Nucleic Acids Research, 1993, Vol. 21, No. 23 5485–5488

NMR evidence for the RNA stem-loop structure involved in the transcription attenuation of *E.coli trp* operon

Vasudevan Ramesh
Department of Biochemistry and Biological NMR Centre, PO Box 138, Medical Sciences Building, University of Leicester, University Road, Leicester LE1 9HN, UK

Received June 22, 1993; Revised and Accepted October 8, 1993

ABSTRACT

High field ^{1}H-NMR studies of a synthetic 21-mer RNA fragment, corresponding to residues +114 to +134 within the *trp* leader mRNA transcript, have been carried out. Seven well resolved imino proton resonances corresponding to six C-G and one A-U hydrogen bonded base pairs, together with their characteristic NOE patterns can be identified in the NMR spectrum. This experimental result provides direct evidence for the postulated stem-loop secondary structure, 3:4, which has been reported to act as a transcription termination signal for RNA polymerase.

INTRODUCTION

The *trp* operon of *E.coli* consists of a transcription regulatory region and five structural genes (*trp* E,D,C,B,A) encoding the enzymes that catalyse the final sequence of reactions in tryptophan biosynthesis (Figure 1) (1). The initiation of transcription is regulated by a tryptophan activated repressor protein at an operator site located within the promoter region of the operon and this mechanism is understood at atomic structural detail (2). The transcription termination site (or attenuator) is located within the 162 base pair transcribed leader region (*trp*L), preceding the first major structural gene (Figure 1) (3, 4). Near the 5′ end of the leader transcript there is a coding region for a short leader peptide which has two tryptophan residues occuring in tandem (Figure 2A). Tryptophan has a fairly low frequency of occurence in proteins so that it is very unusual for two tryptophans to occur side by side. Depending on the levels of charged $tRNA^{Trp}$, which in turn depends on the Trp concentration in the cell, translation of these regulatory codons will or will not proceed. This event determines whether transcription will terminate in the leader region, since translation of this coding region allows formation of a secondary structure in the mRNA that causes termination of transcription at the *trp* attenuator.

By analysis of the sequence of the nuclease resistant fragments (5) in conjunction with secondary structure prediction algorithms (6), it was proposed that *three* crucial secondary structures can form from *four* regions of the *trp* leader RNA transcript (3, 6, 7). Region 1 (residues +54 to +68), which includes the tandem Trp codons, can basepair with region 2 (residues +76 to +91) to form stem-loop 1:2 (Figure 2A). Region 2 (residues +74 to +85) can also basepair with region 3 (residues +108 to +119) to form stem-loop 2:3 (Figure 2B). Finally, region 3 (residues +114 to +121) can also basepair with region 4 (residues +126 to +134) to form stem-loop 3:4 (Fig 2A). The calculated free energies of formation of these three stem-loop structures are: −11.2 (1:2), −11.7 (2:3) and −20 (3:4), kcal mol^{-1} (3, 6), suggesting that these structures should be stable under physiological conditions. The possibility that region 2 can pair with region 3 as well with region 1 led to the suggestion that alternative stem-loop secondary structures in the *trp*L transcript might provide the termination (3:4) or antitermination (i.e. readthrough) (2:3) signals recognised by RNA polymerase (Figure 2A, 2B) (7). While this is an attractive hypothesis, there is as yet no direct physico-chemical experimental evidence to support it. Detailed mutagenesis work has been carried out on the *trp* leader transcript (reviewed in ref 4) whose results show good support for the above model; however, the proposed stem-loop structures have not been solved by either X-ray crystallography or NMR spectroscopy.

NMR spectroscopy can be successfully applied to determine the three dimensional structure of small RNA fragments (~40 nucleotides) (8). In particular, assignment of the low field imino proton resonances of individual base-pairs simultaneously defines the secondary structure of an RNA motif, such as stem-loop (9). In this report, we present NMR results on a synthetic 21-mer RNA fragment whose sequence (inset Figure 3A) corresponds to the above discussed 3:4 termination site (+114 to +134) and show that the fragment is endowed with a stable, well folded stem-loop secondary structure. We have chosen to study this RNA fragment because it corresponds to a region of the leader transcript that has been well characterised by biochemical and mutagenesis experiments. Further, the small molecular mass of the 21-mer RNA is well suited for structure determination by contemporary NMR techniques. In the future, we propose to take up the study of larger RNA fragments corresponding to 1:2 and 2:3 stem-loop structures (Figure 2A, 2B) to gain an understanding of the complex equilibria governing the transcription attenuation mechanism, in three dimensional structural terms.

