Published in final edited form as:

Science. 2010 November 26; 330(6008): 1244–1247. doi:10.1126/science.1195858.

Poly(A) Tail Recognition by a Viral RNA Element Through Assembly of a Triple Helix

Rachel M. Mitton-Fry¹, Suzanne J. DeGregorio¹, Jimin Wang², Thomas A. Steitz^{2,3,4}, and Joan A. Steitz^{1,*}

¹Department of Molecular Biophysics and Biochemistry (MB&B), Howard Hughes Medical Institute (HHMI), Yale University School of Medicine, Boyer Center for Molecular Medicine, 295 Congress Avenue, New Haven, CT 06536–9812, USA.

²Department of MB&B, Yale University, Bass Center for Molecular and Structural Biology, 266 Whitney Avenue, New Haven, CT 06520–8114, USA.

³Department of Chemistry, Yale University, New Haven, CT 06520–8107, USA.

⁴HHMI, Yale University, Bass Center for Molecular and Structural Biology, 266 Whitney Avenue, New Haven, CT 06520–8114, USA.

Abstract

Kaposi's sarcoma–associated herpesvirus produces a highly abundant, nuclear noncoding RNA, polyadenylated nuclear (PAN) RNA, which contains an element that prevents its decay. The 79-nucleotide expression and nuclear retention element (ENE) was proposed to adopt a secondary structure like that of a box H/ACA small nucleolar RNA (snoRNA), with a U-rich internal loop that hybridizes to and protects the PAN RNA poly(A) tail. The crystal structure of a complex between the 40-nucleotide ENE core and oligo(A)₉ RNA at 2.5 angstrom resolution reveals that unlike snoRNAs, the U-rich loop of the ENE engages its target through formation of a major-groove triple helix. A-minor interactions extend the binding interface. Deadenylation assays confirm the functional importance of the triple helix. Thus, the ENE acts as an intramolecular RNA clamp, sequestering the PAN poly(A) tail and preventing the initiation of RNA decay.

Kaposi's sarcoma–associated herpesvirus (KSHV) is the causative agent of Kaposi's sarcoma (KS), the most common AIDS-associated cancer (1,2). Although largely controlled in the developed world by antiretroviral treatment against HIV, KS has become one of the most prevalent cancers in Africa (3). KSHV is a γ -herpesvirus that exists in either a latent or lytic state. During the lytic phase, KSHV produces PAN (polyadenylated nuclear) RNA, a 1.1-kb noncoding RNA with a 5' cap and 3' poly(A) tail (4–6) that is retained in the nucleus of infected cells. Although the function of PAN RNA is not known, it accumulates to extremely high levels, amounting to as much as 80% of the polyadenylated RNA in the cell (4,6). The expression and nuclear retention element (ENE), a 79-nucleotide (nt) element located near the 3' end of PAN RNA, is responsible for this accumulation (7). The ENE prevents deadenylation and decay of PAN RNA through direct, cis-acting sequestration of

*To whom correspondence should be addressed. joan.steitz@yale.edu.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6008/1244/DC1

Materials and Methods

SOM Text

Figs. S1 to S8

Table S1

the poly(A) tail (8). The ENE also abrogates rapid nuclear decay when inserted into polyadenylated mRNAs lacking introns (8), apparently through protection from deadenylation, the first step in degradation of eukaryotic mRNAs (9,10). Secondary structure prediction and analyses of mutants using cellular assays suggested that the ENE forms a hairpin containing a U-rich internal loop flanked by short helices, reminiscent of box H/ACA small nucleolar RNA (snoRNA) hairpins (Fig. 1A) (8,11). The internal loops of box H/ACA snoRNAs base pair with target regions in ribosomal RNA (rRNA) to direct the conversion of specific uridine residues to pseudouridine (12). By analogy, the ENE internal loop was predicted to base pair with the poly(A) tail of PAN RNA (Fig. 1A), protecting it from exonucleolytic digestion (11).

