Instytut Chemii Bioorganicznej

Polskiej Akademii Nauk

- rozprawa doktorska -

Charakterystyka struktury genomowego RNA retrotranspozonu Ty1 w warunkach in vitro oraz na wybranych etapach replikacji w drożdżach

- Małgorzata Festina Zawadzka -

Praca wykonana

w Zakładzie Struktury i Funkcji Retrotranspozonów

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

Poznań 2022

Składam najserdeczniejsze podziękowania:

*Mojej promotorce dr hab. Katarzynie
Pachulskiej-Wieczorek*

*Za przekazaną wiedzę i doświadczenia, za poświęcony czas
oraz wsparcie i wiarę w moje możliwości*

mgr inż. Angelice Andrzejewskiej-Romanowskiej i dr Julicie Gumnej

*Za lata wspólnej pracy, cenne rady, udzieloną pomoc
oraz przyjazną atmosferę w pracy*

Kolegom i Koleżanką z ICHB PAN

Za wszystkie cenne rady, wskazówki i dobre słowa

Marcie, Kasi i Alicji

*Za przyjaźń, nieocenione wsparcie w trudnych chwilach
i wspólną radość z sukcesów*

Dziękuję i dedykuję tę pracę:

*Moim rodzicom, Dorocie i Markowi, oraz siostrze Katarzynie
Za zaszczepienie we mnie miłości do nauki i możliwość rozwoju,
oraz za nieustającą wiarę we mnie i ogromne wsparcie,
które pozwoliło mi przetrwać w chwilach zwątpienia.*

Spis treści

SPIS PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

[1]   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

(IF$_{2020}$ – 19.160; pkt MNiSW - 200)

[2]   **Małgorzata Zawadzka**, Angelika Andrzejewska-Romanowska, Julita Gumna, David J. Garfinkel,   Katarzyna Pachulska-Wieczorek

Cell compartment-specific folding of the Ty1 LTR-retrotransposon RNA genome.

Viruses, 2022; 14(9):2007

(IF$_{2021}$ – 5.818; pkt MNiSW - 100)

[3]   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):6670

(IF$_{2020}$ – 5.923; pkt MNiSW - 140)

*autorzy mieli równorzędny wkład w powstanie pracy

1. STRESZCZENIE

Ty1 jest retroelementem naturalnie występującym w genomie drożdży *Saccharomyces cerevisiae*. Pod wieloma względami przypomina on proste retrowirusy, jednak w czasie cyklu życiowego nie opuszcza komórki gospodarza. Ty1 replikuje poprzez cząsteczkę RNA o długości 5652 reszt nukleotydowych, pełniącą funkcję zarówno matrycy do syntezy białek Ty1, jak i genomu RNA (gRNA), który pakowany jest do cząsteczek wirusopodobnych (VLPs, ang. *virus like particles*), gdzie ulega odwrotnej transkrypcji do cDNA, który może integrować do chromosomalnego DNA gospodarza. W odróżnieniu od gRNA wirusów dla których coraz częściej publikowane są kompletne struktury *in vitro, in virio, ex virio,* a także *in vivo*, bardzo mało wiadomo na temat struktury gRNA retrotranspozonów LTR. Zupełnie niezbadany pozostaje również temat zmian strukturalnych gRNA wirusów i retrotranspozonów, zachodzących podczas ich replikacji w poszczególnych kompartmentach komórki.

Badania podjęte w ramach niniejszej rozprawy doktorskiej umożliwiły poznanie drugorzędowej struktury Ty1 gRNA w warunkach *in vitro*, a także na wybranych etapach replikacji Ty1 w drożdżach. Wykorzystując metodę SHAPE (ang. *selective 2'-hydroksyl acylation and primer extension*) otrzymałam model struktury całego gRNA retrotranspozonu Ty1 w warunkach *in vitro* (**Andrzejewska, Zawadzka et al. *NAR,* 2021**). Wykazałam, że struktura ta różni się istotnie od struktury *in vivo*, jednocześnie pozostaje bardziej stabilna i strukturalnie homogenna. Następnie podjęłam się scharakteryzowania zmian strukturalnych Ty1 gRNA zachodzących w trakcie replikacji retrotranspozonu w komórce (**Zawadzka et al. *Viruses*, 2022**). Metodą SHAPE, otrzymałam po raz pierwszy, jądrową strukturę gRNA retrotranspozonu LTR. Na podstawie dogłębnej analizy porównawczej otrzymanych danych, ze strukturami Ty1 gRNA w cytoplazmie, *in virio* oraz in *vitro* wykazałam, że gRNA retrotranspozonu podlega licznym rearanżacjom strukturalnym, i jego architektura jest ściśle zależna od środowiska kompartmentu komórki, w którym się znajduje. Aby lepiej zrozumieć temat zwijania się cząsteczek RNA w różnych stanach biologicznych wraz z zespołem przygotowaliśmy obszerną pracę przeglądową podsumowującą najnowszą wiedzę na temat struktury RNA w komórkach i wzajemnej korelacji pomiędzy strukturą i funkcją RNA (**Andrzejewska, Zawadzka et al. *IJMS,* 2020**).

2. ABSTRACT

Ty1 is a retroelement that naturally occurs in the genome of the yeast *Saccharomyces cerevisiae.* It resembles simple retroviruses but replicates without leaving the host cells. Ty1 replicates via RNA intermediate, which is 5652nt long and fulfills a dual role: template for proteins synthesis and gRNA that is packed in dimeric form into virus like particles (VLPs) and reverse transcribed to a cDNA copy that later integrates into host DNA. Unlike viral gRNA for which complete structures *in vitro*, *in virio*, *ex virio* and also *in vivo* are increasingly presented, little is known about the gRNA structure of LTR retrotransposons. Also, it remains unexplored how the structure of viral and retrotransposon gRNA changes during replication in the cellular compartments.

The research undertaken in this doctoral dissertation allowed for discovering Ty1 gRNA secondary structure in the *in vitro* condition as well as on the particular steps of Ty1 replication in yeast cells. Using the SHAPE method, I provided the secondary structure model of the entire Ty1 gRNA under defined *in vitro* conditions (**Andrzejewska, Zawadzka et al. *NAR*, 2021**). I showed that this structure differs significantly from *in vivo* state and is more stable and structurally homogenous. I also characterized structural alterations of Ty1 gRNA during retrotransposon replication in the cell (**Zawadzka et al. *Viruses*, 2022**). Using SHAPE probing, I provided the first, nuclear structure of LTR retrotransposon gRNA. Through a detailed comparison with the Ty1 gRNA structures established in the cytoplasm, VLPs, and *in vitro,* I found that LTR retrotransposon gRNA undergoes many significant structural rearrangements, and its architecture is closely dependent on the environment of cellular compartments. To better understand RNA folding in distinct biological states, together with the team, we prepared a comprehensive review to summarize the up-to-date knowledge about the RNA structure in the cell and RNA structure to function correlation (**Andrzejewska, Zawadzka et al. *IJMS*,2020**).

3. WPROWADZENIE

3.1 Zarys historyczny oraz podział elementów transpozycyjnych

W 1950 roku światło ujrzał historyczny artykuł pt. „Pochodzenie i zachowanie mutowalnych loci w kukurydzy" autorstwa Barbary McClintock, przedstawiający zaskakujące wyniki jej długoletniej pracy nad mechanizmem powstawania i niestabilnym dziedziczeniem mozaikowatości nasion kukurydzy [4]. Rezultaty prowadzonych przez McClintock badań sugerowały istnienie elementów genetycznych zdolnych do autonomicznego przemieszczania się w genomie. Ta kontrowersyjna, jak na tamten okres, teoria została zaakceptowana przez szerokie grono naukowców dopiero w latach 70. ubiegłego wieku, po odkryciu mobilnych elementów genetycznych najpierw u bakteriofagów i samych bakterii, a później także u bardziej złożonych organizmów jak *Drosophila* [5-7]. W 1983 roku Barbara McClintock za swoje odkrycie sprzed ponad dwóch dekad, została uhonorowana zasłużoną nagrodą Nobla w dziedzinie fizjologii i medycyny. Dzięki trwającym od tamtej pory intensywnym badaniom nad autonomicznymi elementami genetycznymi, dziś wiadomo że są one istotnym składnikiem genomów niemal wszystkich żyjących obecnie gatunków, począwszy od wirusów, przez prokariotyczne i eukariotyczne organizmy jednokomórkowe, na ssakach w tym człowieku kończąc [8-10]. Z czasem poznano także dwa główne mechanizmy propagacji elementów transpozycyjnych (TE; ang. *transposable elements*). Transpozony DNA (klasa II TE), pierwotnie odkryte przez McClintock w genomie kukurydzy, propagują niereplikacyjnym mechanizmem typu „wytnij-wklej". Wykorzystują przy tym kodowany przez siebie enzym - transpozazę rozpoznającą i przecinającą dwuniciowe DNA w obrębie sekwencji TIR (ang. *terminal inverted repeats*) [11]. Dużo bardziej złożoną grupę mobilnych elementów genetycznych stanowią występujące tylko w genomach eukariontów – retrotranspozony (klasa I TE) propagujące metodą „kopiuj-wklej". Ze względu na ich organizację strukturalną wyróżniono retrotranspozony typu LTR, zawierające długie, terminalne powtórzenia flankujące ich genom (ang. *long terminal repeats*) oraz pozbawione tych powtórzeń retrotranspozony non-LTR do których należą m.in. LINE (ang. *long interspersed nuclear element*) oraz nieautonomiczne elementy SINE (ang. *short interspersed nuclear element*) [12, 13].

3.2 Ogólna charakterystyka i podział retrotranspozonów LTR

Retrotranspozony LTR reprezentowane są przez trzy główne rodziny: Ty1/Copia (*Pseudoviridae*), Ty3/Gypsy (*Metaviridae*) oraz BEL/Pao (*Belpaoviridae*) należące do rzędu (*Ortervirales*) klasyfikującego wszystkie wirusy ssRNA-RT [14]. Pod względem organizacji strukturalnej genomu, jak i mechanizmu propagacji wykazują one istotne podobieństwa do infekcyjnych retrowirusów (*Retroviridae*), należących do tego samego rzędu taksonomicznego (Rys.1) [14]. Oprócz posiadania długich terminalnych powtórzeń, wspólna dla wszystkich wymienionych rodzin jest obecność częściowo nakładających się ramek odczytu dla dwóch genów *gag* oraz *pol*. Gen *gag* koduje białko strukturalne o tej samej nazwie (Gag), odpowiedzialne za formowanie wirionów u wirusów lub cząsteczek wirusopodobnych VLPs u retrotranspozonów. Produktami genu *pol* są natomiast białka enzymatyczne niezbędne dla zachodzenia retrotranspozycji: proteaza (PR), integraza (IN) oraz odwrotna transkryptaza z domeną RNazy H (RT/RH), które w zależności od rodziny, kodowane są w innej kolejności. Istotną cechą odróżniającą retrowirusy i retrotranspozony jest posiadanie przez te pierwsze genu *env* kodującego glikoproteiny otoczki lipidowej, umożliwiające retrowirusom fuzje z błoną komórkową gospodarza [15]. Retrotranspozony LTR w większości nie zawierają funkcjonalnego genu *env* bądź zawierają jego szczątkowe fragmenty i ich cykl życiowy zachodzi wyłącznie w obrębie jednej komórki [16].
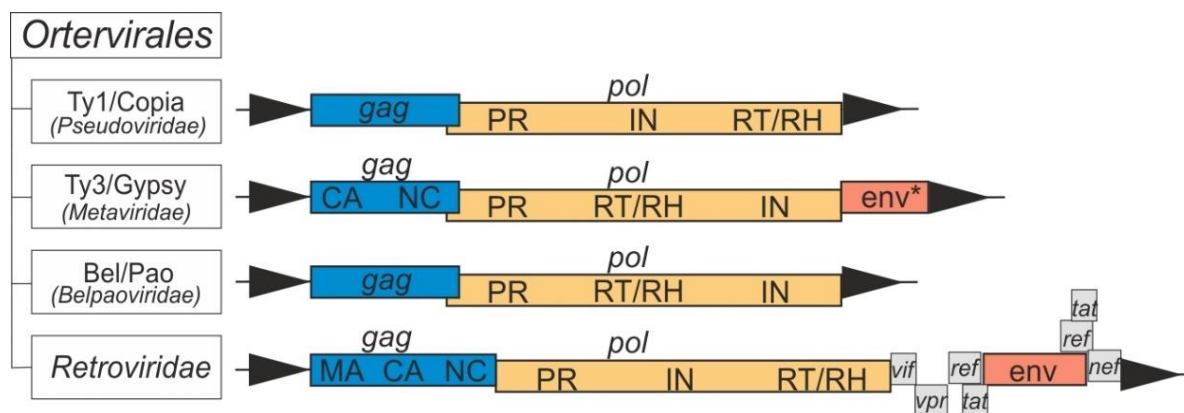
Rysunek 1. Uproszczony schemat organizacji genetycznej retroelementów należących do rzędu *Ortervirales*. We wszystkich rodzinach występują rejony LTR zaznaczone czarnymi trójkątami. Niektóre retrotranspozony z rodziny *Metaviridae* mogą zawierać gen podobny do *env* (ang. *env-like*), przedstawiony na rycinie jako env*. Funkcjonalny gen *env* obecny jest jedynie w rodzinie *Retroviridae*, tak samo jak dodatkowe geny regulatorowe: *vif*, *vpr*, *ref*, *tat* i *nef*.

Fundamentalną cechą łączącą wszystkie retroelementy LTR jest fakt, że replikują one poprzez cząsteczkę RNA [17]. Wewnątrz wirionu lub VLPs, RNA w wieloetapowej reakcji odwrotnej transkrypcji kopiowany jest do dwuniciowej cząsteczki cDNA, przy wykorzystaniu komórkowego tRNA jako startera [18]. Kodowana przez retroelement integraza rozpoznaje i wiąże terminalne powtórzenia na obu końcach nowopowstałej cząsteczki cDNA. Kompleks IN:cDNA transportowany jest następnie do jądra komórki gospodarza gdzie IN przeprowadza cięcie chromosomalnego DNA, a następnie liguje do niego kopię cDNA retroelementu [19]. Procesy te różnią się jednak w szczegółach u różnych retroelementów.

Jednym z pierwszych badanych retrotranspozonów był element Ty1, naturalnie występujący w genomach m.in. drożdży *Saccharomyces cerevisiae,* muszki *Drosophila melanogaster,* a także wielu gatunków roślin [17]. Obecnie należy on do najlepiej scharakteryzowanych retrotranspozonów LTR, pozostającym niezwykle cennym źródłem wiedzy na temat zarówno funkcjonalnych, jak i strukturalnych aspektów retrotranspozycji.

3.3 Charakterystyka oraz cykl życiowy retrotranspozonu Ty1

W genomie referencyjnego szczepu *S. cerevisiae* S288c znajduje się ok. 313 insercji pochodzących od elementu Ty1, z czego większość stanowią jego pozostałości w formie tzw. solo-LTR. Zidentyfikowano jednak 32 insercje pełnej długości, na które oprócz Ty1 składają się także hybrydy Ty1/Ty2 oraz ancestralny element Ty1' [20, 21]. Pełnej długości insercja Ty1 w genomie drożdży liczy 5918 par zasad i na obu swoich końcach zawiera powtórzenia LTR o długości 334 par zasad każde. Fragmenty LTR Ty1, analogicznie do LTR wirusowych, zawierają 3 domeny: U3 (ang. *unique 3' terminus*), U5 (ang. *unique 5' terminus*) i R (ang. *repeat*) o długości odpowiednio 240nt, 38nt i 56nt (Rys.2) . Sekwencje U3 oraz U5 są unikalne dla końców 3' i 5', natomiast sekwencja R jest taka sama na obu końcach genomu. Pomiędzy sekwencjami LTR znajdują się dwie częściowo nakładające się ramki odczytu *gag* oraz *pol* (Rys.2) [22]. Cykl replikacyjny Ty1 (Rys. 3) rozpoczyna się w jądrze komórkowym, gdzie polimeraza RNA II przeprowadza transkrypcję Ty1 zaczynając dokładnie od 1. nukleotydu powtórzenia R z 5'LTR, a kończąc na ostatnim nukleotydzie powtórzenia R z 3'LTR (Rys. 2). Powstałe transkrypty, nie podlegają splicingowi i tylko ok. 15% z nich jest poliadenylowanych [23]. Mimo to, RNA Ty1 cechują się bardzo wysoką

stabilnością i stanowią nawet do 0,8% wszystkich RNA w komórce [24]. Transkrypty Ty1 opuszczają jądro komórkowe w szlaku wykorzystującym receptor Mex67p, a w cytoplazmie kierowane są do specyficznych miejsc nazywanych ciałkami T (ang. *T bodies*) lub retrosomami, będącymi miejscem akumulacji Ty1 RNA i jego białek przed składaniem VLP [25, 26]. W wyniku translacji Ty1 RNA, powstają dwa produkty białkowe: p49-Gag oraz p199-Gag-Pol (Rys. 2).
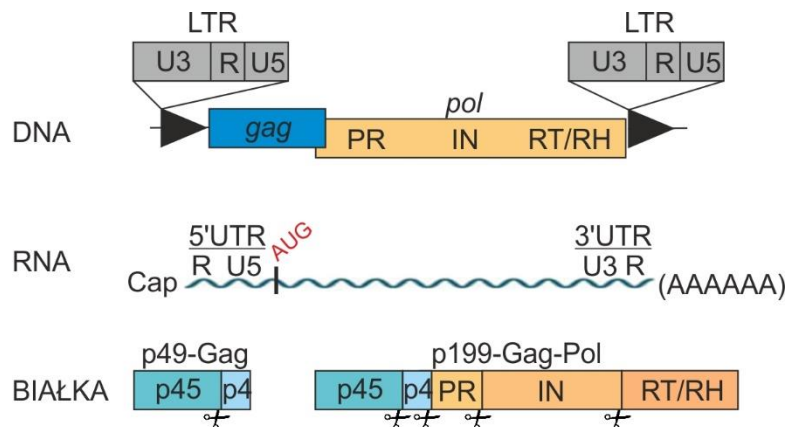
Rysunek 2. Drożdżowy retrotranspozon Ty1. DNA zawiera powtórzenia LTR flankujące ramki odczytu dla genów *gag* i *pol*. Transkrypty Ty1 zawierają na końcu 5' czapeczkę, a część z nich może zawierać także ogon poli(A) na końcu 3'. Na obu końcach Ty1 RNA znajdują się rejony nieulegające translacji UTR (ang. *untranslated region*). W wyniku translacji powstają dwie poliproteiny p49-Gag oraz p199-Gag-Pol, ulegające enzymatycznemu cięciu przez proteazę Ty1 do dojrzałych białek strukturalnych i enzymatycznych .

Gag-Pol jest efektem zaprogramowanego przesunięcia ramki odczytu, do którego dochodzi na siedmionukleotydowej sekwencji CUU-AGG-C [27]. Ty1 wykorzystuje przy tym mechanizm polegający na zapauzowaniu rybosomu, gdy w jego miejscu A znajdzie się kodon AGG, rozpoznawany przez nielicznie występujący w komórce, dekodujący go tRNA^{CUU} [28]. Pauza umożliwia znajdującemu się w miejscu P tRNA^{UAG}, rozpoznającego kodon CUU, przesunięcie się o +1 nukleotyd na kodon UUA. Proces ten zachodzi jednak z niską wydajnością, a stosunek białek Gag do Gag-Pol wynosi ok. 20:1 [29]. W retrosomach, p49-Gag i p199-Gag-Pol składane są w przypominające wirusowy nukleokapsyd cząstki wirusopodobne do których pakowane są po dwie cząsteczki gRNA Ty1 oraz komórkowe tRNA$_i^{met}$ [22]. W niedojrzałych VLPs, w pierwszej kolejności PR ulega autokatalitycznego wycięciu z poliproteiny p199-Gag-Pol, a następnie katalizuje szereg dodatkowych cięć, skutkujących powstaniem dojrzałych produktów białkowych: p20-PR, p71-IN, p63-RT/RH,

p45-Gag, a także p4-Gag o nieznanej funkcji (Rys. 2) [30]. W dojrzałym VLP gRNA Ty1 ulega odwrotnej transkrypcji. Kopia cDNA Ty1 w kompleksie preintegracyjnym, tworzonym razem z p71-IN, importowana jest do jądra komórkowego, gdzie ulega integracji do chromosomalnego DNA gospodarza, w pobliżu genów transkrybowanych przez polimerazę RNA III [31]. Pełna retrotranspozycja zachodzi jednak niezmiernie rzadko, z szacowaną częstotliwością na poziomie $10^{-5} – 10^{-7}$ na element w pokoleniu [32].
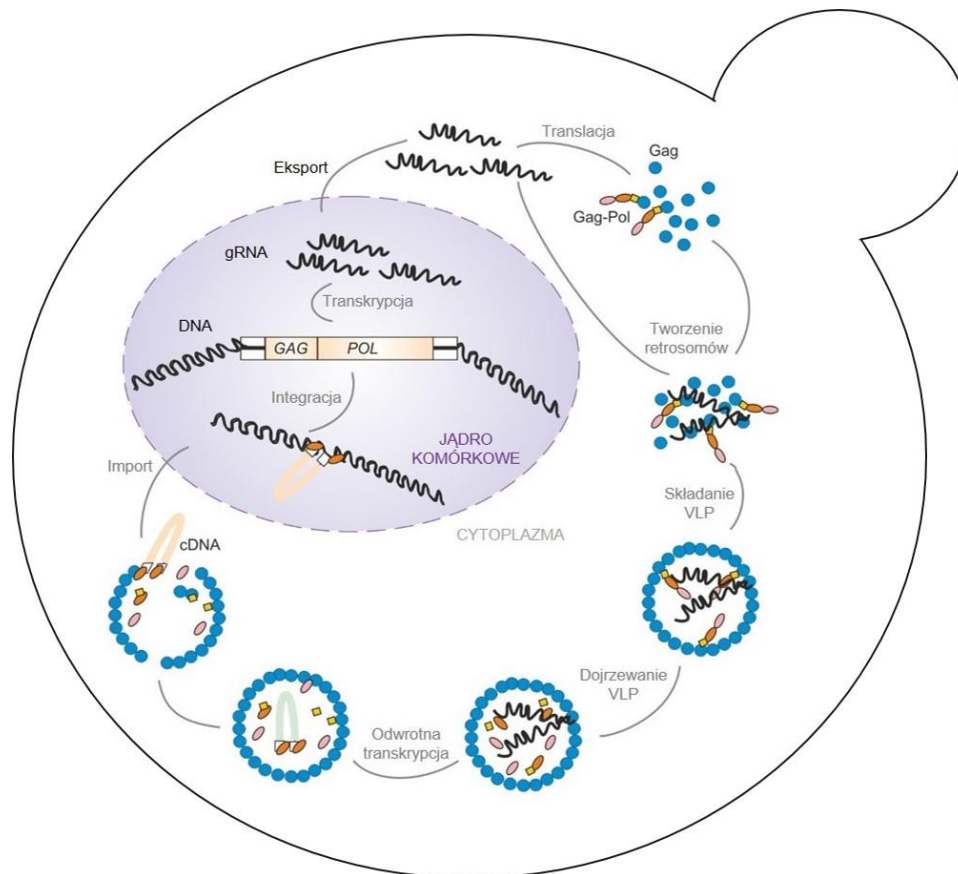
Rysunek 3. Schemat cyklu replikacyjnego retrotranspozonu Ty1. Transkrypcję Ty1 przeprowadza komórkowa polimeraza RNA II. Po eksporcie z jądra transkrypty Ty1 służą jako matryca do syntezy białek Gag i Gag-Pol. W cytoplazmie akumulują one wraz z gRNA Ty1 tworząc retrosomy, będące miejscem składania VLP. W dojrzałym VLP zachodzi odwrotna transkrypcja gRNA, a powstałe cDNA wraz z integrazą importowane jest do jądra komórkowego, gdzie ulega integracji do chromosomalnego DNA gospodarza.

3.4 Rola białka Gag w wybranych aspektach retrotranspozycji Ty1

Pomimo niskiej homologii sekwencji, Gag Ty1 jest funkcjonalnie podobne do retrowirusowych poliprotein Gag [33]. Pełni ono rolę głównego komponentu strukturalnego VLPs, jak również aktywnie uczestniczy w regulacji procesu replikacji Ty1. Wykazano, że bezpośrednie odziaływanie Gag z gRNA Ty1 jest kluczowe dla zachodzenia retrotranspozycji [26, 34]. Mutacje w wiążącej RNA C-końcowej domenie Gag istotnie zakłócały formowanie się retrosomów i składanie VLPs [26, 34]. Co więcej, Gag odgrywa istotną rolę w transporcie jądrowym i stabilności cytoplazmatycznej gRNA Ty1 [35]. Pod nieobecność białka Gag, transkrypty Ty1 akumulują w jądrze komórkowym, a ich stabilność cytoplazmatyczna istotnie spada i retrosomy nie są wykrywane. Dostarczenie Gag w *trans* skutkowało przywróceniem eksportu RNA Ty1 do cytoplazmy i wydajnym formowaniem się retrosomów [35]. Gag Ty1 nie zawiera sygnału lokalizacji jądrowej ale lokalizuje się w sąsiedztwie otoczki jądrowej wychwytując i stabilizując transkrypt Ty1 [25, 35]. Sugeruje się, że oddziaływanie z Gag odpowiada za długi czas życia cząsteczek RNA Ty1 [35]. Ponadto, Gag Ty1 wykazuje aktywność opiekuńczą względem do kwasów nukleinowych. Na podstawie badań *in vitro* zauważono, że wiąże się on do gRNA Ty1 i promuje kluczowe dla retrotranspozycji interakcje RNA:RNA – dimeryzację i cyklizację gRNA Ty1, a także przyłączanie do niego starterowego $tRNA_i^{met}$ [36-39]. Brakuje jednak danych potwierdzających zaangażowanie białka Gag Ty1 w te odziaływania *in vivo*.

