Characterization of RNA hairpin loop stability

Duncan R.Groebe and Olke C.Uhlenbeck

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

Received October 4, 1988; Revised and Accepted November 10, 1988

ABSTRACT

Fifteen RNA hairpins that share the same stem sequence and have homopolymer loops of A, C and U residues which vary in length from three to nine nucleotides were synthesized and their thermal stabilities determined. T_m varies as a function of loop size but is almost independent of loop composition. Loops of four or five nucleotides are found to be the most stable loop size. This is consistent with the observation that four-membered loops are the most prevalent loop size in 16S-like RNAs. The contribution of each loop to hairpin stability was calculated by subtracting the known contribution of the helical stem. These data should be useful for predicting the stability of other hairpins.

INTRODUCTION

Hairpins are the basic units of RNA secondary structure. Although the secondary structure of the 16S RNA can be divided into several large domains, about 70 percent of the molecule consists of 31 small hairpins that vary in sequence and size. A review of 21 16S-like RNA secondary structures from a number of organisms (1) reveals that 268 of the 590 phylogenetically proven RNA hairpins have four nucleotides in the loop. 66 hairpins with eight nucleotides in the loop are the second most common class. Previous thermodynamic studies have determined that the most stable RNA hairpin occurred when six nucleotides were present in the loop (2, 3). These results were consistent with the prevalence of seven-membered loops in the secondary structure of tRNA. Furthermore, Haasnoot *et al.* (4) have pointed out that the physical constraints of the RNA helix suggest that a loop size of seven residues should show maximal stability. Based on these studies, it has generally been thought that hairpins with these somewhat larger loops are the most stable.

This apparent contradiction between the phylogenetic data on ribosomal RNA and the limited physical data on RNA hairpins prompted this study of the effect of loop size and sequence on hairpin stability.

MATERIALS AND METHODS

Oligonucleotide Synthesis

RNA hairpins were synthesized by *in vitro* transcription of synthetic DNA by T7 RNA polymerase (5). A typical 10 ml transcription reaction contained 50 nM DNA template, 1 mM each ribonucleoside triphosphate, 1 mM spermidine, 10 mM MgCl₂, 5 mM dithiothreitol, 0.01% (v/v) Triton X-100, 80 mg/ml polyethylene glycol (8000 MW), 40 mM Tris-HCl (pH 8.1 at 37°C) and 20 µg/ml T7 RNA polymerase. After incubation for four hours at 37°C, the reactions were phenol extracted and ethanol precipitated. The RNA products were purified on a denaturing 20% polyacrylamide gel and visualized by UV-shadowing. In all but two cases (see results) two major transcription products of the approximate full length were obtained. In order to identify which product was the correct one, all ten RNAs with C and U residues in the loop were prepared with [α -³²P] ATP and each of the product bands subjected to total digestion with a mixture of RNase A, RNase T₁ and RNase T₂ (5). The digestion products were identified by two-dimensional thin layer chromatography (6). Since the C-loop and U-loop transcripts contain a unique 3'-terminal CpA sequence, only the full-length [α -³²P]ATP-labeled transcript and longer products should release a radiolabeled Cp upon RNase digestion. The shortest RNA transcript showing the presence of radioactive Cp was presumed to be the full-length product. In all cases, this was the shorter and most abundant of the major transcripts in the transcription reaction. The A-loop transcripts were presumed to be the correct sequence by comparison of their mobility with the C-loop and U-loop transcripts on a 20% denaturing polyacrylamide gel.

The correct transcription products were excised from the preparative gels, electroeluted and ethanol precipitated. From 50 to 100 nmoles of each RNA were recovered for thermal denaturation studies. The purity of each RNA was confirmed on a denaturing 20% polyacrylamide gel.

As a further test of the purity of the samples, 5 pmoles of each RNA was incubated in a 5 µl reaction with 0.1 µg of RNase T1 in 10 mM Tris-HCl, 1 mM EDTA pH 8 at 55°C for 2 hours. The digestion products were [5'-³²P] labeled in a 10 µl reaction containing 20 pmoles [γ -³²P]ATP, 50 mM Hepes pH 8.3, 20 mM MgCl₂, 3 mM DTT and 1 unit T4 polynucleotide kinase. After 37°C for one hour, the labeled RNA fragments were then separated on a 20% denaturing polyacrylamide gel and their location determined by autoradiography.

