Published in final edited form as:

Wiley Interdiscip Rev RNA. 2014 ; 5(1): 15–29. doi:10.1002/wrna.1194.

The emerging role of triple helices in RNA biology

Nicholas K. Conrad*

Department of Microbiology, UT Southwestern Medical Center, Dallas, TX, USA

Abstract

The ability of RNA to form sophisticated secondary and tertiary structures enables it to perform a wide variety of cellular functions. One tertiary structure, the RNA triple helix, was first observed *in vitro* over 50 years ago, but biological activities for triple helices are only beginning to be appreciated. The recent determination of several RNA structures has implicated triple helices in distinct biological functions. For example, the SAM-II riboswitch forms a triple helix that creates a highly specific binding pocket for *S*-adenosylmethionine. In addition, a triple helix in the conserved pseudoknot domain of the telomerase-associated RNA TER is essential for telomerase activity. A viral RNA *cis*-acting RNA element called the ENE contributes to the nuclear stability of a viral noncoding RNA by forming a triple helix with the poly(A) tail. Finally, a cellular noncoding RNA, MALAT1, includes a triple helix at its 3'-end that contributes to RNA stability, but surprisingly also supports translation. These examples highlight the diverse roles that RNA triple helices play in biology. Moreover, the dissection of triple helix mechanisms has the potential to uncover fundamental pathways in cell biology.

INTRODUCTION

RNA is most closely associated with its role in the 'central dogma' of molecular biology, DNA→RNA→protein. That is, it carries genetic information from the gene to the translational machinery where it is decoded into proteins. However, RNA does not serve solely as a molecular middle-man in protein synthesis. Transfer RNAs drive the decoding of the mRNA and, perhaps most impressively, the peptidyl transferase reaction in protein synthesis is performed using an RNA-based active site in the ribosome.¹ Spliceosomal snRNAs and snoRNAs are central players in pre-mRNA splicing and RNA modification, respectively. Small and large noncoding RNAs (ncRNAs) serve as regulators of gene expression by a variety of mechanisms. In addition to these roles in protein production and gene regulation, many viruses use RNA as their genetic material. The unique functional adaptability of RNA is derived in part from its capacity to form elaborate structures. With its unique specificity and reversibility, RNA–RNA hybridization drives complex and dynamic tertiary structure formation in RNA. This review focuses on the discovery of novel functions for the RNA triple helix, a specific RNA tertiary interaction in which double-stranded RNA stems make hydrogen bond contacts with a third strand of RNA.

*Correspondence to: Nicholas.conrad@utsouthwestern.edu.

Conflict of interest: The author has declared no conflicts of interest for this article.

The ability of RNA to adopt a triple-helical confirmation was first reported in 1957 by Alexander Rich, Gary Felsenfeld, and David Davies.^{2,3} When mixing poly(A) and poly(U), the authors noted that two complexes were formed. The first complex, a typical double helix, was the major form at an A:U ratio of approximately 1:1, while the formation of the triple helix peaked when the ratio was 1:2. The reaction was specific for poly(U) as no other homopolymer tested (poly(A), poly(I), or poly(C)) formed a triple helix with A–U double strands. The authors speculated that similar triple helices may form between DNA and a single-stranded RNA in cells and, indeed, such DNA–RNA triple helices have been implicated in a variety of cellular functions.⁴ For a more complete historical perspective, readers are encouraged to see first-person retrospectives on the discovery of triple helices and other early investigations of RNA structural analysis.^{5,6}

The poly(A)–poly(U) RNA triple helix is formed by conventional Watson-Crick base-pairing of poly(A) with poly(U), while the second strand of poly(U) interacts with the double-helix via Hoogsteen contacts in the major groove (Figure 1(a)).⁷ A considerable number of additional nonconventional base triples have been observed in RNA, so not all triple helices are formed by U•A–U base triples [in the text of this review, dashes (–) represent Watson-Crick base pairs while dots (•) represent Hoogsteen base pairs or other non-Watson-Crick base interactions]. Similar hydrogen-bonding patterns can be formed by C•G–C base triples (Figure 1(b)), but these triples require protonation of the cytosine N3 to form the same number of Hoogsteen hydrogen bonds observed in an U•A–U triple. As an experimental approach to verify predicted triple helices, functional replacement of U•A–U base triples with C•G–C bases provides direct evidence for the existence of a triple helix. In cases where a triple helix is essential for activity, point mutations at all three positions (U•A–U→C•G–C) support function, whereas single or double point mutations do not. Moreover, because N3 protonation stabilizes C•G–C base triples, the activity should be enhanced at lower pH in the C•G–C mutant.

Biological functions for RNA triple helices remained largely unreported, but some insights have been gained through structure determination of viral and cellular RNAs. RNA triple helices have now been implicated in RNA stability, translation, RNA–ligand interactions, and RNA-mediated catalysis.^{8–15} Here, the structures and proposed functions for four triple helices are discussed. First, a triple helix in the bacterial SAM-II riboswitch is discussed. In this case, the triple helix forms a binding pocket that modulates ligand-dependent changes in gene expression. Second, the role of the triple helix in the telomerase RNA subunit, TER, and its proposed functions in promoting telomerase activity is covered. Third, the PAN RNA ENE, a *cis*-acting stability element from a viral ncRNA is examined. Fourth, the dual roles for a triple helix from the MALAT1 RNA in RNA stabilization and translation are detailed. Together, these examples illustrate the diversity in composition, structure, and function of RNA triple helices.

THE SAM-II RIBOSWITCH TRIPLE HELIX FORMS A SAM-SPECIFIC BINDING POCKET

Riboswitches are RNA elements that change conformation upon binding to a metabolite and modulate gene expression in *cis*. For example, changes in RNA structures upstream of an

ORF may expose (or hide) a ribosome binding site or transcription terminator.¹⁶ Riboswitches have been found in archaea, plants, fungi, and algae but are most thoroughly characterized in diverse bacterial species. Several different riboswitches modulate gene expression in response to *S*-adenosylmethionine (SAM), a primary methyl donor for biochemical reactions.^{17,18} Several classes of SAM riboswitches have been identified across a taxonomically wide range of bacteria. Each class has a distinct overall structure, although SAM-IV and SAM-V appear to share similar ligand binding pockets with SAM-I and SAM-II, respectively.^{18–20} Interestingly, one of two genes encoding the SAM synthetase in mammals, MAT2A, contains conserved RNA secondary structural elements suggesting that it is also regulated by post-transcriptional mechanisms, but no evidence for a riboswitch has been reported.²¹