To elucidate the mode of interaction between the ENE and the poly(A) tail of PAN RNA, we have determined the crystal structure of a complex formed between the 40-nt ENE core and oligo(A) RNA and studied it biochemically (13). Electrophoretic mobility shift assays showed that the isolated ENE is capable of binding oligo(A) in trans (fig. S1). U to C mutations in the internal loop greatly decrease the interaction between the ENE and oligo(A), whereas truncations of the ENE that retain the helix-loop-helix core (Fig. 1A) have no effect (fig. S1). These results are consistent with those obtained from in vivo and in vitro assays of ENE mutants (11). Nuclear magnetic resonance (NMR) studies confirmed that the ENE core alone adopts a helix-loop-helix conformation (fig. S2A). Increased spectral complexity in the imino region of one-dimensional ^1H spectra upon A_{10} addition suggested that the loop becomes involved in hydrogen-bonding interactions with the oligo(A) (fig. S2B). Crystallization trials were then performed using several ENE core sequences that varied in the helix length mixed with a variety of short A-rich RNA oligonucleotides. Optimal crystallization occurred with the 40-nt ENE core construct and A_9 RNA. Data from crystals soaked in buffer solutions containing iridium hexamine were used to determine the phases, and the final structure was refined to 2.5 Å resolution ($\langle I/\sigma_I \rangle = 1.0$; $\langle I/\sigma_I \rangle = 2.0$ at 2.65 Å resolution) with working and free R factors (R_{work} and R/R_{free}) of 21.8%/23.3%, respectively (figs. S3 to S5 and table S1).

The crystal structure reveals that the ENE core forms a triple-stranded complex with its bound A_9 oligonucleotide (Fig. 1). The ENE assumes the expected secondary structure, with Watson-Crick stems flanking a U-rich internal loop (10). Instead of forming base pairs around this loop, however, the A_9 RNA adopts an extended conformation and interacts with both the ENE loop and the lower stem. Nucleotides A5 to A9 of the A_9 oligonucleotide simultaneously engage both sides of the internal loop in an extended U-A•U major-groove triple helix (Fig. 2A). Here, the five consecutive A nucleotides form Watson-Crick base pairs with the five consecutive U nucleotides on the 3' side of the loop [U27 to U31 (14)]. The 5' side of the ENE internal loop lies in the major groove of the resulting helix, with the Watson-Crick face of nucleotides U8 to U12 base pairing with the Hoogsteen face of the A nucleotides (Fig. 2B). Hydrogen bonds are also observed between the A_9 phosphate backbone and the ribose hydroxyl groups of the Hoogsteen U strand. The base triples formed are nearly planar, and there is no direct contact between the two U strands of the internal loop (Fig. 2, A and B). The helical axis of the complex is nearly straight through the transitions between the triple helix and the flanking ENE stems. The additional two nucleotides of the 3' side of the ENE internal loop (A32 and U33) bulge out, allowing continuous stacking between the lower stem and the A and Hoogsteen strands of the triple helix. Despite the presence of the Hoogsteen strand in the A:U major groove, a deep groove remains (fig S6).

The binding interface is augmented by a triad of A-minor interactions formed between the bound A_9 oligonucleotide and the lower stem of the ENE. Nucleotides A2 to A4 contact the three consecutive G-C base pairs that close the lower stem (Fig. 2C). As previously

described for 23S rRNA in the high-resolution structure of the 50S ribosome (15), the three As of the triad penetrate with increasing depth into the G:C minor groove from the 5' to 3' direction, such that the initial interaction is a type III, the next is a type II, and the last is a type IA-minor interaction (Fig. 2D) (15,16).

The interactions made between the ENE core and A₉ RNA are strikingly different from those determined for the box H/ACA snoRNAs and their target rRNA sequences (Fig. 3, A and B). In the snoRNA structures, the substrate strand is sharply kinked, forming parallel Watson-Crick helices with the two strands of the snoRNA internal loop, which then stack coaxially on the flanking snoRNA hairpin helices (17, 18). The crystal structure described here involves an intermolecular interaction between the core ENE and A₉ RNA. However, as the ENE has only been observed to protect the PAN RNA poly(A) tail in an intramolecular manner (8), these elements formally adopt a pseudoknot structure in the context of the full-length RNA (see supporting online material). Consistent with this, the ENE A₉ complex resembles H-type pseudoknots observed in telomerase RNA (Fig. 3, C and D) (19) and the S-adenosylmethionine-responsive riboswitch SAM-II (fig. S7) (20). Although the topology of the surrounding stems is different in the H-type pseudoknots, the structures of the major-groove triple helices can be superimposed with that of the ENE complex with root mean square deviations (RMSDs) of ~1 Å. Felsenfeld, Davies, and Rich first studied RNA triple helices composed of poly(U)-poly(A)•poly(U) more than half a century ago (21, 22). Denaturation studies later suggested that DNA hairpins with T-rich loops likewise bind oligo(dA) through formation of T-A•T triple helices (23). Analogous interactions between RNA hairpins and A-rich sequences were also proposed (23). However, the ENE is the first natural example of the use of a U-rich internal loop to capture and sequester a poly(A) RNA sequence.