MATERIALS AND METHODS

The 21-mer RNA, 5′-AGCCCGC(CUAAUGA)GCGGGCU-3′, was synthesised, including HPLC purification, by Oswel DNA Service, University of Edinburgh. The sample was heated to

85°C for two minutes and very slowly cooled down, prior to NMR measurements. Preliminary UV absorbance−melting curve experiments (10, 11) were carried out at 0.8 μM and 10 μM RNA concentrations by Dr Tom Brown and Dr Susanne Ebel at University of Edinburgh and the data suggests hairpin formation in solution. The NMR sample (0.45 ml) consisted of 1.2 mg of RNA, 100 mM NaCl, 0.1 mM EDTA, 20 mM PO_4'''(pH 6.8) and dissolved in 90% H_2O + 10% D_2O mixture. NMR spectra of samples containing no salt were also measured at a lower pH of 5.5.

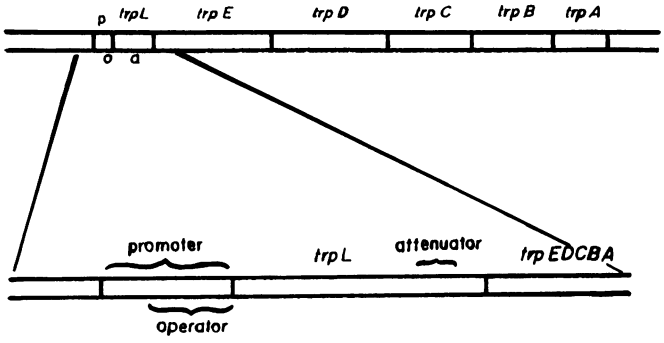

Figure 1. The *trp* operon of *E.coli*, showing the various regulatory and structural regions (adapted from reference 3).

^{1}H-NMR spectra were recorded using Bruker 600 MHz NMR spectrometer at several temperatures in the range of 280−294 K. 1D ^{1}H-NMR spectra were collected with a spectral width of 13200 Hz and is an average of 1000 transients. Phase sensitive NOESY spectra were measured (12) with a series of cross relaxation mixing times of 150, 250, 300 and 400 ms. In order to observe the exchange labile imino proton resonances with sufficient intensity, suppression of water protons in both 1D and 2D experiments was achieved using the 'jump-return' (90_x-t-90_{-x}, t = 50 μs) pulse sequence as the final detecting pulse (13). 256 time incremented FIDs were measured, each over 2 K data points and an average of 256 transients, before Fourier transformation to obtain the 2D spectrum. Baseline correction along only F2 dimension was carried out before plotting the spectrum.

RESULTS AND DISCUSSION

Figure 3A shows the low field region of the 1D spectrum of the 21-mer RNA, where resonances due to labile imino protons of hydrogen bonded Watson−Crick base pairs can be identified. Six well resolved resonances of similar intensity, *b*−*g*, can be observed in the 12.0−13.6 ppm range and peak *a* resonates at

