Nucleic Acids Research, 1996, Vol. 24, No. 14 **2841–2848**

Structural features and stability of an RNA triple helix in solution

Jason A. Holland and David W. Hoffman*

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA

Received January 10, 1996; Revised and Accepted May 24, 1996

ABSTRACT

A 30 nt RNA with a sequence designed to form an intramolecular triple helix was analyzed by one- and two-dimensional NMR spectroscopy and UV absorption measurements. NMR data show that the RNA contains seven pyrimidine–purine–pyrimidine base triples stabilized by Watson–Crick and Hoogsteen interactions. The temperature dependence of the imino proton resonances, as well as UV absorption data, indicate that the triple helix is highly stable at acidic pH, melting in a single sharp transition centered at 62°C at pH 4.3. The Watson–Crick and Hoogsteen pairings are disrupted simultaneously upon melting. The NMR data are consistent with a structural model where the Watson–Crick paired strands form an A-helix. Results of model building, guided by NMR data, suggest a possible hydrogen bond between the 2′ hydroxyl proton of the Hoogsteen strand and a phosphate oxygen of the purine strand. The structural model is discussed in terms of its ability to account for some of the differences in stability reported for RNA and DNA triple helices and provides insight into features that are likely to be important in the design of RNA binding compounds.

INTRODUCTION

Triple helix formation has been suggested as a means for sequence-specific binding of either DNA or RNA (1–6). A relatively short single strand of DNA or RNA can bind to double-stranded DNA, RNA or a DNA/RNA hybrid by forming Hoogsteen-type interactions in the major groove of the duplex. The antiparallel arrangement of one purine strand flanked by two pyrimidine strands (Fig. 1) is referred to as the 'pyrimidine·purine*pyrimidine' triplex motif, which in this case contains C–G–C and U–A–U triples. Molecules with the ability to recognize and bind a specific sequence of double-stranded nucleic acid are of particular interest, since they have the potential to be useful as probes of structure or possibly as inhibitory or therapeutic agents. As examples, triple helix-forming DNA oligonucleotides can compete with transcription factors for their DNA binding sites (3) and RNA oligonucleotides with EDTA on the 5′-end can sequence-specifically cleave duplex RNA or RNA/DNA hybrids (4). Triplex nucleic acids may also have functional roles in natural cellular processes. For example, DNA triplex formation may be involved in regulating the c-*myc* proto-oncogene (7). Previous structural studies have shown that RNA triples exist within transfer RNA (8,9), the *Tetrahymena* group I intron (10–12) and possibly within the human immunodeficiency virus TAR RNA when bound by arginine (13–15).

Recently several biophysical studies of the solution properties of RNA triple helices have been reported (4–6,16,17) and the relative stabilities of RNA, DNA and RNA/DNA hybrid triple helices have been investigated (4–6). Although there are some discrepancies between the various studies, in general it is found that R(DR), R(DD), R(RR) and R(DD) form the most stable triplexes, D(DD) and D(DR) are somewhat less stable and D(RR) and D(RD) are relatively unstable (the Watson–Crick paired strands are in parentheses). These studies have focused primarily on characterizing the relative stabilities of the possible DNA and RNA triplexes, but have not focused on determining the structural factors that actually account for the observed differences in stability. Goals of the work presented here include characterizing the structural features and stability of a pyrimidine–purine–pyrimidine RNA triplex in solution, analyzing the melting behavior of the triplex, identifying ribose conformations and potential hydrogen bonds on the three strands and proposing features that are most likely to contribute to RNA triplex stability. An understanding of the structural basis of the observed triplex stabilities will be helpful in understanding triplex formation at the molecular level and if nucleotide analogs are to be successfully designed to form specific triplexes with improved recognition and binding properties.

MATERIALS AND METHODS

In vitro transcription reactions using T7 RNA polymerase (18) were used to prepare a 30 nt RNA with the sequence 5′-GAGAAGAUUCGUCUUCUCUCUCUCUCUUCU-3′. The two Watson–Crick strands are linked by a UUCG 'tetraloop' sequence (19–21) and the Hoogsteen strand is linked to the Watson–Crick paired strands by a 5 nt pyrimidine loop. Oligonucleotides for the DNA template with sequences of 5′-AGAAGAGAGAGAGAAGAAGACGAATCTTCTCTATAGTGAGTCGTATTAC-3′ and 5′-TAATACGACTCACTAATAG-3′ were synthesized on an Applied Biosystems DNA synthesizer or obtained from Operon Inc. (California). Template DNA was purified by 20% polyacrylamide gel electrophoresis in 8 M urea.