The SAM-II riboswitch was first discovered in a search for structured RNA regulatory elements in proteobacteria.²² Its location upstream of genes involved in methionine and SAM biosynthesis and its structural responsiveness to SAM binding demonstrated that SAM-II is a bona fide riboswitch. Determination of the SAM-II riboswitch structure in complex with SAM revealed a prominent role for a 5-base triple helix⁸ (Figure 2). The three strands of the triple helix (red, yellow, and orange; Figure 2(a–c)) form a specific binding pocket for SAM (blue). In most cases, the three strands of the triple helix form by Watson-Crick base-pairing of nt U40-A46 (red) with A24-U20 (orange) and the triple helix is formed by nucleotides G8-U12 (yellow) resting in the major groove of the Watson-Crick paired double-stranded molecule. Interesting exceptions to this occur at the ‘top’ (as illustrated in Figure 2(a)) of the helix and at the binding pocket for the SAM adenosine. Starting at the top of the triple helix, U40 Watson-Crick pairs with A24, but the triple is formed by a minor groove interaction with A41 (Figure 2(d), top left). Proceeding down the triple helix, the next two base triples are formed by C23–G42•G8 and G22–C43•A9 (Figure 2(d), middle left and bottom left). Below that, binding of SAM mimics a triple helix by forming a reverse Hoogsteen pair between U44 and the adenine base of SAM; an additional hydrogen bond between U10 and U44 completes the base triple (Figure 2(d), top right). The bottom two base triples are typical U-A•U triples (Figure 1(a)) formed by U21-A45•U11 and U20-A46•U12 (Figure 2(d), middle right and bottom right). Additional contacts between the U21-A45•U11 and the positively charged sulfur group of SAM further support specific binding to SAM and aid in the discrimination between SAM and the related metabolite *S*-adenosylhomocysteine (SAH) (not shown).^{8,23} Thus, the SAM-II riboswitch triple helix has evolved to form a highly specific binding pocket for SAM, which then modulates gene expression in response to metabolite concentrations.

The SAM-II triple helix is not simply a static binding pocket for SAM; formation and stabilization of the triple helix is central to the responsiveness of the riboswitch. In the absence of SAM, the Shine-Dalgarno sequence (Figure 2(a), A47–G52, green box) is accessible for ribosome binding and translation. Chemical probing and structural data verify that the riboswitch adopts a more highly structured conformation in the presence of SAM.^{8,22,24–26} The Shine-Dalgarno sequence becomes sequestered in the compact structure resulting in the loss of translation (Figure 2(a)). Further biochemical and biophysical assays led to the proposal that the SAM-II riboswitch employs a conformational capture

mechanism in response to SAM.^{24–26} The stable GC-rich double helix formed between the 5'-end of the RNA with nt G25–G30 forms in the absence of ligand. The presence of magnesium promotes the formation of multiple conformers particularly with respect to the tertiary interactions surrounding that helix (Figure 2(a), dashed lines near the top of the molecule). Because SAM intercalates near the center of the triple helix, it is only in the presence of SAM that the hydrogen bonding networks are strong enough to capture a stable triple helix structure. Formation of the stable triple helix supports the interactions at the bottom of the molecule, resulting in sequestration of the Shine-Dalgarno sequence and repression of gene expression. The SAM-II riboswitch provides a striking example in which an RNA triple helix forms the central component of a dynamic RNA regulatory element.

A TRIPLE HELIX IN TER IS ESSENTIAL FOR TELOMERASE ACTIVITY

The primary mechanism for DNA replication is incapable of faithfully replicating chromosome ends, or telomeres. In the absence of alternate mechanisms for telomere replication, the chromosomes shorten upon each round of cell division ultimately resulting in genome instability and loss of cell viability. In order to ensure proper telomere replication, eukaryotes encode an enzyme called telomerase that extends the ends of DNA using a specialized reverse transcription mechanism.²⁷ Telomerase activity is necessary for continued cell proliferation and, as such, it is strongly implicated in tumorigenesis.^{28,29} In fact, approximately 85–90% of tumor biopsies have activated telomerase. Therefore, understanding the detailed mechanisms of telomerase activity may result in discovery of novel therapeutic targets for cancer.

The human telomerase activity is provided by a ribonucleoprotein (RNP) complex minimally consisting of the protein TERT (Est2 in *Saccharomyces cerevisiae*), and a ncRNA, TER (also known as hTR, TERC; Tlc1 in *S. cerevisiae*).^{27,30} The protein provides the reverse transcriptase activity, and shares structural similarities with known retroviral reverse transcriptases.³¹ The RNA component has multiple functional domains including the template sequence that is copied by TERT using the 3'-end of the chromosome as a primer (Figure 3(a)). Iterative copying of the template yields the short repeated sequence found on telomeric DNA ([TTAGGG]_n in mammals). TER further provides a boundary region essential for stopping reverse transcription past the template. Telomerase-associated RNAs have *cis*-acting sequences important for their processing and stability, but the specific types of sequences are not conserved. For example, human TER assembles into a H/ACA snoRNA-like particle while yeast Tlc1 assembles into an Sm protein complex to promote RNA maturation and stability in cells.³³ Most importantly for this review, TER contains a highly conserved pseudoknot domain (Figure 3(a)).^{34–36} While the precise molecular activity of this pseudoknot remains incompletely understood, mutagenesis studies and its conservation in ciliates, budding yeast, and humans suggest an important function (see below).

Solution NMR revealed that an RNA triple helix is a central structural motif in the TER pseudoknot (Figure 3(b,c)).^{9,32} The triple helix is composed of five base triples (red, orange, and yellow) interrupted by a single A•U Hoogsteen base pair (light blue). Interestingly, the two triples at the top of the helix are formed by minor groove interactions. The first triple is

formed from a single hydrogen bond between A171 and U97 while A117 contacts U97 in a Watson-Crick pair (Figure 3(d), top). Directly below that, A172 sits in the minor groove of the G98–C116 Watson-Crick base pair (Figure 3(d), middle). The triple interactions are broken with a Hoogsteen base pair between A173 and U99 (Figure 3(d), bottom) which induces an approximately 30–38° bend (Figure 3(c)). Immediately below the Hoogsteen pair are three canonical U•A–U base triples (Figure 1(a)) in which A174–A176 Watson-Crick pairs with U113–U115 and Hoogsteen pairs with U100–U102 (yellow). Beyond these U•A–U triples, the Watson-Crick stem continues, but the triple helix is terminated.