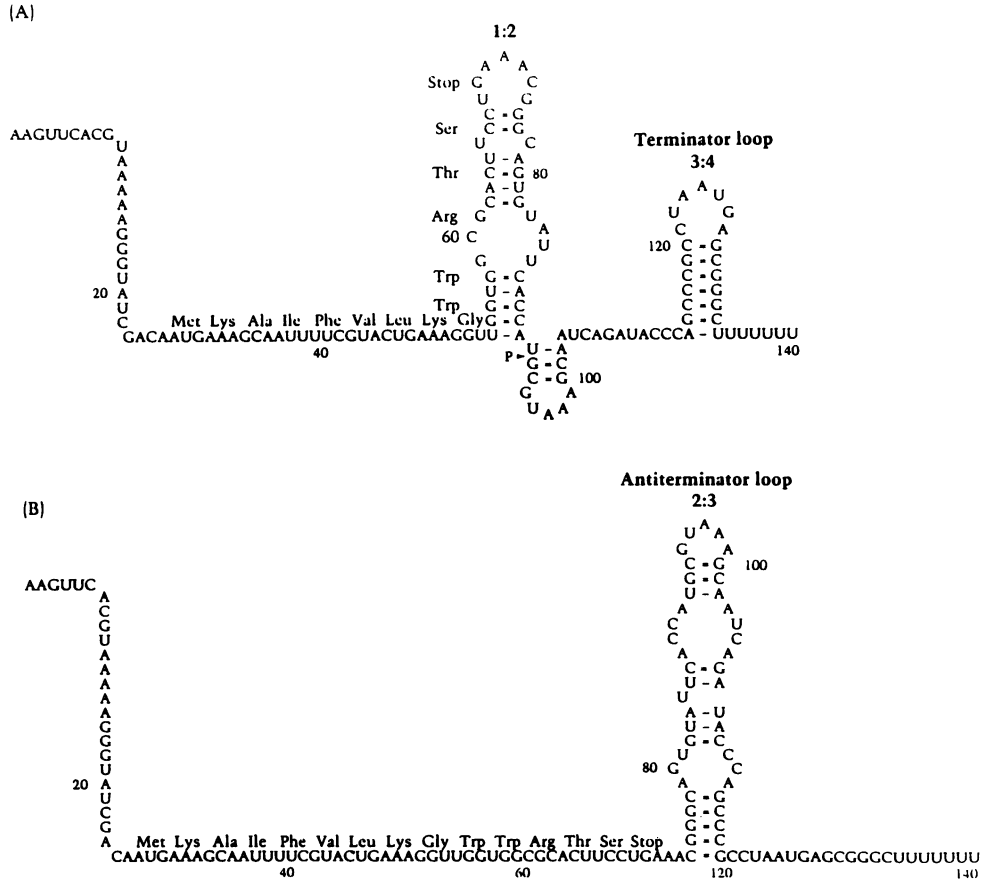

Figure 2. mRNA transcript of the *trp* leader segment of *E.coli* showing the stem-loop secondary structures involved in the regulation of transcription attenuation **A)** Terminator stem-loop 3:4 **B)** Antiterminator stem-loop 2:3 (adapted from reference 4).

a lower field of 14.07 ppm. It may be noted that imino proton resonances corresponding to the non hydrogen bonded loop nucleotides are not observed in the spectrum due to high pH (6.8) of the sample but they are identified at a lower pH of 5.5 as shown below.

The NOESY spectrum (Fig 3B) shows a number of connectivities from imino protons $b-g$ to highfield amino and aromatic protons of RNA. In contrast, imino proton a fails to show any NOE correlated crosspeaks. Close inspection of the spectrum reveals that the pattern of NOE correlations due to the imino protons $b-g$ are very similar, suggesting that they may belong to the same base pair type.

Previous NMR studies of RNA hairpin structures have shown that C-G base pairs in stem regions show the characteristic CG imino-C amino'-(chemical exchange)-C amino''-C_{H5} magnetisation transfer pathway (14, 8). Similarly, A-U base pairs of the stem show the characteristic AU imino-A_{H2} strong NOE correlations. Further, due to ring current shift effect of stacked base pairs, usually, the AU imino protons resonate downfield of CG iminos (15). These observations form the standard procedure that is used to distinguish CG and AU imino proton resonances in the spectrum of RNA. More recently, Heus and Pardi (16) have shown additional magnetisation transfer pathways for CG base pairs: CG imino-G amino'-G amino''-$C_{H1'}$, the last transfer being a weak one. In addition to the above intra base pair NOE connectivities, the imino protons also exhibit characteristic inter basepair sequential imino proton−imino proton NOE correlations which are used for sequence specific assignment purposes (8, 14).