* To whom correspondence should be addressed

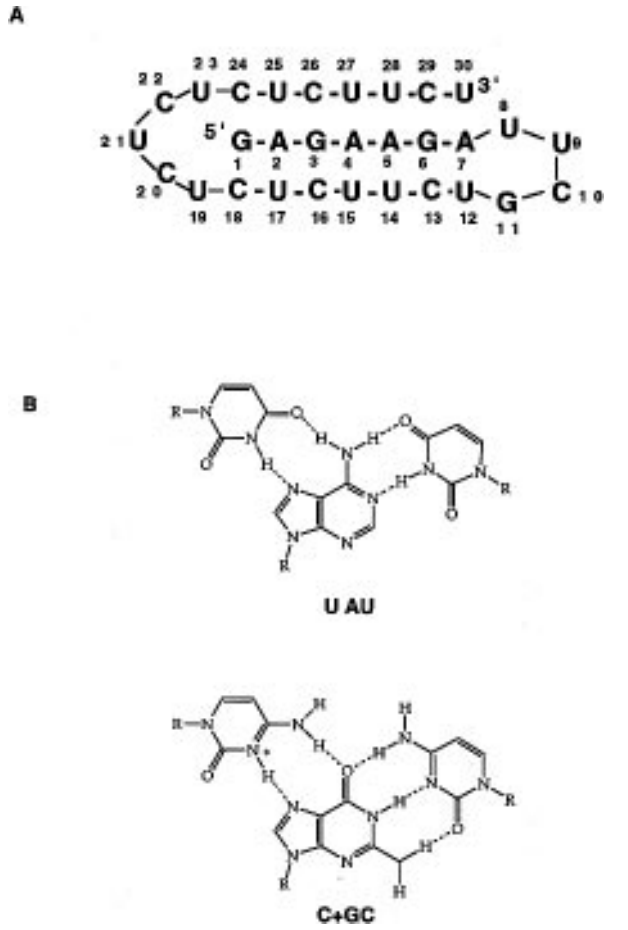

Figure 1. (**A**) Schematic diagram of the nucleotide sequence and topology of the RNA triple helix used in this work. The diagram is drawn to show the base pairing that occurs within the RNA. Nucleotides 1–7 form Watson–Crick pairs with nt 12–18 and nt 24–30 form Hoogsteen pairs with nt 1–7. (**B**) Schematic diagram of the hydrogen bonding patterns within the UAU and CGC triples.

T7 RNA polymerase was purified from an overproducing strain of *Escherichia coli* (pAR1219) provided by John Dunn. Small scale transcriptions were performed to optimize the yield of RNA triplex product, varying the concentrations of template DNA, T7 RNA polymerase, nucleotide triphosphates and magnesium. After optimization of conditions, large scale transcription reactions containing 2.5 mM nucleotide triphosphates, 23 mM magnesium, 730 nM template DNA were incubated at 37°C for 4–5 h, then ethanol precipitated and dissolved in 8 M urea. The RNA transcript was then separated from incorrect sized transcripts by 20% polyacrylamide gel electrophoresis under 8 M urea denaturing conditions. RNA was visualized by UV shadowing and then removed from the gel using a BioRad model 422 electroeluter. The RNA was further purified by ethanol precipitating several times, passed over a Sephadex G25 gel filtration column in 1 mM potassium monophosphate buffer and lyophilized. A typical yield was 1 mg purified RNA per 20 ml transcription reaction.

RNA melting was monitored by UV absorption and was performed using a Gilford System 2600 UV-VIS Spectrophotometer equipped with a Gilford Thermoprogrammer model 2527. The temperature was increased at a rate of 1.0°C/min, typically starting at 35°C and ending at 95°C. The melting behavior of the RNA did not differ significantly when measured at a scan rate of 0.5°C/min. Absorbance readings at 260 nm were recorded every 12 s. The concentration of the RNA sample was 9 μM RNA triplex in 25 mM phosphate buffer, pH 4.3, conditions at which NMR data indicated the RNA was predominantly in the form of a triplex. The phosphate buffer without added RNA was used as a reference and this absorbance was subtracted from the absorbance of the RNA sample. Samples were run through one heating and cooling cycle before data were recorded. Melting data were imported into Kaleidagraph (Synergy Software, Reading, PA). Data were smoothed using a five data point moving average prior to calculating the first derivative of the spectrum.

NMR data were collected at 500 MHz using a Bruker AMX spectrometer equipped with a proton-dedicated probe. Samples of the RNA triplex typically contained 1.6 mM RNA dissolved in 0.6 ml 40 mM potassium phosphate buffer, pH 4.3. Solvent suppression was achieved using the jump-return method (22) for 1D and 2D NOE spectra of samples in 92% H_2O/8% D_2O solvent. 1D spectra were obtained at temperatures between 1 and 70°C and 2D spectra were obtained at 15°C. Presaturation was used for solvent suppression for samples in D_2O solvent and for 2QF-COSY spectra. 2D NOE spectra were obtained in 90% H_2O/10% D_2O and 99.9% D_2O, with mixing times between 225 and 250 ms. For the 2D NOE spectrum in 90% H_2O/10% D_2O, 128 scans were acquired for each of 700 free induction decays of 2048 complex points, with a sweep width of 12195 Hz and recycle time of 1.8 s between scans. All 2D NMR spectra were acquired in the phase-sensitive mode by the method of time-proportional phase incrementation (TPPI). 2D spectra were typically acquired with 700 blocks of 1024 or 2048 complex points, using sweep widths of 4500–5000 Hz for spectra in D_2O and 12195 Hz for spectra in 90% H_2O/10% D_2O.

RESULTS

NMR experiments

NMR methods were used to investigate the structure and stability of the 30 nt intramolecular RNA triplex. For the 21 nt of the base triples, resonance assignments were determined for all of the guanosine and uridine imino and cytosine amino protons, while a smaller fraction (45%) of the non-exchangeable H2, H6, H8 and H1′ protons were assigned (Table 1). Resonance assignments were determined through the use of homonuclear 2D nuclear Overhauser effect (NOE) and 2QF-COSY spectra obtained in D_2O and 92%H_2O/8%D_2O solvents, at 15°C in 40 mM phosphate buffer. The resonance assignments in Table 1 are reported at pH 4.3, where the RNA is predominantly in the triplex form.

