RNA hairpin loop stability depends on closing base pair

Martin J. Serra, Matthew H. Lytle^{1,*}, Theresa J. Axenson, Calvin A. Schadt and Douglas H. Turner^{2,*}

Department of Chemistry, Allegheny College, Meadville, PA 16335, ¹Millipore Corporation, Bedford, MA 01730 and ²Department of Chemistry, University of Rochester, Rochester, NY 14627-0216, USA

Received March 29, 1993; Revised and Accepted June 12, 1993

ABSTRACT

Thermodynamic parameters are reported for hairpin formation in 1 M NaCl by RNA sequences of the type GGXAUAUAYCC, where X and Y are CG, GC, AU, UA, GU, or UG. A nearest neighbor analysis of the data indicates the free energy change for loop formation at 37°C, $\Delta^\circ G_{i,37}$, averages 3.4 kcal/mol for hairpin loops closed with C·G, G·C, and G·U pairs. In contrast, $\Delta^\circ G_{i,37}$ averages 4.6 kcal/mol for loops closed with A·U, U·A, or U·G pairs. Thus the stability of an RNA hairpin depends on the closing base pair. The hairpin with a GA mismatch that is formed by GGCGUAAUAGCC is more stable than the corresponding hairpin with an AA mismatch. Thus hairpin stability also depends on loop sequence. These effects are not included in current algorithms for prediction of RNA structure from sequence.

INTRODUCTION

The ability to predict RNA secondary structure from sequence based on thermodynamics has improved markedly (1,2). To fully utilize the rapid accumulation of nucleic acid sequence information, however, further refinements in structure prediction are necessary. The stabilities of fully base paired duplex regions of RNA have been extensively investigated (3). RNAs contain a wide variety of single stranded regions, however: bulges, internal loops, hairpins and multibranch loops. The influence of sequence on the stability of these secondary structural motifs is only beginning to be explored.

One prominent RNA secondary structure is the hairpin. Nearly 70% of the Escherichia coli 16S rRNA is found in small hairpin structures. An examination of these hairpins reveals that 85% are closed by either a G·C or C·G base pair (4). Due to a lack of data, current structure prediction models consider the stability of RNA hairpins to depend only upon the number of bases in the loop (3,5) or upon the number of bases in the loop and the stacking of the first mismatch in the loop (1). In earlier models, the stability of a hairpin was dependent on the closing base pair (6). This was based on a comparison of hairpins formed by $A_4G_{n-1}CU_4$ (7) and $A_6C_nU_6$ (8), $n=4,5$. Here we present a

study of the folding stability of hairpins with the sequence GGXAUAUAYCC where X and Y are CG, GC, AU, UA, GU, or UG. The results support a model in which the stability of the hairpin is dependent on the closing base pair. This may provide a thermodynamic basis for the preference for G·C and C·G hairpin closures in naturally occurring RNAs.

MATERIALS AND METHODS

RNA synthesis and purification

Most oligomers were synthesized on solid support by the phosphoramidite approach with the 2' hydroxyl protected as the tert-butyl dimethylsilyl ether (9–11). GGUAUAAUAGCC was synthesized with the 2' hydroxyl protected as the tetrahydropyranyl ether (12). pppGGCAUAAUAGCC and pppGGCGUAAUAGCC were synthesized with the T7 RNA polymerase method of Milligan et al. (13). After ammonia deblocking of chemically synthesized oligomers, the crude oligomer was purified by SEP-PAK C18 (Waters) chromatography. After removal of the acid labile protecting group, the oligomer was purified by preparative thin-layer chromatography on a Baker Si500 F plate (n-propanol: ammonium hydroxide: water, 55:35:10). The pure oligomer was desalted by G-10 chromatography. The purities of the oligomers were checked by analytical C-8 HPLC (Beckman) and were determined to be greater than 95%. A typical HPLC profile is shown in Figure 1.

