The solution structure of a d[C(TTCG)G] DNA hairpin and comparison to the unusually stable RNA analogue

Joyce K. James and Ignacio Tinoco, Jr*

Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, CA 94720, USA

Received February 22, 1993; Revised and Accepted May 14, 1993

ABSTRACT

The solution structure of the DNA analogue of the unusually stable r[C(UUCG)G] RNA hairpin, 5'-d[GGA-C(TTCG)GTCC]-3', has been determined by NMR spectroscopy, and its structure has been compared to that of the RNA molecule. The RNA molecule is compact and rigid with a highly structured loop. However, the DNA molecule is much less structured. The DNA hairpin contains a B-form stem of four base pairs. The terminal base pair frays, and the 3'-terminal nucleotides, C11 and C12, are in equilibrium between 2'-endo and 3'-endo conformations. Unlike the RNA loop, the DNA loop contains no syn nucleotides, and there is no evidence for base-base or base-phosphate hydrogen bonding in the loop. The loop is flexible, and reveals no specific internucleotide interactions.

INTRODUCTION

The tetranucleotide loop sequence, r[C(UNCG)G], is exceptionally common in ribosomal and messenger RNAs (1-3). This motif displays unusual thermodynamic stability (2,4) and is thought to be important as a nucleation site for RNA folding or as a protein recognition site (5). However, although hairpin structures do exist in naturally occurring DNAs (cruciforms), exceptionally stable tetraloops are not known to exist.

Previous studies on the loop sequence in RNA, r[C(UNCG)G], or DNA, d[C(TNCG)G], have not revealed the structural differences between the RNA and DNA hairpins. Sakata *et al.*'s, (6) thermodynamic comparison of synthetic oligonucleotides revealed a relation between enhanced stability and nucleic acid type. However, an analysis (5,7) of the solution structure of the RNA hairpin yielded results that do not fully explain Sakata *et al.*'s data. Sakata *et al.* (6), found several factors were important for enhanced thermodynamic stability: an A-form stem and/or the lack of T methyl groups in the loop, the presence of the G amino group in the loop, and the presence of the sugar hydroxyl groups in the loop. The NMR analysis (5,7), however, determined that while the RNA hairpin was highly structured, there was no evidence from either the NMR data or the structural modeling that the G amino group or the sugar hydroxyl groups

were involved in specific structural interactions. An equivalent structure was possible for the DNA analogue. Previous work on synthetic DNA hairpins demonstrates that loop structure in DNA, like RNA, is sequence dependent, and base pairing in DNA tetraloops has been observed (8). Therefore, the question of whether the specific structure of the RNA hairpin could explain the thermodynamic properties still remained. It is conceivable that the RNA and DNA molecules form similar structures, but RNA is simply inherently more stable. RNA helices are more stable thermodynamically than DNA helices (9).

In this report we present the solution structure of the DNA hairpin, 5'-d[GGAC(TTCG)GTCC]-3' (numbering system 1 through 12, 5' to 3') determined by applying NMR-derived inter-proton distance and torsion angle constraints to dynamical simulated annealing and energy minimization. The structure is then compared to the RNA analogue. Experimental data coupled with modeling results reveal the loop is highly dynamic and nonstructured. Each structure obtained from the modeling contains a unique and significantly different loop conformation. Therefore, the precision with which this molecule can be described is reduced compared with the RNA hairpin.

MATERIALS AND METHODS

The oligodeoxynucleotide was synthesized on an Applied Biosystems DNA synthesizer using standard phosphoramidite chemistry. The DNA was deprotected and detached from the support beads by incubation in 28%-30% ammonium hydroxide at 55°C overnight. The ammonia was removed by vacuum (Savant Speedvac), and the sample was rinsed twice with doubly distilled water and concentrated under vacuum. The DNA was purified on a 20% polyacrylamide denaturing gel (tris-borate buffer, 8 M urea, pH 8.5), run at 40 W for 7 hrs. The slowest and widest band was visualized with a UV light source and cut from the gel for electroelution. The eluted samples were pooled, ethanol precipitated twice, and redissolved in water. The sample volume was reduced under vacuum and microdialyzed into 10 mM sodium phosphate and 10 mM EDTA, pH 6.9. The sample was about 2 mM in DNA concentration when dissolved in 600 µL of buffer. The oligodeoxynucleotide exists as a hairpin (or

* To whom correspondence should be addressed

single strand) under these conditions. Even in 1 M NaCl the concentration dependence of the melting curves gave no evidence for double strand formation (4).

