Published in final edited form as:
Nat Struct Mol Biol. 2014 July ; 21(7): 633–640. doi:10.1038/nsmb.2844.

Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix

Jessica A. Brown[1], **David Bulkley**[2], **Jimin Wang**[1], **Max L. Valenstein**[1], **Therese A. Yario**[3], **Thomas A. Steitz**[1,2,3], and **Joan A. Steitz**[1,3]

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

[2]Department of Chemistry, Yale University, New Haven, CT USA

[3]Howard Hughes Medical Institute, Yale University, New Haven, CT USA

Abstract

Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) is a highly-abundant nuclear long noncoding RNA that promotes malignancy. A 3′-stem-loop structure is predicted to confer stability by engaging a downstream A-rich tract in a triple helix, similar to the expression and nuclear retention element (ENE) from the KSHV polyadenylated nuclear RNA. The 3.1-Å resolution crystal structure of the human MALAT1 ENE and A-rich tract reveals a bipartite triple helix containing stacks of five and four U•A-U triples separated by a C^+•G-C triplet and C-G doublet, extended by two A-minor interactions. *In vivo* decay assays indicate that this blunt-ended triple helix, with the 3′ nucleotide in a U•A-U triple, inhibits rapid nuclear RNA decay. Interruption of the triple helix by the C-G doublet induces a “helical reset” that explains why triple-helical stacks longer than six do not occur in nature.

Long noncoding RNAs (lncRNAs) function in myriad cellular processes and are associated with various disease states, including cancer[1,2]. Human metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) is an abundant, ~8-kb lncRNA that is upregulated in multiple cancers[3]. High nuclear levels of MALAT1, with a half-life up to 15 h, promote tumor growth by affecting proliferation, invasion and metastasis, processes associated with altered gene expression in lung cancer[3–9]. The enhancement of oncogenic processes by MALAT1 in colorectal cancer cells has been localized to a ~1500-nucleotide (nt) fragment near the 3′ end (nts 6918–8441)[5]. This region includes the highly-conserved 3′-end sequence

Corresponding author: Joan A. Steitz, Boyer Center for Molecular Medicine, 295 Congress Avenue, New Haven, CT 06536-9812 USA, joan.steitz@yale.edu, Phone: (203) 737 4418, Fax: (203) 624 8213.

SUPPLEMENTARY INFORMATION
Supplementary information includes seven figures, six tables, and references.

Accession codes. Coordinates and structural factors for the reported crystal structure have been deposited in the Protein Data Bank under accession code 4PLX.

AUTHOR CONTRIBUTIONS
J.A.B. and J.A.S. designed research; J.A.B., D.B., M.L.V. and T.A.Y. performed research; J.A.B., D.B. and J.W. analyzed data; T.A.S. and J.A.S. oversaw research; J.A.B. and J.A.S. wrote the paper; all authors discussed the results and commented on the manuscript.

of MALAT1 (nts 8254–8413), which contains a genomically-encoded A-rich tract flanked by two structural elements: a proposed stem-loop structure that resembles an ENE (expression and nuclear retention element, Fig. 1a) and a downstream tRNA-like structure known as mascRNA (MALAT1-associated small cytoplasmic RNA)[10–12]. MascRNA is processed by RNase P, generating the 3′ end of MALAT1 with a 3′-terminal A-rich tract predicted to engage in a triple-helical ENE structure based on its similarity to ENE structures discovered in viral lncRNAs and genomic RNAs[13,14].

The ENE near the 3′ end of polyadenylated nuclear (PAN) RNA (Fig. 1a), a lncRNA produced by the Kaposi's sarcoma-associated herpesvirus (KSHV) during the lytic phase of infection, is the best characterized[13,15–17]. It protects the 3′ end of PAN RNA from rapid deadenylation-dependent nuclear decay, forming a triple helix by sequestration of PAN's 3′-poly(A) tail within the U-rich internal loop, denoted ENE+A[15,17]. Structural components important for robust activity of the PAN ENE+A include: (i) a triple helix of five consecutive U•A-U triples (where • and – represent interactions along the Hoogsteen and Watson-Crick faces, respectively); (ii) canonical Watson-Crick base pairs in the duplexes flanking the triplex; and (iii) A-minor interactions with three consecutive G-C base pairs adjacent to the triplex (Fig. 1a and Supplementary Fig. 1a)[16,17]. Similar structural features were predicted for the MALAT1 ENE+A, including two A-minor interactions; however, the predicted triple helix would be markedly different from that of any viral ENE because G and C nucleotides interrupt the U-rich internal loop (Fig. 1a)[11,12,14,17].

We set out to determine the crystal structure of the MALAT1 ENE+A, presented here at 3.1 Å resolution. It forms a bipartite triple helix that sequesters the 3′ end of the RNA within a U•A-U triple, conferring resistance to rapid RNA decay. The U•A-U triplex is interrupted by a C^+•G-C triplet and adjacent C-G doublet that induces a "helical reset", suggesting that successive base triples are limited to a finite length. This ENE structure is a major determinant of MALAT1 stability, identifying a potential target for reducing MALAT1 levels in cancer cells.

RESULTS

An ENE facilitates accumulation of MALAT1 lncRNA

We tested whether the ENE is responsible for the high levels of cellular MALAT1[3]. An ~80-bp region containing the ENE (ΔENE) was deleted from plasmids expressing full-length (~6.9-kb) mouse MALAT1 (mMALAT1) or a ~2-kb sequence from the 3′ end of human MALAT1 (Fig. 1b, c). Mouse and human MALAT1 ENE+A sequences are ~90% identical (Supplementary Table 1). Upon transient expression in HEK293T cells, transcripts lacking the ENE accumulate to only ~1.5% the level of wild-type (WT) mouse or human MALAT1 RNA (Fig. 1d, e). A single U to C base substitution on the 5′ side of the U-rich internal loop decreases RNA levels to 23% and 15% for mouse U6612C and human U8275C MALAT1, respectively (Fig. 1a–e and Supplementary Table 2). These dramatic reductions underscore the ENE's contribution to the high nuclear accumulation of endogenous MALAT1.

Determination of the MALAT1 ENE+A structure

Stabilization is predicted to result from the ability of the U-rich internal loop of the MALAT1 ENE to sequester the downstream A-rich tract in an RNA triple helix[11,12]. We solved the X-ray crystal structure of this region using a modified MALAT1 ENE engaged in *cis* with its A-rich tract at 3.1 Å resolution (Fig. 1a, f, g and Table 1). The modified MALAT1 ENE+A core has a 5′ G, a six-nucleotide deletion of the predicted single-stranded linker, a hairpin structure with a GAAA tetraloop for crystal contacts (Supplementary Fig. 2a) and cation-binding site (nts G22-C24 and G29-C31) for iridium (III) hexamine trichloride. Its transcript levels are comparable to the WT MALAT1 ENE+A, indicating full functionality (Fig. 1e, WT vs Core).

Initial experimental phases were determined at 3.4 Å resolution from a single iridium-soaked crystal in space group $P3_221$ using the multiwavelength anomalous dispersion (MAD) method (Fig. 1f and Supplementary Fig. 3, see online Methods). Structural models were built for all three RNA molecules in the asymmetric unit (Supplementary Fig. 2). The model was then refined against a native data set at 3.1 Å resolution ($I/\sigma I = 1.28$ for the 3.10–3.18 Å resolution shell) with working and free R factors (R_{work} and R_{free}) of 22.0% and 25.5%, respectively, and an overall coordinate error based on maximum likelihood of 0.34 Å (Fig. 1f, Supplementary Fig. 3c and Table 1). Data analyses and figures are based on the most ordered RNA molecule A, whose conformation lies between molecules B and C, with electron density visible for all 76 nts (Supplementary Fig. 2b).

MALAT1 ENE and A-rich tract form a bipartite triple helix

The MALAT1 ENE+A core assembles into an intramolecular major-groove RNA triple helix formed by the ENE U-rich internal loop and downstream A-rich tract, connected by a single-stranded linker (Figs. 1g and 2a). The triple helix is bipartite, composed of two independent stacks of triples: five consecutive U•A-U triples (#1–5 of triplex I in Fig. 2a) capped with one C^+•G-C triple and four consecutive U•A-U triples (#6–9 of triplex II in Fig. 2a). The stacks are separated by one C-G doublet. The bases in the C^+•G-C and most of the U•A-U triples are within hydrogen-bonding distance (≤3.2 Å) on both the Watson-Crick and Hoogsteen faces (Fig. 2b). The 3′ terminus of the molecule is sequestered within U•A-U[9] of triplex II. The bipartite triplex is flanked by a GG dinucleotide bulge and two double-helical RNA stems (Fig. 2a). G6-C50 and G5-C51 in stem I interact with A65 and A64 of the A-rich tract to form type I and type II A-minor interactions, respectively (Fig. 2c and Supplementary Fig. 4). Together, these double- and triple-helical units assemble into a nearly straight, rod-like structure similar to the KSHV PAN ENE core structure with oligo A_9 (Supplementary Fig. 1)[17].

