Published in final edited form as:
Science. 2021 February 05; 371(6529): . doi:10.1126/science.abe6523.

RNA stabilization by a poly(A) tail 3′-end binding pocket and other modes of poly(A)-RNA interaction

Seyed-Fakhreddin Torabi[1,2], **Anand T. Vaidya**[1,2,3], **Kazimierz T. Tycowski**[1,2], **Suzanne J. DeGregorio**[1,2], **Jimin Wang**[2], **Mei-Di Shu**[1,2], **Thomas A. Steitz**[1,2,‡], **Joan A. Steitz**[1,2,*]

[1]Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06536, USA.

[2]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT 06536, USA.

[3]TIFR Centre for Interdisciplinary Sciences, Tata Institute of Fundamental Research, Hyderabad 500046, India.

Abstract

Poly(A) tail addition to the 3′ end of a wide range of RNAs is a highly conserved modification that plays a central role in cellular RNA function. Elements for nuclear expression (ENEs) are cis-acting RNA elements that stabilize poly(A) tails by sequestering them in RNA triplex structures. A 2.89-Å resolution crystal structure of a double ENE from a rice hAT transposon mRNA complexed with poly(A)$_{28}$ reveals multiple modes of interaction with poly(A), including major-groove triple helices, extended minor-groove interactions with RNA double helices, a quintuple-base motif that transitions poly(A) from minor-groove associations to major-groove triple helices, and a poly(A) 3′-end binding pocket. Our findings both expand the repertoire of motifs involved in long-range RNA interactions and provide insights into how polyadenylation can protect an RNA's extreme 3′ end.

One Sentence Summary:

The structure of an RNA element that protects the poly(A) tail reveals multiple modes of interaction between poly(A) and double-stranded RNA

*Correspondence to: joan.steitz@yale.edu.
‡Deceased October 9, 2018.

Author contributions: S.-F.T. designed and performed biochemical and structural experiments, S.-F.T. and A.V. designed and carried out crystallization, S.-F.T., A.V. and J.W. analyzed data, carried out model building and structure refinement, K.T. performed bioinformatics analysis, S.-F.T. and S.J.D. carried out cloning experiments and RNA preparation, S.-F.T. and M.-D. S. performed β-globin assays, and J.A.S designed experiments. J.A.S. and T.A.S. oversaw research. S.-F.T., K.T., J.A.S. drafted and finalized the paper and all authors discussed the results and provided input on the manuscript.

Supplementary Materials:
Materials and Methods
Supplementary Text
Figures S1-S25
Tables S1 50
References (37–70)

Competing interests:
The authors declare no competing interests.

Cis-acting RNA elements play crucial roles in stabilizing polyadenylated RNAs, which include both mRNAs and long noncoding RNAs (lncRNA) (1–5). One cis-acting RNA element conferring RNA stabilization is the element for nuclear expression (ENE), which contains a U-rich internal loop (URIL) flanked by short double helices. ENEs reside mostly in 3′ untranslated regions (UTRs) of mRNAs or in the 3′-proximal regions of lncRNAs (6–9). The ENE was first identified in polyadenylated nuclear (PAN) lncRNA that is expressed by oncogenic Kaposi's sarcoma-associated herpesvirus (KSHV) (6). The PAN ENE acts by sequestration of the poly(A) tail via formation of a triplex composed of five U-A•U triples that protect the RNA from rapid deadenylation-dependent decay (10–12). This leads to the accumulation of PAN RNA to astonishingly high levels (up to ~500,000 copies per nucleus) in KSHV-infected cells (11, 13).

Discovery of the PAN ENE led to the bioinformatic prediction of numerous ENEs in the genomes of other viruses, as well as in fungi and plants (9, 14). Moreover, ENEs were identified in two abundant vertebrate lncRNAs, *i.e.* human metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) and multiple endocrine neoplasia β (MENβ) (7, 8). Abnormal upregulation and accumulation of MALAT1 reported in various human cancers require the ENE at the 3' end (15–17).

High-resolution crystal structures of ENE triplexes from KSHV PAN (11) and human MALAT1 RNAs (17) revealed that they both contain stacked major-groove triples involving the poly(A) tail (PAN RNA) or 3'-A-rich tract (MALAT1) (fig. S1). In each case, the triplex is fortified by adjacent A-minor interactions (18) with double-stranded RNA. The PAN ENE sequesters only five nucleotides of the poly(A) tail within a triplex, leaving its 3′ end exposed downstream of the triplex (11, 17). In contrast, MALAT1 terminates with an A-rich tract, generated by RNase P processing, and harbors an ENE URIL interrupted by C and G nucleotides, which interact with complementary nucleotides in the A-rich tract, locking its register to form a blunt-ended triplex (fig. S1B) (17). Thereby, the MALAT1 triplex inhibits the rapid nuclear deadenylation-dependent decay pathway more effectively than PAN RNA, which slows only the initial phase of decay (10, 12, 17).

Recently, double ENEs (dENEs), containing two URILs separated by a predicted double-helical region (Fig. 1A and fig. S2), were found in transposable element (TE) mRNAs, mostly in plants and fungi (9). Interestingly, unlike PAN and MALAT1 ENEs, all dENEs possess three highly conserved adenosines (the adenosine triad) (cyan in fig. S2). In the predicted secondary structure of dENEs, two of the adenosines lie in the lower URIL, while the third is bulged from the lower stem. In addition, dENEs exhibit a striking pyrimidine/purine (Y/R) bias in the composition of the URIL-flanking stems that are not expected to interact with poly(A), based on the crystal structures of the PAN and MALAT1 triplexes (9, 19) (figs. S1 and S2).

To gain insight into the molecular architecture and function of dENEs, we determined the crystal structure of the dENE from rice TWIFB1, a hAT DNA transposon, complexed with $poly(A)_{28}$. The TWIFB1 dENE is located 35 nt upstream of the poly(A) tail in the 3′ UTR of the transposase mRNA. The 2.89-Å resolution structure reveals several different modes of interaction between the dENE and poly(A), including: 1) the expected major-groove

triplexes, 2) a previously unrecognized class of minor-groove interactions between poly(A) and double-helices, 3) quintuple-base motifs that transition poly(A) from minor-groove interactions to major-groove triplexes, and 4) a novel poly(A) 3′-end binding pocket. We demonstrated the physiological significance of all these motifs for RNA stabilization and uncovered their conservation in many other RNAs.

Results

Crystallization and structure determination of a dENE+poly(A)$_{28}$ complex

Electrophoretic mobility shift assays (EMSAs) demonstrated that the dENE binds poly(A) in trans and that binding affinity increases with poly(A) length (supplementary text and Fig. 1C and fig. S3A). Surprisingly, poly(A)-binding was lost when the URIL-flanking Y/R biased stems, not predicted to engage in poly(A) interactions, were strand-swapped (fig. S3B). This suggested that interactions of poly(A) with the dENE possibly extend into the middle and upper stems (Fig. 1A). After testing over 40 different variants, a modified dENE construct was eventually obtained (Xtal, Fig. 1B), which contains the dENE core (green in Fig. 1A and B), binds poly (A), and yields crystals that diffract well (supplementary text, Fig. 1A-C and fig. S4). As judged by the cell-based intronless β-globin (βΔ1,2) reporter assay, the Xtal dENE is active (supplementary text and Fig. 1D). The X-ray structure of the 86-nt Xtal dENE bound to a 28-mer poly(A) was determined at 2.89-Å resolution with R_{work} and R_{free} of 0.18 and 0.22, respectively (fig. S5 and table S1).