Size and Conformation Determination

RNA hairpins can form duplex species in equilibrium with the monomolecular conformation. HPLC sieving column analysis of a thawed sample of several hairpins in 1 M NaCl, 10 mM Na₂HPO₄, pH 7.0 and 0.1 mM EDTA reveals the presence of

several conformational species of higher apparent molecular weight (7). However, heating to 99°C for five minutes in the thermal denaturation buffer results in the formation of only the monomer hairpin conformation for the RNAs in this study.

Thermal Denaturation

All thermal denaturation profiles were measured at 260 nm in 10 mM Na₂HPO₄, 0.1 mM EDTA, pH 7.0 with and without 1 M NaCl using the procedure and apparatus described previously (7). At 1 M NaCl, at least six thermal denaturation profiles between 1 and 10 μ M oligonucleotide were determined for each RNA hairpin. In each case, the profiles were identical which is consistent with the expected intramolecular helix-coil transition. Degradation of the RNA samples as a result of the thermal denaturation process could not be detected by gel electrophoresis. The data from each melting profile was analyzed using a two-state model to derive the thermodynamic parameters for the helix-coil transition (7, 8). In each case, there is excellent agreement between the melting data and a curve generated from the six parameters used to fit the data.

On the basis of reproducibility, the estimated error limits are $\pm 5\%$ for ΔH and ΔS and $\pm 2\%$ for ΔG . Additional significant figures are given in the data to allow accurate calculation.

RESULTS

Fifteen RNA hairpins were prepared which share the same stem sequence and have homopolymer loop of A, C and U residues which vary in size from three to nine nucleotides (Figure 1). Loops composed only of G residues were not considered due to the potential for aggregation of oligonucleotides containing G (9, 10). Although the range of loop sizes and the G-C content of the stem is similar to that found in the 16S-like RNAs (1), the sequences of these molecules do not represent any particular RNA sequence and are not associated with any known physiological function. Since

FIGURE 1: Secondary structure of the fifteen RNA hairpins. Hairpins are named according to composition and size of loop. Thus C5 is the hairpin with 5 cytidines in the loop.

FIGURE 2: (A.) Transcription reactions of A9, C9 and U9 labeled with [α - ^{32}P] ATP. The arrows indicate the A9 and U9 transcripts. (B.) Analysis of [$5'$ - ^{32}P] labeled RNase T1 digestion products of several hairpins.

all the molecules share the same stem sequence, any differences in the stability of the RNA hairpins will be due only to differences in the sequence of the loop.

RNA transcription from synthetic DNA templates using T7 RNA polymerase (5) was used to synthesize the RNA hairpins for thermal denaturation studies. However, the transcription pattern for the A9 and U9 templates differ substantially from the other 13 templates. In a typical transcription reaction such as C9, a strong doublet is obtained corresponding to the full-length product and a transcript one nucleotide longer (Figure 2A). However, both the A9 and U9 transcription reactions produce a number of products longer than the presumed full-length transcript. In addition, a low yield of the full-length U9 product was obtained due to substantial termination by the T7 RNA polymerase at the loop region of the U9 DNA template (Figure 2A).

FIGURE 3: T_m versus loop size in 10 mM Na_2HPO_4 0.1 mM EDTA, pH 7 with (above) and without (below) 1.0 M NaCl (Δ) A-loops (O) C-loops; (∇) U-loops.

The purity of the RNA hairpins isolated for thermal denaturation studies was further assessed by digestion with RNase T1 and 5'-phosphorylation with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and T4 polynucleotide kinase. Only two large radiolabeled fragments are expected from this procedure, pAUAC(N_x)Gp from the loop region and pUAUCCA from the 3'-terminus of the RNA hairpin. As shown in Figure 2B, the A3-A7 and C9 molecules produced the expected two major fragments upon RNase T1 digestion. However, both A9 and U9 produced doublets of the two RNase T1 fragments. Based on the relative length of these fragments and the observation that the undigested samples of A9 and U9 migrate as a single band on a denaturing polyacrylamide gel, we conclude that the A9 and U9 samples are a mixture of a full-length N8 transcript and an N9 transcript lacking the 3'-terminal A residue. This heterogeneity is not observed in any of the other RNA preparations.