We used deadenylation assays to confirm the functional importance of the ENE:A₉ triple helix described here (13). Mutation of a single U to C in either side of the ENE U-rich pocket had previously been shown to decrease protection of PAN RNA's poly(A) tail from deadenylation in nuclear extract (11). Simultaneous U to C mutations (mutating one nucleotide from each side of the pocket) were as deleterious as deletion of the entire ENE. Because the two U nucleotides mutated in that experiment contact the same A in the crystal structure (Fig. 4A), we postulated that the loss of ENE function arising from disruption of a single U-A•U base triple might be restored by replacement with a nearly isosteric C-G•C base triple (Fig. 4B) (24). Native gel shift analysis supported this proposal, as ENEs containing simultaneous U to C mutations contacting the same A in the crystal structure can bind an oligo(A) molecule containing a single A to G substitution (A₇GA₂) (fig. S8, A and C), whereas ENE constructs with mutations to C in U residues that contact two different A nucleotides in the crystal structure cannot (fig. S8, B and C).

We thus transcribed PAN RNA deadenylation substrates that contained either wild-type or double-mutant ENE or that lacked the ENE altogether (see legend to Fig. 4C). Each of the substrates terminated in a 60-nt poly(A) tail with or without a single A to G substitution (A₆₀ or A₅₇GA₂). We assessed these constructs for ENE-dependent protection from deadenylation in nuclear extract (Fig. 4C, left and middle panels). As expected, the wild-type ENE protected the A₆₀ tail from deadenylation, whereas the A₆₀ tails of constructs containing no ENE (Δ ENE) or the double-mutant ENE (U903C, U949C) were not protected from deadenylation (Fig. 4C, left panels) (10). In contrast, the double-mutant ENE effectively protected the A₅₇GA₂ poly(A) tail, which has a single A to G substitution close to its 3' end (Fig. 4C, center panel). The presence of a G residue in the poly(A) tail does not on its own confer resistance to deadenylation, as the A₅₇GA₂ poly(A) tail of a construct lacking the ENE was not protected (Fig. 4C, bottom middle panel). These results, supported

by the native gel shift data described above (fig. S8), indicate that the triple helix observed in the structure is critical to ENE function.

Finally, we tested the ability of the double-mutant ENE to protect a poly(A) tail containing a more internal G substitution (A₁₉GA₄₀) (Fig. 4C, right panel). In nuclear extract, this poly(A) tail was rapidly deadenylated to a size consistent with formation of a triple helix that includes the predicted C-G•C base triple (Fig. 4C, right middle panel). The ability of the double-mutant ENE to locate a single G within a long stretch of A nucleotides is striking. The deadenylation data also demonstrate that the ENE does not require the 3' terminus of the poly(A) tail for binding, despite the involvement of the 3' end of the A₉ RNA in the final base triple of the ENE core:A₉ structure. The results further argue that no specific register is required for the interaction of the wild-type ENE with the PAN RNA poly(A) tail. The presence of multiple binding sites for the ENE along the poly(A) sequence may contribute to its ability to protect tails of various lengths from deadenylation by cellular exonucleases (Fig. 4A). How the ENE may collaborate with poly(A)—binding proteins that are known to coat the poly(A) tails of RNA polymerase II transcripts in vivo (25) remains to be determined.

The key feature of the core ENEL:A₉ crystal structure is a functionally important U-A•U major-groove triple helix, which is extended by A-minor interactions. The structure reveals an intramolecular clamp mechanism for recognition of poly(A) RNA and suggests how the ENE sequesters the PAN poly(A) tail from degradation by cellular deadenylases. Since viruses routinely borrow strategies from their hosts, we predict that similar mechanisms may protect some cellular noncoding RNAs from rapid turnover.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank S. Borah, K. Tycowski, K. Herbert, and K. Riley for critical reading of the manuscript and the entire Steitz laboratories for helpful discussion. We thank S. Strobel and G. Conn for the generous gifts of iridium(III) hexamine and 3'-hepatitis delta virus (HDV) plasmid, respectively. Special thanks to Y. Zuo and G. Blaha for crystallography assistance, P. Moore and E. Paulson for NMR assistance, and M.-D. Shu, D. Mishler, and K. Durniak for technical assistance. X-ray data were collected at the National Synchrotron Light Source (X29A) at Brookhaven National Laboratory (13). Financial support for this research was provided by NIH grant CA16038 to J.A.S. and NIH grant GM022778 to T.A.S. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. R.M.M.-F. is supported by a Jane Coffin Childs Memorial Fund Postdoctoral fellowship. J.A.S. and T.A.S. are investigators of the Howard Hughes Medical Institute. Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 3P22.

