Nucleic Acids Research, 1997, Vol. 25, No. 11 **2129–2137**

Structural characterization of three RNA hexanucleotide loops from the internal ribosome entry site of polioviruses

Roscoe Klinck, Tara Sprules and Kalle Gehring*

Department of Biochemistry and Montreal Joint Centre for Structural Biology, McIntyre Medical Science Building, McGill University, 3655 Drummond, Montréal, QC, H3G 1Y6, Canada

Received January 30, 1997; Revised and Accepted April 17, 1997

ABSTRACT

Structural characteristics of three RNA hairpins from the internal ribosome entry site of poliovirus mRNAs have been determined in solution by NMR. Complete proton, phosphorus and carbon resonance assignments were made for the three 16 nt hairpins. The loop sequences, 5′-AAUCCA , AAACCA and GAACCA, have been shown to be essential for viral mRNA translation. NOESY spectra for the three oligomers were very similar indicating a common three dimensional structure. Stems were A-type duplexes with C3′-endo sugar pucker. In the loops, sequential base stacking interactions were detected for all bases except between U8/A8 and C9, indicating a turn in the phosphodiester backbone at this point. Only one nucleotide, U8/A8, had a sugar pucker which deviated appreciably from C3′-endo. The final base in the loop, A11, exhibited an unusual *gauche*(–) gamma angle. An ensemble of 10 structures calculated for one hairpin using restrained molecular dynamics shows that the first three bases of the loop are turned so as to be exposed to the exterior of the molecule, while the remaining three bases are in an orientation approximating a continuation of the stem helix. Structure calculations and NMR relaxation measurements indicate that the loop apex is subject to considerable local dynamics.

INTRODUCTION

Translation of the majority of eukaryotic mRNAs is initiated by a 5′ cap dependent mechanism which involves the scanning of the 5′ untranslated region (5′-UTR) by small ribosomal subunits. Contrary to this, initiation of translation of some viral mRNAs in polioviruses does not require a 5′ cap structure (m^7GpppN). The initiation of viral genomic RNA translation has been reviewed recently (1,2). In members of the *Picornaviridae* family, such as poliovirus, the 5′-UTR is approximately 740 nt long of which ~450 are necessary for cap-independent translation. The terms ribosome landing pad (3) or internal ribosome entry site, IRES (4), have been used to describe this region. The IRES has conserved secondary structural elements which are essential for ribosome recognition and binding (5–7).

Comparative sequence analysis of enteroviruses/rhinoviruses (E/R viruses, a subfamily of the picornaviruses) has shown that loops III and VI are highly conserved. Sonenberg and co-workers reported that loop III in poliovirus is essential for initiation of transcription (7). Le and co-workers (6) have suggested that loop VI participates in the formation of a pseudoknot structure based on sequence comparisons of 22 strains of poliovirus, coxsackievirus and rhinovirus. If present, the pseudoknot structure may serve as a recognition site for components of the translation apparatus (6). Sequence comparisons of these two loops in 28 strains of picornaviruses are described elsewhere (Klinck *et al.*, submitted).

We have studied by NMR three RNA hairpins analogous to loops III and VI of the poliovirus IRES (Fig. 1) in order to determine the structural features that might account for their strong conservation. The hairpins studied were chosen to encompass the variations encountered among members of the poliovirus family. We relate general structural features of the oligomers and sequence related differences. For one of the oligomers, a ^{13}C-labeled sample was prepared allowing the calculation of a structural model as well as measurements of NMR relaxation parameters related to molecular motion.

MATERIALS AND METHODS

RNA preparation

The sequences of the RNA oligomers were based on the proposed secondary structure of nucleotides 140–155 and 493–508 of the 5′-UTR in poliovirus type 2 Lansing PV2la and PV2s (7). The stem sequence of the hairpins was modified to improve transcription yields and provide a constant structural context for studying differences in the loop sequence (Fig. 1). The 16-base oligomers were prepared biosynthetically by DNA template driven T7 RNA polymerase transcription (8). T7 polymerase was prepared following the procedure of Wyatt and co-workers (9). The *Escherichia coli* strain BL21 harboring the plasmid pAR1219 which carries the T7 RNA polymerase gene was kindly provided by William Studier (Brookhaven National Laboratory, Upton, NY).