Resonance assignments were most straightforward for the Watson–Crick base paired regions of the triplex (nt 1–7 and 12–18). The patterns observed in the 2D NOE spectra for these nucleotides are typical of A-form helical RNA, where regular patterns of intra- and internucleotide NOEs were observed. NOEs involving the imino and amino protons were particularly valuable in establishing the resonance assignments in the Watson–Crick base paired strands of the RNA molecule (Fig. 2). The imino proton of each Watson–Crick base paired uridine (U12, U14, U15 and U17) was identified by a very strong NOE peak to the H2 proton of the adenosine with which it is base paired (Fig. 2A). Amino protons of each of the three Watson–Crick base paired

Table 1. Assignments of chemical shifts (p.p.m.) for the 30 nt triple helical RNA

	imino	amino	H2/H5	H8/H6	H1′
G1	12.05	8.44/8.29			
A2		7.53/7.64	7.35	7.42	
G3	12.46	8.39/8.32			5.70
A4		7.85/7.38	7.03	7.33	
A5		8.08/7.47	7.23	7.17	5.85
G6	12.80	8.30/8.45			5.82
A7		7.98/7.38	7.10	7.20	
U8					
U9					
C10					
G11	10.33				
U12	14.01		5.78	7.90	
C13		8.50/7.16	6.01	7.64	5.46
U14	13.82				
U15	13.20				5.51
C16		8.01/6.96	5.51	7.75	5.43
U17	13.79				
C18		7.62/6.91			
U19					
C20					
U21					
C22					
U23					
C24		9.78/8.97	5.69	7.59	
U25	13.47				
C26		9.55/9.04	5.80	7.65	
U27	13.93				
U28	13.42				
C29		9.28/8.94	5.73	7.54	
U30	13.64				

Data for protons are reported at 15°C, pH 4.3, with chemical shifts relative to the residual H_2O resonance at 4.82 p.p.m.

cytosines (C13, C16 and C18) were assigned by the NOEs to their own H5 protons and the imino proton of the guanosine with which it is base paired. In this way it was possible to distinguish the A·U and G·C pairs within the base triples. NOEs from the imino protons of the Watson–Crick pairs to the H1′ protons of the nucleotides in the next sequential (*i + 1*) position relative to each member of the pair were also observed. These imino–H1′ NOEs are also typical of regular A-form helical RNA (23). NOE peaks were observed from the H2 protons of the Watson–Crick paired adenosines (A2, A4, A5 and A7) at position *i* to the H1′ proton of nucleotide *i + 1*, as well as the H1′ of the nucleotide that is base paired with nucleotide i – *1*. The H2–H1′ NOEs were of nearly equal intensities, another feature typical of A-form helical duplex RNA (24). In spectra of RNA double helices it is often possible to observe NOEs between the imino protons of adjacent base pairs and this was the case in the RNA triplex. All six of the possible imino–imino NOEs in the Watson–Crick part of the triplex were observed, although weakly, and were consistent with our assignments. An imino–imino NOE was also observed between the two sequential uridines of the Hoogsteen strand.

The amino protons of the three cytosines in the Hoogsteen-paired third strand (C24, C26 and C29) were identified by their distinctive chemical shifts, downfield of the other amino resonances in the molecule (25,26). These amino protons resonate between 8.94 and 9.78 p.p.m. and were identified by the very strong NOE between the two protons belonging to the same amino group (peaks a, b and c in Fig. 2B). In addition to providing convenient starting points for assigning the resonances of the third strand, these cytosine amino resonances served to unambiguously confirm the formation of the RNA triplex. Strong NOE peaks were observed between the cytosine amino protons and the H5 proton on the same nucleotide and weaker intranucleotide NOE peaks were observed between the cytosine amino and H6 protons. The C24–G1–C18 and C29–G6–C13 triples were assigned by NOE peaks between the H5 proton of cytosine on the third strand and one of the amino protons of the Watson–Crick base paired cytosine within the same triple. An NOE peak between the amino protons of C29 and C13 was also observed and this served to further confirm the assignment of the C29–G6–C13 triple. NOEs were observed between the amino protons of C24 and A2 and between the amino protons of C29 and A7. A sequential imino–imino NOE between U27 and U28 in the Hoogsteen bound third strand was also observed. Weak imino–imino NOE peaks were seen between the U27 and U15 imino protons, the U30 and U12 imino protons and the G3 and U27 imino protons. NOEs involving the imino protons of G3 and G6 were stronger than those associated with G1. This is presumably due to the more rapid exchange rate of the G1 imino proton with the solvent, due to its location at the end of the stem.

It should be noted that some resonances are apparent in the NMR spectra that were not assignable to the triplex (Fig. 3). This indicates the likely presence of some alternate RNA conformations or sequences. However, the major peaks of the 2D spectra are clearly assignable to the triplex (Fig. 2). This provides substantial evidence that the major species in the NMR tube is triplex RNA.

H1′–H2′ scalar coupling constants are sensitive indicators of ribose conformation in RNA. Small H1′–H2′ couplings, indicated by weak or absent correlation peaks in 2QF-COSY spectra, identify riboses that are predominantly in the 3′ *endo* conformation. For each of the assigned H1′ and H2′ resonances of the purine strand, the correlations in the 2QF-COSY spectrum are weak or absent. The H1′ resonances of the pyrimidine strands are mostly unresolved, however the NOE data indicate that these H1′ and H2′ protons resonate in relatively narrow ranges centered at 5.5 and 4.5 p.p.m. respectively. The H1′–H2′ scalar couplings for these riboses are also weak or absent. This suggests that the riboses of the base triples are predominantly in the C3′ *endo* conformation, which is consistent with the observed A-helix-like pattern of NOEs. An alternative explanation, that we cannot entirely rule out at this time, is that the absence of H1′–H2′ 2QF-COSY peaks is due to line broadening as a result of conformational exchange or the high molecular weight of the triplex.

Cytosine nucleotides in the Hoogsteen paired third strand can become protonated at the N3 position upon triple helix formation at low pH. These cytosine imino protons resonate downfield of the other imino protons (16,17,25–28). The three broad resonances observed between 14 and 15 p.p.m. at low pH are probably associated with the imino protons of cytosines C24, C26 and C29 (Fig. 3). The broadness of these putative cytosine imino resonances is most likely due to their relatively rapid exchange with protons of the solvent water.

