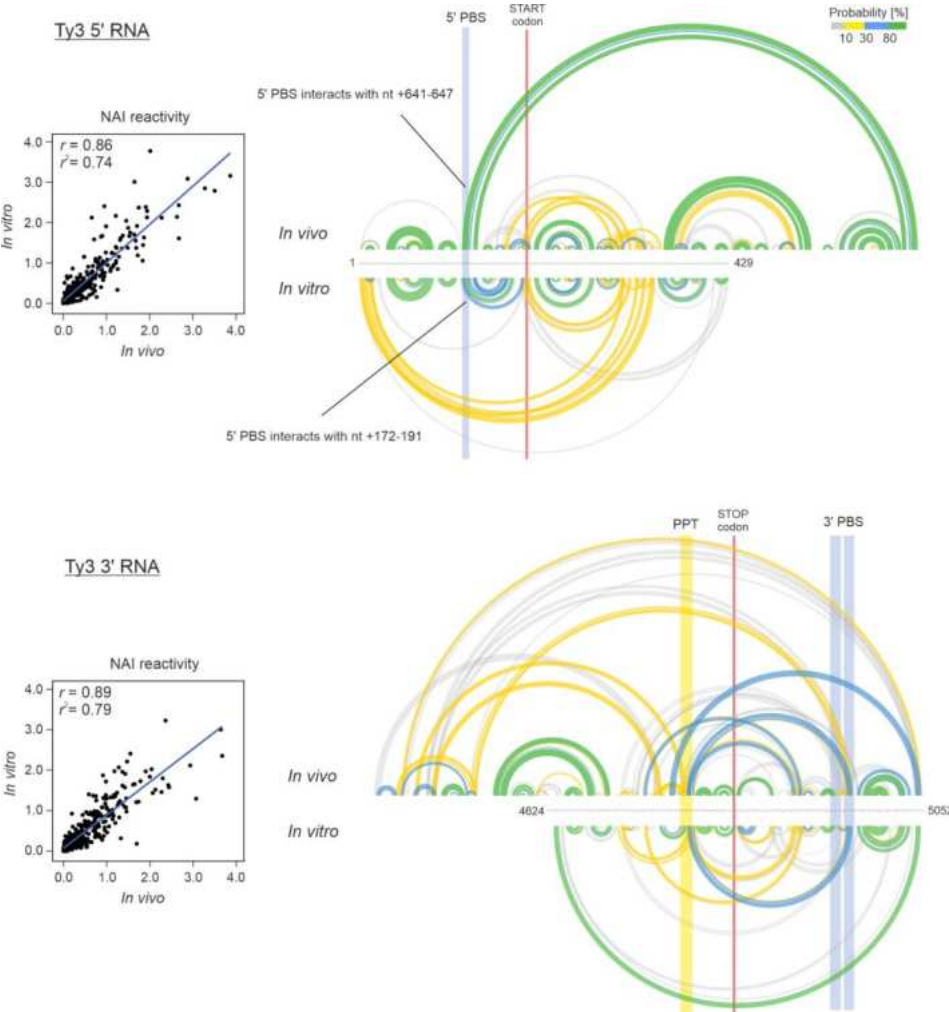


Figure S7.

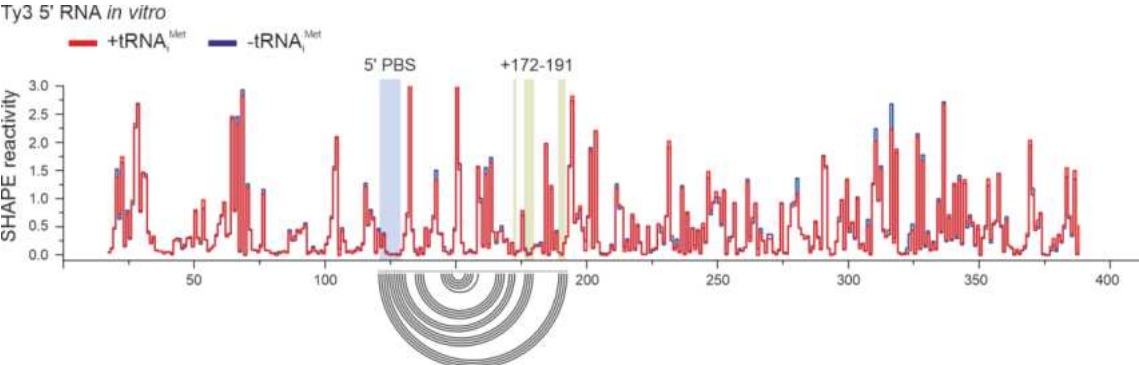


Figure S8.