Table 1: Thermodynamics of Hairpin Formation at 1 M NaCl.

Hairpin	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	T_m (°C)	$-\Delta G^\circ_{37}$ (kcal/mol)
A3	56.2	160.9	76.4	6.3
A4	63.8	181.2	78.9	7.6
A5	64.6	183.4	79.1	7.7
A7	61.5	175.1	78.1	7.2
A9	52.6	151.7	73.3	5.6
C3	40.0	114.0	77.7	4.6
C4	38.2	109.3	77.2	4.3
C5	35.8	102.2	77.1	4.1
C7	26.7	76.7	75.4	2.9
C9	22.2	64.8	69.4	2.1
U3	60.1	171.1	78.1	7.0
U4	67.4	191.0	79.7	8.2
U5	67.4	191.1	79.5	8.1
U7	66.0	189.0	76.1	7.4
U9	73.8	212.0	74.3	8.0

The stability of the RNA hairpins vary as a function of both loop size and loop sequence. A plot of T_m versus loop size for all the RNA hairpins at two ionic strengths (Figure 3) reveals that for every loop sequence, the smaller loops are more stable than the larger loops. This situation is more pronounced at low ionic strength than high ionic strength. At both ionic strengths, loops of A and U residues are very similar in stability and C-loops are somewhat less stable.

Table 1 lists the thermodynamic parameters for hairpin formation at 1 M NaCl as determined by averaging the fitted thermodynamic values from five or more melting profiles of each hairpin. The largest standard deviation among 13 of the 15 molecules is ± 2.8 kcal/mol in ΔH° and ± 8.1 e.u. in ΔS° and is similar to what has been obtained previously (7, 8). The ΔH° and ΔS° values for A9 and U9 display a larger standard deviation than the other loops ($>10\%$), but this is probably a consequence of the heterogeneity of these samples. The values in Table 1 are consistent with the trends seen in the plot of T_m versus loop size (Figure 3) except that U9 has a higher ΔG°_{37} than expected. The thermodynamic parameters for hairpin formation in 10 mM Na^+ are listed in Table 2.

The overall ΔG of the RNA hairpins can be considered as the sum of the individual contributions of the five nearest-neighbor stacking interactions, the terminal G•A mismatch and the loop residues. The contribution of the 5'-triphosphate to RNA secondary structure stability is known to be small (11) and will be ignored.

Table 2: Thermodynamics of Hairpin Formation at 10 mM Na⁺

Hairpin	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	T_m (°C)	$-\Delta G^\circ_{37}$ (kcal/mol)
A3	61.9	184.2	64.2	4.8
A4	60.8	180.3	64.3	4.9
A5	59.4	176.7	62.8	4.6
A7	57.9	174.7	58.4	3.7
A9	43.1	132.7	51.8	1.9
C3	48.7	145.6	61.3	3.5
C4	48.2	144.1	61.3	3.5
C5	48.7	146.7	58.9	3.2
C7	34.2	104.3	54.8	1.9
C9	34.6	108.0	47.4	1.1
U3	59.9	177.4	64.5	4.9
U4	63.4	190.6	64.3	4.3
U5	61.6	183.3	63.1	4.7
U7	64.6	195.7	57.1	3.9
U9	46.0	140.6	54.3	2.4

By subtracting the ΔG values for the five nearest-neighbor stacking interactions and terminal G•A mismatch deduced from the study of RNA duplexes (12, 13) we can calculate the contribution of each loop to the ΔG of the RNA hairpin (Table 3). For all the RNA hairpin loops, the free energy calculated at 37°C is destabilizing. Since the thermodynamic parameters for the nearest-neighbor stacking interactions as a function of ionic strength are not available, the contribution of the loops to the stability of the RNA hairpins at 10 mM Na⁺ could not be assessed.

DISCUSSION

T7 transcription from synthetic DNA templates (5) has proven to be a versatile method for synthesizing oligoribonucleotides for biochemical and physical studies. Despite the template-independent addition of an extra nucleotide to the 3' terminus of some molecules, sufficient amounts of the correct full-length transcripts can be produced from preparative synthesis of the RNA. However, transcription of the A9 and U9 DNA templates result in the production of substantial amounts of transcripts longer than the expected doublet. RNase T1 digestion of purified samples of A9 and U9 reveal that these samples are mixtures of RNA hairpins differing in the length of both the loop and the 3'-terminus. It appears that the T7 RNA polymerase slips at stretches of nine A or T residues in a DNA template during the transcription process. This slippage is template-dependent as no slippage is detected with the other 13 DNA templates.

