Structural basis for transfer RNA mimicry by a bacterial Y RNA

Wei Wang[1], Xinguo Chen[2,3], Sandra L. Wolin[2,3*] and Yong Xiong[1,4,*]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

[2]Department of Cell Biology, Yale School of Medicine, New Haven, CT 06536

[3]RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702

[4]Lead contact

[*]Correspondence: sandra.wolin@nih.gov, yong.xiong@yale.edu

Present address: RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702

SUMMARY

Noncoding Y RNAs are present in both animal cells and many bacteria. In all species examined, Y RNAs tether the Ro60 protein to an effector protein to perform various cellular functions. For example, in the bacterium *Deinococcus radiodurans*, Y RNA tethers Ro60 to the exoribonuclease polynucleotide phosphorylase, specializing this nuclease for structured RNA degradation. Recently, a new Y RNA subfamily was identified in bacteria. Bioinformatic analyses of these YrlA (Y RNA-like A) RNAs predict that the effector-binding domain resembles tRNA. We present the structure of this domain, the overall folding of which is strikingly similar to canonical tRNAs. The tertiary interactions that are responsible for stabilizing tRNA are present in YrlA, making it a close tRNA mimic. However, YrlA lacks a free CCA end and contains a kink in the stem corresponding to the anticodon stem. Since nucleotides in the D and T stems are conserved among YrlAs, they may be an interaction site for an unknown factor. Our experiments identify YrlA RNAs as a new class of tRNA mimics.

Keywords

Y RNA, YrlA, tRNA-like element, noncoding RNA

INTRODUCTION

In addition to the canonical tRNAs that function in protein synthesis, several RNAs depend on structural similarity to tRNA in order to function. These tRNA mimics include the bacterial transfer-messenger RNA (tmRNA) that rescues stalled ribosomes from mRNAs lacking stop codons and the tRNA-like structures that contribute to translation and replication of positive-strand RNA viruses. For tmRNA, the tRNA-like portion undergoes aminoacylation and binds the elongation factor EF-Tu, allowing it to enter the A-site of arrested ribosomes and function as an acceptor for the stalled polypeptide (Keiler, 2015). Although the viral tRNA-like sequences have diverse functions in translation and replication, several are substrates for the CCA-adding enzyme and a tRNA synthetase and interact with the elongation factor EF-1A (Dreher, 2009). Additionally, tRNA-like structures at the 3′ end of several mammalian long noncoding RNAs (lncRNAs) are cleaved by the tRNA 5′ maturation enzyme RNase P, resulting in 3′ end formation of the upstream lncRNA and release of a tRNA-like noncoding RNA (ncRNA) of unknown function (Sunwoo et al., 2009; Wilusz et al., 2008).

Another class of RNAs that are proposed to mimic tRNA consists of bacterial ncRNAs known as YrlA (Y RNA-like A) RNAs (Chen et al., 2014). These ncRNAs are members of the Y RNA family, 80 to ~220 nucleotides (nt) ncRNAs that were initially identified in human cells because they are bound by the Ro60 autoantigen, a major target of autoantibodies in patients with systemic lupus erythematosus (Wolin et al., 2013). Studies in vertebrate cells revealed that Y RNAs regulate the subcellular location of Ro60 and its association with other proteins and RNAs. For

example, Y RNA binding masks a nuclear accumulation signal on Ro60, retaining it in the cytoplasm (Sim et al., 2009). Y RNAs also scaffold the association of Ro60 with other proteins, as binding of the zipcode-binding protein ZBP1 to a mouse Y RNA adapts the Ro60/Y RNA complex for nuclear export (Sim et al., 2012). Moreover, the ring-shaped Ro60 binds the 3′ ends of some misfolded ncRNAs in its central cavity and adjacent structured RNA regions on its outer surface (Fuchs et al., 2006; Stein et al., 2005). Because Y RNAs bind overlapping sites on the Ro60 outer surface, Y RNAs may regulate the access of misfolded ncRNAs to the Ro60 cavity (Stein et al., 2005).

Studies of bacterial Y RNAs have revealed that, as in animal cells, their functions are intertwined with that of the Ro60 protein. In *Deinococcus radiodurans*, the bacterium where Ro60 and Y RNAs have been most extensively characterized, at least two Y RNAs, called Yrn1 (Y RNA 1) and Yrn2, are bound and stabilized by the Ro60 ortholog Rsr (Ro sixty-related) (Chen et al., 2000; Chen et al., 2013; Chen et al., 2007). Consistent with co-regulation, these ncRNAs are encoded upstream of Rsr and on the same DNA strand. One role of Yrn1 is to tether Rsr to the ring-shaped 3′ to 5′ exoribonuclease polynucleotide phosphorylase (PNPase), forming a double-ringed RNA degradation machine called RYPER (Ro60/Y RNA/PNPase Endonuclease RNP) (Chen et al., 2013). In RYPER, single-stranded RNA threads from the Rsr ring into the PNPase cavity for degradation, rendering PNPase more effective in degrading structured RNA (Chen et al., 2013). In addition to its role in RYPER, Rsr assists 23S rRNA maturation by two 3′ to 5′ exoribonucleases, RNase II and RNase PH, during heat stress, where Rsr functions as a free protein, and is

inactive when bound to Y RNA (Chen et al., 2007). Thus, in addition to acting as a tether, Yrn1 may function as a gate to block access of other RNAs to Rsr.

Y RNAs are modular, a feature that is critical for carrying out their functions. All characterized Y RNAs contain a long stem, formed by base-pairing the 5′ and 3′ ends of the RNA, that contains the Ro60 binding site. Although in both metazoans and *D. radiodurans*, the sequences required for Ro60 binding map to a conserved helix (Chen et al., 2013; Green et al., 1998), a structure of a *Xenopus laevis* (*X. laevis*) Ro60/Y RNA complex revealed that Ro60 primarily interacts with the 5′ strand (Stein et al., 2005). Consistent with the idea that base-specific interactions with this strand are critical for Ro60 recognition, only the 5′ strand of the helix is conserved across bacterial species (Chen et al., 2014). The other end of all Y RNAs consists of internal loops and stem-loops that interact with other proteins. For example, to form the mammalian Ro60/Y RNA/ZBP1 complex, ZBP1 interacts with the large internal loop of the Y RNA (Köhn et al., 2010; Sim et al., 2012), while in *D. radiodurans* RYPER, this portion of Yrn1 interacts with the KH and S1 single-stranded RNA-binding domains of PNPase (Chen et al., 2013). Thus, one role of this second Y RNA module is to tether Ro60 to an effector protein.