3.5 Funkcja i struktura Ty1 gRNA

Pełnej długości Ty1 RNA pełni podwójną rolę w retrotranspozycji: matrycy do syntezy białek retrotranspozonu oraz genomu RNA pakowanego do cząstek wirusopodobnych. Genomowy RNA Ty1 zawiera liczne sekwencje *cis*-regulatorowe ulokowane głównie na końcach 5' i 3' [37, 40, 41]. Ich obecność umożliwia zachodzenie ważnych dla retrotranspozycji Ty1 wewnątrz- i między- cząsteczkowych odziaływań RNA:RNA (Rys. 4). Analogicznie do gRNA infekcyjnych retrowirusów, gRNA Ty1 pakowany jest do VLPs w formie dimerycznej [42]. Na przestrzeni lat proponowano kilka różnych modeli powstawania dimeru Ty1 gRNA [36, 43], jednak najnowsze badania przeprowadzone m. in. w naszej grupie wskazują, że dimer ten utrzymywany jest przez

międzycząsteczkowe oddziaływania krótkich sekwencji palindromowych z końca 5' gRNA Ty1 [39, 44]. W rejonie 5' zlokalizowana jest także 10-nukleotydowa sekwencja PBS (ang. *Primer Binding Site*) komplementarna do ramienia akceptorowego komórkowego tRNA$_i^{met}$ pełniącego funkcję startera w odwrotnej transkrypcji [45]. W wiązanie tRNA$_i^{met}$ zaangażowane są także dodatkowe sekwencje zlokalizowane w pobliżu PBS: BOX0 oraz BOX1, komplementarne odpowiednio do ramion TΨC oraz DHU, stabilizujące wiązanie tRNA$_i^{met}$ do Ty1 gRNA i wspomagające jego pakowanie do VLPs [46]. W sąsiedztwie sekwencji wiążących starter znajduje się 14-nukleotydowa sekwencja CYC5 oddziałująca z komplementarną sekwencją CYC3 umiejscowioną na końcu 3' Ty1 gRNA. Oddziaływanie CYC5/CYC3 umożliwia cyklizacje gRNA, wspomagając tym samym transfer nici(-) cDNA w trakcie odwrotnej transkrypcji [37].

Element Ty1 jest pierwszym retrotranspozonem LTR dla którego podjęto próby ustalenia struktury jego genomowego RNA. W 2013 roku na podstawie danych strukturalnych otrzymanych metodą SHAPE (dokładny opis metody przedstawiono w podrozdziale 5.6) zaproponowano model struktury drugorzędowej rejonu 5' Ty1 gRNA (+1 – 1482) upakowanego do cząstek wirusopodobnych (Rys. 4), a także w warunkach *ex virio* oraz *in vitro* [44]. Badania te wykazały, że rejon ten jest wysoce ustrukturyzowany i zawiera szereg złożonych motywów spinkowych o różnej długości trzonów i pętli. Zauważono również, że koniec 5' Ty1 gRNA jest stabilizowany przez odziaływania dalekiego zasięgu, tworzące motyw pseudowęzła typu H, zawiązanego pomiędzy resztami nukleotydowymi, 1-7 a 264-270 (trzon 1) oraz 256-262 a 319-325 (trzon 2). Mutacje zaburzające tworzenie pseudowęzła istotnie hamowały proces retrotranspozycji, głównie na etapie inicjacji odwrotnej transkrypcji, potwierdzając jego istotność funkcjonalną [43, 47]. Na podstawie różnic w reaktywności SHAPE pomiędzy dimerem gRNA w VLPs a monomerem *in vitro* zasugerowano także, że w dimeryzację Ty1 gRNA zaangażowane są trzy palindromowe sekwencje - PAL1, PAL2 oraz PAL3. Co ważne, eksperymenty SHAPE w VLP pozwoliły także ustalić wzór reaktywności charakterystyczny dla Ty1 gRNA w formie cyklicznej i z przyłączonym do niego tRNA$_i^{met}$.

Kiedy rozpoczynałam swoje badania dostępnych było także kilka modeli strukturalnych *in vitro* dla fragmentu +1 – 560, reprezentującego część 5' minimalnego funkcjonalnego elementu Ty1 zdolnego do retrotranspozycji po dostarczeniu niezbędnych

białek in *trans* (tzw. mini-Ty1) [40]. Struktury te zarówno w formie monomerycznej [43, 47], jak i formie indukowanego przez Gag dimeru [39] w wysokim stopniu przypominają analogiczny rejon Ty1 gRNA w VLPs i wszystkie zawierają motyw pseudowęzła. Jednak struktura kompletnego gRNA retrotranspozonu Ty1, podobnie jak innych retrotranspozonów LTR, nie była poznana.
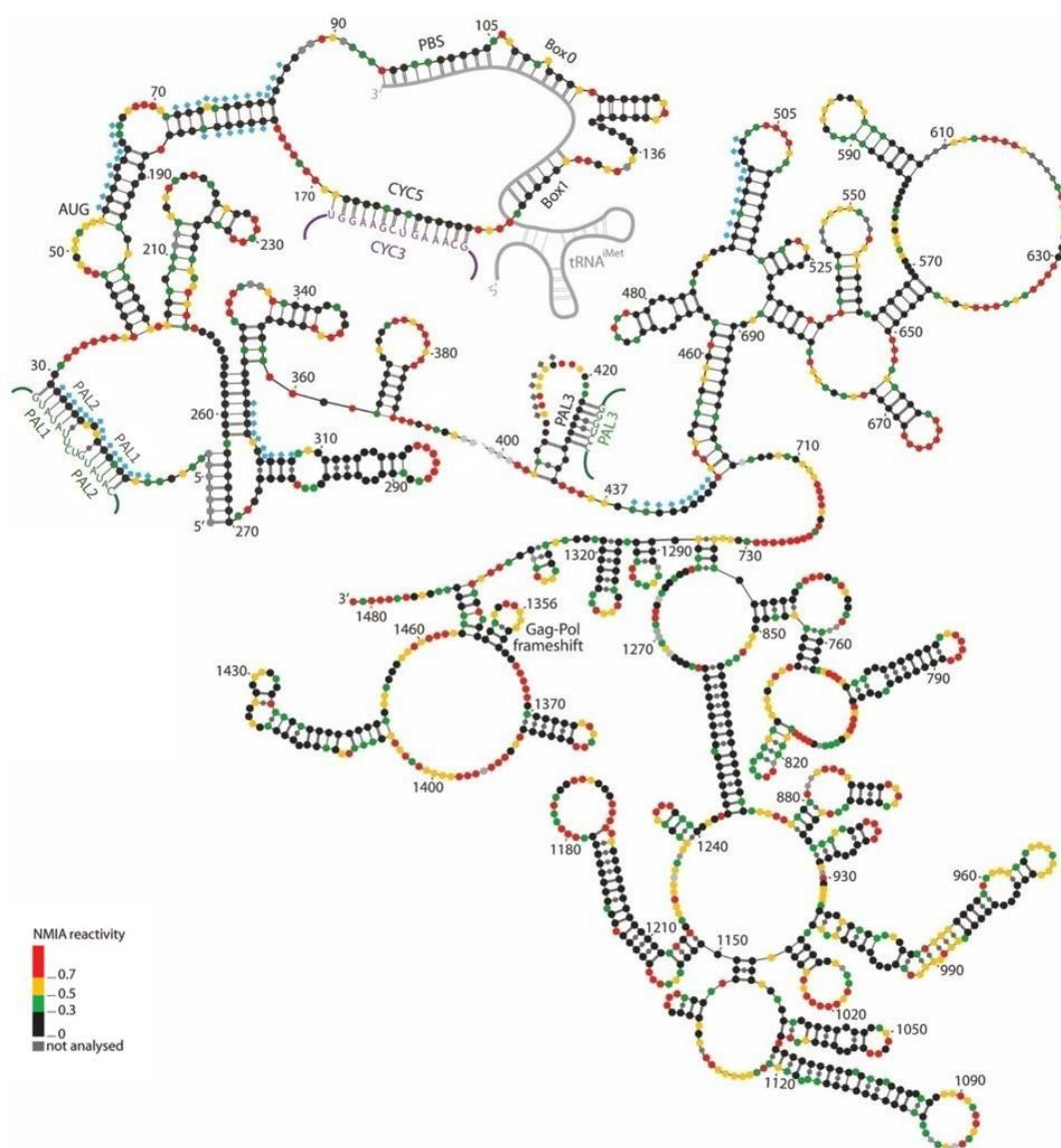
Rysunek 4. Model struktury drugorzędowej rejonu +1 – 1482 Ty1 gRNA wraz z reaktywnościami SHAPE w VLPs, na podstawie [44]. Na rysunku zaznaczono kluczowe dla retrotranspozycji odziaływania RNA:RNA.

3.6 Badanie struktury wirusowych RNA

W ciągu ostatnich dekad zaproponowano szereg podejść pozwalających z różną dokładnością badać strukturę cząsteczek RNA. Zastosowanie większości z nich jest jednak ograniczone do warunków *in vitro*. Przełomową metodą pozwalającą na wydajane mapowanie struktury cząsteczek RNA nie tylko w warunkach *in vitro*, ale także *in vivo* okazała się wciąż rozwijana, metoda SHAPE [48]. Wykorzystuje się w niej niskocząsteczkowe związki chemiczne, które kowalencyjnie modyfikują wolne grupy hydroksylowe przy węglu 2' rybozy niesparowanych lub dynamicznych konformacyjnie reszt nukleotydowych [49, 50]. Odczynniki SHAPE nie wykazują przy tym powinowactwa do zasady azotowej i z porównywalną wydajnością modyfikują wszystkie cztery rodzaje reszt nukleotydowych. Do detekcji miejsc modyfikacji wykorzystuje się odwrotną transkryptazę, która zatrzymuje się lub wprowadza mutację w miejscu powstałego w wyniku modyfikacji adduktu. Pulę różnej długości fragmentów cDNA analizuje się z wykorzystaniem elektroforezy kapilarnej (ang. *capillary electrophoresis)* (SHAPE-CE) lub metodami sekwencjonowania nowej generacji (NGS; ang. *next generation sequencing*) (SHAPE-Seq) [51, 52]. Indukowane adduktem mutacje w cDNA wykrywane są tylko przy życiu technik NGS - SHAPE-MaP [53]. Intensywne prace nad rozwojem metodologii SHAPE, zaowocowały powstaniem szeregu reagentów modyfikujących charakteryzujących się różnym czasem półtrwania i zdolnością do przenikania przez błony komórkowe, np. bezwodnik kwasu 1-metyloizatoinowego (NMIA), imidazolid kwasu 2-metylonikotynowego (NAI) i jego modyfikacja NAI-N3, bezwodnik 1-metylo-7-nitroizatowego (1M7), czy imidazolid kwasu 2-aminopirydyno-3-karboksylowego (2A3) [54-56]. Innym powszechnie stosowanym odczynnikiem znajdującym zastosowanie w badaniach struktur RNA *in vivo* jest siarczan dimetylu (DMS) [57]. W przeciwieństwie jednak do reagentów SHAPE wprowadza on modyfikacje na zasadzie azotowej i w fizjologicznym pH wykazuje silną preferencję w kierunku modyfikowania niesparowanych adenin i cytozyn [58]. Nie mniej jednak SHAPE i mapowanie DMS pozostają obecnie głównymi metodami stosowanymi do badań strukturalnych RNA w żywych komórkach umożliwiając badania struktur zarówno indywidualnych cząsteczek RNA, a dzięki połączeniu z technikami NGS, także całych transkryptomów [3].

Rozwój metod SHAPE i DMS umożliwił poznanie struktur długich cząsteczek wirusowych RNA w różnych stanach eksperymentalnych i biologicznych. Istotnym punktem w badaniach strukturalnych wirusowych RNA było wykorzystanie SHAPE-CE do zmapowania struktury pełnego gRNA HIV-1 wyizolowanego z wirionu [59]. W kolejnych latach SHAPE-CE został wykorzystany do ustalenia struktury *ex virio* całego gRNA wirusa Polio [60] oraz gRNA wirusa żółtaczki (HCV) w warunkach *in vitro* [61]. Rozwój technik SHAPE-MaP oraz SHAPE-Seq ułatwił poznanie struktur kilku kolejnych wirusowych RNA, np. Dengi, Ziki, HCV czy grypy typu A (IAV) nie tylko *in vitro, in virio* i *ex virio*, ale również w zainfekowanych komórkach (*in vivo*) [62-68]. Ponadto, w ciągu ostatnich dwóch lat kilka niezależnych grup badawczych, wykorzystując zarówno techniki SHAPE, jak i DMS, dostarczyło modele struktury dla gRNA SARS-CoV-2, najdłuższego znanego wirusowego RNA (~30000 nt), w zainfekowanych komórkach [69-72], *in virio* [73] i *in vitro* [71].

Powyższe badania wskazują, że wirusowe RNA przyjmują złożoną strukturę i efektywne wykorzystanie kompaktowego genomu RNA możliwe jest, nie tylko dzięki informacji genetycznej zakodowanej w sekwencji nukleotydowej, ale również w strukturze wyższego rzędu, jaką to RNA przyjmuje [74]. Wysoce ustrukturyzowane domeny wirusowych RNA są często bogate w kluczowe sekwencje regulatorowe [59]. Jednocześnie zauważono, że RNA wirusów cechuje mozaikowa budowa, a domeny charakteryzujące się wysokim stopniem ustrukturyzowania są przeplatane rejonami o bardziej dynamicznej strukturze. Wykazano także istotne różnice w zwijaniu wirusowych RNA w różnych stanach eksperymentalnych. Dla przykładu, RNA wirusa Ziki, a także mRNA IAV w zainfekowanych komórkach przyjmują inną strukturę niż *in virio*, *ex virio* oraz *in vitro*, jedocześnie wykazując mniejszy stopień ustrukturyzowania i większą heterogenność strukturalną [64, 68]. Dla kontrastu, RNA wirusa HCV w komórce zachowuje poziom ustrukturyzowania charakterystyczny dla stanu *in vitro* [66]. Przyczyny tych różnic nie zostały jednak jednoznacznie zidentyfikowane, gdyż wciąż jeszcze niewiele jest badań kompleksowo porównujących zwijanie się cząsteczek RNA w różnych warunkach eksperymentalnych. Nie wiadomo także czy, a jeśli tak to jak zmienia się struktura wirusowych RNA w trakcie podróży przez kompartmenty komórki. Wiedza ta jest limitowana również dla komórkowych RNA. Pojedyncze badania poruszające problem zmian strukturalnych RNA w trakcie jego cyklu życiowego dostępne są dla ssaczych

komórek HEK293 i mECSs [75] oraz rośliny *Arabidopsis thaliana* [76, 77]. W komórkach ssaczych zauważono tylko niewielkie i lokalne zmiany strukturalne pomiędzy kompartmentami komórki wynikające z wiązania się RBP oraz modyfikacji RNA, sugerując, że struktura RNA przyjmowana w trakcie transkrypcji nie zmienia się w trakcie cyklu życiowego RNA, łącząc i regulując wszystkie procesy w które dany RNA jest zaangażowany. Natomiast, w *A. thaliana* zauważono istotne różnice strukturalne pomiędzy jądrem a cytoplazmą w obrębie kodonów stop i start, oraz rejonów kluczowych dla potranskrypcyjnej obróbki RNA.

4. CEL NAUKOWY PRZEDSTAWIONEGO CYKLU PRAC

Głównym celem niniejszej rozprawy doktorskiej było pozyskanie nowej wiedzy na temat molekularnych aspektów procesu replikacji retrotranspozonów LTR, poprzez scharakteryzowanie zmian strukturalnych zachodzących w gRNA retrotranspozonu Ty1 w trakcie replikacji i podróży przez kompartmenty komórki. Tym samym, moim celem było również poszerzenie wiedzy na temat procesów zwijania się cząsteczek RNA w różnych stanach eksperymentalnych i biologicznych.

W pracy **Andrzejewska, Zawadzka et al.,** *NAR*, **2021 – Publikacja [1]** moim celem było zbadanie struktury Ty1 gRNA w warunkach *in vitro* oraz szczegółowe porównywanie uzyskanego modelu struktury drugorzędowej do modelu struktury Ty1 gRNA *in vivo*.

Badania zaprezentowane w pracy **Zawadzka et al.,** *Viruses*, **2022 – Publikacja [2]** ukierunkowane były na ustalenie struktury gRNA Ty1 w jądrze komórek drożdżowych, a następnie jej szczegółowe porównanie do struktury w cytoplazmie, VLP oraz *in vitro*, celem poznania zmian strukturalnych zachodzących trakcie cyklu replikacyjnego retrotranspozonu LTR w komórce.

Celem pracy przeglądowej **Andrzejewska, Zawadzka et al.,** *IJMS*, **2020 – Publikacja [3]** było zgłębienie i usystematyzowanie aktualnej wiedzy na temat wpływu czynników komórkowych na zwijanie się struktur RNA, a także wzajemnej regulacji pomiędzy strukturą, a funkcją RNA w komórkach różnych organizmów.

5. SKRÓTOWY OPIS PRAC NAUKOWYCH WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

5.1 In vivo structure of the Ty1 retrotransposon RNA genome

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

Cząsteczki RNA zaangażowane są w liczne mechanizmy komórkowe, a pełnione przez nie funkcje nierozerwalnie związane są z ich strukturą. Dokładne poznanie jak zwija się RNA niezbędne jest więc do pełnego zrozumienia roli RNA w procesach, w które jest on zaangażowany. W ciągu ostatnich lat coraz więcej badań, w tym nad wirusowymi RNA, wskazuje na istotne różnice strukturalne pomiędzy cząsteczkami RNA w warunkach *in vitro* oraz *in vivo*. Chociaż poziom tych różnic oraz ich przyczyna wciąż nie jest do końca jasna, to bezpośrednie porównanie danych otrzymanych w odrębnych kontekstach środowiskowych, umożliwia zaobserwowanie kluczowych dla funkcji RNA zmian strukturalnych wynikających z bezpośredniego odziaływania środowiska komórkowego m.in. wiązania ligandów i białek, czy aktywności rybosomów. Z tego powodu, w pierwszej pracy wchodzącej w skład mojej rozprawy doktorskiej wraz z zespołem podjęliśmy się scharakteryzowania różnic pomiędzy strukturą drugorzędową gRNA retrotranspozonu Ty1 zwijanego w ściśle zdefiniowanych warunkach *in vitro* oraz w natywnym środowisku komórkowym. Wykorzystując metodę SHAPE i zaawansowane narzędzia bioinformatyczne ustaliłam strukturę genomowego RNA retrotranspozonu Ty1 *in vitro*. Stosując analogiczne podejście Angelika Andrzejewska-Romanowska uzyskała dane strukturalne oraz przewidziała strukturę Ty1 gRNA w komórkach drożdży, gdzie główna frakcja Ty1 gRNA znajdowała się w cytoplazmie.

Przeprowadzone przeze mnie eksperymenty pozwoliły na uzyskanie wysokiej jakości danych SHAPE dla 98% nukleotydów Ty1 gRNA. Ustaliłam, że w zdefiniowanych warunkach *in vitro,* badany RNA jest bogaty w niereaktywne nukleotydy (reaktywność SHAPE < 0,4; udział 54%), które zwyczajowo biorą udział w odziaływaniach typu Watsona-Cricka. 35% nukleotydów wykazywało pośrednie reaktywności SHAPE (0,4 - 0,85), sugerujące ich zaangażowanie w mniej stabilne odziaływania, charakterystyczne dla

koegzystujących ze sobą konformerów RNA. Natomiast nukleotydy wykazujące wysoką reaktywność SHAPE (> 0,85), charakterystyczną dla rejonów jednoniciowych, stanowiły tylko 9% wszystkich nukleotydów w analizowanym RNA. Obserwowany przeze mnie rozkład reaktywności SHAPE *in vitro* różnił się od tego uzyskanego *in vivo*, a zależna od pozycji nukleotydów korelacja reaktywności pomiędzy tymi dwoma stanami była umiarkowana (*r* = 0,49) (Rys. 2B,F w [1]). *In vitro* globalne mediany reaktywności SHAPE (0,35) oraz entropii Shannona (0,04) były znacząco niższe niż w warunkach komórkowych (Rys. 2A;C w [1]). Ponadto, na podstawie analizy rozkładu entropii Shanonna wykazałam, że w warunkach *in vitro* dwa razy więcej reszt nukleotydowych zaangażowanych jest w tworzenie rejonów o stabilnej strukturze (Rys. 2H w [1]). Wyniki te sugerują, że Ty1 gRNA w ściśle zdefiniowanych warunkach *in vitro* jest bardziej stabilny i mniej heterogenny strukturalnie niż w warunkach *in vivo*.

W następnym etapie badań, uzyskane dane liczbowe SHAPE wykorzystałam do ustalenia wspartego eksperymentalnie modelu struktury drugorzędowej o najniższej energii swobodnej - MFE (ang. *minimum free energy*) Ty1 gRNA *in vitro* (Rys. 3E;F;G – w [1]). W tym celu użyłam oprogramowania SuperFold, dedykowanego modelowaniu długich cząsteczek RNA [78]. SuperFold wykorzystuje funkcje *Partition* oraz *Fold* zaimplementowane w programie RNAstructure [79] oraz reaktywności SHAPE, jako ograniczenia pseudoenergetyczne. Modelowanie struktury RNA odbywa się w przesuwającym się oknie (ang. *sliding window*), a konsensusowy model MFE struktury drugorzędowej badanego RNA przewidywany jest z uwzględnieniem par zasad o największym prawdopodobieństwie (> 99%), które występują w co najmniej połowie wszystkich okien zwijania. Na podstawie analizy otrzymanego modelu struktury MFE zauważyłam, że prawie 50% wszystkich nukleotydów Ty1 gRNA w warunkach *in vitro* zaangażowanych jest w tworzenie par zasad (Rys. 4E w [1]). Podobna zawartość sparowanych nukleotydów obserwowana była w strukturze *in vivo*, jednak na podstawie miar precyzji (PPV, ang. *positive predictive value*) oraz czułości (ang. *sensitivity*) ustaliłam, że prawie połowa par zasad przewidziana w strukturze *in vitro* jest inna niż w strukturze *in vivo* (Rys. 4A w [1]). Następnie, aby rozważyć pary zasad występujące w większości możliwych konformerów Ty1 gRNA *in vitro*, przeanalizowałam pary przewidziane z wysokim prawdopodobieństwem (> 80%). W strukturze MFE Ty1 gRNA *in vitro,* tego typu pary zasad

stanowiły aż 68% wszystkich przewidzianych par zasad, 30% więcej niż *in vivo* (Rys. 4E,F w [1]). Zauważyłam również, że tylko 39% wysoce prawdopodobnych par zasad przewidzianych *in vitro*, obecnych jest także w strukturze *in vivo* i wspólne dla obu stanów są głównie odziaływania krótkiego zasięgu (Rys. 4G w [1]). Powyższe wyniki pokazują, że analizowane struktury są inne i w warunkach *in vitro* Ty1 gRNA jest wysoce ustrukturyzowany, a liczba współistniejących konformerów strukturalnych jest niższa, niż w środowisku komórkowym.

Dostępne dane wskazują, że rejony RNA bogate w stabilne motywy strukturalne często zawierają istotne funkcjonalnie sekwencje regulatorowe [53]. Na podstawie analizy entropii par zasad (entropii Shanonna) i wartości reaktywności SHAPE, zidentyfikowałam 11 stabilnych strukturalnie rejonów w Ty1 gRNA *in vitro*, z czego 5 częściowo pokrywało się z rejonami silnie ustrukturyzowanymi zidentyfikowanymi *in vivo* (Rys. 3B;E w [1]). Najdłuższy rejon, wspólny dla obu stanów badanego RNA, znajdował się na końcu 5' i pokrywał się z lokalizacją kluczowych dla retrotranspozycji sekwencji regulatorowych jak PAL1, PAL2 oraz CYC5, zaangażowanych odpowiednio w dimeryzację i cyklizację genomu Ty1 RNA, czy PBS, BOX0 i BOX1, będących miejscami przyłączania $tRNA_i^{Met}$ – startera dla odwrotnej transkrypcji. Na podstawie porównania uzyskanego profilu reaktywności SHAPE, do tego otrzymanego w VLP [44], nie zaobserwowaliśmy jednak dimeryzacji i cyklizacji badanego RNA w warunkach *in vitro*. Nie zauważyliśmy także charakterystycznych zmian w reaktywności SHAPE przemawiających za tworzeniem się motywu pseudowęzła (Rys. 5 w [1]).