For all nonexchangeable proton experiments, the dialyzed sample was lyophilized to dryness and dissolved in 99.99% D₂O (Altec) three times, then resuspended under nitrogen in 99.99% D₂O to a final volume of 600 μ L. The NMR sample tube was stored at 4°C in a sealed bottle containing Drierite between experiments. For exchangeable proton experiments, the sample was lyophilized to dryness and resuspended in 90%/10% H₂O/D₂O to a final volume of 600 μ L. Experiments were done on a Bruker AMX 600 MHz spectrometer operating at 600 MHz for protons and 242 MHz for phosphorus. Data were processed with FTNMR.

One-dimensional spectra in D₂O were collected using presaturation of the HDO peak. Nonexchangeable, two-dimensional spectra were collected in the phase-sensitive mode with the TPPI method (10) and presaturation of the residual HDO resonance. A long mixing time NOESY (400 ms) was used for assigning the spectrum. 2K points were collected in t_2 and 400 in t_1 , at 35°C, with a sweep width of 5,000 Hz. NOE buildup NOESY were collected at 15°C, with various mixing times (80, 120, 200 ms). 2K points were collected in t_2 and 360 in t_1 , with a 5,000 Hz sweep width. The data were zero-filled to 1K real points in t_1 and skewed sine bell apodization was applied in both dimensions (30°–50° phase shift).

A double-quantum filtered, phosphorus-decoupled COSY (DQF COSY) was collected at 25°C. 4K points were collected in t_2 and 900 in t_1 , with a sweep width of 4,808 Hz. The data were zero-filled to 2K real points in t_1 for a resolution of 2 Hz per point. Sine bell apodization was applied (30°).

A Purged Exclusive COSY (P. E. COSY) was collected at 35°C in hypercomplex mode. 4K points in t_2 and 900 (450 complex points) in t_1 were collected, with a sweep width of 4,808 Hz. The data were zero filled to 4K in both dimensions for a resolution of 1 Hz per point. Sine bell apodization was applied (50°).

A ¹H-³¹P heteronuclear COSY was collected at 25°C on a Bruker 400 MHz spectrometer (operating at 162 MHz for phosphorus). 4K points were collected in t_2 and 168 in t_1 , with a sweep width of 1,362 Hz in t_2 and 1,600 Hz in t_1 . The data were zero-filled to 4K real points in t_1 and skewed sine bell apodization was applied (30°).

NOESY spectra in H₂O were collected using the 11 jump-return sequence (11) with a 100 ms mixing time, at 5°C. 4K points in t_2 and 288 in t_1 were collected, with a sweep width of 10,870 Hz. The data were zero-filled to 4K real points in t_1 and skewed sine bell apodization was applied (65°).

Interproton distances between nonexchangeable protons were estimated from NOE buildup rate analysis of NOESY spectra collected with mixing times of 80, 120, and 200 ms. NOE volume integration was done with FTNMR. The buildup rates for the cytosine H5-H6 protons agreed for three of the four cytosines in the molecule, C4, C7, and C12. The average value for those three was used to reference all distances. However, the lack of agreement for all the H5-H6 buildup curves is an indication that the assumption of a single correlation time for the molecule is incorrect.

To account for the uncertainties in the distances obtained, distance ranges were used during modeling. The NOEs were classified as strong, medium, weak, and not present. The distance ranges were 1.8–3.0 Å, 2.0–4.0 Å, 2.5–5.0 Å, and >3.5

Å, respectively, which were the same ranges used in the modeling of the analogous C(UUCG)G hairpin (5,7), except the not present NOEs for that molecule were classified as >4 Å. Since the intensity of NOEs involving exchangeable protons also rely on the exchange rate, distance constraints were set to 2.0–5.0 Å between protons when cross peaks were present. No constraints were applied for absent NOEs involving exchangeable protons.

¹H-¹H scalar couplings were measured from high resolution DQF-COSY and P. E. COSY. The sums of the scalar couplings (Σ_i) were measured instead of individual couplings (J_{ij}), except for the H1'-H2' and H1'-H2'' couplings. The coupling data were then analyzed using a graphical approach (12). $\Sigma_{1'}$, $\Sigma_{2'}$, $\Sigma_{2''}$, $J_{1'2'}$, and $J_{1'2''}$ values were used. From the analysis, the five torsion angles that describe the deoxyribose sugar ring, ν_0 – ν_4 , were determined. An error margin of $\pm 5^\circ$ was used during modeling to account for uncertainties in the measurements.