Remarkably, for many bacterial Y RNAs, the effector-binding module bears a striking resemblance to tRNA. The first member of this Y RNA subfamily was identified in the enteric bacterium *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium), where it and a second Y RNA were bound by Rsr and encoded 3′ to this protein (Chen et al., 2013). Because the two Y RNAs in *S.* Typhimurium appeared to represent a separate evolutionary lineage from the more metazoan-like

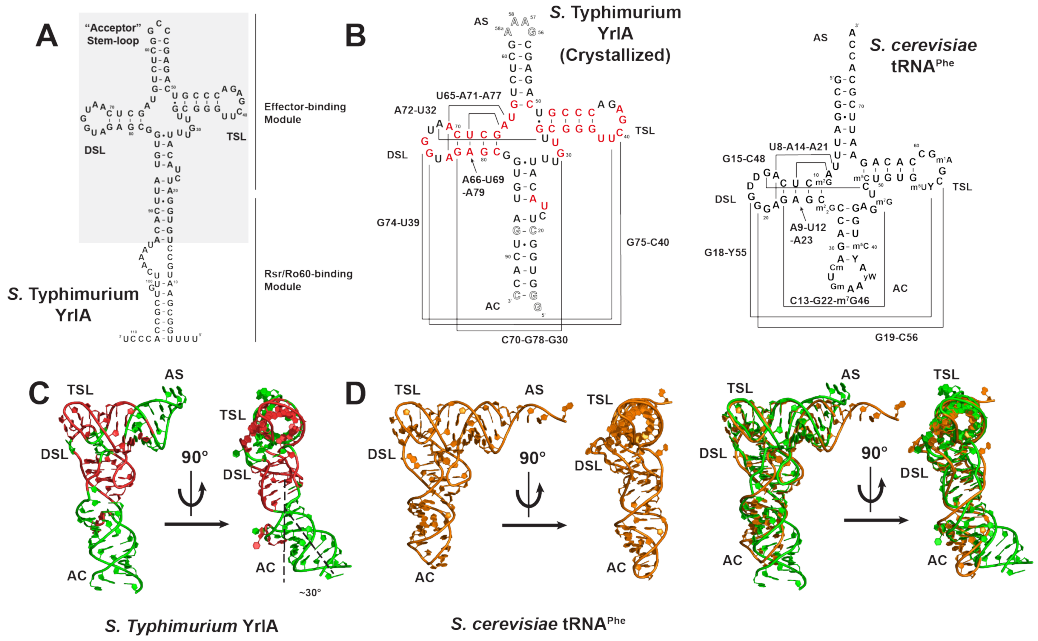
Y RNAs characterized in *D. radiodurans*, these RNAs were designated YrlA and YrlB (Y RNA-like A and B)(Chen et al., 2013). Homology searches revealed that RNAs resembling YrlA were widespread, as they were detected near Rsr in >250 bacterial species representing at least 10 distinct phyla (Chen et al., 2014). Identification of conserved sequences and secondary structures within these RNAs revealed similarities to the D, T and acceptor stem-loops of tRNA (Figure 1) (Chen et al., 2014). Consistent with a tRNA-like fold, *S.* Typhimurium YrlA is a substrate for two tRNA modification enzymes, TruB and DusA, that modify the T and D loops, respectively (Chen et al., 2014).

To test the hypothesis that the YrlA effector-binding domain folds into a tRNA-like structure, we determined the structure of the tRNA-like domain of *S.* Typhimurium YrlA by X-ray crystallography. We show that the YrlA effector-binding domain indeed assumes a similar overall fold as tRNA and that the same tertiary interactions that stabilize tRNA are present in YrlA. In support of a critical role for the tRNA-like module, both the ability to fold into a tRNA-like structure and specific sequences within the structure are conserved in YrlA RNAs from a wide range of bacteria.

RESULTS

We determined the crystal structure of the *S.* Typhimurium YrlA effector-binding module (nucleotides 15-93) at 3 Å resolution (Figure1A, Figure S1). To obtain the high-resolution diffraction data, the 3-nucleotide (nt) loop ^{56}CCG^{58} of YrlA was changed to a ^{56}GAAA^{58a} tetraloop, where the extra nucleotide was numbered as 58a to maintain the original numbering of the subsequent nucleotides (Figure 1B).

133 ^{14}GG^{15} and ^{94}C were added to the construct to increase transcription yield and
134 stabilize the stem. The base pair A20-U88 was also mutated to C20-G88 for stem
135 stabilization. The modified YrlA crystallized in the $C222_1$ space group and diffracted
136 to 3.0 Å resolution. The structure was determined by a combination of single-
137 wavelength anomalous dispersion and molecular replacement methods (see
138 Materials and Methods section for more details). We were able to build all
139 nucleotides into the electron density map and the refinement statistics are
140 summarized in Table 1.



141

142 Figure 1. *S.* Typhimurium YrlA folds into a structure very similar to that of canonical
143 tRNA. (A) Full-length *S.* Typhimurium YrlA consists of an Rsr/Ro60-binding module
144 and a effector-binding module (shaded) that resembles tRNA. (B) Secondary
145 structures of the crystallized *S.* Typhimurium YrlA effector-binding module (nt 15-
146 93) and *S. cerevisiae* tRNA^{Phe}. Important tertiary interactions are connected by lines
147 and are labeled. Conserved nucleotides in YrlA are red. Modified nucleotide
148 sequences to facilitate crystallization are shown as hollow characters. (C) The
149 crystal structure of *S.* Typhimurium YrlA (green, PDB: 6cu1). Conserved nucleotides
150 in YrlA are red. (D) Left, the tertiary structure of *S. cerevisiae* tRNA^{Phe} (orange, PDB:
151 4tna); right, overlay of the YrlA (green) and tRNA^{Phe} (orange) structures.
152

The overall architecture of S. Typhimurium *YrlA resembles that of tRNA*

The YrlA RNA folds into a L-shaped structure that is characteristic of tRNAs. As predicted from the secondary structure (Chen et al., 2014), all tRNA equivalent regions are present in the YrlA structure, including a stem-loop region resembling the acceptor stem (AS, nt 50-64), the T stem-loop (TSL, nt 33-49), the D stem-loop (DSL, nt 67-81), and the anticodon stem (AC, nt 15-27 and 83-93). These structural elements fold into the L-shaped conformation with the two extended stem-loops interacting at an angle very similar to that of tRNAs (Figure 1C). The overall YrlA structure has a backbone RMSD of ~2.4 Å to that of *Saccharomyces cerevisiae* (*S. cerevisiae*) $tRNA^{Phe}$, based on the alignment of 52 phosphorus atoms. It was shown that YrlA RNAs are substrates for tRNA modification enzymes and contain canonical tRNA modifications such as dihydrouridine in the D loop and pseudouridine in the T loop (Chen et al., 2014). The high degree of structural similarity between the two RNAs provides an explanation for the recognition of YrlAs by tRNA modification enzymes.

Deviating from the canonical tRNA structure, there is a kink in the AC region of YrlA, bending the lower portion of the stem by ~30 degrees (Figure 1C). The region extending from this stem contains the Rsr/Ro60-binding module, making YrlA much more elongated than tRNA. The two nucleotides responsible for the kink, $^{23}UA^{24}$, are conserved among YrlA family members. However, sequence alignment indicates that these two nucleotides are predicted to be located in the variable loop (VL) of most YrlA RNAs, rather than being part of the AC stem (Chen et al., 2014). Thus, it is unclear whether the kink is a universal feature of YrlA RNAs.