The precise role of the triple helix in telomerase function remains unclear, but several lines of evidence demonstrate that it is a critical component of functional telomerase. The nucleotides in the triple helix are conserved throughout vertebrates³⁴ and triple helices are found in yeast as well.^{37,38} Importantly, loss of the triple helix leads to a less stable RNA structure and C•G–C substitutes for U•A–U to stabilize the pseudoknot structure in a pH-dependent fashion (Figure 1(b)).^{9,38} Similarly, C•G–C base triples compensate for U•A–U triples in telomerase activity assays *in vitro*.^{9,37} While RNP formation is not affected by mutations disrupting the triple helix, the triple helix is located near the catalytic site.³⁷ The ability of C•G–C to complement U•A–U in telomerase demonstrates that the requirement for the triple helix is based on the triple helix structure, rather than the specific sequence. In fact, Qiao and Cech³⁷ showed that the 2'-OH of A176 is necessary for catalysis *in vitro*, suggesting the triple helix positions A176 2'-OH near the catalytic site for activity. However, it should be noted that methylation of the analogous hydroxyl groups in *S. cerevisiae* TLC1 had minimal or no effects telomerase activity *in vivo* or on structural stability *in vitro*.^{39,40} Rather, loss of a 2'-OH of a uridine nt four nt downstream of the A-rich strand of the budding yeast triple helix had significant deleterious effects on telomerase activity.^{39,40}

Further studies of this domain are required to clarify the specific roles of the TER triple helix in telomerase function. For example, it is important to understand how 2'-OHs in or near the triple helix contribute to telomerase catalysis. Furthermore, it is important to keep in mind that telomerase is a processive enzyme. Thus, structural changes during binding, catalysis, and translocation of the RNA template requires sophisticated enzyme-substrate dynamics. Therefore, it will be interesting to determine whether the triple helix structure changes throughout the catalytic cycle of telomerase. Thus, exciting mechanistic insights on telomerase may be revealed by continued empirical examination of the telomerase triple helix.

THE PAN RNA ENE TRIPLE HELIX PROTECTS POLYADENYLATED RNA FROM NUCLEAR DECAY

The Kaposi's sarcoma-associated herpesvirus (KSHV) encodes approximately 1.1 kb RNA called the polyadenylated nuclear (PAN) RNA. As with other herpesviruses, KSHV is long (approximately 140 kb) nuclear dsDNA virus with a life cycle that includes both latent and lytic phases.^{41–43} PAN RNA is expressed during the lytic phase of infection to very high levels. In fact, it is by far the most abundant RNA expressed in lytic phase, making up as much as 70% of the polyadenylated RNA in lytically reactivated cell lines.^{44–46} As its name

implies, PAN RNA has a poly(A) tail and is primarily found in the nucleus. Consistent with an important role in viral replication, knockdown of PAN RNA in infected cell lines or deletion of PAN RNA from bacmid-based infectious clones severely decreases viral replication and expression of viral lytic phase genes.^{47,48} The mechanisms and functions of PAN RNA are only beginning to be understood, but recent reports suggest that it regulates transcription of viral and host genes to suppress immune responses and promote viral lytic gene expression.^{47–50} Because KSHV is an oncogenic virus, understanding how viral factors modulate host gene expression may provide clues to the pathogenesis of this clinically relevant virus.

PAN RNA is now classified as a viral long noncoding RNA, or lncRNA,^{51,52} but there were few known examples of nuclear polyadenylated ncRNAs when PAN RNA was first described. PAN RNA includes a 79-nt element near its 3'-end that is necessary for PAN RNA accumulation and sufficient to increase the nuclear levels of an inefficiently exported intronless β -globin reporter mRNA, but not its efficiently exported spliced counterpart.⁵³ Because its activity depends on a canonical cleavage and polyadenylation signal, it was initially proposed that the element enhanced RNA levels by increasing the efficiency of 3'-end formation. In addition, because the intronless β -globin mRNA accumulates in the nucleus, the element was dubbed the expression and nuclear retention element or ENE. However, later work showed that the primary function of the ENE is not to promote 3'-end formation or nuclear retention directly. Instead, the ENE is a potent *cis*-acting inhibitor of nuclear RNA decay.^{54,55} Presumably, both PAN RNA and intronless β -globin trigger nuclear RNA quality control pathways^{56–58} due to their inefficient export. Thus, the accumulation of intronless β -globin and PAN RNA in the nucleus is primarily due to protection of nuclear-retained transcripts from nuclear RNA decay pathways. While these findings suggest that 'ENE' is a misnomer, the element itself is a potent regulator of RNA metabolism and useful tool for studying RNA decay pathways and nuclear ncRNAs.

The ENE is predicted to form a secondary structure with two prominent stems (Figure 4(a), light and dark blue) flanking a loop of nearly exclusively uridines (Figure 4(a), yellow and red). The presence of an unpaired U-rich loop and the ENE's dependence on a polyadenylation site led to the hypothesis that the poly(A) tail hybridized with the U-rich portion of the ENE to sequester it from exonucleases.⁵⁵ In support of this model, the poly(A) tail interacts with the ENE *in vitro* and in RNPs purified from cells.^{12,54,55} In addition, the ENE protects transcripts from exonucleases in nuclear extract, in recombinant *in vitro* systems, and robustly stabilizes polyadenylated nuclear reporter transcripts in cells. Furthermore, the U-rich loops and stem structures are all essential for ENE activity. These data provide a strong argument that the ENE functions interact with the poly(A) tail in *cis* to protect transcripts from decay.

To better understand the structural basis for ENE function, Mitton-Fry et al.¹² determined the crystal structure of the 'core' ENE in complex with a nine-adenosine oligonucleotide (oligoA₉). The core ENE consists of the bottom stem, the U-rich loop, and four base pairs from the top stem (Figure 4(a), right). The top stem is extended by one C–G base pair and includes a GAAA tetraloop to stabilize the ENE structure; the uppermost part of the ENE (nt 911–941) is not predicted to form a strong secondary structure and was deleted for structural

analysis. Importantly, the core ENE maintains the ability to interact with poly(A) tails and inhibit decay *in vitro*, but the entire ENE is necessary for full activity in cells.^{12,54}

Determination of the structure of the “core” ENE in complex with oligoA₉ revealed that the U-rich sequences and the adenosines associate through a five-base triple helix (Figure 4(a–c)).¹² The 3′-most five adenosines (oligoA₉ is orange) are effectively “clamped” between Watson-Crick base pairs with the uridines on the right side of the U-rich loop (red) and Hoogsteen interactions with five uridines on the left (yellow) forming typical U•A–U base triples (Figure 1(a)). Tertiary interactions between oligoA₉ and the ENE extend to the bottom stem with A2–A4 interacting in the minor groove through A-minor motifs (Figure 4(c), orange and dark blue).^{12,59,60} The first adenosine of oligoA₉ does not contact the ENE. Based on its proposed function and mechanism, the ENE is predicted to restrict accessibility of the 3′-end of the poly(A) tail. Indeed, it is difficult to imagine enzymatic access to this terminal nucleotide (A₉) based on the ENE-oligoA₉ structure (Figure 4(d)). Thus, the ENE makes extensive tertiary contacts with oligoA₉ to generate a stable structure and an inaccessible poly(A) tail 3′-end.