A ¹H-³¹P heteronuclear COSY provided qualitative information about the backbone torsion angles. The phosphorus resonances were clustered in an 0.7 ppm range around –4.2 ppm, referenced to TMP at 0 ppm. The lack of downfield-shifted resonances in the phosphorus dimension was taken as evidence that none of the α (O3'-P-O5'-C5') or ζ (C3'-O3'-P-O5') backbone torsion angles adopted the trans conformation. These two angles were therefore constrained to the gauche conformation ($0^\circ \pm 120^\circ$). In DNA, the backbone angle ϵ (C4'-C3'-O3'-P) can adopt either the trans or the gauche[–] conformation and they can not be distinguished from one another. Thus, ϵ is constrained to $-120^\circ \pm 120^\circ$. If the backbone angles β and γ adopt the trans, gauche⁺ conformation respectively, the P and H4' nuclei adopt a 'W' conformation and the P to H4' coupling is observable (13). Additionally, the H4' to H5' and H4' to H5'' couplings are small, due to the conformation about γ , and the P to H5' and P to H5'' couplings are small and similar due to the conformation about β (13). This combined effect yields symmetrical P to H5' and P to H5'' cross peaks in the ¹H-³¹P COSY spectrum with the $\Sigma 5'$ and $\Sigma 5''$ couplings of magnitude of ~ 20 Hz or less (the geminal H5' to H5'' coupling is ~ 12 Hz). The presence of P to H4' cross peaks, and of symmetrical P to H5' and H5'' cross peaks of similar magnitude, indicate β (P-O5'-C5'-C4') adopts the trans conformation ($180^\circ \pm 60^\circ$), while γ (O5'-C5'-C4'-C3') is gauche⁺ ($60^\circ \pm 60^\circ$) for most of the residues. The backbone angles β and γ could not be constrained for four of the nucleotides due to severe overlap in the spectrum.

The X-PLOR package, version 2.0 (14), was used for modeling the structure of the DNA hairpin. Interproton distance and torsion angle constraints were applied during dynamical simulated annealing with restrained molecular dynamics followed by energy minimization. Since the terminal base pair is dynamic and the C12 sugar is in C2'-endo-C3'-endo equilibrium, few experimental constraints could be determined for the terminal base pair. Therefore, artificial internucleotide distance and intranucleotide torsion angle constraints were applied to the terminal base pair to ensure a B-form closing base pair. In addition, distance constraints corresponding to the H-bond distances between base paired nucleotides were also included and were the only H-bonding constraints applied. No other artificial constraints were applied. Randomized starting structures were generated, and the molecule was folded in several steps. First, a global fold was done by applying the NMR-determined distance constraints while performing dynamical simulated annealing and energy minimization. The distance constraints were applied in two steps;

Table 1. Assignments for the DNA hairpin in 10 mM sodium phosphate, 10 mM EDTA, pH 6.9*

	G1	G2	A3	C4	T5	T6	C7	G8	G9	T10	C11	C12
H8/6	7.84	7.89	8.22	7.31	7.61	7.55	7.51	7.86	7.95	7.44	7.69	7.78
H2/5	---	---	8.09	5.36	---	---	5.82	---	---	---	5.84	5.95
ME	---	---	---	---	1.86	1.84	---	---	---	1.46	---	---
H1'	5.70	5.63	6.31	6.07	6.10	6.04	5.91	5.68	5.95	6.20	6.22	6.31
H2'	2.45	2.74	2.71	2.10	2.27	2.10	1.81	2.67	2.73	2.27	2.30	2.31
H2''	2.63	2.81	2.90	2.42	2.50	2.31	2.22	2.74	2.80	2.59	2.55	2.36
H3'	4.82	5.03	5.07	4.82	4.85	4.63	4.62	4.86	4.93	4.93	4.89	4.58
H4'	4.18	4.39	4.48	4.23	4.25	4.16	3.69	4.29	4.41	4.31	4.24	4.11
am*	---	---	---	8.37	---	---	---	7.52	---	---	8.64	8.45
am	---	---	---	6.89	---	---	---	6.81	---	---	7.16	7.38
im	12.94	13.06	---	---	---	---	---	---	12.82	13.86	---	---
5'-P	---	-3.97	-4.06	-4.14	-4.16	-4.24	-4.42	-4.24	-3.92	-4.39	-4.10	-3.92

Non exchangeable resonances are assigned at 35°C referenced to HDO at 4.66 ppm, exchangeable resonances at 5°C referenced to H₂O at 5.06 ppm, and phosphorus resonances at 25°C referenced to TMP at 0 ppm. Italic type for the C7 amino assignments indicates uncertainty and that neither amino is involved in H-bonding, AM indicates the H-bonded amino. Blank spaces indicate assignment was not possible, and --- indicates that there is no proton of that type for that residue. The H5'/H5'' and remaining imino and amino chemical shifts are not listed as specific assignments were not made.