A major difference between *S*. Typhimurium YrlA and tRNA is that the AS, which terminates in 3´-CCA in all mature tRNAs, is instead a closed loop in YrlA. (Figure 1A). In addition, YrlA lacks the anticodon loop. The length of the AS of YrlAs varies between species (Chen et al., 2014). For S. Typhimurium YrlA, the stem is only six base pairs, which is one base pair less than that of tRNAs. For other species, such as some cyanobacteria, this YrlA stem is predicted to be much longer (Chen et al., 2014). Interestingly, *Mycobacterium smegmatis* YrlA, which resembles *bona fide* tRNAs in containing a seven-base pair AS, is a substrate for RNase P. Following cleavage, the fragment corresponding to a tRNA 3´ end undergoes exonucleolytic nibbling and CCA addition (Chen et al., 2014). However, since most YrlA RNAs contain acceptor stems that are predicted to be poor RNase P substrates, the majority of YrlA RNAs likely resemble circularly permuted tRNAs with closed loop-containing acceptor stems (Chen et al., 2014).

YrlA is stabilized by the same tertiary interactions as tRNAs

In addition to the similarities in overall folding, the tertiary interactions that stabilize the L-shaped structure of YrlA resemble those of tRNA (Figure 2). A major feature of tRNA folding is the interaction between the DSL and the TSL regions to form the tRNA elbow, which serves as a binding site for numerous enzymes that recognize tRNA (reviewed in (Zhang and Ferré-D'Amaré, 2016)). In YrlA, the interaction between nucleotides U38-A42 of the T loop and two guanines in the D loop closely mimics the elbow region of tRNA (Figure 2A, 2B). The first two nucleotides in the YrlA T loop, U38 and U39, stack with nucleotides in the T stem. The third nucleotide, C40, interacts with G75 in the DSL. There is a gap between the

fourth and fifth nucleotides, G41 and A42, in which G74 intercalates to form a continuous stacking interaction.

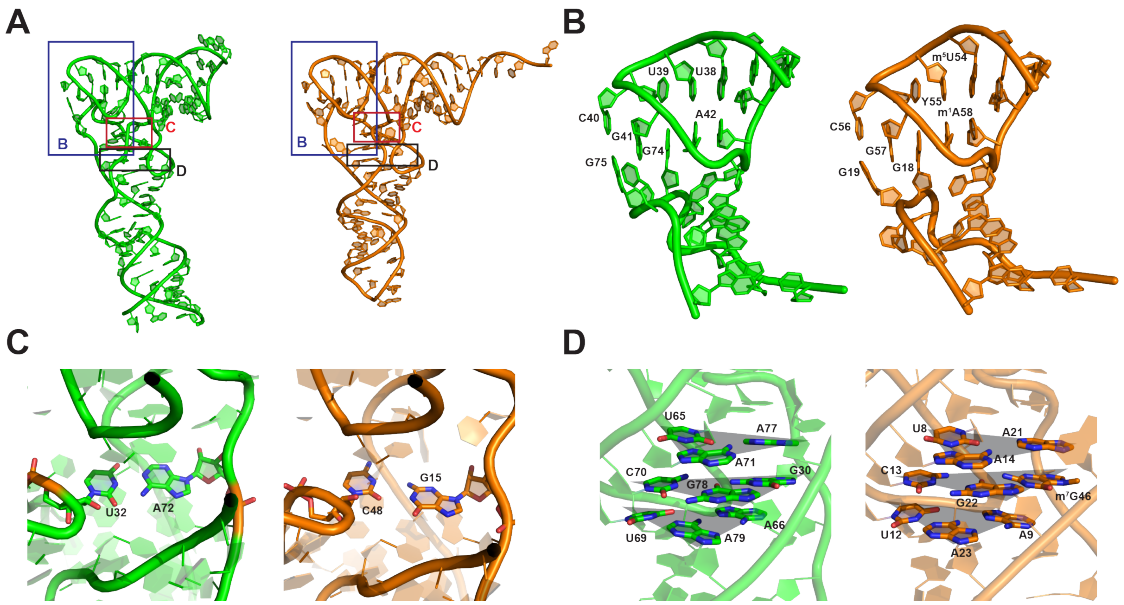


Figure 2. *S.* Typhimurium YrlA is stabilized by tRNA-like interactions. (A) Overall structures of *S.* Typhimurium YrlA (Green, PDB: 6cu1) and *S. cerevisiae* $tRNA^{Phe}$ (Orange, PDB: 4tna). The areas shown in (B), (C) and (D) are boxed. (B) YrlA has DSL-TSL interactions that resemble that of $tRNA^{Phe}$. (C) YrlA contains a Levitt base pair (A72-U32) similar to that of $tRNA^{Phe}$ (G15-C48). (D) U65-A71-A77, C70-G78-G30 and A66-U69-A79 of YrlA closely resemble the base triples of $tRNA^{Phe}$, U8-A14-A21, C13-G22-m^7G46 and A9-U12-A23. The base triples are highlighted by shaded triangles (gray).

Other tertiary interactions important for stabilizing tRNA structure are also present in YrlA. For example, U32:A72 form a base pair equivalent to the Levitt base pair C48:G15 of $tRNA^{Phe}$ (Figure 2C) (Levitt, 1969). This base pair is evolutionarily conserved in YrlA RNAs (Chen et al., 2014). In addition, in YrlA, U65-A71-A77, C70-G78-G30 and A66-U69-A79 form three base triplets that stack on one another. The equivalent base triplets are also found in $tRNA^{Phe}$ (Figure 2D).