* To whom correspondence should be addressed. Tel: +1 514 496 2558; Fax: +1 514 496 5143; Email: kalle@bri.nrc.ca

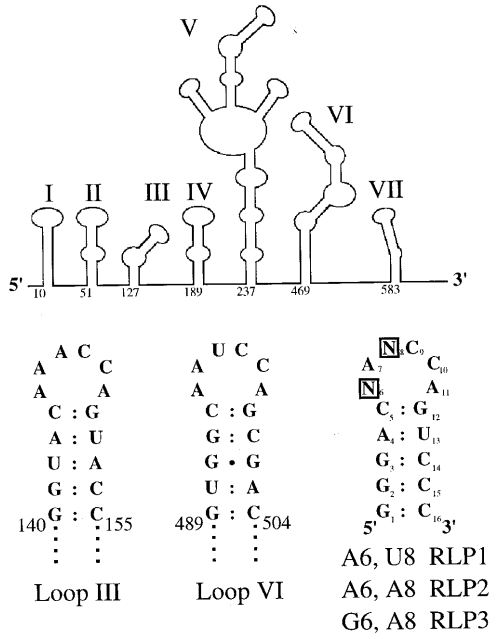

Figure 1. Proposed secondary structure of the poliovirus type 2 Lansing 5′-UTR, redrawn from Nicholson *et al*. (7). Sequences of terminal loops from domains III and VI are shown along with the three RNA oligomers studied.

Initial NMR samples were prepared from transcription reactions which were primed with 5 mM rGMP. As this led to a mixture of 5′ monophosphorylated and triphosphorylated RNAs, rGMP was omitted from later reactions. A ^{13}C labeled sample of RLP1 was prepared from a 30 ml transcription containing 4 mM uniformly ^{13}C labeled rNTPs. Uniformly ^{13}C labeled nucleoside triphosphates were prepared as described from high molecular weight cellular RNA (10,11). Purification was done on 20% polyacrylamide gels containing 7 M urea. The desired bands were viewed by UV shadowing, excised and electroeluted and the RNA concentrated on a 3000 MWCO Centricon micro-concentrator (Amicon). Final sample concentrations were 0.6 mM (RLP1), 0.3 mM (RLP2) and 2 mM (RLP3) for unlabeled samples and 1 mM for ^{13}C labeled RLP1 in ~300 µl volumes. Buffer conditions for all experiments except exchangeable proton spectroscopy were 100% D_2O, 10 mM sodium phosphate, pH 6.4 (uncorrected for deuterium), 25°C.

NMR spectroscopy

All experiments were run on a Bruker DRX 500 spectrometer equipped with a $^{1}H/^{13}C/^{31}P$ triple resonance *x, y, z* gradient probe. Proton resonances were referenced to the water resonance at 4.7 p.p.m. at 25°C. Carbon chemical shifts were referenced to the proton spectrum using the ratio of the gyromagnetic moments (γ^H/γ^C = 3.97693404) obtained from a sample of tetramethylsilane (0 p.p.m. for both nuclei). Phosphorus chemical shifts were referenced in a similar manner to 85% phosphoric acid (0 p.p.m. ^{31}P) using γ^H/γ^P = 2.47031216.

Resonance assignments

NMR assignments were made using a combination of homonuclear and heteronuclear techniques. The synthesis of a uniformly ^{13}C labeled sample of RLP1 allowed more complete chemical shift assignments to be obtained for this oligomer, including stereospecific assignments of the H5′/H5′ resonances. Chemical shift tables for all three RNAs are available as supplementary material.