The two Watson–Crick base paired strands in the intramolecular triple helix are linked by a stable RNA tetraloop with the sequence UUCG (19–21). Several characteristic resonances associated with the UUCG tetraloop were identified within our 30 nt RNA and the chemical shifts of these resonances are

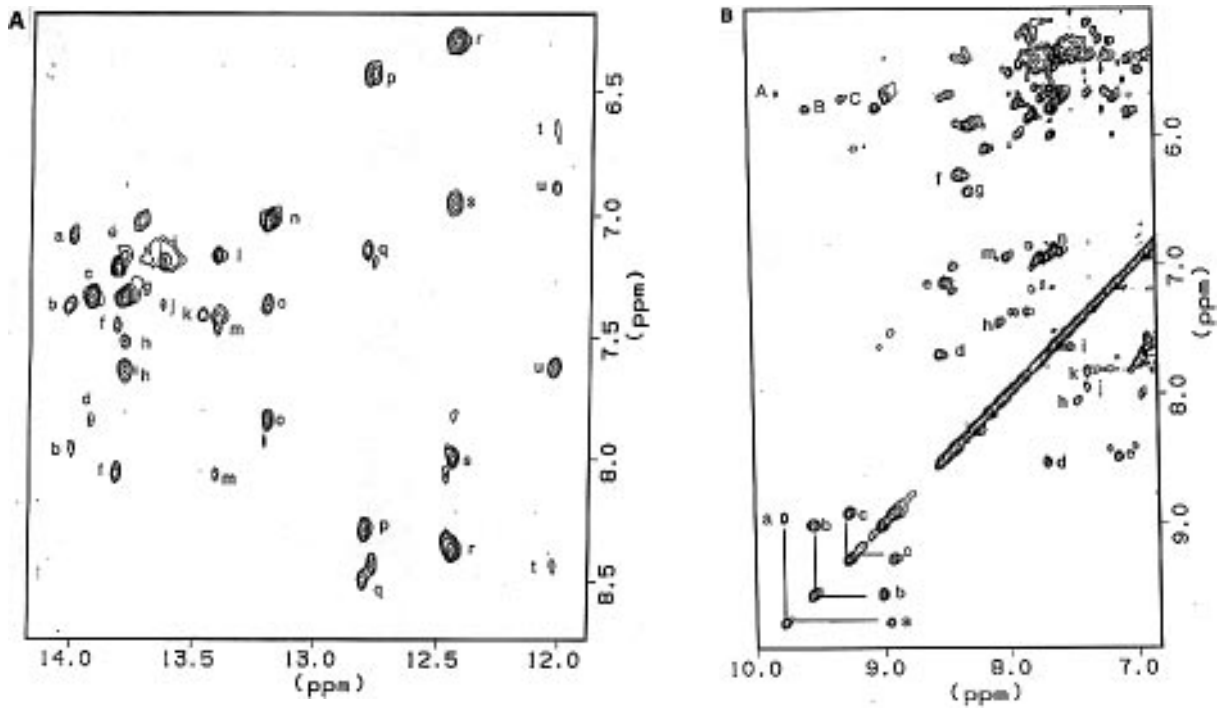

Figure 2. (**A**) Section of the 2D NOE spectrum (280 ms mixing time) of the RNA triple helix in 90% H_2O/10% D_2O solvent, 40 mM phosphate buffer, pH 4.8, at 15°C. The solvent resonance was suppressed by the 1–1 (jump-return) method. Several of the assigned cross-peaks are labeled as follows: (a) U12(H3)–A7(H2); (b) U12(H3)–A7(H6); (c) U27(H3)–A4(H8); (d) U27(H3)–A4(H6); (e) U14(H3)–A5(H2); (f) U14(H3)–A5(H6); (g) U17(H3)–A2(H2); (h) U17(H3)–A2(H6); (i) U30(H3)–A7(H8); (j) U30(H3)–A7(H6); (k) U25(H3)–A2(H8); (l) U28(H3)–A5(H8); (m) U28(H3)–A5(H6); (n) U15(H3)–A4(H2); (o) U15(H3)–A4(H6); (p) G6(H1)–G6(H2); (q) G6(H1)–C13(H4); (r) G3(H1)–G3(H2); (s) G3(H1)–C16(H4); (t) G1(H1)–G1(H2); (u) G1(H1)–C18(H4). (**B**) Section of the same 2D NOE spectrum, with assigned cross–peaks labeled as follows: (A) C24(H4)–C24(H5); (B) C26(H4)–C26(H5); (C) C29(H4)–C29(H5); (a) C24(H4)–C24(H4); (b) C26(H4)–C26(H4); (c) C29(H4)–C29(H4); (d) G11(H2)–G11(H2); (e) C13(H2)–C13(H2); (f) G3(H2)–G3(H2); (g) G6(H2)–G6(H2); (h) A5(H6)–A5(H6); (i) A2(H6)–A2(H6); (j) A7(H6)–A7(H6); (k) A4(H6)–A(H6); (m) C16(H4)–C16(H4); (n) C18(H4)–C18(H4).

similar to those reported for the same tetraloop sequence as part of an RNA stem–loop (19–21). It therefore appears likely that the structure of the tetraloop within the RNA triple helix is similar to its structure when it is part of a smaller stem–loop. We were unable to specifically assign resonances of the UCUCU loop that links the two pyrimidine strands in our intramolecular RNA triplex, although the broad imino resonances near 11 p.p.m. (Fig. 3) are probably associated with the uridine imino protons in this loop, assigned by default. It is therefore not possible to make detailed conclusions on the structure of the UCUCU loop based on our NMR data. However, it is likely that this loop does not contain any particularly stable base pairings, based on the broad imino proton resonances that disappear due to exchange broadening at relatively low temperatures.

In summary, we have found that the patterns and intensities of the assigned NOEs involving the Watson–Crick base paired nucleotides are characteristic of an A-form helical RNA duplex (nt 1–7 and 12–18). NMR data associated with the third RNA strand are consistent with a structural model where this third strand uses Hoogsteen interactions to bind in the major groove of an otherwise A-form helical duplex. The ribose conformations on all three strands of the RNA triplex are most likely C3′ *endo*. It is interesting that Liquier and co-workers (17) also recently reported that riboses of all three strands in an RNA triplex are in the C3′ *endo* conformation, although it is significant that this previous result was based on infrared absorption measurements rather than NMR data.