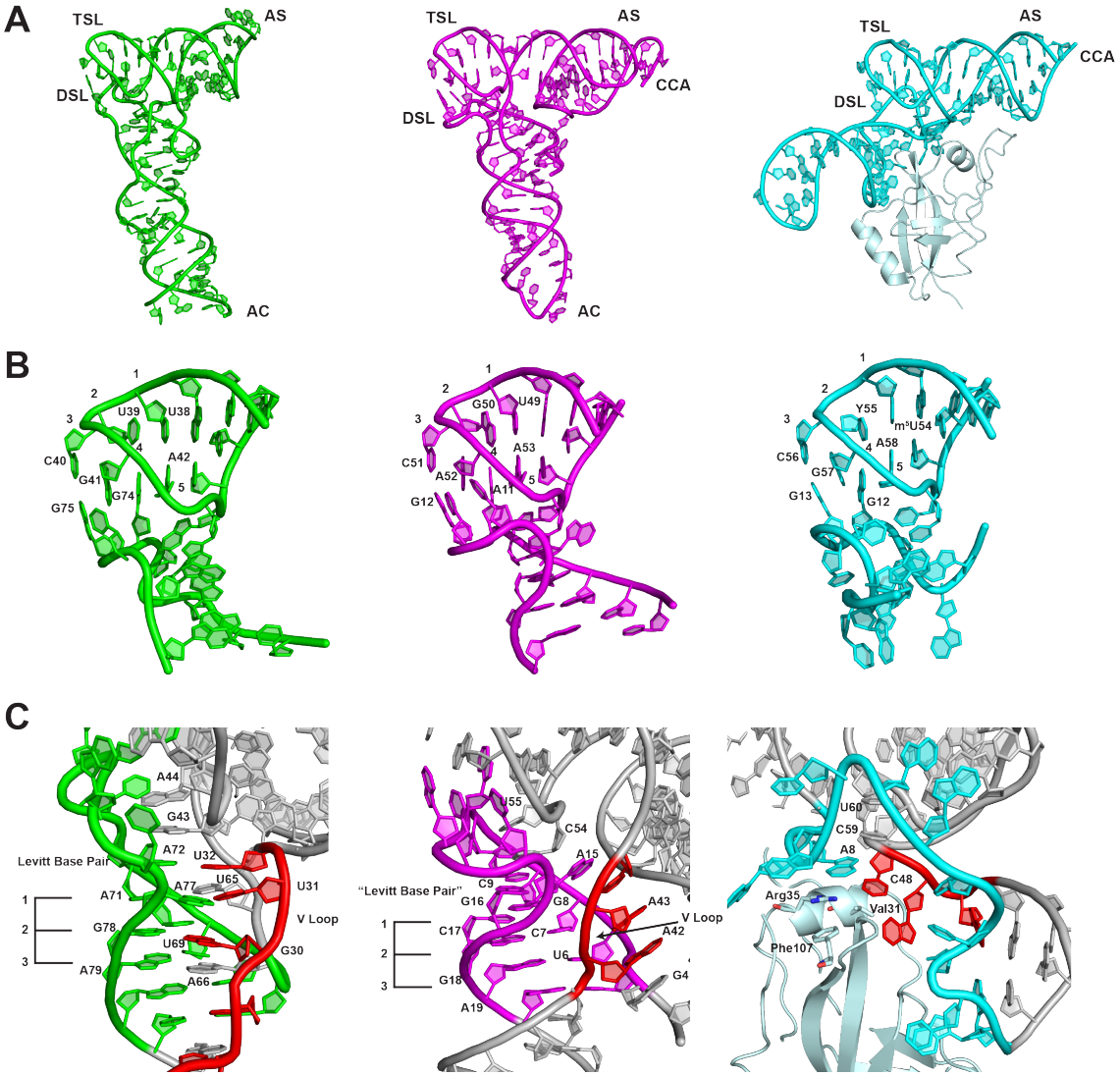


Figure 3. Comparison of the crystal structures of *S.* Typhimurium YrlA (Green, PDB: 6cu1), TYMV TLS (Magenta, PDB: 4p5j) and *T. thermophilus* tmRNA-SmpB complex (Cyan, PDB: 2czj). (A) The three TLEs share the same overall L-shaped fold. (B) The interactions between TSL and DSL are very similar between YrlA, TLS and tmRNA-SmpB. (C) The interactions between the VL and DSL show large differences but maintain the same architecture in the three TLEs. The VLs are colored in red. The three stacking layers of base triples or base pairs are labeled with numbers 1-3. [The numbering of TYMV TLS and *T. thermophilus* tmRNA-SmpB complex are based on previous publications (Bessho et al., 2007; Colussi et al., 2014)]

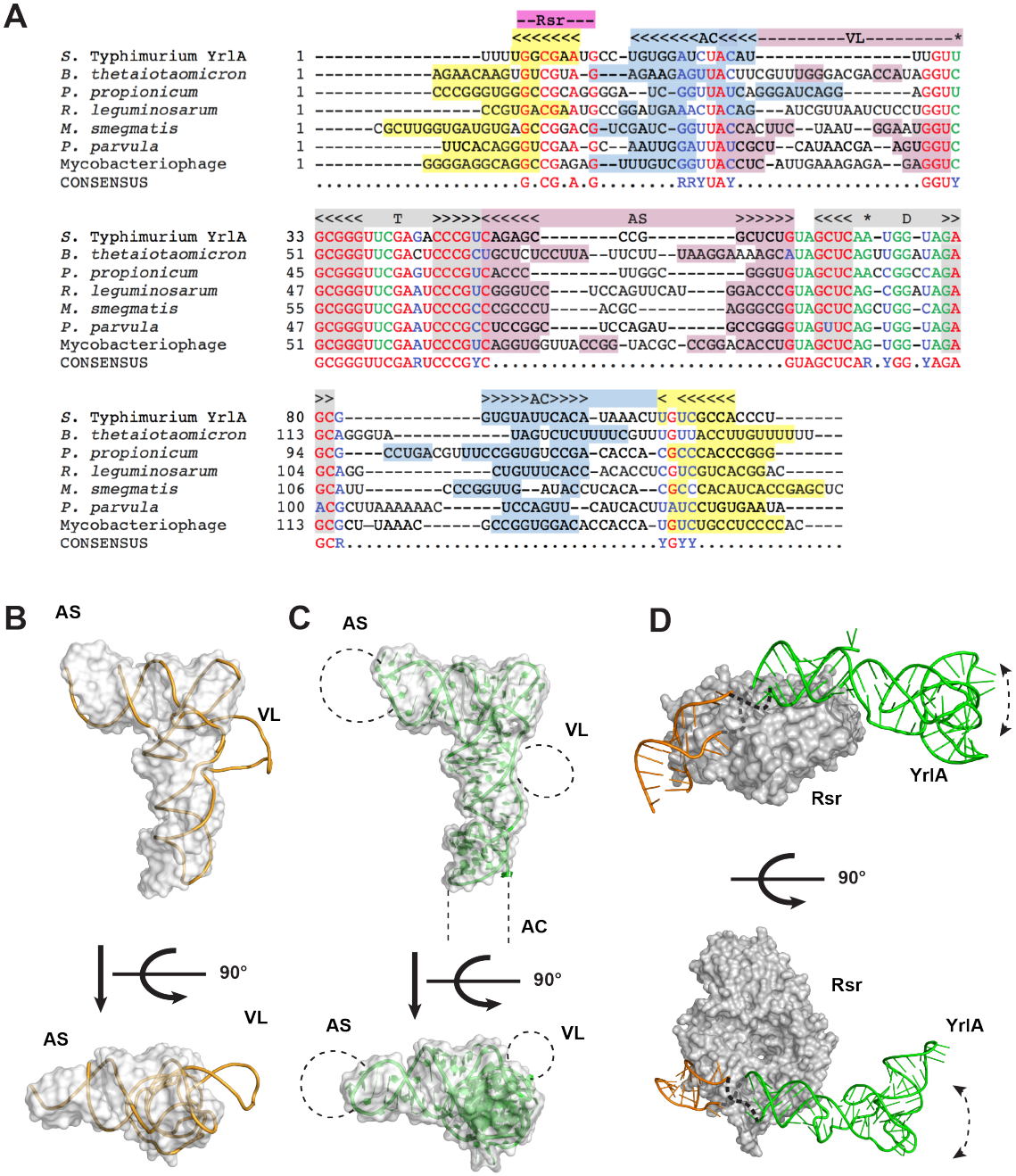
With the *S.* Typhimurium YrlA structure reported here, the crystal structures of three tRNA-like elements (TLE) have now been determined. The other two TLE

structures are *Thermus thermophilus* (*T. thermophilus*) tmRNA and the TLS (tRNA-like structure) at the 3′-end of the turnip yellow mosaic virus (TYMV) genome (Bessho et al., 2007; Colussi et al., 2014; Gutmann et al., 2003). These structures share the L-shaped tRNA fold (Figure 3A) but not all tRNA features are present in all three TLEs. Notably, tmRNA and TYMV TLS contain free 3′-CCA ends and anticodon domain mimics (in the case of tmRNA, a portion of its SmpB protein partner substitutes for the anticodon stem-loop) (Bessho et al., 2007; Gutmann et al., 2003) (Figure 3A). The presence of these structural features is consistent with the ability of these two TLEs to be charged with amino acids and with their biological functions (Dreher, 2009; Keiler, 2015).