Determination of the ENE structure enabled novel bioinformatic strategies to identify ENE-like elements in other viruses. Using a search query that incorporated both structural and mutational analyses, Tycowski et al.⁶¹ identified six new viral ENE-like elements. Two of these elements were found in polydnaviruses, segmented dsDNA insect viruses. If one assumes that the ENE functions to inhibit nuclear RNA decay in these cells, this observation strongly suggests conservation of nuclear decay mechanisms from insects to humans. Two more elements were found in monkey (rhesus rhadinovirus, RRV) and horse (equine herpesvirus-2, EHV-2) gammaherpesviruses. Because of poor conservation at the sequence level, lncRNAs can be difficult to identify based on sequence analysis. Unexpectedly, the identification of these new ENE-like elements uncovered a lncRNA in RRV and supported the existence of a lncRNA in EHV-2. The novel transcripts and PAN RNA are encoded from syntenic positions on their respective viral genomes, strongly suggesting they are viral lncRNA homologs with similar functions in viral replication. The final two ENE-like elements were found in mimivirus, a large dsDNA virus,⁶² and in picorna-like dicistrovirus (PSIV),⁶³ a positive strand RNA virus with only a cytoplasmic phase. This latter discovery supports cytoplasmic functions for ENE-like structures, a finding consistent with potential roles for ENE-like structures in translation (see below). This study demonstrates that ENE-like triple helices may be widespread in viruses and have both nuclear and cytoplasmic functions in viral RNA metabolism.

THE MALAT1 TRIPLE HELIX STABILIZES A MAMMALIAN NUCLEAR RNA

Viruses utilize host cell strategies to perform cellular functions, so Steitz and colleagues proposed that cellular RNAs likely include ENE-like elements that form triple helices.¹² In fact, two reports recently showed that the MALAT1 and MEN β nuclear RNAs have noncanonical 3′-ends that form triple helix structures.^{14,64} The formation of triple helix structures in both transcripts is well supported, but this review focuses on the MALAT1 triple helix.

MALAT1, the **metastasis-associated lung adenocarcinoma transcript 1** (also known as NEAT2), was originally identified due to its overexpression in metastatic lung cancers⁶⁵ and has since been associated with several additional cancers.^{52,66} MALAT1 is a conserved, abundant, nuclear ncRNA that modulates gene expression and cellular function at multiple levels. Mechanistically, MALAT1 regulates alternative splicing in cells by interacting with splicing regulatory proteins⁶⁷ and this regulation can alter cell cycle progression.⁶⁸ MALAT1 has additionally been proposed to regulate cell cycle progression by modulating transcription.⁶⁹ Surprisingly, mouse MALAT1 knockout models have no strong phenotypes, so its physiological roles remain somewhat ambiguous.^{70–72}

The 3'-ends of MALAT1 are generated using a unique endonucleolytic cleavage mechanism (Figure 5(a)).^{73,74} A long (approximately 8.7 kb) polyadenylated MALAT1 transcript is generated using a canonical polyadenylation signal, but this transcript represents only a small (<1%) fraction of the MALAT1 in the cell. Instead, the 3'-end of the abundant MALAT1 is generated by endonucleolytic cleavage of a 61-nt tRNA-like transcript called the MALAT1-associated small cytoplasmic RNA (mascRNA). The mascRNA 5' and 3'-ends are generated by the same endonucleases involved in tRNA production, RNase P, and RNase Z, respectively. MascRNA is exported to the cytoplasm, but its function remains unknown. The RNase P cleavage site generates the mature 3'-end of the stable nuclear MALAT1 ncRNA. This region includes four domains conserved from *Xenopus* to humans: two U-rich elements (Figure 5(a), yellow and red), a predicted stem (white), and a 3'-terminal A-rich motif (orange). The arrangement of U-rich motifs upstream of a terminal A-tract led to the proposal that these elements may serve as an ENE-like element.⁷³ This hypothesis has since been validated by the demonstration that both structures form triple helices and protect transcripts from decay.

Work from the Steitz and Sharp labs independently uncovered the formation of a triple helix at the 3' terminus of MALAT1.^{14,64} To further explore the processing of the MALAT1-mascRNA ncRNAs, Wilusz et al.¹⁴ inserted the 3' elements from mouse MALAT1 into a GFP-based reporter system. They included both the U-rich motifs, the conserved stem, the A-rich tract, and the mascRNA. Together, these elements recapitulate the MALAT1 3' processing pathway and produce a stable transcript that accumulates primarily in the cytoplasm, consistent with the idea that these elements are not strong nuclear retention factors. Extensive mutational analysis further defined a minimal MALAT1 3' stabilizing element (called Comp14) that included the U- and A-rich motifs with a smaller stem (Figure 5(b)). Brown and colleagues used the triple helix structure from the PAN RNA ENE as a guide to identification of similar elements in cellular RNAs.⁶⁴ They focused on abundant nuclear ncRNAs that have the same predicted secondary structural features as the ENE and identified the 3' terminus of MALAT1 as a candidate. Reporter assays confirmed that the identified MALAT1 ENE-like element is sufficient to stabilize RNA *in cis*. Interestingly, unlike PAN RNA, the mascRNA was also needed for full activity; that is, they did not stabilize a polyadenylated transcript produced using the canonical 3'-end formation machinery. Together, these studies defined potential ENE-like triple helices in cellular RNAs.

Even though a structure has not been solved for the MALAT1 ENE-like element, multiple lines of evidence verify that it forms a triple helix between the U- and A-rich conserved sequences. First, a tertiary structural modeling approach predicts that this helix will form.¹⁴ Second, the melting profile of the MALAT1 ENE-like element is biphasic, consistent with melting of the Hoogsteen triple followed by melting of the Watson-Crick strand.⁶⁴ Third, U•A–U→C•G–C substitutions are functional in vivo whereas corresponding double mutations are not.^{14,64} Fourth, the co-conservation of U- and A-rich elements and the similarity to the known structure of the PAN RNA ENE support the formation of the triple helix. Taken together, these data demonstrate that MALAT1 contains a conserved triple-helix structure at its 3'-end after processing and release of the mascRNA (Figure 5(a)).