Although all 10 major-groove triples maintain a similar overall layout (Fig. 2b), the observed lengths for potential hydrogen bonds between the 2′-OH of the Hoogsteen strand and phosphate oxygen (O2P more often than O1P) of the A-rich tract or Watson strand vary (Fig. 2d and Supplementary Table 3). For triplexes I and II, the 2′-OH-O2P distances gradually decrease from U•A-U[1] to [5] (5.1 to 2.5 Å) and U•A-U[6] to [9] (4.1 to 2.6 Å), creating a zipper-like, hydrogen-bonding pattern. These 2′-OH-O2P hydrogen bonds likely stabilize the highly-electronegative RNA triple-stranded structure by minimizing potential electrostatic clashes between the O2 and O4′ of the Hoogsteen strand and O2P of the Watson strand (Supplementary Table 3)[18,19]. Importantly, an abrupt change in the 2′-OH-O2P distance occurs at the internal triplex junction, implicating the C-G doublet in a unique structural role.

MALAT1 ENE+A triple helix is interrupted by a C$^+$•G-C and C-G

The disruption of the U•A-U triplex by a central C$^+$•G-C triplet and C-G doublet in the MALAT1 ENE+A is absent from all known viral ENE+A structures[14,17]. An overlay of C$^+$•G-C, C-G, and U•A-U[6] with U•A-U[2–4] reveals the Hoogsteen strands are structurally different (Fig. 3a). At the C-G doublet, a one-nucleotide gap in the Hoogsteen strand shifts the backbone inwards by ~3 Å, positioning the Hoogsteen C$^+$12 and U13 within hydrogen-bonding distance of the Watson strand. Furthermore, base-stacking interactions are approximately 4-fold greater for the internal C$^+$•G-C/C-G than for any U•A-U stack because both the Watson and Crick (3′ side of U-rich loop) strands participate, rather than predominantly the Watson strand, *e.g.* A68 and A69 from U•A-U[3] and U•A-U[4] (Fig. 3b). At the C-G doublet, the helical axes of triplexes I and II change their relative orientation by ~15° (Fig. 3c). Likewise, re-orientation of ~13° and 21° is observed between the axes of stems I and II, respectively, and the adjacent triplex. These changes in the local helical axes correlate with changes in both major- and minor-groove widths near duplex-triplex transitions (Supplementary Fig. 5 and Supplementary Table 4). Together, it appears that the C-G doublet enables a “reset” of the helical axis. Because a nucleotide is missing in the Hoogsteen strand, we hypothesize that the Hoogsteen strand can interact with the Watson-Crick duplex to form a finite number of successive major-groove triples.

Predicted steric clash between Hoogsteen and Watson strands

A literature survey reveals that experimentally-determined tertiary structures of naturally-occurring RNA triple helices contain three to six consecutively-stacked major-groove triples (Supplementary Table 5)[20]. Thus, helical reset by the C-G doublet may be a mechanism to circumvent destabilizing interactions occurring with more than six triples. To test this possibility, we computationally extended the Hoogsteen and Watson strands by superimposing a nucleotide at positions “n-2” and “n-1” onto a nucleotide at positions “n-1” and “n”, respectively, to advance the nucleotide at position “n” into the extended position “n +1” (*e.g.* U10 overlaid onto U11) for each of the strands in both triplex I and II (Fig. 4a). This analysis revealed steric clashes between the 2′-OH of the Hoogsteen strands and O2P of the Watson strands in positions C12+1 and G71+1 for triplex I and U16+1 and A76+1 for triplex II (Fig. 4a and Supplementary Table 3). Thus, depending on the specific sequence and structural context, successive base triples of more than four to six are destabilizing because of steric hindrance.

MALAT1 ENE+A structure inhibits the rapid phase of RNA decay

Next, we asked if the MALAT1 ENE+A counteracts the rapid phase of RNA decay similar to the ENE+A structure of KSHV PAN RNA[15–17]. We determined the *in vivo* decay profile for a doxycycline-inducible, intronless β-globin transcript (βΔ1,2) containing the MALAT1 ENE+A+mascRNA sequence in its 3′-untranslated region (Fig. 5a, b). Following a transcriptional pulse of 2 h, the transcript (Fig. 5c, d, green line) underwent a single phase of decay with a half-life ($t_{1/2}$) of 3.4 h (Table 2). Interestingly, the U8275C mutant form of βΔ1,2-MALAT1 ENE+A+mascRNA, which accumulates to ~15% of WT under steady-state conditions (Fig. 1e)[11], exhibited a single, but somewhat faster ($t_{1/2}$=1.3 h), phase of decay (Table 2 and Fig. 5d, blue line). These profiles sharply contrast the biphasic decay of βΔ1,2 and βΔ1,2-KSHV PAN ENE transcripts (Fig. 5d, black and purple lines)[15], which undergo rapid ($t_{1/2}$=~7–12 m) followed by slow ($t_{1/2}$=~3–4 h) decay but differ in the percentage of the transcript population undergoing rapid decay: 39% for the βΔ1,2 versus 23% for the βΔ1,2-KSHV PAN ENE transcript (Fig. 5 and Table 2).

The distinctly different decay profiles for the KSHV PAN and MALAT1 ENE+A suggest different 3′→5′ exonucleolytic mechanisms. We tested whether the different 3′-end structures of the substrates, a poly(A) overhang for the KSHV PAN ENE+A versus a blunt end for the MALAT1 ENE+A, are responsible by inserting the KSHV PAN ENE upstream of an A-rich tract and mascRNA sequence (TRP-βΔ1,2-KSHV PAN ENE+A+mascRNA C^+•G-C), which should yield a blunt-ended structure locked into register by the C•G-C triple substitution (Fig. 5b). Indeed, the βΔ1,2-KSHV PAN ENE+A+mascRNA C^+•G-C transcript decayed with a single, slow phase ($t_{1/2}$=3.5 h), almost identical to that of the βΔ1,2-MALAT1 ENE+A+mascRNA WT transcript (Fig. 5d, red line and Table 2). We conclude that the 3′-blunt-ended MALAT1 ENE+A structure effectively abolishes rapid RNA decay *in vivo*.

Sequence requirements for MALAT1 ENE+A function

The contributions of the C-G doublet and C^+•G-C triplet in preventing RNA decay were probed by mutating the βΔ1,2-MALAT1 ENE+A+mascRNA reporter (Fig. 6a, b). First, we replaced the C-G doublet with a G-G mismatch or the three other Watson-Crick base pairs. G-G dramatically reduced stabilization, suggesting a requirement for base pairing at this position (Fig. 6c, lanes 2 and 3). The Watson-Crick base pair replacement showed relative accumulation levels of C-G (WT) > U-A > G-C = A-U (Fig. 6c, lanes 2 and 4–6), revealing higher accumulation for a pyrimidine-purine pair than purine-pyrimidine. We also inverted the C^+•G-C and C-G to C-G and C^+•G-C, which dropped reporter accumulation to ~39% of WT (Fig. 6c, lane 7). Relocating the C^+•G-C and C-G between U•A-U[2] and U•A-U[3] lowered reporter accumulation to ~76% (Fig. 6c, lane 8). We conclude that a pyrimidine-purine arrangement is more effective than the converse for stabilization and that the location of the doublet within the MALAT1 ENE+A triplex affects reporter accumulation.

The C-G doublet stabilizes a protonated C^+•G-C *in vivo*

The C^+•G-C base triple in the MALAT1 ENE+A structure raises the question of whether the Hoogsteen base C8273 is protonated (C^+) *in vivo*. Protonation would increase triplex stability by promoting hydrogen bond formation between N3 of cytosine and N7 of guanine (Fig. 6d). Our X-ray structure shows that the N3 of C12 and N7 of G71 are within hydrogen-bonding distance at 2.7 Å (Fig. 2b). For a protonated C^+•G-C, two hydrogen bonds are predicted along the Hoogsteen face whereas the unprotonated triple would form one hydrogen bond, as in a U•G-C triple (Fig. 6d). Thus, we tested a U•G-C mutant in the βΔ1,2-MALAT1 ENE+A+mascRNA reporter and found ~69% accumulation relative to WT (Fig.

6c, lanes 2 and 9). This moderate decrease with the loss of one hydrogen bond contrasts with reporter levels of ≤15% that we observed previously after complete loss of hydrogen-bonding interactions along the Hoogsteen-Watson interface using the same β-globin reporter system[11].