short range distances were applied first (N and N + 1 distances), then short and long range distances. During this round of calculations, a pure repulsive van der Waals contact and no electrostatic terms were used to reduce calculation time. Structural refinement was then achieved by applying the NMR determined torsion angle constraints as well as the distance constraints while again performing dynamical simulated annealing and energy minimization. The torsion angles were also applied in steps, in the order $\nu_0-\nu_4$, $\nu_0-\nu_4$ and γ , $\nu_0-\nu_4$, γ , β and ϵ , and finally, $\nu_0-\nu_4$, γ , β , ϵ , α and ζ . During refinement a Leonard-Jones 6-12 potential and electrostatics are included. The entire structure was then submitted to energy minimization to ensure good atomic contacts within the structure.

RESULTS AND DISCUSSION

NMR analysis

The nonexchangeable proton spectrum was assigned from a 400 ms mixing time NOESY experiment referenced to the HDO signal at 4.66 ppm (T = 35°C). At this temperature, overlap in the aromatic region is minimized, as determined from one-dimensional spectra collected as a function of temperature. The aromatic H8, H6, and H5 resonances and the sugar H1', H2', and H2'' resonances were assigned by following connectivity pathways through the H8/H6 to H5/H1' and H8/H6 to H2'/H2'' regions. These two regions were taken together to ensure consistency in the assignment patterns of the H8 and H6 resonances. These assignments were consistent with other regions of the spectrum and used to assign the H3' and H4' resonances. Additionally, a DQF-COSY experiment verified the H2', H2'' and cytosine H5 resonances. All of the non-exchangeable protons except for the H5' and H5'' protons were assigned. Several exchangeable amino and imino assignments were made from a NOESY experiment collected in H₂O (τ_m = 100 ms, T = 5°C, referenced to H₂O at 5.06 ppm). For G·C base pairs, the C(H5)-C(amino)-G(imino) pathway allowed the assignment of the three stem G(imino) protons. The base paired A·T A(H2)-T(imino) pathway identified the stem T(imino) proton. Three upfield loop iminos were observed, but a lack of identifiable NOEs to these resonances prevented assignment of the loop imino protons. The G and A aminos were not observed. The phosphorus

Figure 1. The aromatic H8/H6 to sugar H1' region of the 400 ms mixing time NOESY spectrum at 35°C. Aromatic H8/H6 resonances are assigned. The solid line indicates the connectivity pathway between the base and sugar protons.

chemical shifts were assigned from a ¹H-³¹P COSY (25°C) through the H3'-P cross peaks, referenced to TMP at 0 ppm. Table 1 lists the assigned resonances for the DNA hairpin.

At T = 35°C and τ_m = 400 ms, an assignment connectivity pathway could be followed in the H8/H6-H1' region originating at G2 and continuing to C4. A connectivity could be traced from G9 to C11 as well. Intranucleotide H8/H6-H1' NOEs were present for the remaining residues (Figure 1). The H8/H6-H2'/H2'' region of the NOESY spectrum provided a more complete connectivity pathway, originating at G1 and continuing to C7, and from G8 to C12 (Figure 2).

NOESY spectra were collected as a function of the mixing time, t_m , for NOE buildup analysis and distance determination. These spectra were collected at 15°C to reduce the internal loop

dynamics suggested by the lack of internucleotide NOEs in the loop. Lower temperatures slow the dynamics but increase the spectral overlap; the H8 resonances of G1, G2, and G8 collapse into a single peak by 5°C. Although the resonances shift at the new temperature, the resonance patterns are the same. The connectivity pathway described above could be followed, except for the step between T10 and C11, which was obscured by overlap in the spectrum. A connectivity between C11 and C12 appears at this temperature, however. Mixing times of 80, 120, and 200 ms were used.

Figure 2. The aromatic H8/H6 to sugar H2'/H2'' region of the 400 ms mixing time NOESY. Aromatic H8/H6 resonances are assigned. The solid line is the connectivity pathway traced in this region.

Figure 3. The NOE buildup rate for the C(H5-H6) protons is given. Data were collected at 15°C. Mixing times of 80, 120, and 200 ms were used. Note the buildup rate for C11 is larger than for the remaining cytosines. This is evidence that the loop nucleotides and the terminal base pair are more dynamic than the nucleotides in the stem. The average value for C4, C7, and C12 was used in the distance calculations and distance ranges were used during modeling which reflect and account for the uncertainty in the determined values.

