T-hairpin structure found in the RNA element involved in piRNA biogenesis

NAOMI TAKASE,[1] MAINA OTSU,[1] SHIGEKI HIRAKATA,[2] HIROTSUGU ISHIZU,[2,3] MIKIKO C. SIOMI,[2] and GOTA KAWAI[1]

[1]Department of Life Science, Graduate School of Advanced Engineering, Chiba Institute of Technology, Chiba 275-0016, Japan
[2]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0032, Japan
[3]Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan

ABSTRACT

PIWI-interacting RNAs (piRNAs) repress transposons to protect the germline genome from DNA damage caused by transposon transposition. In *Drosophila*, the *Traffic jam* (*Tj*) mRNA is consumed to produce piRNA in its 3′-UTR. A *cis* element located within the 3′-UTR, *Tj-cis*, is necessary for piRNA biogenesis. In this study, we analyzed the structure of the *Tj-cis* RNA, a 100-nt RNA corresponding to the *Tj-cis* element, by the SHAPE and NMR analyses and found that a stable hairpin structure formed in the 5′ half of the *Tj-cis* RNA. The tertiary structure of the 16-nt stable hairpin was analyzed by NMR, and a novel stem–loop structure, the T-hairpin, was found. In the T-hairpin, four uridine residues are exposed to the solvent, suggesting that this stem–loop is the target of Yb protein, a Tudor domain-containing piRNA biogenesis factor. The piRNA biogenesis assay showed that both the T-hairpin and the 3′ half are required for the function of the *Tj-cis* element, suggesting that both the T-hairpin and the 3′ half are recognized by Yb protein.

Keywords: T-hairpin; RNA structure motif; piRNA biogenesis; *Tj-cis* element; Yb protein; NMR

INTRODUCTION

PIWI-interacting RNAs (piRNAs) are endogenous small noncoding RNAs that protect the genome from transposable elements in the germline (Czech et al. 2018; Yamashiro and Siomi 2018). Animals lacking piRNA function exhibit defects in gametogenesis and infertility, and their descendants are predisposed to inheriting mutations. The piRNA pathway has evolved to repress transposons transcriptionally and/or post-transcriptionally. In addition, piRNAs have also been found in neuronal cells, suggesting that the function of piRNAs is not limited to reproductive tissues (Lee et al. 2011; Rajasethupathy et al. 2012).

piRNAs are generated through the primary processing pathway and assemble RNA-induced silencing complexes (piRISCs) with PIWI proteins. In germ cells of the ovary of many animals such as *Drosophila*, piRNAs are subsequently amplified by a piRNA amplification loop known as the Ping-Pong cycle. However, the Ping-Pong cycle is inert in the somatic cells that surround the germ cells of the ovary (Hirakata and Siomi 2016). In the primary processing pathway, piRNAs are generated from primary transcripts from piRNA clusters, intergenic elements filled with transposon fragments. In contrast, the transcripts from transposons are cleaved by piRISCs in the Ping-Pong cycle. This suggests that the two processes, secondary piRNA generation and transposon silencing by cleaving their transcripts, occur simultaneously.

piRNA biogenesis requires many factors other than PIWI proteins, such as Tudor domain-containing proteins and RNA helicases (Hirakata and Siomi 2016). In *Drosophila* ovarian somatic cells, primary piRNA biogenesis occurs through Yb bodies, cytoplasmic nonmembranous organelles, and mitochondria. Several piRNA biogenesis factors localize to Yb bodies. A Tudor protein, Yb, is central to the formation of Yb bodies (Hirakata et al. 2019). Primary piRNA precursors are processed into piRNA intermediates by unknown nuclease(s) in Yb bodies. The intermediates are then bound with Piwi, one of three PIWI proteins in *Drosophila*, within Yb bodies, and the complex then translocates to the mitochondrial surface with the RNA helicase Armitage, where an endonuclease Zucchini processes Piwi-bound piRNA intermediates to mature piRNAs (Ipsaro et al. 2012; Nishimasu et al. 2012; Ishizu et al. 2019; Munafò et al.

2019). Two other mitochondrial proteins, Gasz and Daedalus, function as the scaffold for Zucchini processing (Munafò et al. 2019; Yamashiro et al. 2020).

As described above, piRNAs protect the germline genomes from transposons. In ovarian somatic cells of *Drosophila*, transposon-targeting piRNAs are mostly generated from transcripts of piRNA clusters. However, piRNAs are also produced from a subset of mRNAs, especially their 3′-UTR (Robine et al. 2009; Saito et al. 2009). Factors necessary for producing transposon-targeting piRNAs are also required for production of genic piRNAs, which also assemble piRISCs with Piwi. A representative example of an mRNA that contains genic piRNA is *Traffic jam* (*Tj*) mRNA (Fig. 1A), which encodes TJ protein, a transcription regulator. Recently, a 100-nt-long *cis*-acting element within the 3′-UTR of *Tj* mRNA was found to be essential for the generation of *Tj*-piRNAs (Ishizu et al. 2015). Similar *cis*-acting elements were found in piRNA cluster transcripts, the major source of transposon-targeting piRNAs (Homolka et al. 2015; Ishizu et al. 2015). It was also found that the generation of primary piRNA strongly depends on the binding of Yb protein to the *cis* element (Ishizu et al. 2015) and that RNA fragments inserted downstream from the *cis* element were able to produce artificial piRNAs while RNA fragments inserted upstream of the *cis* element were not (Homolka et al. 2015; Ishizu et al. 2015).

In the present study, to elucidate how Yb protein recognizes and binds to the *Tj-cis* element, the secondary structure of *Tj-cis* RNA, a 100-nt RNA corresponding to the *Tj-cis* element, was analyzed. Several fragments were designed from the *Tj-cis* RNA based on the results of SHAPE analysis and secondary structure predictions. NMR spectra of these fragments suggested that a stable structure formed in the 5′ region of *Tj-cis* RNA. The mutation and NMR analyses revealed that a stable structure is formed for U17–G32, and the tertiary structure of the region was determined as a novel stem–loop struc-

