Nucleic Acids Research, 1996, Vol. 24, No. 20 **4015–4022**

Secondary structure of the r(CUUCGG) tetraloop

Jean E. Kanyo, Jean Duhamel[+] and Ponzy Lu*

Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104-6323, USA

Received June 17, 1996; Revised and Accepted August 28, 1996

ABSTRACT

The oligonucleotide r(GGACUUCGGUCC) has been observed to adopt a hairpin conformation by solution NMR and a double helical conformation by X-ray diffraction. In order to understand this apparent conflict, we used time-resolved fluorescence depolarization and ^{19}fluorine NMR to follow the secondary structure of this dodecamer as the solution composition was changed stepwise from the NMR experimental conditions to those used for crystallization. Calculation of the dodecamer concentration in the crystal (180 mM strands) and the cation concentration needed for neutrality (>2 M) prompted investigation of a tethered species, in which two dodecamers are connected by a string of 4 nt, geometrically equivalent to ~100 mM strands, in 2.5 M NaCl. The RNA tetraloop and its DNA analog maintain a single-strand hairpin conformation in solution, even under the conditions used to grow the crystal. Under high salt conditions, the tethered RNA and DNA analogs of this sequence yield secondary components which could be the double helical conformation. Crystal contacts in addition to solvent changes and high RNA concentrations are needed to obtain the double helix as the predominant species.

INTRODUCTION

Our understanding of RNA function, stability and catalytic activity is hampered by a paucity of three-dimensional structures. Nucleic acid structures, particularly of folded RNA, are difficult to obtain using solution NMR and single crystal X-ray diffraction methods. The NMR study of RNA molecules larger than 30 nt requires isotope enrichment (1–3), while X-ray diffraction requires large, ordered RNA crystals, which have only been successfully produced for tRNA and a few other small RNAs. Some notable RNA structures have recently been solved, including: the P1 helix (20 nt) from Group I self-splicing introns, solved by heteronuclear multidimensional NMR (4); a 34 nt hammerhead ribozyme complexed to a DNA inhibitor, solved by X-ray diffraction (5); two complexes (38 and 40 nt) between RNA hairpins with complementary loops (6), termed 'kissing hairpins' (7); a 30 nt RNA construct modeling a regulatory element of the human U1A pre-mRNA, both free and complexed with the RNA binding domain of the U1A protein, solved by NMR (8); a 21 nt hairpin complexed to the U1A domain, studied by X-ray diffraction (9).

Although solution NMR and single crystal X-ray diffraction are the methods of choice for macromolecular structure determination, solvent composition for the two methods are very different and so too may be the results. We ask why a solution NMR structure and an X-ray diffraction structure disagree in the case of an RNA dodecamer, r(GGACUUCGGUCC) (Fig. 1). This sequence is of interest because it is expected to adopt a particularly stable hairpin structure (10). Approximately half of all hairpins in rRNAs are tetraloops, with over two thirds of these sequences belonging to two motifs, UNCG and GNRA, where N is any nucleotide and R is a purine (11). In fact, the most frequent tetranucleotide loop sequence in *Escherichia coli* 16S rRNA is C(UUCG)G. When Tuerk *et al.* (12) searched a bacteriophage T4 sequence library for this hexanucleotide sequence, they found that it occurred, flanked by inverted repeats, amazingly frequently in intercistronic regions. Many of these predicted hairpins efficiently terminated DNA synthesis by avian myeloblastosis virus (AMV) reverse transcriptase, suggesting that the enzyme could not traverse the secondary structure. Thermal denaturation of model hairpins (12,13) demonstrated that C(UUCG)G tetraloops have melting points ~20°C higher than homopolymeric tetraloops, indicating that the UUCG loop sequence enhances the stability of the constructs. In fact, when an RNA structure prediction algorithm assigns 2 kcal/mol extra stability to hairpins with the eight most common rRNA tetraloop sequences (all belonging to the UNCG and GNRA families and including UUCG), the results become more consistent with phylogenetic analysis (14). Hairpins with UUCG are also abundant in many other RNAs, both eukaryotic and prokaryotic, including precursor and mature tRNAs and the catalytic M1 RNA of *E.coli* RNase P (12).

The solution structure of the RNA dodecamer r(GGACUUCG-GUCC) was determined by Tinoco and co-workers using NMR NOE distance constraints and scalar coupling constants, followed by distance geometry and constrained energy minimization (4,15,16). The dodecamer did indeed adopt a hairpin conformation. Structural features in the loop convincingly explained the unusual stability, including extended base stacking, a U–G base pair closing the loop and a contact from the loop cytidine to a phosphate oxygen. The results agreed with chemical modification experiments of CUUCGG hairpins in 16S rRNA (17) in that the more reactive, and variable, second residue of the loop lacked specific contacts and stuck out into the solvent in the NMR structure. The solution structure, together with the biophysical and biochemical data, painted a satisfying, convincing picture of this stable UUCG hairpin. A year later, however, Holbrook *et al.* (18) published the single crystal X-ray diffraction structure of the same oligomer. The data were collected to 2.0 Å resolution and the structure was solved by molecular replacement with a 4 bp probe and refined to a crystallographic *R* factor of 18%. Surprisingly, the RNA

* To whom correspondence should be addressed

[+]Present address: Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada

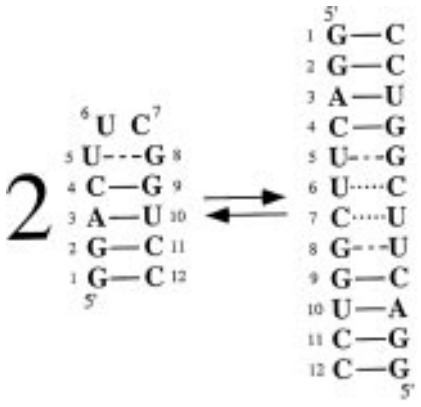


Figure 1. Published structures of r(GGACUUCGGUCC). (Left) Hairpin solution structure as determined by NMR in 10 mM sodium phosphate, pH 6.5, 0.01 mM EDTA (4,5). (Right) Duplex single crystal X-ray diffraction structure (6). Crystallization conditions were 100 mM sodium citrate, 50 mM Tris, pH 8.5, 2.5 mM $MgCl_2$, 12–32% PEG400 (S. R. Holbrook, personal communication). Dashed line, U–G base pair; dotted line, U–C base pair.