Overall architecture of the dENE+poly(A)$_{28}$ complex

Similar to ENEs from PAN RNA and MALAT1 (11, 17), the dENE forms major-groove RNA triplexes with poly(A)$_{28}$ (fig. S6). However, the structure reveals several unanticipated features. One is that the 5′ to 3′ direction of poly(A) interacting with the dENE is opposite to that predicted based on structural homology with the PAN and MALAT1 ENEs (9, 19): poly(A) instead interacts with the upper dENE domain and then the lower dENE domain in the 5′ to 3′ direction (Fig. 1E and fig. S7). Between the two triple-stranded features, electron density is missing for five adenylates from the poly(A)$_{28}$ and they do not appear in the model (Fig. 1E). Presumably, this region is conformationally disordered as the intact complex can be recovered from the crystals (fig. S8).

A second unexpected feature is that multiple consecutive adenosines associate with the minor groove of the upper stem to form a previously-unrecognized category of A-minor motif (Fig. 2A-C and figs. S9 and S10). In contrast to the few classical A-minor interactions observed adjacent to the PAN (11) and MALAT1 (17) triplexes, which form by insertion of the minor-groove edge of adenines into the minor groove of an RNA stem (18), these minor-groove interactions involve the Watson-Crick edge or both the Watson-Crick and Hoogsteen edges of adenines. Therefore, we call them WC/H A-minor interactions. While the N1 and N6 atoms of poly(A) A5, A6 and A8 hydrogen bond to dENE residues A82, G6 and U80, respectively, within the minor groove of the upper stem to form base triple interactions, A7 is twisted with respect to the upper stem and associates with two adjacent base pairs through formation of a quintuple-nucleotide interaction involving its N1, N6 and N7 (Fig. 2B-C, and fig. S11A). After forming several WC/H A-minor interactions,

poly(A) makes a transition to an extended U-A•U major-groove triplex through another quintuple-base interaction, which involves two poly(A) nucleotides: A9 and A10, and three dENE nucleotides: U9 and the C79-G8 base pair (Fig. 2D-F and figs. S9 and S11B). The structure of the upper dENE domain and that of its counterpart in the lower domain (i.e. the quintuple-base transition motif and major-groove triplex) are so similar to each other that they superimpose with root mean square deviation (RMSD) of ~0.8 Å (fig. S12).

A poly(A) 3′-end binding pocket composed solely of RNA facilitates formation of a blunt-ended triple helix

The crystal structure further reveals that the lower dENE domain contains a novel structural feature for transitioning to the lower stem which we call a “pocket motif”. This motif is absent from the upper dENE domain (figs. S9 and S10) and was not seen in any previous ENE crystal structure (11, 17). The pocket resides in the border between the lower URIL and the lower stem of the dENE and is formed through the stacking of the bases of A30, A31 and A56, (cyan in Fig. 3 and fig. S13) within the major groove of the lower stem directly abutting the URIL. The base of A56 is splayed apart from its two adjacent nucleotides in the primary sequence, U55 and C57, and is stacked between A30 and A31 (Fig. 3A and fig. S14). In addition, the 2′-OH group and O4′ oxygen of A56 ribose are hydrogen bonded to N6 of A30 and N6 of A31, respectively, and N1 of A31 is hydrogen bonded to the C57-G35 pair (Fig. 3A-B, and figs. S2 and S14). To form such unprecedented architecture three nucleotides in the pocket motif, A31, C32 and A56, adopt a C2′-*endo* sugar pucker (figs. S14). Moreover, C32 adopts a *syn* conformation and flips its base out into solution, which allows positioning of G33 to form the C59-G33 base pair (Fig. 3A-B). Collectively, these rare structural features bend the RNA backbone of the pocket motif into a Z-shape (Fig. 3A). By forming a U-A•U triple, the 3′-most adenosine of poly(A) thereby becomes poised to form a hydrogen bond between its 3′-OH group and the G33 phosphate within the backbone of the pocket (Fig. 3 and fig. S14B).

The pocket motif is one feature distinguishing the TWIFB1 dENE from previous ENE structures. As shown in surface representation, in the PAN and MALAT1 triplexes the poly(A) tail or the A-rich tract follows a path that points towards an accessible space, i.e. the minor groove of the flanking stem (Fig. 4A and fig. S15). Thereby, the 3′ end of poly(A) could be easily extended. Yet, MALAT1 forms a 3′-blunt ended triplex because the register of its A-rich tract is fixed (fig. S6C) (17). Similar to the PAN and MALAT1 ENEs, the TWIFB1 upper dENE domain does not recognize the 3′-most adenosine of poly(A). Indeed, a nucleotide (A15) emerges from the major-groove triplex (Fig. 4A). In contrast, poly(A) encounters a ‘blockade’ – the pocket motif – at its 3′ end in the TWIFB1 lower dENE domain, such that the minor-groove pathway in the flanking stem is not open (Fig. 4A and fig. S15). We superimposed the major-groove triplexes and flanking 3-bp stems from the TWIFB1 upper dENE domain on those of the PAN and MALAT1 ENEs (RMSD = 1.1 and 1.2 Å, respectively) (Fig. 4B). None of these three ENEs exhibits structural constraints that should prevent emergence of a 3′ poly(A) overhang. Superposition of the upper and lower major-groove triplexes from the TWIFB1 dENE clearly shows that insertion of the adenosine-triad bases into the major groove of the 3-bp stem flanking the lower triplex results in movement of the closing base pair of the lower domain compared to that of the

upper domain (~10 Å) (Fig. 4B). The exit path of poly(A) is blocked by creation of the pocket for the poly(A) 3′ end. Therefore, the unique architecture of the pocket sequesters the poly(A) 3′-end and facilitates the formation of a 3′-blunt ended triplex.

We used quantitative EMSAs to determine the functional importance of the pocket in poly(A) binding by the dENE. First, we compared the $poly(A)_{28}$-binding affinity of the wild-type dENE and pocket-deleted (Δpocket) mutant (fig. S16). The Δpocket mutant showed an ~4-fold increase in its apparent dissociation constant (K_d) for $poly(A)_{28}$ (fig. S16). Furthermore, we examined the effect of creating a steric clash between the dENE G33 phosphate and a bulky group (e.g. a 3′-overhanging nucleotide) appended to the 3′ end of $poly(A)_{28}$. Such constructs exhibited: 1) lower poly(A) binding affinity, and 2) significant smear formation below the complex band (fig. S16-S17). Compensatory mutations in the dENE or poly(A), which - based on the structure - should abolish the introduced steric clash, enhanced poly(A) binding affinity and resolved smear formation (supplementary text and figs. S16-S17). Finally, we found that substituting the 3′-OH of poly(A) with a 2′,3′-cyclic phosphate resulted in an ~2-fold decrease in K_d (fig. S18). Thus, our biochemical data confirm our crystallographic observations that the 3′-most adenosine of poly(A) is tightly bound in the pocket with its 3′-OH group contributing to binding.

To obtain more insights into the function of the pocket in poly(A) 3′-end protection, we performed in vitro deadenylation assays (11, 12) on substrates containing either the wild-type or the Δpocket dENE extended with $poly(A)_{72}$. While the wild-type dENE protects the substrate from decay, mutational deletion of the pocket renders the RNA susceptible to degradation (Fig. 4D). To show that the observed degradation represents deadenylation, we used a double U to C mutant dENE (M26 in fig. S17) terminating with $poly(A)_{69}G(A)_{22}$, which is expected to lock the poly(A) register by forming a C-G•C^+ triple (Fig. 4C). We observed rapid initial substrate shortening and then accumulation of a product whose length is consistent with the size of the wild-type dENE-$poly(A)_{72}$ (Fig. 4D and supplementary text). Together, these results demonstrate the importance of the pocket in formation of a 3′-blunt ended triplex structure through a steric mechanism, which inhibits deadenylation in vitro.