ture, the T-hairpin. The piRNA biogenesis assay showed that both the T-hairpin and the 3′ half of the *Tj-cis* RNA are required for the function of the *Tj-cis* element, suggesting that both the T-hairpin and the 3′ half are recognized by Yb protein.
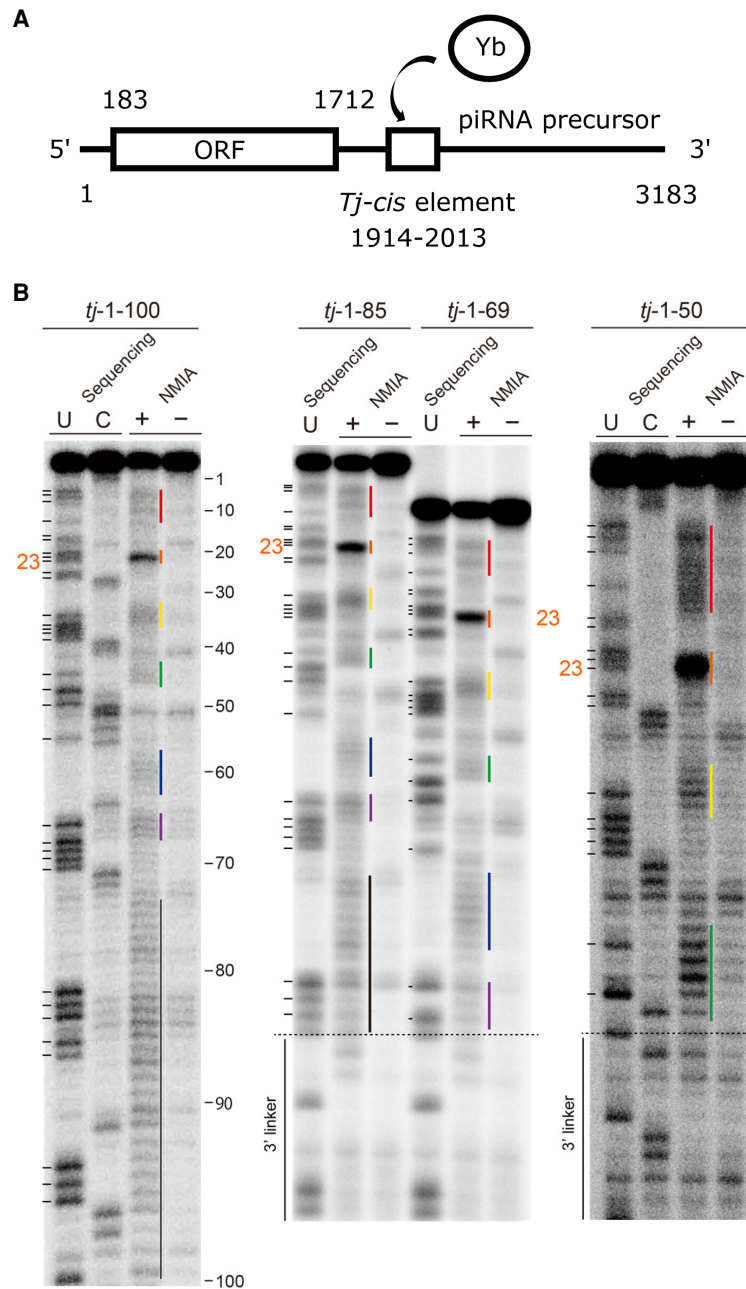

FIGURE 1. Structure of *Traffic jam* (*Tj*) mRNA and SHAPE analysis of *Tj-cis* RNA. (*A*) The *Tj-cis* element is located in the 3′-UTR of the *Tj* mRNA. The element is specifically recognized by Yb protein, and piRNA is generated from the 3′ side of the element. (*B*) A single-stranded region with structural fluctuation was detected by the effect of NMIA for the RNA corresponding to the *Tj-cis* element, *Tj-cis* RNA. Colored bars indicate regions affected by NMIA; the 4–16, 21–25, 33–37, 45–50, 58–64, and 66–69 regions are indicated by red, orange, yellow, green, blue, and purple, respectively. An intense band was observed for U23. Residues at the 3′ side of position 75, indicated by black bars, were not clearly analyzed.

RESULTS

Secondary structure analysis of *Tj-cis* RNA

To analyze the structure of *Tj-cis* RNA, the SHAPE analysis was applied (Wilkinson et al. 2006). The estimated single-stranded regions are indicated by lines in Figure 1B. The single-stranded regions of the 5′ half (1–50) were not affected by deletion of 3′ regions including 86–100, 70–100, and 51–100, suggesting that the 5′ half forms a structural domain. Among them, an intense band was observed for U23, suggesting the formation of a specific structure in this region. For the 75–100 region, shown by the black line, bands for all residues were observed and the effect of NMIA was not clearly judged. The present study focused on the specific structure in the 5′ half.

Then, the secondary structure of full-length *Tj-cis* RNA was predicted to consist of five stem–loops, as shown in Figure 2A. In the predicted structure, the 5′ half (1–50) is structurally independent from the 3′ half (51–100), consistent with the result of the SHAPE analysis described above. Then, the secondary structure of the 5′ half was further analyzed to specify the region forming the specific structure including U23. For the 5′ half, a three stem–loop structure, as predicted for the full-length RNA, was predicted with the lowest energy, and a two stem–loop structure was also predicted with the second lowest energy (Fig. 2B). The free energy for the secondary structures (ΔG) was calculated with the program vs_subopt were −7.46 and −4.87 kcal/mol, respectively. The SHAPE analysis indicated that U23 is highly reactive (Figs. 1B, 2, orange), and this suggests the formation of SL2′, because U23 is in a stem structure for SL2, whereas it is in a loop structure for SL2′ (Fig. 2B). On the other hand, the SHAPE analysis suggested that the region for 4–16 is single stranded (Figs. 1B, 2, red). Although the region mostly corresponds to the loop structure of SL1, U4 and A16 are in the stem structure in the lowest structure. In contrast, this region corresponds to the single-stranded region in the second lowest structure (Fig. 2B, right), also supporting the formation of SL2′.
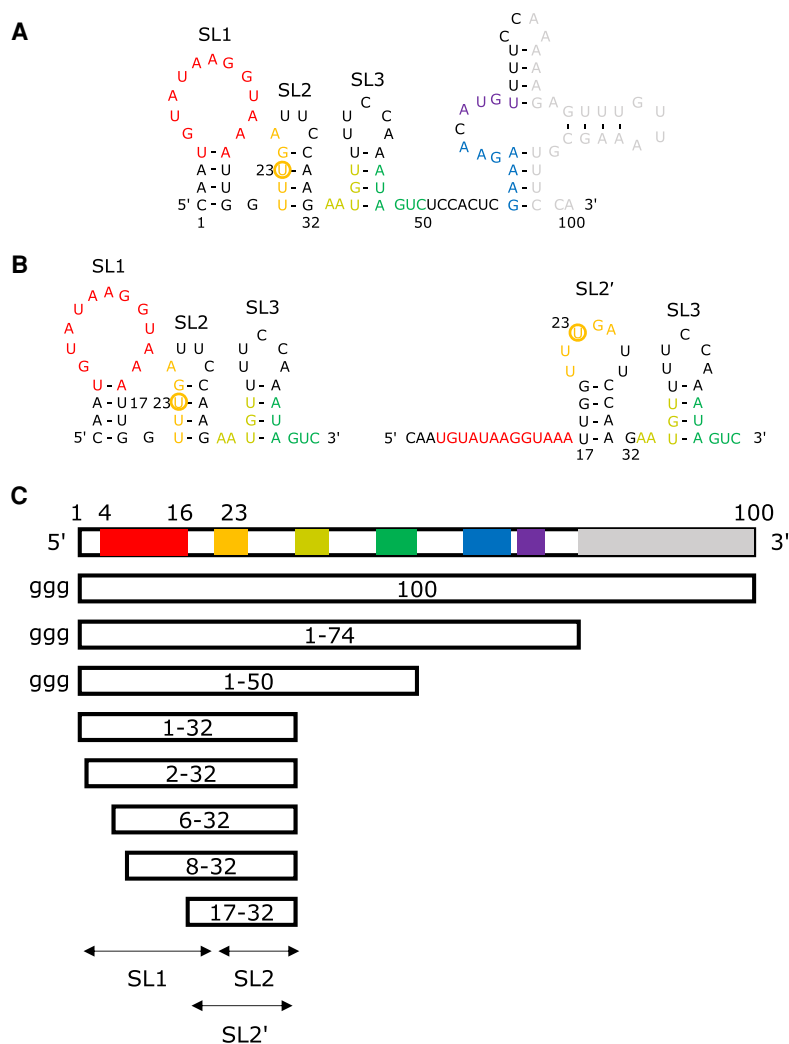