adopted a double helical conformation with a remarkably regular structure, despite four consecutive non-Watson–Crick base pairs in the center of this helix. This tract consisted of two U–G wobble base pairs and two U–C base pairs, with the latter composed of one base–base hydrogen bond and one water-mediated hydrogen bond. In the crystal, the duplexes stacked end-to-end forming pseudo-infinite helices. These interacted through water-mediated hydrogen bonds and four direct hydrogen bonds per dodecamer, which all involved ribose 2′-hydroxyls.

Thus, the two primary techniques for macromolecular structure determination, NMR and X-ray diffraction, yielded disparate results for this RNA sequence. This result was echoed by Fujii *et al.* (19,20) upon investigation of a tridecanucleotide containing the CUUCGG sequence. NMR and thermal denaturation studies indicated that the RNA forms a stable hairpin in solution, while the crystal conformation was a double helix with the same tract of non-Watson–Crick base pairs seen by Holbrook and co-workers. Kennard and co-workers reported a very similar mispaired RNA duplex when they crystallized a nonamer containing the CUUCGG sequence (21). Holbrook *et al.* (18) state that high RNA and salt concentrations favored an intermolecular double helix over an intramolecular hairpin loop conformation, although no data other than the diffraction-based structure has been published.

We attempt here to identify the cause of the transition between the unimolecular hairpin and double helical conformations. The NMR solvent and crystallization conditions were quite different: 10 mM sodium phosphate, 0.01 mM EDTA, pH 6.5, versus 100 mM sodium citrate, 50 mM Tris, pH 8.5, 2.5 mM $MgCl_2$ and 12–32% PEG400 (S. J. Holbrook, personal communication) respectively. Therefore, we monitored the conformation of the RNA dodecamer and several analogs when the solvent was changed stepwise from that used in the solution structure determination to that of the crystal growth. We note that the RNA concentration in the crystal is 180 mM (in strands), which requires >2 M accompanying monovalent cations for charge neutrality. We constructed tethered species in which two dodecamers are connected by a string of 4 nt, geometrically equivalent to ~100 mM strands, in order to mimic these oligonucleotide concentrations. This tethering not only increases the apparent local nucleic acid concentration, but also makes the hairpin to duplex transition a unimolecular event, circumventing analysis of concentration dependence, with concomitant economic benefit. Furthermore, the unusual stability of the unimolecular hairpin or tetraloop is a property of the RNA sequence, not seen in an analogous DNA (13). Thus, to potentially favor the double helical form, we repeated these experiments with DNA.

These investigations required a direct probe of nucleic acid conformation under a wide range of solution conditions. We have used fluorescence depolarization to examine the hydrodynamic behavior of the RNA, since we are exploring large size and shape changes. The anisotropy function $r(t)$ and rotational correlation time τ_{rot} of the RNA tetraloop and several analogs are compared with those of control molecules of defined size and shape. We also used ^{19}F NMR to monitor the RNA molecule conformation in the same range of solvents, including the crystallization conditions. Nucleic acid molecules of interest were ^{19}F-labeled at specific sites and the chemical shift and width of the resonance was monitored as a function of solvent conditions.

We demonstrate that in solution the RNA and DNA dodecamers do not adopt the duplex conformation seen in the RNA crystal in appreciable amounts, even under the conditions used to grow the crystal. Under high salt conditions, the tethered species, particularly the RNA, display a significant second component, presumably the extended hairpin which mimics the double helical conformation in the crystal. Although high RNA and salt concentrations do shift the solution equilibrium toward the bimolecular double helix, they are not sufficient to make the duplex conformation the major solution component.

MATERIALS AND METHODS

Sample preparation

Oligoribonucleotides were made on a Cruachem Model PS250 Synthesizer using 2′-*O*-fluorophenyl methoxypiperidinyl-protected phosphoramidites (Cruachem) in 1 μM batches. Labeling is achieved by selectively replacing uridines on the RNA (or thymidines on DNA analogs) with 5-fluoro-2′-deoxyuridine (FdU). FdU-protected phosphoramidite monomers were purchased from American Bionetics, while all of the canonical nucleotide monomers were obtained from Cruachem. The DNA oligonucleotides were made in automated solid-phase syntheses on a Milligen Biosearch Cyclone DNA Synthesizer. Fifteen micromole (for NMR samples) or 1 μmol syntheses were performed using standard phosphoramidite methodologies. All oligomers were separated from failure sequences and shorter products by trityl-on reverse phase HPLC over a Beckman C18 column, with the RNA purified prior to 2′-OH deprotection. The oligonucleotide sequences utilized in this research are listed in Figure 2.

L buffer (Loop; solution NMR conditions) consisted of 10 mM sodium phosphate, 0.01 mM EDTA, pH 6.5 (4); D buffer (Duplex; crystal growth buffer) consisted of 100 mM sodium citrate, 50 mM Tris, pH 8.5. Actual crystal growth conditions are D buffer plus 2.5 mM $MgCl_2$ and 12–32% PEG400. Sample buffers were prepared as 2× stock solutions. Any additives are noted, as are any pH changes. Magnesium chloride was prepared as a 100 mM stock and sodium chloride as a 5 M stock, treated with diethylpyrocarbonate, and autoclaved prior to use. Polyethylene glycol (PEG) with an average molecular weight of 400 g/mol

(PEG400) was deionized over an Amberlite mixed bed resin and filtered prior to use.