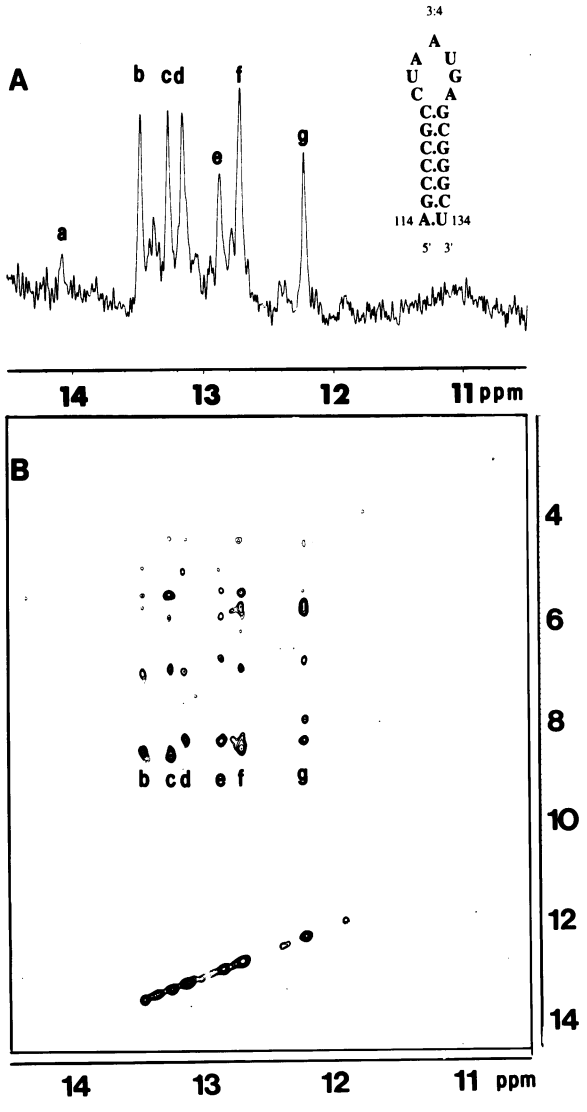

Figure 3. A) 1D ^{1}H-NMR spectrum (low field) of 21-mer RNA at pH 6.8, 280 K. B) Phase sensitive NOESY spectrum of the same sample, measured with a mixing time of 300 ms.

On the basis of the above methodology, analysis of the NOESY spectrum (Figure 3B) shows that resonances $b-g$ should belong to imino protons of CG base pairs while peak a can be assigned to the lone AU terminal base pair of the stem (see inset Figure 3A). Each imino proton ($b-g$) gives the characteristic strong NOE correlation to C amino' (~8.3−8.8 ppm) and C amino'' (~6.6−7.3 ppm) and a set of three crosspeaks (~4.4−6.2 ppm) corresponding to the $C_{H5,H1'}$ region. In addition, crosspeaks arising from correlation between the C amino geminal protons of each CG base pair can be separately identified in the upfield region of the spectrum (not shown). It is interesting to find that resonance g is the only imino proton to give NOE correlation to two amino protons at 8.0 and 8.3 ppm.

Although the NOESY spectrum measured with a mixing time of 300 ms shows strong intra base pair NOE crosspeaks, there are no sequential inter base pair imino proton−imino proton NOE connectivities in the spectrum (Figure 3B). The pH of the sample was then lowered from 6.8 to 5.5 to diminish the exchange rate of the imino protons with the solvent (i.e. water) and NOESY experiments were carried out at various mixing times (150, 250, 300, 400 ms), without success. 1D steady state NOE experiments with a irradiation time of 500 ms also failed to elicit NOEs between imino protons. This suggests that the base pairs of the stem may be undergoing rapid conformational exchange such that the dipolar interactions among the imino protons are not observed. However, as the sequential imino protons of A-RNA about 3.5−4.0 Å apart (15), NOEs between them will be intrinsically weaker and difficult to observe at lower RNA concentrations. Lack of NOE correlations among sequentially related base pairs in a helical stem of RNA is not uncommon (11, 17, 18). For example, in the Hammerhead RNA domain complex (18), one of the helical regions (helix III) that is rich in CG base pairs failed to show CG imino−CG imino proton NOE connectivities in the NOESY spectrum measured with a mixing time of 200 ms mixing time.

Decreasing the pH of the sample, however, resulted in the identification of imino proton resonances of non hydrogen bonded loop nucleotides in the characteristic lowfield region of the NMR spectrum at 11.1 ppm (Figure 4A). Further, two other resonances at 9.20 and 9.35 ppm can be observed in the highfield region of the same spectrum and these may be assigned to the amino protons of the loop nucleotides. Upon raising the temperature to 294 K, peak a of the AU base pair and the imino proton resonance of the loop nucleotides vanish from the spectrum due to rapid exchange with the solvent (Figure 4B). Except for a change in chemical shift of 0.07 ppm of resonance b, the chemical shift of resonances $c-f$ remain unchanged (Figure 4A,4B). The intensities of resonances $b-e$ remain largely unaffected suggesting that they may occupy interior positions of the stem. The intensity of peak f (12.71 ppm) is much reduced due to