Significance of dENE+poly(A) structural features in RNA stabilization

Cellular accumulation of the βΔ1,2 reporter transcript has previously been used to assess the in vivo functionality of ENEs as cis-acting RNA stabilization elements (6, 7, 9, 14). To evaluate the functional significance of the structural features of the dENE+$poly(A)_{28}$ complex, a series of βΔ1,2 reporters was generated by inserting either wild-type or mutant dENE variants into the 3′ UTR and assaying β-globin transcript accumulation in human HEK293T cells (Fig. 5). The data show that, overall, disruption of any of the identified structural features (Fig. 5B) reduces the RNA stabilization activity of the dENE (Fig. 5C). We tested identical mutations in features common to the two dENE domains (e.g. M1 versus M3, M2 versus M4, etc) and observed that mutations in the lower domain are better tolerated. This observation is consistent with the presence of the poly(A) 3′-end binding pocket in the lower dENE domain, which may partially compensate for mutations disrupting the major-groove triplex of the lower domain (supplementary text). Mutations in any of the

three residues involved in the quintuple-base transition motif formation (M1, M5 and M7 in the upper, and M3, M6 and M8 in the lower dENE domain) reduce the dENE stabilization activity to an extent comparable to mutations in the major-groove triplexes (M16 and M19 in the upper, and M17 and M20 in the lower dENE domain). More interestingly, mutations that disrupt only the poly(A) 3′-end binding pocket (M10 to M13), despite leaving other poly(A)-dENE interactions unperturbed, reduce the accumulation of β-globin mRNA by ~50%. Thus, the pocket is important for the stabilization activity of the dENE. In addition, our structure reveals molecular interactions that support the frequent occurrence of a non-canonical G•A pair as the closing base pair of upper dENE domains (fig. S2). The reason is that the adenine of the G74•A14 base pair provides a platform for stacking interactions with the adenosine in the last major-groove triple of the upper domain (supplementary text and fig. S19). Finally, we evaluated the effect of WC/H A-minor interactions in the upper dENE domain by mutating their receptor base pairs (fig. S20). Mutations that disrupted Y/R bias reduced accumulation of the reporter transcript by ~20 to 30%, indicating that WC/H A-minor interactions contribute to stabilization activity (supplementary text). As expected, mutations in the portion of the middle stem that is not involved in any interaction with poly(A) have minimal effects (M14: < 20%; M15: no effect) on dENE stabilization activity.

Functional significance of blunt-ended triplexes, formed through fixing the register of downstream A-rich tracts, has been studied before (7, 17). For example, it was shown that addition of 3′ overhangs reduces MALAT1 ENE stabilization activity (7). We tested stabilization activity of the TWIFB1 dENE when inserted upstream of a 54- or 72-mer poly(A) tract using the βΔ1,2 reporter assay. Indeed, the dENE exhibits robust stabilization activity resembling the activity of a blunt-ended triplex (MALAT1 ENE) rather than those that have 3′ overhang (Δpocket dENE or PAN ENE) (fig. S21 and supplementary text). We conclude that formation of a pocket-mediated blunt-ended triple helix counteracts RNA decay in vivo.

WC/H A-minor interactions and the quintuple-base transition motif are common to other RNAs

The WC/H A-minor and quintuple-base transition motifs observed in the TWIFB1 dENE+poly(A)$_{28}$ structure are not unique. Sequence alignment argues that such motifs also appear in many single domain (fig. S22) and other dENEs (fig. S2A). We searched ribosomal RNA and small RNA structures in the Protein Data Bank (PDB) and found several equivalent WC/H A-minor and quintuple-base transition motifs. For example, the S-adenosylhomocysteine (SAH) riboswitch possesses three consecutive adenosines that form a WC/H A-minor motif (20) superimposable on the A7, A8 and A9 WC/H A-minor motif of the dENE+poly(A)$_{28}$ complex with an RMSD = 0.9 Å (Fig. 6A and fig. S23A). The quintuple-base transition motif is also found in other RNAs (fig. S24), such as the guide RNA trans-activating CRISPR RNA from *Campylobacter jejuni* CRISPR-Cas9 system (Cas9 trans-activating crRNA) (21), where this motif connects a WC/H A-minor triple to a G-A•U major-groove triple (Fig. 6B and fig. S23B). Strikingly, in the human telomerase RNA (hTR) pseudoknot, an RNA major-groove triple helix composed of three stacked U-A•Us is flanked by a quintuple-base transition motif (22) that is superimposable on that of the dENE complex with an RMSD of ~1.5 Å (Fig. 6C and fig. S23C). In conclusion, the

structural motifs identified in the dENE+poly(A)$_{28}$ complex represent types of long-range interactions that are characteristic of many RNA structures.

Discussion

The high-resolution crystal structure of the dENE+poly(A)$_{28}$ complex demonstrates that poly(A) sequestration is widely employed for cellular stabilization of polyadenylated transcripts. Although RNA stabilization by clamping the poly(A) tail or 3′-terminal A-rich tract within a triplex has been described before for PAN or MALAT1 lncRNAs (11, 17), the structure of the dENE+poly(A)$_{28}$ complex reveals novel structural features that contribute to poly(A) binding and RNA stabilization. These novel features are; 1) WC/H A-minor interactions formed between poly(A) and double-stranded helices, 2) conserved quintuple-base motifs that transition poly(A) from minor-groove interactions to major-groove triplexes, and 3) a unique poly(A) 3′-end binding pocket that facilitates formation of a blunt-ended triplex structure through a steric mechanism. Although some of these features, i.e. WC/H A-minor and the quintuple-base motifs exist in RNAs whose structures were previously determined (Fig. 6 and fig. S24), they were not recognized earlier as distinct structural motifs.

Several types of adenosine-minor groove interactions have been previously described, including: the A-minor motif, which is the most frequently-occurring tertiary structure interaction in RNAs (18), the twisted A-minor (23), the inclined A-minor (24, 25), and the A-amino kissing motifs (26). In the dENE structure, the WC/H A-minor motif comprises four consecutive adenosines tracking in the minor groove by interaction with the bases and ribose moieties. This interaction mode makes it possible for several consecutive adenosines to associate with a double-helical stem by alternating contacts with the two strands (Fig. 1F), distinct from previously-described minor-groove interactions. Moreover, extended WC/H A-minor motifs apparently require Y/R biased helices as judged by their high degree of evolutionary conservation (figs. S3, S22 and S25). Moreover, WC/H A-minor interactions are followed by quintuple-base transition motifs if flanked by a major-groove triple(s) (Fig 6, B and C, and fig. S23, B and C).

New structural motifs observed in the dENE+poly(A)$_{28}$ complex are bioinformatically predicted to occur also in many single-domain ENEs identified in transcripts produced by TEs (mostly retrotransposons) and in genomes of positive-sense RNA viruses belonging to the order *Picornavirales* (fig. S22 and S25) (9). Interestingly, the RNA 3′-ends of both classes of these cell-invading agents are targeted by cellular RNA decay pathways to restrict retrotransposition or viral infection, respectively (27–29). For instance, 3′-end uridylation by terminal uridyltransferase 4/7 (TUT4/7), which usually leads to RNA degradation, was demonstrated to potently restrict the proliferation of both retrotransposons and RNA viruses (30, 31). Thus, it appears that these agents evolved ENEs to counter host defense mechanisms acting on RNA 3′-ends. The poly(A) 3′-end binding pocket is particularly well suited for this function.

