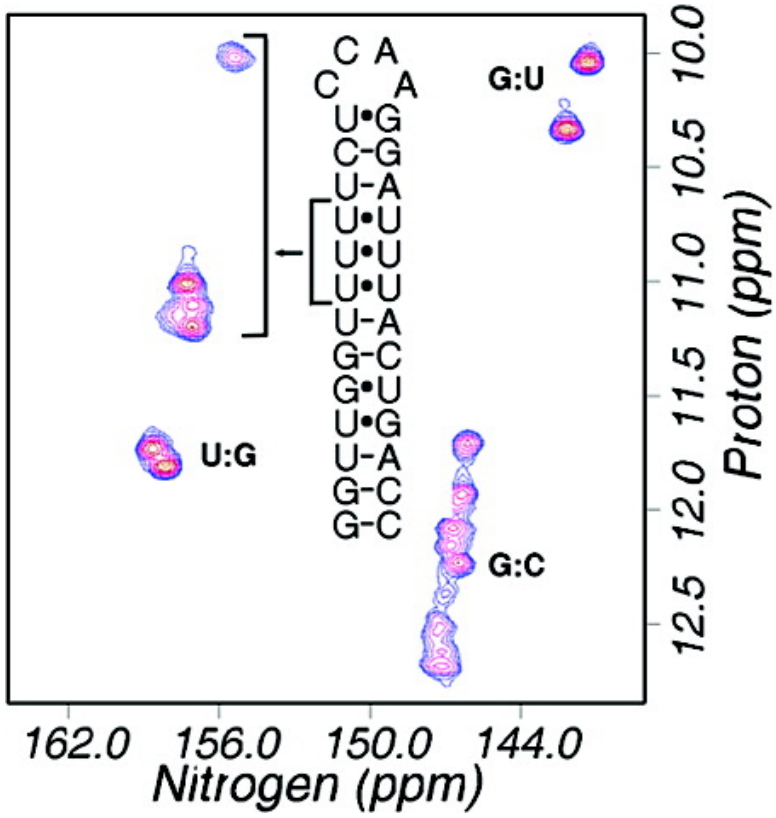
Communication

Stable Triplet of Uracil–Uracil Basepairs in a Small Antisense RNA

Yiping Zhan, and Gordon S. Rule

J. Am. Chem. Soc., **2005**, 127 (45), 15714-15715• DOI: 10.1021/ja055675m • Publication Date (Web): 22 October 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

J|A|C|S
COMMUNICATIONS

Published on Web 10/22/2005

Stable Triplet of Uracil−Uracil Basepairs in a Small Antisense RNA

Yiping Zhan and Gordon S. Rule*

Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213

Received August 18, 2005; E-mail: rule@andrew.cmu.edu

Antisense RNAs are regulatory RNAs that are complementary to their target RNAs and interfere with the activities of their targets due to the formation of duplex RNA.[1,2] The interaction between the ant (antirepressor) mRNA and its antisense RNA, sar (small antisense RNA), is important in regulation of the translation of the ant protein in bacteriophage P22.[3] The biochemical properties of sar−ant pairing have been extensively investigated.[3−5] Schaefer and McClure[4] used limited RNase digestion to determine the secondary structure of sar. On the basis of these studies, they proposed that sar contains two hairpins (Figure 1). A remarkable feature of this structure is the presence of a triple U:U base stack within the first hairpin. Hydrogen-bonded triple U:U base stacks are predicted to be unstable by RNA secondary structure prediction programs[6] and are extremely rare in known structures. There has been only one reported structure of a triple U:U stack in the RNA component of telomerase.[7]

The proposed secondary structure of sar cannot satisfactorily explain all the RNase cleavage patterns.[4] In the case of the first hairpin, the 3′ U in the triple U:U stack is cleaved by a single-strand specific ribonuclease, whereas the adjacent U base, which is part of the central U:U basepair, is cleaved only weakly by a double-strand specific nuclease, suggesting that one, or possibly two, of the U:U basepairs may not be fully paired in the triple U:U stack. Furthermore, the bases within the lower part of the first hairpin are susceptible to cleavage by single-strand specific nucleases. In the case of the second hairpin, single-strand cleavage occurs within the putative helical region at position 36; however, the bulk of the residues in the second hairpin are cleaved by double-strand specific nucleases, albeit weakly.

We have used NMR techniques to unambiguously confirm the proposed secondary structure of sar, and we show that the triple U:U basepairs appear to be a stable feature of this RNA.

Labeled RNA for NMR studies was generated by in vitro transcription using T7 RNA polymerase.[8,9] Templates and reaction conditions for RNA synthesis are described in the Supporting Information. Final yields of purified ^{15}N-labeled RNA were 0.01 mg/mL for the first hairpin, 0.15 mg/mL for the second hairpin, and 0.073 mg/mL for intact sar. Molecular weights of the first hairpin were obtained using MALDI-TOF mass spectrometry;[10] see Supporting Information for details.