The MALAT1 ENE-like element has nine predicted U•A–U base triples, but these U•A–U triples are interrupted by C–G and G–C base pairs on the Watson-Crick strands and a single C on the Hoogsteen strand (Figure 5(b), light blue). It is not immediately obvious why the MALAT1 ENE-like element would include these GC nucleotides instead of having an extended, and presumably more stable, U•A–U triple helical structure. While the structural model does not predict the formation of a C•G–C triple helix,¹⁴ pH-dependent changes in biphasic melting curves suggest the formation of a C•G–C triple within the MALAT1 structure.⁶⁴ Mutational analyses suggest that U•A–U triples at the top and bottom are sufficient for function in the Comp14 element, but triples in the central portion, including the putative C•G–C triple, are less critical.¹⁴ Even so, the Watson-Crick pairing throughout the eleven nucleotides is absolutely required for activity. One proposed function for the G–C base pairs is to ensure that the 3'-most A forms a blunt end with the triple helix. In fact, insertion of additional adenosines to the 3'-end of the A-rich region to create a structure with a single-stranded 3' end extending from the triple helix decreases RNA accumulation in vivo (see Box 1).⁶⁴

THE MALAT1 TRIPLE HELIX PROMOTES TRANSLATION

Perhaps, the most surprising revelation regarding the function of the MALAT1 triple helix is that it supports efficient translation.¹⁴ The GFP reporters used to define the element produce ample levels of GFP protein. In fact, protein production is comparable to corresponding reporters with a conventional polyadenylation signal in two different contexts. Moreover, the translational stimulation is not driven exclusively by increased RNA stability. In fact, mutations of the stems adjacent to the triple helix abrogate translational stimulation but have no effect on cytoplasmic RNA accumulation (shaded green, Figure 5(b)).

The discovery that the MALAT1 triple helix supports translation was unexpected for two primary reasons. First, a poly(A) tail promotes efficient translation by stabilizing mRNAs in the cytoplasm and by recruiting the poly(A) binding protein, an important translation factor.⁷⁹ Second, the triple-helix containing MALAT1 transcript is primarily nuclear and presumed to be noncoding. However, sequences near the 5'-end of MALAT1 are protected in ribosomal profiling assays.¹⁴ Therefore, a fraction of MALAT1 transcript may in fact be translated, even though no conserved ORFs have been identified. Other suspected ncRNAs have been shown to encode small polypeptides.^{80–82} Perhaps, these observations regarding

MALAT1 further question the assumption that all RNAs without an apparent coding region are bona fide ncRNAs.

RNA TRIPLE HELICES AS TOOLS FOR DISCOVERY

Each of the four RNA triple helices described above has unique features that highlight the diversity of RNA structure and function. Moreover, each example provides essential mechanistic insights into the particular biological process for which it has evolved. However, triple helices may have significantly more widespread potential as tools to further our understanding of biological processes. For example, a triple helix structure has been identified in the catalytic domain of the *Oceanobacillus iheyensis* Group II self-splicing intron.¹³ In addition to elucidating the mechanism of the Group II intron, the structure yields important predictions regarding the role of tertiary structural elements at the core of the human spliceosome.⁸³ Thus, the structure of the *O. iheyensis* Group II self-splicing intron can serve as a framework for testing specific and detailed mechanistic questions for human pre-mRNA splicing. In addition, examination of the PAN RNA ENE has led to important insights into the nature of nuclear RNA decay pathways, particularly with respect to the role of the poly(A) tail.⁵⁵ The MALAT1 ENE-like element was cleverly used to demonstrate that the presence of a poly(A) tail was not an absolute requirement for miRNA-mediated translational silencing.¹⁴ Comparison of the ends of GFP reporter mRNAs containing or lacking functional MALAT1 ENE-like element transcripts implicated uridylation as a factor in the decay of nonpolyadenylated transcripts.¹⁴ Given the relatively unexplored arena of RNA triple helix structures, exploration of the TER triple helix in atomic detail could conceivably lead to novel drug development opportunities. Thus, mechanistic dissection of RNA triple helices is important to understand the functions for each specific triple helix, but may additionally uncover unexpected mechanisms in cell biology.

SUMMARY

The pioneering work of Felsenfeld, Davies, and Rich revealed that poly(A) and poly(U) could efficiently form triple helix structures *in vitro*. More recently, RNA triple helices have been found in a number of functional RNAs. The triple helix in the SAM-II riboswitch forms a highly specific ligand-binding pocket for SAM and its formation is central to *cis*-acting gene regulation in the SAM-II riboswitch. The function of the TER triple helix remains somewhat ambiguous, but it is clear that this structure is essential for telomerase activity. Given the importance of telomerase to cancer development, unraveling the mechanistic details of the triple helix function of TER has important implications for human health. The PAN RNA ENE and the MALAT1 ENE-like triple helices both contribute to the stability of long nuclear ncRNAs by sequestration of the RNA 3'-ends. The MALAT1 ENE-like element has an additional activity that allows it to support translation. The examination of these triple helix structures has contributed to our knowledge of cell biology and has strong prospects for uncovering future discoveries. Over 50 years after their initial description, it seems we are only beginning to understand the importance of RNA triple helices in biology.

Acknowledgments

Thanks to Drs. Neal Alto and Yunsun Nam for help with PyMOL. Dr. Conrad is funded by the NIH-NIAID grant AI081710, by the Cancer Prevention and Research Institute of Texas RP110132 and by Welch Foundation Grant I-1732. N.K.C. is a Southwestern Medical Foundation Scholar in Biomedical Research.