NOESY spectra were used for initial assignments. Aromatic-H1′ resonances were assigned from sequential connectivities at long NOESY mixing times. These assignments were confirmed and extended using ^{1}H-^{13}C correlation spectra at natural ^{13}C abundance to identify the different carbon types (12). H2′ protons were assigned from the H1′ protons using COSY-DQF and short mixing time (40–60 ms) NOESY spectra. Comparison of ^{1}H-^{31}P-HETCOR spectra with the aromatic-sugar region of NOESY spectra allowed assignment of the H3′ protons. H4′ assignments were obtained by comparison of the H1′-sugar region of the 400 ms NOESY with the H4′–C4′ region of the ^{1}H-^{13}C HMQC spectra and the H5′/H5′ protons were assigned non-stereospecifically by comparative analysis of NOESY, COSY-DQF, HMQC and HETCOR spectra. The sugar resonance assignments of ^{13}C-labeled sample of RLP1 were confirmed using HCCH-COSY, HCCH-TOCSY and 3D-HCP (carbon–phosphorus) spectra (13–17).

Two methods were used for the stereospecific assignment of the H5′ (pro-S) and H5″ (pro-R) resonances of RLP1. For residue A11, the signs of the C4′–H5′ and C4′–H5″ two bond coupling constants were measured by an HSQC-30 experiment (a fully coupled HSQC in which the final carbon π/2 was replaced with a π/6 pulse) and the magnitudes of the three bond H4′–H5′ and H4′–H5″ couplings were measured by a directed-HCC-TOCSY-CCH-E.COSY experiment (18). These couplings were then interpreted to give the stereospecific assignments as described by Hines *et al*. (19). For the remaining residues, the large chemical shift difference between H5′ and H5″ allowed the assignments to be made using the recently described correlation between H5′/H5″ and C5′ chemical shifts (18).

Phosphorus assignments of unlabeled samples were made using ^{31}P-^{1}H-hetero-TOCSY-NOESY spectra (20) and ^{1}H-^{31}P-HETCOR spectra (21). Assignments of the ^{13}C-labeled sample were carried out using a P(CC)H-TOCSY experiment (22).

NMR constraints

H1′–H2′ coupling constants in unlabeled samples were measured by comparing COSY-DQF and NOESY crosspeaks as described by Poulsen and co-workers (23). Values for $^{3}J_{H3'–H4'}$, $^{3}J_{H4'–H5'}$ and $^{3}J_{H4'–H5''}$ were estimated directly from a phosphorus decoupled COSY-DQF spectrum. Proton–proton coupling constants in the ^{13}C labeled RLP1 sample were extracted from a directed-HCC-TOCSY-CCH-E.COSY spectrum (18,24). Although it was reported that the H1′–H2′ and H4′–H5′/H5″ coupling data should be extracted from two separate experiments with different TOCSY mixing times, we found that one 3D experiment using a 13.5 ms 4 kHz TOCSY spinlock was sufficient to obtain the required information for the entire spin system.

Semi-quantitative measurements of proton phosphorus and carbon phosphorus coupling constants were used for generating backbone angle dihedral constraints as described by Varani and co-workers (12,25). Very weak or absent crosspeaks in the HP-HETCOR (proton–phosphorus) (21) or 3D-HCP (carbon–phosphorus) spectra (13,14) were assumed to reflect a coupling constant of <5 Hz. Values for large sequential H3′-P couplings were estimated by subtracting the passive H2′–H3′ and H3′–H4′ couplings from the H3′-P crosspeak multiplet width in the

HP-HETCOR spectrum (12). The relative intensities of the C4′-P and C2′-P crosspeaks in the HCP experiment were used to assign a lower limit of 5 Hz to the more intense peak. Strong intranucleotide P-C4′ cross peaks was interpreted to represent a P-C4′ coupling constant of the order of 5 Hz or greater. Additional dihedral constraints for structural modeling were derived from NOESY cross peak intensities (to constrain the glycosidic angle, χ), and from phosphorus chemical shifts (α and ζ angles) (26).

NOESY crosspeaks were classified into four groups representing distance ranges based on a visual inspection of their intensities. Strong (1.8–3 Å) and medium (2–4 Å) ranges were used for peaks detectable at 60 ms NOESY mixing time. Weak (3–5 Å) and very weak (3–6 Å) classifications were applied to peaks only detectable at 400 ms NOESY mixing time. The broad distance ranges for weak and very weak NOEs were used to allow for the effects of spin diffusion at long NOESY mixing times.