NMR samples initially contained 3–4 mM DNA in L buffer containing 10% D_2O in a volume of 350–400 µl. The pH of each sample was measured (without regard to D_2O content) and adjusted by adding submicroliter amounts of 1 M NaOH or H_3PO_4 in sample buffer. For studies at higher salt concentrations, the samples were lyophilized to dryness and redissolved in 1× D buffer plus the additional salt. Samples for fluorescence studies typically contained 2.5 OD oligonucleotide in 250 µl 1× buffer (~100 µM). Ethidium bromide (EtBr) in the same buffer was added to a final concentration of 5 µM. All samples were heated in sample buffer to 90°C and allowed to reach a thermodynamically stable conformation by slow cooling to room temperature over 12 h.

Time-resolved fluorescence depolarization

The hydrodynamics of the oligonucleotides were probed by bound EtBr, with time-resolved fluorescence decays collected on a Single Photon Counting apparatus in the Regional Laser and Biology Laboratories at the University of Pennsylvania.

Exciting a population of chromophores with a pulse of vertically polarized light selectively excites those fluorophores whose absorption dipoles are nearly aligned with the *z*-axis. Subsequent fluorescence emission is polarized with anisotropy defined as

$$r(t) = I_{||}(t) - I_{\perp}(t)/I_0(t) \quad \mathbf{1}$$

where $r(t)$ is the time-dependent anisotropy, $I_{||}(t)$ is the time-dependent decay of the vertical component of the fluorescence, $I_{\perp}(t)$ is the time-dependent decay of the horizontal component of the fluorescence and $I_0(t)$ is the time-dependent total fluorescence intensity. If the chromophore exhibits significant rotational Brownian diffusion before fluorescence emission occurs, the rate of anisotropy decay will be proportional to the rotational mobility of the particle. For spherical particles,

$$r(t) = r_0 e^{-t/\tau_{rot}} \quad \mathbf{2}$$

where r_0 is the initial anisotropy and τ_{rot} is the rotational correlation time.

Pulsed excitation was obtained from a cavity dumped dye laser, synchronously pumped by a Nd:YAG laser (Coherent Antares 76-s). Two time-dependent fluorescence decays were collected through a polarization filter mechanically rotated in the horizontal and vertical orientations. The data were stored into two memory addresses in an IBM PC. A third decay was collected at the magic angle (54.7° to the polarization orientation of the exciting beam).

In the data analysis, the fluorescence decay taken at the magic angle was fitted to two exponentials. This data represents the fluorescence decay in the absence of polarization effects and yields the fluorescence lifetime of ~22 ns, in good agreement with previous measurements for EtBr bound to oligonucleotides (22). This represented ~80% of the total decay. The complete fit required a 20% contribution of a 5–15 ns component. The parameters computed from this fit of the magic angle decay (lifetimes and pre-exponential factors) were then fixed in the simultaneous analysis of the vertically and horizontally polarized fluorescence decays, where the anisotropy function is expressed as a single exponential.

$$I_{||}(t) = 1/3\, I_0(t)[1 + 2r(t)] \quad \mathbf{3}$$

$$I_{\perp}(t) = 1/3\, I_0(t)[1 - r(t)] \quad \mathbf{4}$$

A simultaneous analysis of the polarized fluorescence decays insured an improved accuracy on the recovered parameters. The χ^2 parameter, employed for judging the quality of the fit, was optimized by our analysis software using the Marquardt–Levenberg algorithm. A value of <1.4 is reasonable, while a value >2.0 indicates a poor fit.

Care was taken in these experiments to account for any residual free EtBr and to ensure negligible depolarization due to energy transfer by minimizing the number of molecules having more than one bound EtBr (Duhamel,J., Kanyo,J., Dinter-Gottlieb,G. and Lu,P., manuscript submitted).

The viscosities (η) of the solutions were measured on a Brookfield Rheometer with a CP-40 cone. The temperature was regulated by a thermostated circulating water bath. (L buffer: η(4°C) = 1.67 cP, η(22°C) = 0.93 cP, η(40°C) = 0.65 cP; D buffer: η(4°C) = 1.92 cP, η(22°C) = 1.06 cP, η(40°C) = 0.74 cP).

^{19}Fluorine NMR spectroscopy

NMR data were obtained on a Bruker AMX500 spectrometer equipped with a 5 mm $^1H/^{19}F$ probe. The ^{19}F base frequency was 470.27 MHz, with D_2O serving as an internal lock. Typically 8192 complex data points were collected over a frequency window of 4000 Hz. A relaxation delay of 1.0 s was utilized and sufficient (16–256) transients were signal averaged to obtain an acceptable signal-to-noise ratio. The free induction decay was multiplied by an exponential window function having a time constant of 5.00 Hz prior to Fourier transformation. All ^{19}F spectra were referenced to free FdU (20 mM) in L buffer, pH 6.7, as 0.0 p.p.m.

RESULTS

Tetraloop analogs and control molecules

We have investigated the effect of solution conditions on the hairpin/duplex equilibrium (Scheme 1) for the r(CUUCGG) tetraloop and several analogs.

The oligonucleotides used in these studies are shown in Figure 2. The original RNA dodecamer is TinRNA (r-I). Thermal melting data indicate that, although the RNA C(UUCG)G tetraloop is unusually stable, its DNA analog, C(TTCG)G, is not (relative to homopolymeric tetraloops) (3). Also, the loop region of the DNA hairpin is unstructured in solution, as examined by a complete solution NMR analysis (23). There is no evidence for hydrogen bonding in the DNA loop, nor for stacking of T6 on the C7 sugar, as in the RNA. Varani points out that the U–G base pair in the RNA dodecamer involves not only a base–base hydrogen bond, but also a base–sugar hydrogen bond (to the 2′-OH of the first loop uridine residue), thus explaining to some degree the loss of stability in analogs of the hairpin containing deoxyribose sugars (10). We therefore performed parallel experiments with the DNA analog (d-I), thinking that, by removing the stabilization energy of the loop, we would be able to manipulate the transition to the double helical conformation in solution.