The DSL-TSL interaction is well conserved among the three TLEs (Figure 3B). This is perhaps not surprising as this is a major interaction defining the tRNA fold. In all cases, two guanine nucleotides in the DSL interact with a T loop through hydrogen bonding and base stacking interactions (Figure 3B). Since these three molecules have completely different evolutionary trajectories, the highly similar DSL-TSL interaction and hence the tRNA-like fold must be a key feature of their function, such that all three TLEs have either retained it (in the case of YrlA and tmRNA, which may have evolved from tRNA), or convergently evolved to acquire it (viral TLEs).

In contrast, the interactions between the DSL and the VL regions are quite different for the three TLEs. In YrlA, the nucleotides in the DSL region interact with the VL region and form three base-triplet layers, which are very similar to that of canonical tRNAs. These three base layers further stack with the Levitt base pair

A72-U32 and with two nucleotides in the T loop, G43 and A44, stabilizing the L-shaped fold (Figure 3C). TYMV TLS uses a different strategy to stabilize the tRNA-like fold. Its VL does not interact with the DSL. Instead, all three residues in the VL (A42, A43 and U44) flip out to form a continuous stack with A3, G4 and A15. As a consequence, the nucleotides in the D stem form three base pairs instead of base triplets. Nonetheless, the DSL base pairs maintain a stacking interaction with two nucleotides in the T loop (C54 and U55) via a Levitt base pair equivalent (G16-C9) (Figure 3C). Strikingly, these interactions are also mimicked in the case of tmRNA but are conveyed through the SmpB protein, where Val 31, Arg35 and Phe107 bridge the base-pairing and stacking interactions. The completely different strategies used by the three TLEs to stabilize the VL-DSL connection highlight the importance of this region in maintaining the tRNA-like fold. These differences are also consistent with the fact that most tRNA-binding factors do not recognize this portion of tRNAs.



268

269 Figure 4. YrlA from various species likely share the same tRNA-like fold. (A)
270 Sequences of representative YrlA RNAs were aligned using Clustal Omega (Goujon et
271 al., 2010; Sievers et al., 2011) and further adjusted manually. The positions of the
272 DSL and TSL are indicated in gray above the alignments and nucleotides that
273 basepair to form the D- and T-stems are shaded gray. The positions of the AS and VL
274 are indicated in pink, the AC is labeled light blue and the Rsr binding site is indicated
275 with magenta. Nucleotides with the potential to form helices within the stem

created by basepairing the 5′ and 3′ ends are highlighted light blue and yellow. Nucleotides conserved in YrlA RNAs are in red, with conserved nucleotides important for maintaining tertiary tRNA-like structure in green. Conserved nucleotides that are either purines or pyrimidines in YrlA RNAs are in blue. The shaded nucleotides in the VL region can potentially form duplexes. Nucleotides forming Levitt base pair are indicated with a star. (B) The structure of a representative class II tRNA, *T. thermophilus* $tRNA^{Tyr}$ (1h3e, chain B, yellow) was superimposed onto the *S.* Typhimurium YrlA structure (gray surface). (C) Schematic model of a generalized YrlA RNA. The *S.* Typhimurium YrlA structure is shown as both cartoon (green) and gray surface. The different sizes of the AS, VL and AC regions are represented with dashed lines. (D) Model of the *S.* Typhimurium Rsr-YrlA complex. *X. laevis* Ro in complex with its Y RNA fragment (PDB: 1yvp) was used as a model for Rsr (gray surface) and the YrlA Rsr/Ro60 binding module (bright orange). The effector-binding module of YrlA is shown in green.

The YrlA effector domain consists of a conserved tRNA core with variable stems and loops

A sequence alignment of YrlAs from various bacteria shows that the nucleotides involved in stabilizing the tertiary interactions are highly conserved (Figure 4A). These nucleotides include those in the D and T loops, the pyrimidine at the last position of the VL and the ^{65}UA^{66} dinucleotide that connects the AS and the DSL. Most YrlA species also maintain the Levitt base pair (Levitt, 1969) that is important for stabilizing the L-shaped structure (^{32}U and ^{72}A in *S.* Typhimurium YrlA). Thus, we predict that YrlAs from other species will have the same overall tRNA-like fold as observed in the current structure.

Although *S.* Typhimurium YrlA has a short VL, many YrlA species contain the long VLs (Figure 4A) that are characteristic of class II tRNAs. When a representative class II tRNA structure was superimposed on the structure of *S.* Typhimurium YrlA, only the VL protrudes from the L-shaped volume defined by this YrlA (Figure 4B). Class II tRNA VLs contain short duplexes, which could also exist in YrlA species with

306 long VLs (Figure 4A). In addition, the sequence alignment indicates that the AS of
307 YrlA RNAs vary in both sequence and length, as does the portion of the YrlA that
308 corresponds to the AC. These comparisons support a model in which the YrlA core
309 folds into a conserved tRNA-like structure, while the sizes of the AS, AC and VL
310 regions vary (Figure 4C).

311 ***Potential interactions of YrlA RNA with cellular factors***

312 Interestingly, although sequences within the D and T stems are not required
313 to maintain the tRNA L-shape, these nucleotides are conserved in YrlA RNAs (Figure
314 4A). This suggests that the YrlA D and T stems are under evolutionary selection
315 pressure. Most aminoacyl-tRNA synthetases recognize the AC and the AS (Giegé and
316 Eriani, 2014). In addition, the tRNA elbow, where the D and T loops interact, is
317 recognized by the ribosome, ncRNAs and many tRNA-interacting proteins (Zhang
318 and Ferré-D'Amaré, 2016). To our knowledge, no cellular factors have been shown
319 to function by recognizing specific sequences within the T and D stems of tRNAs or
320 TLEs. The sequence conservation in this region suggests that YrlA RNAs may
321 interact with a novel factor that recognizes specific sequences in these stems.

322 We compared available structural information to gain insight into the
323 possible architecture of the Rsr-YrlA. YrlA RNAs contain a conserved Rsr/Ro60-
324 binding module (Figure 1A); hence the structure of the Rsr-YrlA module is believed
325 to resemble that of the *X. laevis* Ro60-Y RNA complex (PDB: 1yvp) (Stein et al., 2005).
326 For *S.* Typhimurium YrlA, the tRNA-like effector-binding module, with structure
327 reported herein, is connected to the Rsr/Ro60-binding module by unpaired short
328 loops of 5 nucleotides. This allows us to model the Rsr-YrlA complex by positioning

329 our structure close to the upper surface of the *X. laevis* Ro60-Y RNA structure
330 (Figure 4D, upper panel). This surface of Ro60 presents positive charges that were
331 shown to be important for both Y RNA and misfolded 5S rRNA binding (Fuchs et al.,
332 2006; Stein et al., 2005). Thus, as predicted for vertebrate Y RNAs and Ro60, YrlA
333 could serve as a gatekeeper to regulate access of other RNAs to Rsr.