In vitro studies of intramolecular DNA triple helices composed of C•G-C and T•A-T triples by UV melting, iron-affinity cleavage and DNase I footprinting assays have found that adjacent C•G-C triples are destabilizing at neutral pH[21–25]. In the MALAT1 ENE+A structure, the C-G doublet could stabilize C^+•G-C by increasing solvent accessibility to satisfy the positive charge or by eliminating electrostatic interference arising from the presence of a third nucleotide. Therefore, we created several βΔ1,2-MALAT1 ENE+A +mascRNA mutants, which delete (Δ) or substitute the C-G doublet with the putative triples C•G-C, U•A-U, C•C-G or U•C-G. Removing the C-G doublet decreased reporter accumulation to ~63% relative to WT (Fig. 6c, lanes 2 and 10), indicating that the doublet stabilizes the C•G-C triple more than a U•A-U triple. Likewise, accumulation was reduced to ~34% and ~66% relative to WT when the C-G doublet was replaced with C•G-C and U•A-U, respectively (Fig. 6c, lanes 2 and 11–12). Because lowered accumulation may be partially due to the loss of the C-G doublet, which confers greater stability than A-U or G-C (Fig. 6c, lanes 2 and 4–5), we examined reporters having C•C-G and U•C-G triples in the C-G position. We observed ~56% and 88% levels relative to WT, respectively (Fig. 6c, lanes 2 and 13–14). Thus, the stabilization activity of the MALAT1 ENE+A is lower when the C•G-C triple is adjacent to a C rather than U in the Hoogsteen strand, in agreement with previous *in vitro* studies of DNA triplexes[21–25]. We conclude that the C-G doublet stabilizes the C^+•G-C triple in the MALAT1 ENE+A *in vivo*, achieving optimal activity when the Hoogsteen C^+ base is in a non-base-stacking position.

DISCUSSION

We have determined the structural basis for the stability of the lncRNA MALAT1: its 3′ end forms a triple-helical ENE+A structure, whose deletion or mutation greatly reduces MALAT1 levels (Fig. 1). The triple helix involves the U-rich internal loop of the MALAT1 ENE and a genomically-encoded downstream A-rich tract, interrupted by a C^+•G-C triple and C-G doublet (Fig. 1a, g). The C^+•G-C and C-G were not predicted by computer-generated modeling of a minimal mMALAT1 ENE+A, which proposed a U•A-U triplex interrupted by C-G and G-C doublets and an unpaired C in the Hoogsteen strand; the two A-minor interactions we have identified were also missing[12].

The 10 major- and two minor-groove base triples revealed by our X-ray analysis create an extensive network of hydrogen bonds that confer *in vivo* accumulation of transcripts terminating with the MALAT1 ENE+A (Figs. 1, 2, 6a and Supplementary Fig. 3)[11]. The U•A-U triples of the MALAT1 ENE+A core closely resemble those of the KSHV PAN ENE core and oligo A_9, overlaying with root mean square deviation (RMSD) values of 0.55 and 0.71 Å for MALAT1 ENE+A triplexes I and II, respectively (Supplementary Fig. 1)[17].

How the MALAT1 ENE+A counteracts decay

Despite their structural similarity, the MALAT1 and KSHV PAN ENE+A differentially counteract decay *in vivo*, exhibiting monophasic and biphasic decay, respectively (Fig. 5 and Supplementary Fig. 1). The KSHV PAN RNA exhibits a rapid followed by a slow phase of decay, with the ENE+A structure reducing the percentage of transcripts undergoing rapid decay[15,16]. Here, we discovered that the MALAT1 ENE+A eliminates the rapid decay phase; instead a single phase with a $t_{1/2}$ similar to the slow phase of PAN ENE+A decay is observed (Fig. 5 and Table 2). Perhaps, the MALAT1 ENE+A is more effective at preventing rapid 3′→5′ exonucleolytic decay because its 3′-A is engaged in a U•A-U triple to form a blunt-ended structure, whose register is determined by the GC dinucleotide, G8350 and C8351, in the A-rich tract of the triple helix (Fig. 6 and Supplementary Fig. 1).

The blunt-end of the MALAT1 triplex appears to be a poor substrate for exonucleases compared to the 3′-poly(A) overhang of KSHV PAN RNA (Fig. 5). When an A_7 overhang was added onto the 3′ end of the βΔ1,2-MALAT1 ENE+A+mascRNA reporter, transcript levels decreased to ~23% relative to WT, indicating accelerated decay[11]. The internal GC in the A-rich tract of the MALAT1 ENE+A may directly slow the rate of deadenylation because purified deadenylases, poly(A)-specific ribonuclease and CCR4, prefer A-containing over G-, C-, and U-containing substrates *in vitro*[26,27]. Based on βΔ1,2-MALAT1 ENE+A+mascRNA reporter assays with Watson-Crick base pairs A-U, G-C, U-A and C-G at the C-G doublet position, the relative substrate preference appears to be A = G > U > C (Fig. 6c, lanes 2 and 4–6). 3′-RACE studies of extracts from cells transfected with the mMALAT1 ENE+A at the 3′ end of a GFP reporter suggested that the A-rich tract is degraded up to the GC followed by oligouridylation[12].

We propose that the MALAT1 ENE+A interferes with RNA decay by reducing enzyme binding to the 3′ end and inhibiting exonuclease activity upon encountering the GC dinucleotide. Enzymes targeting RNA 3′ ends may act by engaging directly in 3′→5′ degradation or by synthesizing a single-stranded tail (oligouridylation or polyadenylation) to initiate exonucleolytic decay. The major- and minor-groove widths of the ENE+A are similar to those of ideal A-form double-stranded RNA (Supplementary Fig. 5 and Supplementary Table 4) and thus suitable for binding decay enzymes or other proteins. The same stabilization mechanism is likely employed by the multiple endocrine neoplasia beta (MENβ) lncRNA because its proposed 3′-triple-helical ENE+A structure also contains a C^+•G-C and C-G interruption (Supplementary Table 5)[11].

The MALAT1 ENE+A is further stabilized by strong base-stacking interactions provided by the optimal C^+•G-C and C-G configuration and the additional hydrogen bond within the apparent C^+•G-C triple (Fig. 6). We argue that the C^+•G-C triple is protonated *in vivo* because the pKa of an internal C^+•G-C triple in DNA triplexes is greater than 7 and accumulation of the βΔ1,2-MALAT1 ENE+A+mascRNA reporter is lower when the C^+•G-C is adjacent to C rather than U in the Hoogsteen strand (Fig. 6c, lanes 2 and 9–14)[28–30]. These conclusions are consistent with studies of DNA triplexes, wherein destabilization is proposed to result from (i) electrostatic repulsion of adjacent positive charges at the N3 position, and (ii) deprotonation or partial protonation because of a change in the local pKa, yielding fewer hydrogen bonds in a C•G-C versus T•A-T triple[21–25].

Implications for RNA triple-helical structures

With 10 major-groove triples, the MALAT1 ENE+A is the most elaborate, naturally-occurring RNA triple-helical structure to date (Supplementary Table 5)[20]. Its unique bipartite triple helix uses a C-G doublet to reset the helical axis, maintaining alignment between the Hoogsteen and Watson-Crick strands (Figs. 2d and 4). The “helical reset” appears necessary; several βΔ1,2-MALAT1 ENE+A+mascRNA reporter mutants designed to form more than six successive canonical base triples all exhibited decreased accumulation *in vivo* (Fig. 6c, lanes 8, 10–12). We speculate that such MALAT1 ENE+A structures are more susceptible to decay because hydrogen bonds along the Hoogsteen face are disrupted. Furthermore, re-locating or deleting the C-G doublet, the major determinant of the “helical reset”, reduces accumulation of the βΔ1,2-MALAT1 ENE+A+mascRNA reporter transcript (Fig. 6c, lanes 2, 7, 8, 10). The “helical reset” may also depend on the C^+•G-C triple as a βΔ1,2-MALAT1 ENE+A+mascRNA reporter with a C^+•G-C to U•A-U substitution exhibits accumulation of ~60% relative to WT[11].

Our studies of the MALAT1 ENE+A triplex suggest that triple helices are restricted to a finite length because extension induces a steric clash between the Hoogsteen and Watson strands. We explored whether this structural feature might be common to known RNA triplexes and found that superposition analysis likewise predicted a steric clash between the extended Hoogsteen (U12+1) and Watson (A9+1) strands in the KSHV PAN ENE core +oligo A_9 structure (Fig. 4b and Supplementary Table 3)17. In contrast, the extended Hoogsteen and Watson strands for the $PreQ_1$-II riboswitch, SAM-II riboswitch, human telomerase and *K. lactis* telomerase (representative example in Fig. 4c) structures revealed that the 2′-OH-O2P groups at the n+1 positions do not clash; instead, the bases undergo gross misalignment, leading to sub-optimal hydrogen bonding along the Hoogsteen face (Supplementary Table 6)[31–34]. Variations in strand incompatibility may arise from dissimilarities among these triple-helical structures. Superposition analyses of the MALAT1 ENE+A triplex I and other known triplexes yielded RMSD values greater than 1 Å except for the KSHV PAN ENE+A (Supplementary Fig. 6 and Supplementary Table 6). Yet, the distances between the 2′-OH and O2P groups exhibit a widening-to-narrowing trend for all known RNA triplexes (Supplementary Table 6). However, only the MALAT1 and KSHV PAN ENE+A triplexes (RMSD < 1 Å), which favor a “zippered” state with more than 50% of the triples poised to form a hydrogen bond between the 2′-OH and O2P groups, undergo a steric clash upon computational strand extension. Other triplexes (RMSD > 1 Å), which favor an “unzippered” or non-hydrogen-bonded state, show strand misalignment (Fig. 4 and Supplementary Tables 3 and 6). These findings suggest that stacked base triples are restricted to a finite length of three to six due to the difficulty of accommodating the third strand.

