On the occurrence of the T-loop RNA folding motif in large RNA molecules

ANDREY S. KRASILNIKOV and ALFONSO MONDRAGÓN

Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA

ABSTRACT

The T-loop RNA folding motif may be considered as a five-nucleotide motif composed of a U-turn flanked by a noncanonical base pair. It was recently proposed that the flanking noncanonical base pair is always a UA *trans* Watson-Crick/Hoogsteen base pair stacked on a Watson-Crick base pair on one side. Here we show that structural analysis of several large RNA molecules, including the recently solved crystal structure of the specificity domain of *Bacillus subtilis* RNase P, combined with sequence analysis, indicates a broader sequence consensus for the motif. Additionally, we show that the flanking base pair does not necessarily stack on a Watson-Crick base pair and the 3' terminus of the five-nucleotide motif is often followed by a sharp turn in the phosphate backbone rather than just a bulged base or bases.

Keywords: RNA; structure; motif; U-turn; T-loop

Three-dimensional organization of large RNA molecules involves, apart from regular double helices, a variety of complex conformational motifs (for reviews, see Hermann and Patel 1999; Moore 1999). Rapid progress in the determination of structures of large RNA molecules has allowed characterization of new structural motifs, including the tetraloop–tetraloop receptor (Pley et al. 1994; Cate et al. 1996a), ribose zippers (Cate et al. 1996a), dinucleotide platforms (Cate et al. 1996b), A-minor motifs (Doherty et al. 2001; Nissen et al. 2001), kink-turns (Klein et al. 2001), and others (Hermann and Patel 1999; Moore 1999). The T-loop RNA folding motif was recently identified in a number of RNA structures (Nagaswamy and Fox 2002). It contains a U-turn (both UNR and GNR U-turns are allowed) flanked by a noncanonical base pair and has a distinctive three-dimensional structure (Fig. 1) similar to the one found in the T ψ C-loop of tRNAs, and hence its name.

In the original work of Nagaswamy and Fox (2002) describing the occurrence of the T-loop motif, the search for the T-loop motif was confined to sequences in which the base pair flanking the U-turn (Fig. 1, bases 1 and 5) forms a UA base pair, thus excluding any other base pair that might occur at this position. We believe that when there is sufficient structural data for analysis, a motif should be

defined not only in terms of sequence similarity, but also in terms of structural similarity. Furthermore, the three-dimensional organization of the motif should be independent of the structural context where it is found and should be consistent with sequence variation analysis. Indeed, initially a motif may be observed from sequence analysis alone (as the number of sequences available is much higher than the still limited number of structures), but later it may be found that a similar structure is observed with a sequence somewhat different from the original definition. It is then only natural to extend the definition of the motif to include those sequence variations rather than introduce a “new” motif or exclude those sequences that do not conform to the original definition. For instance, consider the tetraloop receptor motif. The original definition contained an AA platform as part of the motif (Costa and Michel 1995; Cate et al. 1996a), but later studies showed that the AA was interchangeable with an AC (Costa and Michel 1997) and that, furthermore, the structure in the presence of either sequence is identical (Krasilnikov et al. 2003). Structural motifs are best identified and described when both structural and sequence information is included.

In the structure of the specificity domain of *Bacillus subtilis* RNase P (Krasilnikov et al. 2003), we noted the presence of the T-loop motif in different contexts that do not necessarily include a UA pair. This prompted us to analyze the occurrence of the T-loop motif in the structures of the 23S rRNA (Protein Data Bank 1JJ2; Ban et al. 2000), the 16S rRNA (PDB 1J5E; Wimberly et al. 2000), the yeast initiator tRNA (PDB 1YFG; Basavappa and Sigler 1991),

Reprint requests to: Alfonso Mondragón, Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2205 Tech Drive, Evanston, IL 60208, USA; e-mail: a-mondragon@northwestern.edu.

Article and publication are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2202703>.