Structural modeling of RLP1

The program MC-SYM version 1.3 (27) was used to generate a folded starting structure using the assumption of an A type helix for the stem nucleotides, G1–C5 and G12–C16. The input file consisted of explicit base pairing and stacking terms to generate an A type helix for the stem nucleotides. Terms permitting a large degree of sampling were used for the loop nucleotides. A set of 44 distance constraints derived from the NOE data was used to reduce the number of solutions generated. One of the 13 structures generated by the MC-SYM run was used as an input structure for a four stage structural refinement protocol using X-PLOR, version 3.843 (28). The nucleic acid all hydrogen force field provided with the program was used for all calculations. Electrostatic and attractive van der Waals terms were off for all but the final minimization. A harmonic potential was used to restrain the first four base pairs of the stem to A-type geometry generated by MC-SYM in the first two stages of the refinement protocol. For the remaining two stages, the observed NOEs and standard A type helix dihedral angles (29) (α 190 ± 30, β 180 ± 30, γ 60 ± 20, ε 210 ± 40, ζ 190 ± 30, χ 210 ± 30) were used to constrain the stem nucleotides (G1–A4 and U13–C16).

A total of 149 NOE constraints were used in the structure calculations: 91 for nucleotides C5–G12 (of which 45 were interresidue NOEs), and a less comprehensive set of 39 for the remaining stem nucleotides. The five base pairs in the stem were constrained with 19 pseudo NOE constraints (three for the A·U pair and four for each G·C pair), using distances of 2.9 Å ± 0.2 between heavy atoms and 1.7 Å ± 0.3 between the imino proton donor and acceptor. In addition, four negative NOEs were included, imposing a minimum distance of 4.5 Å for C9H6 to U8H6, C9H6 to U8H1′, U8H6 to A7H8, and U8H6 to A7H1′. These NOEs were justified by the fact that no NOEs between these protons were detected in any of the NOESY spectra recorded and that *other* protons did show NOEs to these protons. Their inclusion improved the rate of convergence in the early stages of the structure refinement. Six very weak NOEs were also included between C5H5 and the N1, C2, N3, C4, C5 and C6 atoms of A6 to describe the ring current effect responsible for the marked upfield shift of C5H5 (30,31).

All sugar puckers were constrained to C3′-endo conformation, with the exception of U8. The α and ζ angles for nucleotides C5 to G12 were constrained to 0 ± 120 degrees to exclude the *trans* conformation. Beta angle constraints of 180 ± 30 degrees were imposed for all nucleotides. On the basis of the small H4′–H5′ and H4′–H5″ couplings, the gamma angles for all nucleotides except A11 were constrained to the *gauche*(+) range (60 ± 20 degrees). A constraint of 270 ± 30 degrees was applied to the A11 gamma angle. Epsilon angles were constrained to a *trans* range, 210 ± 40 degrees, for all nucleotides.

In the first stage of refinement, 40 ‘randomized’ structures were generated by forcing random α, β, γ, ε and ζ dihedral angles of the six loop nucleotides. This procedure involved 250 steps of energy minimization, 1.5 ps of dynamics at 300 K followed by 250 steps of energy minimization. Subsequent stages were similar to those described previously by Tinoco and co-workers (32). A simulated annealing stage was carried out for 13 ps at 1000 K using only NOE derived distance constraints. The repulsive-only van der Waals forces and a soft square well NOE potential were slowly increased during this period. The resulting structures were then subjected to additional simulated annealing, 6 ps at 1000 K, using a square well NOE potential, with the gradual introduction of the dihedral angle constraints, followed by cooling to 300 K. The protocol was completed with a final minimization (1000 steps) during which attractive van der Waals and electrostatic terms were introduced.

Relaxation measurements

Proton T_1 data were generated by a series of inversion-recovery [^{1}H-^{13}C]-HSQC experiments. Peak intensities were fit to a single exponential using the Marquardt–Levenberg algorithm (GNUPLOT v. 3.5). ^{1}H-^{13}C heteronuclear NOEs were measured by comparing two constant-time reverse INEPT experiments with and without proton saturation. The two experiments were acquired in interleaved fashion with careful attention to maintaining constant levels of radio frequency heating.