334 The unknown effector factor that binds YrlA is presumably bound to the
335 elbow region of the YrlA tRNA-like domain, more specifically, the T stem and the D
336 stem. In the presence of misfolded RNA and/or the effector protein, the tRNA-like
337 domain of YrlA may reposition. For instance, the architecture of the *D. radiodurans*
338 RYPER has been determined by single particle electron microscopy (EM) and the Y
339 RNA is predicted to bend downward from the Rsr surface, allowing the stemloop-
340 containing module to contact one or more S1/KH domains of PNPase (Chen et al.,
341 2013). The architecture of the *S.* Typhimurium Rsr-YrlA-effector complex remains to
342 be determined.

343 **DISCUSSION**

344 Although the structure of the metazoan Y RNA module that binds the Ro60
345 autoantigen was elucidated using X-ray crystallography (Stein et al., 2005), high
346 resolution structures of the effector-binding domains of these RNAs have been
347 lacking. Our crystal structure of the tRNA-like module of *S.* Typhimurium YrlA RNA
348 reveals that this module not only adopts an overall L-shaped structure similar to
349 tRNA, but is also stabilized by the same tertiary interactions. Since all sequences
350 involved in critical tertiary interactions are strongly conserved in YrlA RNAs, we
351 predict that the ability to fold into the canonical tRNA L-shape is a general feature of

this Y RNA family. Moreover, the high degree of sequence conservation at the YrlA region corresponding to the tRNA elbow, particularly within the T and D stems, contrasts with the variable sizes of the AS, AC and VL regions. Based on the extreme conservation of these T and D stem sequences, we expect that the effector(s) that bind YrlA RNAs will be one or more tRNA-binding protein(s) that recognize the elbow region with some sequence specificity.

The strong resemblance of YrlA RNAs to tRNAs lends support to the proposal that YrlA and other Y RNAs evolved from tRNA (Chen et al., 2014). Consistent with this hypothesis, Y RNAs are encoded adjacent to one or more tRNAs in some bacteria (Chen et al., 2014). Additionally, the finding that YrlA RNAs differ from *bona fide* tRNAs in that the TSL occurs 5′ to the DSL supports a recent model in which these RNAs originated from dimeric tRNA transcripts (Sim and Wolin, 2018). In this model, the YrlA TSL derived from the first tRNA, while the DSL derived from the second tRNA. Since the YrlA AS would originate from the spacer between the two tRNAs, a model in which YrlA evolved multiple times in distinct bacteria would provide an explanation for the variable length of this stem. Alternatively, if YrlA evolved from a single primordial dimeric tRNA, there may have been less pressure to maintain the length of the AS. In either case, the additional sequences in a dimeric pre-tRNA, such as the DSL and AC of the first tRNA and the AC and TSL of the second tRNA, could potentially basepair to form a stem containing a sequence recognized by Ro60.

In certain algae and at least one archaeal species, some tRNAs are transcribed as circularly permuted variants that are processed to mature tRNAs

(Chan et al., 2011; Maruyama et al., 2010; Soma et al., 2007). Some of these pre-tRNAs resemble YrlA in that the AS is initially a closed loop and the AC stem is initially formed by base pairing the 5′ and 3′ ends of the newly made RNA. These unusual pre-tRNAs are processed to canonical tRNAs by excising and ligating the extended AC stem to form a circular intermediate, followed by opening of the AS by endonucleases such as RNase P and/or RNase Z (Soma et al., 2007). Although enzymes equivalent to the eukaryotic and archaeal splicing endonucleases have not been reported in bacteria, the resemblance of YrlA RNAs to circularly permuted tRNAs raises the possibility that some YrlA RNAs could undergo processing to more closely resemble canonical tRNAs.

The tRNA resemblance is less evident for Yrn1 RNAs: however, these ncRNAs also contain some tRNA-like features. Yrn1 can be folded to contain a TSL that conserves the T stem sequences of YrlA RNAs (Chen et al., 2014). This TSL likely forms in vivo, as it contains pseudouridine at the position corresponding to the pseudouridine in TSLs of all canonical tRNAs (Chen et al., 2014). Since structures have not been reported for the effector-binding domain of Yrn1 or any metazoan Y RNAs, it remains possible that the three stem loops in the Yrn1 effector-binding domain fold in three-dimensions to mimic tRNA.

Although the exact role of YrlA RNA is unknown, it is likely that it functions in RNA degradation and/or repair. Consistent with a role in RNA degradation, some YrlA and Rsr co-purify with PNPase in *S*. Typhimurium (Chen et al., 2013). If, as described for Yrn1 (Chen et al., 2013), the YrlA effector-binding domain interacts with the S1 and KH domains of PNPase, the highly folded tRNA domain could serve

to protect the RNA from endonucleolytic nicks that would render it a substrate for PNPase or other exoribonucleases.

Rsr, YrlB and YrlA have also been proposed to function in RNA repair, since they are encoded adjacent to the RtcB RNA ligase in many bacteria (Burroughs and Aravind, 2016; Chen et al., 2013; Das and Shuman, 2013). In some bacteria, including *S.* Typhimurium, this operon (*rsr-yrlB -rtcBA*) encodes both RtcB, the ligase that joins pre-tRNA halves following intron excision in Archaea and metazoans (Englert et al., 2011; Popow et al., 2011; Tanaka et al., 2011; Tanaka and Shuman, 2011), and RtcA, an RNA terminal phosphate cyclase (Das and Shuman, 2013; Filipowicz et al., 1985). Although the substrates of RtcB in bacteria are largely unknown, *E. coli* RtcB repairs 16S rRNA following cleavage by the MazF toxin (Temmel et al., 2017). Because Rsr and Y RNAs are encoded adjacent to RtcB in bacteria from multiple phyla, it was proposed that Rsr and one or more Y RNAs function as cofactors to enhance RtcB activity (Burroughs and Aravind, 2016). Consistent with a more general role in RNA ligation, Rsr and YrlA are occasionally encoded adjacent to members of other RNA ligase families.

Interestingly, in certain other bacteria, RtcB is encoded adjacent to a protein containing a Band-7 domain and a predicted ncRNA, called band 7-associated tRNA (b7a-tRNA), that strongly resembles an authentic tRNA (Burroughs and Aravind, 2016). Consistent with functional redundancy between b7a-tRNA and YrlA, these bioinformatics searches predict that occasionally b7a-tRNA is encoded adjacent to Rsr and YrlA is adjacent to the Band-7 domain protein. Although the existence of the putative b7a-tRNA has not been validated experimentally, these predictions,

421 together with our finding that the YrlA effector-binding domain folds similarly to
422 tRNA, support the hypothesis that tRNA-like molecule(s) contribute, directly or
423 indirectly, to RNA ligation.

424 **MATERIALS AND METHODS**

425 ***Plasmid construction and RNA purification***

426 The tRNA-like domain of *S.* Typhimurium YrlA was cloned into the EcoRI and
427 NheI sites of plasmid pHDV4 (Walker et al., 2003), such that the YrlA coding
428 sequence was followed by the HDV ribozyme. *S.* Typhimurium YrlA was transcribed
429 from HindIII-linearized plasmid using T7 RNA polymerase (Milligan et al., 1987).
430 The transcription reaction was mixed with an equal volume of ribozyme denaturing
431 buffer (8M Urea and 0.5M $MgCl_2$) and incubated at 37 °C for 2h to increase HDV
432 ribozyme cleavage efficiency (Rosenstein and Been, 1990). Following T7
433 transcription and HDV cleavage, the sequence of the resulting RNA is:
434 GGGUGGCUCUACAUUUGUUGCGGGUUCGAGACCCGUCAGAGCCCGGCUCUGUAGCUC
435 AAUGGUAGAGCGGUGUAGUCACC. The underlined nucleotides indicate modifications
436 to the original YrlA sequence in order to stabilize the AC stem. The YrlA product was
437 precipitated with ethanol and purified by polyacrylamide-urea gel electrophoresis.
438 YrlA variants were made using QuikChange Site-Directed Mutagenesis (Agilent).