A single correlation time for the molecule was assumed for NOE buildup analysis using the simple equality, $\text{NOE}_{ij}/\text{NOE}_{\text{known}} = r_{\text{known}}^6/r_{ij}^6$. Buildup curves were first calculated for the cytosine H5-H6 NOEs. The buildup rate for C11 is 20% larger than the buildup rates for the remaining C residues (Figure 3), indicating that the assumption of a single correlation time is not correct. The difference in the buildup rates could be due to dynamics in the loop and at the terminus of the stem. The average value for the initial slope of the buildup curves for C4, C7, and C12 was used as the standard, known value in this analysis. Interproton distance ranges were determined for the DNA hairpin, which reflect and account for the uncertainty in the calculated distances from this analysis. Ranges used were strong (1.8–2.5 Å), medium (2.0–4.0 Å), weak (3.5–5.0 Å). In some nonambiguous cases where two protons are clearly not close (where dynamic processes could not be responsible for the lack of an NOE), a distance restraint of >3.5 Å was used.

Figure 4. The H₂O NOESY spectrum of the imino to amino region at 5°C. In ω_2 , the G(imino) resonances are identified. In ω_1 , the A(H2) and the non-H-bonded C(amino) resonances are identified. Three upfield imino resonances (between 10.6 to 11.3 ppm) corresponding to the loop nucleotides were also observed, but not assigned.

Table 2. Coupling constants (Hz) for the DNA hairpin*

	$\Sigma H_{1'}$	$\Sigma H_{2'}$	$\Sigma H_{2''}$	$J_{1'2'}$	$J_{1'2''}$
G1	14.5	29.8	19.9	9.5	5.0
G2	14.5	30.1	19.3	9.5	5.0
A3	14.2	29.1	21.6	9.5	4.7
C4	14.4	30.0	22.3	8.8	5.6
T5	14.3	28.2	20.7	8.8	5.6
T6	14.1	28.5	20.5	9.4	4.7
C7	15.2	29.4	22.0	9.2	6.0
G8	14.3	—	—	—	—
G9	14.0	27.3	22.0	9.7	4.3
T10	14.5	39.0	21.7	—	—
C11	13.9	28.0	25.7	—	—
C12	12.4	—	—	—	—

*The values in *italic* are uncertain. Entries containing — indicate that the coupling constants could not be determined.

The NOESY spectrum collected in H₂O provided information about H-bonding in the molecule. Downfield imino chemical shifts (12.82–13.86 ppm) and NOE connectivities between the C(amino) and G(imino) protons of G1 and C12, G2 and C11, and G9 and C4, and between the A3(H2) proton to the T10(imino) proton, provided evidence that base pairing was occurring in the stem of the hairpin (Figure 4). There were three imino resonances

Figure 5. Coupling constants measured from the DQF COSY are plotted on graphs of coupling constants versus P , the pseudorotation angle, for two values of the pucker amplitude: $\phi_m = 35^\circ$ (solid line) and $\phi_m = 40^\circ$ (dotted line). The sugar rings in the hairpin are B-form, except for C11 and C12, which are in C2'-endo/C3'-endo equilibrium. The graph was taken from Rinkel & Altona (9).

Figure 6. The ^1H - ^{31}P heteronuclear COSY at 25°C. The assigned COSY peaks correspond to the P resonances.

between 10.6 and 11.3 ppm, however, that correspond to the imino protons of T5, T6, and G8. These imino protons are not involved in base pairing interactions, as indicated by the chemical shift.

A COSY analysis was done for the hairpin to gather information about torsion angles. The graphical analysis designed by Rinkel & Altona (12) was applied. Sums of couplings, Σ_i , were measured instead of the individual couplings, except for $J_{1'2'}$ and $J_{1'2''}$, which were determined both from this experiment and a P. E. COSY. Table 2 contains the coupling constants determined. The sums of the couplings are plotted versus the pseudorotation angle, P , as a function of the sugar pucker amplitude, ϕ , yielding the values of these parameters (Figure 5). P and ϕ were then converted to the torsion angles ν_0 – ν_4 ($\nu_n = \phi \cos[P + 144(n-2)]$ where $\nu = 0, 1, 2, 3, 4$), which also describe the sugar ring conformation (as well as the backbone angle δ), and can be used as constraints during modeling. It was determined that the sugars in the molecule all adopt the C2'-endo conformation, except for C11 and C12, which are equilibrium mixtures of C2'-endo and C3'-endo conformations (~ 50% C2'-endo). This provides further evidence that the terminal base pair is dynamic.