Unlike base pairs that can form double helices of unlimited length, the irregular geometrical conformation of RNA base triples (Fig. 2 and Supplementary Table 3) indicates that a “helical reset” is required to correct for the geometrical misalignment and steric clashes that accumulate in consecutively stacked base triples. Such discontinuities may explain why attempts to crystallize DNA or RNA triplexes have failed[35]. Alternatively, ENE+A structures may be limited to six or fewer triples (Supplementary Table 5) because extension does not confer additional stabilization due to poor base-stacking interactions or interference during triplex assembly. Finally, the length of U•A-U triplexes may be limited because homopolymeric sequences are mutational hotspots for DNA polymerases[36].

Although major-groove RNA triplexes were discovered almost six decades ago, telomerase and PAN were the only eukaryotic RNAs shown to contain triplexes[20,37,38]. The structure of the MALAT1 ENE+A triplex has revealed that natural RNA triple helices are restricted to a finite number of successive base triples, typically three to six depending on the sequence and structural context. To accommodate more than six triples, a “helical reset” is needed, an important consideration for identifying triplexes. Because MALAT1 has recently emerged as a therapeutic target for cancer[3,8], it has not escaped our attention that the unique triple-helical structure of the MALAT1 ENE+A may serve as an attractive drug target.

Online METHODS

Plasmids and mutagenesis

The full-length mouse MALAT1 (mMALAT1, nts 1–6982) sequence was generated by PCR using the pSV40-mMALAT1 vector[41], a kind gift from K. Prasanth (University of Illinois at Urbana-Champaign), as a template. A fragment of human MALAT1 (MALAT1, nts 6681–8708 from accession NR_002819.2) was created by PCR using human genomic DNA as a template. The mMALAT1 and MALAT1 PCR fragments were subsequently inserted into the XhoI and NotI sites of AVA2136 to generate pCMV-mMALAT1 and pCMV-MALAT1, respectively, using standard molecular biology techniques. Note, AVA2136 is a minimal vector (A. Alexandrov, Yale University) that contains a CMV promoter, a short multiple cloning site, an origin of replication and an ampicillin resistance gene. Site-directed mutagenesis was employed per the manufacturer’s (QuikChange) protocol to create mutant forms of pCMV-mMALAT1 (ΔENE has nts 6591–6672 deleted and U6612C) and pCMV-MALAT1 (ΔENE has nts 8254–8336 deleted and U8275C). For crystallization studies, the pHDV-MALAT1 ENE+A core (5′-GGAAGGTTTTTCTTTTCCTGAGGCGAAAG TCTCAGGTTTTGCTTTTTGGCCTTTCTTAAAAAAAAAAAAAGCAAAA-3′) was created using site-directed mutagenesis to modify the previously-described[11] pHDV-MALAT1 ENE+A-rich tract (nts 8263–8355) plasmid, in which the HDV ribozyme sequence is located downstream[11,42]. The pCMV-βΔ1,2, pCMV-βΔ1,2-KSHV PAN ENE 1xF, pCMV-βΔ1,2-MALAT1 ENE+A-rich tract+mascRNA (nts 8254–8424 of MALAT1) and pTRP-βΔ1,2 plasmids have a pcDNA3 backbone and were previously described[11,13,15,43]. The pTRP-βΔ1,2 plasmids were created using site-directed mutagenesis to replace the CMV promoter with a tetracycline-responsive promoter (TRP) in the pCMV-βΔ1,2 constructs. All mutant forms of pCMV-βΔ1,2-MALAT1 ENE+A-rich tract +mascRNA, pTRP-βΔ1,2-MALAT1 ENE+A-rich tract+mascRNA and pTRP-βΔ1,2-KSHV PAN ENE 1xF were created using site-directed mutagenesis.

RNA preparation and crystallization

The 76-nt MALAT1 ENE+A core RNA was prepared from the pHDV-MALAT1 ENE+A core plasmid template, which was linearized using HindIII and transcribed by T7 RNA polymerase; RNA products were gel purified and exchanged into crystallization buffer (5 mM sodium cacodylate pH 6.5, 50 mM KCl, 1 mM $MgCl_2$ and 0.1 mM EDTA) as described previously[17]. RNA (~8 mg/ml) was heated at 95 °C for 3 m, snap-cooled on ice for 10 m, and allowed to equilibrate at room temperature for at least 1 h before preparing crystal trays. Crystals grew at 20 °C using the sitting-drop vapor diffusion method; folded RNA was combined with an equal volume of the reservoir solution (50 mM sodium cacodylate pH 6.5, 18 mM $MgCl_2$, 2.5 mM spermine and 9% isopropanol). Crystals (final size ~300 μm × 300 μm × 50 μm) appeared within 2 to 14 d and were stabilized by adding increasing amounts of methyl-2,4-pentanediol (MPD) to a final concentration of 45%. Heavy-atom derivatives were soaked with 2 mM iridium (III) hexamine trichloride (a kind gift from S. Strobel, Yale University) for 2 h as increasing amounts of MPD were added for stabilization. All crystals were flash frozen in liquid nitrogen.

Data collection and processing

Native diffraction data were collected under cryocooled conditions (100 K) at the Advanced Photon Source, Argonne National Laboratory on beamline 24-ID-C; iridium diffraction data were collected under cryocooled conditions (100 K) at the National Synchrotron Light Source, Brookhaven National Laboratory on beamline X-25 at iridium’s peak (1.1053 Å), inflection (1.106 Å) and remote (1.1046 Å) wavelengths. All crystals belonged to the $P3_221$ space group and all data were processed using XDS[44]. Non-isomorphic crystals (cross-crystal R values of approximately 20%–30%) prevented the merging of data from different crystals. The ShelXC/D program located the initial heavy-atom sites by MAD[45], yielding experimental maps that were interpretable for two of the three RNA molecules (A and B) in the asymmetric unit (Supplementary Figs. 2a and 3a). Resulting anomalous difference Fourier maps confirmed the presence of iridium hexamine bound to RNA with at least three strong iridium hexamine-binding sites for each RNA molecule. Using the geometric configuration of the three strong iridium-binding sites, we identified molecule C in the weak density (Supplementary Fig. 2a). Initial electron density maps were further improved by refining heavy-atom locations using MLPHARE[46], sharpening amplitudes by $B = -60$ Å^2 (ref.[47]), and performing multidomain electron density averaging using DMMULTI in the CCP4 suite[48] followed by multicrystal averaging of the Ir and native data sets (Supplementary Fig. 3b). Using these experimental maps, a model for all three RNA molecules was built through iterative cycles using Coot[49].

The model was refined against the sharpened amplitudes ($B = -60$ Å^2) of the native data at 3.1 Å resolution using Refmac5 with translation/liberation/screw motions included in the final step[47,50]. Our final model contains three molecules with the following nucleotides: nts 1–76 for molecule A, nts 1–55 and 59–76 for molecule B and nts 1–54 and 60–76 for molecule C. Due to crystal packing, molecules A and B were more ordered (average B factor of 30.2 Å^2 for A and 47.9 Å^2 for B) than molecule C (average B factor 88.9 Å^2). Electron density for ions and water molecules was observed in the final refined maps. The identities of these ions and water molecules were not verified; therefore, they were deposited as water molecules in the final PDB file. Figures were created using PyMOL.

β-globin reporter assays and Northern blots

HEK293T cells were maintained in DMEM supplemented with 10% FBS, 2 mM L-glutamine, and 1× penicillin streptomycin at 37 °C and 5% CO_2. β-globin reporter assays were performed as previously described[11]. The pCMV-mMALAT1 and pCMV-MALAT1 constructs were co-transfected into HEK293T cells with pmaxGFP and performed as described for the β-globin reporter assays with the following modifications after RNA isolation: RNA was resolved on a 1% agarose/6.5% formaldehyde gel and 5′-[^{32}P]-labeled oligonucleotides were used to detect mMALAT1 (5′-TGCCTCCCAAGTGCTAGGAT-3′, 5′-CCATTCATTCC CCTCTGAGC-3′, 5′-CTCGTGGCTCAAGTGAGGTG-3′ and 5′-TTCTGGAAAAGCTGGGGAAA-3′[41]), MALAT1 (5′-GCATTGGAGATCAGCTTCCGCTAAGATGCTAGCTTGGCC AAGTCTGTTATGTTCACC-3′) and GFP (5′-CGTACTTCTCGATGCGGGTGTTGG-3′) on the Northern blot. Original images of Northern blots used in this study can be found in Supplementary Figure 7.