439 ***Crystallization and data collection***

440 The purified YrlA RNAs were folded by heating to 95 °C for 2 min,
441 transferring to 60 °C and incubating for 2 min. $MgCl_2$ was added at 60°C to a final
442 concentration of 10 mM and the reaction was quenched on ice for 30 min before
443 use. The folded RNA was then buffer exchanged to RNA crystallization buffer (50

444 mM sodium cacodylate, 50 mM KCl, 1 mM $MgCl_2$ and 0.1 mM EDTA) using Amicon
445 Ultra Centrifugal Unit (Merck Millipore, Billerica, MA). The RNA was concentrated to
446 a final concentration of 1.2-1.5 mg/ml and screened for crystals using the
447 microbatch under oil method using the Nucleix Suite (Qiagen, Germantown, MD).
448 Crystals were readily formed under numerous conditions overnight.

449 Three YrlA variants were crystallized, namely YrlA WT, YrlA Tetraloop and
450 YrlA Tetraloop 3C, in which an additional cytosine was added to the 3´-end of YrlA
451 Tetraloop. The best crystallization conditions for each YrlA constructs were
452 summarized in Table S1. Crystals were cryoprotected by Paratone oil (Hampton
453 Research, Aliso Viejo, CA) or 30% glycerol. Diffraction data was collected at the
454 Advanced Photon Source beamlines 24ID-C and 24ID-E. The data statistics are
455 summarized in Table 1.

456 ***Structure determination and refinement***

457 The structure of YrlA RNA was determined by a combination of single-
458 wavelength anomalous dispersion (SAD) and molecular replacement. The initial
459 phase information was obtained by SAD phasing using SHELX C/D/E (Sheldrick,
460 2008) from a data set collected on an Iridium hexamine derivative of the YrlA-
461 Tetraloop-3C crystal. A preliminary model was built and used as the search model
462 for the YrlA-Tetraloop data set using Phaser (McCoy et al., 2007). There is one RNA
463 molecule in the asymmetric unit of the crystal. The initial model was refined by
464 iterative rounds of restrained refinement using refmac5 (Vagin et al., 2004)
465 followed by manual rebuilding with Coot (Emsley and Cowtan, 2004). B-factor
466 sharpening was performed to facilitate model building (Liu and Xiong, 2014). The

467 structure was refined to final R/R-free of 21.3%/23.4% with excellent electron
468 density (Figure S1). Refinement statistics are summarized in Table 1.
469

470 **ACKNOWLEDGEMENTS**

471 The authors would like to thank the staff at the Advanced Photon Source beamline
472 24-ID for assistance in data collection. This work was supported by National
473 Institutes of Health Grants R01AI116313 (Y.X.) and R01GM073863 (to S.L.W.) and
474 by the Intramural Research Program, Center for Cancer Research, National Cancer
475 Institute, National Institutes of Health (X.C. and S.L.W.).
476 The authors declare no conflicts of interest.

477 **AUTHOR CONTRIBUTIONS**

478 X.C., S.W., and Y.X. conceived the project. W.W., X.C., S.W. and Y.X. designed the
479 experiments. W.W. and X.C. purified the YrlA RNAs and W.W. crystalized the RNA
480 and collected the X-ray diffraction data. W.W. and Y.X. determined the structure.
481 W.W., X.C., S.W. and Y.X. analyzed the data and prepared the manuscript.

482 **DECLARATION OF INTERESTS**

483 The authors declare no competing interests.

484 **REFERENCES**

485 Bessho, Y., Shibata, R., Sekine, S.-i., Murayama, K., Higashijima, K., Hori-
486 Takemoto, C., Shirouzu, M., Kuramitsu, S., and Yokoyama, S. (2007). Structural basis
487 for functional mimicry of long-variable-arm tRNA by transfer-messenger RNA. Proc.
488 Natl. Acad. Sci. *104*, 8293-8298.

Burroughs, A.M., and Aravind, L. (2016). RNA damage in biological conflicts and the diversity of responding RNA repair systems. Nucleic Acids Res. *44*, 8525-8555.

Chan, P.P., Cozen, A.E., and Lowe, T.M. (2011). Discovery of permuted and recently split transfer RNAs in Archaea. Genome Biology *12*, R38.

Chen, X., Quinn, A.M., and Wolin, S.L. (2000). Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation. Genes Dev. *14*, 777-782.

Chen, X., Sim, S., Wurtmann, E.J., Feke, A., and Wolin, S.L. (2014). Bacterial noncoding Y RNAs are widespread and mimic tRNAs. RNA *20*, 1715-1724.

Chen, X., Taylor, David W., Fowler, Casey C., Galan, Jorge E., Wang, H.-W., and Wolin, Sandra L. (2013). An RNA Degradation Machine Sculpted by Ro Autoantigen and Noncoding RNA. Cell *153*, 166-177.

Chen, X., Wurtmann, E.J., Van Batavia, J., Zybailov, B., Washburn, M.P., and Wolin, S.L. (2007). An ortholog of the Ro autoantigen functions in 23S rRNA maturation in D. radiodurans. Genes Dev. *21*, 1328-1339.

Colussi, T.M., Costantino, D.A., Hammond, J.A., Ruehle, G.M., Nix, J.C., and Kieft, J.S. (2014). The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA. Nature *511*, 366-369.

Das, U., and Shuman, S. (2013). 2′ -Phosphate cyclase activity of RtcA: a potential rationale for the operon organization of RtcA with an RNA repair ligase RtcB in Escherichia coli and other bacterial taxa. RNA *19*, 1355-1362.

511 Dreher, T.W. (2009). Role of tRNA-like structures in controlling plant virus
512 replication. Virus Res. *139*, 217-229.
513 Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular
514 graphics. Acta Crystallogr. D *60*, 2126-2132.
515 Englert, M., Sheppard, K., Aslanian, A., Yates, J.R., and Söll, D. (2011). Archaeal
516 3′ -phosphate RNA splicing ligase characterization identifies the missing
517 component in tRNA maturation. Proc. Natl. Acad. Sci. *108*, 1290-1295.
518 Filipowicz, W., Strugala, K., Konarska, M., and Shatkin, A.J. (1985). Cyclization
519 of RNA 3'-terminal phosphate by cyclase from HeLa cells proceeds via formation of
520 N(3')pp(5')A activated intermediate. Proc Natl Acad Sci U S A *82*, 1316-1320.
521 Fuchs, G., Stein, A.J., Fu, C., Reinisch, K.M., and Wolin, S.L. (2006). Structural
522 and biochemical basis for misfolded RNA recognition by the Ro autoantigen. Nat.
523 Struct. Mol. Biol. *13*, 1002-1009.
524 Giegé, R., and Eriani, G. (2014). Transfer RNA Recognition and
525 Aminoacylation by Synthetases. In eLS (John Wiley & Sons, Ltd).
526 Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and
527 Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI.
528 Nucleic Acids Res. *38*, W695-W699.
529 Green, C.D., Long, K.S., Shi, H., and Wolin, S.L. (1998). Binding of the 60-kDa
530 Ro autoantigen to Y RNAs: evidence for recognition in the major groove of a
531 conserved helix. RNA *4*, 750-765.