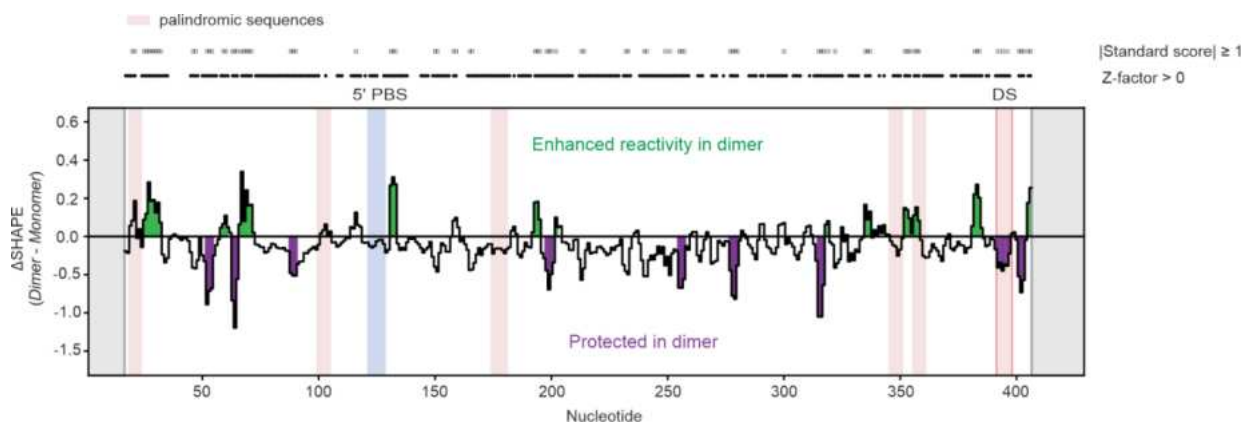


Figure S9.