FIGURE 2. Predicted secondary structures and design of the fragments for NMR measurements. Predicted secondary structures of RNA sequences without additional G residues are shown. (*A*) The predicted secondary structure of *tj*-1–100. (*B*) The lowest-energy (*left*) and second lowest-energy (*right*) structures of the 5′ half (*tj*-1–50). Colors correspond to Figure 1B. The U23 residues are indicated by orange circles. Stem–loops predicted in the 5′-half were indicated as SL1, SL2, SL3, and SL2′. (*C*) Design of the fragments for NMR measurements. Colored bars correspond to Figure 1B. Additional G residues added for the in vitro transcription reaction are indicated by lower letters. The fragments without additional G residues were prepared by chemical synthesis.

Identification of the region for the specific structure

Based on the information described above, *Tj-cis* RNA was fragmented for NMR analysis (Fig. 2C). The full-length RNA (1–100) was fragmented into *tj*-1–74 to exclude the region where the structure was not well defined by the SHAPE analysis, and into the 5′ half, *tj*-1–50. Then, fragments corresponding to the region for SL1 and SL2, *tj*-1–32, and the region for SL2′, *tj*-17–32, were prepared. Furthermore, to identify the region where the specific structure is formed, 5′-deleted fragments of *tj*-1–32 were prepared as *tj*-2–32, *tj*-6–32, and *tj*-8–32. The nucleotide

sequences of the designed fragments are summarized in Supplemental Table S1.

The NMR spectra in the imino proton region for the fragments were shown in Figure 3. The two sharp signals at 12.1 and 13.2 ppm were observed for all spectra, suggesting that a specific structure is formed for *tj*-17–32 and the structure formed in all fragments and full-length RNA. Signals at 13.9 and 13.1 ppm were observed for *tj*-1–32, *tj*-2–32, *tj*-6–32, and *tj*-8–32. As described below, these two signals were due to two A–U base pairs. In the case of *tj*-17–32, the signal at 13.9 was broadened, and the signal at 13.1 ppm was slightly shifted to a lower field probably due to the lack of flanking residues of the specific structure. Thus, the two A–U base pairs are formed for all fragments. Signals around 11.5–10.5 ppm were probably due to unpaired G and U residues, and the intensities of signals in this region were decreased according to shortening of the RNA length for *tj*-1–32, *tj*-2–32, *tj*-6–32, *tj*-8–32, and *tj*-17–32, suggesting that the specific structure is formed in the region of 17–32, whereas residues of 1–16
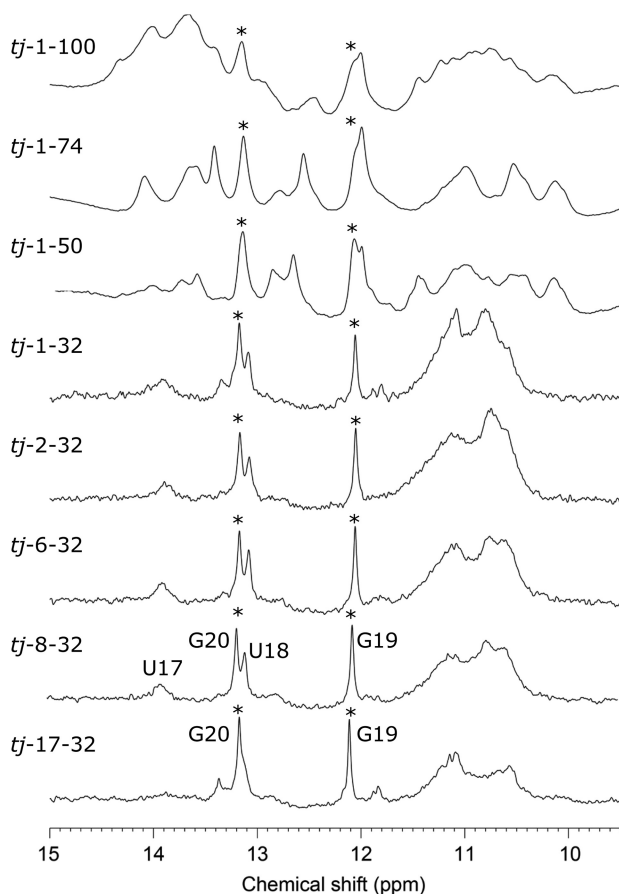

FIGURE 3. Comparison of NMR spectra of *Tj-cis* RNA and its fragments. Imino proton spectra of RNA fragments are shown to indicate the structural similarity with the full-length *Tj-cis* RNA (*tj*-1–100). Asterisks indicate the signals of G19 and G20; for *tj*-1–32 and shorter fragments, the assignments were confirmed by NOESY spectra.

form nonspecific structures. Spectra for *tj*-1–100, *tj*-1–74, and *tj*-1–50 show more signals compared with that of *tj*-1–32, indicating that some structures are formed in the downstream from 1–32. Notably, formation of the SL3 stem–loop was confirmed by NMR analysis of the corresponding 13-nt fragment (data not shown).

These results indicated that a specific structure is formed in the region of U17–G32. As shown in Figure 4A, signals for NOEs between H5 and amino protons, H41 and H42, were observed for C28 and C29, indicating the formation of two G–C base pairs. By the analysis of NOESY spectra, the two sharp imino proton signals could be assigned to the two G–C base pairs. Interstrand NOEs for H2 and H1′ for A30–G19 and A31–U18 were also observed, indicating the formation of two A–U pairs, A30–U18 and A31–U17 (Fig. 4A). Thus, the formation of SL2′ with two G–C base pairs and two A–U base pairs is supported. To confirm this, three mutations, G19U, C28U, and C29U were applied. On mutation of C28 to U28 or C29 to U29, the sharp signals in the imino proton region disappeared (Fig. 4B). For C28U and C29U mutants, imino proton signals around 12–14 ppm indicate the formation of SL1 stem–loop (Fig. 2B) and these signals disappeared by the deletion of the C residue at the 5′ terminal, which destabilizes the SL1 stem–loop (data not shown). Notably, ΔG values of the SL1 and SL2 formation were not affected by these mutations but ΔG of the SL2′ formation is increased (data not shown), supporting the formation of SL1 and SL2 for these mutants. The variant of *tj*-1–32 in which G19 was replaced by U showed no signals in the imino proton region (Fig. 4B). These results also support the formation of SL2′ in the fragments possessing the U17–G32 sequence.

Tertiary structure determination of the novel stem–loop

In NOESY spectra of *tj*-17–32, NOE signals due to the stable stem structure were observed; NOE signals for two G–C base pairs and two A–U base pairs (Fig. 4A). In addition, a signal for H2 of an adenine base was observed in the higher field at 6.81 ppm, suggesting contiguous A–U base pairs (Keane et al. 2015). These results indicate the formation of the alternative stem–loop, SL2′, as shown in Figure 2B. Then, the tertiary structure of *tj*-17–32 was analyzed by the NMR method.