The dENE+poly(A)$_{28}$ structure coupled with current and previous bioinformatics searches (9) has revealed extensive widespread interactions of poly(A) with highly structured RNAs

derived from transposons. The question arises whether such interactions occur more broadly in non-transposon derived cellular polyadenylated RNAs. Recent genome-wide probing of mRNA structure and decay led to the conclusion that double-strandedness of poly(A) tail-proximal regions enhances mRNA stability in both yeast and human cells (1, 32, 33). Is this enhancement brought about by engaging in interactions with poly(A) tails akin to those found for ENEs? Our previous bioinformatics screens for ENE-like elements (designed based on the PAN ENE structure) did not reveal canonical ENEs in either non-repetitive regions of eukaryotic genomes or mRNA and lncRNA transcriptomes except for MALAT1 and MENβ lncRNAs (7, 9, 14). The dENE+poly(A)$_{28}$ structure has now revealed the modular nature of this complex. Separate module(s) appearing in isolation or in combination with other yet unknown features may have escaped detection. The identification of novel structural motifs in this study raises new possibilities for designing more targeted bioinformatic searching strategies. Likewise, new experimental approaches will be needed to probe such interactions.

In conclusion, the dENE+poly(A)$_{28}$ structure in conjunction with other documented poly(A)-RNA interactions (1, 11, 17, 22, 32) demonstrates that poly(A) is capable, perhaps more than any other homo-ribopolymer, to engage in extensive tertiary interactions to form structured RNA modules. Therefore, it is tempting to speculate that the ancestral role of poly(A) tails was to protect RNA from 3′-end degradation through such poly(A)-RNA interactions. Only later in evolution, with the advent of poly(A)-binding proteins, was this function relegated significantly to proteins (34, 35).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank P. Moore, R. Mitton-Fry and K. Zahn for critical reading of the manuscript, A. Miccinello for editorial work and both Steitz laboratories for helpful discussion. We thank A. Ferré-D'Amaré for the generous gift of iridium(III) hexamine. Special thanks to W. Eliason, T.A. Yario and P. Raccuia for technical assistance, and J.A. Brown and R. Mitton-Fry for thoughtful discussions, and the staff of the Advanced Photon Source beamline NECAT 24-ID, where X-ray data were collected.

Funding:

Financial support for this research was provided by NIH grant P01CA16038 to J.A.S. and NIH grant P01GM022778 to T.A.S. S.-F. T. is supported by a Damon Runyon Cancer Research Foundation Postdoctoral fellowship. J.A.S. is investigator of the Howard Hughes Medical Institute.

Data and materials availability:

The atomic coordinates and the structure factors can be accessed in the Protein Data Bank (ID 7JNH).

References and Notes:

1. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K, Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156, 812–824 (2014). [PubMed: 24529382]

2. Kim M, Kogan N, Slack FJ, Cis-acting elements in its 3' UTR mediate post-transcriptional regulation of KRAS. Oncotarget 7, 11770–11784 (2016). [PubMed: 26930719]
3. Mizrahi O et al. , Virus-Induced Changes in mRNA Secondary Structure Uncover cis-Regulatory Elements that Directly Control Gene Expression. Mol Cell 72, 862–874 e865 (2018). [PubMed: 30318442]
4. Vejnar CE et al. , Genome wide analysis of 3' UTR sequence elements and proteins regulating mRNA stability during maternal-to-zygotic transition in zebrafish. Genome Res 29, 1100–1114 (2019). [PubMed: 31227602]
5. Akiyama BM, Eiler D, Kieft JS, Structured RNAs that evade or confound exonucleases: function follows form. Curr Opin Struct Biol 36, 40–47 (2016). [PubMed: 26797676]
6. Conrad NK, Steitz JA, A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. EMBO J 24, 1831–1841 (2005). [PubMed: 15861127]
7. 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 109, 19202–19207 (2012). [PubMed: 23129630]
8. Wilusz JE et al. , A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. Genes Dev 26, 2392–2407 (2012). [PubMed: 23073843]
9. Tycowski KT, Shu MD, Steitz JA, Myriad Triple-Helix-Forming Structures in the Transposable Element RNAs of Plants and Fungi. Cell Rep 15, 1266–1276 (2016). [PubMed: 27134163]
10. 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 24, 943–953 (2006). [PubMed: 17189195]
11. 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 330, 1244–1247 (2010). [PubMed: 21109672]
12. 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 104, 10412–10417 (2007). [PubMed: 17563387]
13. Borah S, Darricarrere 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 7, e1002300 (2011). [PubMed: 22022268]
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 2, 26–32 (2012). [PubMed: 22840393]
15. Gutschner T, Hammerle M, Diederichs S, MALAT1 -- a paradigm for long noncoding RNA function in cancer. J Mol Med (Berl) 91, 791–801 (2013). [PubMed: 23529762]
16. Dhamija S, Diederichs S, From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis. Int J Cancer 139, 269–280 (2016). [PubMed: 26875870]
17. Brown JA et al. , Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat Struct Mol Biol 21, 633–640 (2014). [PubMed: 24952594]
18. 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 98, 4899–4903 (2001). [PubMed: 11296253]
19. Brown JA, Unraveling the structure and biological functions of RNA triple helices. Wiley Interdiscip Rev RNA, e1598 (2020). [PubMed: 32441456]
20. Edwards AL, Reyes FE, Heroux A, Batey RT, Structural basis for recognition of S-adenosylhomocysteine by riboswitches. RNA 16, 2144–2155 (2010). [PubMed: 20864509]
21. Yamada M et al. , Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. Mol Cell 65, 1109–1121 e1103 (2017). [PubMed: 28306506]
22. Kim NK et al. , Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. J Mol Biol 384, 1249–1261 (2008). [PubMed: 18950640]
23. Gilbert SD, Rambo RP, Van Tyne D, Batey RT, Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol 15, 177–182 (2008). [PubMed: 18204466]