Melting curves

The buffer for the melting studies was 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na₂-EDTA, pH 7. Single stranded extinction coefficients were calculated from the extinction coefficients for dinucleoside monophosphates and nucleosides, as described previously (14,15). In units of 10⁴ M⁻¹ cm⁻¹, the calculated extinction coefficients at 280 nm are as follows: GGCAUAAUAGCC, 5.20; GGGUAUAAUACCC, 5.59; GGAAUAAUAUCC, 4.98; GGUAUAAUAAACC, 4.94; GGGUAUAAUAUCC, 5.36; GGUAUAAUAGCC, 5.15; AGG-AAUAAUAUCCU, 5.53; AGGUAAUAAUAGCCU, 5.69; GGCGUAAUAGCC, 5.46. Strand concentrations were determined from absorbance at 280 nm at 90°C. Absorbance

* To whom correspondence should be addressed

+ Present address: Terrapin Technologies, 750H Gateway Blvd., South San Francisco, CA 94080, USA

Figure 1. Reverse-phase HPLC profile. Typical elution profile for pure oligomers using a 4.6 mm×25 cm Ultrasphere Octyl (Beckman) reverse-phase column. The flow rate was 1 mL/min. Elution was with 0.1 M NaH₂PO₄ with a linear gradient of 0–20% methanol.

versus temperature melting curves were measured at 280 nm with a heating rate of 1.0°C min⁻¹, on a Gilford 250 spectrophotometer as described previously (16). Oligomer concentrations were in the 10⁻⁶–10⁻⁴ M range and were varied over at least a 20-fold range for each sequence.

Data analysis

Absorbance versus temperature profiles were fit to a two-state unimolecular model with sloping baselines by using a nonlinear least-squares program (16,17). Thermodynamic parameters for hairpin formation were obtained from averages of the fits of 4 melting curves for sequences with a 5' triphosphate and at least 5 melting curves for all other sequences. Melting temperatures and thermodynamic parameters were concentration independent as expected for unimolecular hairpin formation.

RESULTS

Nine RNA hairpins were prepared that share the AUAUA loop but vary by the identity of the base pair closing the loop or the length of the stem (Table I). One hairpin closed with C·G had a 5' triphosphate at the 5' side of the stem since it was prepared with T7 RNA polymerase (13). This is expected to make the ΔG_{37}° for hairpin folding more favorable by about 0.3 kcal/mol (18). Six of the molecules share the same sequence except the base pair closing the loop; therefore, differences in stability of these hairpins are due to the effects of the closing base pair. In addition, two hairpins were prepared that have a GUAAUA loop that is commonly found in ribosomal RNA (19).

Typical melting curves are shown in Figure 2, and the measured thermodynamic parameters are listed in Table I. Also listed in Table I are ΔH° 's, ΔS° 's and ΔG_{37}° 's predicted for the stems (3) and ΔG_{37}° 's predicted by the model of Jaeger et al. (1) in which stacking of the first mismatch in the loop on the closing base pair is a contributor to hairpin stability. For example, for the chemically synthesized hairpin having an AA mismatch next to the C·G closing base pair, ΔG_{37}° (stem) = ΔG_{37}° ($\frac{CG}{CC}$) + ΔG_{37}° ($\frac{GC}{CC}$) = -2.9 - 3.4 = -6.3 kcal/mol, and the predicted ΔG_{37}° (hairpin) = ΔG_{37}° (loop of 6) + ΔG_{37}° (stem) + ΔG_{37}°