NMR signals were successively assigned by assuming the SL2′ structure (Supplemental Fig. S1). NOESY spectra of *tj*-1–32 and *tj*-6–32 measured at 288, 298, and 303 K were also used for signal assignments, and residue-specific 10% stable-isotopic labeled RNAs, *tj*-1–32-G5L and *tj*-1–32-A25L, were used to confirm the assignment (data not shown). NOEs between signals of the two G–C base pairs were observed, indicating that the two G–C base pairs stack together, and these G–C base pairs were assigned to G19–C29 and G20–C28. The imino proton signals of
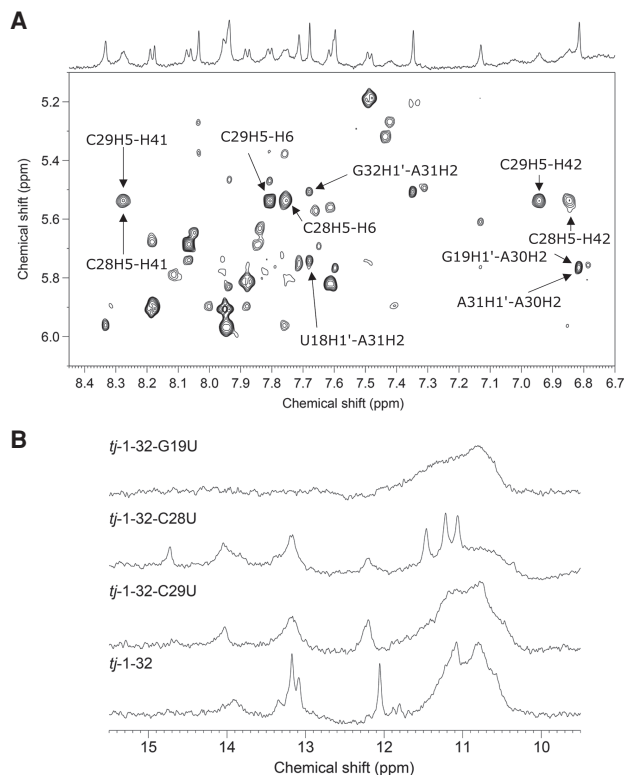
FIGURE 4. NMR spectra of 5′-half fragments. (*A*) NOESY spectra of *tj*-17–32 showing the formation of two G–C base pairs and A–U base pairs. For G–C base pair, NOEs between H5 and amino protons, H41 and H42, can be observed. It is noted that NOE signals for C29H5-H41 and C29H5-H41 were overlapped to each other. For A–U base pair, interstrand and intrastrand NOEs for H2 and H1′ can be observed. (*B*) Imino proton spectra of *tj*-1–32 and its mutants.

these G–C base pairs did not show any change in the chemical shift for 288–303 K, indicating that the structure is stable at least in the temperature range. Notably, no signals in the region for H6/H8/H2–H1′/H5 showed chemical shift changes >0.1 ppm in the NOESY spectra measured at 288 and 298 K. Based on HOHAHA spectra, U17–G20 and C28–G31 were found to be the C3′-*endo* form, suggesting that U17–G20 and C28–G31 form a stem structure. For the loop region, U21 and G24 were in the C3′-*endo* form, and A25 and U27 were in the C2′-*endo* form (Supplemental Fig. S2). U22 and U23 were not fixed to the C3′-*endo* form (Supplemental Fig. S2); these residues are not involved in a stem structure. The sugar pucker for U26 was unknown because the H1′ signal of U26 could not assigned due to signal overlap. Surprisingly, NOEs between A25 and C28 were observed (Supplemental Fig. S1), suggesting that A25 and C28 stack together and U26 and U27 bulge out. H2 of A25 showed NOEs with H1′ of G24 and C28 (Supplemental Fig. S1), also suggesting the specific structural connectivity's for G24–A25–C28. On the other hand, H2′ and H3′ of A25 did not show any interresidual NOEs, indicating the specific conformation of A25.

Notably, H8 of A25 showed NOE with the amino proton of C28 (data not shown), which is also not observed for the RNA-A type geometry. In addition, an NOE between H6 of U27 and H5 of C29 was observed, suggesting that U27 is located in the major groove of the stem. It is noted that residues of U6 to A16 were not fixed to the C3′-*endo* form in *tj*-6–32, suggesting that the 5′ region, 6–16, is flexible.

Based on the structural restraints obtained from the NMR analysis, structure calculations were performed (Supplemental Table S2). The NOE distance restraints were obtained from the NOESY spectrum with a mixing time of 200 ms measured at 298 K. As a result of structure calculation trials, U21 and A25 were found to be in the position forming a reverse Hoogsteen base pair and the structure agrees with the specific NOEs of A25 described above. Furthermore, a sharp imino proton signal of G20 indicates that the G20–C28 base pair is located inside of a stem and is stacked by base pairs from both sides. Thus, the three constraints for the reverse Hoogsteen base pair, two hydrogen-bond distances and one planarity, for the U21–A25 base pair were included in the final calculation to obtain chemically appropriate structures. In the final structure, U21–A25 stacked to G20–C28 and, as a result, a stem consisting of 5 bp was formed (Fig. 5A,B). U22, U23, and U26 were located outside of the molecule and U27 was in the major groove of the stem. Again, the obtained structures agree with the characteristic NOEs of A25 (Supplemental Fig. S1). Hereafter, the determined structure of *tj*-17–32 is called the T-hairpin (Fig. 5C). The secondary structure probing by the SHAPE method showed that U23 is highly sensitive, and this is consistent with the obtained structure in which U23 forms a sharp turn and is exposed to the solvent. As shown in Figure 3, the imino proton signals of G19 and G20 were observed in all fragments sharing the sequence U17–G32, indicating that the T-hairpin structure is formed in all these fragments including *tj*-1–100, the full-length *tj-cis* RNA. Notably, the pyrimidine H5–H6 cross-peaks in HOHAHA spectra for *tj*-17–32 overlapped well with those of *tj*-1–100, except for U17 and U18 (data not shown), also supporting that the T-hairpin structure formed in *tj*-1–100.