Podsumowując, dzięki swoim badaniom zaprezentowałam model struktury genomowego RNA retrotranspozonu Ty1 w warunkach *in vitro*. Zgodnie z dostępną mi wiedzą, na chwilę obecną jest to jedyna poznana struktura *in vitro* pełnej długości gRNA retrotranspozonu LTR. Na podstawie szczegółowego porównania uzyskanych danych do struktury *in vivo* wykazałam, że w warunkach *in vitro* gRNA Ty1 przyjmuje odmienną strukturę drugorzędową, a także jest ona bardziej stabilna i homogenna. Omówiona praca w istotny sposób wzbogaca wiedzę dotyczącą różnic w zwijaniu się RNA w odmiennych warunkach eksperymentalnych. Jest to szczególnie ważne zważywszy, że wciąż dostępnych jest niewiele tego typu analiz porównawczych dla innych wirusowych, a także komórkowych cząsteczek RNA.

5.2 Cell compartment-specific folding of Ty1 Long Terminal Repeat retrotransposon RNA genome

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

W komórce, cząsteczki RNA przechodzą przez złożony cykl życiowy obejmujący ich podróż z miejsca powstania w jądrze komórkowym do docelowego miejsca przeznaczenia. Przeważająca większość dostępnych struktur RNA *in vivo*, przedstawia jednak uśredniony obraz struktury danego RNA we wszystkich kompartmentach, a wiedza dotycząca zmian strukturalnych zachodzących podczas cyklu życiowego RNA i jego podróży w komórce jest wiąż znikoma. Skłoniło mnie to do zgłębienia tej tematyki i zbadania zmian strukturalnych zachodzących w Ty1 gRNA podczas jego replikacji w drożdżach. W tym celu wykorzystałam metodę SHAPE i ustaliłam drugorzędową strukturę Ty1 gRNA zakumulowanego w jądrze komórki drożdżowej. Następnie przeprowadziłam szczegółową analizę porównawczą jądrowej struktury Ty1 gRNA z otrzymanymi wcześniej strukturami tego RNA w cytoplazmie, cząstce wirusopodobnej (VLP) oraz *in vitro* [1, 44].

Reagenty SHAPE różnią się czasem półtrwania i zdolnością przenikania przez błony komórkowe, ponadto niektóre z nich mogą wykazywać pewne preferencje w modyfikowaniu konkretnych reszt nukleotydowych [55]. Wszystkie dotychczasowe badania strukturalne Ty1 RNA zostały przeprowadzone z użyciem NMIA jako odczynnika modyfikującego [1, 39, 43, 44, 47]. Zatem aby ograniczyć wprowadzanie zmiennych między eksperymentami oraz umożliwić najbardziej wiarygodną analizę otrzymanych wyników, zdecydowałam się na użycie NMIA także do próbkowania struktury jądrowego Ty1 gRNA. Wbrew wcześniejszym doniesieniom literaturowym [54] w naszej poprzedniej pracy udowodniliśmy, że NMIA wydajnie modyfikuje RNA *in vivo* [1], jednak jego zdolność przenikania do jądra komórkowego nie była wcześniej wykazana. Dlatego, w pierwszym etapie badań porównałam efektywność NMIA oraz NAI (odczynnik wydajnie modyfikujący RNA w jądrze komórek ssaczych [80]) w mapowaniu U1 snRNA w *S. cerevisiae*. Dla obu użytych odczynników otrzymałam porównywalnie wysoki stosunek sygnału modyfikacji do tła, wykazując tym samym, że zarówno NMIA, jak i NAI, wydajnie modyfikują RNA w jądrze

komórek drożdżowych (Rys. 1A;B w [2]) . Do docelowego mapowania struktury jądrowego Ty1 gRNA wykorzystałam szczep drożdżowy *S. paradoxus* DG3408, dostarczony przez naszego współpracownika prof. Davida Garfinkel'a. W tym szczepie ekspresja Ty1 zachodzi wyłącznie z indukowanego galaktozą plazmidu pGfs, zawierającego w sekwencji Ty1 delecję jednej adeniny w bezpośrednim sąsiedztwie kodonu „start" białka Gag (Rys. 1C w [2]) [35]. W wyniku tej mutacji powstaje pełnej długości transkrypt Ty1, natomiast nie powstaje funkcjonalne białko Gag, co silnie wpływa na lokalizację Ty1 RNA w komórce. Pod nieobecność Gag, Ty1 RNA akumuluje w jądrze komórkowym, a retrosomy i VLPs w cytoplazmie nie są wykrywane. Dodatkowo, Ty1fs zawiera przedwczesny kodon „stop" i w przypadku ewentualnego przedostania się do cytoplazmy jest on jest rozpoznawany i degradowany przez system kontroli RNA - NMD (ang. *Nonsense-Mediated Decay*) [35].

Wykorzystując mapowanie NMIA uzyskałam wysokiej jakości dane strukturalne dla 93% reszt nukleotydowych Ty1 RNA w jądrze komórkowym. Globalna mediana oraz rozkład reaktywności SHAPE jądrowego RNA porównywalne były do tych zaobserwowanych wcześniej w cytoplazmie, natomiast znacząco różniły się od tych charakterystycznych dla Ty1 RNA *in vitro* czy upakowanego do VLPs (Rys. 2A w [2]). Także procentowy rozkład reaktywności SHAPE w jądrze (11% nukleotydów wysoce reaktywnych; 43% średnioreaktywnych; 39% niereaktywnych) podobny był do tego obserwowanego w cytoplazmie, ale istotnie odbiegał od uzyskanego w warunkach *in vitro* (Rys. 1B w [2]). Sugeruje to, że niezależnie od kompartmentu, Ty1 gRNA w komórce jest mniej ustrukturyzowany niż po spakowaniu do VLPs czy w warunkach *in vitro*. Następnie obliczyłam zależną od pozycji nukleotydów korelację reaktywności SHAPE dla wszystkich analizowanych stanów RNA (Rys. 2C w [2]). Korelacja reaktywności pomiędzy RNA jądrowym a cytoplazmatycznym była umiarkowana ($r = 0,55$), a pomiędzy RNA jądrowym a tym z VLPs znacząco niższa ($r = 0,22$), sugerując zachodzenie istotnych rearanżacji strukturalnych po eksporcie Ty1 gRNA z jądra i podczas jego przechodzenia przez kolejne etapy cyklu życiowego. Przeprowadzone przez nas analizy wskazują, że różnice strukturalne w Ty1 RNA nie wynikają z aktywności rybosomów podczas translacji – brak wzrostu korelacji między stanem jądrowym i cytoplazmatycznym po zablokowaniu translacji. Nie zauważyłam także zwiększonej korelacji reaktywności SHAPE między stanem jądrowym i *in vitro*. Ponadto, na podstawie bezwzględnej różnicy reaktywności SHAPE pomiędzy

jądrem i cytoplazmą zidentyfikowałam jedenaście, rejonów o długości ponad 100nt wykazujących największą zmienność (Rys. 1E w [2]). Obserwacja ta przemawia za specyficznym dla kompartmentu zwijaniem się Ty1 RNA, ponieważ różnice wynikające z odmiennego odziaływania z białkami czy występowania modyfikacji RNA indukują raczej dużo mniejsze, bardziej lokalne zmiany [75].

Analogicznie do swoich poprzednich badań wykorzystałam oprogramowanie SuperFold i uzyskane dane SHAPE do predykcji modelu struktury MFE jądrowego Ty1 gRNA (Rys. 3A [2]). Analiza reaktywności w rejonach jedno- i dwuniciowych wykazała wysoką zgodność przewidzianego modelu z eksperymentalnie uzyskanymi danymi (Rys. 3B w [2]). W celu uwidocznienia zmian strukturalnych gRNA w trakcie replikacji Ty1, otrzymany model struktury jądrowej porównałam do modelu MFE struktury cytoplazmatycznej oraz modelu MFE *in vitro*. Wykazałam, że we wszystkich analizowanych modelach, podobna ilość nukleotydów zaangażowana jest w tworzenie par zasad (Rys. 3C w [2]). Nie zauważyłam znaczących różnic w medianie zasięgu sparowań oraz ilości parowań dalekiego zasięgu (> 100 nt) (Rys. 3D w [2]). Pomimo tych podobieństw, na podstawie parametru PPV i miary czułości ustaliłam, że około 40% par zasad przewidzianych w strukturze jądrowego Ty1 gRNA nie zostaje zachowanych po eksporcie RNA do cytoplazmy. Wykazałam, że pary zasad wspólne dla obu kompartmentów stanowią głównie odziaływania krótkiego zasięgu, natomiast oddziaływania dalekiego zasięgu zdają się być specyficzne dla danego kompartmentu (Rys. 3A;D w [2]). Natomiast w porównaniu do struktury *in vitro,* miary precyzji i czułości wyniosły ok 50% wskazując, że połowa przewidzianych w niej par zasad występuje również w strukturze jądrowej (Sup. Rys. 1 w [2]).

Następnie, w celu analizy całej puli możliwych konformerów Ty1 RNA w jądrze, obliczyłam prawdopodobieństwo występowania wszystkich możliwych par zasad, a następnie powtórzyłam analizę porównawczą uwzględniając tylko wysoce prawdopodobne pary zasad, występujące w większości możliwych konformerów. Na tej podstawie zidentyfikowałam 407 wysoce prawdopodobnych par zasad w Ty1 RNA w jądrze, co stanowi 31% wszystkich par zasad przewidzianych w strukturze MFE. Dla porównania struktura cytoplazmatyczna zawierała około 100, a *in vitro* ponad 535 więcej wysoce prawdopodobnych par zasad, co stanowiło odpowiednio 41% i 66% wszystkich sparowań w danych modelach MFE (Rys. 3F w [2]). Co ciekawe, zauważyłam, że Ty1 RNA w jądrze jest

zubożony o wysoce prawdopodobne odziaływania dalekiego zasięgu (> 100 nt) (Rys. 3E w [2]). W zgodzie z powyższymi wynikami, mediana entropii Shanonna w jądrze komórkowym była wyższa niż w cytoplazmie i w warunkach *in vitro*.

Aby lepiej zrozumieć zaobserwowaną heterogenność struktury jądrowego Ty1 gRNA, na podstawie entropii Shanonna i wartości reaktywności SHAPE, wyznaczyłam rejony charakteryzujące się wysokim stopniem ustrukturyzowania. Zidentyfikowałam 11 takich rejonów, z których większość, przynajmniej częściowo, pokrywała się wysoko ustrukturyzowanymi rejonami zidentyfikowanymi wcześniej w cytoplazmie, a cztery z rejonami zidentyfikowanymi *in vitro* (Rys. 3H oraz Sup. Rys. S1 w [2]) . Podobnie do innych stanów eksperymentalnych, najdłuższy silnie ustrukturyzowany rejon jądrowego Ty1 gRNA znajdował się na istotnym funkcjonalnie końcu 5'. Analiza rejonów stabilnych strukturalnie zarówno w jądrze jak i cytoplazmie wskazała na ich wyższe podobieństwo strukturalne, niż w przypadku rozpatrywania całych struktur MFE (Rys. 3G w [2]). Razem, otrzymane rezultaty sugerują, że Ty1 gRNA w jądrze wykazuje niższy poziom ustrukturyzowania i jest bardziej heterogenny strukturalnie niż w cytoplazmie. Niemniej jednak, pewne stabilne i wysoce ustrukturyzowane motywy RNA są zachowane pomiędzy kompartmentami komórki.

Analizy *in vitro* sugerują, że białko Gag Ty1 zaangażowane jest w promowanie kluczowych dla retrotranspozycji odziaływań RNA:RNA [37-39, 81]. Brakowało jednak danych potwierdzających rolę Ty1 Gag w tych oddziaływaniach *in vivo*. Dlatego też, przeprowadziłam szczegółową analizę porównawczą profili reaktywności SHAPE uzyskanych dla Ty1 RNA próbkowanego w szczepie drożdży pozbawionym białka Gag, w cząstkach wirusopodobnych, gdzie zachodzi indukowana przez Gag dimeryzacja i cyklizacja gRNA a także przyłączanie $tRNA_i^{Met}$, oraz w warunkach *in vitro*, gdzie nie dochodzi do wymienionych interakcji RNA:RNA. Skupiłam się na poszukiwaniu specyficznych zmian reaktywności w obrębie istotnych dla retrotranspozycji sekwencji *cis*-regulatorowych, które mogą świadczyć o zachodzeniu tych odziaływań, bądź je wykluczyć (Rys. 4A w [2]). Przeprowadzone analizy wykazały, że w jądrze komórkowym nie dochodzi zarówno do przyłączania $tRNA_i^{Met}$ do trójczłonowego miejsca wiązania startera, cyklizacji Ty1 gRNA, ani jego dimeryzacji. Nie zauważyłam także zmian reaktywności przemawiających za powstawaniem funkcjonalnie ważnej struktury pseudowęzła na końcu 5' gRNA Ty1 (Rys.

4B w [2]). Tym samym dostarczyłam dowodów wskazujących, że białko Gag Ty1 odgrywa istotną rolę w promowaniu kluczowych dla retrotranspozycji interakcji RNA:RNA *in vivo*.

Podsumowując, prowadzone przeze mnie badania pozwoliły na otrzymanie pierwszej struktury gRNA retrotranspozonu LTR w jądrze komórkowym. Dzięki obszernej analizie porównawczej otrzymanej struktury z poprzednio rozwiązanymi strukturami Ty1 gRNA w cytoplazmie, cząstce wirusopodobnej oraz *in vitro*, po raz pierwszy zaprezentowałam, jak zmienia się struktura RNA retroelementu w trakcie cyklu replikacyjnego w komórce. Zaobserwowane przeze mnie rozległe zmiany strukturalne wskazują na istotny wpływ lokalnego środowiska kompartmentów komórkowych na zwijanie się cząsteczki RNA, podkreślając ważność podobnych badań w przyszłości. Uzyskane przeze mnie wyniki dostarczają także pierwszych dowodów *in vivo* na kluczową rolę białka Gag w ważnych dla retrotranspozycji Ty1 oddziaływaniach RNA-RNA.

5.3 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
*autorzy mieli równorzędny wkład w przygotowanie pracy

W ciągu ostatniej dekady stało się jasne, że struktura w jaką zwija się jednoniciowe RNA odgrywa krytyczną rolę w regulacji wielu procesów komórkowych takich jak transkrypcja, obróbka posttranskrypcyjna, translacja i dojrzewanie białek, a także w lokalizacji, stabilności, czy degradacji RNA. Postępujący rozwój technologii wysokoprzepustowych badań strukturalnych RNA pozwolił na poznanie struktur całych transkryptomów w żywych komórkach, umożliwiając przy tym spojrzenie na wzajemną korelację pomiędzy strukturą RNA, a procesami w które jest on zaangażowany. Rosnąca liczba transkryptomowych analiz strukturalnych *in vivo*, skłoniła nas do usystematyzowania dostępnej wiedzy, czego efektem jest niniejsza praca przeglądowa. Podsumowujemy i dyskutujemy w niej najnowsze odkrycia dotyczące wpływu czynników komórkowych na zwijanie się struktur RNA, a także relacji pomiędzy strukturą RNA, a jego funkcją w komórce. W pierwszej kolejności zaprezentowaliśmy i omówiliśmy dostępne metody umożliwiające wysokoprzepustowe badania struktury transkryptomów *in vivo*. W następnych rozdziałach dokonaliśmy przeglądu najnowszych prac na temat korelacji pomiędzy strukturą RNA a translacją, odziaływaniem z białkami oraz stabilnością RNA i jego degradacją. Poruszyliśmy także temat wpływu naturalnie występujących modyfikacji RNA, w szczególności najbardziej rozpowszechnionej w mRNA N^6-methyloadenozyny, na strukturę mRNA. Nasza praca pokazuje również, że pomimo rosnącej liczby doniesień naukowych prezentujących wyniki transkryptomowych badań strukturalnych *in vivo,* wiele kwestii wciąż pozostaje niejasnych i do pełnego zrozumienia korelacji miedzy strukturą RNA a jego funkcją w żywych komórkach niezbędne są dalsze badania.

WYKAZ SKRÓTÓW

CE	elektroforeza kapilarna (ang. *capillary electrophoresis*)
DMS	siarczan dimetylu
DNA	kwas deoksyrybonukleinowy
gRNA	genomowy RNA
IN	integraza
LINE	długie rozproszone elementy jądrowe (ang. *long interspersed nuclear element*)
LTR	długie terminalne powtórzenia na obu końcach DNA retrowirusów i retrotranspozonów LTR (ang. *long terminal repeat*)
mRNA	informacyjny RNA (ang. messenger RNA)
NGS	sekwencjonowanie nowej generacji (ang. *next generation sequencing*)
NMIA	bezwodnik kwasu 1-metyloizatoinowego
nt	nukleotyd
PBS	sekwencja wiążąca starter (ang. *primer biding site*)
PR	proteaza
RNA	kwas rybonukleinowy
RT	odwrotna transkryptaza (ang. reverse transcriptase)
RT/RH	odwrotna transkryptaza z domeną RNazy H
SHAPE	selektywna acylacja grup 2′-OH rybozy analizowana metodą odwrotnej transkrypcji (ang. *selective 2'-hydroxyl acylation analyzed by primer extension*)
SINE	krótkie rozproszone elementy jądrowe (ang. *short interspersed nuclear element*)
TE	elementy transpozycyjne (TE; ang. *transposable elements*)
TIR	odwrócone terminalne powtórzenia (ang. *terminal inverted repeats*)
tRNA	transportujący RNA (ang. *transfer RNA*)
UTR	niepodlegające translacji końcowe rejony RNA (ang. *untranslated region*)
VLPs	cząstki wirusopodobne (ang. *virus-like particles*)

REFERENCJE

1.   Andrzejewska, A., et al., *In vivo structure of the Ty1 retrotransposon RNA genome.* Nucleic Acids Res, 2021. **49**(5): p. 2878-2893.
2.   Zawadzka, M., et al., *Cell Compartment-Specific Folding of Ty1 Long Terminal Repeat Retrotransposon RNA Genome.* Viruses 2022. **14**(9).
3.   Andrzejewska, A., M. Zawadzka, and K. Pachulska-Wieczorek, *On the Way to Understanding the Interplay between the RNA Structure and Functions in Cells: A Genome-Wide Perspective.* Int J Mol Sci, 2020. **21**(18).
4.   McClintock, B., *The origin and behavior of mutable loci in maize.* Proc Natl Acad Sci U S A, 1950. **36**(6): p. 344-55.
5.   Engels, W.R. and C.R. Preston, *Identifying P factors in Drosophila by means of chromosome breakage hotspots.* Cell, 1981. **26**(3 Pt 1): p. 421-8.
6.   Taylor, A.L., *Bacteriophage-Induced Mutation in Escherichia Coli.* Proc Natl Acad Sci U S A, 1963. **50**: p. 1043-51.
7.   Shapiro, J.A., *Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli.* J Mol Biol, 1969. **40**(1): p. 93-105.
8.   Makalowski, W., et al., *Transposable elements and their identification.* Methods Mol Biol, 2012. **855**: p. 337-59.
9.   Mita, P. and J.D. Boeke, *How retrotransposons shape genome regulation.* Curr Opin Genet Dev, 2016. **37**: p. 90-100.
10.   Zawadzka, M. and K. Pachulska-Wieczorek, *For good and for bad: the role of endogenous retroelements in humans.* Postepy Biochem, 2019. **65**(3): p. 217-223.
11.   Munoz-Lopez, M. and J.L. Garcia-Perez, *DNA transposons: nature and applications in genomics.* Curr Genomics, 2010. **11**(2): p. 115-28.
12.   Havecker, E.R., X. Gao, and D.F. Voytas, *The diversity of LTR retrotransposons.* Genome Biol, 2004. **5**(6): p. 225.
13.   Goodier, J.L. and H.H. Kazazian, Jr., *Retrotransposons revisited: the restraint and rehabilitation of parasites.* Cell, 2008. **135**(1): p. 23-35.
14.   Krupovic, M., et al., *Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses.* J Virol, 2018. **92**(12).
15.   Lerat, E. and P. Capy, *Retrotransposons and retroviruses: analysis of the envelope gene.* Mol Biol Evol, 1999. **16**(9): p. 1198-207.
16.   Eickbush, T.H. and V.K. Jamburuthugoda, *The diversity of retrotransposons and the properties of their reverse transcriptases.* Virus Res, 2008. **134**(1-2): p. 221-34.
17.   Boeke, J.D., et al., *Ty elements transpose through an RNA intermediate.* Cell, 1985. **40**(3): p. 491-500.
18.   Marquet, R., et al., *tRNAs as primer of reverse transcriptases.* Biochimie, 1995. **77**(1-2): p. 113-24.
19.   Sultana, T., et al., *Integration site selection by retroviruses and transposable elements in eukaryotes.* Nat Rev Genet, 2017. **18**(5): p. 292-308.
20.   Carr, M., D. Bensasson, and C.M. Bergman, *Evolutionary genomics of transposable elements in Saccharomyces cerevisiae.* PLoS One, 2012. **7**(11): p. e50978.
21.   Kim, J.M., et al., *Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence.* Genome Res, 1998. **8**(5): p. 464-78.
22.   Curcio, M.J., S. Lutz, and P. Lesage, *The Ty1 LTR-Retrotransposon of Budding Yeast, Saccharomyces cerevisiae.* Microbiol Spectr, 2015. **3**(2): p. MDNA3-0053-2014.

23. Morillon, A., et al., *Differential effects of chromatin and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons.* Mol Cell Biol, 2002. **22**(7): p. 2078-88.

24. Curcio, M.J., et al., *Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in Saccharomyces cerevisiae.* Mol Gen Genet, 1990. **220**(2): p. 213-21.

25. Malagon, F. and T.H. Jensen, *The T body, a new cytoplasmic RNA granule in Saccharomyces cerevisiae.* Mol Cell Biol, 2008. **28**(19): p. 6022-32.

26. Checkley, M.A., et al., *P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles.* Mol Cell Biol, 2010. **30**(2): p. 382-98.

27. Clare, J.J., M. Belcourt, and P.J. Farabaugh, *Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon.* Proc Natl Acad Sci U S A, 1988. **85**(18): p. 6816-20.

28. Xu, H. and J.D. Boeke, *Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency.* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8360-4.

29. Farabaugh, P.J., *Post-transcriptional regulation of transposition by Ty retrotransposons of Saccharomyces cerevisiae.* J Biol Chem, 1995. **270**(18): p. 10361-4.

30. Garfinkel, D.J., et al., *Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1.* J Virol, 1991. **65**(9): p. 4573-81.

31. Devine, S.E. and J.D. Boeke, *Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III.* Genes Dev, 1996. **10**(5): p. 620-33.

32. Curcio, M.J. and D.J. Garfinkel, *Single-step selection for Ty1 element retrotransposition.* Proc Natl Acad Sci U S A, 1991. **88**(3): p. 936-40.

33. Pachulska-Wieczorek, K., et al., *Characterizing the functions of Ty1 Gag and the Gag-derived restriction factor p22/p18.* Mob Genet Elements, 2016. **6**(2): p. e1154637.

34. Malagon, F. and T.H. Jensen, *T-body formation precedes virus-like particle maturation in S. cerevisiae.* RNA Biol, 2011. **8**(2): p. 184-9.

35. Checkley, M.A., et al., *Ty1 gag enhances the stability and nuclear export of Ty1 mRNA.* Traffic, 2013. **14**(1): p. 57-69.

36. Cristofari, G., D. Ficheux, and J.L. Darlix, *The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins.* J Biol Chem, 2000. **275**(25): p. 19210-7.

37. Cristofari, G., et al., *A 5'-3' long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition.* EMBO J, 2002. **21**(16): p. 4368-79.

38. Nishida, Y., et al., *Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions.* Nucleic Acids Res, 2015. **43**(15): p. 7414-31.

39. Gumna, J., et al., *Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA.* RNA Biol, 2019. **16**(12): p. 1749-1763.

40. Xu, H. and J.D. Boeke, *Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements.* Mol Cell Biol, 1990. **10**(6): p. 2695-702.

41. Bolton, E.C., et al., *Identification and characterization of critical cis-acting sequences within the yeast Ty1 retrotransposon.* RNA, 2005. **11**(3): p. 308-22.

42. Feng, Y.X., et al., *The genomic RNA in Ty1 virus-like particles is dimeric.* J Virol, 2000. **74**(22): p. 10819-21.

43. Gamache, E.R., et al., *Structure-Function Model for Kissing Loop Interactions That Initiate Dimerization of Ty1 RNA.* Viruses, 2017. **9**(5).

44.   Purzycka, K.J., et al., *Exploring Ty1 retrotransposon RNA structure within virus-like particles.* Nucleic Acids Res, 2013. **41**(1): p. 463-73.

45.   Chapman, K.B., A.S. Bystrom, and J.D. Boeke, *Initiator methionine tRNA is essential for Ty1 transposition in yeast.* Proc Natl Acad Sci U S A, 1992. **89**(8): p. 3236-40.

46.   Wilhelm, M., et al., *Yeast Ty1 retrotransposon: the minus-strand primer binding site and a cis-acting domain of the Ty1 RNA are both important for packaging of primer tRNA inside virus-like particles.* Nucleic Acids Res, 1994. **22**(22): p. 4560-5.

47.   Huang, Q., et al., *Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription.* RNA, 2013. **19**(3): p. 320-32.

48.   Merino, E.J., et al., *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J Am Chem Soc, 2005. **127**(12): p. 4223-31.

49.   Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.

50.   McGinnis, J.L., et al., *The mechanisms of RNA SHAPE chemistry.* J Am Chem Soc, 2012. **134**(15): p. 6617-24.

51.   Mortimer, S.A., et al., *SHAPE-Seq: High-Throughput RNA Structure Analysis.* Curr Protoc Chem Biol, 2012. **4**(4): p. 275-97.