Before examining molecular details we monitored the solution conformation by the hydrodynamics of these molecules. We

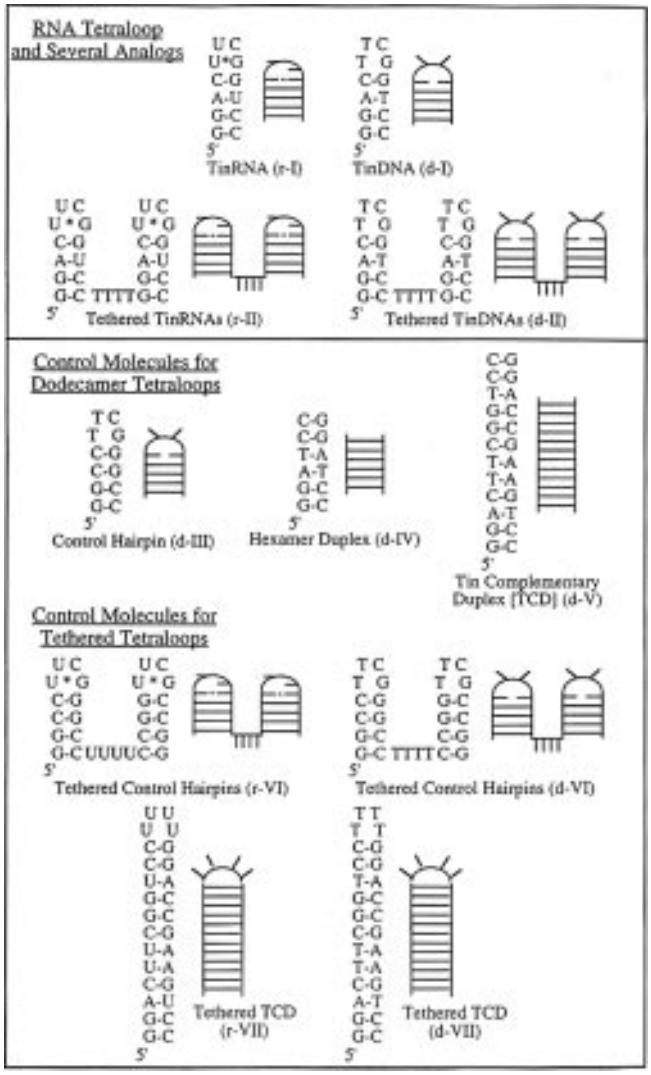
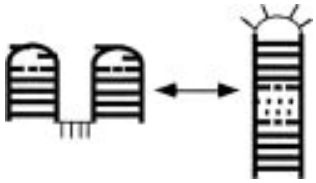


Figure 2. Tetraloop analogs and control molecules used in this publication. Sequences for the oligonucleotides, as well as a schematic for each species shown in its predominant solution conformation, are shown. RNA and DNA analogs of the same sequence have the same Roman numeral. (Tin designates the original dodecamer, since the Tinoco group was involved with both the NMR and X-ray RNA structure papers, as well as the structure determination of the DNA analog.)

synthesized molecules of defined size and shape as comparisons and controls: d-III is a dodecamer tetraloop, while d-IV and d-V are control duplexes of 6 and 12 bp respectively. The 12 bp duplex is composed of the DNA analog of the tetraloop sequence (d-I) with an equimolar amount of a fully Watson–Crick complementary strand.

Since we are looking at the equilibrium between unimolecular and bimolecular species (Scheme 1), concentration is an important consideration. A calculation of the dodecamer concentration in the crystal (180 mM) prompted the investigation of tethered species (r-II and d-II) in an effort to increase the apparent local 'dodecamer concentration'. Two RNA or DNA dodecamers were connected by a string of 4 nt. Assuming a center-to-center distance of 20–50 Å for the tethered hairpins, the calculated local tetraloop concentration is equivalent to 20–300 mM, which brackets the 180 mM seen in the crystal. This tethering not only increases the apparent local nucleic acid concentration, but also makes the hairpin to duplex transition a unimolecular event, circumventing analysis of concentration dependence. The new equilibrium is as depicted in Scheme 2, where the extended hairpin on the right mimics the crystal conformation. Molecules for size/shape comparison are also shown in Figure 2. Species r-VI and d-VI form two independent hairpins; an extended hairpin conformation is not possible since there are no Watson–Crick pairs. The opposite is true for the tethered complementary duplexes, which have 12 Watson–Crick pairs (r-VII and d-VII, the tethered version of d-V).



Hydrodynamics of tetraloop analogs

Since we are exploring large size and shape changes, we examined the hydrodynamic behavior of the oligonucleotides when the solvent was changed from that used in the solution structure determination to that of the crystal growth. Rotational correlation times (τ_{rot} values) for the RNA tetraloop and various analogs were compared with those of control molecules of defined size and shape.

The K_d values for EtBr dissociation from the DNA analogs of the dodecamer hairpin (d-I) and the fully complementary duplex (d-V) suggest that EtBr binds equally tightly to the hairpin and duplex conformations. Therefore, addition of the dye to the samples does not perturb the hairpin–dimer equilibrium.

The τ_{rot} values were calculated by fitting decays of fluorescence anisotropy to a single exponential (equation **2**), i.e. the molecules were modeled as spheres. The rotational correlation time (τ_{rot}) for a spherical molecule is directly proportional to the volume of the rotating unit (V)

$$\tau_{rot} = \eta V/RT \qquad \mathbf{5}$$

where η is the viscosity of the solution and T is the absolute temperature. The term V is both size and shape dependent.