RESULTS

Exchangeable proton spectra

Thermal denaturation studies by 1D spectroscopy showed that all three RNA hairpins became single-stranded between 65 and 70°C. This is consistent with UV studies performed on 100–1000-fold more dilute samples (Klinck *et al.*, submitted) and confirms that the hairpins form intramolecular hairpins. In each case the number of NMR imino proton resonances was consistent with the formation of 5 base pairs in the stem.

NOESY spectra, recorded in H_2O, were used to assign the exchangeable protons. Complete connectivities for the stem imino protons were obtained for all three oligomers. Amino protons involved in base pairing were assigned from imino–amino correlations. RLP1 and RLP3 both showed an extra imino resonance from a loop nucleotide. Based on chemical shifts, these were assigned to the U8 imino proton of RLP1 (10.62 p.p.m.) and the G6 imino proton of RLP3 (9.89 p.p.m.). Aside from chemical exchange with solvent and intrabase imino–amino crosspeaks, no NOESY crosspeaks were observed for these resonances. The absence of other imino protons in the loop and the large linewidths of the amino resonances limited the usefulness of NOESY spectra in H_2O for structural characterization of the RNA hairpins.

Non-exchangeable proton spectra

Strong overall similarities were evident in the NMR spectra of the three hairpins and chemical shifts were generally similar (Fig. 2). Unusual upfield proton chemical shifts were observed for the H5

and H6 protons of C5 for all three oligomers. We attribute this to a ring current effect exerted by the first loop nucleotide, A6 (RLP1 and RLP2), or G6 (RLP3) (30). The size and sign of the shift suggests that the C5H5 proton must be nearly centered under the base of residue 6. In all three hairpins, the A11 aromatic and A11 H1′ chemical shifts were consistently downfield.

The one nucleotide change at position 8 (RLP1 to RLP2) produced only subtle changes in the NMR spectra, whereas more significant differences were observed when nucleotide 6 was changed from A to G (RLP2 to RLP3). In RLP3, the H8 resonance of G12 was shifted downfield by 0.4 p.p.m. The RLP3 C5 aromatic protons showed smaller upfield chemical shifts compared with the large shifts observed in RLP1 and RLP2 and the H1′ resonance of G12 was markedly broadened. We suspect that this broadening may be due to slow conformational averaging.

At short NOESY mixing times, the stem nucleotide resonances displayed correlations typical of A-type helices. Sequential H2′–H6/H8, H3′–H6/H8 and H2–H1′, intranucleotide H6/H8–H3′ correlations and an interstrand A4H2–C14H1′ NOE were observed in all cases except where spectral overlap hindered unambiguous assignment. Weaker NOEs, including sequential H2′–H1′ correlations were also observed throughout the stem nucleotides of each oligomer.

Sequential NOEs were also observed for all of the loop nucleotides. The recurrence of sequential H2′/H3′-aromatic and H2′–H1′ NOEs suggested that base stacking interactions were continuous from G1 to U8 or A8 on the 5′ side, and from C16 to C9 on the 3′ side. Although a number of NOEs were identified between nucleotides 8 and 9, the consistent absence of sequential U8H2′ or A8H2′–C9H1′ correlations was interpreted as an interruption in base stacking interactions. The three hairpins showed medium to strong intranucleotide H6/8–H2′ NOEs within the loop at short NOESY mixing times (40–60 ms). These NOEs are generally associated with C2′-endo sugar pucker. The fact that none of the H1′–H2′ coupling constants was consistent with C2′-endo conformation suggests that these NOEs arise from conformational averaging in the loop sugars. Three non-standard NOEs were observed for RLP1, A11H1′–G12H1′, A11H2–A6H1′ and A11H2–A6H2, each of which was included in the structure calculations as a very weak NOE.