In vivo decay assays and data analysis

HeLa-Tet Off cells were maintained in DMEM supplemented with 10% tetracycline-approved FBS and 2 mM L-glutamine at 37 °C and 5% CO_2. *In vivo* decay assays were modified from the procedure described previously[15]. Briefly, a pTRP-βΔ1,2 plasmid was transfected into HeLa-Tet Off cells using Lipofectamine 2000 (Life Technologies) per the manufacturer's suggested protocol. Medium was changed after ~12 h and the transcriptional pulse was started ~24 h after transfection by washing the cells twice with 1× phosphate-buffered saline (PBS) and adding medium lacking doxycycline. At the end of the pulse period, the medium was removed, cells were washed once with 1× PBS and medium with 50 ng/ml doxycycline was added. Cells were harvested at various time points, RNA isolated using Trizol (Life Technologies) and analyzed by Northern blot. Blots were probed for β-globin mRNA using a uniformly [^{32}P]-labeled antisense RNA probe and 7SL RNA using a 5′-[^{32}P]-labeled oligonucleotide (5′ TGCTCCGTTTCCGACCTGGG CCGGTTCACCCCTCCTT-3′). Blots (original images in Supplementary Figure 7) were exposed to a phosphorimager screen and scanned using a Storm 860 (GE Healthcare). Northern signals were quantitated using ImageQuant Software (Molecular Dynamics); the β-globin signal was divided by the 7SL signal and values were further normalized to time zero, which was set at 100%. Plots of percent RNA remaining versus time were fit to either a single- (Equation 1) or double-exponential (Equation 2) decay using a non-linear regression program (Kalediagraph Software) and half-lives ($t_{1/2}$) were calculated by entering the appropriate rate constant (k, k_{fast} or k_{slow}) from Equations 1 or 2 into Equation 3. The RNA population undergoing rapid decay was extrapolated directly from the amplitude, A_{fast}, in Equation 2. Curves were fit to the double-exponential equation when the residuals showed a substantial improvement compared to the residuals for a single-exponential fit.

$$\mathrm{RNA\ remaining}=\mathrm{A}(\exp(-k\mathrm{t})) \quad (1)$$

$$\mathrm{RNA\ remaining}=\mathrm{A}_{\mathrm{fast}}(\exp(-k_{fast}\mathrm{t}))+\mathrm{A}_{\mathrm{slow}}(\exp(-k_{slow}\mathrm{t})) \quad (2)$$

$$\mathrm{t}_{1/2}=\ln 2/k \quad (3)$$

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for staff assistance at Advanced Photon Source beamline 24-ID and National Syncrotron Light Source beamline X-25, plasmids from K. Prasanth (University of Illinois at Urbana-Champaign, pSV40-mMALAT1) and A. Alexandrov (Yale University, AVA2136), and iridium (III) hexamine trichloride from S. Strobel (Yale University). We thank P. Moore, K. Tycowski and J. Withers for critical review of the manuscript, A. Miccinello for editorial work, and all Steitz lab members for thoughtful discussions. This work was supported by US National Institutes of Health grants GM026154 (J.A.S.) and GM022778 (T.A.S.), a Postdoctoral Fellowship (grant 122267-PF-12-077-01-RMC) from the American Cancer Society (J.A.B.) and the Steitz Center for Structural Biology, Gwangju Institute of Science and Technology, Republic of Korea (J.W.). J.A.S. and T.A.S. are investigators of the Howard Hughes Medical Institute.

References

1. Qiu MT, Hu JW, Yin R, Xu L. Long noncoding RNA: an emerging paradigm of cancer research. Tumour Biol. 2013; 34:613–20. [PubMed: 23359273]
2. Batista PJ, Chang HY. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. Cell. 2013; 152:1298–1307. [PubMed: 23498938]
3. Gutschner T, Hammerle M, Diederichs S. MALAT1 - a paradigm for long noncoding RNA function in cancer. J Mol Med (Berl). 2013
4. Schmidt LH, et al. The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol. 2011; 6:1984–92. [PubMed: 22088988]
5. Xu C, Yang M, Tian J, Wang X, Li Z. MALAT-1: a long non-coding RNA and its important 3′ end functional motif in colorectal cancer metastasis. Int J Oncol. 2011; 39:169–75. [PubMed: 21503572]
6. Feng J, et al. Expression of long non-coding ribonucleic acid metastasis-associated lung adenocarcinoma transcript-1 is correlated with progress and apoptosis of laryngeal squamous cell carcinoma. Head Neck Oncol. 2012; 4:46.
7. Ying L, et al. Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. Mol Biosyst. 2012; 8:2289–94. [PubMed: 22722759]
8. Gutschner T, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res. 2013; 73:1180–9. [PubMed: 23243023]
9. Friedel CC, Dolken L, Ruzsics Z, Koszinowski UH, Zimmer R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Res. 2009; 37:e115. [PubMed: 19561200]
10. Wilusz JE, Freier SM, Spector DL. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell. 2008; 135:919–32. [PubMed: 19041754]
11. Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA. Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENbeta noncoding RNAs. Proc Natl Acad Sci U S A. 2012; 109:19202–7. [PubMed: 23129630]

12. Wilusz JE, et al. A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. Genes Dev. 2012; 26:2392–407. [PubMed: 23073843]

13. Conrad NK, Steitz JA. A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. EMBO J. 2005; 24:1831–41. [PubMed: 15861127]

14. Tycowski KT, Shu MD, Borah S, Shi M, Steitz JA. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. Cell Rep. 2012; 2:26–32. [PubMed: 22840393]

15. Conrad NK, Mili S, Marshall EL, Shu MD, Steitz JA. Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. Mol Cell. 2006; 24:943–53. [PubMed: 17189195]

16. Conrad NK, Shu MD, 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 U S A. 2007; 104:10412–7. [PubMed: 17563387]

17. 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; 330:1244–7. [PubMed: 21109672]

18. Roberts RW, Crothers DM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science. 1992; 258:1463–6. [PubMed: 1279808]

19. Holland JA, Hoffman DW. Structural features and stability of an RNA triple helix in solution. Nucleic Acids Res. 1996; 24:2841–8. [PubMed: 8759020]

20. Conrad NK. The emerging role of triple helices in RNA biology. Wiley Interdiscip Rev RNA. 2014; 5:15–29. [PubMed: 24115594]

21. Kiessling LL, Griffin LC, Dervan PB. Flanking sequence effects within the pyrimidine triple-helix motif characterized by affinity cleaving. Biochemistry. 1992; 31:2829–34. [PubMed: 1547224]

22. Jayasena SD, Johnston BH. Oligonucleotide-directed triple helix formation at adjacent oligopurine and oligopyrimidine DNA tracts by alternate strand recognition. Nucleic Acids Res. 1992; 20:5279–88. [PubMed: 1437547]

23. Volker J, Klump HH. Electrostatic effects in DNA triple helices. Biochemistry. 1994; 33:13502–8. [PubMed: 7947759]

24. Roberts RW, Crothers DM. Prediction of the stability of DNA triplexes. Proc Natl Acad Sci U S A. 1996; 93:4320–5. [PubMed: 8633063]

25. Leitner D, Weisz K. Sequence-dependent stability of intramolecular DNA triple helices. J Biomol Struct Dyn. 2000; 17:993–1000. [PubMed: 10949166]

26. Henriksson N, Nilsson P, Wu M, Song H, Virtanen A. Recognition of adenosine residues by the active site of poly(A)-specific ribonuclease. J Biol Chem. 2010; 285:163–70. [PubMed: 19901024]

27. Viswanathan P, Chen J, Chiang YC, Denis CL. Identification of multiple RNA features that influence CCR4 deadenylation activity. J Biol Chem. 2003; 278:14949–55. [PubMed: 12590136]

28. Plum GE, Breslauer KJ. Thermodynamics of an intramolecular DNA triple helix: a calorimetric and spectroscopic study of the pH and salt dependence of thermally induced structural transitions. J Mol Biol. 1995; 248:679–95. [PubMed: 7752233]

29. Asensio JL, Lane AN, Dhesi J, Bergqvist S, Brown T. The contribution of cytosine protonation to the stability of parallel DNA triple helices. J Mol Biol. 1998; 275:811–22. [PubMed: 9480771]

30. Leitner D, Schroder W, Weisz K. Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. Biochemistry. 2000; 39:5886–92. [PubMed: 10801340]

31. Cash DD, et al. Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. Proc Natl Acad Sci U S A. 2013; 110:10970–5. [PubMed: 23776224]

32. Gilbert SD, Rambo RP, Van Tyne D, Batey RT. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol. 2008; 15:177–82. [PubMed: 18204466]

33. Liberman JA, Salim M, Krucinska J, Wedekind JE. Structure of a class II preQ1 riboswitch reveals ligand recognition by a new fold. Nat Chem Biol. 2013; 9:353–5. [PubMed: 23584677]

34. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. Mol Cell. 2005; 17:671–82. [PubMed: 15749017]

35. Rhee S, Han Z, Liu K, Miles HT, Davies DR. Structure of a triple helical DNA with a triplex-duplex junction. Biochemistry. 1999; 38:16810–5. [PubMed: 10606513]

36. Paoloni-Giacobino A, Rossier C, Papasavvas MP, Antonarakis SE. Frequency of replication/transcription errors in (A)/(T) runs of human genes. Hum Genet. 2001; 109:40–7. [PubMed: 11479734]

37. Felsenfeld G, Davies DR, Rich A. Formation of a 3-stranded polynucleotide molecule. Journal of the American Chemical Society. 1957; 79:2023–2024.