52.   Low, J.T. and K.M. Weeks, *SHAPE-directed RNA secondary structure prediction.* Methods, 2010. **52**(2): p. 150-8.

53.   Siegfried, N.A., et al., *RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP).* Nat Methods, 2014. **11**(9): p. 959-65.

54.   Spitale, R.C., et al., *RNA SHAPE analysis in living cells.* Nat Chem Biol, 2013. **9**(1): p. 18-20.

55.   Busan, S., et al., *Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies.* Biochemistry, 2019. **58**(23): p. 2655-2664.

56.   Marinus, T., et al., *A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy.* Nucleic Acids Res, 2021. **49**(6): p. e34.

57.   Wells, S.E., et al., *Use of dimethyl sulfate to probe RNA structure in vivo.* Methods Enzymol, 2000. **318**: p. 479-93.

58.   Waduge, P., Y. Sakakibara, and C.S. Chow, *Chemical probing for examining the structure of modified RNAs and ligand binding to RNA.* Methods, 2019. **156**: p. 110-120.

59.   Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.

60.   Burrill, C.P., et al., *Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity.* J Virol, 2013. **87**(21): p. 11670-83.

61.   Pirakitikulr, N., et al., *The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures.* Mol Cell, 2016. **62**(1): p. 111-20.

62.   Dethoff, E.A., et al., *Pervasive tertiary structure in the dengue virus RNA genome.* Proc Natl Acad Sci U S A, 2018. **115**(45): p. 11513-11518.

63.   Huber, R.G., et al., *Structure mapping of dengue and Zika viruses reveals functional long-range interactions.* Nat Commun, 2019. **10**(1): p. 1408.

64.   Li, P., et al., *Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity.* Cell Host Microbe, 2018. **24**(6): p. 875-886 e5.

65.   Mauger, D.M., et al., *Functionally conserved architecture of hepatitis C virus RNA genomes.* Proc Natl Acad Sci U S A, 2015. **112**(12): p. 3692-7.

66.   Wan, H., et al., *The In Vivo and In Vitro Architecture of the Hepatitis C Virus RNA Genome Uncovers Functional RNA Secondary and Tertiary Structures.* J Virol, 2022. **96**(8): p. e0194621.

67.   Dadonaite, B., et al., *The structure of the influenza A virus genome.* Nat Microbiol, 2019. **4**(11): p. 1781-1789.

68.   Simon, L.M., et al., *In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs.* Nucleic Acids Res, 2019. **47**(13): p. 7003-7017.

69.   Manfredonia, I., et al., *Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements.* Nucleic Acids Res, 2020. **48**(22): p. 12436-12452.

70.   Huston, N.C., et al., *Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms.* Mol Cell, 2021. **81**(3): p. 584-598 e5.

71.   Sun, L., et al., *In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs.* Cell, 2021. **184**(7): p. 1865-1883 e20.

72.   Lan, T.C.T., et al., *Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells.* Nat Commun, 2022. **13**(1): p. 1128.

73.   Cao, C., et al., *The architecture of the SARS-CoV-2 RNA genome inside virion.* Nat Commun, 2021. **12**(1): p. 3917.

74.   Boerneke, M.A., J.E. Ehrhardt, and K.M. Weeks, *Physical and Functional Analysis of Viral RNA Genomes by SHAPE.* Annu Rev Virol, 2019. **6**(1): p. 93-117.

75.   Sun, L., et al., *RNA structure maps across mammalian cellular compartments.* Nat Struct Mol Biol, 2019. **26**(4): p. 322-330.

76.   Liu, Z., et al., *In vivo nuclear RNA structurome reveals RNA-structure regulation of mRNA processing in plants.* Genome Biol, 2021. **22**(1): p. 11.

77.   Gosai, S.J., et al., *Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus.* Mol Cell, 2015. **57**(2): p. 376-88.

78.   Smola, M.J., et al., *Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis.* Nat Protoc, 2015. **10**(11): p. 1643-69.

79.   Reuter, J.S. and D.H. Mathews, *RNAstructure: software for RNA secondary structure prediction and analysis.* BMC Bioinformatics, 2010. **11**: p. 129.

80.   Lee, B., et al., *Comparison of SHAPE reagents for mapping RNA structures inside living cells.* RNA, 2017. **23**(2): p. 169-174.

81.   Gumna, J., et al., *RNA Binding Properties of the Ty1 LTR-Retrotransposon Gag Protein.* Int J Mol Sci, 2021. **22**(16).

<h1 style="text-align:center">ŻYCIORYS NAUKOWY</h1>

EDUKACJA

2017-2022   Studia III stopnia; profil biologiczny
                   Środowiskowe Studium Doktoranckie ICHB PAN

2015   **Tytuł magistra biotechnologii**
                   Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu
                   Praca magisterska pt. „IRF8-dependent gene regulation in cells from
                   vasculature" realizowana w Zakładzie Genetyki Molekularnej
                   Człowieka, pod kierunkiem prof. dr hab. Johannes'a A. R. Bluijssen'a.

2013-2015   Studia II stopnia na kierunku: biotechnologia w języku angielskim;
                   Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu

2013   **Tytuł licencjata biologii molekularnej**
                   Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu.
                   Praca licencjacka pt. „Analiza aktywności gamma-sekretazy w
                   komórkach roślinnych" realizowana w Zakładzie Biologii
                   Komórkowej i Molekularnej, pod kierunkiem prof. dr hab.
                   Przemysława Wojtaszka

2010-2013   Studia I stopnia na kierunku: biologia molekularna; Wydział Biologii
                   Uniwersytetu im. Adama Mickiewicza w Poznaniu

PRZEBIEG KARIERY ZAWODOWEJ

09.2015-08.2017   Asystent w Katedrze Biochemii i Biologii Molekularnej Uniwersytetu
                   Medycznego im. Karola Marcinkowskiego w Poznaniu

07.2013-10.2013   Praktyki Zawodowe w Zakładzie Genetyki Molekularnej Człowieka,
                   UAM

11.2011-07.2012   Praktyki Zawodowe w Zakładzie Biologii Molekularnej i
                   Komórkowej UAM

PUBLIKACJE

<u>Małgorzata Zawadzka</u>, Angelika Andrzejewska-Romanowska, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek **(2022)** *Cell compartment-specific folding of Ty1 Long Terminal Repeat retrotransposon RNA genome* **Viruses** 14(9):2007

Andrzejewska Angelika, <u>Małgorzata Zawadzka</u>, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek **(2021)** *In vivo structure of the Ty1 retrotranspozon RNA genome*, **Nucleic Acid Research** 49(5):2878-2893

Andrzejewska Angelika*, <u>Małgorzata Zawadzka</u>*, Katarzyna Pachulska-Wieczorek **(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.

<u>Małgorzata Zawadzka</u>, Katarzyna Pachulska Wieczorek **(2019)** *Na dobre i na złe: rola endogennych retroelementów u człowieka*, **Postępy Biochemii** 65(3):217-223

Małgorzata Szczepańska, Przemysław Wirstlein, <u>Małgorzata Zawadzka,</u> Ewa Wender-Ozegowska, Paweł Jagodziński **(2018)** *Alternation of ten-eleven translocation 1, 2, and 3 expression in eutopic endometrium of women with endometriosis-associated infertility*, **Gynecological Endocrinology** 34(2):1084-1090

<u>Małgorzata Zawadzka</u>, Paweł Jagodziński **(2017)** *Exercise-induced epigenetic regulations in inflammatory related cells*, **Journal of Applied Biomedicine** 15(1):63-70

UDZIAŁ W PROJEKTACH BADAWCZYCH

- o Wykonawca w projekcie NCN OPUS 2020/39/B/NZ3/03020 „ Analiza struktury RNA w komórce i jej kompartmentach w skali transkryptomowej oraz identyfikacja wpływu czynników komórkowych na strukturę RNA w S. cerevisiae", kierownik dr hab. Katarzyna Pachulska-Wieczorek

- o Wykonawca w projekcie NCN SONATA BIS 2016/22/E/NZ3/00426 „Struktura RNA i jego oddziaływania z białkiem Gag warunkujące funkcje RNA Ty1 na różnych etapach replikacji retrotranspozonu", kierownik dr hab. Katarzyna Pachulska-Wieczorek

- o Wykonawca w projekcie NCN OPUS 2013/11/B/NZ9/00783 "Symbiotics-activated TLR-signalling as a measure of improved immunity in chickens", kierownik dr. Anna Sławińska

DONIESIENIA KONFERENCYJNE

- Referat: „*Does retroelement work as a signaling molecule? - Inside the Arc*" RNA&Computing – 17-18.05.2018 Biedrusko

- Referat: „*Epigenetyczna Regulacja Ekspresji Genów Indukowana Wysiłkiem Fizycznym*" IV Ogólnopolska Konferencja Młodych Naukowców „Wychowanie Fizyczne, Sport, Turystyka oraz Zdrowie w Badaniach Naukowych" – Szczecin 2.06.2016

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

- 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 D., Pachulska-Wieczorek K. The RNA 2019 Meeting – 11-16.09.2019 Kraków

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

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

- Poster: „*Molecular mechanisms of Toll-like receptors activation by probiotic bacteria (Lactococcus lactis subsp. cremoris)*"; Slawinska A., Plowiec A., Zawadzka M., Siwek M., Bluyssen H.; XXVII Międzynarodowe Sympozjum Drobiarskie PB WPSA "Nauka praktyce - praktyka nauce" – 14-16.09.2015 Bydgoszcz

- Poster: „*In vitro model for prebiotics selection based on immune-related gene expression in chickens*"; Płowiec A., Zawadzka M., Siwek M., Sławińska A., Bluijssen H.A.R.; Central European Conference on Regenerative Medicine – 14-15.03.2015 Bydgoszcz

- Poster: „*Towards the understanding of the plants presenilins functions*"; Skrzypczak T., Smolarkiewicz M., Wasik K., Zawadzka M., Wojtaszek P; III Konferencja Naukowo-Dydaktyczna Wydziału Biologii UAM – 10-12.04.2014 Poznań

- Poster: „*Plant as a model for functional studies of the Alzheimer disease related presenilin and gamma-secretase complex*"; Skrzypczak T., Smolarkiewicz M., Wasik K., Zawadzka M., Wojtaszek P; Bioconnect 2013 – 15-16.05.2013 Nickel Technology Park Poznań Złotniki

PRACE WCHODZĄCE W SKŁAD

ROZPRAWY DOKTORSKIEJ

In vivo structure of the Ty1 retrotransposon RNA genome

Angelika Andrzejewska [ID][1], **Małgorzata Zawadzka** [ID][1], **Julita Gumna** [ID][1], **David J. Garfinkel** [ID][2] **and Katarzyna Pachulska-Wieczorek** [ID][1,*]

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

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), synthetize 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 gRNA's 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 virio*, *in virio* 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) (*MATα his3-Δ200hisG ura3 gal3 Spo⁻*) containing pBDG202 (pGTy1-H3CLA/URA3/2μ), a multicopy plasmid with Ty1 under the control of the *GAL1* 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 *GAL1* 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 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) 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 DH5α 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 μl of PBS or PBS with 1 mg/ml kasugamycin. Each sample was divided into two equal amounts and treated with 40 μl SHAPE reagent in DMSO (10 mM NMIA, 10 mM 1M7 or 20 mM NAI, final concentration) or 40 μl 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 μl of 1 M DTT. Cells were collected at 8000 × *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 × *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 μl 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 synthetized 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× 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$_i^{Met}$ binding, dimerization and cyclization sequences forced to be single-stranded. The MFE structure model of the *in virio* Ty1 RNA + 1–1482 region was generated based on previously published SHAPE data (33), using the RNAstructure.
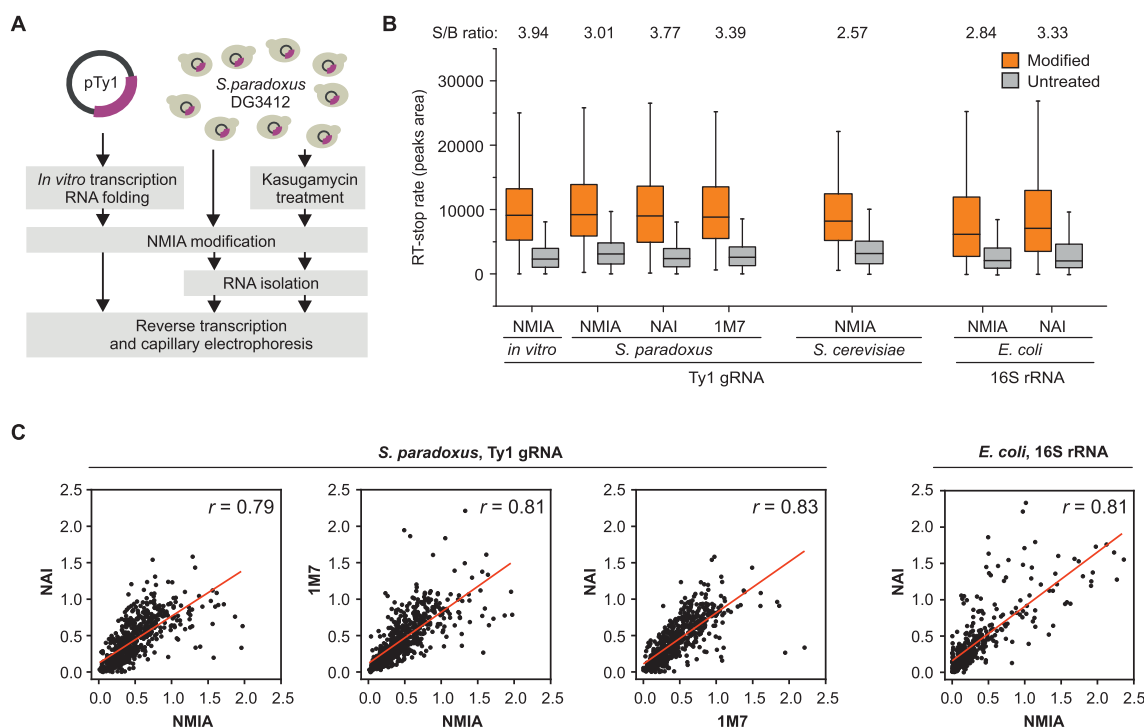
Figure 1. Experimental workflow and SHAPE reagent comparison. (**A**) Experimental workflow. (**B**) Box plot analysis of RT-stop rate (CE peaks area) measurement with medians for signal and background from capillary electrophoresis experiments for different SHAPE reagents and different cells. Plots 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 r RNA in *E. coli*. (**C**) Correlation of position-dependent NMIA, NAI and 1M7 reactivies 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 RNAstructure). 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 *GAL1* 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 virio* and *ex virio* (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 (ΔSHAPE, 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
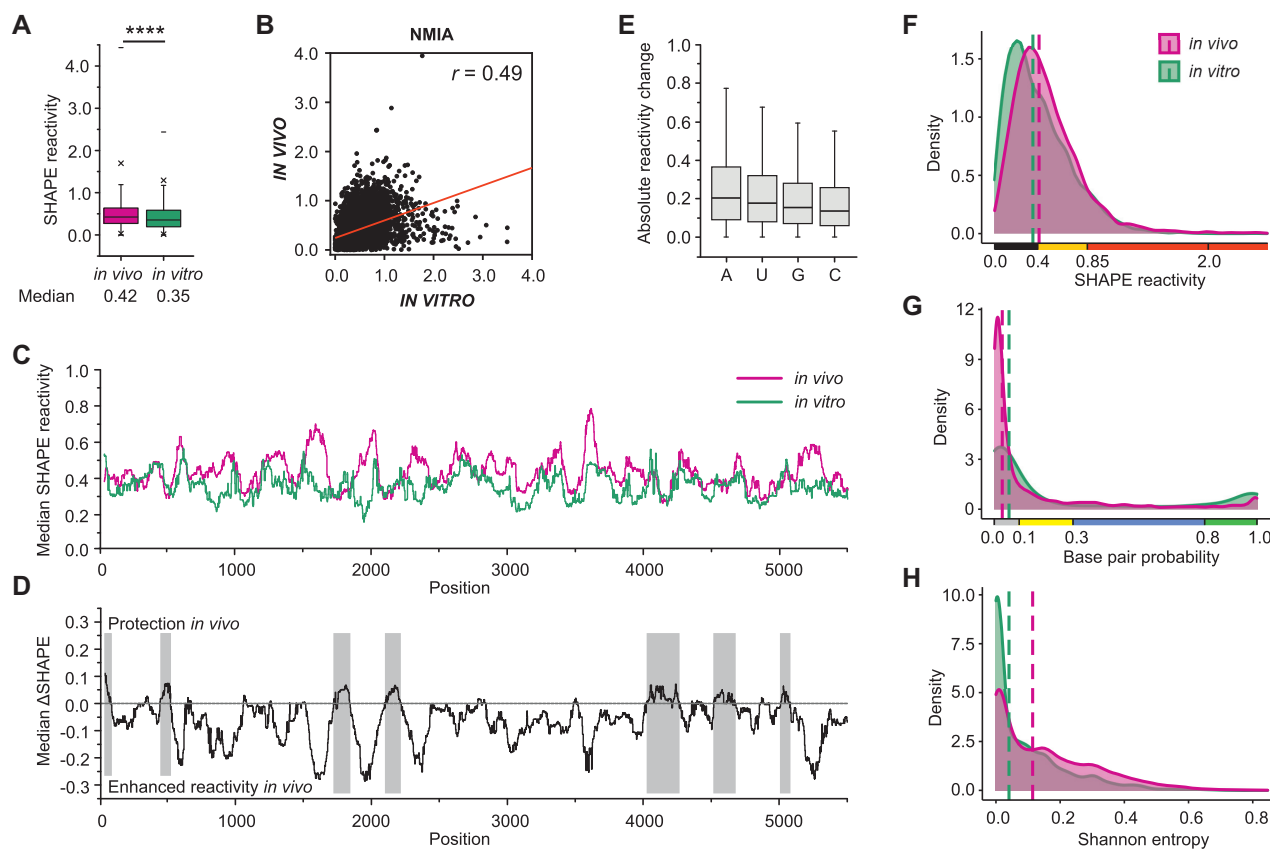
Figure 2. SHAPE reactivities, pairing probabilities and Shannon entropies for *in vivo* and *in vitro* Ty1 gRNAs. (**A**) Box plot analysis of SHAPE reactivity distributions with medians. Significance was computed by unpaired two-tailed Mann–Whitney test; ****$P < 0.0001$. (**B**) Correlation of position-dependent 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 ΔSHAPE 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
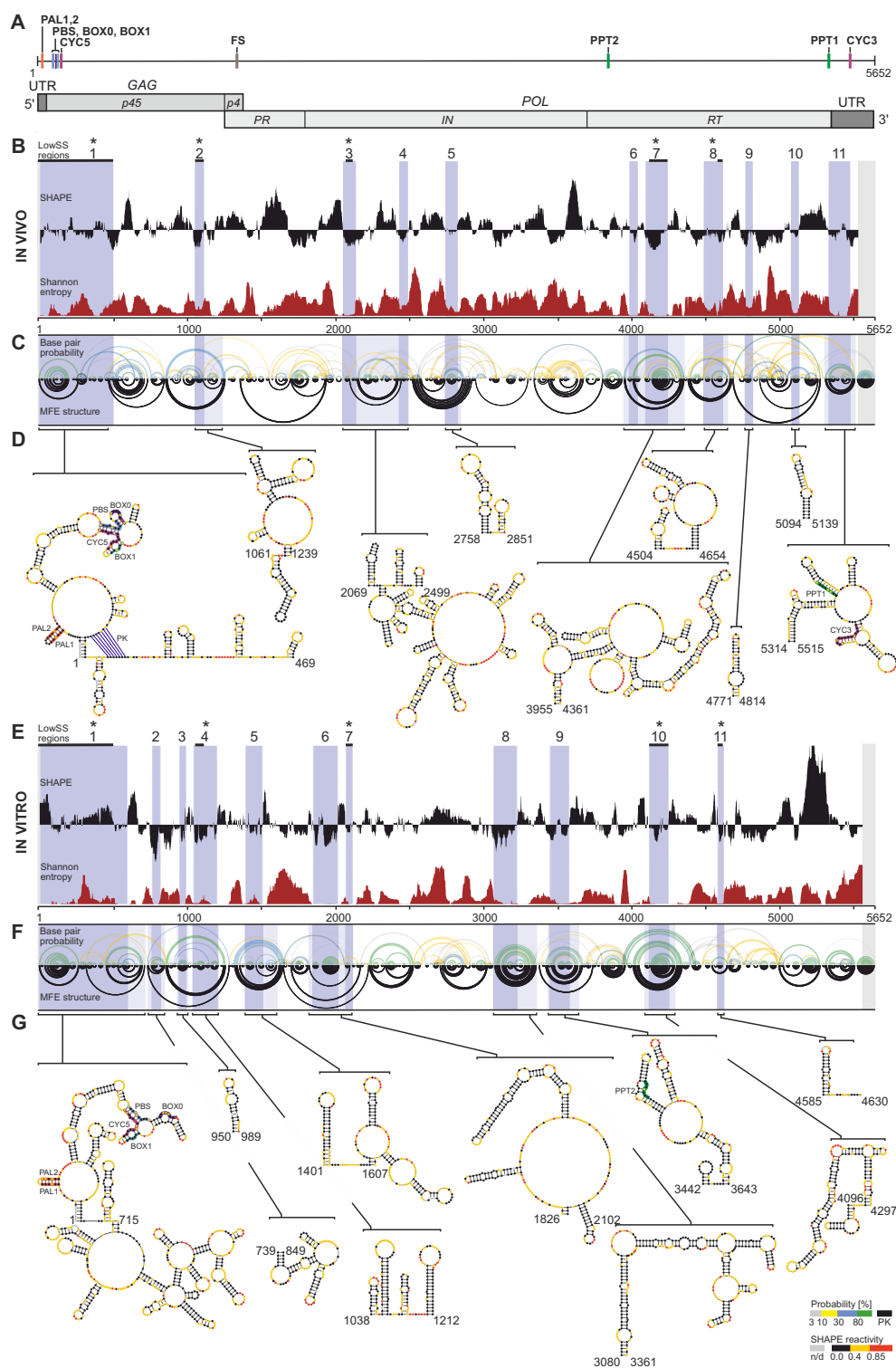
Figure 3. SHAPE-directed structure models and structurally well-defined regions for *in vivo* and *in vitro* Ty1 gRNAs. (**A**) Functional organization of Ty1 gRNA. The known *cis*-acting sequences are indicated: dimerization sequences (PAL1, PAL2), primer-binding sites (PBS, BOX0, BOX1), cyclization 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* Ty 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 parings 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-
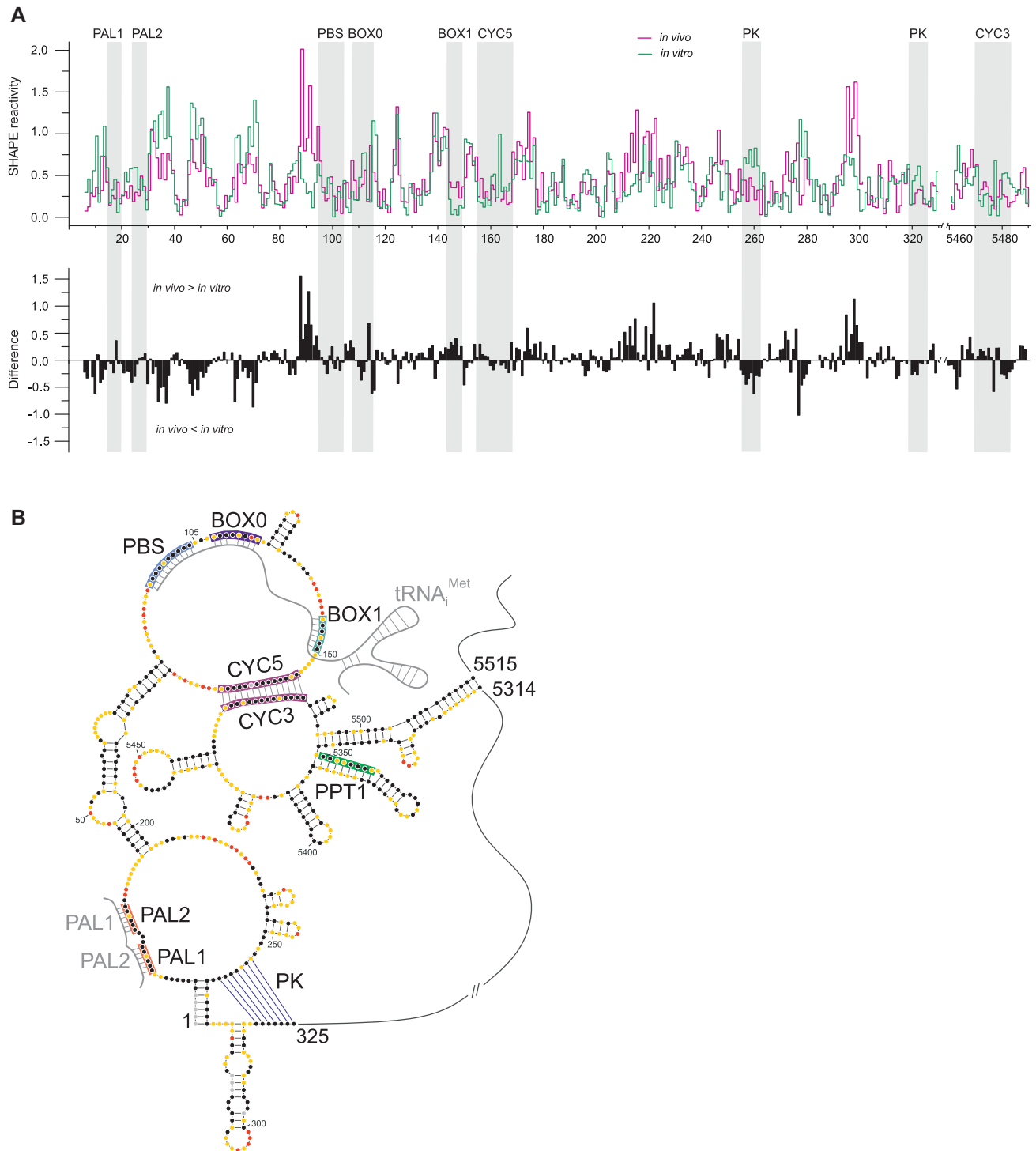
Figure 5. SHAPE-based analysis of Ty1 gRNA dimerization, cyclization and tRNA$_i^{Met}$ annealing. (**A**) The step plot (top) of NMIA reactivity for *in vivo* and *in vitro* gRNAs and the difference plot (bottom) calculated by subtracting the *in vitro* reactivities from those of the *in vivo* gRNA. 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$_i^{Met}$. *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$_i^{Met}$ 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 virio*, 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
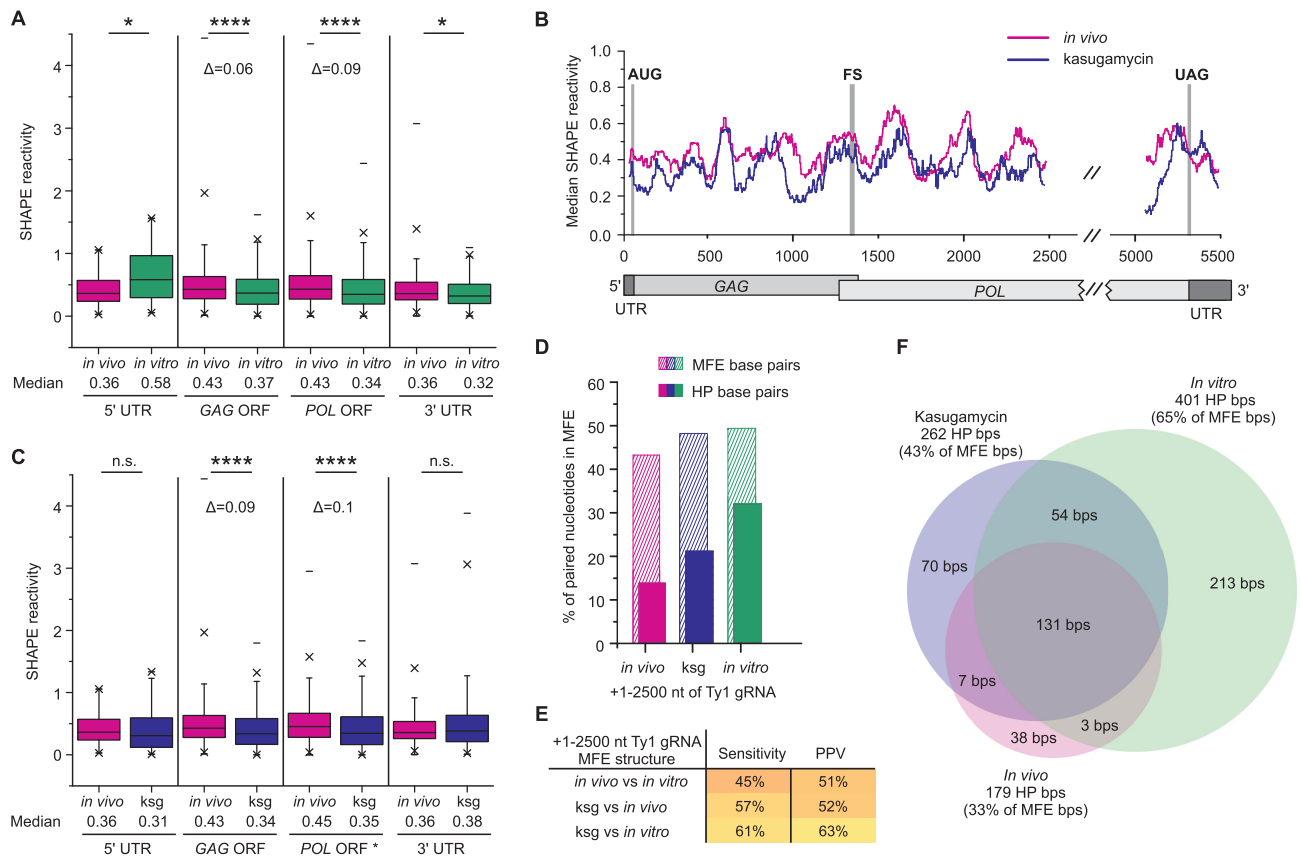
Figure 6. Correlating Ty1 RNA structure and translation. (**A**) Box plot analysis of SHAPE reactivity distributions with medians for coding and untranslated regions of *in vivo* and *in vitro* Ty1 gRNAs. (**B**) Median SHAPE reactivity profiles (smoothed with a 75-nt window) for Ty1 gRNA from cells 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 Δ*G*) of 5′UTRs and the high structure (low Δ*G*) 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 support the formation of a long stem-loop structure within the 3′UTR of Ty1 RNA *in vivo* and *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