To determine the conformation of the backbone a ^1H - ^{31}P COSY was collected (Figure 6). The ^{31}P chemical shifts are all clustered within a 0.7 ppm range about -4.2 ppm referenced to TMP at 0 ppm, indicating all of the α and ζ torsion angles adopt the gauche conformations. The presence of P-H4' couplings and symmetrical P-H5' and P-H5'' couplings indicated that β is trans and γ is gauche⁺ for most of the residues. This region of the spectrum, however, is severely overlapped and β and γ could not be determined for four of the nucleotides. The angle ϵ was constrained to the trans and gauche⁻ conformations, as both can be adopted by this angle, and they can not be distinguished from one another due to the symmetry of the P to the H3' proton in both conformations. Table 3 contains the conformational ranges determined for the torsion angles.

From the NMR analysis, the following structural characteristics for the DNA hairpin can be described. This molecule forms a hairpin under the conditions of 2 mM DNA concentration in the presence of 10 mM Na⁺, as determined from H₂O NOESY spectra. Downfield imino and amino chemical shifts and interstrand NOE connectivities provide evidence of H-bonding interactions between the stem nucleotides. However, upfield imino chemical shifts of the loop residues indicate that there are no H-

Table 3. The structural constraints determined from NMR experiments*

	P°	ϕ°	%S	α°	β°	γ°	ϵ°	ζ°	χ°	ν_0°	ν_1°	ν_2°	ν_3°	ν_4°
G1	179	40	---	na	na	na	t,g-	g	-148	-13	33	-40	32	-12
G2	174	40	> 90	g	t	g+	t,g-	g	-90	-16	35	-40	30	-8
A3	186	40	> 85	g	t	g+	t,g-	g	-100	-8	30	-40	35	-16
C4	193	40	> 90	g	t	g+	t,g-	g	-90	-4	26	-39	37	-21
T5	189	40	> 90	g	t	g+	t,g-	g	-70	-6	28	-40	36	-18
T6	182	40	---	g	t	g+	t,g-	g	-80	-11	32	-40	33	-14
C7	193	40	> 85	g	t	g+	t,g-	g	-90	-4	26	-39	37	-21
G8	198	---	---	g	---	---	t,g-	g	-130	0	24	-38	38	-24
G9	198	40	> 85	g	---	---	t,g-	g	-90	0	24	-38	38	-24
T10	187	40	> 90	g	t	g+	t,g-	g	-100	-8	29	-40	35	-17
C11	144	40	~ 50	g	---	---	t,g-	g	-90	-32	40	-32	12	12
C12	---	---	~ 50	g	---	---	na	na	anti	---	---	---	---	---

*Values are given in degrees. Conformational ranges were determined for the backbone angles (g = gauche and t = trans). Dashed lines indicate values that could not be determined and na = not applicable.

bonding interactions in the loop; there are four bases in the loop, T5, T6, C7, and G8, as predicted from the sequence. There are no intense H8/H6-H1' NOEs. Therefore, all of the glycosidic torsion angles adopt the anti conformation. The stem is B-form, with a terminal base pair that is somewhat dynamic, based on NOE buildup and sugar conformation analysis. Breaks in the H8/H6-H1' connectivity pathway that are not present in the H8/H6-H2'/H2'' pathway indicate that irregular stacking

Figure 7. The nine accepted structures with the stems superimposed. Conformational deviations in the stem are less than 1 Å RMSD. The loop is highly variable, with 2–11 Å RMSD/nucleotide. The flexibility depicted in this figure is probably an accurate representation of this molecule. However, the possibility of a single loop conformation that can not be defined from NMR determined constraints can not be ruled out.

interactions continue into the loop from the 5'-end of the stem through to C7, but the loop is largely unstructured and at least partially dynamic. Finally, an almost total lack of internucleotidal NOEs to G8 indicate that either this residue is far from C7 and G9, or it is highly dynamic.

Molecular modeling with X-PLOR

Molecular modeling was done on the hairpin using the X-PLOR package (version 2) by Axel Brünger (14). Eleven modeling runs were done, of which all were energetically reasonable. Three of these eleven, however, violated too many experimental constraints (> 6 violations) to be accepted as having reached convergence. The three rejected models were resubmitted to refinement and minimization, after which one satisfied both the experimental constraints and energy requirements satisfactorily, one diverged, and the third remained essentially the same. Thus, nine structures were accepted.

The accepted structures were compared and the stems for all nine structures superimpose to within less than 1 Å RMSD (Figure 7). If the stem and loop regions are superimposed together, however, the fit for the stem is about 2 Å RMSD, and the fit for the loop nucleotides varies between 2 to 11 Å RMSD/nucleotide. The nucleotides in the loop are underconstrained due to the lack of internucleotide NOEs between these residues. This is probably due to the inherent flexibility of the loop, and thus is reflected quite well in the modeling. It is also possible that the loop adopts a single conformation that is not defined by the data because of the 5 Å distance limitation of NMR. This is probably not the case, as the cytosine H5-H6 buildup curves provide direct evidence for dynamics in the loop.