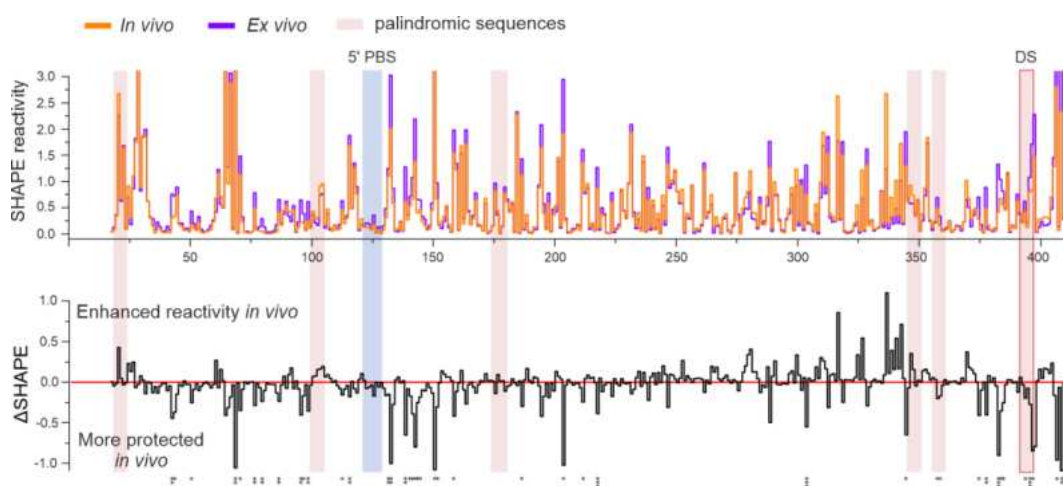
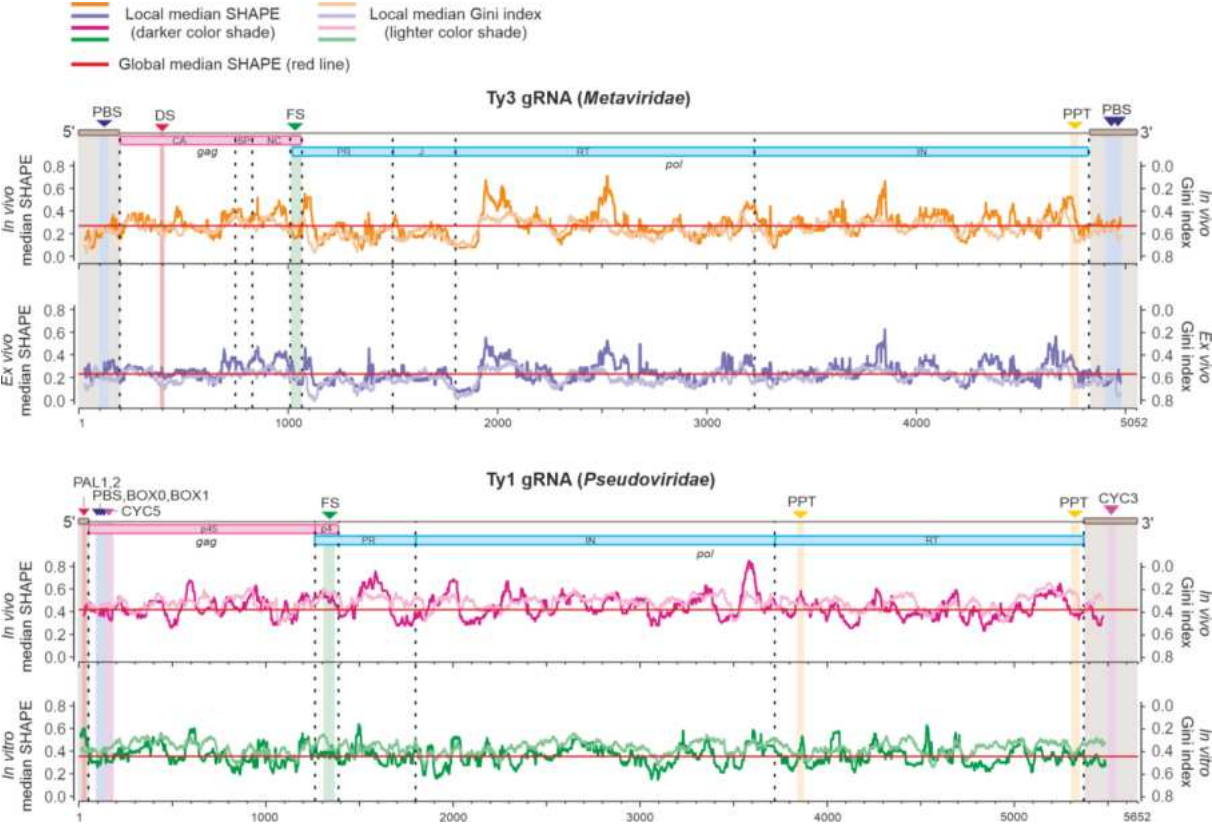


Figure S10.



Załącznik nr 3 do rozprawy doktorskiej pt.

„Charakterystyka dynamiki strukturalnej i funkcjonalnej genomów RNA aktywnych retrotranspozonów LTR”

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

Poznań, 26.06.2024

**Oświadczenie doktoranta dotyczące jego udziału w powstawaniu
pracy naukowej wchodzącej w skład rozprawy doktorskiej**

Angelika Andrzejewska*, Małgorzata Zawadzka*, Katarzyna Pachulska-Wieczorek[§]. **On the Way to Understanding the Interplay between the RNA Structure and Functions in Cells: A Genome-Wide Perspective.** *International Journal of Molecular Sciences*, 2020; 21(18), 6770.