The NOEs identified for the three hairpins showed strong similarities. The results for the aromatic and H1′, H2′ and H3′ protons are presented schematically in Figure 3. Some differences were observed, as described below; however, the overall resemblance of the NOE patterns implies that the three oligonucleotides adopt a common three dimensional structure. Two non-structural sources of differences in the NOE tables are sample concentration and chemical shift overlap. At 2 mM, RLP3 was three times more concentrated than RLP1 or RLP2 and clearly gave more numerous and stronger NOEs. It should be noted, however, that the relative strengths of most of the critical cross loop NOEs observed for RLP3 (see below) was such that their detection would have been possible at concentrations in the 0.5–1 mM range. Occasionally, chemical shift degeneracies also prevented the identification of NOEs. For example, due to the overlap of the G12H8 and U13H6 resonances in RLP3 at all temperatures studied, only two sequential NOEs could be assigned between these nucleotides.

The most significant NOE related differences between the three oligomers were the non-sequential ‘cross-loop’ NOEs of A11. Several interesting long range NOEs involving this nucleotide

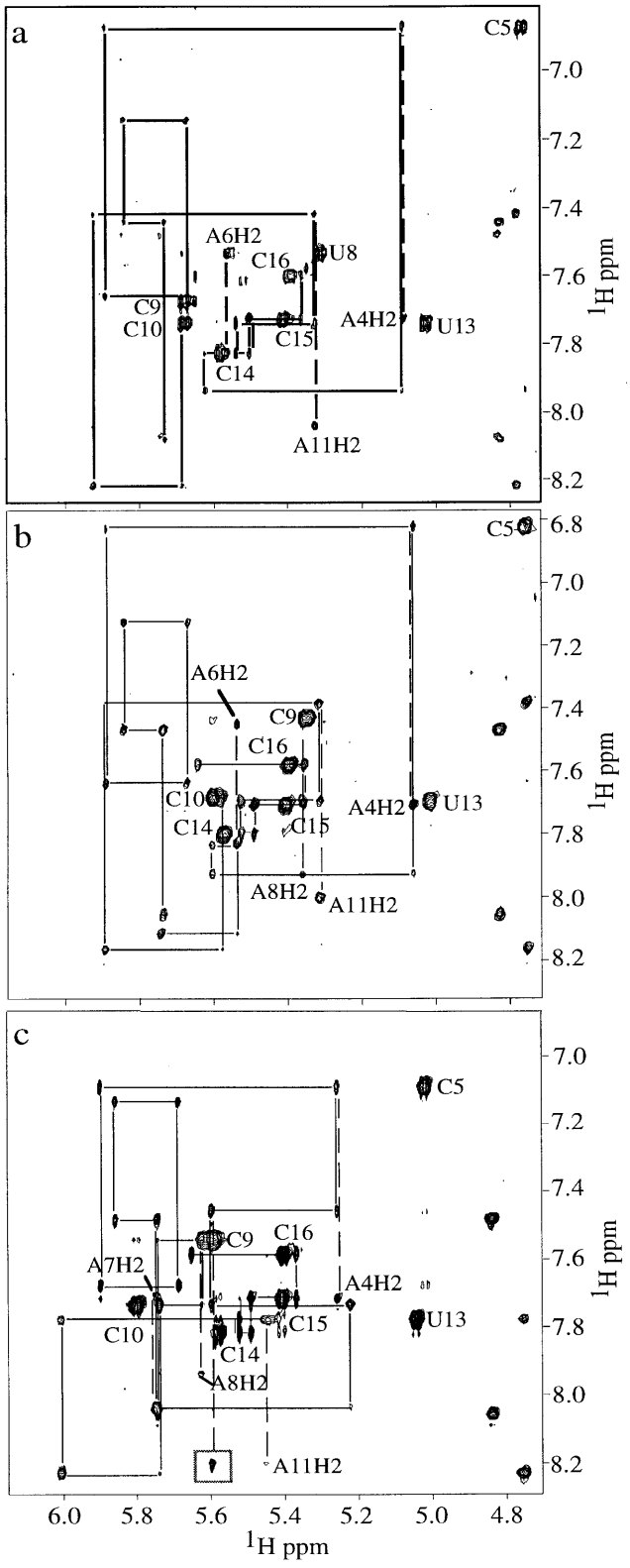

Figure 2. Aromatic to H5/H1′ region of 400 ms NOESY spectra in D_2O. Solid lines trace the sequential base to sugar H1′ connectivities. Dashed lines indicate adenosine H2 to sugar H1′ connectivities. (**a**) RLP1, (**b**) RLP2, (**c**) RLP3. The cross-loop A11H2–G6H1′ NOE of RLP3 is boxed in (c). This NOE is present in the 60 ms NOESY spectrum (data not shown) and critically defines the position of the two bases (Fig. 7).