The average values for the torsion angles, pseudorotation angles, and pucker amplitudes for the nine accepted structures are given in table 4. The standard deviations reflect the degree to which the data constrained the angles. Large ranges were used during modeling for the backbone angles α , β , γ , ϵ , and ζ (Table 3). Therefore, large standard deviations are not surprising. The backbone angle, δ , however, was defined by the sugar angles, $\nu_0 - \nu_4$, which were constrained to $\pm 5^\circ$. Therefore, the

Table 4. The average values for the torsion angles, the pseudorotation angle, and the pucker amplitude in degrees (average/standard deviation)*

	α	β	γ	δ	ϵ	ζ	χ	Π	ϕ
G1	—	—	193/103	150/.7	175/1	—112/4	—107/1	182/.7	29/.5
G2	—47/1	174/4	49/4	146/.9	177/4	—116/7	—103/3	167/.9	32/.5
A3	—67/13	—174/12	55/7	150/.8	136/70	—140/79	—98/9	181/2	31/.8
					171/6	—101/5			
C4	—65/34	—158/10	25/10	150/1	—169/7	—64/19	—101/10	188/3	30/.7
	—48/5								
T5	—99/68	—177/26	92/107	150/1	—178/13	—117/54	—65/9	177/2	33/.9
	—70/5								
T6	149/107	—173/36	80/22	149/1	—155/30	—128/94	—110/5	178/1	31/.9
						—76/14			
C7	—176/94	147/43	100/35	152/1	—118/35	—124/96	—122/10	188/2	31/.4
G8	—155/88	180/87	—155/48	154/.7	—150/36	—141/59	—145/20	196/2	31/.4
					—146/7	—112/59			
G9	—86/83	—116/73	—175/59	154/.7	170/6	—102/3	—112/6	193/3	31/1
T10	—55/21	—173/21	50/27	149/2	—178/61	—148/68	—103/13	178/1	31/1
	—48/5		41/4		—178/3	—115/6			
C11	—75/83	159/14	99/104	135/1	177/6	—115/4	—107/9	144/2	32/0
	—47/7		52/.9						
C12	—49/27	172/16	52/28	150/1	—	—	—98/9	178/.7	32/1
	—40/8		39/3						

*If a structural subset was clearly discernible, the average value for the set is given below the main entry. The standard deviations reflect the data ranges used to constrain the backbone angles. The backbone angle δ is well defined as $\nu_0 - \nu_4$, which were calculated from ρ and ϕ (see text), were used to constrain this angle.

standard deviations for δ are small. All calculated structures agree with the experimental constraints, except for the sugar pucker amplitudes, which were experimentally determined from the scalar couplings as being at the high end of the range, 40° , but calculated from XPLOR as being at the low end of the range, $\sim 30^\circ$. This phenomenon is probably due to the bias of the electrostatic potential and force field parameters used during modeling, and is an artifact. In general, the limitations of molecular modeling should not be forgotten, and although this procedure allows one to visualize the conformation of a molecule, it cannot provide an accurate view of a dynamic species. At best, modeled structures provide a snapshot of a conformation range. A robust, or universal, protocol has not been found that is capable of modeling a wide range of molecules. For example, although the protocols used for modeling the r[C(UUCG)G] hairpin and the d[C(TTCG)G] hairpin are similar, they are not the same. It was found that the distance constraints had to be applied in two steps during the global fold for the DNA hairpin, while only one step was necessary for the RNA molecule (15).

Comparison of the r[C(UUCG)G] and d[C(TTCG)G] hairpins: RNA versus DNA

Having determined the structure of the DNA analogue of the r[C(UUCG)G] hairpin, the two molecules can be compared structurally. The numbering system is the same for both molecules, numbered 1 through 12, 5' to 3'.

r[C(UUCG)G] hairpin	d[C(TTCG)G] hairpin
The stem is A-form	The stem is B-form.
G8 adopts the syn conformation.	G8 is anti, there are no syn nucleotides
U5 and G8 form a reverse-wobble base pair.	There is no evidence of H-bonding in the loop.
The C7 amino group forms a favorable contact with the phosphate between U5 and U6.	No interactions of bases with phosphate groups are seen.
U6 stacks on the C7 sugar.	T5, T6, and C7 have irregular stacking interactions with each other and the stem.
The loop is rigid.	The loop is flexible.