The spherical model is a useful first approximation for following nucleic acid conformation in the range of solutions of interest using fluorescence depolarization, as demonstrated by studies of double helical DNA control molecules. Calibration curves of τ_{rot} versus size (*BP*, number of base pairs) were constructed for eight DNA duplexes ranging in length from 6 to 32 bp (axial ratios of 1:1–5:1). These lengths span the sizes of interest for the tetraloop analogs. The experiments were carried out in both L and D buffers and at three different temperatures (4, 22 and 40°C). For each set of buffer/temperature conditions, rotational times were found to be linearly proportional to the number of duplex base pairs over the range of helix lengths studied (Fig. 3A and B). When temperature and viscosity of the solutions are taken into account according to equation **5**, all of the experimental rotational times fall on a single line, as illustrated in Figure 3C. The average equivalent spherical volume occupied by 1 bp, calculated from the slope of this plot, is found to be ~1700 Å^3, consistent with the volume of hydrated DNA helices (and equivalent to the volume of a cylinder with a height of 3.4 Å and a diameter of 25 Å). Thus, the decay of fluorescence anisotropy of bound EtBr with a spherical approximation is an internally consistent approach for

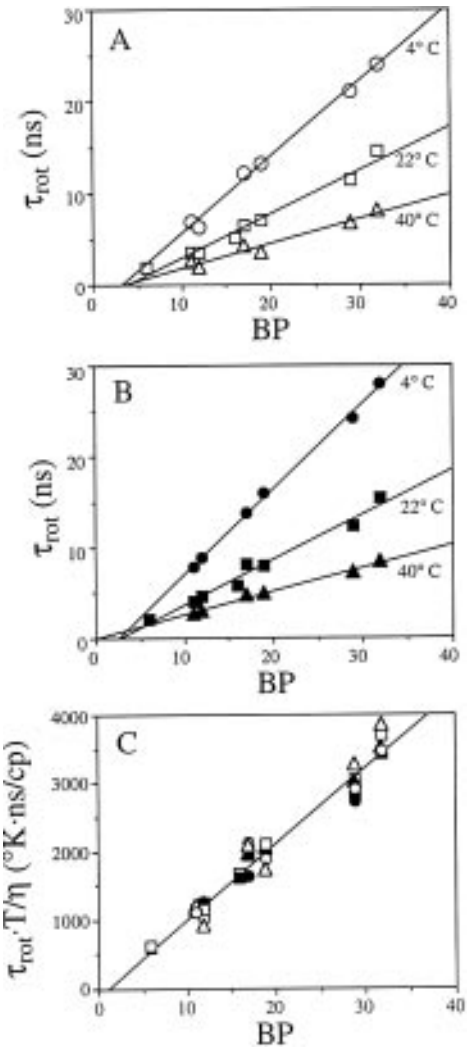


Figure 3. Hydrodynamics of DNA duplexes modeled as spheres. Anisotropy decays of short, defined DNA duplexes were fitted to a single exponential as described in Materials and Methods. The actual DNA sequences used for these curves are in Duhamel *et al*. (23). The measured values are precise to within 10%, based on reproducibility of the calculated rotational correlation times (even when using new samples). (**A**) Plot of τ_{rot} versus number of base pairs in L buffer at three temperatures. (**B**) Plot of τ_{rot} versus number of base pairs in D buffer at three temperatures. (**C**) Plot of τ_{rot} T/η versus number of base pairs for the data from (A) and (B). The convergence of the data on a straight line confirms that equation **3** is an adequate approximation.

following conformation in the solvents of interest for these short oligonucleotides.

The apparent size of the RNA dodecamer and its analogs can, in principle, be calculated from the calibration curves in Figure 3. Both the RNA (r-I) and the DNA (d-I) dodecamers (with τ_{rot} values of 2.2 and 2.0 ns respectively in D buffer at 22°C) behave like a 6 bp duplex, indicating a hairpin structure under the crystal growth conditions as well as in the solution NMR buffer.

Since we are looking for size changes and possible shape changes, particularly for the tethered tetraloop analogs (Scheme 2), the rotational correlation times of the tetraloop analogs were compared with those of reference molecules in Table 1. The RNA (r-I) (Table 1A) and DNA (d-I) (Table 1B) dodecamers are unimolecular hairpins in both L and D buffers, with rotational

Table 1. Hydrodynamics of tetraloop analogs

A	d-III	r-I	d-V
L	1.8	1.8	3.5
D+Mg	2.1	2.2	4.4
B	d-III	d-I	d-V
L	1.8	1.7	3.5
D+Mg	2.1	2.0	4.4
C	r-VI	r-II	r-VII
L	4.4	4.6	5.5
D+Mg	5.8	5.4	6.9
D+Mg+Na	8.8	8.2	11.2
D	d-VI	d-II	d-VII
L	3.0	3.2	4.5
D+Mg	3.5	3.8	5.1
D+Mg+Na	5.8	6.1	7.6

Rotational correlation time τrot in nanoseconds of the tetraloop analogs are compared with those of molecules of defined size and shape in L buffer (L), D buffer plus 2.5 mM $MgCl_2$ (D+Mg) and D buffer plus 2.5 mM $MgCl_2$ plus 2.5 M NaCl (D+Mg+Na). The measured values are precise to within 10%, based on reproducibility of the calculated rotational correlation times. The middle column shows the equilibrium of interest, with the predominant solution conformation adopted by the tetraloop analog boxed. The arrows indicate that the rotational correlation times of the tin RNA analogs are in all cases the same (within experimental error) as those of the control molecules with hairpin conformations.
(**A**) RNA tetraloop (r-I).
(**B**) DNA tetraloop (d-I).
(**C**) Tethered RNA tetraloops (r-II).
(**D**) Tethered DNA tetraloops (d-II).

correlation times equivalent to that of a control hairpin (d-III), slightly shorter than a control hexamer DNA duplex (d-IV) and ~50% that of a control fully complementary 12 bp DNA duplex (d-V).