24. Klein DJ, Edwards TE, Ferre-D'Amare AR, Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. Nat Struct Mol Biol 16, 343–344 (2009). [PubMed: 19234468]
25. Liberman JA et al. , Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. Proc Natl Acad Sci U S A 112, E3485–3494 (2015). [PubMed: 26106162]
26. Kang M, Peterson R, Feigon J, Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. Mol Cell 33, 784–790 (2009). [PubMed: 19285444]
27. Molleston JM, Cherry S, Attacked from All Sides: RNA Decay in Antiviral Defense. Viruses **9**, (2017).**9**
28. Yeo J, Kim VN, U-tail as a guardian against invading RNAs. Nat Struct Mol Biol 25, 903–905 (2018). [PubMed: 30242214]
29. Zinder JC, Lima CD, Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. Genes Dev 31, 88–100 (2017). [PubMed: 28202538]
30. Warkocki Z et al. , Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. Cell 174, 1537–1548 e1529 (2018). [PubMed: 30122351]
31. Le Pen J et al. , Terminal uridylyltransferases target RNA viruses as part of the innate immune system. Nat Struct Mol Biol 25, 778–786 (2018). [PubMed: 30104661]
32. Moqtaderi Z, Geisberg JV, Struhl K, Secondary structures involving the poly(A) tail and other 3' sequences are major determinants of mRNA isoform stability in yeast. Microb Cell 1, 137–139 (2014). [PubMed: 25279376]
33. Wu X, Bartel DP, Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability. Cell 169, 905–917 e911 (2017). [PubMed: 28525757]
34. Schafer IB et al. , Molecular Basis for poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylase. Cell 177, 1619–1631 e1621 (2019). [PubMed: 31104843]
35. Wigington CP, Williams KR, Meers MP, Bassell GJ, Corbett AH, Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. Wiley Interdiscip Rev RNA 5, 601–622 (2014). [PubMed: 24789627]
36. Leontis NB, Westhof E, Geometric nomenclature and classification of RNA base pairs. RNA 7, 499–512 (2001). [PubMed: 11345429]
37. Avis JM, Conn GL, Walker SC, Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5' and 3' ends. Methods Mol Biol 941, 83–98 (2012). [PubMed: 23065555]
38. Golden BL, Kundrot CE, RNA crystallization. J Struct Biol 142, 98–107 (2003). [PubMed: 12718923]
39. Kabsch W, Xds. Acta Crystallogr D Biol Crystallogr 66, 125–132 (2010). [PubMed: 20124692]
40. Otwinowski Z, Minor W, Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276, 307–326 (1997).
41. Sheldrick GM, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr 66, 479–485 (2010). [PubMed: 20383001]
42. McCoy AJ et al. , Phaser crystallographic software. J Appl Crystallogr 40, 658–674 (2007). [PubMed: 19461840]
43. Winn MD et al. , Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67, 235–242 (2011). [PubMed: 21460441]
44. Emsley P, Lohkamp B, Scott WG, Cowtan K, Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501 (2010). [PubMed: 20383002]
45. Murshudov GN, Vagin AA, Dodson EJ, Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53, 240–255 (1997). [PubMed: 15299926]
46. Ford LP, Watson J, Keene JD, Wilusz J, ELAV proteins stabilize deadenylated intermediates in a novel in vitro mRNA deadenylation/degradation system. Genes Dev 13, 188–201 (1999). [PubMed: 9925643]

47. Fuchs AL, Neu A, Sprangers R, A general method for rapid and cost-efficient large-scale production of 5’ capped RNA. RNA 22, 1454–1466 (2016). [PubMed: 27368341]
48. Dignam JD, Lebovitz RM, Roeder RG, Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11, 1475–1489 (1983). [PubMed: 6828386]
49. Bock RM, Controlled partial hydrolysis of RNA. Method Enzymol 12, 218–221 (1967).
50. Pijlman GP et al. , A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. Cell Host Microbe 4, 579–591 (2008). [PubMed: 19064258]
51. Rauhut R, Klug G, mRNA degradation in bacteria. FEMS Microbiol Rev 23, 353–370 (1999). [PubMed: 10371038]
52. Akiyama BM et al. , Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. Science 354, 1148–1152 (2016). [PubMed: 27934765]
53. Steckelberg AL et al. , A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure. Proc Natl Acad Sci U S A 115, 6404–6409 (2018). [PubMed: 29866852]
54. Steckelberg AL, Vicens Q, Costantino DA, Nix JC, Kieft JS, The crystal structure of a Polerovirus exoribonuclease-resistant RNA shows how diverse sequences are integrated into a conserved fold. RNA, (2020).
55. Meyer S, Temme C, Wahle E, Messenger RNA turnover in eukaryotes: pathways and enzymes. Crit Rev Biochem Mol Biol 39, 197–216 (2004). [PubMed: 15596551]
56. Yu S, Kim VN, A tale of non-canonical tails: gene regulation by post-transcriptional RNA tailing. Nat Rev Mol Cell Biol, (2020).
57. Lingaraju M et al. , To Process or to Decay: A Mechanistic View of the Nuclear RNA Exosome. Cold Spring Harb Symp Quant Biol 84, 155–163 (2019). [PubMed: 32493762]
58. Wasmuth EV, Januszyk K, Lima CD, Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. Nature 511, 435–439 (2014). [PubMed: 25043052]
59. Zinder JC, Wasmuth EV, Lima CD, Nuclear RNA Exosome at 3.1 A Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3. Mol Cell 64, 734–745 (2016). [PubMed: 27818140]
60. Olson WK et al. , Effects of Noncanonical Base Pairing on RNA Folding: Structural Context and Spatial Arrangements of G.A Pairs. Biochemistry 58, 2474–2487 (2019). [PubMed: 31008589]
61. Proctor DJ et al. , Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. Biochemistry 43, 14004–14014 (2004). [PubMed: 15518549]
62. Hall KB, Mighty tiny. RNA 21, 630–631 (2015). [PubMed: 25780168]
63. Battle DJ, Doudna JA, Specificity of RNA-RNA helix recognition. Proc Natl Acad Sci U S A 99, 11676–11681 (2002). [PubMed: 12189204]
64. Vi SL et al. , Target specificity among canonical nuclear poly(A) polymerases in plants modulates organ growth and pathogen response. Proc Natl Acad Sci U S A 110, 13994–13999 (2013). [PubMed: 23918356]
65. Minvielle-Sebastia L, Preker PJ, Wiederkehr T, Strahm Y, Keller W, The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3’-end formation. Proc Natl Acad Sci U S A 94, 7897–7902 (1997). [PubMed: 9223284]
66. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP, Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508, 66–71 (2014). [PubMed: 24476825]
67. Chang H, Lim J, Ha M, Kim VN, TAIL-seq: genome-wide determination of poly(A) tail length and 3’ end modifications. Mol Cell 53, 1044–1052 (2014). [PubMed: 24582499]
68. Giedroc DP, Cornish PV, Frameshifting RNA pseudoknots: structure and mechanism. Virus Res 139, 193–208 (2009). [PubMed: 18621088]
69. Karplus PA, Diederichs K, Linking crystallographic model and data quality. Science 336, 1030–1033 (2012). [PubMed: 22628654]

70. Karplus PA, Diederichs K, Assessing and maximizing data quality in macromolecular crystallography. Curr Opin Struct Biol 34, 60–68 (2015). [PubMed: 26209821]