Table 3: Thermodynamic Contribution of Loops to Hairpin Stability at 1M NaCl

Loop Length	$-\Delta G_{37}^{\circ}(\text{kcal/mol})$		
	A	C	U
3	-4.9	-6.6	-4.2
4	-3.6	-6.9	-3.0
5	-3.5	-7.1	-3.1
7	-4.0	-8.3	-3.8
9	-5.6	-9.1	-3.2

A possible mechanism for the slippage of T7 RNA polymerase may involve pausing at the junction between the A-U rich and the second G-C rich regions of the DNA template (14). A realignment of the RNA transcript on the DNA template can occur during the pause and resumption of the transcription results in some transcripts having loop sizes shorter or longer than specified by the DNA template. The heterogeneity of loop sizes detected in the purified A9 and U9 transcripts is then a result of template-independent addition of a nucleotide to the 3'-terminus of those molecules which have lost a nucleotide in the loop. The pausing of T7 RNA polymerase would also explain the substantial number of aborted products in the U9 transcription due to termination at the stretch of nine A residues in the DNA template.

Regardless of loop sequence, smaller loop sizes (3 to 5 nucleotides) are more stable than larger loop sizes (7 and 9 nucleotides). These data differ from previous experiments using loops with C residues which showed that six membered loops were clearly more stable than other loop sizes (2, 3). The disagreement may be a result of the very different composition of the stems in the two cases. In this study, we used stems with a G•C base pair at each end and a 50 percent G•C content which results in hairpins which melt with T_m s between 70°C and 80°C. The previous hairpins had stems composed mostly of A•U base pairs and melted with T_m s between 10°C and 30°C. It is possible that either the structure of the AU rich helix or its tendency to fray could lead to the different optimal loop size. In addition, if single strand stacking in either the loop or the random coil is an important determinant in the maximal loop stability, then the very different T_m s may be the cause of the differing results. Single strand stacking is minimal at 80°C, but is still considerable at 30°C.

In a structural analysis of RNA hairpin loops Haasnoot *et al.* (4) have pointed out that the physical constraints of the RNA helix suggest two favored conformations for loop residues. The first would contain 7 or 8 loop residues and cross the major

groove of the helix in a 5' stack similar to what is found in the anticodon loop of tRNA (15, 16). The second would contain fewer residues and cross the minor groove. Although the phosphate-phosphate distance required to cross the minor groove is somewhat greater than to cross the major groove (17Å versus 10Å), the minor groove could still easily be spanned by three or even two nucleotides. Minor groove loops could also explain the stability of the three and four membered loops seen in this work. It is even possible that two-membered loops could occur in the secondary structure of large RNAs. In a previous study (3) an oligoribonucleotide that could potentially form a two-membered hairpin loop instead formed a duplex with two internal mismatches at all conditions tested. However, a two-membered DNA hairpin loop has been reported (17). Because of the propensity to dimerize, two-membered RNA loops would be expected to be difficult to form as isolated hairpins, but may occur within the structure of a larger RNA where the possibility for duplex formation would be inhibited by the presence of additional secondary and tertiary structure.

It is notable that the thermal stability of A and U-loops are very similar despite the ability of A residues to stack much better than U residues. The effect of loop sequence on stem stability has been difficult to predict. Stacking between residues in a loop has been shown to destabilize the stem (18, 19). However, loop residues adjacent to the stem could be considered "dangling" and stabilize the helix (20). Although it is possible that these two effects compensate for each other and thereby eliminate composition effects, we consider this unlikely for all sequences and loop sizes. It seems more likely that loop nucleotides adopt conformations that cannot be approximated by dangling nucleotides at the ends of bimolecular helical structures.

The C-loops are slightly less stable than the corresponding A and U-loops and have broader melting profiles. However, the hairpins containing C residues have the potential to form alternate structures during the denaturation process. Base pairing between the three 5'-terminal G residues and the C residues in the loop of C3-C9 would result in an alternate hairpin that may predominate at higher temperatures. This additional melting transition would explain the observed broader melting profiles for C-loops and, since longer loops would have more alternate structures, even explain the increased breadth with increasing chain length. Thus, we feel that the more positive thermodynamic values determined when C-loops are fit with the two state model in Tables 1 and 2 are incorrect. With a different stem sequence, we expect that C-loops would display similar stabilities to A and U-loops in an RNA hairpin.