Stability studies

pH dependence of exchangeable proton resonances. The imino and amino resonances exhibit a strong pH dependence, with the spectrum at low pH being considerably more complex (Fig. 3). The low pH spectrum is attributed to the stable RNA triple helix, based on our resonance assignments. At pH 7, the resonances associated with the hydrogen bonded imino and amino resonances of the Hoogsteen strand are not observed and the remaining exchangeable resonances are attributed to the Watson–Crick base pairs and UUCG tetraloop (Fig. 3). The RNA therefore has a stem–loop structure at pH 7, with the pyrimidine resonances of the Hoogsteen strand being unpaired. This pH-dependent behavior of the RNA triplex stem–loop equilibrium is essentially the same as observed for DNA triple helices (27).

Temperature dependence of exchangeable proton resonances. The temperature dependence of the imino proton resonances can provide insight into the stabilities of the base pairings within the triple helix. As the temperature is increased, the exchange rate of the imino protons with those of the solvent increases and the resonances broaden and disappear. The exchange rates of the imino protons are related to the frequency with which the base pairs open. This frequency increases as the melting point of the RNA is approached.

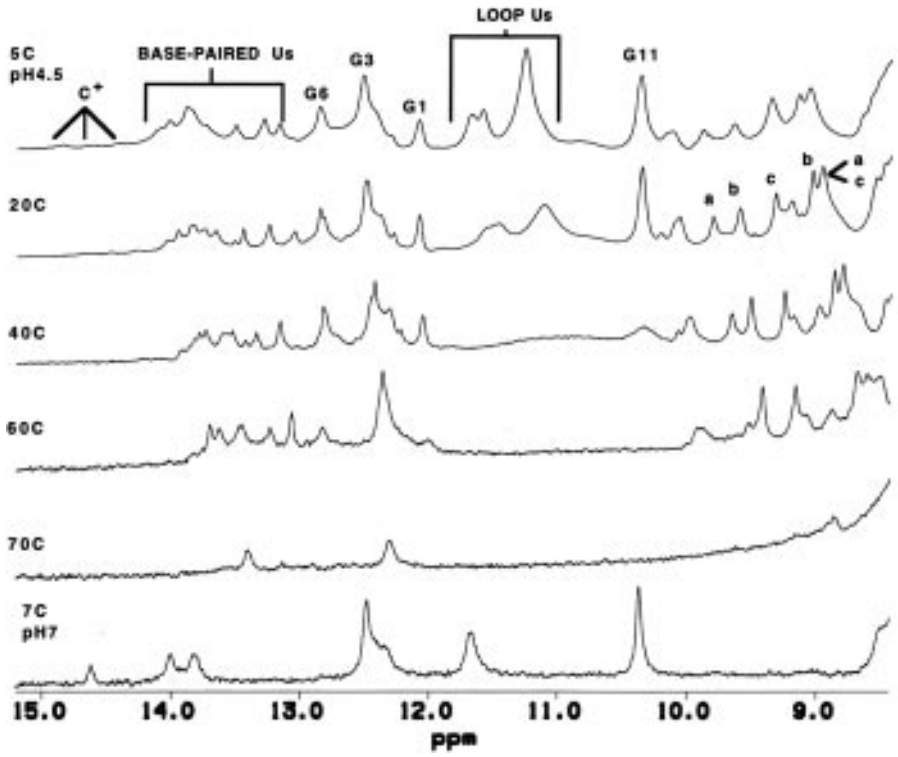

Figure 3. The top five spectra show imino and amino proton resonances of the RNA triple helix, recorded at temperatures between 1 and 70°C in 92% H_2O/8% D_2O solvent, 40 mM phosphate buffer, pH 4.8. Peaks labeled a, b and c are the amino protons of residues C24, C26 and C29 respectively. The bottom spectrum shows the imino proton resonances for the same RNA sequence at pH 7, where the resonances are attributed to imino protons of the G·C and A·U Watson–Crick base pairs in the hairpin and G11 of the tetraloop. The RNA forms a triple helix at the lower pH and forms a hairpin at the higher pH.

The RNA triple helix appears to be quite stable as judged by the temperature dependence of the imino resonances. First to disappear as the temperature is increased are the relatively broad resonances with chemical shifts between 10.5 and 11.5 p.p.m., assigned to the imino protons of the loop uridines, all of which are expected to be relatively solvent accessible. The imino resonance of G11, in the UUCG tetraloop, is observed at temperatures of up to 40°C. This imino proton is hydrogen bonded as part of the stable tetraloop structure (21). The remaining imino and amino resonances are associated with the hydrogen bonded protons of the Watson–Crick and Hoogsteen pairs. The spectrum obtained at 60°C shows that the hydrogen bonded protons associated with all three strands remain significantly protected from rapid exchange, although the resonances are broadened and reduced in intensity. As the temperature is increased, the resolved guanosine imino protons at the ends of the Watson–Crick part of the triplex (G1 and G6) are less protected from solvent exchange than G3 at the center of the triplex. Of particular interest, the amino resonances of the protonated cytosines are clearly observed, even at this relatively high temperature. This suggests that the triplex conformation is the major species at 60°C. The exchangeable proton resonances almost completely disappear in an abrupt transition between 60 and 70°C. This phenomenon is completely reversible upon cooling the sample. The temperature dependence of the exchangeable proton resonances is most consistent with denaturation in a single transition, from triplex to a single-stranded structure, rather than a two-step transition, proceeding from triplex to stem–loop to single strand. In an NMR study of a 28 base DNA intramolecular triplex, Feigon and co-workers (28) found that the imino proton resonances disappear at slightly above 60°C, a temperature that is similar to that observed for our RNA triplex.