FUNDING

REFERENCES

1. 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.
2. 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.
3. 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.
4. 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., Luganini,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., Lusvarghi,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., Eizentstat,L.D., Lusvarghi,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., Mazor,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., Ozsolak,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	TTGGTGTTGTCCTGGAAGTGA
Ty1 PR_5	TCGAATTGTCTGTGCATTGG
Ty1 PR_6	AGACACTGGCCTGAAACTGG
Ty1 PR_7	GCCACGAGGATGTATTTTGG
Ty1 PR_8	GGAGGTGTGGAATCGGTTGG
Ty1 PR_9	GGTACGTTTGTATGATTAGTCTCA
Ty1 PR_10	GTCTCGTGATACCTTAATTTCAG
Ty1 PR_11	TGCTGAATATCACCTCTTGC
Ty1 PR_12	CAAATTGTCACTTGACTG
Ty1 PR_13	TGCATTTCATGTACCTTCTC
Ty1 PR_14	GCTTCCGTAGTTGAAGTACAT
Ty1 PR_15	TCTATTCCAACATACCACCC
Ty1 PR_16	CTCCTCAAGGATTTAGGAATCC
Ty1 PR_17	CATTGTTGATAAAGGCTATA
16S rRNA PR_559	CTTTACGCCCAGTAATT

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
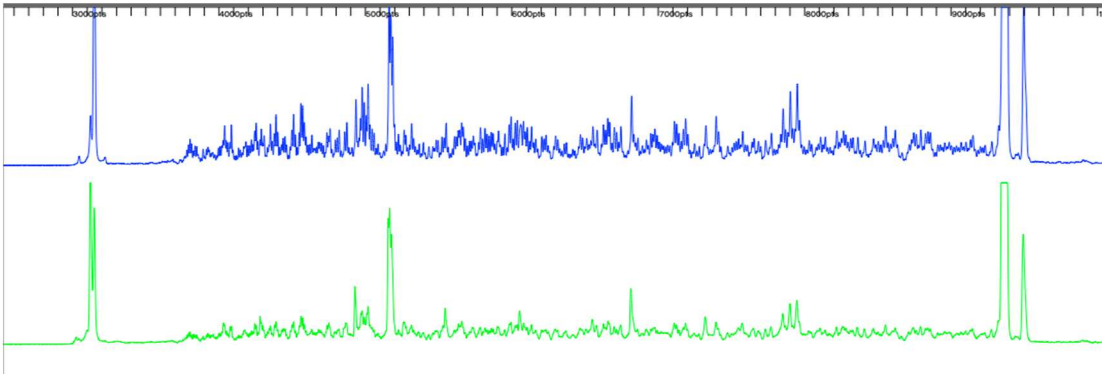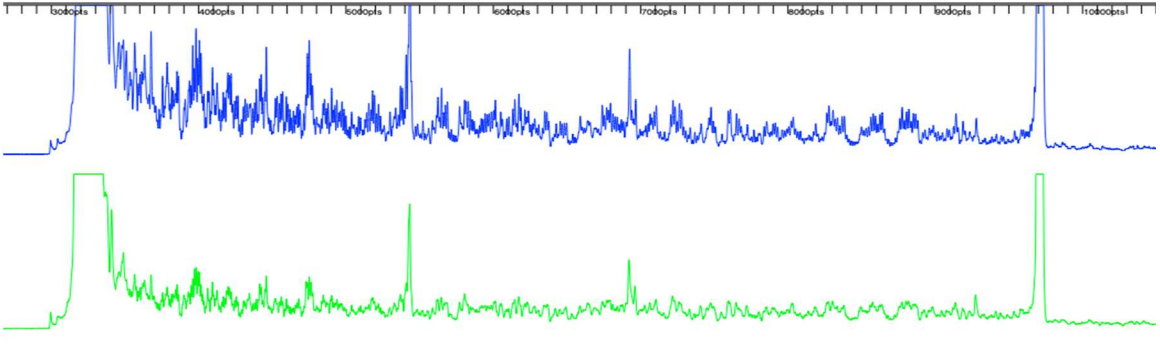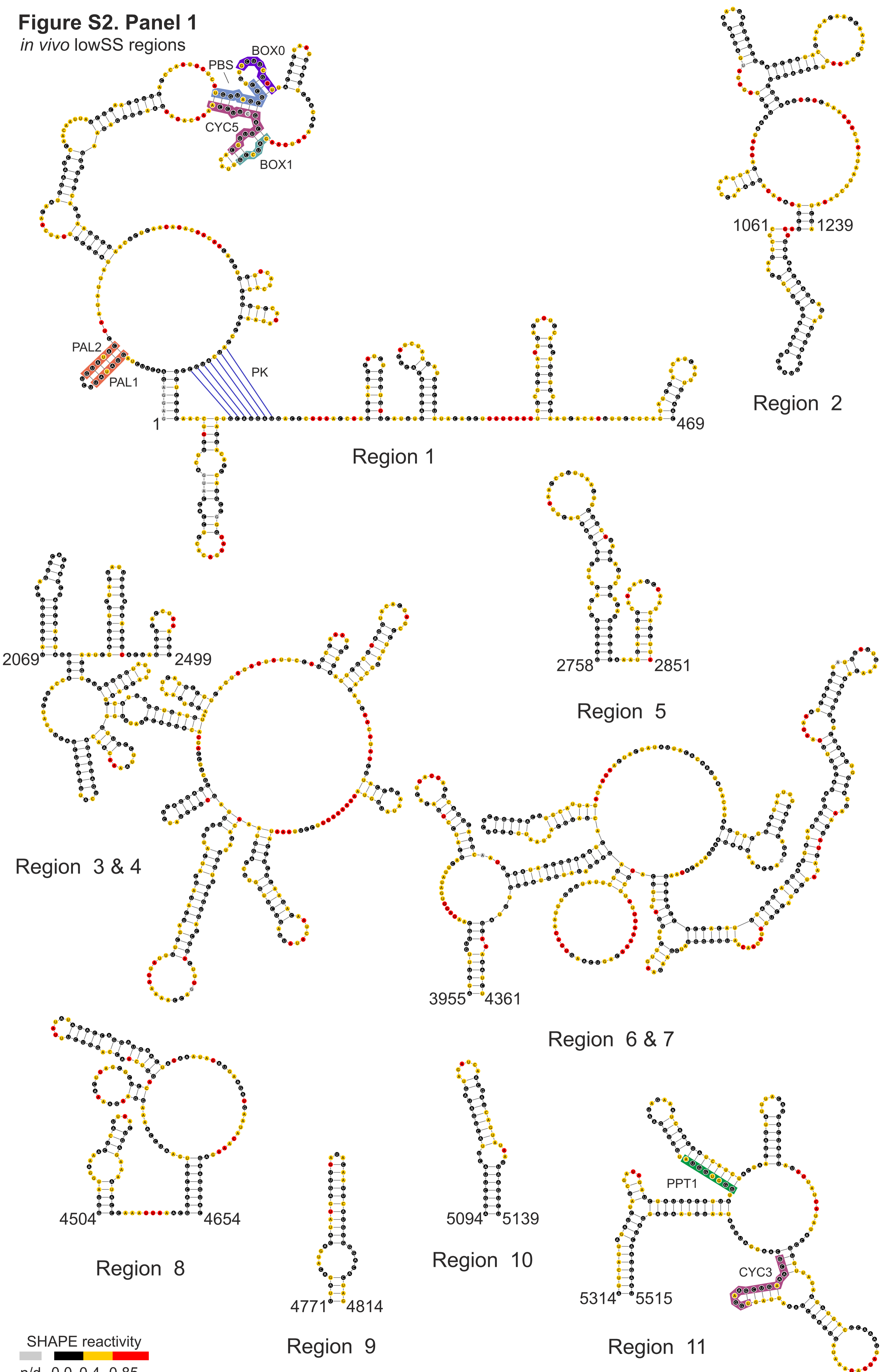

NAI *in vivo*, Ty1 PR_1
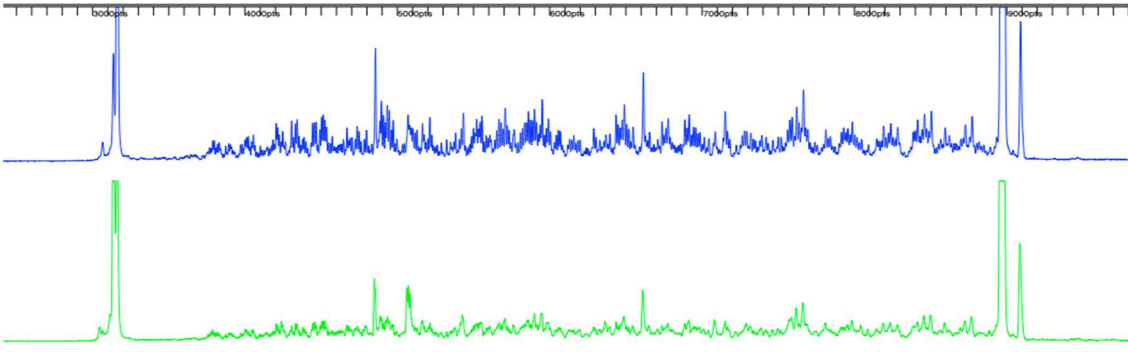
1M7 *in vivo*, Ty1 PR_1
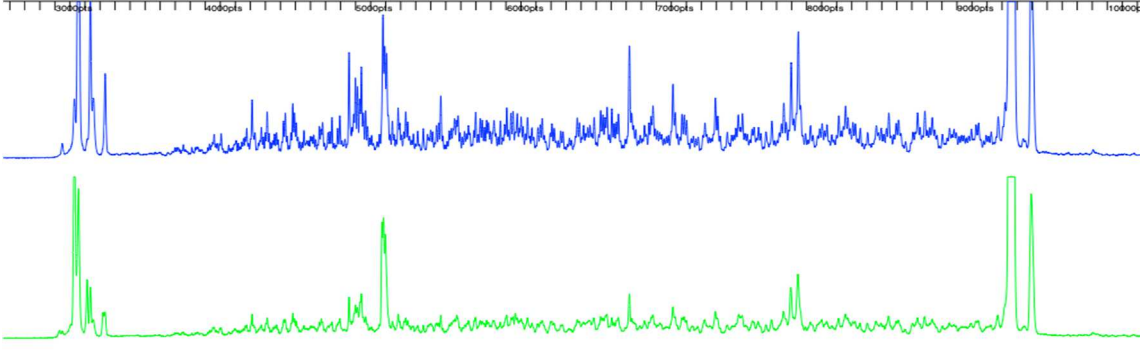
NMIA *in vitro*, Ty1 PR_1

Figure S2. Panel 1
in vivo lowSS regions
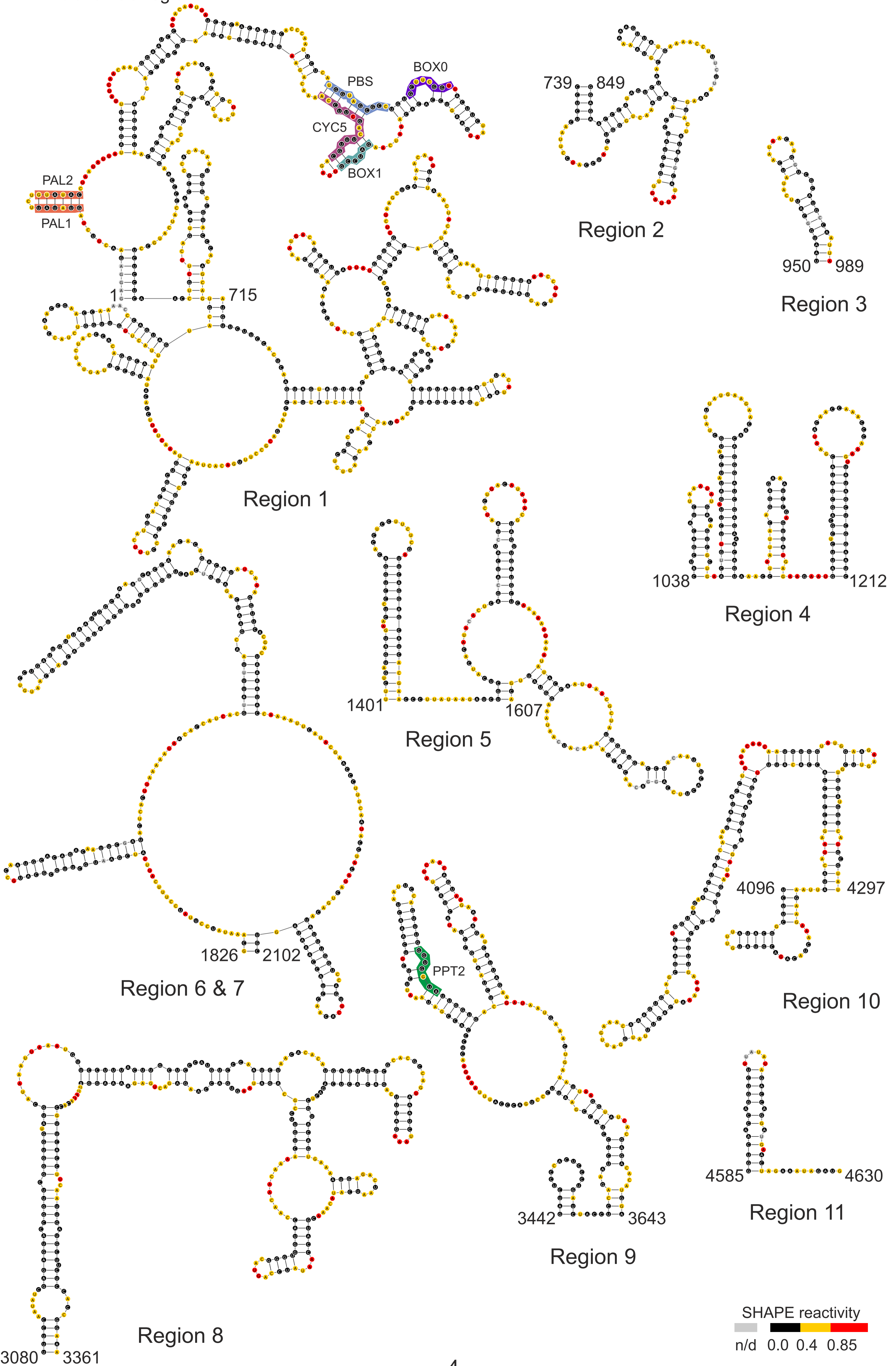
PBS
BOX0
CYC5
BOX1
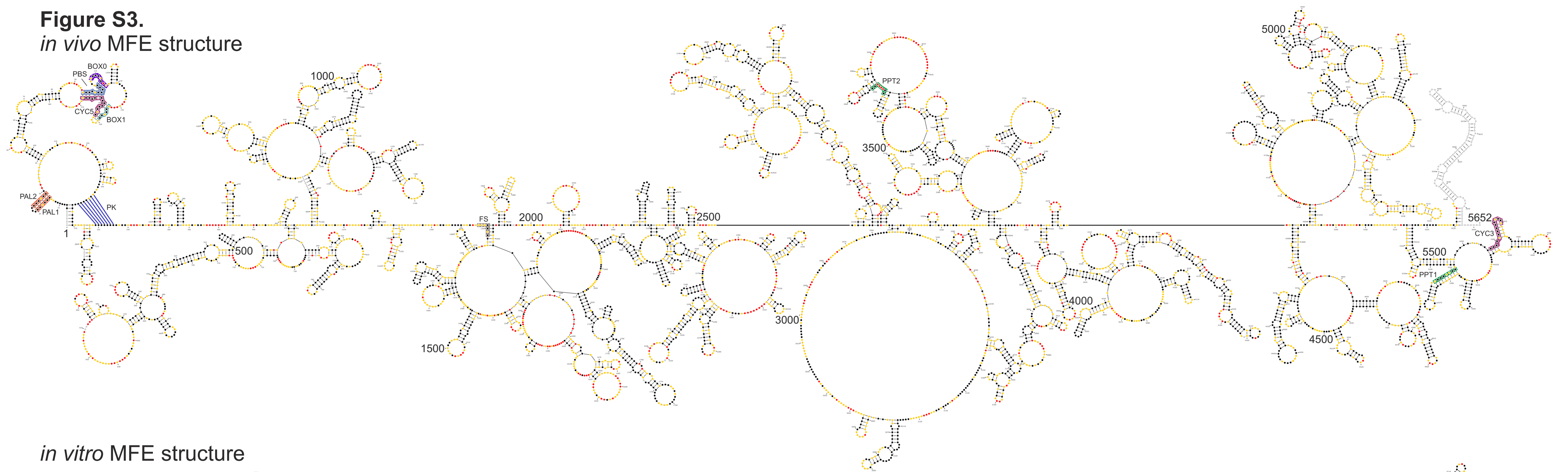PAL2
PAL1
PK
1
469
Region 1

1061 1239
Region 2

2069 2499
Region 3 & 4

2758 2851
Region 5

3955 4361
Region 6 & 7

4504 4654
Region 8

4771 4814
Region 9

5094 5139
Region 10

PPT1
CYC3
5314 5515
Region 11

SHAPE reactivity

n/d 0.0 0.4 0.85

3

Figure S2. Panel 2. *in vitro* lowSS regions.

Region 1

Region 2

Region 3

Region 4

Region 5

Region 6 & 7

Region 8

Region 9

Region 10

Region 11

SHAPE reactivity

n/d 0.0 0.4 0.85

Figure S3.
in vivo MFE structure

in vitro MFE structure

SHAPE reactivity

n/d 0.0 0.4 0.85

5

Figure S4.