The thermodynamic results suggest that composition can be ignored in determining the loop contribution to RNA hairpin loop stability. The data is consistent with the assumption of Freier *et al.* (12) that loop sequence does not

affect loop stability and roughly agree with the free energy values associated with different loop sizes that they derived from a limited amount of available data. However, the free energy values for the loops found here are somewhat lower than those estimated by Freier *et al.* (12). We believe the data in Table 3 to be more accurate due to the larger set of RNA hairpin loops used in this study. It is important to note that not all loop sequences are expected to have free energies similar to those found in this study. Presumably as a result of a particular structure, some loop sequences are exceptionally stable. A four-membered loop with a UUCG sequence is much more stable than the same hairpin loop with a UUUG loop sequence (21). Dimethylation of the A residues in the GGAA-loop of the 3'-terminal colicin fragment from *E. coli* 16S RNA destabilizes a hairpin loop (19). Interestingly, both the UUCG and GGAA loop sequences are conserved in certain hairpin loops of 16S-like RNAs. It will therefore be necessary to explore the effect of loop sequence on RNA hairpin loop stability in greater detail.

ACKNOWLEDGMENTS

We thank Mr. Jerry Haney for his assistance in this study. This work was supported by a grant from the National Institutes of Health (GM36944).

REFERENCES

1. Guttell, R.R., Weiser, B., Woese, C.R. and Noller, H.F. (1985) *Prog. Nucleic Acids Res. Mol. Biol.* **32**, 155.
2. Uhlenbeck, O.C., Borer, P.N., Dengler, B. and Tinoco, Jr., I. (1973) *J. Mol. Biol.* **73**, 483.
3. Gralla, J. and Crothers, D.M. (1973) *J. Mol. Biol.* **73**, 497.
4. Haasnoot, C.A.G., Hilbers, C.W., van der Marel, G.A. and van Boom, J.H. (1985) *J. Biosci.* **8**, 767.
5. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) *Nucleic Acids Res.* **15**, 8783.
6. Nishimura, S. (1979) In Schimmel, P.R., Söll, D. and Abelson, J. (eds), *Transfer RNA: Structures, Properties and Recognition*, Cold Spring Harbor Laboratory, New York, pp. 551.
7. Groebe, D.R. and Uhlenbeck, O.C. (1988) *Biochemistry*, in press.
8. Petersheim, M. and Turner, D.H. (1983) *Biochemistry* **22**, 256.
9. Jaskunas, S.R., Cantor, C.R. and Tinoco, Jr., I. (1968) *Biochemistry* **7**, 3164.
10. Studencki, A. (1981) Ph.D. Thesis, University of Illinois.
11. Freier, S.M., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) *J. Biomol. Struct. Dyn.* **1**, 1229.
12. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9397.
13. Hickey, D.R. and Turner, D.H. (1985) *Biochemistry* **24**, 3987.
14. Gilbert, W., Maizels, N. and Maxam, A. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 845.
15. Kim, S.H., Sussman, J.C., Suddath, F.L., Quigley, G.J., McPherson, A. and Rich, A.V. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4990.

16. Fuller, W. and Hodgson, A. (1967) *Nature* **215**, 817.
17. Orbons, L.P.M., van der Marel, G.A., van Boom, J.H. and Altona, C. (1986) *Nucleic Acids Res.* **10**, 4187.
18. van Charldorp, R., Verhoeven, J.J., van Knippenberg, P.H., Haasnoot, C.A.G. and Hilbers, C.W. (1982) *Nucleic Acids Res.* **10**, 4237.
19. van Charldorp, R., Heus, H.A., van Knippenberg, P.H., Jourdan, J., deBruin, S.H. and Hilbers, C.W. (1981) *Nucleic Acids Res.* **9**, 4413.
20. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) *Biochemistry* **26**, 4554.
21. Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Guild, N., Stormo, G., Gayle, M., d'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, Jr., I., Brody, E.N. and Gold, L. (1987) *Proc. Natl. Acad. Sci. USA* **85**, 1364.