MD simulations of 100 nsec were performed with the lowest energy, second lowest energy and third lowest energy structures obtained by the structure calculation as the initial structures. For all three simulations, the four base pairs in the stem formed stably during the simulation, except for U17–A31, which opened during the simulation for the second lowest energy structure (data not shown). In most cases, the U21–A25 base pair was formed for the simulation time (Supplemental Fig. S3), except for the lowest energy structure for which one of the two hydrogen bonds of the U21–A25 base pair opened at 65 nsec and the base pair opened at 90 nsec. To evaluate the dynamics of the T-hairpin structure, the atomic fluctuations during the simulation were calculated for the periods when the
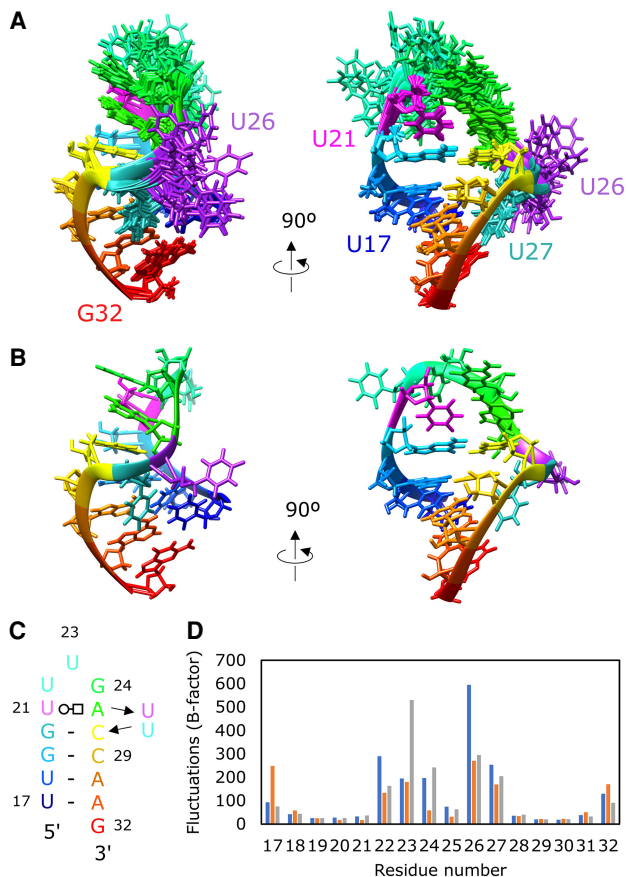
FIGURE 5. Solution structures of *tj*-17–32, T-hairpin. (*A*) Superposition of ten lowest energy structures. The stem region, U17–G20, and C28–A31 are superposed. Each molecule was colored gradually from 5′ in blue to 3′ in red, except for U21, U26, and U27 in magenta, purple, and light-sea green, respectively. (*B*) The minimized average structure. (*C*) The secondary structure of *tj*-17–32, T-hairpin. Each residue was colored as in *A* and *B*. (*D*) Structural fluctuations during the MD simulations. The lowest energy structure (1–65 nsec, blue), the second lowest energy structure (1–100 nsec, orange) and the third lowest energy structure of *tj*-17–32 (1–100 nsec, gray) are shown.

U21–A25 base pair was maintained. Figure 5D showed that U22, U23, U26, and U27 showed higher fluctuations than the other residues in the loop, as expected. Fluctuation of G24 is rather small probably due to the stacking with A25.

Function of the novel stem–loop, T-hairpin

To analyze the structure–function relationship, sequence modifications and mutations were applied to the piRNA biogenesis assay (Fig. 6A). Figure 6B shows the effect of modification of each structural region, the T-hairpin, SL3, or 3′ half, indicating that the T-hairpin and the 3′ half are required for piRNA biogenesis. Mutations in the T-hairpin and 3′ regions affected the piRNA biogenesis activity of the *Tj-cis* element, whereas a mutation in the SL3 region did not affect the activity. The structure–activity relation-

ship was further examined by mutagenesis (Fig. 6C). A mutation, C28C29 to G28G29, which disrupts the T-hairpin structure, decreased the activity. Thus, the T-hairpin structure is involved in piRNA biogenesis. A mutation, C41C42 to G41G42, which altered the SL3 structure, did not affect the activity, indicating that the SL3 region is not necessary for piRNA biogenesis. The SHAPE analysis confirmed the disruption and formation of the T-hairpin structure for mutations C28C29 to G28G29 and C41C42 to G41G42, respectively (Supplemental Fig. S4). These results indicate that the T-hairpin is required for the function of the *Tj-cis* element.

DISCUSSION

Secondary structure analysis of *Tj-cis* RNA showed the formation of a characteristic stem–loop and a possible stem–loop in the 5′ half, as summarized in Figure 5C. In the 5′ half, a unique and stable stem–loop structure, T-hairpin, is formed with four uridine residues exposed to the solvent. The exposed uridine residues may be recognized by other molecules as described below. Because the piRNA biogenesis assay indicated the requirement of the 3′ half, it is possible that some specific structures in the 3′ half could be a key structure for recognition in the pathway. Further structural analysis for the 3′ half is in progress.

The T-hairpin, the characteristic stem–loop structure found in the 5′ half, contains a structural motif with a 7-nt loop. A schematic drawing of the 7-nt loop of the T-hairpin with the closing base pair G20–C28 is shown in Figure 7A. The U and A in position 1 and 5, respectively, of the 7-nt loop form a reverse Hoogsteen base pair, and the 2 nt in positions 6 and 7 are flipped out from the structure. In *tj*-17–32, the G residue in position 4 is stacked to the A residue in position 5. Similar motifs were found in a 3D motif data base, RNA 3D Motif Atlas (Petrov et al. 2013), as HL_33239.1 for the T-loop of tRNAs and HL_08002.1 for loops in riboswitches and rRNAs (Fig. 7B–E). For these motifs, the 7-nt loops are interacting with other loops. For example, the T-loop interacts with the D-loop in tRNA; a base from the interacting loop (D-loop for tRNA) is stacked between two purine bases in positions 4 and 5 of the 7-nt loop. The flipped-out residues in positions 6 and 7 also interact with other loops in these motifs. In contrast, the loop of *tj*-17–32 forms the structure independently. The purine bases in positions 4 and 5 stack together and the uridine residue in position 7 is located in the major groove of the adjacent stem. Thus, the stem–loop structure determined in this study is structurally different from the motifs described above, and thus, the T-hairpin is a novel RNA structural motif. The superposition of the T-hairpin and T-loop (Fig. 7A, right) indicates the structural similarity between the two motifs especially for the G–C and UA base pairs. Because the T-hairpin is formed independently, the T-loop motif can be a variant of the T-hairpin motif. The combination of the reverse Hoogsteen
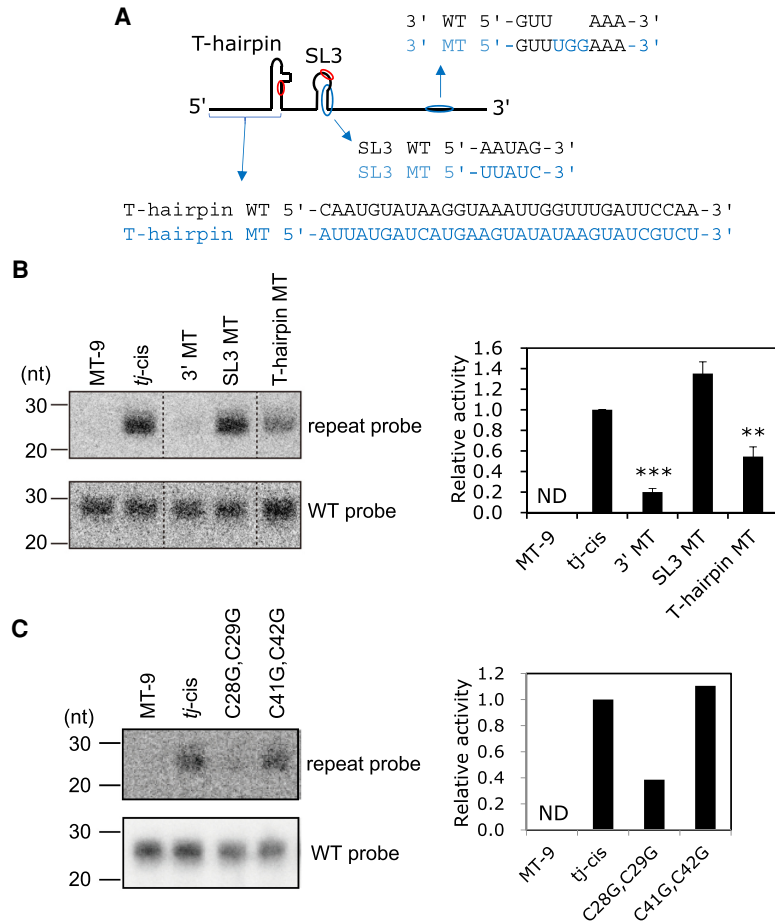
FIGURE 6. piRNA biogenesis assay. (*A*) Location of modifications and mutations in the *Tj-cis* element. The blue letters indicate the replaced or inserted sequences. Red circles indicate the location of point mutations. For T-hairpin MT, the sequence corresponding to residues 1–31 were randomized. (*B*) Replacement and insertion analysis of the three structural elements. (*Left*) MT-9 and *tj-cis* indicate without and with the plasmid carrying the *Tj-cis* element and the repeat sequences, respectively. The 3′ MT, SL3 MT, and T-hairpin MT indicate the mutations in the corresponding region in the *Tj-cis* element. Repeat probe shows the expression of repeat sequences inserted downstream from the *Tj-cis* element. WT probe shows expression of intrinsic piRNA production. (*Right*) Relative activity was calculated by normalizing the repeat probe signal to the WT probe signal. Bars and error bars represent means ± SEM values of four independent experiments. *P*-values were calculated by bootstrap resampling. (**) $P < 0.01$, (***) $P < 0.001$, (ND) not detected. (*C*) Point mutations in the three structural elements. (*Left*) MT-9 and *tj-cis* indicate without and with the plasmid carrying the *Tj-cis* element and the repeat sequences, respectively. (*Right*) Relative activities obtained from the image shown in the *left* panel.