Table I. Thermodynamic parameters for hairpin formation^a

RNA Hairpin	T _m °C	ΔH° meas. (pred. stem) ^b kcal/mol	ΔS° meas. (pred. stem) ^b e.u.	ΔG_{37}° meas. (pred. stem) ^b kcal/mol	pred. ^c ΔG_{37}° kcal/mol	exptl. ΔG_{37}° (loop formation) ^d kcal/mol
A U 5'GGU A ... CCA A A U	40.7	-25.9 ± 1.3 (-22.4) ^b	-82.4 ± 3.9 (-55.9) ^b	-0.34 ± 0.02 (-5.0) ^b	-0.9	4.7
A U 5'AGGA A ... UCCU A A U	57.0	-39.1 ± 1.6 (-33.1) ^b	-118.4 ± 5.3 (-84.4) ^b	-2.19 ± 0.08 (-6.9) ^b	-2.6	4.7
A U 5'GGA A ... CCU A A U	44.2	-29.2 ± 1.6 (-25.5) ^b	-91.8 ± 5.4 (-65.2) ^b	-0.73 ± 0.04 (-5.2) ^b	-0.9	4.5
A U 5'GGU A ... CCG A A U	42.8	-27.3 ± 0.1 (-21.7) ^b	-86.5 ± 0.5 (-53.6) ^b	-0.47 ± 0.02 (-5.0) ^b	-0.4	4.5
A U 5'AGGU A ... UCCG A A U	56.6	-35.4 ± 0.9 (-29.3) ^b	-107.0 ± 3.0 (-72.8) ^b	-2.21 ± 0.06 (-6.7) ^b	-2.1	4.5
A U pppGGC A ... CCG A A U	66.1	-34.8 ± 1.2 (-27.5) ^b	-102.7 ± 3.7 (-67.1) ^b	-2.99 ± 0.09 (-6.6) ^b	-2.9	3.6
A U 5'GGC A ... CCG A A U	64.0	-34.3 ± 1.0 (-26.4) ^b	-101.7 ± 3.2 (-64.6) ^b	-2.74 ± 0.13 (-6.3) ^b	-2.6	3.6
A U 5'GGG A ... CCC A A U	58.7	-33.5 ± 1.6 (-24.4) ^b	-101.0 ± 6.7 (-59.4) ^b	-2.17 ± 0.08 (-5.8) ^b	-1.8	3.6
A U 5'GGG A ... CCU A A U	50.8	-34.3 ± 2.7 (-18.5) ^b	-106.0 ± 8.2 (-45.5) ^b	-1.42 ± 0.18 (-4.3) ^b	0.0	2.9
G U pppGGC A ... CCG A A U	70.8	-37.5 ± 1.3 (-27.5) ^b	-109.1 ± 3.8 (-67.1) ^b	-3.68 ± 0.09 (-6.6) ^b	-2.9	2.9
G U 5'GGC A ... CCG A A U	68.7	-36.9 ± 2.4 (-26.4) ^b	-107.8 ± 7.2 (-64.6) ^b	-3.42 ± 0.24 (-6.3) ^b	-2.6	2.9

^aMeasurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na₂EDTA, pH 7. Errors in ΔH° , ΔS° , and ΔG° are standard deviations. ΔG_{37}° was calculated from ΔH° and ΔS° before rounding off.

^b ΔH° , ΔS° , and ΔG° for stems were predicted from parameters in S.M.Freier, R.Kierzek, J.A.Jaeger, N.Sugimoto, M.H.Caruthers, T.Neilson, and D.H.Turner, Proc. Natl. Acad. Sci. 83, 9373–9377 (1986) and L.He, R.Kierzek, J.SantaLucia, Jr., A.E.Walter, and D.H.Turner, Biochemistry 30, 11124–11132 (1991). The 5' triphosphates were treated as 5' monophosphates (S.M.Freier, D.Alkema, A.Sinclair, T.Neilson, and D.H.Turner, Biochemistry 24, 4533–4539 (1985)).
^cPredicted from model of J.A.Jaeger, D.H.Turner, and M.Zuker, Proc. Natl. Acad. Sci. U.S.A. 86, 7706–7710 (1989), with $\Delta G_{37}^\circ = -1.4$ kcal/mol for AA mismatch on C·G closing base pair (J.SantaLucia, Jr., R.Kierzek, and D.H.Turner, J. Am. Chem. Soc. 113, 4313–4322 (1991)).

^dCalculated as: $\Delta G_{i,37}^\circ = \Delta G_{37}^\circ$ (hairpin formation) - ΔG_{37}° (pred. stem).

($\frac{CA}{GA}$) = 5.1 - 6.3 - 1.4 = -2.6 kcal/mol. Since no measurements are available for ΔG_{37}° ($\frac{CA}{GA}$), it is assumed to be equal to the measured value of -1.4 kcal/mol for both $\frac{CG}{GA}$ and $\frac{CG}{GP}$, where P is purine (20).

The hairpin structure was dissected into its two secondary

Figure 2. Normalized melting curves. Sequences for highest to lowest melting temperature are: A) pppGGCAUAAUAGCC, AGGUAAUAAUAGCCU, GGUAAUAAUAGCC, and GGUAAUAAUACC. B) GGGAAUAAUACCC, AGGAAUAAUACCU, GGGAAUAAUAUCC, GGAUAAUUAUCC. Melting curves were normalized at 90°C. Melting buffer was 1 M NaCl, 10 mM cacodylate, 0.1 mM Na₂EDTA, pH 7.0.

structural motifs, the double helical stem and the loop of six bases. The free energy for loop formation at 37°C, $\Delta G_{l,37}^{\circ}$, can be calculated from (7): $\Delta G_{l,37}^{\circ} = \Delta G_{i,37}^{\circ}$ (measured for folding of

hairpin) $- \Delta G_{37}^{\circ}(\text{stem})$. For example, for GGUAAUAAUACC, $\Delta G_{l,37}^{\circ} = -0.34 - (-5.0) = 4.7$ kcal/mol. The results of this analysis are in Table I.