Analysis of UV absorption changes upon thermal denaturation. UV absorption spectroscopy as a function of temperature was used to monitor the unfolding of the RNA triplex in a solution of 25 mM potassium phosphate, pH 4.3. A non-linear least squares fit of the derivative data to a single unimolecular transition yielded excellent agreement with a two-state model (data not shown), with a melting temperature T_m of 62°C. In this respect, the NMR and UV data are in agreement, with each experimental method indicating essentially the same melting point for the RNA triplex. The data obtained using each method (NMR and UV absorption) is most consistent with the melting transition occurring in a single step, without a stem–loop intermediate.

Modeling of the RNA triplex structure

As expected for a 30 nt RNA, the problems encountered in interpreting the rather complex NMR spectra are formidable and this does limit the resolution of the structural model we obtain. However, the substantial amount of data that we have been able to interpret provides significant information regarding the triplex structure. In particular, the hydrogen bonds identified by slow exchanging imino and amino protons and ribose conformations, supported by our NMR data, provide significant constraints upon

the possible structure of the triplex. We therefore constructed a model of the RNA triplex that is consistent with 99 interproton distance restraints derived from our assigned NMR data, corresponding to five restraints per nucleotide for the 21 nt triplex model. Clearly, this model does not represent a finished high resolution structure. However, this preliminary model is useful in identifying some factors that may contribute to the stability of the RNA triplex.

The structural model was constructed using the X-PLOR version 3.1 program suite (29). An initial model was constructed as follows. Nucleotides 1–7 and 12–18 were placed exactly as they would be in an ideal A-form double helix (30). This is entirely consistent with our observed NOEs and hydrogen bonds for these nucleotides. Nucleotides C24–U30 were placed along the major groove of the double helical stem so that they spanned from G1 to A7, also consistent with our observed NOEs. Loop nucleotides were not included in the model. The initial model had a relatively high value for the X-PLOR energy function, primarily due to van der Waal's violations involving close contacts of the backbones of the purine strand and the pyrimidine third strand. The initial model was then subjected to simulated annealing with the following constraints: distances between hydrogen bonded atoms identified in our NMR data were restricted to be <2.4 Å; torsion angle restraints were used to restrict ribose conformations for nt 1–7, 12–18 and 24–30 to C3′ *endo* to be consistent with our NMR results; bases within each triple were penalized for deviating >4° from co-planar. The near co-planarity of the bases is consistent with fiber diffraction data on triple helices (31,32) and was included in our model building to compensate for the relatively sparse NOE information. Interatomic distances between the atoms of nt 1–7 and 12–18 were restricted to be within 1.5 Å of ideal A-helical values (30), consistent with our NMR data indicating these nucleotides are in A-helical conformation. The 1.5 Å deviation from ideal A-helical interatomic distances for the Watson–Crick strands was found to be sufficient to permit the third strand to be accommodated in the major groove. The target function that was minimized during the simulated annealing process contained the following: (i) quadratic harmonic potential terms for covalent geometry, including bonds, angles, planes and chirality; (ii) square-well potentials for interatomic distance constraints and torsion angle restraints. The simulated annealing procedure produced a model of the structure that has chemically reasonable values for bond lengths and angles, no significant van der Waal's violations, a low value of the X-PLOR energy function and is consistent with our NMR data. Hydrogen bonds involving the 2′-hydroxyl groups were not included as restraints; these hydrogen bonds emerged during the simulated annealing process.

Upon close inspection of the triple helix model, it can be seen that the ribose phosphate backbones of the purine strand (strand 1) and the Hoogsteen strand (strand 3) are in rather close proximity. Of particular interest, the 2′-hydroxyl proton of each nucleotide of the third strand is positioned only 1.8 Å from a phosphate oxygen of the purine strand (Fig. 4). This implies that there may be an 'extra' hydrogen bond linking the first and third strands of the triplex, in addition to the expected hydrogen bonds in the Hoogsteen pairings of the bases. This hydrogen bond can account for the C3′ *endo* conformation of the ribose groups on the third strand. If the third strand ribose groups were C2′ *endo*, the 2′-hydroxyl proton would not be in a position to hydrogen bond with the phosphate oxygen of the purine strand. The existence of this hydrogen bond has also been suggested in a previous model building study done in the absence of NMR-derived ribose conformation and hydrogen bonding restraints (6).

DISCUSSION

Our studies demonstrate that the 30 nt RNA with the sequence shown in Figure 1 forms an intramolecular antiparallel triple helix, with the ribose groups on all three strands most likely in the C3′ *endo* conformation. In addition to the Watson–Crick and Hoogsteen hydrogen bonding between the bases, our structural modeling provides evidence for a possible hydrogen bond between the 2′-hydroxyl proton and a phosphate oxygen on the backbone of the purine strand. A DNA triplex lacking a 2′-hydroxyl group could not form this last hydrogen bond. This could, at least partially, account for the higher stability reported for triplexes that contain RNA rather than DNA as the Hoogsteen paired third strand (4–6).

NMR studies of intramolecular DNA triple helices have indicated that the riboses are predominantly in the C2′ *endo* (28), as opposed to the C3′ *endo*, conformation that we observe for the RNA triple helix. This apparent difference in sugar conformations between DNA and RNA triple helices may be a consequence of the formation of the hydrogen bond between the 2′-hydroxyl group on the third strand and the purine strand. Since the DNA third strand cannot form this hydrogen bond, it does not have as much to gain in terms of stability by having its riboses adopt the C3′ *endo* conformation. Riboses of the RNA third strand can also be stabilized by a hydrogen bond between the 2′-hydroxyl proton of nucleotide *i* and the O4′ oxygen of nucleotide *i + 1* (33). This hydrogen bond also cannot occur in DNA and also requires that the riboses be in the C3′ *endo* conformation. Base stacking probably makes a substantial contribution to stabilizing the third strand in triple helices and is of course a common factor in both DNA and RNA.