FIGURE 1. The T-loop motif (yellow) is formed by a U-turn (bases 2–4) flanked by a non-Watson-Crick base pair (bases 1 and 5) and is followed by a sharp turn or a bulged base or bases at its 3' terminus. Red dots represent interactions crucial for the stability of the motif (Nagaswamy and Fox 2002). The estimation of the rate of occurrence of the particular nucleotides was based on the analysis of more than 4000 sequences. The 217–223 fragment of the S-domain of the RNase P (Krasilnikov et al. 2003) is shown as an example.

and the S-domain of RNase P (PDB 1NBS; Krasilnikov et al. 2003). Rather than searching for sequence patterns, we used structural similarity in our search: namely, we superposed the T-loop motif in phe-tRNA (PDB 1EHZ; Shi and Moore 2000) on all possible continuous five-nucleotide stretches in these structures and scored them by using the root mean square deviation (r.m.s.d.) between the two sets of coordinates. As expected, besides the motifs identified by Nagaswamy and Fox (2002), we found additional occurrences (Table 1). For some occurrences, including some reported by Nagaswamy and Fox (2002), it was difficult to ascertain from the structure whether these examples really contained the required base pair, and hence they were excluded from the comparison. In the case of the 335–339 loop in 23S rRNA, which conforms to the sequence consensus, it was hard to conclude that it forms a similar structure to the one in the T-loop motif in tRNA; therefore, it was excluded from our analysis (Fig. 2).

Table 1 shows that a fold similar to the one observed in the T ψ C-loop of tRNA is abundant in large RNA molecules (12 examples with average r.m.s.d. of 1.3 Å). Table 1 also shows that the non-Watson-Crick base pair that flanks the U-turn is not necessarily a *trans* W.-C./Hoogsteen UA,

as proposed by Nagaswamy and Fox (2002). For example, a *trans* W.-C./Hoogsteen CA base pair is found in the flanking position of the 505–509 T-loop of the 23S rRNA (1JJ2), and *trans* W.-C./Hoogsteen AA base pairs are observed in the 187–191 T-loop motif of the S-domain of the RNase P (PDB 1NBS) and in the T ψ C-loop of the yeast initiator tRNA (PDB 1YFG). Although the abundance and well-defined three-dimensional structure of the T-loop motif (Fig. 3) definitely warrant its designation as a separate structural motif, the consensus sequence is looser than the one previously proposed (Nagaswamy and Fox 2002).

In order to derive the sequence consensus for the T-loop motif, we aligned the sequences of the two T-loop motifs

corresponding to the J11/12–J12/11 module of *B. subtilis* RNase P (bases 187–191 and 218–222) and six T-loop motifs in the 23S rRNA from *Haloarcula marismortui* (Table 1) with corresponding regions from different organisms. The sequence analysis was done using sequences and sequence

TABLE 1. Examples of T-loop motifs found in large RNA structures

Structure (PDB accession code and resolution)	Position	Sequence	r.m.s.d.	Intercalating base	Flanking base pair
phe-tRNA (1EHZ, 1.93 Å)	54–58	UUCGA	-	G	U-A, WC/H ^a (A-A, C-A) ^b
Initiator tRNA (1YFG, 3.0 Å)	54–58	AUCGA	1.46 Å	G	A-A, WC/H ^a
S-domain (1NBS, 3.15 Å)	187–191	AGUGA	1.38 Å	A	A-A, WC/H ^a (G-A) ^b
S-domain (1NBS, 3.15 Å)	218–222	UGGAA	1.26 Å	A	U-A, WC/H ^a
23S rRNA (1JJ2, 2.40 Å)	313–317	UGGAA	1.25 Å	A	U-A, WC/H ^a
23S rRNA (1JJ2, 2.40 Å)	481–485	UGCAA	1.10 Å	A	U-A, WC/H ^a (C-A) ^b
23S rRNA (1JJ2, 2.40 Å)	505–509	CGAAA	1.20 Å	A	C-A, WC/H ^a (U-A) ^b
23S rRNA (1JJ2, 2.40 Å)	624–628	UUUGA	1.27 Å	C	U-A, WC/H ^a (U-U) ^b
23S rRNA (1JJ2, 2.40 Å)	1388–1392	UGAGA	1.29 Å	U	U-A, WC/H ^a (C-A) ^b
23S rRNA (1JJ2, 2.40 Å)	2597–2601	UUAAA	1.82 Å	None	U-A, WC/H ^a (C-U) ^b
16S rRNA (1J5E, 3.05 Å)	323–327	UGAGA	1.30 Å	A	U-A, WC/H ^a
16S rRNA (1J5E, 3.05 Å)	1315–1319	UGCAA	1.13 Å	A	U-A, WC/H ^a

^a(WC/H) *trans* Watson-Crick/Hoogsteen base pair.

^bPossible alternative flanking base pairs suggested by sequence variation analysis.
(r.m.s.d.) Root mean square deviation.