References and Notes

1. Viejo-Borbolla A, Ottinger M, Schulz TF. *Curr. HIV/AIDS Rep.* 2004; 1:5. [PubMed: 16091217]
2. Mesri EA, Cesarman E, Boshoff C. *Nat. Rev. Cancer.* 2010; 10:707. [PubMed: 20865011]
3. Parkin DM, et al. *Lancet Oncol.* 2008; 9:683. [PubMed: 18598933]
4. Sun R, Lin SF, Gradoville L, Miller G. *Proc. Natl. Acad. Sci. U.S.A.* 1996; 93:11883. [PubMed: 8876232]
5. Zhong W, Wang H, Herndier B, Ganem D. *Proc. Natl. Acad. Sci. U.S.A.* 1996; 93:6641. [PubMed: 8692871]
6. Song MJ, Brown HJ, Wu TT, Sun R. *J. Virol.* 2001; 75:3129. [PubMed: 11238840]
7. Conrad NK, Steitz JA. *EMBO J.* 2005; 24:1831. [PubMed: 15861127]
8. Conrad NK, Mili S, Marshall EL, Shu M-D, Steitz JA. *Mol. Cell.* 2006; 24:943. [PubMed: 17189195]

9. Parker R, Song H. *Nat. Struct. Mol. Biol.* 2004; 11:121. [PubMed: 14749774]
10. Goldstrohm AC, Wickens M. *Nat. Rev. Mol. Cell Biol.* 2008; 9:337. [PubMed: 18334997]
11. Conrad NK, Shu M-D, Uyhazi KE, Steitz JA. *Proc. Natl. Acad. Sci. U.S.A.* 2007; 104:10412. [PubMed: 17563387]
12. Kiss T, Fayet-Lebaron E, Jády BE. *Mol. Cell.* 2010; 37:597. [PubMed: 20227365]
13. Materials and methods are available as supporting material on *Science* Online.
14. Numbering from the 5' end of the crystallization construct. ENE core nucleotides 1 to 16 correspond to PAN nucleotides 894 to 909, and 23 to 39 correspond to PAN nucleotides 943 to 959.
15. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. *Proc. Natl. Acad. Sci. U.S.A.* 2001; 98:4899. [PubMed: 11296253]
16. Doherty EA, Batey RT, Masquida B, Doudna JA. *Nat. Struct. Biol.* 2001; 8:339. [PubMed: 11276255]
17. Jin H, Loria JP, Moore PB. *Mol. Cell.* 2007; 26:205. [PubMed: 17466623]
18. Wu H, Feigon J. *Proc. Natl. Acad. Sci. U.S.A.* 2007; 104:6655. [PubMed: 17412831]
19. Theimer CA, Blois CA, Feigon J. *Mol. Cell.* 2005; 17:671. [PubMed: 15749017]
20. Gilbert SD, Rambo RP, Van Tyne D, Batey RT. *Nat. Struct. Mol. Biol.* 2008; 15:177. [PubMed: 18204466]
21. Felsenfeld G, Rich A. *Biochim. Biophys. Acta.* 1957; 26:457. [PubMed: 13499402]
22. Felsenfeld G, Davies D, Rich A. *J. Am. Chem. Soc.* 1957; 79:2023.
23. D'Souza DJ, Kool ET. *J. Biomol. Struct. Dyn.* 1992; 10:141. [PubMed: 1418737]
24. The C-G•C base triple is strongest upon protonation of the Hoogsteen C nucleotide, but the triple helical structure can increase the pK_a of the C base, such that full protonation occurs at neutral pH (19).
25. Kühn U, Wahle E. *Biochim. Biophys. Acta.* 2004; 1678:67. [PubMed: 15157733]
26. Leontis NB, Westhof E. *RNA.* 2001; 7:499. [PubMed: 11345429]
27. Kim N-K, et al. *J. Mol. Biol.* 2008; 384:1249. [PubMed: 18950640]
28. Dignam JD, Lebovitz RM, Roeder RG. *Nucleic Acids Res.* 1983; 11:1475. [PubMed: 6828386]

Fig. 1.