Short RNA and DNA helices adopt the A and B forms respectively and thus may exhibit different hydrodynamics. For the RNA and DNA dodecamers (r-I and d-I) this is not of great concern, since their rotational correlation times are within 10% of each other in both L and D buffers. Therefore, these tetraloop analogs were both compared with DNA control molecules. For larger molecules, however, analogous RNA and DNA species did indeed exhibit significantly different hydrodynamics, with τ_{rot} values up to 66% different (e.g. r-VI and d-VI in D buffer have τ_{rot} values of 5.8 and 3.5 ns respectively). The tethered RNA tetraloop analogs must, therefore, be compared directly with RNA control molecules, while the tethered DNA must be compared with DNA controls. Under the solvent conditions investigated, the hydrodynamic behavior of the tethered RNA

tetraloop analog (r-II) (Table 1C) is consistent with both halves of the molecule adopting independent hairpin conformations. For example, under the crystal growth conditions minus PEG (D+Mg), the tethered RNA species (r-II) has a rotational correlation time of 5.4 ns, significantly less than that of the control tethered fully complementary duplex (r-VII, 6.9 ns) and very close to that of a control tethered hairpin, which cannot form a stable duplex (r-VI, 5.8 ns). The tethered DNA analog (d-II) also apparently does not adopt the extended hairpin conformation, but remains as two independent hairpins (Table 1D). It is interesting to note that the six tethered species are all 28 bases long, but have very different hydrodynamics.

Further evidence that the tethered tetraloop analogs form two independent hairpins comes from an analysis in which the oligonucleotides are modeled as cylinders (as in Duhamel,J., Kanyo,J., Dinter-Gottlieb,G. and Lu,P., manuscript submitted). In such a case, two rotational correlation times are calculated, where τ_1 corresponds to rotation parallel to the helical axis and τ_2 corresponds to rotation about an axis perpendicular to the helical axis. For the DNA duplexes examined in Figure 3, rotation perpendicular to the helix axis occurs with a diffusion coefficient $D_\perp$ of a cylinder with a diameter of 19 Å and a height equal to 3.4 Å times the number of duplex base pairs (plus 2.7 Å for the bound EtBr molecule). Rotation parallel to the helix axis also occurs at the rate predicted by theory for a cylinder of equivalent diameter and length. When the same analysis was applied to the tethered DNA analog (d-II), we computed the τ_2 value expected for an equivalent 14mer duplex, indicating an extended conformation, as expected by electrostatic repulsion. However, the calculated τ_1 value was approximately one third smaller than that expected if all the base pairs were coaxially stacked. This data is consistent with each half of the molecule adopting a hairpin conformation, with the hairpins having independent motions parallel to the helical axis.

The fluorescence anisotropy decay results demonstrate that increasing the apparent oligonucleotide concentration and changing the solvent from that used in the solution structure determination to that used in the crystal growth did not induce a noticable shift in the equilibrium toward the duplex conformation for either the RNA or DNA tetraloop analogs.

We note that, since we are using a monoexponential fit, we are actually calculating a weighted average of the rotational correlation times of the individual species or conformers present. We have performed experiments in order to assess the sensitivity of fluorescence depolarization to mixtures of the hairpin and duplex conformers. Mixtures with various molar ratios of TinDNA (d-I, τ_{rot} = 1.7 ns) and its fully complementary duplex (d-V, τ_{rot} = 3.7 ns) were prepared. The equilibrium constants for EtBr binding to both the hairpin and the duplex were approximately equal. When even 5% of the molecules were duplexes, the rotational correlation time calculated when the anisotropy decay was fitted to one exponential increased over 10%, beyond the range of experimental error. Therefore, we estimate that the lower limit of detectable duplexes is ~5%.

Since the RNA concentration in the crystal is 180 mM (in strands), over 2 M accompanying cations must be present to ensure charge neutrality. Therefore, we also measured τ_{rot} under high ionic strength conditions (1–2.5 M NaCl). Fluorescence depolarization provided no evidence for a change toward the duplex conformation even under these high salt conditions. However, we note that data analysis became somewhat more complicated for samples in higher salt buffers, because the EtBr–oligonucleotide association constant changes, resulting in more unbound dye, particularly for the RNA analogs. Estimates of the binding constants of EtBr to the DNA hairpin (d-I) in both buffers, calculated from a combination of UV absorption and steady-state fluorescence, indicates that the dye binds less tightly to the hairpin at higher salt concentrations.

Lest we forget, the crystal growth conditions also contained up to 32% PEG. When PEG400 was included in the solutions, the fluorescence data were too complicated to analyze, since the blank (run in the absence of RNA or DNA) could not be fitted reliably to a single exponential. Furthermore, addition of PEG to the samples induced changes in binding of the dye to our oligonucleotides, resulting in an increase in free EtBr.

^{19}Fluorine NMR spectroscopy

We turned to ^{19}F NMR to confirm and extend the fluorescence anisotropy decay data. This allowed observations in the actual solvent environment of the crystal mother liquor. Furthermore, because we have incorporated only one fluorouracil per strand, ^{19}F NMR spectra would show the presence of multiple conformations in slow exchange, which would have been hidden by the exponential fitting routines used for the fluorescence data analysis.

There are three sites (uracils or thymines) on each dodecamer which can be followed using ^{19}F NMR. We would expect the greatest change upon duplex formation for the ^{19}F nucleus at uridine 6 (see Fig. 1), which is unpaired in the solution structure and stacked as a U–C base pair in the crystal structure. In the hairpin conformation, that base is stacked on a sugar, while in the duplex conformation, it is stacked between two other bases. Finally, since the ribose of U6 is in a C2′-*endo* conformation in the RNA tetraloop, substitution with a deoxyribose residue should not be deleterious to the structure. We found that the fluorine label at that position has little effect on either the RNA (r-I.F6) or the DNA (d-I.F6) structure by comparison of NOESY cross-peaks in the aromatic H6/H8 to sugar H2′/H2″ region for the labeled versus unlabeled oligonucleotides (data not shown).