38. Felsenfeld G, Rich A. Studies on the formation of two- and three-stranded polyribonucleotides. Biochim Biophys Acta. 1957; 26:457–68. [PubMed: 13499402]

39. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. RNA. 2001; 7:499–512. [PubMed: 11345429]

40. 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 U S A. 2001; 98:4899–903. [PubMed: 11296253]

41. Tripathi V, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010; 39:925–38. [PubMed: 20797886]

42. Walker SC, Avis JM, Conn GL. General plasmids for producing RNA in vitro transcripts with homogeneous ends. Nucleic Acids Res. 2003; 31:e82. [PubMed: 12888534]

43. Lykke-Andersen J, Shu MD, Steitz JA. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell. 2000; 103:1121–31. [PubMed: 11163187]

44. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–32. [PubMed: 20124692]

45. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr. 2010; 66:479–85. [PubMed: 20383001]

46. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011; 67:235–42. [PubMed: 21460441]

47. DeLaBarre B, Brunger AT. Considerations for the refinement of low-resolution crystal structures. Acta Crystallogr D Biol Crystallogr. 2006; 62:923–32. [PubMed: 16855310]

48. Cowtan K. DM: an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM newsletter on protein crystallography. 1994; 31:34–38.

49. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–32. [PubMed: 15572765]

50. Vagin AA, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallographica Section D: Biological Crystallography. 2004; 60:2184–2195.

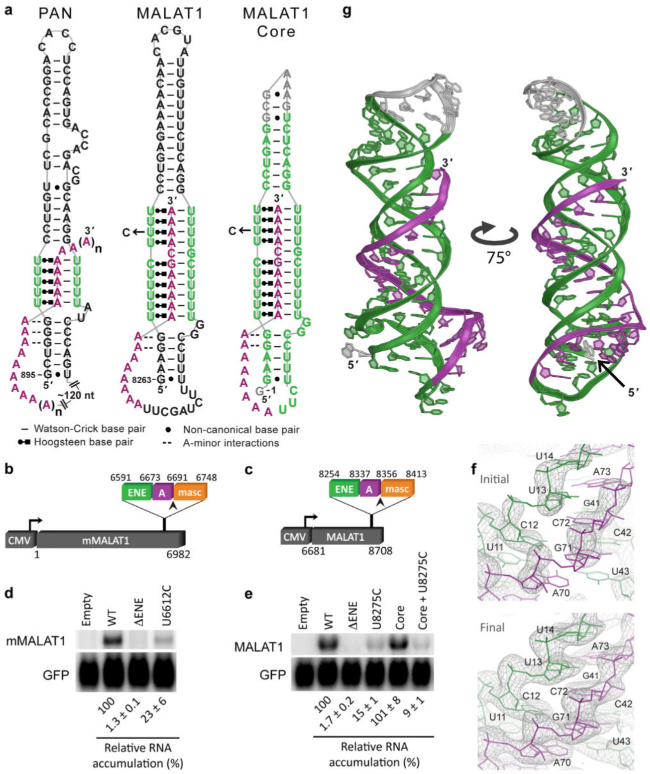

Figure 1. Overview of ENE+A structures and their importance for RNA accumulation

(a) Schematic diagrams of ENE+A structures from KSHV PAN RNA and human MALAT1 alongside the MALAT1 ENE+A core, which was used for structure determination (panel g). The U-rich internal loops are green and the poly(A) tail or A-rich tracts are purple. In the MALAT1 ENE+A core, the non-native sequence is gray. U→C denotes the U8275C mutation in the MALAT1 ENE. The U6612C mutant in the mMALAT ENE is analogous to U8275C in the human MALAT1 ENE as shown in Supplementary Tables 1 and 2. Hydrogen-bonding interactions are indicated as a dash (−) for Watson-Crick base pairs, a dot (•) for non-canonical base pairs, Leontis-Westhof notation[39] for Hoogsteen base pairs, and double dashes (--) for A-minor interactions with the corresponding G-C base pair. (b) and (c) Schematic diagrams of the constructs containing either (b) full-length mouse MALAT1 (mMALAT1) or (c) a ~2-kb fragment of the 3′ end from human MALAT1. The ENE (green), A-rich tract (purple) and mascRNA (orange, masc) sequences are expanded in the insets. Expression is driven by a cytomegalovirus (CMV) promoter while 3′-end processing occurs via RNase P cleavage (arrowhead). (d) and (e) Northern blots were probed for GFP and either (d) mMALAT1 or (e) human MALAT1 RNAs. Relative RNA accumulation was quantitated by normalizing the MALAT1 signal to the GFP signal, which served as a loading and transfection control. The WT plasmid level was set at 100%. Values are the average of three biological replicates ± standard deviation. Uncropped blot images are in Supplementary Figure 7. (f) Initial experimental electron density for a region of the triple helix in molecule A at 3.4 Å resolution contoured at 1.5 (upper panel) and final $2F_o$-F_c electron density for the same region at 3.1 Å resolution contoured at 1.5 σ (lower panel). (g) The crystal structure of the MALAT1 ENE+A core RNA is depicted with the ENE in green, A-rich tract in purple and non-native sequence in gray. Two cartoon representations (75° rotation) are shown.

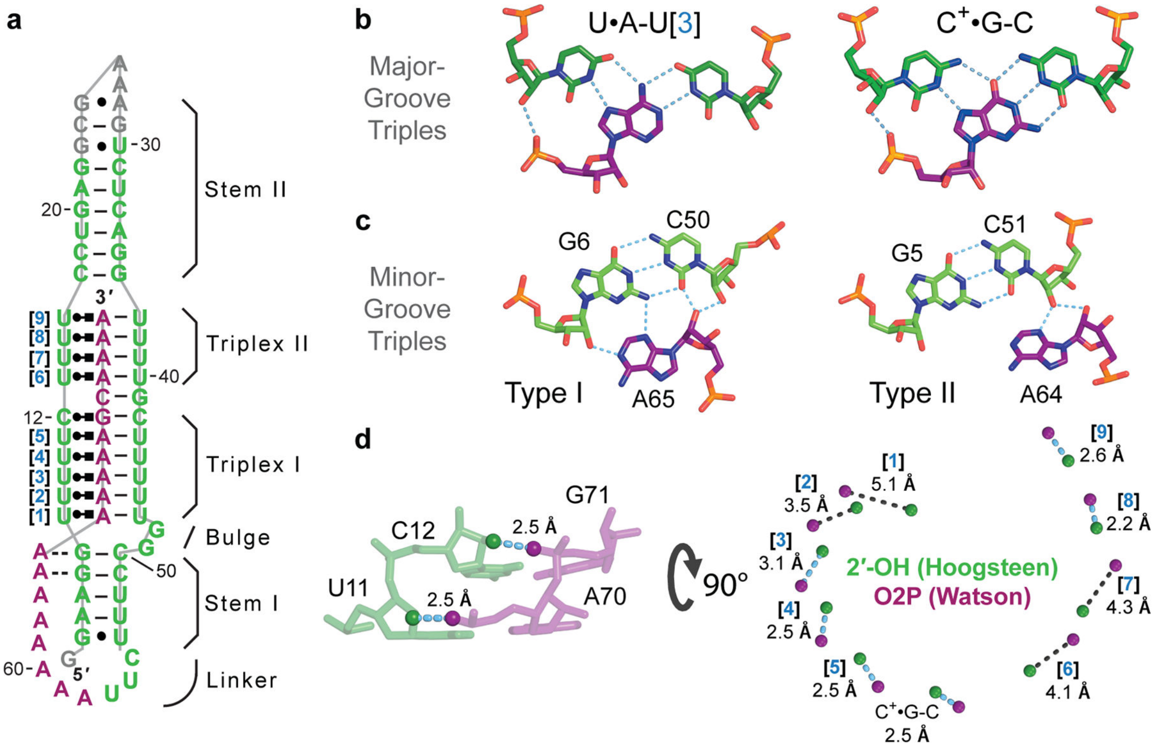

Figure 2. Hydrogen-bonding interactions in the triple helix of the MALAT1 ENE+A core RNA

(a) Schematic diagram of the MALAT1 ENE+A core structure using the notation for hydrogen bonds defined in Figure 1a. The major structural regions are labeled on the right in black; the U•A-U triples are numbered on the left in blue. (b) Potential hydrogen-bonding interactions (≤ 3.2 Å) for two different major-groove base triples, U•A-U[3] and C^+•G-C, are shown with blue dashed lines. (c) Stick representation of A-minor interactions with hydrogen bonds represented by blue dashed lines. The A-minor interactions mediated by A65 and A64 are of type I and type II, respectively[40]. (d) Potential hydrogen bonds (blue dashed line) between the 2′-OH (green spheres) of the Hoogsteen strands and O2P (purple spheres) of the Watson strands (A-rich tract) are shown in a close-up view for U11•A70 and C12•G71 (left panel) and in a view down the helical axis for all triples numbered in (a) (right panel). Distances too long for hydrogen bonding are indicated by black dashed lines.