REFERENCES

1. Cech TR. Structural biology. The ribosome is a ribozyme. *Science*. 2000; 289:878–879. [PubMed: 10960319]
2. Rich A, Felsenfeld G. Formation of a threestranded polynucleotide molecule. *J Am Chem Soc*. 1957; 79:2023–2024.
3. Felsenfeld G, Rich A. Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim Biophys Acta*. 1957; 1000:87–98. [PubMed: 2673412]
4. Buske FA, Mattick JS, Bailey TL. Potential in vivo roles of nucleic acid triple-helices. *RNA Biol*. 2011; 8:427–439. [PubMed: 21525785]
5. Rich A. Discovery of the hybrid helix and the first DNA-RNA hybridization. *J Biol Chem*. 2006; 281:7693–7696. [PubMed: 16547011]
6. Rich A. The era of RNA awakening: structural biology of RNA in the early years. *Q Rev Biophys*. 2009; 42:117–137. [PubMed: 19638248]
7. Hoogsteen K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Cryst*. 1963; 16:907–916.
8. Gilbert SD, Rambo RP, Van Tyne D, Batey RT. Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. *Nat Struct Mol Biol*. 2008; 15:177–182. [PubMed: 18204466]
9. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell*. 2005; 17:671–682. [PubMed: 15749017]
10. Chastain M, Tinoco I. Nucleoside triples from the group I intron. *Biochemistry*. 1993; 32:14220–14228. [PubMed: 8260508]
11. Su L, Chen L, Egli M, Berger JM, Rich A. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat Struct Biol*. 1999; 6:285–292. [PubMed: 10074948]
12. Mitton-Fry RM, DeGregorio SJ, Wang J, Steitz TA, Steitz JA. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science*. 2010; 330:1244–1247. [PubMed: 21109672]
13. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science*. 2008; 320:77–82. [PubMed: 18388288]
14. Wilusz JE, JnBaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, Sharp PA. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev*. 2012; 26:2392–2407. [PubMed: 23073843]
15. Chastain M, Tinoco I. A base-triple structural domain in RNA. *Biochemistry*. 1992; 31:12733–12741. [PubMed: 1463745]
16. Serganov A, Nudler E. A decade of riboswitches. *Cell*. 2013; 152(1–2):17–24. [PubMed: 23332744]
17. Roje S. *S*-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry*. 2006; 67:1686–1698. [PubMed: 16766004]
18. Wang JX, Breaker RR. Riboswitches that sense *S*-adenosylmethionine and *S*-adenosylhomocysteine. *Biochem Cell Biol*. 2008; 86:157–168. [PubMed: 18443629]
19. Poiata E, Meyer MM, Ames TD, Breaker RR. A variant riboswitch aptamer class for *S*-adenosylmethionine common in marine bacteria. *RNA*. 2009; 15:2046–2056. [PubMed: 19776155]
20. Weinberg Z, Regulski EE, Hammond MC, Barrick JE, Yao Z, Ruzzo WL, Breaker RR. The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA*. 2008; 14:822–828. [PubMed: 18369181]

21. Parker BJ, Moltke I, Roth A, Washietl S, Wen J, Kellis M, Breaker R, Pedersen JS. New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. *Genome Res.* 2011; 21:1929–1943. [PubMed: 21994249]
22. Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, Mandal M, Rudnick ND, Breaker RR. Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol.* 2005; 6:R70. [PubMed: 16086852]
23. Doshi U, Kelley JM, Hamelberg D. Atomic-level insights into metabolite recognition and specificity of the SAM-II riboswitch. *RNA.* 2012; 18:300–307. [PubMed: 22194311]
24. Haller A, Rieder U, Aigner M, Blanchard SC, Micura R. Conformational capture of the SAM-II riboswitch. *Nat Chem Biol.* 2011; 7:393–400. [PubMed: 21532598]
25. Chen B, Zuo X, Wang Y-X, Dayie TK. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.* 2012; 40:3117–3130. [PubMed: 22139931]
26. Kelley JM, Hamelberg D. Atomistic basis for the on-off signaling mechanism in SAM-II riboswitch. *Nucleic Acids Res.* 2010; 38:1392–1400. [PubMed: 19969538]
27. Blackburn EH, Collins K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol.* 2011; 3:a003558–a003558. [PubMed: 20660025]
28. Shay JW, Wright WE. Role of telomeres and telomerase in cancer. *Semin Cancer Biol.* 2011; 21:349–353. [PubMed: 22015685]
29. Günes C, Rudolph KL. The role of telomeres in stem cells and cancer. *Cell.* 2013; 152:390–393. [PubMed: 23374336]
30. Wyatt HDM, West SC, Beattie TL. InTERTpreting telomerase structure and function. *Nucleic Acids Res.* 2010; 38:5609–5622. [PubMed: 20460453]
31. Gillis AJ, Schuller AP, Skordalakes E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature.* 2008; 455:633–637. [PubMed: 18758444]
32. Kim N-K, Zhang Q, Zhou J, Theimer CA, Peterson RD, Feigon J. Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J Mol Biol.* 2008; 384:1249–1261. [PubMed: 18950640]
33. Egan ED, Collins K. Biogenesis of telomerase ribonucleoproteins. *RNA.* 2012; 18:1747–1759. [PubMed: 22875809]
34. Chen JL, Blasco MA, Greider CW. Secondary structure of vertebrate telomerase RNA. *Cell.* 2000; 100:503–514. [PubMed: 10721988]
35. Chen J-L, Greider CW. An emerging consensus for telomerase RNA structure. *Proc Natl Acad Sci USA.* 2004; 101:14683–14684. [PubMed: 15466703]
36. Theimer CA, Feigon J. Structure and function of telomerase RNA. *Curr Opin Struct Biol.* 2006; 16:307–318. [PubMed: 16713250]
37. Qiao F, Cech TR. Triple-helix structure in telomerase RNA contributes to catalysis. *Nat StructMol Biol.* 2008; 15:634–640.
38. Shefer K, Brown Y, Gorkovoy V, Nussbaum T, Ulyanov NB, Tzfati Y. A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol Cell Biol.* 2007; 27:2130–2143. [PubMed: 17210648]
39. Huang C, Yu Y-T. Targeted 2'-O methylation at a nucleotide within the pseudoknot of telomerase RNA reduces telomerase activity in vivo. *Mol Cell Biol.* 2010; 30:4368–4378. [PubMed: 20647541]
40. Liu F, Theimer CA. Telomerase activity is sensitive to subtle perturbations of the TLC1 pseudoknot 3' stem and tertiary structure. *J Mol Biol.* 2012; 423:719–735. [PubMed: 22954661]
41. Mesri EA, Cesarman E, Boshoff C. Kaposi's sarcoma and its associated herpesvirus. *Nat Rev Cancer.* 2010; 10:707–719. [PubMed: 20865011]
42. Cai Q, Verma SC, Lu J, Robertson ES. Molecular biology of Kaposi's sarcoma-associated herpesvirus and related oncogenesis. *Adv Virus Res.* 2010; 78:87–142. [PubMed: 21040832]
43. Wen KW, Damania B. Kaposi sarcoma-associated herpesvirus (KSHV): molecular biology and oncogenesis. *Cancer Lett.* 2010; 289:140–150. [PubMed: 19651473]