DISCUSSION

Current models for predicting the secondary structural stability of RNA employ many simplifying assumptions, particularly for loop regions (1,5). Improved methods for the synthesis of RNA molecules of defined sequence allow testing the validity of these assumptions (9–13). This work examines the effect of the closing base pair upon the stability of RNA hairpins with an AUAUA loop.

Free energy changes for AUAUA hairpin loop formation at 37°C, $\Delta G_{i,37}^\circ$, are listed in Table I. These values fall within two ranges, loops closed with G·C, C·G and G·U have an average $\Delta G_{i,37}^\circ$ of 3.4 kcal/mol. This is similar to the $\Delta G_{i,37}^\circ$ of 3.6 kcal/mol obtained by interpolation of the results of Groebe and Uhlenbeck (21) for A_n and U_n loops closed by C·G, where $n = 5$ and 7. Loops closed with A·U, U·A, and U·G have an average $\Delta G_{i,37}^\circ$ for loop formation of 4.6 kcal/mol. Evidently, loops closed by C·G, G·C, or G·U are about 1 kcal/mol more stable than loops closed by U·A, A·U, or U·G. The reliability of the thermodynamic values is seen by the similarities in $\Delta G_{i,37}^\circ$ for the pairs of hairpins that have identical base pair closures but different length stems; each of the values is within 0.2 kcal/mol of the other. This includes stems that are extended by either a base pair or a 5' triphosphate. (The 5' triphosphate was assumed to make the stem more favorable by 0.3 kcal/mol, the free energy increment measured for a 5 monophosphate (18).) This suggests the nearest neighbor model (6) is reasonable for hairpin stem-loop motifs. Antao and Tinoco (22) reached the same conclusion based on the dependence of hairpin stability on stem sequence.

No simple model explains all the sequence dependence observed for hairpin loop stability. The most popular current algorithm for prediction of RNA secondary structure from sequence (23) uses the model of Jaeger et al. (1) in which the sequence dependence of loops larger than 4 is attributed to stacking of the first mismatch on the closing base pair. Comparison of the predicted and measured $\Delta G_{i,37}^\circ$'s in Table I, however, indicates this model overestimates stabilities of loops closed by AU pairs and underestimates stabilities of loops starting with GA mismatches. Groebe and Uhlenbeck (21) suggested that $\Delta G_{i,37}^\circ$ for loop formation is independent of loop sequence. Comparison of loops with first mismatches of AA and GA (Table I) suggests this model is also oversimplified. Studies of RNA and DNA hairpin loops with 4 nucleotides also reveal a significant sequence dependence (22,24–26). Early models of loop initiation suggested that loops closed by GC pairs are 2 kcal/mol more stable than loops closed by AU pairs (6). The results in Table I support the dependence of loop closure on base pair, although the difference between GC and AU closure is only about 1 kcal/mol. The difference between loop closure by GC and AU could be due to the extra hydrogen bond in a GC pair (18,27,28). The loop closed by G·U, however, is more stable than the loops closed by G·C or C·G. Since G·U pairs have only 2 hydrogen bonds, the effect of closing base pair probably depends on more than the number of hydrogen bonds between the bases closing the loop. Evidently, it will be necessary to characterize several factors before it is possible to accurately predict stabilities of hairpin loops. It is likely that other types of loops will have similar complexity (20,29).

The data in Table I also provide insight into the temperature dependence of hairpin loop stability. In all cases, the ΔH° for hairpin formation is more favorable than the ΔH° predicted for

the stems based on the results of Freier et al. (3). The average difference is -8 kcal/mol. Thus, closure of these hairpins is associated with a favorable ΔH° . Of course, the measured ΔH° is derived from a two-state model. If hairpin formation is not a two-state process, then the true ΔH° 's will be even more favorable than those reported in Table I. This is because the presence of intermediates broadens the transition curve leading to a fitted ΔH° with a smaller magnitude than the real ΔH° (30). It is expected, however, that hairpins with only 3 base pairs in the stem will melt in a two-state manner. Favorable ΔH° 's of loop closure have also been observed for RNA and DNA hairpins with other lengths (21,22,26), although examples of unfavorable ΔH° 's have also been reported (8,26).