U–A base pair and two flipped out residues (U26, U27) may be important for the stability of these structural motifs. Notably, the ΔG difference between the two predicted secondary structures shown in Figure 2B was 2.59 kcal/mol, and this is smaller than the energy for the reverse Hoogsteen U–A base pair and additional stackings. The exposed four U residues in the T-hairpin may contribute to interactions with other molecules, including Yb protein, during piRNA biogenesis. The piRNA biogenesis assay suggested that the T-hairpin, as well as the 3′ half, are necessary for the function of *Tj-cis* RNA.

The method of elucidating the RNA secondary structure by comparing the NMR spectra of its fragments has been successfully applied for several RNAs (Okui et al. 2016; Ohyama et al. 2020). In the previous work, RNAs with stable structures and giving sharp NMR signals were selected as targets for analyses. In contrast, RNAs, such as full length *Tj-cis* RNA with fluctuating structures and broad NMR signals, have not been subjected to NMR analyses. In the present study, the secondary structures were extensively analyzed by a combination of NMR and fragmentation/base-substitution methods, and a stable stem–loop, the T-hairpin, was discovered. In many eukaryotes, the G–C contents are low, and most RNA molecules show structural fluctuation, like *Tj-cis* RNA. Thus, the method applied in this study, in which structures were fragmented and substituted based on the predicted secondary structures, are especially useful for the analysis of functional and fluctuating RNAs.

MATERIALS AND METHODS

Design of RNA fragments for NMR analysis

RNA fragments of *Tj-cis* RNA were designed based on the results of SHAPE analysis (Wilkinson et al. 2006) and secondary structure predictions by the programs vsfold5 and vs_subopt (Dawson et al. 2007, 2014) for the full-length RNA. For each fragment, the secondary structure was analyzed by vsfold5 and/or vs_subopt to confirm that similar secondary structures to the full-length RNA were predicted. For vsfold5 and vs_subopt, default parameters were used. It was confirmed that the predicted secondary structures were not affected by the addition of G residues for the in vitro transcription.

Preparation of RNA samples for NMR measurements

The nucleotide sequences of the fragments used for NMR measurements are summarized in Supplemental Table S1. RNA samples were prepared by in vitro transcription or chemical synthesis. For *tj-1–100*, *tj-1–74*, and *tj-1–50*, RNA samples were synthesized
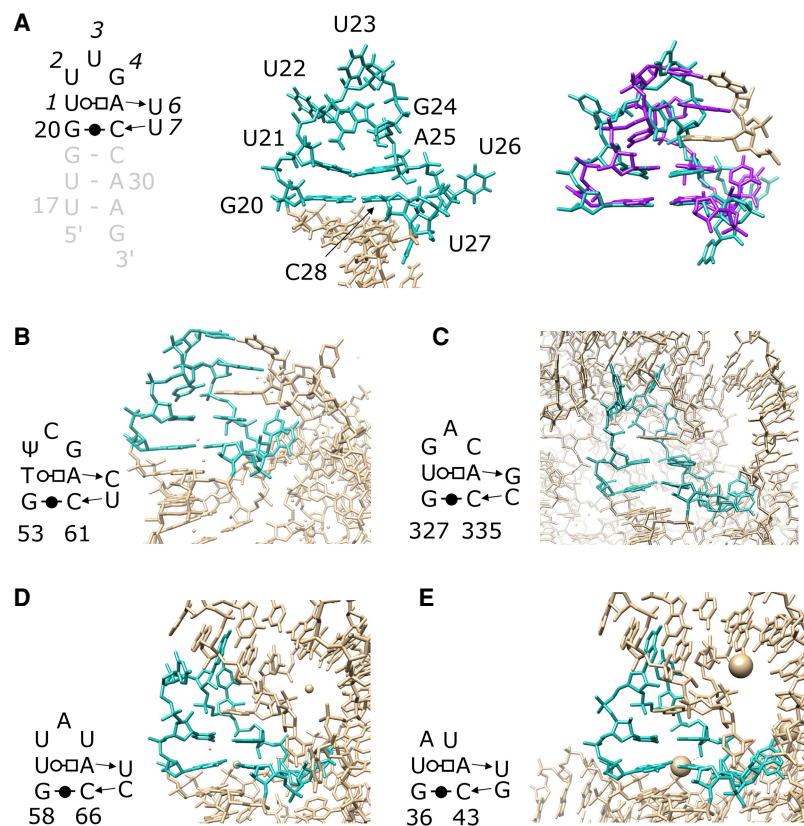
FIGURE 7. Structural comparison between the T-hairpin of *Tj-cis* RNA and similar structures. (*A*) The T-hairpin of *Tj-cis* RNA. The minimized averaged structure (*middle*) and a superposition with yeast tRNA[Phe] (PDB ID: 6TNA) (*right*) are shown. T-loop of the tRNA was colored in magenta and, G18 and G19 of D-arm were colored in tan. (*B*) Yeast tRNA[Phe] (PDB ID: 6TNA). (*C*) *E. coli* 23S rRNA (5J7L). (*D*) Cobalamin riboswitch (4FRG). (*E*) FMN Riboswitch (4YCO). For each panel, the schematic drawing of the 7(6)-nt loop structure with the closing G–C base pair is shown. Positions in the 7-nt loop are indicated by italics.