FIGURE 2. Superposition of the T-loop motif in tRNA (PDB 1EHZ, shown in gold) and the 335–339 region of 23S rRNA (PDB 1JJ2, shown in pink). The flanking base pair is the same in both loops, but the overall structure is very different. Note that the important hydrogen bonds that serve to stabilize the U-turn cannot be made in the 335–339 region.

alignments found in two large public domain databases (Brown 1999; Wuyts et al. 2001). The sequence alignment yielded 4301 sequences that are likely to adopt the T-loop motif conformation. The results of the sequence comparison are summarized in Figure 1.

As mentioned earlier, although the flanking base pair (Fig. 1, bases 1 and 5) is indeed mostly a UA, other base pairs can be found at this position as well. Both analysis of known large RNA structures and sequence comparison show that a *trans* W.-C./Hoogsteen UA base pair can be substituted with a *trans* W.-C./Hoogsteen CA base pair. Substitution of the UA for a CA base pair can be found in T-loop motifs of some tRNAs (e.g., eukaryotic met-tRNAs [Lowe and Eddy 1997]) as well as in several T-loops in 23 S RNA (see Table 1 and Ban et al. 2000). This is not surprising given that *trans* W.-C./Hoogsteen UA and CA base pairs are isosteric (Leontis and Westhof 1998). More surprising is the presence of a *trans* W.-C./Hoogsteen AA in the T-loop mo-

tif of the TψC loop of yeast initiator tRNA (PDB 1YFG), as this base pair is not isosteric to a *trans* W.-C./Hoogsteen UA. Additionally, a *trans* W.-C./Hoogsteen AA also flanks the T-loop motif in a different context: the S-domain of the RNase P (PDB 1NSB). Apparently, the T-loop motif is flexible enough to accommodate some other flanking *trans* base pairs, which are not necessarily isosteric to the *trans* W.-C./Hoogsteen UA originally identified. Sequence analysis (Fig. 1) indicates another possible flanking base pair for the T-loop motif: a UU base pair. The presence of the UU base pair has not been observed in a T-loop of any RNA structure and hence the possible presence of this base pair is still speculative, although likely.

The second base is almost always a G or a U, as proposed (Nagaswamy and Fox 2002). It is possible that a similar fold might be formed with a C replacing a U in the second position, but this clearly does not happen often (in <2% of the aligned sequences of potential T-loop motifs) and it has not been observed in known structures so far. The third base of the motif is in most cases an A. The presence of an A does not appear to be required for the stability of the motif; therefore, the overrepresentation of adenosines in the third position is probably due to their involvement in tertiary interactions. The fourth base of the motif is strictly a purine, as it is required for the stability of the fold (Nagaswamy and Fox 2002), with a preference toward an A, which is probably due to the same reason as for the third base: participation in tertiary interactions.

The T-loop motif is typically observed in phylogenetically well-conserved regions and is involved in a variety of tertiary interactions (Nagaswamy and Fox 2002). A large gap

FIGURE 3. T-loop motifs from different RNA molecules are structurally very similar. The fragments were superposed using the program *lsqkab* (CCP4; Collaborative Computational Project 4 1994). The lines at the 3' termini show the direction of the phosphate backbone following the T-loop motif. In many instances, it is a sharp turn rather than just a bulged base or bases. The backbone may travel different paths beyond the 3' terminus of the motif, as it is not a part of the motif, yet the motif itself is never incorporated directly into an A-helix.

between the fourth and fifth nucleotides in the motif is capable of accepting an intercalating base, thus linking separate parts of a molecule together. In this manner the two T-loop motifs in the S-domain of RNase P (Krasilnikov et al. 2003) bring together the J11/12 and J12/11 strands to form one rigid structural module. In most of the T-loop motifs, this intercalated base (mostly an A) is present in between the fourth and fifth nucleotides (Table 1). The presence of this base slightly changes the geometry of the motif: compared with the T-loop motif from the phe-tRNA (which has an intercalating base) the r.m.s.d. in the single unambiguous example without the intercalating base is 1.82Å, whereas

the average r.m.s.d. for the motifs containing an intercalating base is 1.26Å.