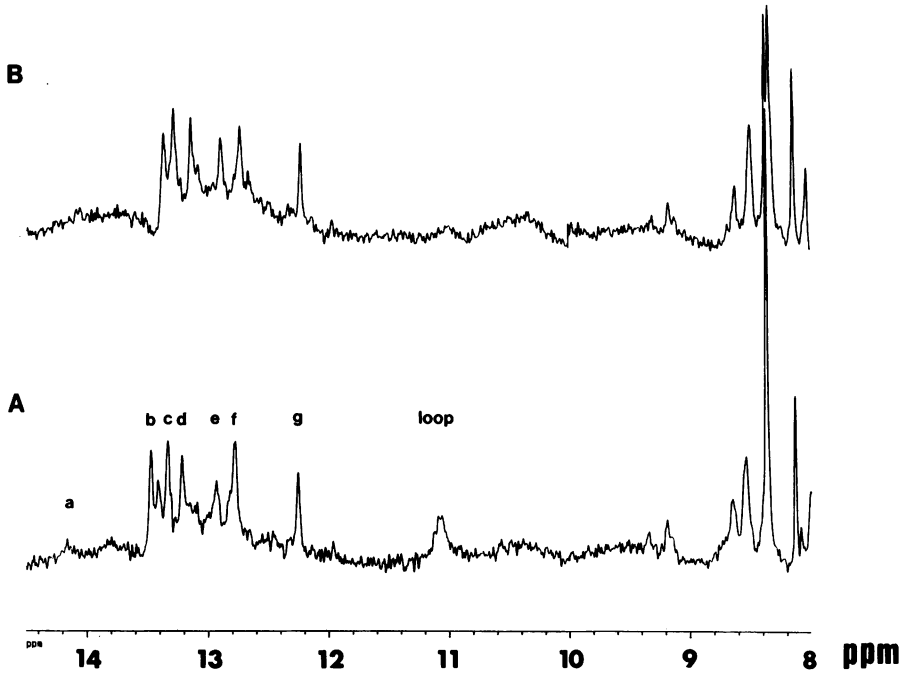

Figure 4. 1D ^{1}H-NMR spectra of 21-mer RNA at pH 5.5. **A)** 280K **B)** 294 K.

elevation of temperature presumably due to increased exposure to the solvent.

In summary, the NMR results described above provide strong evidence for the hitherto postulated 3:4 stem-loop as a stable, well folded structure. This significant observation should pave the way for more detailed studies on specific interactions of individual bases and changes in the phosphodiester backbone arising from such interactions . It is proposed to use heteronuclear (^{13}C,^{15}N,^{31}P) multi-dimensional NMR techniques (19) to fully characterise the 3D structure of this stem-loop and understand its role in the regulatory event.

ACKNOWLEDGEMENTS

I thank Professor G.C.K.Roberts, University of Leicester, for his valuable support and access to the NMR facilities. I also thank Dr Tom Brown and Dr Susanne Ebel (University of Edinburgh) for assistance with RNA sample preparation and UV melting curve experiments.

REFERENCES

1. Yanofsky,C. and Crawford,I.P. (1987) In Neidhardt,F. (ed.) *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, Vol 2. American Society for Microbiology, Washington, DC, Chapter 90, pp. 1453–1472.
2. Luisi,B.F. and Sigler,P.B. (1990) *Biochim. Biophys. Acta.* **1048**, 113–126.
3. Yanofsky,C. (1981) *Nature* **289**, 751–758.
4. Landick,R. and Yanofsky,C. (1987) In Neidhardt,F. (ed) *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, Vol 2. American Society for Microbiology, Washington, DC, Chapter 77, pp 1276–1301.
5. Lee,F. and Yanofsky,C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4365–4369.
6. Oxender,D.L., Zurawski,G. and Yanofsky,C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5524–5528.
7. Yanofsky,C. (1988) *J. Biol. Chem.* **263**, 609–612.
8. Varani,G. and Tinoco,I. (1991) *Quart. Rev. Biophys.* **24**, 479–532.
9. Draper,D.E. (1992) *Acc. Chem. Res.* **25**, 201–207.
10. Puglisi,J.D. and Tinoco,I. (1989) *Methods Enzymol.* **180**, 304–325.
11. Davis,P.W., Thurmes,W. and Tinoco,I. (1993) *Nucleic Acid. Res.* **21**, 537–545.
12. Williamson,M.P., Marion,D. and Wuthrich,K. (1984) *J. Mol. Biol.* **173**, 341–359.
13. Mateau,P. and Gueron,M. (1982) *J. Amer. Chem. Soc.* **104**, 7310–7311.
14. Puglisi,J.D., Wyatt,J.R. and Tinoco,I. (1990) *Biochemistry* **29**, 4215–4226.
15. Reid,B.R. (1981) *Annu. Rev. Biochem.* **50**, 969–996.
16. Heus,H.A. and Pardi,A. (1991) *J. Amer. Chem. Soc.* **113**, 4360–4361.
17. Wu,J. and Marshall,A.G. (1990) *Biochemistry* **29**, 1722–1730.
18. Heus,H.A. and Pardi,A. (1991) *J. Mol. Biol.* **217**, 113–124.
19. Nikonowicz,E.P. and Pardi,A. (1993) *J. Mol. Biol.* **232**, 1141–1156.