532 Gutmann, S., Haebel, P.W., Metzinger, L., Sutter, M., Felden, B., and Ban, N.
533 (2003). Crystal structure of the transfer-RNA domain of transfer-messenger RNA in
534 complex with SmpB. Nature *424*, 699-703.
535 Keiler, K.C. (2015). Mechanisms of ribosome rescue in bacteria. Nat Rev
536 Micro *13*, 285-297.
537 Köhn, M., Lederer, M., Wächter, K., and Hüttelmaier, S. (2010). Near-infrared
538 (NIR) dye-labeled RNAs identify binding of ZBP1 to the noncoding Y3-RNA. RNA *16*,
539 1420-1428.
540 Levitt, M. (1969). Detailed Molecular Model for Transfer Ribonucleic Acid.
541 Nature *224*, 759-763.
542 Liu, C., and Xiong, Y. (2014). Electron Density Sharpening as a General
543 Technique in Crystallographic Studies. J. Mol. Biol. *426*, 980-993.
544 Maruyama, S., Sugahara, J., Kanai, A., and Nozaki, H. (2010). Permuted tRNA
545 Genes in the Nuclear and Nucleomorph Genomes of Photosynthetic Eukaryotes. Mol.
546 Biol. Evol. *27*, 1070-1076.
547 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and
548 Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658-674.
549 Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987).
550 Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA
551 templates. Nucleic Acids Res. *15*, 8783-8798.
552 Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K.,
553 Trowitzsch, S., Will, C.L., Lührmann, R., Söll, D*., et al.* (2011). HSPC117 Is the
554 Essential Subunit of a Human tRNA Splicing Ligase Complex. Science *331*, 760.

555 Rosenstein, S.P., and Been, M.D. (1990). Self-cleavage of hepatitis delta virus
556 genomic strand RNA is enhanced under partially denaturing conditions.
557 Biochemistry *29*, 8011-8016.
558 Sheldrick, G. (2008). A short history of SHELX. Acta Crystallogr. A *64*, 112-
559 122.
560 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R.,
561 McWilliam, H., Remmert, M., Söding, J.*, et al.* (2011). Fast, scalable generation of
562 high‐quality protein multiple sequence alignments using Clustal Omega. Mol. Syst.
563 Biol. *7*.
564 Sim, S., Weinberg, D.E., Fuchs, G., Choi, K., Chung, J., and Wolin, S.L. (2009).
565 The Subcellular Distribution of an RNA Quality Control Protein, the Ro Autoantigen,
566 Is Regulated by Noncoding Y RNA Binding. Molecular Biology of the Cell *20*, 1555-
567 1564.
568 Sim, S., and Wolin, S.L. (2018). Bacterial Y RNAs: Gates, Tethers and tRNA
569 Mimics. Microbiology Spectrum, in press.
570 Sim, S., Yao, J., Weinberg, D.E., Niessen, S., Yates, J.R., and Wolin, S.L. (2012).
571 The zipcode-binding protein ZBP1 influences the subcellular location of the Ro 60-
572 kDa autoantigen and the noncoding Y3 RNA. RNA *18*, 100-110.
573 Soma, A., Onodera, A., Sugahara, J., Kanai, A., Yachie, N., Tomita, M.,
574 Kawamura, F., and Sekine, Y. (2007). Permuted tRNA Genes Expressed via a Circular
575 RNA Intermediate in <em>Cyanidioschyzon merolae</em&gt. Science *318*,
576 450.

577 Stein, A.J., Fuchs, G., Fu, C., Wolin, S.L., and Reinisch, K.M. (2005). Structural
578 Insights into RNA Quality Control: The Ro Autoantigen Binds Misfolded RNAs via Its
579 Central Cavity. Cell *121*, 529-539.

580 Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector,
581 D.L. (2009). MEN ε / β nuclear-retained non-coding RNAs are up-regulated upon
582 muscle differentiation and are essential components of paraspeckles. Genome Res.
583 *19*, 347-359.

584 Tanaka, N., Meineke, B., and Shuman, S. (2011). RtcB, a Novel RNA Ligase, Can
585 Catalyze tRNA Splicing and HAC1 mRNA Splicing in Vivo. J. Biol. Chem. *286*, 30253-
586 30257.

587 Tanaka, N., and Shuman, S. (2011). RtcB Is the RNA Ligase Component of an
588 Escherichia coli RNA Repair Operon. J. Biol. Chem. *286*, 7727-7731.

589 Temmel, H., Müller, C., Sauert, M., Vesper, O., Reiss, A., Popow, J., Martinez, J.,
590 and Moll, I. (2017). The RNA ligase RtcB reverses MazF-induced ribosome
591 heterogeneity in Escherichia coli. Nucleic Acids Res. *45*, 4708-4721.

592 Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F.,
593 and Murshudov, G.N. (2004). REFMAC5 dictionary: organization of prior chemical
594 knowledge and guidelines for its use. Acta Crystallogr. D *60*, 2184-2195.

595 Walker, S.C., Avis, J.M., and Conn, G.L. (2003). General plasmids for producing
596 RNA in vitro transcripts with homogeneous ends. Nucleic Acids Res. *31*, e82-e82.

597 Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3′ End Processing of a
598 Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. Cell
599 *135*, 919-932.

Wolin, S.L., Belair, C., Boccitto, M., Chen, X., Sim, S., Taylor, D.W., and Wang, H.-W. (2013). Non-coding Y RNAs as tethers and gates. RNA Biology *10*, 1602-1608.

Zhang, J., and Ferré-D'Amaré, R.A. (2016). The tRNA Elbow in Structure, Recognition and Evolution. Life *6*.

605 Table 1 Data collection and refinement statistics (See also Figure S1)

Data collection	YrlA Tetraloop	YrlA Tetraloop 3C
Space group	$C222_1$	$P6_3$
Cell dimensions		
a, *b*, *c* (Å)	59.41, 147.11, 101.57	146.46, 146.46, 52.53
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	50.0-3.0 (3.1-3.0)[a]	50.0-4.3 (4.37-4.30)
R_{merge} (%)	9.6 (80.2)	9.6 (>100)
$\langle I \rangle / \langle \sigma(I) \rangle$	13.8 (2.3)	14.4 (1.0)
Completeness (%)	98.0 (97.0)	99.1 (97.4)
Redundancy	4.1 (4.1)	5.7 (4.8)
$CC_{1/2}$	1.00(0.97)	1.00(0.67)
Refinement		
Resolution (Å)	30.8-3.0 (3.1-3.0)	
No. reflections	9072 (843)	
R_{work}/ R_{free} (%)	21.3/23.4 (43.3/47.2)	
No. atoms		
RNA	1742	
Ion	12	
Water	4	
Mean *B*-factors (Å^2)		
RNA	150	
Ion	198	
Water	129	
R.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (°)	1.3	

606 [a]Values in parentheses are for highest-resolution shell.

607