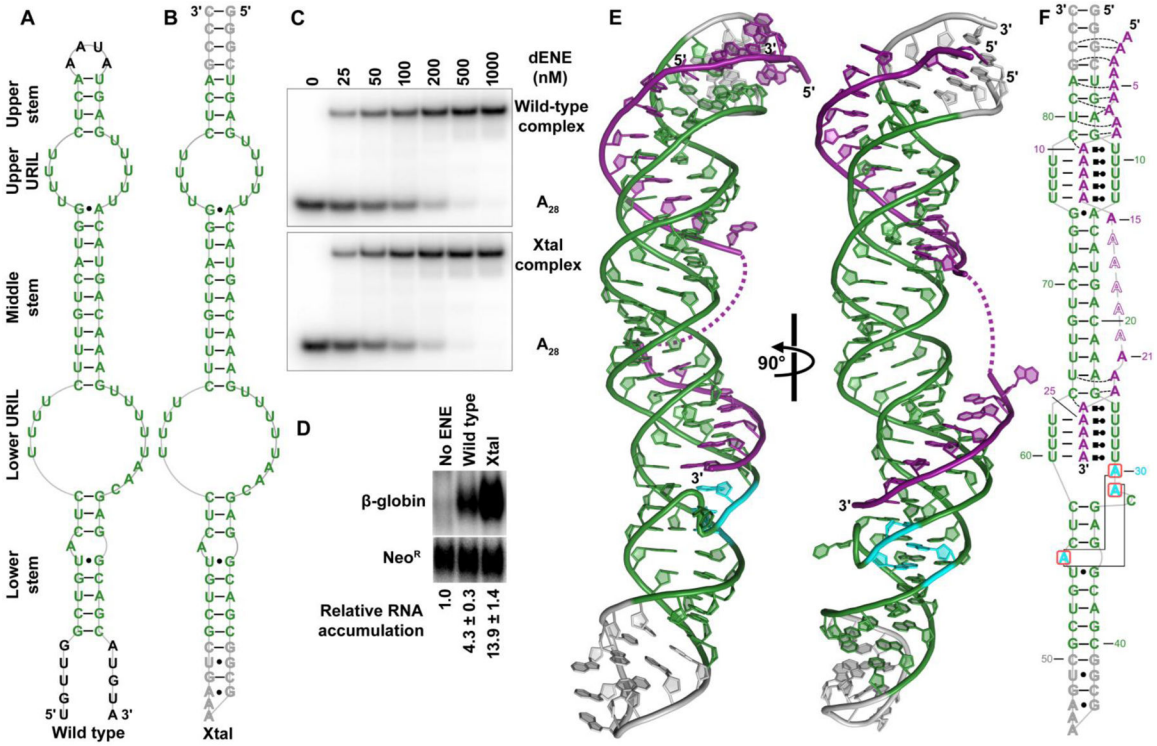

Fig. 1.

Structural overview of the TWIFB1 dENE+poly(A)$_{28}$ complex. Predicted secondary structures of the wild-type dENE (A) and of the crystallization construct (Xtal) (B), which was used to determine the X-ray structure in (E). The dENE core is shown in green. Non-native sequences in Xtal are shown in gray. (C) Interaction between ^{32}P 5′-end labeled poly(A)$_{28}$ and dENE constructs was evaluated by native gel shift assays. (D) Northern blot analysis of the βΔ1,2 reporter transcript abundance in HEK293T cells. Values are the average of at least four biological replicates ± standard deviation. (E) Cartoon representation of the crystal structure, with poly(A)$_{28}$ in purple, the ENE core in green, the three adenosines engaged in base stacking interactions (adenosine triad) in cyan, and non-native sequences in gray. (F) Schematic diagram of the complex tertiary structure with colors as in (E). Leontis-Westhof notation (36) is used for Hoogsteen base pairs; bullets for noncanonical base pairs and dashed lines for poly(A) interactions with the minor grooves of the stems. Long-range interactions observed in the pocket structure between stacked adenines (boxed in red) are indicated by solid lines. Disordered nucleotides, A$_{16}$-A$_{20}$, are indicated by dashes in (E) or outlined letters in (F).

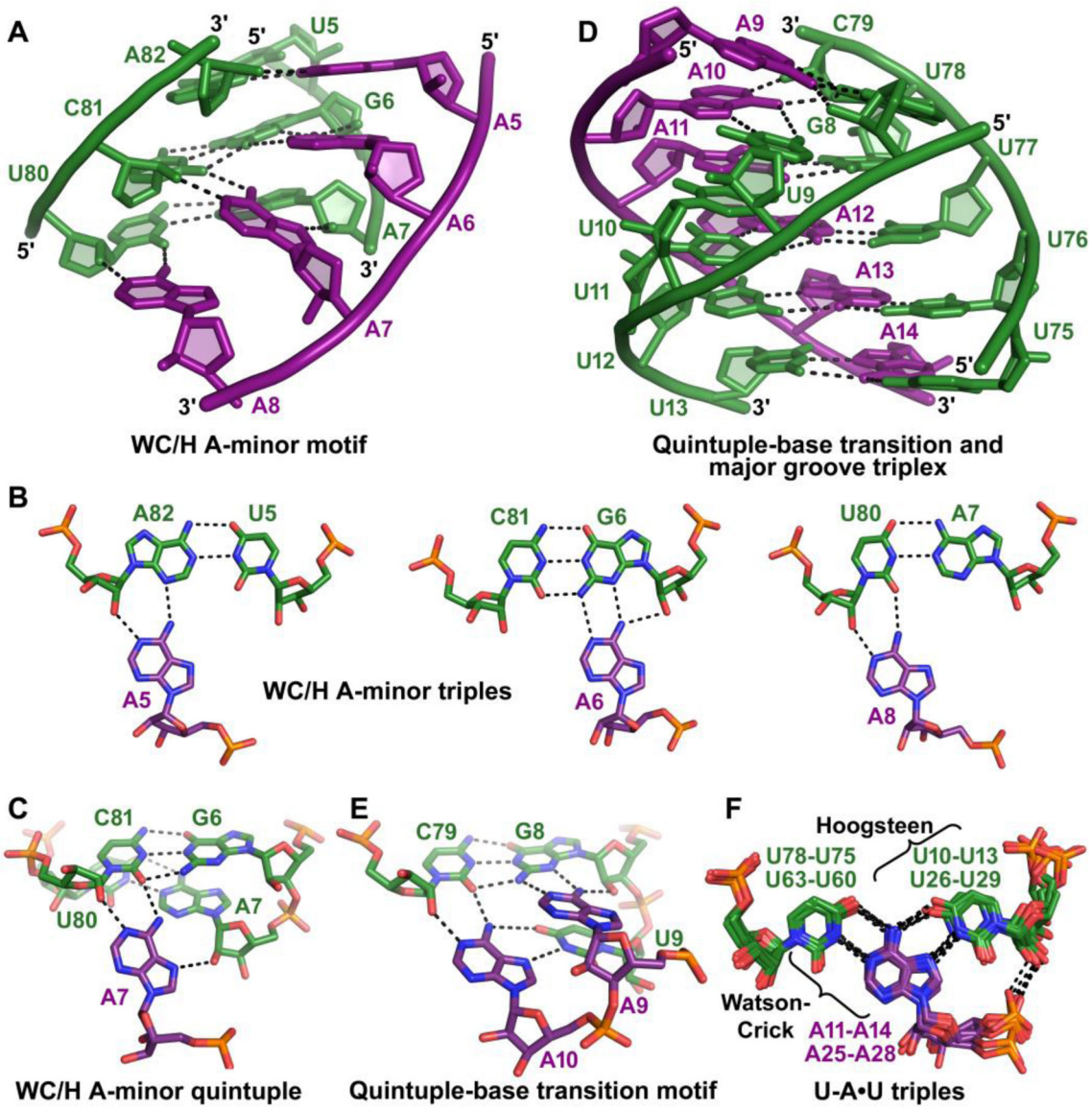

Fig. 2.

Hydrogen-bonding interactions in the upper ENE domain of the dENE+poly(A)$_{28}$ complex. Colors as in Fig. 1, with nucleotide labels also color coded. Watson-Crick base pairing and hydrogen bonds involving poly(A) are represented by black dashed lines. (A) Close-up of the WC/H A-minor motif that poly(A) forms with the upper stem. WC/H A-minor motif include three base triples (B) and one quintuple-nucleotide interaction (C), in which poly(A) adenines use their Watson-Crick and/or Hoogsteen edges to interact with the minor groove of the upper stem. (D) Close-up of the poly(A) transition from the WC/H A-minor motif to the major-groove triple helix. (E) Poly(A) makes this transition through formation of a network of hydrogen bonds involving A9 and A10 from poly(A), the C79-G8 base pair and U9. The Hoogsteen face of A10 base pairs with the Watson-Crick face of U9, while its Watson-Crick face interacts with the C79-G8 base pair minor groove. (F) Superposition of canonical U-A•U triples that form major-groove triplexes in the structure.