Oświadczam, że mój udział w tworzeniu niniejszej pracy polegał na:

- współudziale w zgromadzeniu i analizie stosownej literatury,
- przygotowaniu pierwszej wersji rozdziału *The Correlation between mRNA Structure and Translation*,
- przygotowaniu wraz z M. Zawadzką rozdziału *The Relationship between RNA Structure and Proteins Binding* oraz Tabeli 1 i 2,
- przygotowaniu rysunków do manuskryptu,
- współuczestniczeniu w przygotowaniu pozostałych części manuskryptu, jego edycji oraz opracowaniu jego finalnej wersji.

Angelika
Andrzejewska - Romanowska

.....
Podpis

[§]Autor korespondencyjny, *równorzędny pierwszy autor

Poznań, 26.06.2024

Oświadczenie autora do korespondencji dotyczące udziału doktorantki w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

Angelika Andrzejewska*, Małgorzata Zawadzka*, Katarzyna Pachulska-Wieczorek[§]. **On the Way to Understanding the Interplay between the RNA Structure and Functions in Cells: A Genome-Wide Perspective.** *International Journal of Molecular Sciences*, 2020; 21(18), 6770.

Oświadczam, że wkład mgr inż. Angeliki Andrzejewskiej-Romanowskiej w powstanie wyżej wymienionej pracy naukowej polegał na:

- współudziale w zgromadzeniu i analizie stosownej literatury,
- przygotowaniu pierwszej wersji rozdziału *The Correlation between mRNA Structure and Translation*,
- przygotowaniu wraz z M. Zawadzką rozdziału *The Relationship between RNA Structure and Proteins Binding* oraz Tabeli 1 i 2,
- przygotowaniu rysunków do manuskryptu,
- współuczestniczeniu w przygotowaniu pozostałych części manuskryptu, jego edycji oraz opracowaniu jego finalnej wersji.



dr hab. Katarzyna Pachulska-Wieczorek

[§]Autor korespondencyjny, *równorzędny pierwszy autor

Poznań, 26.06.2024

**Oświadczenie doktoranta dotyczące jego udziału w powstawaniu
pracy naukowej wchodzącej w skład rozprawy doktorskiej**

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

Oświadczam, że mój udział w tworzeniu niniejszej pracy polegał na:

- współudziale w opracowaniu szczegółowego planu badań,
- przeprowadzeniu hodowli drożdży *S. paradoxus* szczep DG3412, w tym w warunkach zahamowanej translacji,
- optymalizacji oraz wykonaniu eksperymentów próbkowania strukturalnego gRNA Ty1 metodą SHAPE-CE w warunkach *in vivo*,
- współudziale w eksperymentach próbkowania strukturalnego RNA w komórkach *E. coli* metodą SHAPE-CE,
- analizie surowych danych otrzymanych z przeprowadzonych eksperymentów SHAPE-CE *in vivo*,
- otrzymaniu modeli struktury drugorzędowej gRNA Ty1 *in vivo*,
- współudziale w analizach statystycznych i interpretacji otrzymanych wyników,
- współudziale w przygotowaniu pierwszej wersji manuskryptu oraz jego finalnej wersji,
- przygotowaniu rysunków, *Supplementary Tables and Figures* oraz *Supplementary Dataset*.

Angelika
Andrzejewska - Romanowska

[§]Autor korespondencyjny

.....
Podpis


Poznań, 26.06.2004

Oświadczenie autora do korespondencji dotyczące udziału doktorantki w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

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

Oświadczam, że wkład mgr inż. Angeliki Andrzejewskiej-Romanowskiej w powstanie wyżej wymienionej pracy naukowej polegał na:

- współdziałanie w opracowaniu szczegółowego planu badań,
- przeprowadzeniu hodowli drożdży *S. paradoxus* szczep DG3412, w tym w warunkach zahamowanej translacji,
- optymalizacji oraz wykonaniu eksperymentów próbkowania strukturalnego gRNA Ty1 metodą SHAPE-CE w warunkach *in vivo*,
- współdziałanie w eksperymentach próbkowania strukturalnego RNA w komórkach *E.coli* metodą SHAPE-CE,
- analizie surowych danych z eksperymentów SHAPE-CE *in vivo*,
- otrzymaniu modeli struktury drugorzędowej gRNA Ty1 *in vivo*,
- współdziałanie w analizach statystycznych i interpretacji otrzymanych wyników,
- współdziałanie w przygotowaniu pierwszej wersji manuskryptu oraz jego finalnej wersji,
- przygotowaniu rysunków, *Supplementary Tables and Figures* oraz *Supplementary Dataset*.