An examination of the T-loop motifs presented in Table 1 also shows that the non-Watson-Crick base pair flanking the U-turn (Fig. 1, bases 1 and 5) is not necessarily stacked on a Watson-Crick base pair on one side, as was suggested by Nagaswamy and Fox (2002). Although both bases are always involved in stacking interactions with neighboring nucleotides, these neighboring nucleotides do not always form a Watson-Crick pair (for example, 187–191, 218–222 in the S-domain of the RNase P [Krasilnikov et al. 2003], 505–509 in 23S rRNA [Ban et al. 2000], and 323–327 in 16S rRNA [Wimberly et al. 2000]). It is true that in all the structural examples available, the phosphate backbone changes direction at the 3' terminus of the flanking base, but in many instances it is a sharp turn rather than just a bulged base or bases.

Our analysis indicates that the T-loop motif is more universal than was previously realized and has a general consensus sequence that does not have an absolute requirement for a UA *trans* W.-C./Hoogsteen base pair in the flanking position. The T-loop motif seems to be widely spread in large RNA structures in many different structural contexts. Although in most instances this motif accepts an intercalating base, this is not always the case. Finally, the presence of the T-loop motif always forces a sharp turn in the 3' terminus of the phosphate backbone, precluding a simple continuation into a stacked stem. The motif is never found simply capping a stem. The RNA strand may travel different paths beyond the 3' terminus of the motif, and may eventually continue into a stacked stem, but always after a sharp turn of the backbone. This property of the T-loop motif, namely, the presence of a strictly single base pair flanking a three-nucleotide loop, has recently allowed for the identification, based on sequence covariation analysis, of what has been termed the Lonepair Triloop RNA motif (Lee et al. 2003). Although the Lonepair Triloop RNA motif is defined differently, it actually overlaps greatly with the T-loop motif.

ACKNOWLEDGMENTS

We thank Tao Pan and Anita Changela for comments and suggestions. Research was supported by an NIH grant (GM58443) to A.M. and an NRSA Fellowship (GM63417) to A.K. Support from the R.H. Lurie Cancer Center of Northwestern University to the Structural Biology Center is acknowledged.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received December 20, 2002; accepted April 3, 2003.

REFERENCES

- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905–920.
- Basavappa, R. and Sigler, P.B. 1991. The 3 Å crystal structure of yeast initiator tRNA: Functional implications in initiator/elongator discrimination. *EMBO J.* **10**: 3105–3111.
- Brown, J.W. 1999. The ribonuclease P database. *Nucleic Acids Res.* **27**: 314.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996a. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273**: 1678–1685.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996b. RNA tertiary structure mediation by adenosine platforms. *Science* **273**: 1696–1699.
- Collaborative Computational Project 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**: 760–763.
- Costa, M. and Michel, F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**: 1276–1285.
- . 1997. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: Comparison with in vivo evolution. *EMBO J.* **16**: 3289–3302.
- Doherty, E.A., Batey, R.T., Masquida, B., and Doudna, J.A. 2001. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**: 339–343.
- Hermann, T. and Patel, D.J. 1999. Stitching together RNA tertiary architectures. *J. Mol. Biol.* **294**: 829–849.
- Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. 2001. The kink-turn: A new RNA secondary structure motif. *EMBO J.* **20**: 4214–4221.
- Krasilnikov, A.S., Yang, X., Pan, T., and Mondragón, A. 2003. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**: 760–764.
- Lee, J.C., Cannone, J.J., and Gutell, R.R. 2003. The lonepair triloop: A new motif in RNA structure. *J. Mol. Biol.* **325**: 65–83.
- Leontis, N.B. and Westhof, E. 1998. Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.* **31**: 399–455.
- Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.
- Moore, P.B. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**: 287–300.
- Nagaswamy, U. and Fox, G.E. 2002. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **8**: 1112–1119.
- Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. 2001. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci.* **98**: 4899–4903.
- Pley, H.W., Flaherty, K.M., and McKay, D.B. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**: 68–74.
- Shi, H. and Moore, P.B. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA* **6**: 1091–1105.
- Wimberly, B.T., Brodersen, D.E., Clemons Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vornheim, C., Hartsch, T., and Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**: 327–339.
- Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., and De Wachter, R. 2001. The European Large Subunit Ribosomal RNA Database. *Nucleic Acids Res.* **29**: 175–177.

RNA

A PUBLICATION OF THE RNA SOCIETY

On the occurrence of the T-loop RNA folding motif in large RNA molecules

ANDREY S. KRASILNIKOV and ALFONSO MONDRAGÓN

RNA 2003 9: 640-643

References

This article cites 22 articles, 8 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/9/6/640.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>