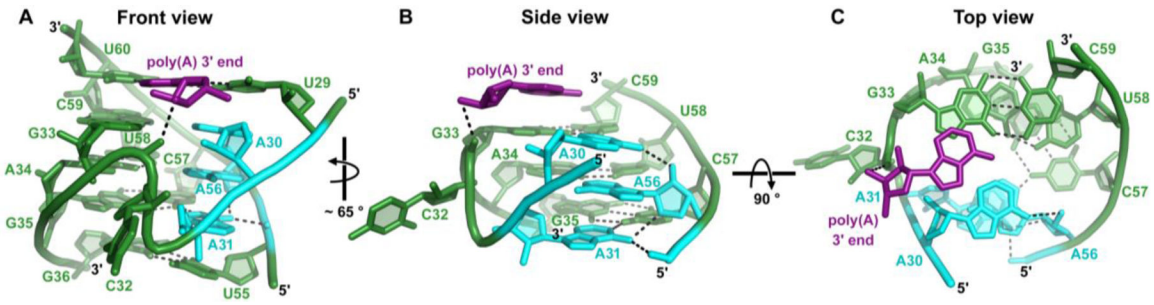

Fig. 3.

Key interactions that form the poly(A) 3′-end binding pocket. (A) Close-up overview of the poly(A) transition from the lower dENE major-groove triplex to the flanking double helix (dENE sequence in green). The poly(A) 3′-end binding pocket formed by the stacked nucleobases of the adenosine triad (cyan) lies in the major groove of a 3-base-pair long flanking double helix. The U60-A28•U29 triple and the U55•G36 base pair flanking the pocket are shown only in the front view. (B) and (C) show side and top views of the pocket, respectively. The 3′ end of poly(A) is shown in purple. The base of A56 stacks between those of A30 and A31, while forming multiple hydrogen bonds with these two dENE nucleotides. The flipped out nucleobase of C32 contributes to the flexibility of the backbone and allows formation of the Z-shaped structure of the phosphate backbone. The poly(A) 3′ end is poised above the pocket with its 3′-OH within hydrogen-bonding distance of the G33 phosphate oxygen.

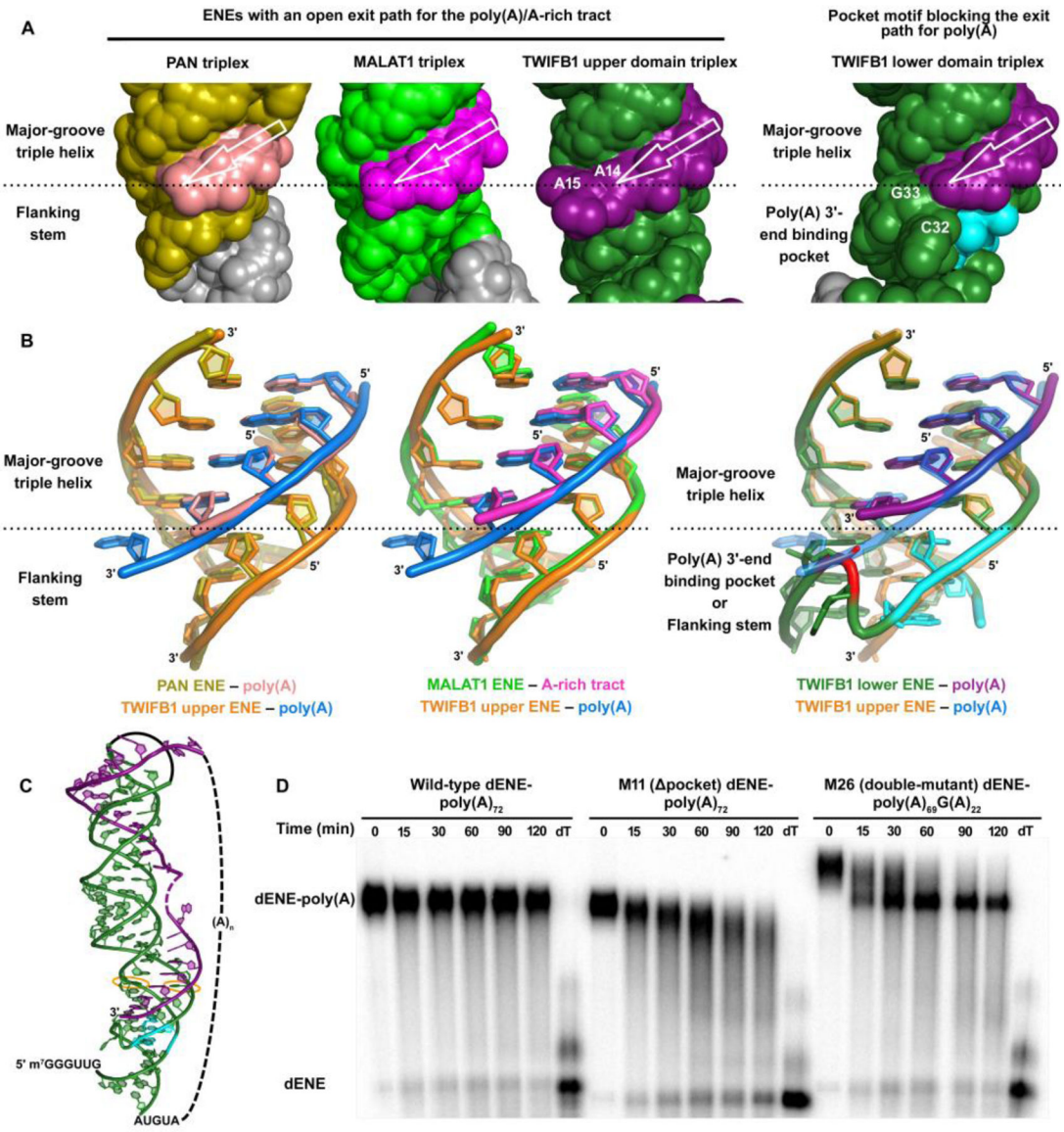

Fig. 4.

The TWIFB1 dENE pocket motif blocks the poly(A) exit path, forming a 3′ blunt-ended triplex. (A) Molecular surface representations of different ENEs oriented to show the direction of poly(A)/A-rich tract (white open arrow), the 3′-most adenosine of each major-groove triplex and its surrounding region. (B) Superposition of the triple helices from various ENEs using the last four U-A•U triples and three base pairs of their flanking stems. The upper dENE domain is superimposed over the PAN or MALAT1 ENE with an RMSD = 1.1 or 1.2 Å, respectively. In the superposition of the two dENE domains, only the major-groove triple helices were used (RMSD = 0.7 Å). The upper dENE triplex on right is shown in transparent colors. The red-colored region represents predicted steric clash that should occur between the backbone of poly(A) and the pocket motif if the poly(A) strand leaves the major-groove triplex in the lower dENE domain. PAN and MALAT1 triplex structures were derived from PDB 3P22 (11) and 4PLX (17), respectively. (C) Cartoon representation

of the TWIFB1 dENE with a 72-mer poly(A) tail. Us changed to Cs in M26 are circled in orange. (D) In vitro deadenylation assays show that wild-type dENE protects its downstream poly(A) from deadenylation, while the absence of the pocket in M11 abrogates this ability. In M26, part of the poly(A) tail is trimmed rapidly and a product of the same size as wild-type dENE-poly(A)$_{72}$ accumulates. dT lanes show transcripts fully deadenylated by endogenous RNase H in the HeLa cell nuclear extract upon addition of oligo(dT)$_{40}$ to the reaction.