44. Song MJ, Brown HJ, Wu TT, Sun R. Transcription activation of polyadenylated nuclear RNA by RTA in human herpesvirus 8/Kaposi's sarcoma-associated herpesvirus. *J Virol.* 2001; 75:3129–3140. [PubMed: 11238840]
45. Sun R, Lin SF, Gradoville L, Miller G. Polyadenylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. *Proc Natl Acad Sci USA.* 1996; 93:11883–11888. [PubMed: 8876232]
46. Zhong W, Wang H, Herndier B, Ganem D. Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. *Proc Natl Acad Sci USA.* 1996; 93:6641–6646. [PubMed: 8692871]
47. Rossetto CC, Pari G. KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome. *PLoS Pathog.* 2012; 8:e1002680. [PubMed: 22589717]
48. Borah S, Daricarrère N, Darnell A, Myoung J, Steitz JA. A viral nuclear noncoding rna binds re-localized poly(A) binding protein and is required for late KSHV gene expression. *PLoS Pathog.* 2011; 7:e1002300. [PubMed: 22022268]
49. Rossetto CC, Pari GS. KSHV noncoding PAN RNA interacts with virus and cellular-encoded proteins and suppresses expression of genes involved in immune modulation. *J Virol.* 2011
50. Rossetto CC, Tarrant-Elorza M, Verma S, Purushothaman P, Pari GS. Regulation of viral and cellular gene expression by Kaposi's sarcoma associated herpesvirus (KSHV) PAN RNA. *J Virol.* 2013
51. Shamovsky I, Nudler E. Gene control by large noncoding RNAs. *Sci STKE.* 2006; 2006:pe40. [PubMed: 17018852]
52. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell.* 2013; 152:1298–1307. [PubMed: 23498938]
53. Conrad NK, Steitz JA. A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. *EMBO J.* 2005; 24:1831–1841. [PubMed: 15861127]
54. Conrad NK, Shu M-D, Uyhazi KE, Steitz JA. Mutational analysis of a viral RNA element that counteracts rapid RNA decay by interaction with the polyadenylate tail. *Proc Natl Acad Sci USA.* 2007; 104:10412–10417. [PubMed: 17563387]
55. Conrad NK, Mili S, Marshall EL, Shu M-D, Steitz JA. Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. *Mol Cell.* 2006; 24:943–953. [PubMed: 17189195]
56. Schmid M, Jensen TH. Quality control of mRNP in the nucleus. *Chromosoma.* 2008; 117:419–429. [PubMed: 18563427]
57. Fasken MB, Corbett AH. Mechanisms of nuclear mRNA quality control. *RNA Biol.* 2009; 6:237–241. [PubMed: 19574733]
58. Doma MK, Parker R. RNA quality control in eukaryotes. *Cell.* 2007; 131:660–668. [PubMed: 18022361]
59. Doherty EA, Batey RT, Masquida B, Doudna JA. A universal mode of helix packing in RNA. *Nat Struct Biol.* 2001; 8:339–343. [PubMed: 11276255]
60. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci USA.* 2001; 98:4899–4903. [PubMed: 11296253]
61. Tycowski KT, Shu M-D, Borah S, Shi M, Steitz JA. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *CellReports.* 2012; 2:26–32.
62. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. A giant virus in amoebae. *Science.* 2003; 299:2033. [PubMed: 12663918]
63. Bonning BC, Miller WA. Dicistroviruses. *Annu Rev Entomol.* 2010; 55:129–150. [PubMed: 19961327]
64. Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MEN β noncoding RNAs. *Proc Natl Acad Sci.* 2012
65. Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and

- survival in early-stage nonsmall cell lung cancer. *Oncogene*. 2003; 22:8031–8041. [PubMed: 12970751]
66. Gutschner T, Hämmerle M, Diederichs S. MALAT1—a paradigm for long noncoding RNA function in cancer. *J Mol Med*. 2013
 67. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010; 39:1–14. [PubMed: 20603069]
 68. Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, Zhang Y, Gorospe M, Prasanth SG, Lal A, et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet*. 2013; 9:e1003368. [PubMed: 23555285]
 69. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*. 2011; 147:773–788. [PubMed: 22078878]
 70. Eißmann M, Gutschner T, Hämmerle M, Günther S, Caudron-Herger M, Groß M, Schirmacher P, Rippe K, Braun T, Zörnig M, et al. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol*. 2012; 9:1076–1087. [PubMed: 22858678]
 71. Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, et al. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *CellReports*. 2012; 2:111–123.
 72. Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X, Hirose T, Prasanth KV. Malat1 is not an essential component of nuclear speckles in mice. *RNA*. 2012; 18:1487–1499. [PubMed: 22718948]
 73. Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell*. 2008; 135:919–932. [PubMed: 19041754]
 74. Wilusz JE, Spector DL. An unexpected ending: noncanonical 3' end processing mechanisms. *RNA*. 2010; 16:259–266. [PubMed: 20007330]
 75. Pijlman GP, Funk A, Kondratieva N, Leung J, Torres S, van der Aa L, Liu WJ, Palmenberg AC, Shi P-Y, Hall RA, et al. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe*. 2008; 4:579–591. [PubMed: 19064258]
 76. Mullen TE, Marzluff WF. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev*. 2008; 22:50–65. [PubMed: 18172165]
 77. Scott DD, Norbury CJ. RNA decay via 3' uridylation. *Biochim Biophys Acta*. 2013
 78. Anderson JT. RNA turnover: unexpected consequences of being tailed. *Curr Biol*. 2005; 15:R635–R638. [PubMed: 16111937]
 79. Kühn U, Wahle E. Structure and function of poly(A) binding proteins. *Biochim Biophys Acta*. 2004; 1678(2–3):67–84. [PubMed: 15157733]
 80. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol*. 2007; 5:e106. [PubMed: 17439302]
 81. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; 147:789–802. [PubMed: 22056041]
 82. Xu Y, Ganem D. Making sense of antisense: seemingly noncoding RNAs antisense to the master regulator of Kaposi's sarcoma-associated herpesvirus lytic replication do not regulate that transcript but serve as mRNAs encoding small peptides. *J Virol*. 2010; 84:5465–5475. [PubMed: 20357088]
 83. Keating KS, Toor N, Perlman PS, Pyle AM. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA*. 2009; 16:1–9. [PubMed: 19948765]