by in vitro transcription and purified by denaturing polyacrylamide gel electrophoresis (PAGE). T7 RNA polymerase solution was purchased from Taiyo-Nippon Sanso Corporation. DNA templates for in vitro transcription were prepared by PCR with cDNA from *Tj-cis* mRNA. The in vitro transcription reactions were performed under conventional conditions, including 4 h at 37°C. After denaturing PAGE purification, transcripts were extracted from the gel, concentrated by ethanol precipitation, and washed by water using ultrafiltration with Vivaspin 2 (molecular weight cutoff of 3000, Sartorius). Finally, the solvent was exchanged to NMR buffer, 20 mM sodium phosphate buffer (pH 6 .5) with 50 mM NaCl, by ultrafiltration. For NMR measurements, 5% D_2O was added for each sample. Chemically synthesized RNA samples for *tj-1–32*, *tj-2–32*, *tj-6–32*, *tj-8–32*, and *tj-17–32* including their mutants were purchased from Hokkaido System Sciences. Each sample was dissolved in water, then NMR buffer and D_2O were added. For residue-specific labeled RNA samples, 10% $^{13}C/^{15}N$-labeled phosphoramidite units were purchased from Taiyo-Nippon Sanso Corporation, and oligonucleotides were synthesized by Hokkaido System Sciences. For each sample, the concentration was evaluated by UV absorption at 260 nm for a sample diluted 100 times. The purities and

conformational homogeneities were evaluated by denatured and native PAGEs, respectively.

SHAPE analysis

The SHAPE analysis was performed as previously described (Wilkinson et al. 2006). DNA templates for *Tj-cis* RNAs were amplified by PCR and transcribed into RNA using T7 RNA polymerase. The RNAs were purified using 6% denaturing PAGE, excised from the gel, and eluted in 400 µL of 0.4 M NaCl on rotation overnight at 4°C. The eluted RNA was ethanol precipitated, then 200 ng of RNA was refolded in 100 mM HEPES (pH 8.0), 100 mM NaCl, and 10 mM $MgCl_2$. After folding, all RNAs were modified in the presence of 13 mM *N*-methylisatoic anhydride (NMIA) and incubated for 140 min at 25°C. No-reagent controls, containing neat dimethyl sulfoxide rather than NMIA, were performed in parallel. Following modification and precipitation with ethanol, both reagent and control RNAs were subjected to reverse transcription with SuperScript III (Invitrogen) using a radiolabeled primer that targeted the 3′ structure cassette. A reverse transcription sequencing reaction using ddATP or ddGTP and the radiolabeled primer was also performed to allow sequence alignment. We performed the separation step using 10% denaturing PAGE. The primers used are listed in Supplemental Table S3.

Measurements and analyses of NMR spectra

All NMR spectra were measured by an Avance 600 NMR spectrometer (Bruker BioSpin). Sample concentrations were 0.1–0.37 mM. The measuring temperature was 288 K unless otherwise mentioned. Water signals were suppressed by a jump-and-return pulse (Plateau and Guéron 1982) with a delay of 50 µsec for one-dimensional imino proton spectra. For other spectra, a 3–9–19 pulse (Piotto et al. 1992) was used for water suppression with a delay of 70 µsec for measuring all regions and 139 µsec for focusing on the CH region. The NMR spectra were measured and analyzed with the conventional method (Wüthrich 1986; Varani et al. 1996; Sakamoto et al. 2018). Homonuclear Hartmann–Hahn spectroscopy (HOHAHA) spectra were measured with a mixing time of 50 msec and nuclear Overhauser effect spectroscopy (NOESY) spectra were measured with mixing times of 150 to 300 msec. For some measurements of the HOHAHA spectra, the clean-HOHAHA method was used (Griesinger et al. 1988). For structural determination, a NOESY with a mixing time of 200 msec and data points of 8192 and 1980 for the F2 and F1 dimensions, respectively, was measured at 298 K and processed with data points of 8192 and 4096. The observed data were processed by

TopSpin (Bruker BioSpin) and analyzed by Sparky (Goddard and Kneller 2008). The window functions of half sine bell or gaussian function were used. For the residue-specific labeled RNAs, ^{13}C-^{1}H single quantum coherence spectroscopy (SQC) spectra and/or ^{15}N-^{1}H heteronuclear multiple quantum coherence spectroscopy (HMQC) spectra were measured.

Structure calculations

For the determination of the structure of *tj*-17–32, 88 nuclear Overhauser effect (NOE) distance restraints for nonexchangeable protons were obtained using the NOESY spectrum (a mixing time of 200 msec) measured at 298 K in H_2O (5% D_2O). Twelve distance constraints to maintain the hydrogen bonds were used for the base pairs for which the formation was confirmed by the imino proton analysis, except for U21–A25 as described below. Constraints for the planarity of base pairs were also used. Sugar pucker was judged by the signal between H1′ and H2′ in the HOHAHA spectra measured with a mixing time of 50 msec. Dihedral constraints for sugar packer (C2′-*endo* for A25 and U27 and C3′-*endo* for U21 and G24) and the RNA-A stem structure including the *anti*-conformation were used for U17–G20 and C28–A31. Structure calculations were performed 100 times by CNS_SOLVE (Brünger et al. 1998) with the protocol described previously (Otsu et al. 2017) to obtain 72 accepted structures. Ten structures with the lowest energy were selected and a minimized average structure was calculated. In the final stage of the structure calculation of *tj*-17–32, 89 constraints for van der Waals distances were used to prevent a clash between atoms. Molecular images were produced with UCSF Chimera (Pettersen et al. 2004). Atomic coordinates and NMR information for the reported structure have been deposited with the Protein Data Bank under accession number 7EXY and BMRB ID 36422.

Molecular dynamics simulation

Molecular dynamics (MD) simulations were performed with AMBER12 (Case et al. 2012) with the protocol described previously (Otsu et al. 2017). The calculated structures with lowest energies of *tj*-17–32 were used as the initial structures. The force field ff12SB was used. Productive simulations in constant volume without positional restraints were performed for 100 nsec (10,000,000 steps). For the MD simulation of *tj*-17–32, 15 sodium ions were added to the system. For the lowest energy, the second lowest energy and third lowest energy structures of *tj*-17–32, 3501, 3586, and 3068 water molecules were added, respectively. The trajectories of the productive simulation were processed by the program ptraj in the AMBER suite. The atomic fluctuations were shown as the temperature factor in the X-ray crystallography (B-factor, $Å^2 × 8π/3$).

piRNA biogenesis assay

Plasmids used for the piRNA biogenesis assay were produced by inverse PCR, using EGFP-*tj-cis* vector (Ishizu et al. 2015) as templates. MT-9 (Ishizu et al. 2015) was used as a negative control. A piRNA biogenesis assay was performed as described previously (Ishizu et al. 2015). Briefly, ovarian somatic cells (OSCs) were main-

tained as described previously (Saito et al. 2009) and transfected with plasmids using Xfect Transfection Reagent (Takara) following the manufacturer's instructions. RNAs were isolated from OSCs with ISOGEN II (Nippon Gene), followed by DNase treatment. piRNAs were detected by northern blotting as described previously (Saito et al. 2006) and quantified using ImageJ (National Institutes of Health). Bootstrap tests were performed with BootstRatio (Clèries et al. 2012). All primers and probes used in the assay are listed in Supplemental Table S3.