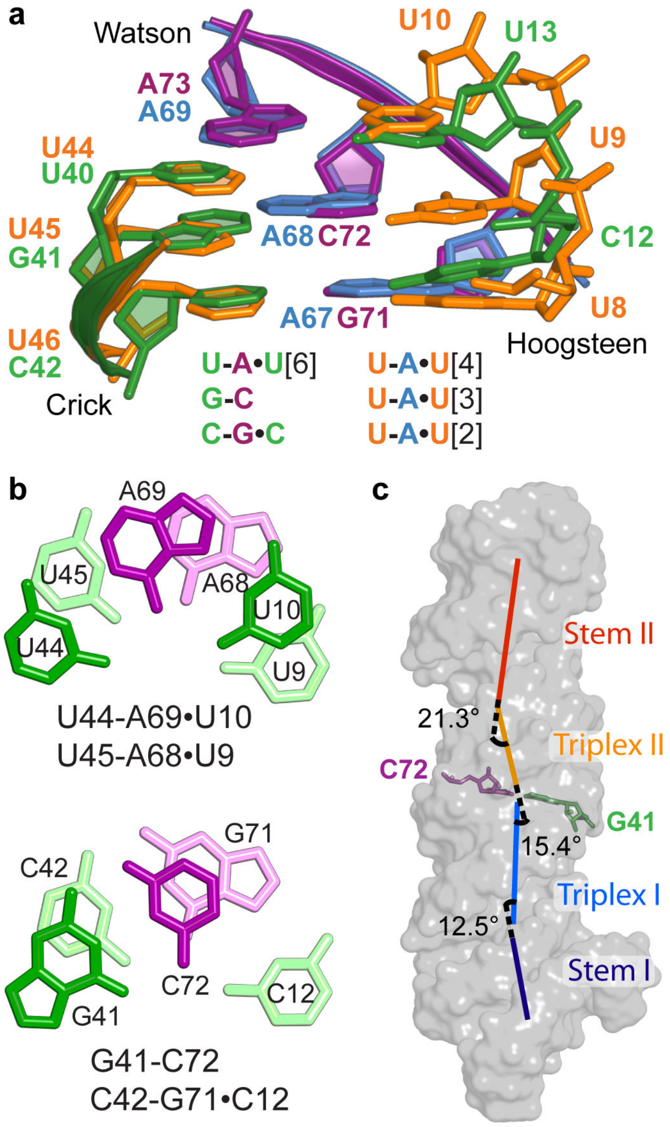

Figure 3. Structural features of the C•G-C/C-G in the MALAT1 ENE+A core

(a) Overlay of C•G-C, C-G and U•A-U[6] (ENE is green and A-rich tract is purple) with U•A-U[2], U•A-U[3] and U•A-U[4] (ENE is yellow and A-rich tract is blue) from the MALAT1 ENE+A core structure. Individual nucleotides are labeled. The similar Watson and Crick strands are displayed as cartoons while the dissimilar Hoogsteen strands are shown as sticks. (b) Base-stacking interactions are shown for U9•A68-U45 and U10•A69-U44 and C12•G71-C42 and C72-G41. The light and dark colors represent bases in the background and foreground, respectively. (c) Schematic of the local helical axes for stem I (dark blue), triplex I (light blue), triplex II (yellow) and stem II (red) overlaid on the MALAT1 ENE+A core structure (gray surface). The C72-G41 doublet is shown in purple and green sticks, respectively. Angles between the helical axes are shown.

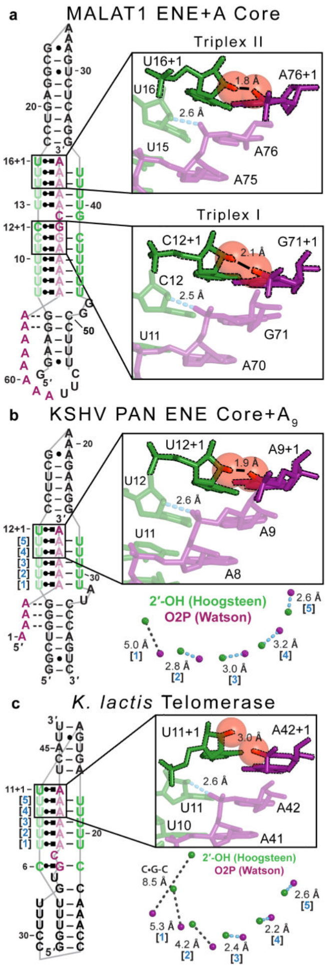

Figure 4. Destabilizing structural features predicted between extended Hoogsteen and Watson strands of RNA triplexes

(a) Computational extension of the Hoogsteen (green sticks) and Watson (purple sticks) strands in both triplex I (lower) and triplex II (upper) of the MALAT1 ENE+A core structure generates a steric clash between the 2′-OH of the Hoogsteen strand and O2P of the Watson strand at the n+1 position (black dashed outline) for C•G-C and U•A-U[9]. Coordinates for the extended Hoogsteen and Watson strands were generated by superimposing the respective strand offset by one nucleotide (*e.g.* residue “n-2” onto residue “n-1” so that residue “n” advances to “n+1”). RMSD values for the superimposed strands were 0.52 and 0.63 Å for extension of the Hoogsteen strands (light green) in triplex I and II, respectively, and 0.45 and 0.91 Å for extension of the Watson strands (light purple) in triplex I and II, respectively. (b) and (c) Extension of the Hoogsteen and Watson strands (upper right) for the (b) KSHV PAN ENE core and oligo A_9 (ref. [17]) and (c) *K. lactis* telomerase [31] structures were performed as described in panel (a). For KSHV PAN ENE core+A_9, RMSD values for the superimposed Hoogsteen and Watson strands were 0.33 and 0.59 Å, respectively. For *K. lactis* telomerase, RMSD values were 0.98 and 0.85 Å for the superimposed Hoogsteen and Watson strands, respectively. The lower right diagrams in (b) and (c) display the distances between the 2′-OH (green sphere) of the Hoogsteen strand and O2P (purple sphere) of the Watson strand in a view down the axis; dashed lines are as defined in Figure 2d. Blue numbers correspond to the numbered triples in the schematic (left). The advanced Hoogsteen base pair (“n+1”) is included in the schematic diagrams.