Stabilities of RNA·DNA*DNA triplexes where the ribose O2′ is methylated have been reported (34,35). The O2′ methylated triplex possessed a higher thermal stability than a triplex containing standard RNA or DNA. Hydrogen bonding interactions between the 2′-hydroxyl proton of the third strand and phosphate oxygen of the purine strand are not possible due to the methylation at the O2′ position. The loss of possible hydrogen bonding interactions is believed to be compensated for by an increase in favorable Van der Waals interactions with the methyl groups (34,35). These results are therefore not entirely inconsistent with a hydrogen bond between the 2′-hydroxyl of strand 3 and the phosphate oxygen of the purine strand in unmodified RNA.

Recently, preliminary NMR and stability studies were reported for an RNA intramolecular triplex with a substantially different sequence from the RNA used in our study (16,17). A comparison of the results for the two molecules gives some additional insight into the factors that are important for stabilizing the RNA triplexes. The RNA investigated by Liquier *et al.* (17) contained seven triples (four UAU and three CGC triples), however, the order of the triples within the stem was different from that used in the work reported here, as were the sequences of the connecting loops. Liquier *et al.* (17) used four cytidines for their first connecting loop, compared with our RNA triplex, which uses a tetraloop sequence (UUCG) that is known for its high stability in RNA stem–loops. Liquier and co-workers (17) used four uridines for their second connecting loop, compared with our study, where

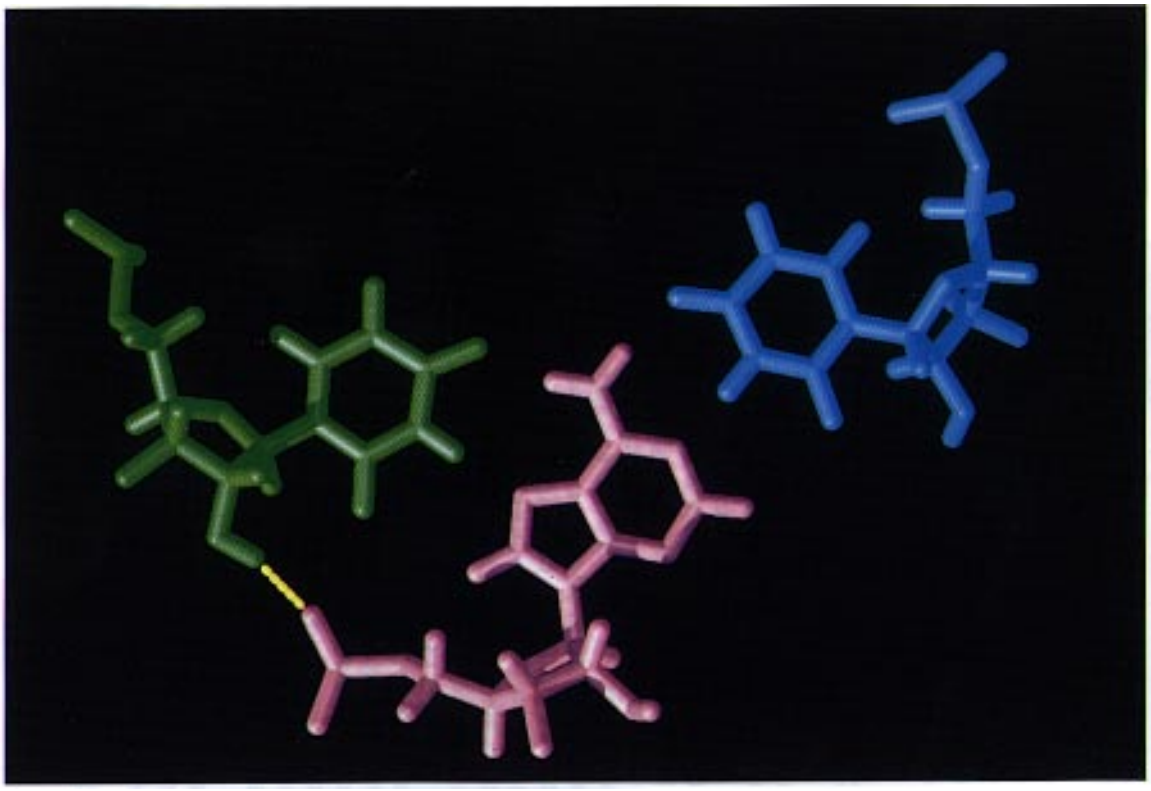

Figure 4. A diagram of a triple, showing the possible hydrogen bond (yellow) between the 2′-hydroxyl proton of the Hoogsteen strand and a phosphate oxygen of the purine strand. The pyrimidine of a cytosine in the third strand is green and the Watson–Crick GC base pair is pink and blue.

a loop of 5 nt (UCUCU) was used. Despite these difference in the sequences, the melting point reported by Liquier *et al.* (17) differed from the melting point found in our RNA triplex by only 2°C under conditions (20 mM NaCl, pH 4.8) similar to ours. The similar stabilities observed in each study are significant in that they indicate that the stable intramolecular RNA triplexes are not artifacts induced by the choice of loop or stem sequences. The similarity of the melting points found for the different RNA triplexes suggests that the loop sequences may not be particularly important in governing intramolecular triplex stability. This is somewhat surprising, when one considers that there is a very large increase of 8.6–11.3°C in melting points for duplex RNAs with a G·C base pair prior to the stable UUCG tetraloop sequence versus a UUUU tetraloop (36). An RNA duplex connected by a UUUU tetraloop has been shown to have a higher stability than a CCCC tetraloop (37), used as the first connecting loop by Liquier and co-workers (17). The choice of loop sequence has also been shown to influence the melting points of DNA triplexes (38). A comparison of the study by Liquier *et al.* (17) with our work also suggests that the length of the second connecting loop may have little effect on the stability of the third strand, at least in the case of loops of 4 or 5 nt.