A

	Median
in vivo	0.42
in vitro	0.37
in virio	0.29
ex virio*	0.32
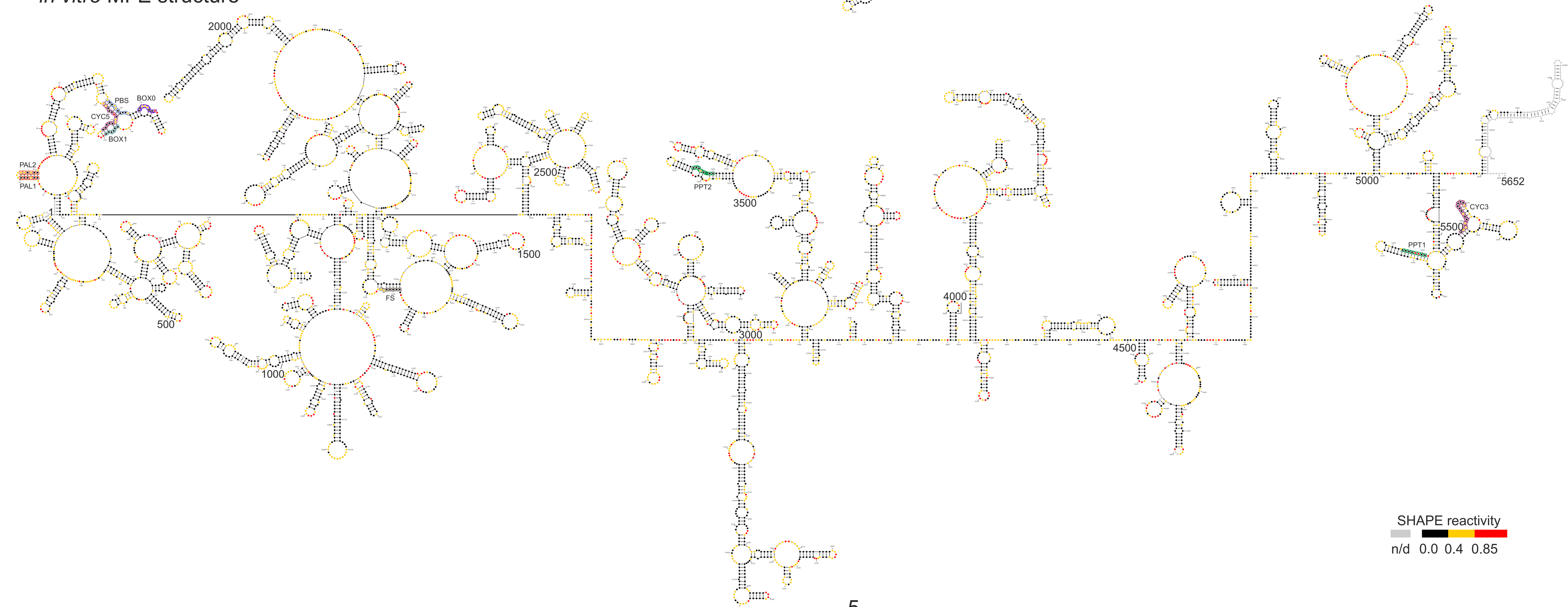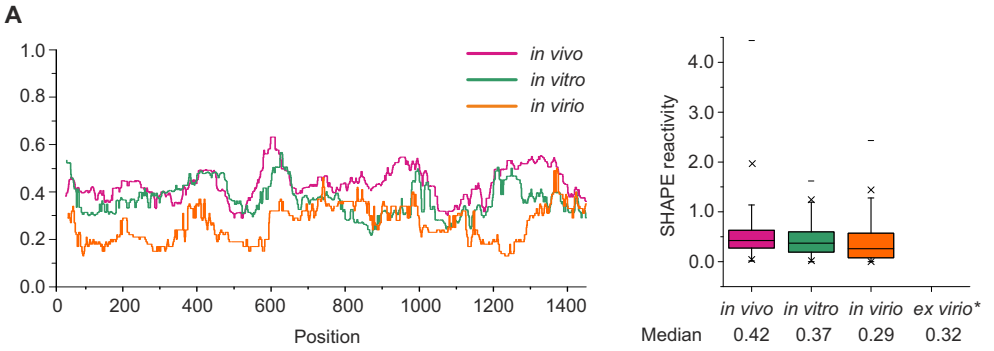
B

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

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 SHAPE dla poszczególnych nukleotydów badanych RNA dostępny jest pod adresem:

https://academic.oup.com/nar/article/49/5/2878/6148176#supplementary-data

Article

Cell Compartment-Specific Folding of Ty1 Long Terminal Repeat Retrotransposon RNA Genome

Małgorzata Zawadzka [1] **, Angelika Andrzejewska-Romanowska** [1] **, Julita Gumna** [1] **, David J. Garfinkel** [2] **and Katarzyna Pachulska-Wieczorek** [1,*]

1   Department of Structure and Function of Retrotransposons, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland
2   Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA
*   Correspondence: kasiapw@ibch.poznan.pl; Tel.: +48-61-852-85-03; Fax: +48-61-852-05-32

Abstract: The structural transitions RNAs undergo during trafficking are not well understood. Here, we used the well-developed yeast Ty1 retrotransposon to provide the first structural model of genome (g) RNA in the nucleus from a retrovirus-like transposon. Through a detailed comparison of nuclear Ty1 gRNA structure with those established in the cytoplasm, virus-like particles (VLPs), and those synthesized in vitro, we detected Ty1 gRNA structural alterations that occur during retrotransposition. Full-length Ty1 gRNA serves as the mRNA for Gag and Gag-Pol proteins and as the genome that is reverse transcribed within VLPs. We show that about 60% of base pairs predicted for the nuclear Ty1 gRNA appear in the cytoplasm, and active translation does not account for such structural differences. Most of the shared base pairs are represented by short-range interactions, whereas the long-distance pairings seem unique for each compartment. Highly structured motifs tend to be preserved after nuclear export of Ty1 gRNA. In addition, our study highlights the important role of Ty1 Gag in mediating critical RNA–RNA interactions required for retrotransposition.

Keywords: RNA genome; RNA structure; cell compartment-specific folding; LTR-retrotransposon; gRNA dimerization; gRNA cyclization; tRNA annealing; Ty1; Gag

1. Introduction

Eukaryotic genomes are rich in transposable elements [1]. One prevalent group contains the long terminal repeat (LTR)-retrotransposon families Ty1/Copia (*Pseudoviridae*), Ty3/Gypsy (*Metaviridae*), and BEL/Pao (*Belpaoviridae*) [2]. Their genomic organization resembles that of simple retroviruses, but most lack a functional envelope gene required for extracellular transmission [3]. One of the best characterized LTR-retrotransposons is Ty1, which is present in many *Saccharomyces cerevisiae* strains [4,5] and propagates via an RNA intermediate [6,7]. The integrated Ty1 element is transcribed by the RNA polymerase II, producing a gRNA that comprises two partially overlapped *GAG* and *POL* ORFs flanked by untranslated regions [8]. After export to the cytoplasm, Ty1 gRNA fulfills a dual role, serving as the template for Gag and Gag–Pol protein synthesis and as a genome that is packaged in a dimeric form into VLPs where reverse transcription occurs. In VLPs, Gag and Gag–Pol undergo proteolytic maturation, gRNA is reversed transcribed using tRNA$_i^{Met}$ as a primer, and the resulting cDNA is imported into the nucleus and integrated into host DNA, usually upstream of the genes transcribed by RNA polymerase III [9–11].

Ty1 Gag is the structural component of VLPs and mediates RNA transactions during the process of Ty1 retrotransposition. Direct interactions with Gag are needed for Ty1 RNA trafficking to preassembly sites termed retrosomes, as well as VLP assembly [12]. Gag also enhances the stability and nuclear export of the Ty1 gRNA [13]. Like retroviral Gag polyproteins, Ty1 Gag displays nucleic acid chaperone activity in vitro and facilitates RNA–RNA interactions such as dimerization and cyclization of Ty1 gRNA, annealing of tRNA$_i^{Met}$, and initiation of reverse transcription [14–17].

Ty1 gRNA is 5.6-kb long, is not spliced, and ~15% of the transcripts undergo polyadenylation [18]. However, Ty1 transcripts are highly abundant and can comprise ~0.8% of total RNA, in part because Ty1 RNA is more stable than other mRNAs [19,20]. Like retroviruses, the Ty1 gRNA 5′ and 3′ termini contain cis-acting sequences directly involved in retrotransposition [15,16,21–23]. Recently, we showed that cytoplasmic Ty1 gRNA adopts a significantly different and more heterogeneous structure than when synthesized under in vitro conditions, but it retains specific well-structured regions containing functional cis-acting sequences [24]. We also provided evidence that critical RNA–RNA interactions required for retrotransposition occur prior to VLP assembly.

Here, we used SHAPE (Selective 2′ Hydroxyl Acylation analyzed by Primer Extension) mapping to explore the Ty1 gRNA structure inside the nucleus of living yeast cells. Using complementary and synergistic experimental and bioinformatics approaches, we compared structural elements of nuclear Ty1 gRNA with that in the cytoplasm, VLPs, and in vitro. These analyses reveal cell compartment-specific structural transitions that Ty1 gRNA undergoes during the retrotransposon replication cycle. We also demonstrate the importance of Ty1 Gag in functionally important RNA–RNA interactions. Together, our work provides the first RNA structure model for a retroviral-like nuclear gRNA and the structural rearrangements that occur during its cellular journey, further increasing our understanding of RNA folding.

2. Materials and Methods

2.1. Yeast Strains, Media, and Growth Conditions

Strains DG3408 (MAC483) and DG3412 (MAC520) [13] are derived from Ty1-less *S. paradoxus* strain DG1768 (MATα his3-Δ200hisG ura3 gal3 Spo⁻) and contain multicopy plasmids pGTy1fs (pBDG1112, pGTy1gag-ATGfs) and pGTy1 (pBDG202, wild type), respectively, under the control of GAL1 promoter [25–27]. DG3408 was grown overnight to OD600 of 1.0 in SC-Ura 2% raffinose broth at 30 °C with shaking at 250 rpm. The saturated yeast culture was centrifuged at 7500× *g* for 5 min, the pellet was washed once in PBS buffer, and then it was resuspended in SC-Ura 2% galactose broth. The culture was grown for 2–4 h at 22 °C with 250 rpm shaking to induce "Ty1fs" expression from the GAL1 promoter. The *S. cerevisiae* reference strain BY4742 (MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) [28,29] was used for testing the effectivity of RNA modification by NMIA, and NAI in the nucleus was grown in conditions similar to those mentioned above using SC-complete broth instead of SC-Ura.

2.2. RNA Modification

A total of 100 mL of galactose-induced yeast culture (DG3408 or BY4742) was centrifuged at 7500× *g* for 5 min, the cell pellet was washed with PBS, and then it was resuspended in 720 μL of fresh PBS. Samples were divided into two equal parts and treated with either 40 μL of SHAPE reagent in DMSO [final concentration: 10 mM NMIA, 10 mM NAI; (+)] or 40 μL DMSO alone (-). The modification reactions were carried out at 37 °C for 15 min. RNA modification reactions with NAI were additionally quenched by adding 1 M DTT at a 1:1 ratio to the reaction volume. Cells were collected by centrifugation at 8000× *g* for 5 min.

2.3. Isolation of Total RNA

RNA isolation was performed as previously described [24]. Briefly, the cell pellet was 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. Then, samples were centrifuged at 12,000× *g* for 5 min, supernatants were transferred to fresh tubes, and RNA was extracted twice with phenol (pH 8.0) and twice with acid phenol:chloroform (pH 4.5). RNA was recovered by LiCl precipitation overnight at −20 °C, pellets were washed twice with 70% ethanol, and then they were resuspended in an appropriate amount of water.

2.4. Primer Extension and Data Processing

The optimal amount of SHAPE modified (+) or control (−) RNA sample was combined with 1 μL of 10 μM fluorescently labeled primer [Cy5 (+) or Cy5.5(−)] and 1 μL 2 mM EDTA pH 8.0, 3.84 μL of 5 M betaine and water to a final volume of 12 μL. Primer-template solutions were initially incubated at 95 °C for 3 min, 37 °C for 10 min, and 55 °C for 2 min. A total of 8 μL of reverse transcriptase mix (SuperSript III, Invitrogen, Waltham, MA, USA) was added to each reaction tube, followed by an additional 90 min incubation at 50 °C. Sequencing ladders were prepared using labeled primer: WellRed D2 (ddA) or IRD-800 (ddT) and a Thermo Sequenase Cycle Sequencing kit (Applied Biosystems, Waltham, MA, USA) as recommended by the manufacturer. Reverse transcribed samples and sequencing ladders were purified using ZR DNA Sequencing Clean-Up Kit (Zymo Research, Irvine, CA, USA) and analyzed on a GenomeLab GeXP Analysis System (Beckman-Coulter, Pasadena, CA, USA). Raw electropherograms were processed using ShapeFinder software [30] and normalized using RNAthor software [31]. Data for each read were obtained from at least two independent experiments. All primers are listed in Table S1.

2.5. RNA Secondary Structure Modeling and Analysis

A SHAPE-derived minimum free energy (MFE) secondary structure model for nuclear Ty1 gRNA was predicted using an updated version of SuperFold (v1.1) [32]. The previously obtained SHAPE data for Ty1 gRNA in the cytoplasm and the in vitro state [24] were recalculated using the SuperFold v1.1 to avoid biases resulting from algorithm differences. In brief, SuperFold v1.1 software uses Partition and Fold functions implemented in RNA structure together with the SHAPE reactivities input as pseudo-energy constraints [32,33] for prediction of the MFE structure model, calculation of base pairing probabilities for all possible canonical base pairs, and identification of well-defined regions (lowSS). Slope and intercept folding parameters were set as the default (1.8 and −0.6 kcal mol^{-1}, respectively) for each calculation. A partition function was computed in 1200-nt sliding windows with the increment of 100 nts, the pairing distance was limited to 600 nts, and the four additional calculations were performed on the 5′ and 3′ ends of the nuclear Ty1 gRNA. Moreover, 300 nts were trimmed from each window to exclude the end effect. Base-pair probabilities from multiple windows were combined. Shannon entropies were calculated from individual base-pairing probabilities and combined into a single, 55-nt sliding windows profile. The MFE structure was predicted with the SHAPE data as pseudo-energy constraints, and base pairs with >99% probability were added as hard constraints. Fold was run in 3000-nt sliding windows with 300 nt-increments, and the pairing distance was set to 600 nts. Four additional fold calculations were performed on the 5′ and 3′ ends, as described above. The final MFE structure was combined from multiple windows with the requirement that base pairs must appear in a majority of the windowed folds. Well-defined regions (lowSS) were identified by selecting regions with at least 40 nts with both median SHAPE and median Shannon entropy values below their respective medians. Some lowSS regions were combined or expanded to include the entirety of helices predicted in MFE. RNA structures were visualized using VARNA [34]. Additionally, potential pseudoknot formation was analyzed with ShapeKnots [35], in which full-length nuclear Ty1 gRNA was folded in 600-nt sliding windows with 100-nt increments. Pseudoknots occurring in >50% of the windows were included in the MFE structure.

2.6. Signal-To-Background Ratio and Correlation Calculation

The signal-to-background (S/B) ratio for NMIA and NAI modification in the yeast nucleus was calculated based on ShapeFinder raw output files. The signal and background peak area values for each nucleotide were averaged, and their medians were estimated. S/B was then calculated as the ratio between the medians of signal and background. Spearman's correlations and linear regressions were computed with GraphPad Prism 8 software.

2.7. Comparison of SHAPE Reactivities and SHAPE-Directed Structural Models of Ty1 gRNA

Sensitivity and PPV values for the entire nuclear Ty1 gRNA MFE model were calculated relative to the cytoplasmic or in vitro Ty1 gRNA MFE using the Scorer function implemented in RNAStructure. To ensure that compared structures have identical sequence, the adenine at position 58 in the frameshift mutant Ty1fs was manually deleted from cytoplasmic and in vitro Ty1 MFE structures. For identified lowSS regions, sensitivity and PPV values were computed relative to the nuclear Ty1 gRNA MFE. If the region of interest holds only a half of the predicted base pairings, nucleotides from the disrupted helix were added manually to the calculations. SHAPE reactivities for Ty1 gRNA in virio were reported previously [36]. For comparative analysis of nuclear and in virio Ty1 gRNAs, deconvolution of in virio SHAPE reactivities was performed by multiplying them by 0.6, the ratio between SHAPE reactivity medians for nuclear and in virio gRNAs.

2.8. Computing Regions of Large Absolute SHAPE Reactivities Changes

The absolute value of the SHAPE reactivity difference between nuclear and cytoplasmic compartments was summed over 75-nt sliding windows. Spans of at least 100 consecutive nucleotides in which absolute reactivity changes were above the global median change were characterized as regions with the largest SHAPE reactivity differences.

3. Results

3.1. NMIA Successfully Modifies RNA in the Yeast Cell Nucleus

SHAPE mapping using an appropriate electrophile to acylate the ribose 2′-hydroxyl groups efficiently interrogates the flexibility of local RNA structures in vitro and in vivo [37]. Several well-validated SHAPE reagents with different cell permeabilities, half-lives, and nucleotide biases are currently used for RNA structure mapping in cells [38–40]. In the case of Ty1 gRNA, all prior structural studies were performed using N-methylisatoic anhydride (NMIA) [15,22,24,36,41,42]. For structural comparisons, we also used NMIA to determine the secondary structure of nuclear Ty1 gRNA. Application of the same probing reagent allowed more reliable comparisons of Ty1 gRNA structural transitions between cell compartments or biological states. Recently, we showed that NMIA penetrates yeast cells and effectively modifies cytoplasmic transcripts [24], but its capacity to modify nuclear RNA was not investigated. To that end, we performed structural mapping of the U1 small nuclear RNA (snRNA) in the well-characterized *S. cerevisiae* strain BY4742 using NMIA and 2-methylnicotinic acid imidazolide (NAI)—SHAPE reagent previously shown to be suitable for RNA probing inside the nucleus of mammalian cells [40]. Using capillary electrophoresis to detect modification-induced reverse transcriptase-stops, we obtained high and comparable signal-to-background (S/B) ratios for both SHAPE reagents, showing that NMIA robustly modifies nuclear RNAs in yeast with efficiency similar to NAI (Figure 1A). The correlation between position-dependent NMIA and NAI reactivities in U1 snRNA (r = 0.77) showed that there is a strong similarity in the modification patterns (Figure 1B), but some nucleotide biases were also observed.
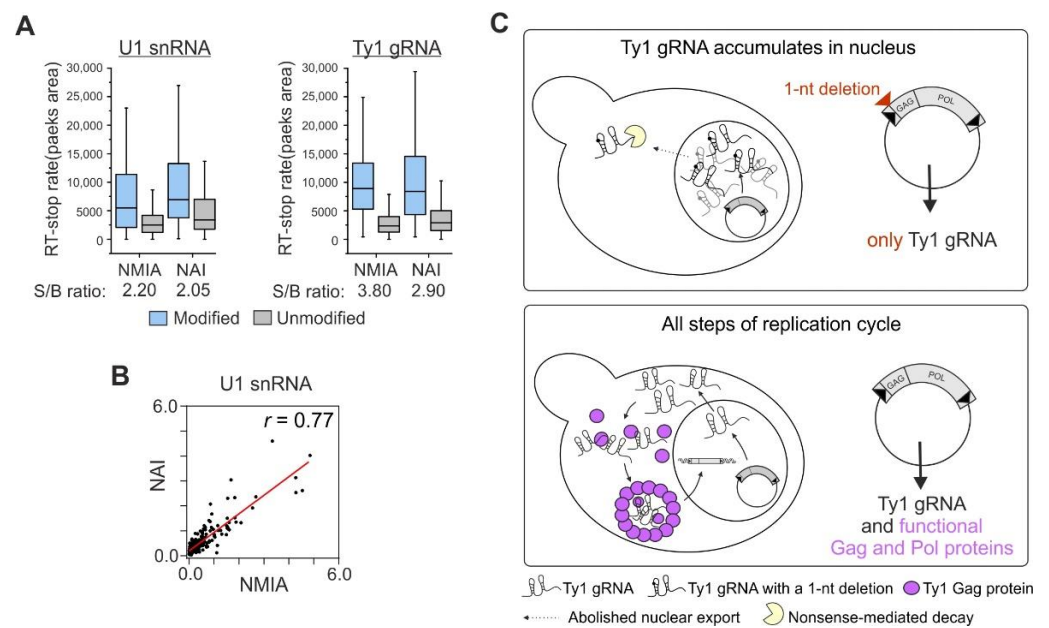
Figure 1. Yeast strains used for Ty1 gRNA structural studies and analysis of the ability of NMIA to modify yeast nuclear RNA. (**A**) Box plot analysis of RT-stop rate measurement with medians for signal and background for U1 snRNA and Ty1 gRNA. Plots present data for approximately 340 nts and 400 nts of U1 snRNA and Ty1 gRNA, respectively. (**B**) Correlation of position-dependent NMIA and NAI reactivities for U1 snRNA (data for 230 nts). (**C**) Comparison of DG3408 (pGTy1fs) and DG3412 (pGTy1 wild type) yeast strains.

3.2. The Experimental Strategy Used to Explore Ty1 gRNA Structure in the Nucleus

To determine the structure of Ty1 gRNA in the yeast nucleus, we used a previously-characterized yeast strain—DG3408—that naturally lacks chromosomal Ty1 elements, and expression of Ty1 occurs exclusively from a galactose inducible plasmid pGTy1fs (pBDG1112, pGTy1gag-ATGfs) [13] (Figure 1C). Ty1fs contains a single "A" deletion adjacent to the GAG start codon, which profoundly affects Ty1 gRNA localization in the cell. After galactose induction, pGTy1gag-ATGfs produces full-length Ty1fs transcripts (5651 nts) but, as expected for a frame-shift mutation, does not synthesize complete Gag and Pol proteins. Due to the lack of functional Gag, Ty1 gRNA accumulates in the nucleus, its cytoplasmic level and stability are greatly decreased, and retrosomes and VLPs are not detected [13]. This strain enabled us to obtain a more homogenous pool of correctly initiated nuclear Ty1 transcripts that were subjected to SHAPE analysis using NMIA. We obtained high-quality SHAPE data (S/B = 3.8) for over 93% of the nucleotides in nuclear Ty1 gRNA, further demonstrating that NMIA is suitable for mapping RNA structure in the yeast nucleus (Figure 1A).

To investigate how the structure of LTR-retrotransposon gRNA changes during replication in host cells, we performed a detailed comparison of SHAPE data for the nuclear and cytoplasmic Ty1 gRNA states [24]. We can exclude the impact of strain-specific cellular factors on gRNA folding because yeast strains applied here and for cytoplasmic Ty1 gRNA studies are based on the same parental yeast strain (DG1768) [43] and differ only by the point mutation in the pGTy1fs plasmid. In contrast, wild type pGTy1 transcripts expressed in DG3412 are efficiently exported from the nucleus, the majority of gRNA is in the cytoplasmic compartment, and small fractions are detected in the nucleus or in VLPs (Figure 1C). We also compared the structure of nuclear Ty1fs gRNA to those determined in VLPs or under in vitro conditions [24,36].

3.3. Analysis of SHAPE Structural Data for Nuclear Ty1 Grna

We compared the overall SHAPE reactivity distribution in different Ty1 gRNA states (Figure 2A). The distribution of reactivity in the nuclear RNA was more similar to the cytoplasmic state than the in vitro or in virio (within a VLP) states [24,36]. Accordingly, the median SHAPE reactivity of the nuclear Ty1 gRNA was comparable to cytoplasmic gRNA (0.45 and 0.42, respectively), and they were higher than in vitro gRNA (0.35) or in VLPs (0.29) (Figure 2A). Analysis of the percentage reactivity distributions showed that ~11% of nucleotides displayed a high SHAPE reactivity (>0.85), 43% were intermediate (0.4–0.85), and the remaining 39% were unreactive (< 0.4) in the nuclear state (Figure 2B). This SHAPE reactivity content is comparable to the cytoplasmic Ty1 gRNA but different from gRNA transcribed and folded in vitro. Nucleotides with a low SHAPE reactivity are likely base-paired, and highly reactive nucleotides are placed in single-stranded RNA regions, whereas those with intermediate SHAPE reactivity can be involved in alternate pairing. Together, these results suggest that independent of the compartment, Ty1 gRNA is less structured than after packaging into VLPs or under in vitro conditions.
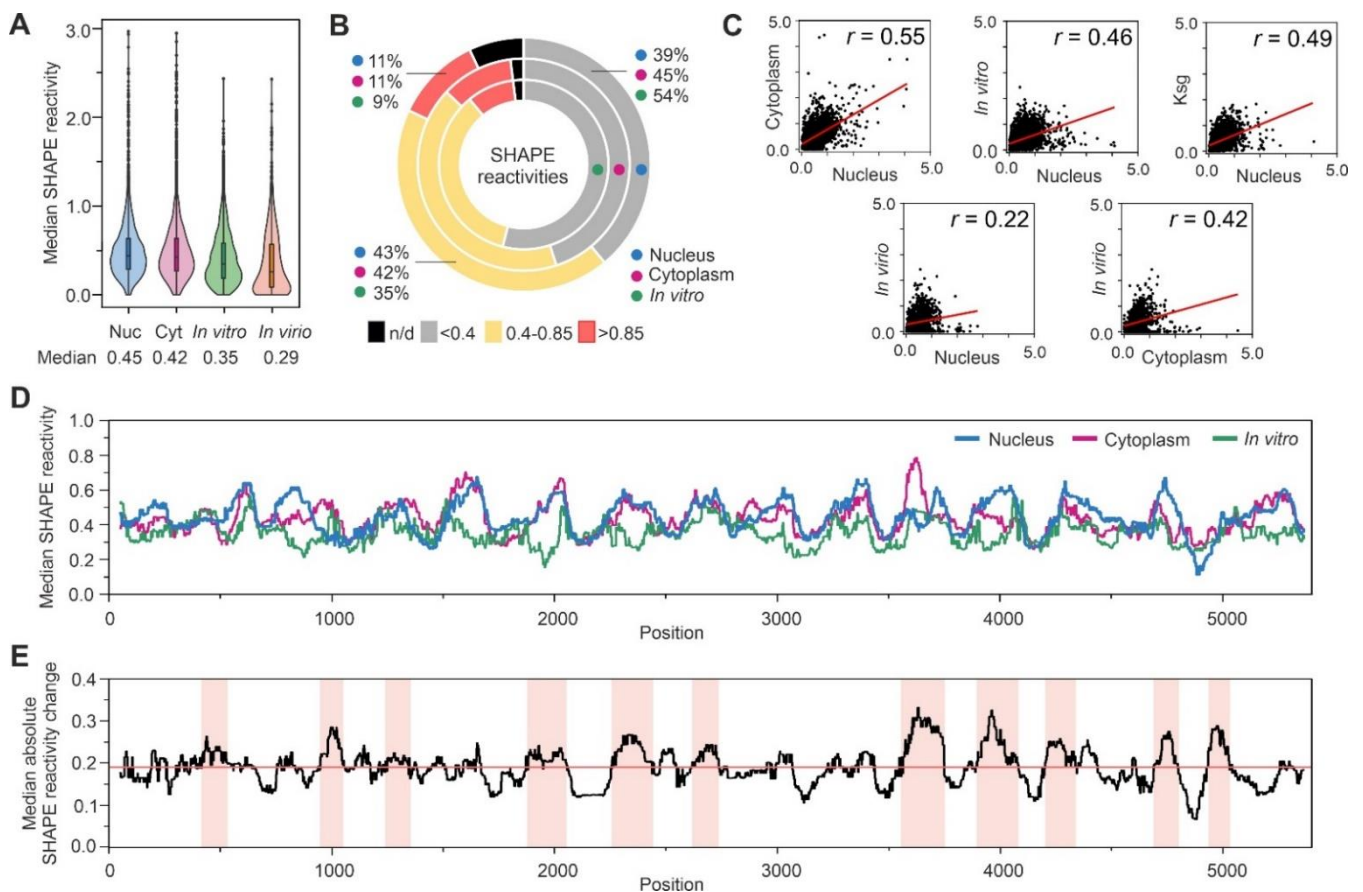
Figure 2. Analysis of SHAPE reactivities for nuclear Ty1 gRNA compared to cytoplasmic and in vitro full-length Ty1 gRNAs, the inhibition of translation initiation state (for 2500 nts), and the **in virio** state (for 1482 nts). (**A**) Violin plot analysis of SHAPE reactivity distributions with medians. (**B**) Pie chart analysis of SHAPE reactivity percentage distribution. (**C**) Correlation of position-dependent NMIA reactivities. (**D**) The median SHAPE reactivity profiles smoothed within a 75-nt window. (**E**) The median profile of absolute SHAPE reactivity change between nuclear and cytoplasmic datasets, smoothed within a 75-nt window. Light-red shadings indicate regions with the greatest absolute differences above the global median change (marked in red) (see Section 2).