Initial attempts to prepare the first hairpin of sar, using a template that was designed to produce exactly the 28 nucleotide hairpin, generated a major aberrant transcript that was approximately 45 nucleotides long (see Figure 1, Supporting Information). Although it is well-known that T7 RNA polymerase produces transcripts with one or two extra nucleotides at the 3′ end of the sequence,[11] the generation of longer aberrant products at very high yield is relatively uncommon. In the case of the first hairpin the five consecutive U's could cause the polymerase to "slip" when synthesizing the RNA.[12] However, this explanation is unlikely because the amount of aberrant product is much lower for whole sar, which also contains the same sequence. It is more likely that the newly synthesized

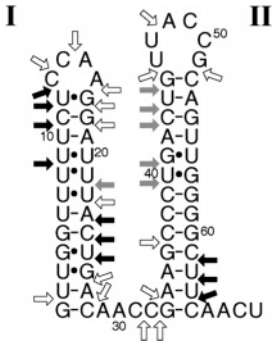


Figure 1. Secondary structure of sar proposed by Schaefer and McClure.[4] Hairpins I and II are labeled. Open arrows correspond to sites sensitive to single-strand specific nucleases.[4] Arrows filled with black are sites that are strongly cleaved by a double-strand specific nuclease (RNase V1). Arrows filled with gray are slowly cleaved by RNase V1.

RNA self-primes due to a free 3′-OH. Products of this type are more likely to occur with transcripts that lack a stable secondary structure at the 3′ end.[11] Indeed, the addition of another GC pair to the end of the stem structure of the first hairpin reduced the level of the aberrant transcript (Figure 1, Supporting Information).

NMR experiments were performed at proton frequencies of 500 or 600 MHz. The concentrations of RNA in the NMR samples were 0.2, 1.0, and 0.26 mM for the first hairpin, the second hairpin, and whole sar, respectively. The samples were boiled for 1 min and annealed to room temperature prior to acquiring spectra. The spectra were independent of the RNA concentration used during annealing.

^{15}N HMQC pulse sequences with selective water flip-back pulses[13] were used for HMQC experiments and long-range HMQC experiments. Multiple-bond couplings, which have small N−H coupling constants, were detected by optimizing the magnetization transfer delays to ∼25 ms. Three-dimensional NOESY−HMQC experiments,[13] with mixing times of 110 or 200 ms, were used to observe dipolar coupling between imino and amino protons.

The isolated first hairpin is heterogeneous in structure. The HMQC spectrum of hairpin I showed more peaks for G:C and A:U imino protons than predicted from the secondary structure (Figure 2, Supporting Information). For example, there are at least eight resonance peaks in the region of the spectrum that corresponds to imino protons in A−U basepairs, while only three A−U basepairs are predicted from the structure. This structural heterogeneity does not appear to involve the three U:U basepairs, as the number of U:U imino protons (six) is consistent with the proposed triple of U:U basepairs (see below).

The second hairpin adopts a single well-defined structure, either as an isolated fragment or in intact sar. Resonance assignments of the amino and imino protons of the isolated second hairpin were obtained from 2D ^{15}N HMQC and 3D NOESY−HMQC experiments, as well as from the detection of internuclear N−H···N hydrogen bonds.[14] All 11 expected hydrogen bonds in the stem of the hairpin were detected in long-range HMQC spectra (Figure 3,

10.1021/ja055675m CCC: $30.25

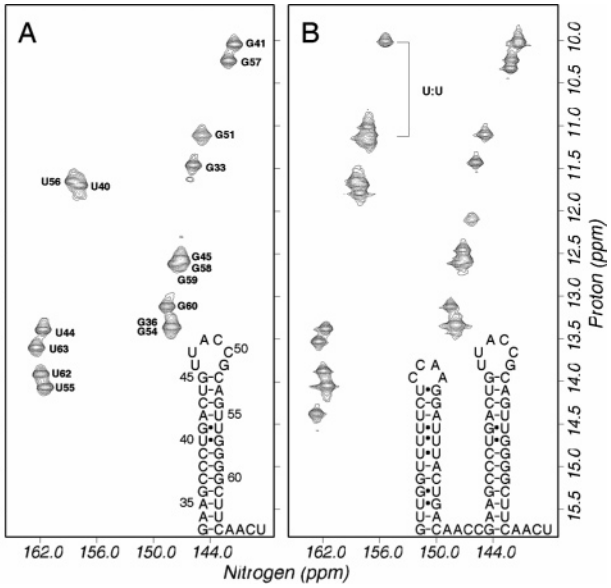


Figure 2. HSQC spectra of the imino region of the second hairpin of sar (A) and of whole sar (B). Resonance assignments for the isolated second hairpin are given in panel A. The "U:U" label in panel B indicates the location of U:U imino protons in the spectrum.

Supporting Information). The assignments of imino resonances are given in panel A of Figure 2. The observed NOE transfers and patterns of hydrogen bonds are consistent with the double-helix stem of hairpin II shown in Figure 1. The positions of resonance peaks from the second hairpin in whole sar do not differ significantly from their positions in the isolated hairpin, indicating that the structure of the second hairpin of sar does not depend on the presence of the nucleotides from the first hairpin. Observed NOE peaks in the NOESY spectra of intact sar are consistent with this conclusion.