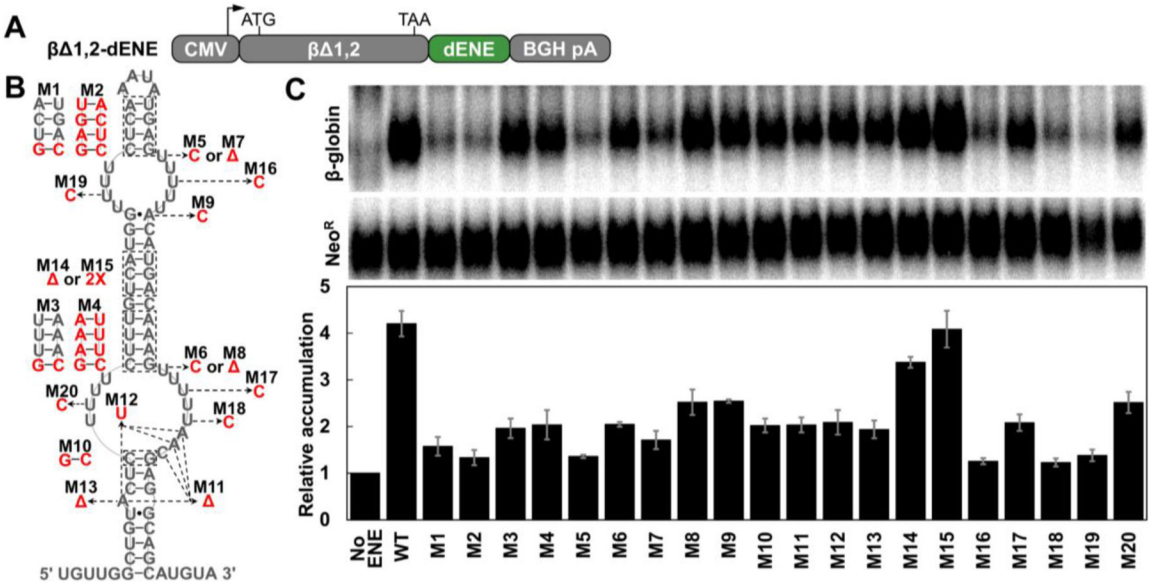

Fig. 5.

Functional significance of dENE structural features contributing to the stabilization and accumulation of the β-globin reporter mRNA containing different dENE mutations. (A) Schematic of the βΔ1,2-dENE construct containing a CMV promoter, a human βΔ1,2 gene and a bovine growth hormone poly(A) site (BGH pA). (B) Schematic of the dENE predicted secondary structure highlighting the mutations studied in the assay. Gray is wild-type sequence, mutated nucleotides are red and Δ represents nucleotide deletion/s. In M12, all three As of the adenosine triad are substituted with U. (C) Northern blots (top) probed for β-globin and neomycin resistance (Neo^R) sequences. Quantification of the Northern blots (bottom) with the β-globin signals normalized to those of Neo^R. Accumulation of No ENE was set at 1. Relative accumulation is the average of at least three independent experiments ± standard deviation.

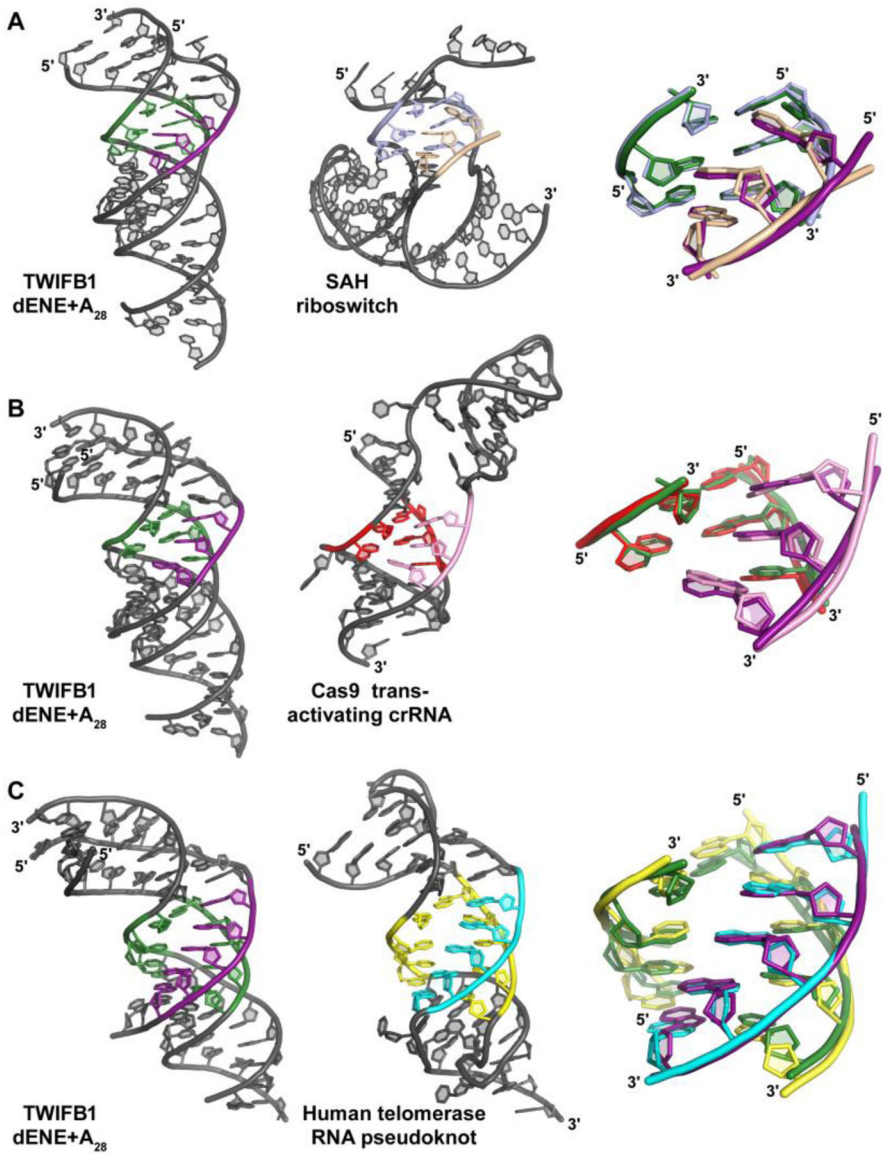

Fig. 6.

Similarity between structural motifs identified in the dENE+poly(A)$_{28}$ complex and equivalent motifs from other RNAs. Only the motifs that are common between compared structures are highlighted with color. (A) Structures of dENE+poly(A)$_{28}$, the SAH riboswitch (PDB 3NPQ (20)) and superposition of their WC/H A-minor motifs (only interactions involving A7-A9 from the dENE) with an RMSD = 0.9 Å. (B) Structures of dENE+poly(A)$_{28}$, the Cas9 trans-activating crRNA (PDB 5X2G (21)) and superposition of their WC/H A-minor interaction and quintuple-base transition motif (only interactions involving A8-A10 from the dENE) with RMSD = 0.9 Å. (C) Structures of dENE+poly(A)$_{28}$, the hTR pseudoknot (PDB 2K95 (22)) and superposition of their quintuple-base transition and major-groove triplex (only interactions involving A9-A13 from the dENE) with RMSD = 1.5 Å.