To follow Ty1 gRNA structural changes between cell compartments, we calculated the position-dependent correlation between their SHAPE reactivities (Figure 2C). We

found that this correlation was moderate (r = 0.55), suggesting that significant Ty1 gRNA structure remodeling occurred after nuclear export. The reactivity correlation between nuclear and VLPs states was even lower (r = 0.22), implying increased structural differences. Interestingly, the correlation between reactivities in the cytoplasm and VLPs was 0.42 and suggested a slightly higher similarity in Ty1 gRNA structure than that observed between nuclear and in virio states. In an attempt to identify factors contributing to the differences observed between nuclear and cytoplasmic Ty1 gRNA structures, we compared SHAPE data for nuclear Ty1 gRNA with those obtained in cells partially inhibited for translation by kasugamycin treatment [24]. However, we did not detect an increase in the position-dependent correlation between SHAPE reactivities that would indicate a higher structural similarity between nuclear and cytoplasmic Ty1 gRNA in the presence of kasugamycin, which blocks translation initiation [44] (Figure 2C). Moreover, we did not detect any increase in correlation when nuclear data were compared to those from the in vitro state.

Analysis of SHAPE reactivity profiles confirmed the substantial differences detected between many regions of the nuclear, cytoplasmic, or in vitro Ty1 gRNA states (Figure 2D). To estimate the range of SHAPE reactivity alterations between compartments, we calculated the absolute reactivity changes between them, and we searched for Ty1 gRNA regions of at least 100 consecutive nucleotides with the absolute change exceeding the median change (Figure 2E). Based on this approach, we detected 11 regions evenly distributed across the *GAG* and *POL* ORFs. Such extensive reactivity alterations support compartment-specific folding of Ty1 gRNA rather than differences in RNA/protein interactions or RNA modification patterns that usually induce more local changes [45].

3.4. The MFE Structure Model of the Nuclear Ty1 RNA Genome

To better understand Ty1 gRNA folding inside the nucleus, we used the SuperFold pipeline to perform SHAPE-directed windowed modeling of the minimum free energy (MFE) RNA structure [32] (Figure 3A). Examination of SHAPE reactivities in single- and double-stranded regions confirmed the agreement between the predicted structural model and the SHAPE data (Figure 3B). By comparing the nuclear MFE structure to that of cytoplasmic Ty1 gRNA in terms of sensitivity (sens) and positive predictive value (PPV), we found that about 60% of MFE bps predicted for the nuclear state also appeared in the cytoplasmic state (Figure 3A). Most of these shared bps represented short-range interactions, whereas the long-distance pairings seemed unique for each compartment. About 46% of nucleotides are involved in base-pairing in both MFE structure models (Figure 3C), and we did not find significant differences in the median pairing distances or numbers of long-distance pairings (Figure 3D).

To consider the complete ensemble of possible Ty1 gRNA structures in the nucleus, we calculated the probability of each base-pairing in the MFE model (Figure 3A). Highly probable base pairs (HP bps; >80% of probability) are the most likely to be present in all coexisting structural conformers of an RNA molecule, whereas those with a low probability are much rarer. We identified 407 HP bps, which constitute 31% of all MFE bps (Figure 3F). For comparison, about 100 more HP bps have been identified in the cytoplasmic Ty1 gRNA, and they constituted 41% of all MFE bps [24]. Interestingly, we did not observe a tendency to preserve HP bps between compartments (63% of shared HP bps) compared to other bps. Also, nuclear Ty1 gRNA was depleted for long-range HP bps (>100 nt) (Figure 3A). The median HP pairing distances in both compartments remained similar (20 and 22, respectively), and they were much lower than the median distances calculated for MFE bps (Figure 3E). In agreement with the above findings, the median Shannon entropy of nuclear Ty1 gRNA was higher than that calculated for the cytoplasmic state (0.13 and 0.10, respectively) (Figure 3H) [24]. Taken together, our data suggests that nuclear Ty1 gRNA is less structured and in a more heterogenous state.

Figure 3. SHAPE-directed structure model of the nuclear Ty1 gRNA and its comparison with MFE structures predicted for cytoplasmic or in vitro Ty1 gRNAs. (**A**) Genomic organization of Ty1 gRNA. (**B**) Box plot analysis of SHAPE reactivities mapped to single- and double-stranded regions of the nuclear Ty1 gRNA MFE structure. (**C**) Pie chart analysis of percentage distributions of nucleotide pairings in the MFE structures, including high probability base pairs (HP bps, pairing probability > 80%). (**D**) Box plot analysis of MFE bps and (**E**) HP bps distance with medians. (**F**) Venn diagram showing overlap of HP bps. (**G**) Sensitivity and PPV parameters for lowSS regions identified in nuclear Ty1 gRNA. (**H**) LowSS regions analysis. Light shadings indicate extending to encompass entire intersecting helices from MFE structures. The bottom bars present locations of lowSS regions in the cytoplasmic Ty1 gRNA. Overlapped regions are marked by *. Significance was computed by unpaired two-tailed Mann–Whitney test; **** $p < 0.0001$. * $p < 0.05$; n.s., not significant.

To better understand the structural heterogeneity of nuclear Ty1 gRNA, we determined the most compact and thermodynamically stable RNA regions. Based on low SHAPE reactivity and Shannon entropy values (lowSS framework) [46], we identified 11 well-structured regions that cover about 19% of Ty1 gRNA in the nucleus (Figure 3H). We found that most overlap, at least partially, with lowSS regions detected in cytoplasmic Ty1 gRNA. For most lowSS regions, the PPV and sensitivity metrics are higher than for the entire MFE structure, indicating that stable, highly structured motifs tend to be preserved after the export of Ty1 gRNA from the nucleus to the cytoplasm (Figure 3G).

We also compared the nuclear Ty1 gRNA MFE structure to that determined under defined in vitro conditions [24]. Calculated PPV and sensitivity parameters were 51% and 47%, respectively, indicating that only about half of MFE bps are shared between the nuclear and the in vitro Ty1 gRNA structure models (Figure S1). Even though the percentage of nucleotides engaged in base-pairings in both structures was similar, the nuclear MFE model has 535 fewer HP bps than the in vitro model (Figure 3C,F). Consistently, the median Shannon entropy for gRNA in the nucleus was much higher than in vitro ($\Delta_{Shannon}$ = 0.09). Based on the lowSS approach, we were able to identify 4 lowSS regions that were partially shared in these states of Ty1 gRNA (Figure S1). Therefore, our results further support the idea that there are substantial differences in Ty1 gRNA folding in vivo and in vitro.

3.5. RNA–RNA Interactions Mediated by Gag In Vivo

The 5′ terminus of Ty1 gRNA contains well-characterized cis-acting sequences required for the efficient gene expression and propagation of the element. These include the following motifs: PBS, BOX0, and BOX1 necessary for tRNA$_i^{Met}$ binding; CYC5 involved in genome cyclization; and palindromes PAL1 and PAL2 required for Ty1 gRNA dimerization [9,11] (Figure 4). In vitro analyses suggest that these crucial RNA–RNA interactions are promoted by Ty1 Gag [14,15,17]. However, Ty1 Gag's role during retrotransposition has not been demonstrated. To address this issue based on changes in Ty1 gRNA structure, we compared SHAPE data obtained in vivo in the absence of Gag to that from VLPs and the cytoplasm, where Ty1 gRNA is dimerized, cyclized, and tRNA is annealed, and to the in vitro state, where Gag-mediated RNA–RNA interactions do not occur [24,36].

In the absence of tRNA$_i^{Met}$ in vitro, the PBS is unreactive due to an intramolecular interaction with CYC5, and segments of BOX0 and BOX1 remain unpaired and accessible to SHAPE modification [36]. Although the PBS remains unreactive independent of primer binding, SHAPE can detect disruption of helical regions adjacent to PBS resulting from tRNA$_i^{Met}$ hybridization [24,36]. Therefore, increased reactivity in these regions can serve as an indicator for tRNA$_i^{Met}$ binding. In the nucleus, we did not detect increases in reactivity in the region proceeding PBS or in the 3-nt sequence linking PBS and BOX0 (Figure 4A,B—black frames). Also, we observed increased reactivity in BOX0 and BOX1. These data do not support Ty1 gRNA/tRNA$_i^{Met}$ complex formation in the nucleus and strongly agree with other studies suggesting that complex formation occurs in the cytoplasm [47–49] (Figure 4B). Our results also argue against Ty1 gRNA cyclization occurring in the nucleus, as engagement of CYC5 in interaction with CYC3 would lead to increased reactivity of the PBS when tRNAiMet is not annealed.

Ty1 gRNA dimerization is mediated by cis-acting palindromic sequences PAL1 and PAL2 [15,36]. In the dimeric state, these sequences remain unreactive due to Gag-mediated PAL1/PAL2 intermolecular interactions. In monomeric gRNA, PALs are also unreactive due to their intramolecular interaction, leading to stem-loop formation. These two distinct interactions can be distinguished based on reactivity changes in nucleotides linking PAL1 and PAL2 (21UCU23). For nuclear Ty1 gRNA, we observed 21UCU23 reactivity similar to that in the dimeric state; however, we also detected increased SHAPE-reactivity in the PALs (Figure 4A). Thus, our data suggest that in the absence of Gag, PAL1/PAL2 intermolecular duplexes are not formed, and the stem-loop with PALs is labile in the Ty1 gRNA monomer (Figure 4B).

In VLPs and the cytoplasmic state, the 5′end of Ty1 gRNA forms a functionally important H-type pseudoknot (PK), where nucleotides at positions 1–7 and 256–262 form base pairs with nucleotides 264–270 and 319–325, respectively [36,41]. The PK has not been detected under in vitro conditions [24,36]. With nuclear Ty1 gRNA, we observed slightly increased reactivity in the sequences involved in PK formation when compared to VLPs (Figure 4A). Also, we did not detect this important PK interaction in nuclear Ty1 gRNA using ShapeKnot software and nuclear SHAPE data as constraints (Figure 4B).
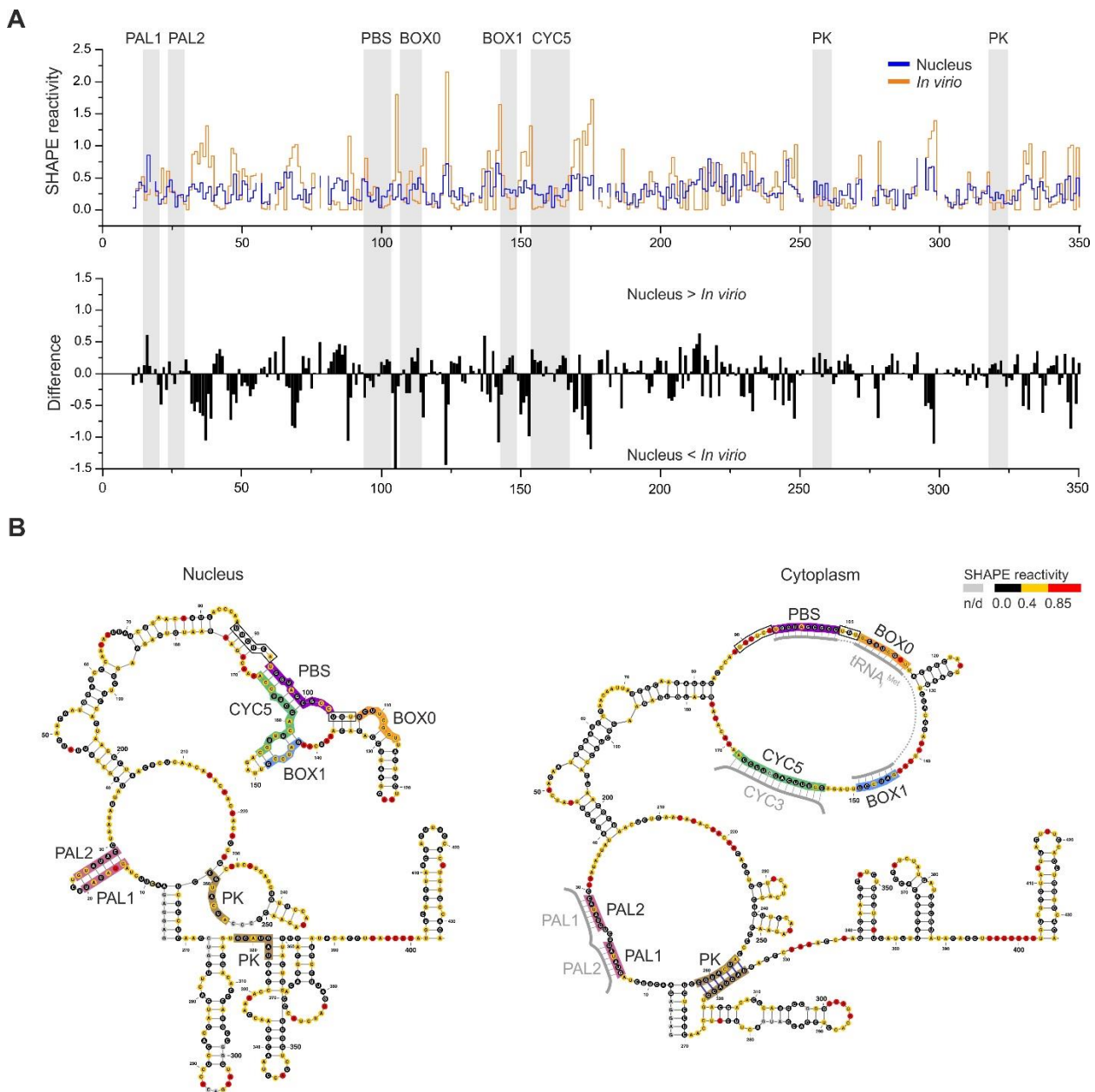
Figure 4. Comparative analysis of Ty1 gRNA regions containing cis-acting sequences. (**A**) The step plot (top) and the difference plot (down) calculated by subtracting the in virio NMIA reactivities from those of the nuclear Ty1 gRNA. (**B**) The MFE models of nuclear and cytoplasmic Ty1 gRNA 5′ end. Cis-acting sequences are marked with colored boxes. Neighboring regions affected by tRNA$_i$^Met annealing are boxed.

4. Discussion

Due to the rapid development of RNA probing techniques, the structures of several viral RNAs in distinct experimental conditions have been resolved, e.g., [50–54]. These studies reveal meaningful differences between RNA folding in vitro and in cells and highlight the importance of studying RNA structures in their native cellular context. However, they provide an averaged "steady-state" picture of viral RNA in the cell that may fail to resolve functionally important structural transitions during replication. Consequently, little is known about whether and how the structure of viral RNAs change in specific cellular compartments.

Here, we took advantage of a well-developed Ty1 experimental system to resolve the first structure of a retroviral-like RNA in the nucleus. To avoid problems concerning inefficient purification of intact yeast nuclei and the low level of wild type nuclear Ty1 RNA, we utilized a genetically modified element that does not produce Ty1 Gag and accumulates GAL1-promoted Ty1fs RNA in the nucleus [13]. In the absence of Ty1 Gag, Ty1fs RNA that enters the cytoplasm is subjected to robust RNA-turnover pathways including nonsense-mediated decay and degradation by P-body components. We assumed that only a very small, if any, fraction of cytoplasmic Ty1fs RNA could be mapped together with nuclear Ty1fs RNA and strongly believe that the modification signals from this small cytoplasmic fraction do not significantly interfere with our data. Through a direct comparison of SHAPE-derived nuclear Ty1 gRNA structure with those available for Ty1 gRNA in the cytoplasm, within purified VLPs, or synthesized and folded in vitro [24,36], we characterized gRNA structural alterations that occur during Ty1 retrotransposition. Finally, we confirm and extend the role of Gag during RNA–RNA interactions crucial to Ty1 replication in budding yeast.

In support of our previous findings that the cellular environment greatly impacts Ty1 RNA architecture [24], we observe significant structural differences between nuclear and in vitro Ty1 gRNAs. We also discovered large regions with extensive structural alterations between nuclear and cytoplasmic Ty1 gRNAs, indicating compartment-specific folding. mRNA structural changes between compartments have been identified by transcriptome-wide study in the mammalian cells, but in contrast to our observations, most are relatively small and mainly local, suggesting that RNA structures formed during transcription remain largely unchanged during the mRNA life span [45]. However, transcriptome-wide analyses may not reveal the extent of all structural variations for individual RNAs. Thus, Ty1 gRNA is an example of a transcript sensitive to the local nuclear and cytoplasmic environments, and its structure undergoes significant remodeling during the transition between cell compartments. Based on base-pairing probabilities and Shannon entropy calculations, we show that Ty1 gRNA inside the nucleus adopts a more heterogeneous structure than in the cytoplasm. Interestingly, we show that short-range interactions represent the majority of nuclear HP bps, whereas long-range HP bps likely form once Ty1 gRNA exits the nucleus. Our observations are consistent with previous findings regarding cotranscriptional RNA folding, suggesting that short-range base pairs are formed immediately after transcription, whereas long-range interactions are more likely to form transitory, less-probable structures, which refold during RNA maturation and transition between cell compartments [55–57].

Although Ty1 gRNA displays a higher degree of structural heterogeneity in the nucleus, we identified 11 thermodynamically stable regions (lowSS regions) that significantly overlap with those in the cytoplasm [24]. Prior studies show that such structurally stable motifs are often overrepresented with functionally important cis-acting sequences [46,53,58–60]. Most of the cis-acting Ty1 sequences are also placed in lowSS regions of the gRNA, but the functions of other stable regions remain unknown. Our observations suggest that gRNA regions likely necessary for Ty1 retrotransposition adopt well-defined structures during or just after transcription, and they remain unchanged across compartments. However, we also noticed that several nuclear lowSS regions are no longer detected in the cytoplasm where new lowSS regions are detected. Therefore, our results also support a recent study suggesting that some functionally important RNA structural motifs fold differently in the nucleus and cytoplasm [61].

The exact mechanisms by which RNA structures are modulated in vivo are not well understood. While some studies suggest that translation is the primary force responsible for the RNA structural remodeling, others highlight the importance of RNA binding proteins or RNA modifications [45,62]. We previously found that the ribosomes are an important factor shaping the Ty1 RNA structure in the cytoplasm [24]. However, here we show that active translation does not explain Ty1 gRNA structural alterations in different cellular compartments. On the other hand, we detected greater structural similarities between Ty1 gRNA in the cytoplasm and VLPs than between nuclear Ty1 gRNA and VLPs. As Ty1

Gag protein displays RNA chaperone activity and is present in the cytoplasm and VLPs, but not in the nucleus, we conclude that Gag may account for the higher gRNA structural similarity between the cytoplasm and VLPs. Analogous to HIV-1 Gag, interactions between Ty1 Gag and retrotransposon gRNA may vary at different stages of retrotransposition and induce diverse structural effects [63]. Although we cannot fully explain a Gag's role in Ty1 gRNA structure remodeling, our comparative analyses of SHAPE data confirmed and extended the critical role for Gag in Ty1 gRNA dimerization, cyclization, and primer tRNA annealing in vivo. Ty1 Gag does not contain a nuclear localization signal and is detected near but not within the nuclei of cells expressing Ty1 [13]. Thus, we show RNA–RNA interactions essential for retrotransposition cannot occur prior to Ty1 gRNA nuclear export to the cytoplasm and interaction with Gag.

Studying RNA structures in cellular compartments is essential to gain a comprehensive view of the complex interplay between RNA structure and function. Our work is the first to present the nuclear structure of a viral-like gRNA, although it may be premature to conclude whether compartment-specific gRNA folding is a general feature of LTR-retrotransposons or other reverse-transcribing single-strand RNA viruses. However, additional RNA structural studies across cellular compartments are needed to define the common or specific mechanisms of viral and retrotransposon gRNA structure modulation.

References

1. Eichler, E.E.; Sankoff, D. Structural dynamics of eukaryotic chromosome evolution. *Science* **2003**, *301*, 793–797. [CrossRef] [PubMed]
2. Krupovic, M.; Blomberg, J.; Coffin, J.M.; Dasgupta, I.; Fan, H.; Geering, A.D.; Gifford, R.; Harrach, B.; Hull, R.; Johnson, W.; et al. Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. *J. Virol.* **2018**, *92*, e00515-18. [CrossRef] [PubMed]
3. Malik, H.S.; Henikoff, S.; Eickbush, T.H. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **2000**, *10*, 1307–1318. [CrossRef] [PubMed]
4. Bleykasten-Grosshans, C.; Fabrizio, R.; Friedrich, A.; Schacherer, J. Species-Wide Transposable Element Repertoires Retrace the Evolutionary History of the Saccharomyces cerevisiae Host. *Mol. Biol. Evol.* **2021**, *38*, 4334–4345. [CrossRef]
5. Czaja, W.; Bensasson, D.; Ahn, H.W.; Garfinkel, D.J.; Bergman, C.M. Evolution of Ty1 copy number control in yeast by horizontal transfer and recombination. *PLoS Genet.* **2020**, *16*, e1008632. [CrossRef]
6. Boeke, J.D.; Garfinkel, D.J.; Styles, C.A.; Fink, G.R. Ty elements transpose through an RNA intermediate. *Cell* **1985**, *40*, 491–500. [CrossRef]

7. Mellor, J.; Fulton, S.M.; Dobson, M.J.; Wilson, W.; Kingsman, S.M.; Kingsman, A.J. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. *Nature* **1985**, *313*, 243–246. [CrossRef]

8. Elder, R.T.; Loh, E.Y.; Davis, R.W. Rna from the Yeast Transposable Element Ty1 Has Both Ends in the Direct Repeats, a Structure Similar to Retrovirus Rna. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 2432–2436. [CrossRef]

9. Curcio, M.J.; Lutz, S.; Lesage, P. The Ty1 LTR-retrotransposon of budding yeast, Saccharomyces cerevisiae. *Microbiol. Spectr.* **2015**, *3*(2), 1–35.