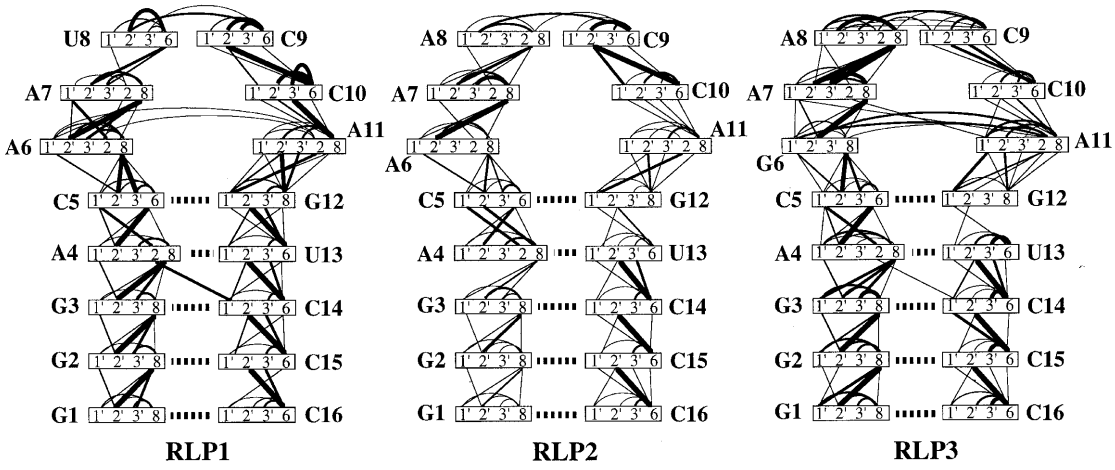

Figure 3. Schematic representation of the 16mer hairpin loops, summarizing the observed NOEs. The four line styles correspond in order of decreasing thickness to strong, medium, weak and very weak NOEs, as defined in the text. For clarity, only NOEs involving aromatic and H1′, H2′ or H3′ protons are shown.

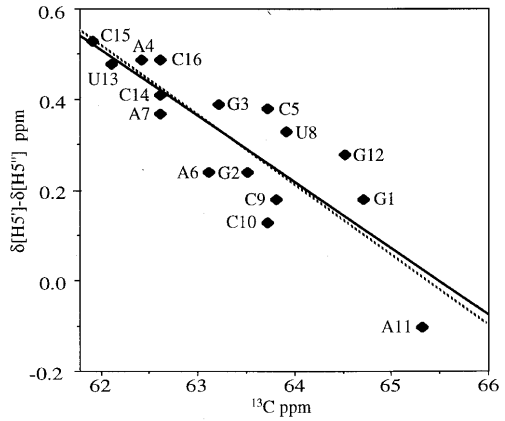

Figure 4. Correlation of H5′–H5″ chemical shift difference with C5′ chemical shift used for the stereospecific assignment of H5′ and H5″. Our data (◆) and best linear fit (solid black line) are shown with the linear fit (dotted line) reported by Griesinger and co-workers (18).

were identified in RLP3. Medium range NOEs were observed from A11H2 to G6H1′ and G6H2′, and weaker mutual cross strand H2–H1′ correlations were detected between A11 and A7 (Fig. 2). One very weak A11H2–G6H8 NOE was also observed. The A11–G6 NOEs introduced the possibility of base pairing between these two nucleotides. The A7–A11 correlations observed would be expected for an A-type helix. This suggests that in RLP3, the stem helix continues into the loop region to include at least nucleotides G6, A7 and A11. Analogous NOEs were either very weak or absent in RLP1 and RLP2. The presence of A7–A11 NOEs and the different ring current effect of A6 on the C5 aromatic protons, described above, suggests that RLP3 differs slightly from RLP1 and RLP2 with respect to the positioning of nucleotides A6 and A7.