Structural overview of the complex formed between the ENE core and A₉ RNA. **(A)** Schematic diagram of PAN RNA, showing the interaction between the ENE and the poly(A) tail that was previously proposed (8,11), with the ENE core in green. For crystallization, the upper stem of this central sequence was capped with a C-G base pair and a GAAA tetraloop and the lower stem was followed by a 3' C nucleotide (to produce a blunt-ended helix) (Fig. 1D). **(B)** Ribbon representation of the crystal structure, with the ENE core in green, the non-native sequence in gray, and the A₉ oligonucleotide in magenta. Four copies of the ENE complex are found in the asymmetric unit of the crystal, two pairs of complexes related by noncrystallographic symmetry (NCS). The two non-NCS-related complexes are almost identical in the central region [RMSD = 0.57 Å over all nonhydrogen atoms in the triple helix (fig. S3)]. **(C)** 90° rotation of the image shown in (B). **(D)** Schematic diagram of the complex [colors as in (B and C)], with Leontis-Westhof notation (**26**) indicating major-groove triple helix contacts and dotted lines indicating A-minor interactions. The base of the bulged nucleotide A32 (outlined) is disordered.

Fig. 2.

Detailed views of key structural interactions between the ENE and A₉ RNA. **(A)** Close-up of the major-groove triple helix formed by 5 nt of A₉ and the internal U-rich loop of the ENE hairpin (colors as in Fig. 1, with the bases involved in Watson-Crick pairing in yellow, hydrogen bonds within base triples in cyan, and the A₉ nucleotides labeled in italics). The U strand that makes Hoogsteen interactions in the major groove is shown in the foreground. **(B)** Superposition of the five ENE:A₉ U-A•U base triples showing the regularity of the triple-helical structure. **(C)** Close-up of the A-minor triad, with Watson-Crick base pairing and hydrogen bonds involving the A strand in cyan. **(D)** A-minor interactions with the lower ENE stem. Crystal-packing interactions involving nucleotide A1 (fig. S3A) most likely pull the N2 of nucleotide A2 just beyond hydrogen-bonding distance from the 2' OH of C36 in the Type III interaction.

Fig. 3.

Comparison of the ENE:A₉ complex with other structures. **(A)** The complex between the ENE and A₉ RNA (top three base triples boxed). **(B)** The solution structure of a complex between a truncated H/ACA snoRNA and a 14-nt sequence corresponding to its rRNA substrate [PDB 2P89, model 1 (18)]. The snoRNA hairpin is shown in green, with the bound substrate in yellow. **(C)** The solution structure of the human telomerase pseudoknot [PDB 2K95, model 1 (27)]. The triple helix (boxed) is composed of three U-A•U base triples, and the strand contributing these A nucleotides is shown in cyan, whereas the strands that contribute the U nucleotides are shown in brown. **(D)** Superposition of the triple helix from the ENE:A₉ complex with the triple helix from the telomerase RNA pseudoknot. Colors and orientations are as in (A) and (C). The top three U-A•U base triples (involving A7 to A9) from the ENE:A₉ complex were used for superposition (RMSD = 1.1 Å over all nonhydrogen atoms).

Fig. 4.

Triple-helix assembly protects the PAN RNA poly(A) tail from deadenylation. (A) Cartoon of the ENE:A₉ complex structure in the context of full-length PAN, highlighting the locations of U903 and U949 (numbering from the PAN RNA 5' end). (B) Comparison of U-A•U and C-G•C⁺ base triples [colors as in (A), with the hydrogen bond formed upon protonation of the Hoogsteen C nucleotide in blue]. (C) In vitro deadenylation assays show that single G substitutions in the poly(A) tail rescue a nonfunctional double-mutant ENE by formation of C-G•C base triples (24). The substrates consist of the 327-nt 3' terminus of PAN RNA followed by a 60-nt tail either composed of all adenylate (A₆₀) or with single G substitutions 3 or 41 nucleotides from the 3' end of the poly(A) tail (A₅₇GA₂ or A₁₉GA₄₀, respectively); tail identity is designated above the panels. After incubation in HeLa cell nuclear extract (28) for the indicated times, products were separated by denaturing gel electrophoresis. RNAs containing the wild-type ENE, double-mutant (U903C, U949C) ENE, and Δ ENE are shown in the upper, middle, and lower panels, respectively. +dT lanes refer to transcripts in which the poly(A) tail was removed by endogenous ribonuclease H after addition of oligo(dT)₄₀ to the reaction mix. The A₆₀ and A₀ labels on the left show the migration of fully adenylated and deadenylated substrates.