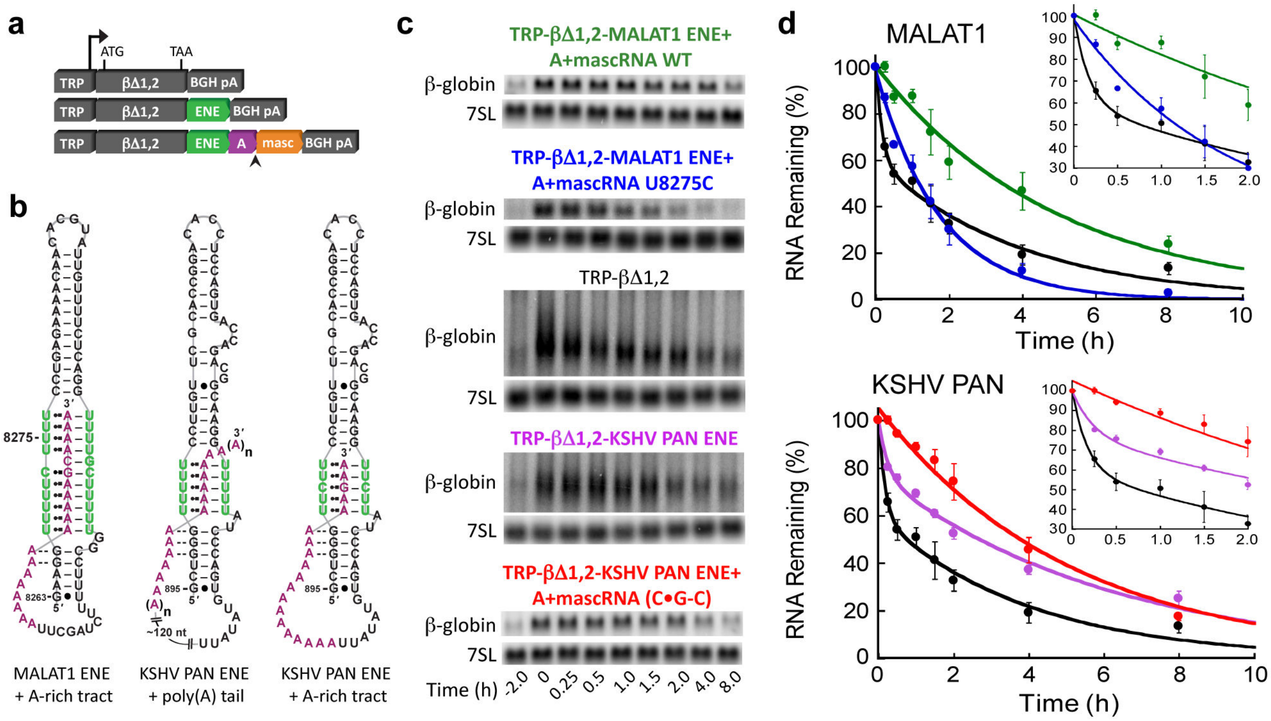

Figure 5. The MALAT1 ENE+A exhibits a single phase of RNA decay *in vivo*

(a) Schematic diagrams are shown for the βΔ1,2 constructs containing a tetracycline-responsive promoter (TRP) and bovine growth hormone polyA signal (BGH pA). Transcripts containing the mascRNA segment (orange, masc) undergo 3′-end processing via RNase P cleavage whereas the other transcripts undergo cleavage and polyadenylation. The resulting ENE (green) structures engage a genomically-encoded A-rich tract or the poly(A) tail (purple), respectively. (b) Schematic diagrams of the KSHV PAN ENE and MALAT1 ENE structures tested in the decay assays. Hydrogen-bonding interactions are as defined in Figure 1a. (c) Representative Northern blots probed for β-globin and 7SL RNAs show the amount of mRNA remaining at different times after the pulse for each construct: TRP-βΔ1,2-MALAT1 ENE+A+mascRNA WT (green), TRP-βΔ1,2-MALAT1 ENE+A +mascRNA U8275C (blue), TRP-βΔ1,2 (black), TRP-βΔ1,2-KSHV PAN ENE (purple) and TRP-βΔ1,2-KSHV PAN ENE+A+mascRNA with C•G-C (red). Uncropped blot images are in Supplementary Figure 7. (d) Northern blot data were quantitated by normalizing the β-globin signal to the 7SL signal, which serves as a loading control. The end of the transcriptional pulse is labeled as time zero and set at 100%. Values are the average of three biological replicates ± standard deviation. Curve colors correspond to the constructs in (c). The inset of each plot shows the first 2 h to emphasize biphasic nature of the curves for βΔ1,2 (black) and βΔ1,2-KSHV PAN ENE (purple) transcripts.

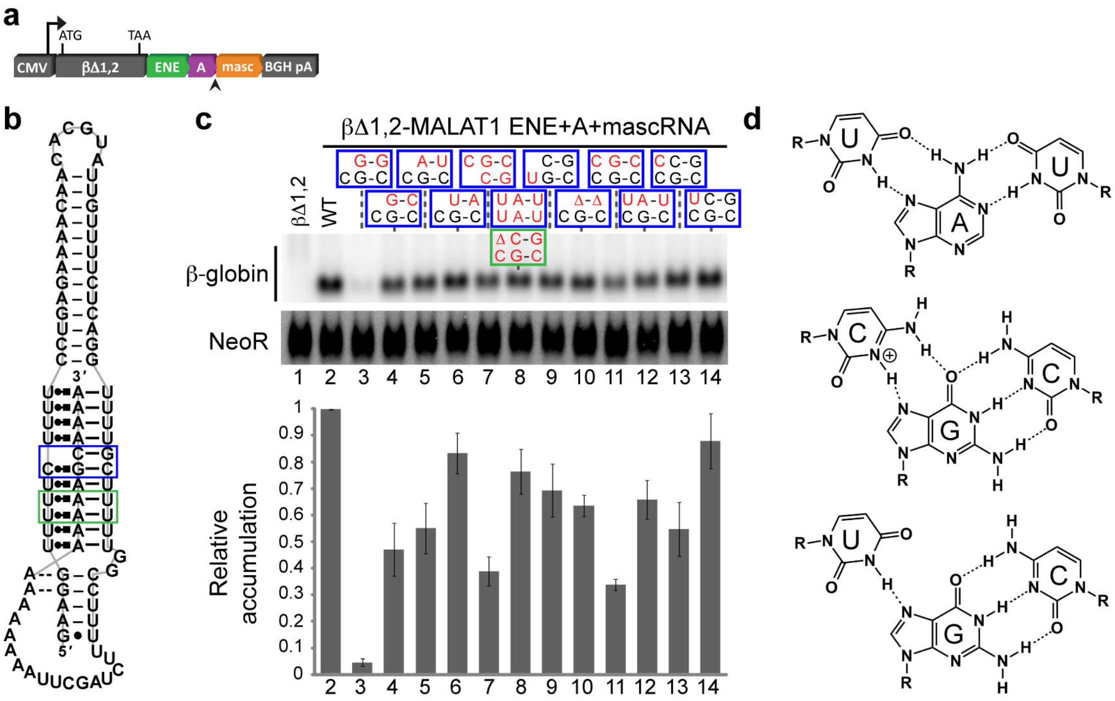

Figure 6. Accumulation levels of βΔ1,2-MALAT1 ENE+A+mascRNA containing mutations in the C•G-C and C-G nucleotides

(a) A schematic of the βΔ1,2-MALAT1 ENE+A+mascRNA construct containing a CMV promoter and BGH pA. The transcript containing the MALAT1 ENE (green), A-rich tract (purple) and mascRNA (orange, masc) undergoes 3′-end processing via RNase P cleavage (arrowhead). (b) Schematic of the MALAT1 ENE+A structure with interactions indicated as in Figure 1a. The blue and green boxes highlight mutation sites. (c) Northern blots (upper) were probed for β-globin and Neomycin resistance (NeoR) mRNAs. Black font denotes WT sequence, mutated nucleotides are red and Δ represents a nucleotide deletion. Results were quantitated (lower) by normalizing the β-globin signal to the NeoR signal, which served as a loading and transfection control. The WT βΔ1,2-MALAT1 ENE+A+mascRNA reporter level was set at an arbitrary value of 1. Relative accumulation is the average of five biological replicates; error bars represent standard deviation. Uncropped blot images are in Supplementary Figure 7. (d) Chemical structures showing the hydrogen-bonding interactions of three different bases triples: U•A-U, C^{+}•G-C, and U•G-C.

Table 1

Data collection, phasing and refinement statistics for MAD (Ir) structures

	Native		**Iridium Derivative**	
Data collection				
Space group	$P3_221$		$P3_221$	
Cell dimensions				
a, b, c (Å)	162.8, 162.8, 65.9		164.9, 164.9, 64.3	
α, β, γ (°)	90.0, 90.0, 120.0		90.0, 90.0, 120.0	
		Peak	*Inflection*	*Remote*
Wavelength		1.1053	1.1060	1.1046
Resolution (Å)	50–3.1 (3.18–3.10) [*]	50–3.4 (3.49–3.40)	50–3.4 (3.49–3.40)	50–3.4 (3.49–3.40)
R_{merge}	9.6 (239.7)	8.5 (368)	6.6 (192)	9.3 (390)
$I/\sigma I$	17.2 (1.28)	15.1 (0.73)	18.6 (1.25)	14.3 (0.67)
Completeness (%)	99.7 (99.8)	100 (100)	99.9 (100)	100 (100)
Redundancy	11.2 (11.3)	8.7 (8.9)	8.7 (8.2)	8.6 (8.2)
Refinement				
Resolution (Å)	50–3.1 ($I/\sigma I$ =1.28)			
No. reflections	17565	26893	26879	27106
$R_{\text{work}}/R_{\text{free}}$	22.0/25.5			
No. atoms				
Protein	7072			
Ligand/ion	0			
Water	35			
B factors				
Protein	54.8			
Ligand/ion	N/A			
Water	31.7			
r.m.s deviations				
Bond lengths (Å)	0.0097			
Bond angles (°)	1.03			

[*] Data sets were collected from one native crystal and one iridium-soaked crystal.

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

Table 2

Calculated half-lives and percent of transcripts rapidly degraded

Construct	Fast (%)	$t_{1/2\ (fast)}$ (m)	$t_{1/2\ (slow)}$ (h)	$t_{1/2\ (single)}$ (h)
Tet-βΔ1,2	39 ± 8	6.5 ± 2.7	2.8 ± 0.9	--
Tet-βΔ1,2-KSHV PAN ENE 1xF	23 ± 2	12 ± 6	4.4 ± 0.7	--
Tet-βΔ1,2-KSHV PAN ENE+A+mascRNA (C•G-C)	--	--	--	3.5 ± 0.2
Tet-βΔ1,2-MALAT1 ENE+A+mascRNA WT	--	--	--	3.4 ± 0.6
Tet-βΔ1,2-MALAT1 ENE+A+mascRNA U8275C	--	--	--	1.3 ± 0.1

Values are an average of three biological replicates ± standard deviation.