The first hairpin adopts a single structure within the context of whole sar. The presence of hairpin II appears to force hairpin I into a single conformation, either by direct hairpin−hairpin interactions[4] or from excluded volume effects. The additional peaks that arise from imino protons in the first hairpin (Figure 2, panel B), 10 U imino protons (2 AU, 2 GU, 6 UU) and 4 G imino protons (2 GC, 2 GU), indicate that most of the duplex region of the first hairpin is stable in solution. However, the number of peaks from the first hairpin is less than expected from the proposed structure shown in Figure 1. The single missing AU, GC, and UG imino resonance lines suggest that the ends of the hairpin I are less protected from solvent exchange than the rest of the hairpin, in agreement with RNase digestion studies.[4]

The existence of the three U:U basepairs in intact sar was further substantiated by integration of the region of the spectrum containing the U:U imino resonances. This region of the spectrum integrates to six protons (Figure 4, Supporting Information). The U:U imino protons are also in close proximity to each other, as shown by the presence of cross-peaks between the imino protons in the NOESY spectra (see Figure 3).

Our results indicate that the model of sar presented in Figure 1 is an excellent representation of the structure of sar in solution. Remarkably, the triple of U:U basepairs appears to be stable in both the isolated first hairpin and in intact sar. The only other known occurrence of a triple U:U basepair is in telomerase RNA.[7] In the

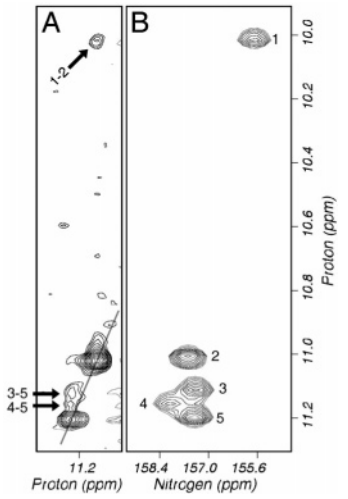


Figure 3. Dipolar coupling between U:U imino protons. A slice from a 3D NOESY spectrum, at $\delta_N = 157.06$ ppm, is shown in panel A (mixing time of 200 ms). Panel B shows the corresponding region from a $^1H-^{15}N$ HSQC spectrum. The peaks in this spectrum are arbitrarily labeled 1−5; these peaks arise from six protons. In the NOESY spectrum (A), three imino−imino cross-peaks are indicated by labeled arrows.

case of the telomerase RNA, the 5′ and 3′ basepairs that flank the triple U:U stack are C:G and C:U, respectively. While in the case of sar, the corresponding basepairs are both U:A. The different flanking bases in these two RNAs suggest that the triple U:U stack may be stable in a number of contexts, and further studies are warranted to investigate the stability of this unique structural motif.

Acknowledgment. We thank Dr. W. McClure for advice regarding optimization of in vitro transcription. The T7 RNA polymerase expression system was provided by Dr. F. W. Studier. Supported by a grant from the American Heart Association.

Supporting Information Available: A description of the transcription reactions, conditions used for mass spectrometry, a gel of the aberrant transcription product for hairpin I, HSQC spectrum of hairpin I, long-range HSQC spectrum of hairpin II, and integration of the U:U HSQC peaks are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lee, A. J.; Crothers, D. M. *Structure* **1998**, *6*, 993−1005.
(2) Wagner, D. G.; Simons, R. W. *Annu. Rev. Microbiol.* **1994**, *48*, 713−742.
(3) Liao, S. M.; Wu, T. H.; Chiang, C. H.; Susskind, M. M.; McClure, W. R. *Genes Dev.* **1987**, *1*, 197−203.
(4) Schaefer, K. L.; McClure, W. R. *RNA* **1997**, *3*, 141−156.
(5) Schaefer, K. L.; McClure, W. R. *RNA*, **1997**, *3*, 157−174.
(6) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner D. H. *J. Mol. Biol*. **1999**, *288*, 911−940.
(7) Theimer, C. A.; Finger, L. D.; Trantirek, L.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 449−454.
(8) Milligan, J. F.; Uhlenbeck, O. C. *Methods Enzymol.* **1989**, *180*, 51−62.
(9) Davanloo, P.; Rosenberg, A.; Dunn, J.; Studier, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 2035−2039.
(10) Ball, R. W.; Packman, L. C. *Anal. Biochem.* **1997**, *246*, 185−194.
(11) Triana-Alonso, F. J.; Dabrowski, M.; Wadzack, J.; Nierhaus, K. H. *J. Biol. Chem.* **1995**, *270*, 6298−6307.
(12) Macdonald, L. E.; Zhou, Y.; McAllister, W. T. *J. Mol. Biol*. **1993** *232*, 1030−1047.
(13) Gruschus, J. M.; Ferretti, J. A. *J. Magn. Reson.* **1998**, *135*, 87−92.
(14) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293−8297.

JA055675M