10. Sultana, T.; Zamborlini, A.; Cristofari, G.; Lesage, P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.* **2017**, *18*, 292–308. [CrossRef]

11. Pachulska-Wieczorek, K.; Le Grice, S.F.J.; Purzycka, K.J. Determinants of Genomic RNA Encapsidation in the Saccharomyces cerevisiae Long Terminal Repeat Retrotransposons Ty1 and Ty3. *Viruses-Basel.* **2016**, *8*, 193. [CrossRef] [PubMed]

12. Malagon, F.; Jensen, T.H. T-body formation precedes virus-like particle maturation in S-cerevisiae. *RNA Biol.* **2011**, *8*, 184–189. [CrossRef]

13. Checkley, M.A.; Mitchell, J.A.; Eizenstat, L.D.; Lockett, S.J.; Garfinkel, D.J. Ty1 gag enhances the stability and nuclear export of Ty1 mRNA. *Traffic* **2013**, *14*, 57–69. [CrossRef]

14. Nishida, Y.; Pachulska-Wieczorek, K.; Blaszczyk, L.; Saha, A.; Gumna, J.; Garfinkel, D.J.; Purzycka, K.J. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic Acids Res.* **2015**, *43*, 7414–7431. [CrossRef] [PubMed]

15. Gumna, J.; Purzycka, K.J.; Ahn, H.W.; Garfinkel, D.J.; Pachulska-Wieczorek, K. Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biol.* **2019**, *16*, 1749–1763. [CrossRef] [PubMed]

16. Cristofari, G.; Bampi, C.; Wilhelm, M.; Wilhelm, F.X.; Darlix, J.L. A 5′—3′ long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *EMBO J.* **2002**, *21*, 4368–4379. [CrossRef] [PubMed]

17. Cristofari, G.; Ficheux, D.; Darlix, J.L. The Gag-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. *J. Biol. Chem.* **2000**, *275*, 19210–19217. [CrossRef]

18. Malagon, F.; Jensen, T.H. The T body, a new cytoplasmic RNA granule in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **2008**, *28*, 6022–6032. [CrossRef]

19. Nonet, M.; Scafe, C.; Sexton, J.; Young, R. Eukaryotic Rna-Polymerase Conditional Mutant That Rapidly Ceases Messenger-Rna Synthesis. *Mol. Cell. Biol.* **1987**, *7*, 1602–1611. [CrossRef]

20. Curcio, M.J.; Hedge, A.M.; Boeke, J.D.; Garfinkel, D.J. Ty Rna Levels Determine the Spectrum of Retrotransposition Events That Activate Gene-Expression in Saccharomyces-Cerevisiae. *Mol. Gen. Genet.* **1990**, *220*, 213–221. [CrossRef]

21. Friant, S.; Heyman, T.; Bystrom, A.S.; Wilhelm, M.; Wilhelm, F.X. Interactions between Ty1 retrotransposon RNA and the T and D regions of the tRNA(iMet) primer are required for initiation of reverse transcription in vivo. *Mol. Cell. Biol.* **1998**, *18*, 799–806. [CrossRef] [PubMed]

22. Gamache, E.R.; Doh, J.H.; Ritz, J.; Laederach, A.; Bellaousov, S.; Mathews, D.H.; Curcio, M.J. Structure-Function Model for Kissing Loop Interactions That Initiate Dimerization of Ty1 RNA. *Viruses* **2017**, *9*, 93. [CrossRef] [PubMed]

23. Xu, H.; Boeke, J.D. Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: Analysis of mini-Ty1 elements. *Mol. Cell. Biol.* **1990**, *10*, 2695–2702. [CrossRef]

24. Andrzejewska, A.; Zawadzka, M.; Gumna, J.; Garfinkel, D.J.; Pachulska-Wieczorek, K. In vivo structure of the Ty1 retrotransposon RNA genome. *Nucleic Acids Res.* **2021**, *49*, 2878–2893. [CrossRef]

25. Garfinkel, D.J.; Nyswaner, K.; Wang, J.; Cho, J.Y. Post-transcriptional cosuppression of Ty1 retrotransposition. *Genetics* **2003**, *165*, 83–99. [CrossRef] [PubMed]

26. Moore, S.P.; Liti, G.; Stefanisko, K.M.; Nyswaner, K.M.; Chang, C.; Louis, E.J.; Garfinkel, D.J. Analysis of a Ty1-less variant of Saccharomyces paradoxus: The gain and loss of Ty1 elements. *Yeast* **2004**, *21*, 649–660. [CrossRef] [PubMed]

27. Garfinkel, D.J.; Mastrangelo, M.F.; Sanders, N.J.; Shafer, B.K.; Strathern, J.N. Transposon tagging using Ty elements in yeast. *Genetics* **1988**, *120*, 95–108. [CrossRef] [PubMed]

28. Winston, F.; Dollard, C.; Ricupero-Hovasse, S.L. Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. *Yeast* **1995**, *11*, 53–55. [CrossRef]

29. Brachmann, C.B.; Davies, A.; Cost, G.J.; Caputo, E.; Li, J.; Hieter, P.; Boeke, J.D. Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **1998**, *14*, 115–132. [CrossRef]

30. Vasa, S.M.; Guex, N.; Wilkinson, K.A.; Weeks, K.M.; Giddings, M.C. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **2008**, *14*, 1979–1990. [CrossRef]

31. Gumna, J.; Zok, T.; Figurski, K.; Pachulska-Wieczorek, K.; Szachniuk, M. RNAthor—fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis. *PLoS ONE* **2020**, *15*, e0239287.

32. Smola, M.J.; Rice, G.M.; Busan, S.; Siegfried, N.A.; Weeks, K.M. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **2015**, *10*, 1643–1669. [CrossRef]

33. Reuter, J.S.; Mathews, D.H. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinform.* **2010**, *11*, 129. [CrossRef] [PubMed]

34. Darty, K.; Denise, A.; Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **2009**, *25*, 1974–1975. [CrossRef]
35. Hajdin, C.E.; Bellaousov, S.; Huggins, W.; Leonard, C.W.; Mathews, D.H.; Weeks, K.M. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5498–5503. [CrossRef]
36. Purzycka, K.J.; Legiewicz, M.; Matsuda, E.; Eizentstat, L.D.; Lusvarghi, S.; Saha, A.; Le Grice, S.F.J.; Garfinkel, D.J. Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res.* **2013**, *41*, 463–473. [CrossRef] [PubMed]
37. Weeks, K.M. SHAPE Directed Discovery of New Functions in Large RNAs. *Acc. Chem. Res.* **2021**, *54*, 2502–2517. [CrossRef]
38. Marinus, T.; Fessler, A.B.; Ogle, C.A.; Incarnato, D. A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy. *Nucleic Acids Res.* **2021**, *49*, e34. [CrossRef]
39. 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]
40. Lee, B.; Flynn, R.A.; Kadina, A.; Guo, J.K.; Kool, E.T.; Chang, H.Y. Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA* **2017**, *23*, 169–174. [CrossRef]
41. Huang, Q.; Purzycka, K.J.; Lusvarghi, S.; Li, D.H.; LeGrice, S.F.J.; Boeke, J.D. Retrotransposon Ty1 RNA contains a 5′-terminal long-range pseudoknot required for efficient reverse transcription. *RNA* **2013**, *19*, 320–332. [CrossRef] [PubMed]
42. Blaszczyk, L.; Biesiada, M.; Saha, A.; Garfinkel, D.J.; Purzycka, K.J. Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression. *Viruses* **2017**, *9*, 74. [CrossRef]
43. Chen, J.; McQueary, H.; Hall, D.W.; Philippsen, P.; Garfinkel, D.J.; Bergman, C.M. Genome Assembly of the Ty1-Less Saccharomyces paradoxus Strain DG1768. *Microbiol. Resour. Announc.* **2022**, *11*, e0086821. [CrossRef] [PubMed]
44. Schuwirth, B.S.; Day, J.M.; Hua, C.W.; Janssen, G.R.; Dahlberg, A.E.; Cate, J.H.; Vila-Sanjurjo, A. Structural analysis of kasugamycin inhibition of translation. *Nat. Struct. Mol. Biol.* **2006**, *13*, 879–886. [CrossRef] [PubMed]
45. Sun, L.; Fazal, F.M.; Li, P.; Broughton, J.P.; Lee, B.; Tang, L.; Huang, W.Z.; Kool, E.T.; Chang, H.Y.; Zhang, Q.F.C. RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **2019**, *26*, 322–330.
46. Siegfried, N.A.; Busan, S.; Rice, G.M.; Nelson, J.A.E.; Weeks, K.M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **2014**, *11*, 959–965. [CrossRef]
47. Chapman, K.B.; Bystrom, A.S.; Boeke, J.D. Initiator methionine tRNA is essential for Ty1 transposition in yeast. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3236–3240. [CrossRef]
48. Keeney, J.B.; Chapman, K.B.; Lauermann, V.; Voytas, D.F.; Astrom, S.U.; von Pawel-Rammingen, U.; Bystrom, A.; Boeke, J.D. Multiple molecular determinants for retrotransposition in a primer tRNA. *Mol. Cell. Biol.* **1995**, *15*, 217–226. [CrossRef]
49. Kawakami, K.; Pande, S.; Faiola, B.; Moore, D.P.; Boeke, J.D.; Farabaugh, P.J.; Strathern, J.N.; Nakamura, Y.; Garfinkel, D.J. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. *Genetics* **1993**, *135*, 309–320. [CrossRef]
50. 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.Y.; Aw, J.G.; et al. Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Na.t Commun.* **2019**, *10*, 1–13. [CrossRef]
51. Simon, L.M.; Morandi, E.; Luganini, A.; Gribaudo, G.; Martinez-Sobrido, L.; Turner, D.H.; Oliviero, S.; Incarnato, D. In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs. *Nucleic Acids Res.* **2019**, *47*, 7003–7017. [CrossRef] [PubMed]
52. 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. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **2020**, *48*, 12436–12452. [CrossRef] [PubMed]
53. Huston, N.C.; Wan, H.; Strine, M.S.; Tavares, R.D.A.; Wilen, C.B.; Pyle, A.M. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell.* **2021**, *81*, 584–598.e5. [CrossRef] [PubMed]
54. Wan, H.; Adams, R.L.; Lindenbach, B.D.; Pyle, A.M. The In Vivo and In Vitro Architecture of the Hepatitis C Virus RNA Genome Uncovers Functional RNA Secondary and Tertiary Structures. *J. Virol.* **2022**, *96*, e0194621. [CrossRef] [PubMed]
55. Herschlag, D.; Bonilla, S.; Bisaria, N. The Story of RNA Folding, as Told in Epochs. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*. [CrossRef] [PubMed]
56. Lai, D.; Proctor, J.R.; Meyer, I.M. On the importance of cotranscriptional RNA structure formation. *RNA* **2013**, *19*, 1461–1473. [CrossRef] [PubMed]
57. Incarnato, D.; Morandi, E.; Anselmi, F.; Simon, L.M.; Basile, G.; Oliviero, S. In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding. *Nucleic Acids Res.* **2017**, *45*, 9716–9725. [CrossRef]
58. Mauger, D.M.; Golden, M.; Yamane, D.; Williford, S.; Lemon, S.M.; Martin, D.P.; Weeks, K.M. Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3692–3697. [CrossRef]
59. 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]
60. 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. Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11513–11518. [CrossRef]

61. Liu, Z.S.; Liu, Q.; Yang, X.F.; Zhang, Y.Y.; Norris, M.; Chen, X.X.; Cheema, J.; Zhang, H.K.; Ding, Y.L. In vivo nuclear RNA structurome reveals RNA-structure regulation of mRNA processing in plants. *Genome Biol.* **2021**, *22*, 1–22. [CrossRef] [PubMed]
62. 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] [PubMed]
63. Kutluay, S.B.; Zang, T.; Blanco-Melo, D.; Powell, C.; Jannain, D.; Errando, M.; Bieniasz, P.D. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* **2014**, *159*, 1096–1109. [CrossRef] [PubMed]

SUPPLEMENTARY TABLES AND FIGURES

Cell compartment-specific folding of Ty1 retrotransposon RNA genome

Małgorzata Zawadzka[1], Angelika Andrzejewska-Romanowska[1], Julita Gumna[1], David J.

Garfinkel[2] and Katarzyna Pachulska-Wieczorek[1,*]

Table S1 . Reverse transcription primers.

PRIMER	SEQUENCE (5'-3')
Ty1 PR_1	TCAGGTGATGGAGTGCTCAG
Ty1 PR_2	TCGTTTGCCTCTTGGGTATC
Ty1 PR_3	CATCACTCGGATTTCTCCTG
Ty1 PR_4	TTGGTGTTGTCCTGGAAGTGA
Ty1 PR_5	TCGAATTGTCTGTGCATTGG
Ty1 PR_6	AGACACTGGCCTGAAACTGG
Ty1 PR_7	GCCACGAGGATGTATTTTGG
Ty1 PR_8	GGAGGTGTGGAATCGGTTGG
Ty1 PR_9	GGTACGTTTGTATGATTAGTCTCA
Ty1 PR_10	GTCTCGTGATACCTTAATTTCAG
Ty1 PR_11	TGCTGAATATCACCTCTTGC
Ty1 PR_12	CAAATTGTCACTTGACTG
Ty1 PR_13	TGCATTTCATGTACCTTCTC
Ty1 PR_14	GCTTCCGTAGTTGAAGTACAT
Ty1 PR_15	TCTATTCCAACATACCACCC
Ty1 PR_16	CTCCTCAAGGATTTAGGAATCC
Ty1 PR_17	CATTGTTGATAAAGGCTATA
U1 PR	TAAGATCCACCCGTTCTAC
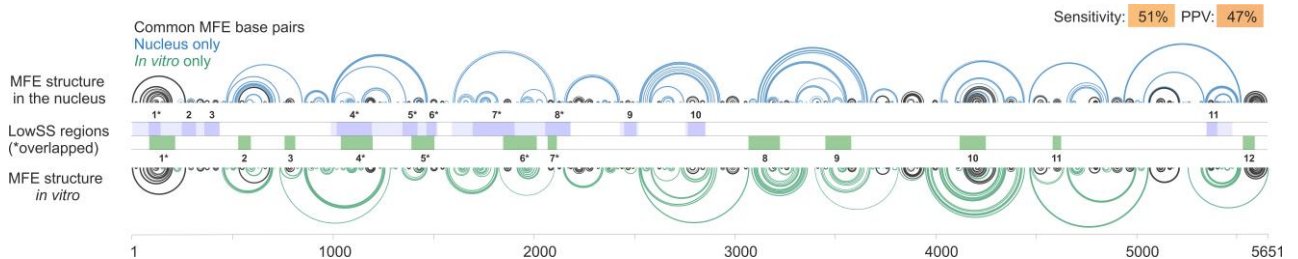
Figure S1. Comparison of SHAPE-directed MFE model predicted for nuclear and Ty1 gRNA *in vitro*. Location of lowSS regions identified for nuclear and in vitro Ty1 gRNA are presented. Overlapped regions are marked by asterix *.

Dodatek *Supplementary Data* do pracy naukowej: M. Zawadzka, A. Andrzejewska, J. Gumna, D.J. Garfinkel, K. Pachulska-Wieczorek; *Cell Compartment-Specific Folding of Ty1 Long Terminal Repeat Retrotransposon RNA Genome*; Viruses, 2022, zawierający średnie wartości reaktywności SHAPE dla poszczególnych nukleotydów badanych RNA dostępny jest pod adresem:

https://www.mdpi.com/1999-4915/14/9/2007#supplementary

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.

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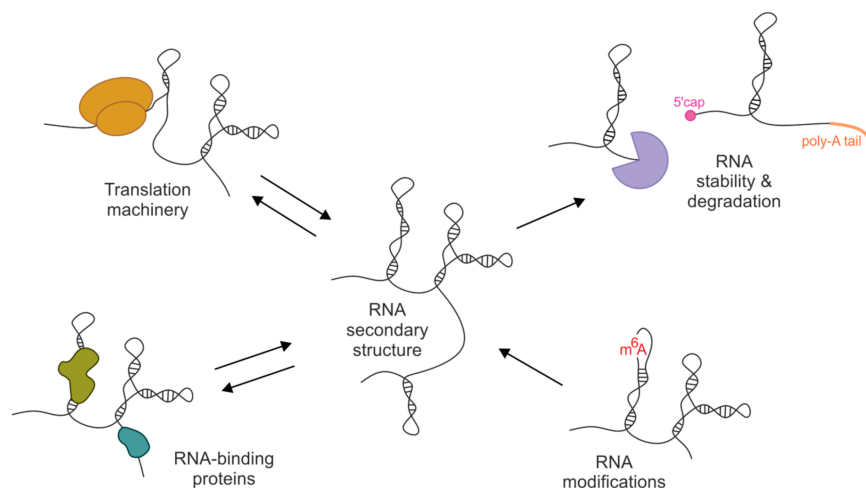
Figure 1. Schematic view on the relationship between the RNA secondary structure and cellular environment elements.

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 to 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 synthetized 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], *E. coli* [40]	DMS	RT-stop	In vivo and in vitro
SHAPE-Structure-seq	*A. thaliana* [41]	NAI	RT-stop	In vivo and in vitro
Structure-seq	*A. thaliana* [8], rice [42]	DMS	RT-stop	In vivo
SHAPE-MaP	*E. coli* [43]	1M7	RT-mutate	In vivo and in vitro
SPLASH	Human and yeast [44]	Biotinylated psoralen	Mapping of ligated junctions	In vivo
PARS	*E. coli* [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]	*E. coli*	The unstructured sequence upstream of the start codon is a general feature of *E. coli* genes and is positively correlated with gene expression.
Mustoe et al., 2018 [43]	*E. coli*	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]	*E. coli*	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^6A 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^6A 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]	*A. thaliana*	Less structured regions immediately upstream the start codon region facilitate ribosome binding and increase TE.
Liu et al., 2019 [41]	*A. thaliana*	Nuclear mRNAs fold differently from cytosolic mRNAs.
Sanchez De Groot et al., 2019 [53]	Various	Highly structured RNAs bind a large amount of proteins.
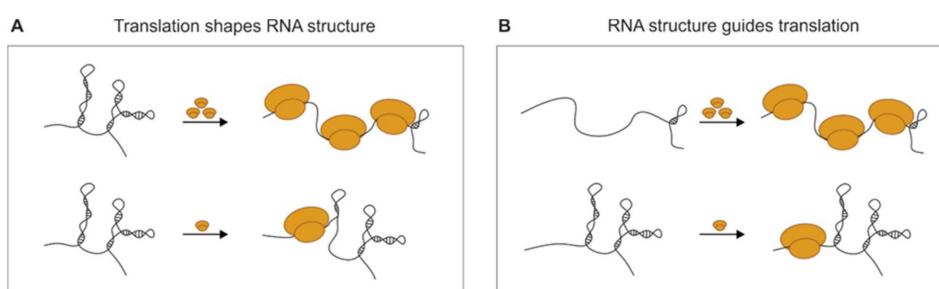
Figure 2. Schematic representation of the interplay between the translation and RNA structure. (**A**) Translation shapes the RNA structure in the cells by ribosome helicase activity and (**B**) the RNA structure guides translation by modulating ribosome binding.

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 a 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 structuromes 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 m^6A 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 N^6-methyladenosine (m^6A). Thus, it is not surprising that m^6A modification is of particular interest in research. Development of an m^6A RNA immunoprecipitation approach followed by high-throughput sequencing (MeRIP–seq) allowed to study m^6A modification landscape in a transcriptome-wide manner [98,99]. These studies identify over 12,000 m^6A sites, mainly in the context of the sequence GGm^6ACU, in more than 7000 human transcripts. The sites of m^6A 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 m^6A 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 m^6A influences the RNA structure because of the rotation of the methlyloamino 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 m^6A 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 m^6A modification sites, with a strong tendency for unwinding RNA secondary structure [51]. Therefore, the appearance of m^6A 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 m^6A-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.

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 Siederdissen, 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.; Ozsolak, 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.; Aeschimann, 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.; Błaszczyk, 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 Epistructurome: 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]
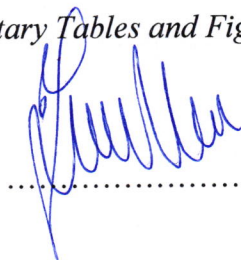
OŚWIADCZENIA AUTORA

Poznań, 14.09.2002

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:

- Zaplanowaniu przebiegu prac badawczych pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek,

- otrzymaniu transkryptu genomowego RNA Ty1 metodą transkrypcji *in vitro*

- Wykonaniu eksperymentów próbkowania strukturalnego Ty1 RNA *in vitro* metodą SHAPE-CE

- Współudziale w przeprowadzeniu eksperymentów próbkowania strukturalnego RNA w komórkach *E.coli* metodą SHAPE-CE

- Normalizacji i analizie surowych danych otrzymanych z przeprowadzonych przeze mnie eksperymentów SHAPE

- Otrzymaniu modelu struktury drugorzędowej Ty1 RNA *in vitro* przy zastosowaniu oprogramowania SuperFold oraz jej wizualizacji przy użyciu aplikacji VARNA

- Współudziale w przeprowadzeniu analizy statystycznej i interpretacji wyników pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek

- Przygotowaniu wraz z mgr inż. Angeliką Andrzejewską-Romanowską pierwszej wersji manuskryptu oraz jego finalnej wersji pod kierunkiem dr hab. K. Pachulskiej-Wieczorek i prof. D. J. Garfinkel

- Przygotowaniu *Supplementary Dataset*

- Pomocy w przygotowaniu figur, tabel oraz *Supplementary Tables and Figures*

[§]Autor korespondencyjny

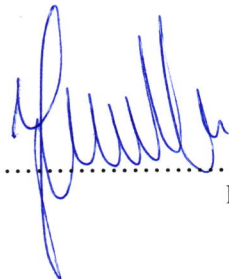..
Podpis

Poznań, 14.09.2002

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

Małgorzata Zawadzka, Angelika Andrzejewska, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek[§]. **Cell compartment-specific folding of Ty1 Long Terminal Repeat retrotransposon RNA genome**. *Viruses* , 2022; 14(9):2007

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

- Zaplanowaniu wspólnie z dr hab. Katarzyną Pachulską-Wieczorek przebiegu prac badawczych,

- Optymalizacji warunków i przeprowadzeniu hodowli drożdży *S. paradoxus* szczep DG3408 i *S. cerevisiae* szczep BY4742

- Wykonaniu eksperymentów próbkowania strukturalnego jądrowego Ty1 RNA w drożdżach (DG3408) metodą SHAPE-CE

- Wykonaniu eksperymentów próbkowania strukturalnego U1 snRNA w drożdżach (BY4742) metodą SHAPE-CE

- Normalizacji i analizie surowych danych otrzymanych ze wszystkich eksperymentów

- Otrzymaniu modelu struktury drugorzędowej jądrowego Ty1 RNA przy zastosowaniu oprogramowania SuperFold oraz jej wizualizacji przy użyciu aplikacji VARNA

- Przeprowadzeniu analizy statystycznej i wraz z mgr inż. Angeliką Andrzejewską-Romanowską interpretacji wyników pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek

- Przygotowaniu, pod kierunkiem dr hab. Katarzyną Pachulską-Wieczorek, pierwszej wersji manuskryptu oraz jego finalnej wersji wspólnie z pozostałymi współautorami

- Przygotowaniu *Supplementary Data* oraz *Supplementary File*

[§]Autor korespondencyjny

....
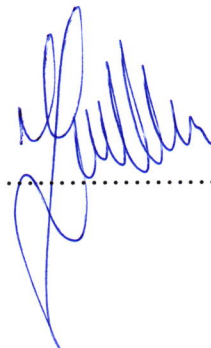Podpis

Poznań, 14.09.2002

**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),
6670

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

- Zgromadzeniu i analizie stosownej literatury
- Przygotowaniu pierwszej wersji rozdziału *RNA Structure in Relation to RNA Stability and Degradation* oraz *Impact of RNA Modifications on RNA Structure In Vivo*
- Współtworzeniu rozdziału *The Relationship between RNA Structure and Proteins Binding*
- Współuczestniczeniu w wykonaniu tabel
- Współuczestniczeniu w przygotowaniu pozostałych części manuskryptu, jego edycji oraz opracowaniu jego finalnej wersji pod kierunkiem dr hab. Katarzyny Pachulskiej-Wieczorek

[§]Autor korespondencyjny

Podpis

OŚWIADCZENIA AUTORA KORESPONDUJĄCEGO

Poznań, 14.09.2002

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 Małgorzaty Zawadzkiej w powstanie wyżej wymienionej pracy naukowej polegał na:

- współudziale w przygotowaniu planu badań,
- otrzymaniu transkryptu genomowego RNA Ty1 metodą transkrypcji *in vitro*
- wykonaniu eksperymentów próbkowania strukturalnego Ty1 RNA *in vitro* metodą SHAPE-CE
- współudziale w eksperymentach próbkowania strukturalnego RNA w komórkach *E.coli* metodą SHAPE-CE
- normalizacji i analizie surowych danych otrzymanych z eksperymentów SHAPE-CE
- otrzymaniu modelu struktury drugorzędowej Ty1 RNA *in vitro* przy zastosowaniu oprogramowania SuperFold oraz jej wizualizacji przy użyciu aplikacji VARNA
- współudziale w analizie statystycznej i interpretacji wyników
- współudziale w przygotowaniu pierwszej wersji manuskryptu oraz jego finalnej wersji,
- przygotowaniu *Supplementary Dataset*
- współudziale w przygotowaniu figur, tabel oraz *Supplementary Tables and Figures*

[§]Autor korespondencyjny

Podpis

Poznań, 14.09.2002

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

Małgorzata Zawadzka, Angelika Andrzejewska, Julita Gumna, David J. Garfinkel, Katarzyna Pachulska-Wieczorek[§]. **Cell compartment-specific folding of Ty1 Long Terminal Repeat retrotransposon RNA genome**. *Viruses*, 2022; 14(9):2007

Oświadczam, że wkład mgr Małgorzaty Zawadzkiej w powstanie wyżej wymienionej pracy naukowej polegał na:

- współudziale w opracowaniu planu badań,
- optymalizacji warunków i przeprowadzeniu hodowli drożdży *S. paradoxus* szczep DG3408 i *S. cerevisiae* szczep BY4742,
- wykonaniu eksperymentów próbkowania strukturalnego jądrowego Ty1 RNA w drożdżach (DG3408) metodą SHAPE-CE,
- wykonaniu eksperymentów próbkowania strukturalnego U1 snRNA w drożdżach (BY4742) metodą SHAPE-CE,
- normalizacji i analizie surowych danych otrzymanych ze wszystkich eksperymentów
- otrzymaniu modelu struktury drugorzędowej jądrowego Ty1 RNA przy zastosowaniu oprogramowania SuperFold oraz jej wizualizacji przy użyciu aplikacji VARNA,
- przeprowadzeniu analizy statystycznej i współudziale w interpretacji wyników
- przygotowaniu pierwszej wersji manuskryptu oraz współudziale w przygotowaniu i edycji finalnej wersji
- przygotowaniu *Supplementary Data* oraz *Supplementary File*.

Podpis

[§]Autor korespondencyjny

Poznań, 14.09.2002

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), 6670

Oświadczam, że wkład mgr Małgorzaty Zawadzkiej w powstanie wyżej wymienionej pracy naukowej polegał na

- współudziale w opracowaniu koncepcji pracy
- zgromadzeniu i analizie stosownej literatury
- przygotowaniu pierwszej wersji rozdziału *RNA Structure in Relation to RNA Stability and Degradation* oraz *Impact of RNA Modifications on RNA Structure In Vivo*
- współudziale w przygotowaniu wraz z mgr inż. Angeliką Andrzejewską-Romanowską rozdziału *The Relationship between RNA Structure and Proteins Binding*
- współudziale w przygotowaniu tabel
- współuczestniczeniu w przygotowaniu pozostałych części manuskryptu, jego edycji oraz opracowaniu wersji finalnej.

[§]Autor korespondencyjny

Podpis