dr hab. Katarzyna Pachulska-Wieczorek

Poznań, 26.06.2024

**Oświadczenie doktoranta dotyczące jego udziału w powstawaniu
pracy naukowej wchodzącej w skład rozprawy doktorskiej**

Angelika Andrzejewska-Romanowska, Julita Gumna, Ewa Tykwińska, Katarzyna Pachulska-Wieczorek[§]. **Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome.** *Nucleic Acids Research*, 2024, doi: 10.1093/nar/gkae494.

Oświadczam, że mój udział w tworzeniu niniejszej pracy polegał na:

- współudziale w opracowaniu planu badań,
- przeprowadzeniu transformacji drożdży *S. cerevisiae* szczep BY4741 plazmidem zawierającym sekwencję Ty3 oraz hodowli drożdży do mapowania strukturalnego,
- optymalizacji oraz wykonaniu eksperymentów SHAPE-MaP w warunkach *in vivo* oraz *ex vivo*,
- wykonaniu eksperymentów próbkowania strukturalnego transkryptów Ty3 *in vitro* (we współpracy z J. Gumną),
- analizie wszystkich surowych danych otrzymanych z przeprowadzonych eksperymentów SHAPE-MaP,
- otrzymaniu, wspartych danymi eksperymentalnymi, modeli struktury drugorzędowej gRNA Ty3 *in vivo* oraz *ex vivo*, a także transkryptów Ty3 *in vitro*,
- przeprowadzeniu analiz statystycznych,
- współudziale w interpretacji otrzymanych wyników z eksperymentów SHAPE-MaP,
- przygotowaniu pierwszej wersji manuskryptu, współudziale w jego późniejszej edycji oraz w przygotowaniu finalnej wersji,
- przygotowaniu rysunków oraz *Supplementary Tables and Figures*,
- pozyskaniu funduszy na badania w ramach projektu Preludium NCN.

Angelika
Andrzejewska-Romanowska

.....

Podpis

[§]Autor korespondencyjny

Poznań, 26.06.2024

Oświadczenie autora do korespondencji dotyczące udziału doktorantki w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

Angelika Andrzejewska-Romanowska, Julita Gumna, Ewa Tykwińska, Katarzyna Pachulska-Wieczorek[§]. **Mapping the structural landscape of the yeast Ty3 retrotransposon RNA genome.** *Nucleic Acids Research*, 2024, doi: 10.1093/nar/gkae494.

Oświadczam, że wkład mgr inż. Angeliki Andrzejewskiej-Romanowskiej w powstanie wyżej wymienionej pracy naukowej polegał na:

- opracowaniu wspólnie z autorką korespondencyjną planu badań,
- przeprowadzeniu transformacji *S. cerevisiae* (BY4741) pDLC201 oraz hodowli drożdży do mapowania strukturalnego,
- optymalizacji oraz wykonaniu eksperymentów SHAPE-MaP *in vivo* oraz *ex vivo*,
- wykonaniu eksperymentów próbkowania strukturalnego transkryptów Ty3 *in vitro* (we współpracy z J. Gumną),
- analizie surowych danych z eksperymentów SHAPE-MaP,
- otrzymaniu wspartych danymi eksperymentalnymi modeli struktury drugorzędowej RNA *in vivo*, *ex vivo* oraz *in vitro*,
- wykonaniu analiz statystycznych,
- współudziale w interpretacji otrzymanych wyników,
- przygotowaniu pierwszej wersji manuskryptu oraz współudziale w przygotowaniu jego finalnej wersji,
- przygotowaniu rysunków oraz *Supplementary Tables and Figures*,
- pozyskaniu funduszy na badania w ramach projektu Preludium NCN.



dr hab. Katarzyna Pachulska-Wieczorek

[§]Autor korespondencyjny