The thermodynamic properties of these molecules are reflected in the structures. The melting temperature for the RNA hairpin ($T_m = 76.2^\circ\text{C}$ in 0.5 M NaCl) is nearly 20°C higher than normal for tetraloop hairpins of its size, while that for the DNA hairpin ($T_m = 60.4^\circ\text{C}$) is average (4). This is partly due to the specific H-bonding and stacking interactions that occur in the RNA loop which contribute to the thermodynamic stability. The DNA molecule, however, has an unstructured loop which does not contribute to stability. The specific contributions to the structural differences, however, are not understood.

CONCLUSION

We have determined the structural differences between the r[C(UUCG)G] and d[C(TTCG)G] hairpins, and have found that the thermodynamic properties of these molecules is reflected by the loop structures. The RNA loop is highly structured and rigid, while the DNA loop is nonstructured and flexible. Yet, only two chemical features differ between RNA and DNA, the hydroxyl group at the 2' position of the ribose sugar and the methyl group

on thymidine. An effort has been made to understand the contribution of these functional groups to structure.

RNA helices naturally adopt A-form geometry due to the presence of the hydroxyl groups on the sugars, and DNA adopts B-form geometry. It is thought the stem plays a role in the placement of the loop nucleotides which determines the potential for specific interactions to occur (the U·G base pair). In addition, the loop hydroxyls clearly play a structural role (6), whether through specific interactions or through electrostatic contributions. There is no evidence for specific interactions involving the loop hydroxyls in the r[C(UUCG)G] hairpin structure (5,7), but these contacts are difficult to trace from NMR data. Therefore, specific interactions can not be excluded.

The T methyl groups may inhibit loop structure, but it is unlikely. The H5 position of U5 is in the major groove, away from the base pairing face of the residue, and there appears to be ample space to accommodate a methyl group from the model. However, this hypothesis has not been tested by modeling. There is no steric hindrance for a methyl group on U6. It is noted, however, that the hydrophobicity of the methyl groups could affect structure, especially since U6 adopts an exposed position.

ACKNOWLEDGEMENTS

We thank Mr David Koh for synthesizing the DNA molecule and Ms Barbara Dengler for managing the laboratory. This work was supported in part by National Institutes of Health grant GM10840, by the Department of Energy, Office of Energy Research, Office of Health and Environment Research grant DE-FG03-86ER60406, and through instrumentation grants from the US Department of Energy, DE FG05-86ER75281, and the National Science Foundation, DMB 86-09305 and BBS 87-20134. J.J. was supported by NIH training grant ES07075-15.

REFERENCES

1. Woese, C.C., Winker, S., Gutell, R.R. (1990) *Proc. Natl. Acad. Sci. U.S.A.*, 87, 8467–8471.
2. 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., Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, 85, 1364–1368.
3. Carafa, Y.d'A., Brody, E., Thermes, C. (1990) *J. Mol. Biol.*, 216, 835–858.
4. Antao, V.P., Lai, S.Y., Tinoco, I.Jr. (1991) *Nucleic Acids Res.*, 19, 5901–5905; Antao, V.P., Tinoco, I.Jr. (1992) *Nucleic Acids Res.* 20, 819–824.
5. Varani, G., Cheong, C., Tinoco, I.Jr. (1991) *Biochemistry*, 30, 3280–3289.
6. Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M., Uesugi, S. (1990) *Nucleic Acids Res.*, 18, 3831–3839.
7. Cheong, C., Varani, G., Tinoco, I.Jr. (1990) *Nature*, 346, 680–682.
8. Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G., Aelen, J.M.A., van der Marel, G.A., van Boom, J.H., Hilbers, C.W. (1989) *Biochemistry*, 28, 7491–7498.
9. Turner, D.H., Sugimoto, N., Freier, S.M. (1988) *Annu. Rev. Biophys. Biophys. Chem.*, 17, 167–192.
10. Marion, D., Wüthrich, K. (1983) *Biochem. Biophys. Rev. Commun.*, 113, 967–974.
11. Plateau, P., Gueron, M. (1982) *J. Am. Chem. Soc.*, 104, 7310–7311.
12. Rinkel, L.J., Altona, C. (1987) *J. Biomol. Struct. Dyn.*, 4, 621–649.
13. Varani, G., Tinoco, I.Jr. (1991) *Quart. Rev. Biophys.*, 24, 479–532.
14. Brünger, A. (1992) *X-PLOR version 3.0 Manual: A System for Crystallography and NMR*.
15. Wimberly, B. (1992) Ph. D. Thesis, University of California, Berkeley.