REFERENCES

Brünger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges N, Pannu NS, et al. 1998. Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54:** 905–921. doi:10.1107/S0907444499800 3254

Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, et al. 2012. *AMBER 12.* University of California, San Francisco.

Clèries R, Galvez J, Espino M, Ribes J, Nunes V, López de Heredia M. 2012. BootstRatio: a web-based statistical analysis of fold-change in qPCR and RT-qPCR data using resampling methods. *Comput Biol Med* **42:** 438–445. doi:10.1016/j.compbiomed.2011.12.012

Czech B, Munafò M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, Hannon GJ. 2018. piRNA-guided genome defense: from biogenesis to silencing. *Annu Rev Genet* **52:** 131–157. doi:10.1146/annurev-genet-120417-031441

Dawson W, Fujiwara K, Kawai G. 2007. Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PLoS ONE* **2:** 905. doi:10.1371/journal.pone.0000905

Dawson W, Takai T, Ito N, Shimizu K, Kawai G. 2014. A new entropy model for RNA: part III. Is the folding free energy landscape of RNA funnel shaped? *J Nucleic Acids Invest* **5:** 2652. doi:10.4081/JNAI.2014.2652

Goddard TD, Kneller DG. 2008. *SPARKY 3.* University of California, San Francisco.

Griesinger C, Otting G, Wüthrich K, Ernst RR. 1988. Clean TOCSY for ^{1}H spin system identification in macromolecules. *J Am Chem Soc* **110:** 7870–7872. doi:10.1021/ja00231a044

Hirakata S, Siomi MC. 2016. piRNA biogenesis in the germline: from transcription of piRNA genomic sources to piRNA maturation. *Biochim Biophys Acta* **1859:** 82–92. doi:10.1016/j.bbagrm.2015 .09.002

Hirakata S, Ishizu H, Fujita A, Tomoe Y, Siomi MC. 2019. Requirements for multivalent Yb body assembly in transposon silencing in *Drosophila. EMBO Rep* **20:** e47708. doi:10.15252/embr .201947708

Homolka D, Pandey RR, Goriaux C, Brasset E, Vaury C, Sachidanandam R, Fauvarque MO, Pillai RS. 2015. PIWI slicing and RNA elements in precursors instruct directional primary piRNA biogenesis. *Cell Rep* **12:** 418–428. doi:10.1016/j.celrep.2015.06.030

Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. 2012. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491:** 279–283. doi:10.1038/nature11502

Ishizu H, Iwasaki YW, Hirakata S, Ozaki H, Iwasaki W, Siomi H, Siomi MC. 2015. Somatic primary piRNA biogenesis driven by *cis*-acting RNA elements and *trans*-acting Yb. *Cell Rep* **12:** 429–440. doi:10.1016/j.celrep.2015.06.035

Ishizu H, Kinoshita T, Hirakata S, Komatsuzaki C, Siomi MC. 2019. Distinct and collaborative functions of Yb and Armitage in transposon-targeting piRNA Biogenesis. *Cell Rep* **27:** 1822–1835. doi:10.1016/j.celrep.2019.04.029

Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, et al. 2015. Structure of the HIV-1 RNA packaging signal. *Science* **348:** 917–921. doi:10.1126/science.aaa9266

Lee EJ, Banerjee S, Zhou H, Jammalamadaka A, Arcila M, Manjunath BS, Kosik KS. 2011. Identification of piRNAs in the central nervous system. *RNA* **17:** 1090–1099. doi:10.1261/rna.2565011

Munafò M, Manelli V, Falconio FA, Sawle A, Kneuss E, Eastwood EL, Seah JWE, Czech B, Hannon GJ. 2019. Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. *Genes Dev* **33:** 844–856. doi:10.1101/gad.325662.119

Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, et al. 2012. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491:** 284–287. doi:10.1038/nature11509

Ohyama T, Takahashi H, Sharma H, Yamazaki T, Gustincich S, Ishii Y, Carninci P. 2020. An NMR-based approach reveals the core structure of the functional domain of SINEUP lncRNAs. *Nucleic Acids Res* **48:** 9346–9360. doi:10.1093/nar/gkaa598

Okui S, Ushida C, Kiyosawa H, Kawai G. 2016. Sequence and structure analysis of a mirror tRNA located upstream of the cytochrome oxidase I mRNA in mouse mitochondria. *J Biochem* **159:** 341–350. doi:10.1093/jb/mvv106

Otsu M, Kajikawa M, Okada N, Kawai G. 2017. Solution structure of a reverse transcriptase recognition site of a LINE RNA from zebrafish. *J Biochem* **162:** 279–285. doi:10.1093/jb/mvx026

Petrov AI, Zirbel CL, Leontis NB. 2013. Automated classification of RNA 3D motifs and the RNA 3D motif atlas. *RNA* **19:** 1327–1340. doi:10.1261/rna.039438.113

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605–1612. doi:10.1002/jcc.20084

Piotto M, Saudek V, Sklenár V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2:** 661–665. doi:10.1007/BF02192855

Plateau P, Guéron M. 1982. Exchangeable proton NMR without baseline distortion, using new strong-pulse sequences. *J Am Chem Soc* **104:** 7310–7311. doi:10.1021/ja00389a067

Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, Kandel ER. 2012. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149:** 693–707. doi:10.1016/j.cell.2012.02.057

Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC. 2009. A broadly conserved pathway generates 3′UTR-directed primary piRNAs. *Curr Biol* **19:** 2066–2076. doi:10.1016/j.cub.2009.11.064

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* **20:** 2214–2222. doi:10.1101/gad.1454806

Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. 2009. A regulatory circuit for *piwi* by the large Maf gene *traffic jam* in *Drosophila*. *Nature* **461:** 1296–1299. doi:10.1038/nature08501

Sakamoto T, Otsu M, Kawai G. 2018. NMR studies on RNA. In *Experimental approaches of NMR spectroscopy* (ed. Naito A, et al.), pp. 439–459. Springer, Singapore.

Varani G, Aboul-ela F, Allain FHT. 1996. NMR investigation of RNA structure. *Prog. NMR Spectrosc* **29:** 51–127. doi:10.1016/0079-6565(96)01028-X

Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1:** 1610–1616. doi:10.1038/nprot.2006.249

Wüthrich K. 1986. *NMR of proteins and nucleic acids.* John Wiley & Sons, New York.

Yamashiro H, Siomi MC. 2018. PIWI-interacting RNA in *Drosophila*: biogenesis, transposon regulation, and beyond. *Chem Rev* **118:** 4404–4421. doi:10.1021/acs.chemrev.7b00393

Yamashiro H, Negishi M, Kinoshita T, Ishizu H, Ohtani H, Siomi MC. 2020. Armitage determines Piwi-piRISC processing from precursor formation and quality control to inter-organelle translocation. *EMBO Rep* **2:** e48769. doi:10.15252/embr.201948769