BOX 1**HOW DO ENE-LIKE ELEMENTS PROTECT TRANSCRIPTS FROM DECAY?**

The observation the MALAT1 ENE-like element is significantly more active when the triple helix has a blunt end strongly supports the model that the triple helix functions by sequestration of the RNA 3'-end from the decay machinery. This is in contrast to a flavivirus pseudoknot that stops the progression of the XRN1 5'-to-3' exonuclease with a 'roadblock' mechanism.⁷⁵ If having the 3'-most adenosine blunt ended in the triple helix is essential to inhibit decay, then it is perplexing how the PAN RNA ENE, which contacts long poly(A) tails using exclusively U•A–U triples, inhibits processive degradation. Perhaps, regions outside of the core ENE,⁵⁴ surrounding sequences in PAN RNA,⁵³ or interplay with poly(A) binding proteins⁴⁸ contributes to the full activity of PAN RNA ENE-mediated inhibition of RNA decay. It is also not clear which enzyme(s) is inhibited by sequestration of the RNA 3'-ends. For example, GFP mRNAs with mutated MALAT1 triple helices undergo uridylation-dependent decay.^{14,76,77} Deletion of the ENE from PAN RNA leads to poly(A) tail extension⁵⁴ and unpublished data strongly support the idea that this hyperadenylation is coupled to transcript decay (Bresson and Conrad, *PlosGenetics* in press). Taken together, these results raise the possibility that the sequestration of the poly(A) tail is not protecting it from the exonucleases directly, but may be keeping it from tailing activities that are linked to exonucleolytic decay.^{77,78} Further work with the ENE and ENE-like elements can therefore reveal important mechanistic insights into cellular RNA decay pathways.

FIGURE 1.

U•A–U and C•G–C base triples. (a) U•A–U and (b) C•G–C base triples with the Hoogsteen and Watson-Crick interactions shown. Hydrogen bonds are shown as broken lines. The cytosine N3 of the Hoogsteen interaction is in the protonated state to yield an additional hydrogen bond.

FIGURE 2.

The SAM-II triple helix. (a) Schematic depiction of the SAM-II riboswitch structure. The triple helix is shown as yellow, red, and orange nucleotides. Watson-Crick interactions are shown with solid lines whereas nonconventional and Hoogsteen base interactions within the triple helix are shown with solid circles. Tertiary interactions and nonconventional base pairs outside the triple helix are shown with dashed lines. SAM is blue (A_{SAM}), and the Shine-Delgarno sequence (A47–G52) is shaded with a green box. (b) Surface representation of the SAM-II riboswitch. The three strands corresponding to the triple helix and SAM are color coded as in (a). All structures for SAM-II were from PDB ID: 2QWY.⁸ (c) Stick

representation of the triple helix domain in the SAM-II riboswitch showing the fit of SAM in the ligand-binding pocket. The three strands corresponding to the triple helix and SAM are color coded as in (a). (d) Base triples and hydrogen bonds between the SAM-II triple helix and the SAM adenosine. Nitrogens are dark blue, oxygens are bright red, hydrogens are white, carbons are color coded by triple helix strand as in (a). Hydrogen bonds are shown with dashed lines.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 3.

The TER triple helix. (a) The secondary structure of the TER. The line drawing depicts the full-length TER secondary structure and highlights the positions of the box H/ACA snoRNA-like elements and the template. The sequence of the pseudoknot used for the NMR structure is also shown.³² Note that nt 121–165 were deleted in the structure determination of the pseudoknot. Watson-Crick pairs are shown with solid lines and Hoogsteen or other tertiary interactions in the triple helix are shown with dotted lines. The three strands that compose the triple helix are colored in yellow, red, and orange. The U99•A173 Hoogsteen

base pair is light blue. (b) Surface representation of the TER pseudoknot. The three strands corresponding to the triple helix are colored as in (a). All structures for the TER pseudoknot were derived from PDB ID: 2 K95.³² (c) The TER triple helix including the conventional U•A–U triples at the bottom, the U99•A173 Hoogsteen base pair, and the unconventional base triples at the top. The colors are the same as (a). (d) Base triples and hydrogen bonds between nonconventional triples at the top of the TER triple helix. Nitrogens are dark blue, oxygens are bright red, hydrogens are white, and carbons are color coded as in (a). Hydrogen bonds are shown as dashed lines.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 4.

The PAN ENE triple helix. (a) *Left* Predicted secondary structure of the 79-nt PAN ENE. The numbers are relative to the PAN RNA transcription start site; the polyadenylation site is at position 1077 (not shown). The upper and lower stems are colored in light blue and dark blue, respectively. The five-uridine stretches that participate in the triple helix are shown in yellow and red. *Right* : Schematic of the triple helix formed by the core ENE in complex with oligoA₉ (orange). The Hoogsteen interactions in the triple helix are shown with filled circles and Watson-Crick base pairs are depicted with solid lines. A-minor interactions are shown as dashed lines. (b) Surface representation of the PAN RNA ENE. Color-coding is the same as in (a). All PAN RNA ENE structures were derived from PDB ID: 3P22.¹² (c) The triple helix and lower stem of the PAN RNA ENE in complex with A2–A9 of oligoA₉. The U•A–U triple helix and the interface of the A-minor interactions between the minor groove of the lower G–C stem with A2–A4 are indicated. A952–U953 (gray) are ‘flipped out’ relative to the bases in the helix. (d) Close-up view of the 3′-most adenosine, A9

(orange, stick representation) as it fits in the pocket formed from the U-rich strands (yellow and red, surface representation) and the upper stem (light blue, surface representation).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 5.

The MALAT1 ENE-like triple helix. (a) RNA processing pathway for MALAT1 and mascRNA. For simplicity, the polyadenylated species is shown as the primary transcript, but it is not clear whether the RNase P cleavage occurs on polyadenylated MALAT1 or co-transcriptionally.⁷³ The diagram is not to scale; the highlighted elements compose only a small fraction of the total MALAT1 length. The numbers refer to the human MALAT1, but all the relevant domains are conserved. (b) Secondary structure and triple helix formation in the MALAT1 ENE-like element. The structure on the left corresponds to the full-length

human MALAT1 ENE-like element, while the structure on the right is a minimal element derived from murine MALAT1, Comp14.¹⁴ The U-rich sequence that forms Watson-Crick base pairs is shown in red, the A-rich stretch is orange, and the U-rich tract that forms Hoogsteen base triples is yellow. The Hoogsteen interactions in the triple helix are shown with filled circles and Watson-Crick base pairs or triples are depicted with solid lines; predicted A-minor interactions are shown as dashed lines. The central C and G nucleotides are shown in light blue. The C•G–C is depicted as part of the helix based on pH profiles,⁶⁴ but this triple was not observed in a structural model.¹⁴ The lower G–C stem is shown in dark blue. The sequences important for translation are highlighted in the green box.