We looked for evidence of a conformational change in the oligonucleotide when the solvent was changed in steps from that used in the solution NMR structure determination to that of the crystal growth. Under the various conditions, spectra were acquired as a function of pH (6–10 range) for the fluorinated RNA and DNA dodecamers (r-I and d-I), for the tethered species (r-II and d-II) and for the control samples FdU and the fully complementary duplex (d-V). In this pH range each species gives one peak per fluorine probe. Therefore, we found no evidence for a double-stranded duplex coexisting with a hairpin in the 6–10 pH range in any of the buffers for either the RNA or DNA dodecamer of the Tinoco RNA sequence or for the tethered analogs. For all of the tetraloop analogs, the apparent pK_a of the imino proton of the FdU base, the midpoint of the titration curve, was shifted lower by salt addition (100 mM sodium citrate and 50 mM Tris). Addition of 2.5 mM $MgCl_2$ and 20% PEG400 brought about no further change to the titration curves. ^{19}F NMR chemical shift versus pH curves of the tethered DNA and RNA are superimposable on those of the non-tethered species. Thus, the tethering did not bring about a transition to the extended duplex conformation for either the RNA or DNA, but each half of the molecule adopted an independent hairpin structure, as the fluorescence data showed.

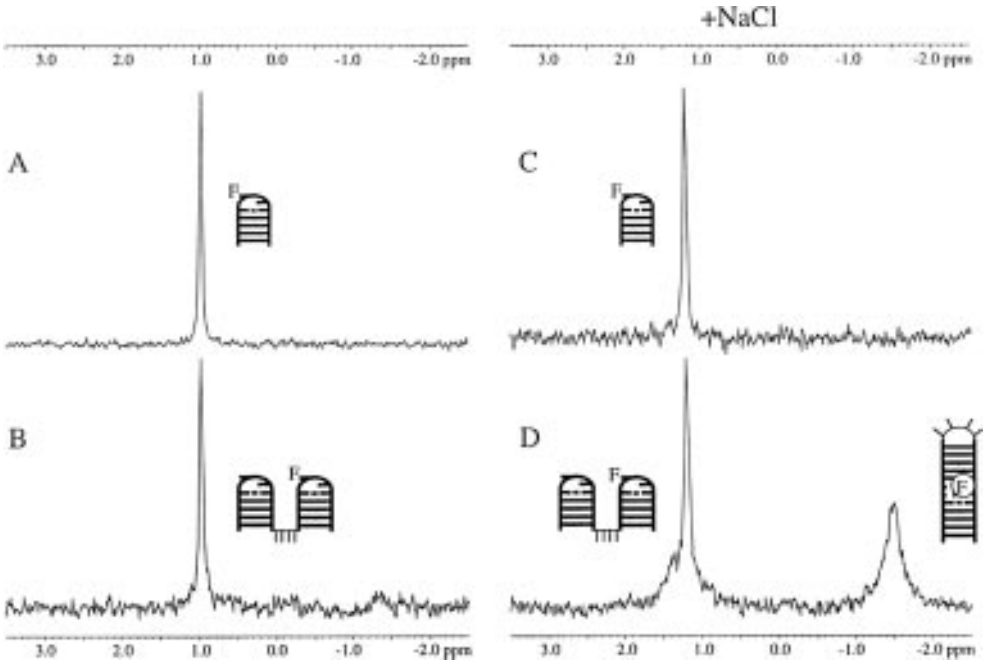


Figure 4. Selected ^{19}F NMR spectra of labeled RNA tetraloop analogs collected and processed as in Materials and Methods, referenced to free FdU in L buffer, pH 6.7, as 0.0 p.p.m. Although the normal pH of D buffer is 8.5, spectra taken at pH 6.7 are presented. The results are the same, but at the higher pH all resonances are broadened due to exchange of the FdU imino proton. (**A**) The original RNA sequence labeled at position U6 (r-I.F6) in D buffer at pH 6.85. (**B**) The tethered RNA tetraloops (r-II.F6′) in D buffer, pH 6.70. (**C**) The original RNA sequence labeled at position U6 (r-I.F6) in D buffer plus 2.5 M NaCl, pH 6.61. (**D**) The tethered RNA tetraloops (r-II.F6′) in D buffer plus 2.5 M NaCl, pH 6.69. Integration of the area under each peak indicates that the second component comprises 40% of the sample.

Since over 2 M monovalent cations must accompany the RNA in the crystal to ensure charge neutrality, we subjected tethered RNA and DNA species, geometrically equivalent to ~100 mM strands, to 2.5 M NaCl in D duffer. Selected RNA spectra are shown in Figure 4. Upon tethering two RNA analogs and adding 2.5 M NaCl, the RNA shows a significant second component at –1.5 p.p.m. (Fig. 4D) not seen under identical conditions for the RNA dodecamer (Fig. 4C), indicating that the tethering to increase the apparent concentration of the dodecamer was necessary to obtain the second spectral component. The DNA analog of the tethered sequence shows a broad second component upfield, although the effect is less pronounced than for the tethered RNA. The relative proportions of the two components is not altered in 20% PEG (data not shown). To summarize, evidence for a second solution conformation, presumably the extended hairpin which mimics the double helix in the crystal, was obtained only after tethering two dodecamers together in order to increase the local oligonucleotide concentration and after the addition of 2.5 M NaCl to mimic the cation concentration in the crystal.

DISCUSSION

We demonstrate that an RNA CUUCGG tetraloop and its DNA analog maintain hairpin conformations in a continuum of solvent conditions, including conditions closely approximating those used to grow crystals. This contrasts with the double helices observed by single crystal X-ray diffraction.