ACKNOWLEDGEMENTS

This research was supported by NIH Grant GM22939 (D.H.T.) and by the Pew and Dana Foundations.

REFERENCES

- Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) *Proc. Nat. Acad. Sci.* **86**, 7706–7710.
- Zuker, M., Jaeger, J.A. and Turner, D.H. (1991) *Nucleic Acids Res.* **19**, 2707–2714.
- 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**, 9373–9377.
- Woese, C.R., Winker, S. and Gutell, R.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8467–8471.
- Turner, D.H., Sugimoto, N. and Freier, S.M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167–192.
- Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, M.O., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.* **246**, 40–41.
- Gralla, J. and Crothers, D.M. (1973) *J. Mol. Biol.* **73**, 497–511.
- Uhlenbeck, O.C., Borer, P.N., Dengler, B. and Tinoco, I., Jr. (1973) *J. Mol. Biol.* **73**, 483–496.
- Usman, N., Ogilvie, K.K., Jiang, M.Y. and Cedergren, R.L. (1987) *J. Am. Chem. Soc.* **109**, 7845–7854.
- Wang, Y.Y., Lytle, M.H. and Borer, P.N. (1990) *Nucleic Acids Res.* **18**, 3347–3352.
- a) Lytle, M.H., Wright, P.B., Sinha, N.D., Bain, J.D. and Chamberlin, A.R. (1991) *J. Org. Chem.* **56**, 4608–4615. b) Lytle, M.H. (1993) *Nucleosides & Nucleotides* **12**, 95–106.
- Kierzek, R., Caruthers, M.H., Longfellow, C.E., Swinton, D., Turner, D.H. and Freier, S.M. (1986) *Biochemistry* **25**, 7840–7846.
- Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) *Nucleic Acids Res.* **15**, 8783–8798.
- Borer, P.N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, C.D., Ed.) 3rd ed., Vol I, p. 589, CRC, Cleveland, OH.
- Richards, E.G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, C.D., Ed.) 3rd ed., Vol I, p. 197, CRC, Cleveland, OH.
- Petersheim, M. and Turner, D.H. (1983) *Biochemistry* **22**, 256–263.
- SantaLucia, J., Jr., Kierzek, R. and Turner, D.H. (1992) *Science* **256**, 217–219.
- Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985) *Biochemistry*, **24**, 4533–4539.
- Gutell, R. R. and Fox, G. E. (1988) *Nucleic Acids Res.* **16** (supplement), r175–r269.
- SantaLucia, J., Jr., Kierzek, R. and Turner, D.H. (1991) *J. Am. Chem. Soc.* **113**, 4313–4322.
- Groebe, D.R. and Uhlenbeck, O.C. (1988) *Nucleic Acids Res.* **16**, 11725–11735.
- Antao, V.P. and Tinoco, I., Jr. (1992) *Nucleic Acids Res.* **20**, 819–824.
- Zuker, M. (1989) *Science* **244**, 48–52.
- Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stormo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, I., Jr.,

- Brody, E.N. and Gold, L. (1988) *Proc. Natl. Acad. Sci. USA* 85, 1364–1368.
25. Antao, V.P., Lai S.Y. and Tinoco, I., Jr. (1991) *Nucleic Acids Res.* 19, 5901–5905.
26. Senior, M.M., Jones, R.A. and Breslauer, K.J. (1988) *Proc. Natl. Acad. Sci.* 85 6242–6246.
27. Freier, S.M., Kierzek, R., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) *Biochemistry* 25, 3209–3213.
28. Turner, D.H., Sugimoto, N., Kierzek, R. and Dreiker, S.D. (1987) *J. Am. Chem. Soc.* 109, 3783–3785.
29. SantaLucia, J., Jr., Kierzek, R. and Turner, D.H. (1991) *Biochemistry* 30, 8242–8251.
30. Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical Chemistry Part III*, Chapter 21, W. H. Freeman, San Francisco, CA.