ACKNOWLEDGEMENT

This material is based in part upon work supported by the Texas Advanced Research Program under grant no. 003658-185 and is in partial fulfillment of the requirements for the PhD degree at the University of Texas at Austin (JAH).

REFERENCES

1 Moser,H.E and Dervan,P.B. (1987) *Science*, **238**, 645–650.
2 LeDoan,T., Perrouault,L., Praseuth,D., Habboub,N., Decout,J.-L., Thoung,N.T., Lhomme,J. and Helene,C. (1987) *Nucleic Acids Res.*, **15**, 7749–7760.
3 Maher,L.J., Wold,B. and Dervan,P.B. (1989) *Science*, **245**, 725–730.
4 Han,H. and Dervan,P. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3806–3810.
5 Roberts,R.W. and Crothers,D.M. (1992) *Science*, **258**, 1463–1465.
6 Escudé,C., Francois,J., Sun,J., Günther,O., Sprinzl,M., Garestier,T. and Hélène,C. (1993) *Nucleic Acids Res.*, **21**, 5547–5553.
7 Kinniburgh,A.J., Firulli,A.B. and Rukmini,K. (1994) *Gene*, **149**, 93–100.
8 Kim,S.-H., Quigley,G.J., Suddath,F.L., McPherson,A., Sneden,D., Kim,J.J., Weinzierl,J. and Rich,A. (1973) *Science*, **179**, 285–288.
9 Robertus,J.D., Ladner,J.E., Finch,J.R., Rhodes,D., Brown,R.S., Clark,B.F. and Klug,A. (1974) *Nature*, **250**, 546–551.
10 Michel,F., Ellington,A.D., Couture,S. and Szostak,J.W. (1990) *Nature*, **347**, 578–580.
11 Green,R. and Szostak,J.W. (1994) *J. Mol. Biol.*, **235**, 140–155.
12 Chastain,M. and Tinoco,I. (1992) *Biochemistry*, **31**, 12733–12741.
13 Puglisi,J.D., Chen,L., Frankel,A.D. and Williamson,J.R. (1992) *Science*, **257**, 76–80.
14 Puglisi,J.D., Chen,L., Frankel,A.D. and Williamson,J.R. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3680–3684.
15 Aboul-ela,F. and Varani,G. (1995) *J. Mol. Biol.*, **253**, 313–332.
16 Klinck,R., Guittet,E., Liquier,J., Taillandier,E., Gouyette,C. and Huynh-Dinh,T. (1994) *FEBS Lett.*, **355**, 297–300.
17 Liquier,J., Taillandier,E., Klinck,R., Guittet,E., Gouyette,C. and Huynh-Dinh,T. (1995) *Nucleic Acids Res.*, **23**, 1722–1728.
18 Milligan,J., Groebe,D., Witherell,G. and Uhlenbeck,O. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.
19 Cheong,C., Varani,G. and Tinoco,I. (1990) *Nature*, **346**, 680–682.
20 Varani,G., Cheong,C. and Tinoco,I. (1991) *Biochemistry*, **30**, 3280–3289.
21 Allain,F.H.T. and Varani,G. (1995) *J. Mol. Biol.*, 250, 333–353.
22 Guéron,M., Plateau,P. and Decorps,M. (1991) *Prog. Nucl. Magn. Resonance Spectrosc.*, **23**, 135–219.

23 Heus,H and Pardi,A. (1991) *J. Am. Chem. Soc.*, **113**, 4360–4361.
24 Hoffman,D.W., Colvin,R.A., Garcia-Blanco,M.A. and White,S.W. (1993) *Biochemistry*, **32**, 1096–1104.
25 Rajagopal,P. and Feigon,J. (1989) *Nature*, **339**, 637–640.
26 Mooren,M., Pulleyblank,D., Wijmenga,S., Blommers,M. and Hilbers,C. (1990) *Nucleic Acids Res.*, **18**, 6523–6529.
27 Radhakrishnan,I., Gao,X., Santos,C., Live,D. and Patel,D. (1991) *Biochemistry*, **30**, 9022–9030.
28 Macaya,R., Wang,E., Schultze,P., Sklenar,V. and Feigon,J. (1992) *J. Mol. Biol.*, **225**, 755–773.
29 Brünger,A.T. (1990) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR.* Yale University Press, New Haven, CT.
30 Arnott,S., Hukins,D.W.L. and Dover,S.D. (1972) *Biochem. Biophys. Res. Commun.*, **48**, 1392–1398.
31 Arnott,S. and Bond,P.J. (1973) *Nature New Biol.*, **244**, 99–101.
32 Arnott,S. and Selsing,E. (1974) *J. Mol. Biol.*, **88**, 509–521.
33 Saenger,W. (1984) *Principles of Nucleic Acid Structure.* Springer-Verlag, New York, NY.
34 Escudé,C., Sun,J.S., Rougée,M., Garestier,T. and Hélène,C. (1992) *C.R. Acad. Sci. Paris, Ser. III*, **315**, 521–525.
35 Shimizu,M., Konishi,A., Shimada,Y., Inoue,H. and Ohtsuka,E. (1992) *FEBS Lett.*, 302, 155–158.
36 Antao,V.P., Lai,S.Y. and Tinoco,I. (1991) *Nucleic Acids Res.*, **19**, 5901–5905.
37 Groebe,D.R. and Uhlenbeck,O.C. (1988) *Nucleic Acids Res.*, **16**, 11725–11735.
38 Wang,S., Booher,M.A. and Kool,E.T. (1994) *Biochemistry*, **33**, 4639–4644.