In order to mimic oligonucleotide concentrations in the crystal, we constructed tethered RNA and DNA species, geometrically equivalent to ~100 mM strands. Although the tethered RNA remains as two independent hairpins under both the solution NMR and crystallization conditions, a significant second component, presumably the extended hairpin which mimics the double helical conformation in the crystal, appears in our NMR spectra if additional NaCl (2.5 M) is added to the crystallization buffer. The same holds true, although to a lesser extent, for the tethered DNA analog. Although counterion condensation theory predicts that the cation concentration at the surface of the DNA helix is constant, regardless of bulk ion concentration, Honig and Nicholls (24) remind us that the distribution near the molecular surface (which is a function of bulk ion concentration) does play a role in polynucleotide conformation.

Careful attention must be paid to the conditions under which nucleic acid structures are solved. There are at least two other cases where related RNA tetraloops show double helices in the diffraction-derived structure. Fujii *et al.* (20) claimed that r(UGAGCUUCGGCUC) is a double helix with 5′ U residues unpaired in their crystal structure and a hairpin with an unpaired 5′ U in solution when examined by NMR spectroscopy. Although solvent conditions were not given for either experiment, they note that the ‘duplex is favored by high RNA and salt concentrations’, an identical conclusion to that drawn by Holbrook *et al.* (18). Similarly, the sequence r(GCUUCGGC) is a hairpin in solution and a double helix in the crystal form, even though it is short and thus presumably has less base pairing stabilization in the stem (21). A similar dimorphism occurs in the case of DNA. X-ray diffraction (using 20 mM potassium cacodylate, pH 7.0, 10 mM $MgCl_2$, 6 mM spermine, 40 mM KCl and 5–40% 2-methly-2,4-pentane diol) and solution NMR (with 50 mM NaCl, pH 7.0) yielded structures for the G quartet with very different pathways for the phosphodiester backbone (25,26).

Holbrook and co-workers (18) state that high RNA and salt concentrations favor an intermolecular double helix over an intramolecular hairpin loop conformation, without suggesting how high the concentrations should be. Our experiments demonstrate that although high RNA and salt concentrations do shift the solution equilibrium toward the bimolecular double helix, they are not sufficient to make the duplex the major conformation. The duplex conformation adopted by the RNA dodecamer in the crystal must also be a consequence of crystal contacts. Since the predominant species under all solution conditions is the hairpin, we believe that the duplex conformation seen in the crystal results from crystallization of a secondary component of the crystallization mixture. Perhaps the ability of the duplexes to stack end-to-end in an orderly array selectively precipitates that conformation.

REFERENCES

1 Nikonowicz,E.P. and Pardi,A. (1992) *Nature*, **355**, 184–186.
2 Batey,R.T., Inada,M., Kujawinsky,E., Puglisi,J.D. and Williamson,J.R. (1992) *Nucleic Acids Res.*, **20**, 4515–4523.
3 Dieckmann,T. and Feigon,J. (1994) *Curr. Opin. Struct. Biol.*, **4**, 745–749.
4 Allain,F.H.-T. and Varani,G. (1995) *J. Mol. Biol.*, **250**, 333–353.
5 Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994) *Nature*, **372**, 68–74.
6 Marino,J.P., Gregorian,R.S.,Jr, Csankovszki,G. and Crothers,D.M. (1995) *Science*, **268**, 1448–1454.
7 Chang,K.-Y. and Tinoco,I.,Jr (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 8705–8709.
8 Gubser,C.C. and Varani,G. (1996) *Biochemistry*, **35**, 2253–2267.
9 Oubridge,C., Nobutoshi,I., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) *Nature*, **372**, 432–438.
10 Uhlenbeck,O.C. (1990) *Nature*, **346**, 613–614.
11 Woese,C.R., Winkler,S. and Gutell,R.R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8467–8471.
12 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.
13 Antao,V.P., Lai,S.Y. and Tinoco,I.,Jr (1991) *Nucleic Acids Res.*, **19**, 5901–5905.
14 Jaeger,J A., Turner,D.H. and Zuker,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7706–7710.
15 Cheong,C., Varani,G. and Tinoco,I.,Jr (1990) *Nature*, **346**, 680–682.
16 Varani,G., Cheong,C. and Tinoco,I.,Jr (1991) *Biochemistry*, **30**, 3280–3289.
17 Moazed,D., Stern,S. and Noller,H.F. (1986) *J. Mol. Biol.*, **187**, 399–416.
18 Holbrook,S.R., Cheong,C., Tinoco,I.,Jr and Kim,S.-H. (1991) *Nature*, 353, 579–581.
19 Fujii,S., Tanaka,Y., Tomita,K., Sakata,T., Hiroaki,H., Tanaka,T. and Uesugi,S. (1991) *Nucleic Acids Symp. Ser.*, **25**, 181–182.
20 Fujii,S., Tanaka,Y., Uesugi,S., Tanaka,T., Sakata,T. and Hiroaki,H. (1992) *Nucleic Acids Symp. Ser.*, **27**, 63–64.
21 Cruse,W.B.T., Saludjian,P., Biala,E., Strazewski,P., Prang,T. and Kennard,O. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4160–4164.
22 Lakowicz,J.R. (1983) *Principles of Fluorescence Spectroscopy.* Plenum Press, New York, NY.
23 James,J.K. and Tinoco,I.,Jr (1993) *Nucleic Acids Res.*, **21**, 3287–3293.
24 Honig,B. and Nicholls,A. (1995) *Science*, **268**, 1144–1149.
25 Smith,F.W. and Feigon,J. (1992) *Nature*, **356**, 164–168.
26 Kang,C., Zhang,X., Ratliff,R., Moyzis,R. and Rich,A. (1992) *Nature*, **356**, 126–131.