The H5′ and H5″ resonances of RLP1 were initially assigned stereospecifically using the correlation of ΔH5′/H5″ and C5′ chemical shifts described by Marino *et al.* (18). Our data are in very good agreement with those reported previously when the downfield proton was assigned to H5′ for all residues except A11 (Fig. 4). For A11, the chemical shift difference between H5′ and H5″ is too small (0.1 p.p.m.) for a reliable attribution. Therefore, these H5′/H5″ protons were assigned independently using two and three bond J couplings (Materials and Methods). Assignments made using ^{1}H-^{13}C correlation spectra are a timely improvement over Shugar and Remin's rule (33) as they account for cases where H5′ protons resonate upfield of H5″ protons. Stereospecific assignment of A11 was critical for identification of the unusual *gauche*(–) backbone dihedral angle γ.

Coupling constant analysis

Two methods were used for measurement of the H1′–H2′ coupling constants of RLP1 with very similar results (Table 1). The HCC-TOCSY-CCH-E.COSY method (18,24) has the advantage that very weak coupling constants can be measured but requires uniform ^{13}C labeling. In all three oligomers, most of the loop nucleotides showed detectable DQF-COSY H1′–H2′ cross-peaks which indicates deviation from C3′-endo sugar geometry. None of the $J_{H1'-H2'}$ constants were characteristic of full C2′-endo pucker (8–10 Hz) which suggests that C3′-endo-C2′-endo conformational averaging was taking place. For U8, the possibility of a stable O4′-endo conformation could be ruled out by the absence of a strong H1′–H4′ NOE.

The values measured for RLP3 were very similar to those measured for RLP1 with the exception of residue A11. The measured $J_{H1'-H2'}$ of RLP3 were 4.9, 2.8, 2.3 and 4.5 Hz for A8, C9, C10 and A11 respectively. The larger coupling observed for A11 in RLP3 suggests that its sugar undergoes more conformational averaging. Coupling constants of RLP2 were difficult to measure accurately because of the lower sample concentration but were qualitatively the same as those of RLP1.

Preliminary assessment of the gamma dihedral angles was done by inspection of the H5′–H5″ crosspeaks of the phosphorus decoupled COSY-DQF spectrum. For both RLP1 and RLP3, nucleotide A11 exhibited a pronounced H5′/H5″ to H4′ passive coupling and a strong H4′–H5′/H5″ crosspeak. For RLP1, this could be attributed to a *gauche*(–) γ conformation for A11. This conformation was confirmed by the presence of medium A11H8–A11H5′ and weak A11H8–A11H5″ NOEs. For RLP3, the H5′ and H5″ resonances were degenerate (overlapped) which prevented exact determination of the γ backbone angle. However, it seems likely that all three hairpins adopt a similar conformation at A11.

Table 1. Loop nucleotide coupling constants for RLP1

	$^3J_{H1'–H2'}$ (Hz)		$^3J_{H2'–H3'}$ (Hz)	$^3J_{H3'–H4'}$ (Hz)		$^3J_{H4'–H5'}$ (Hz)	$^3J_{H4'–H5''}$ (Hz)
	E.COSY[a]	COSY-DQF[b]	COSY-DQF[c]	E.COSY[a]	COSY-DQF[c]	E.COSY[a]	E.COSY[a]
A6	2.0	–[d]	4.5	–	8	1.2	0.3
A7	1.5	1.6	4.5	8.3	8	3.6	0.3
U8	4.4	4.6	5	5.7	5.5	1.3	3.1
C9	2.4	2.3	4	5.0	7	0.6	1.3
C10	2.3	1.7	4	–	7	1.0	–
A11	2.6	1.7	4	–	8	5.2	0.7

[a]Measured coupling constants (Hz) in directed-HCC-TOCSY-CCH-E.COSY, 25°C.
[b]Measured coupling constants (Hz) in COSY-DQF spectrum at 34°C (cf. Materials and Methods).
[c]Observed splittings in COSY-DQF spectrum at 34°C.
[d]Coupling not measured due to spectral overlap.

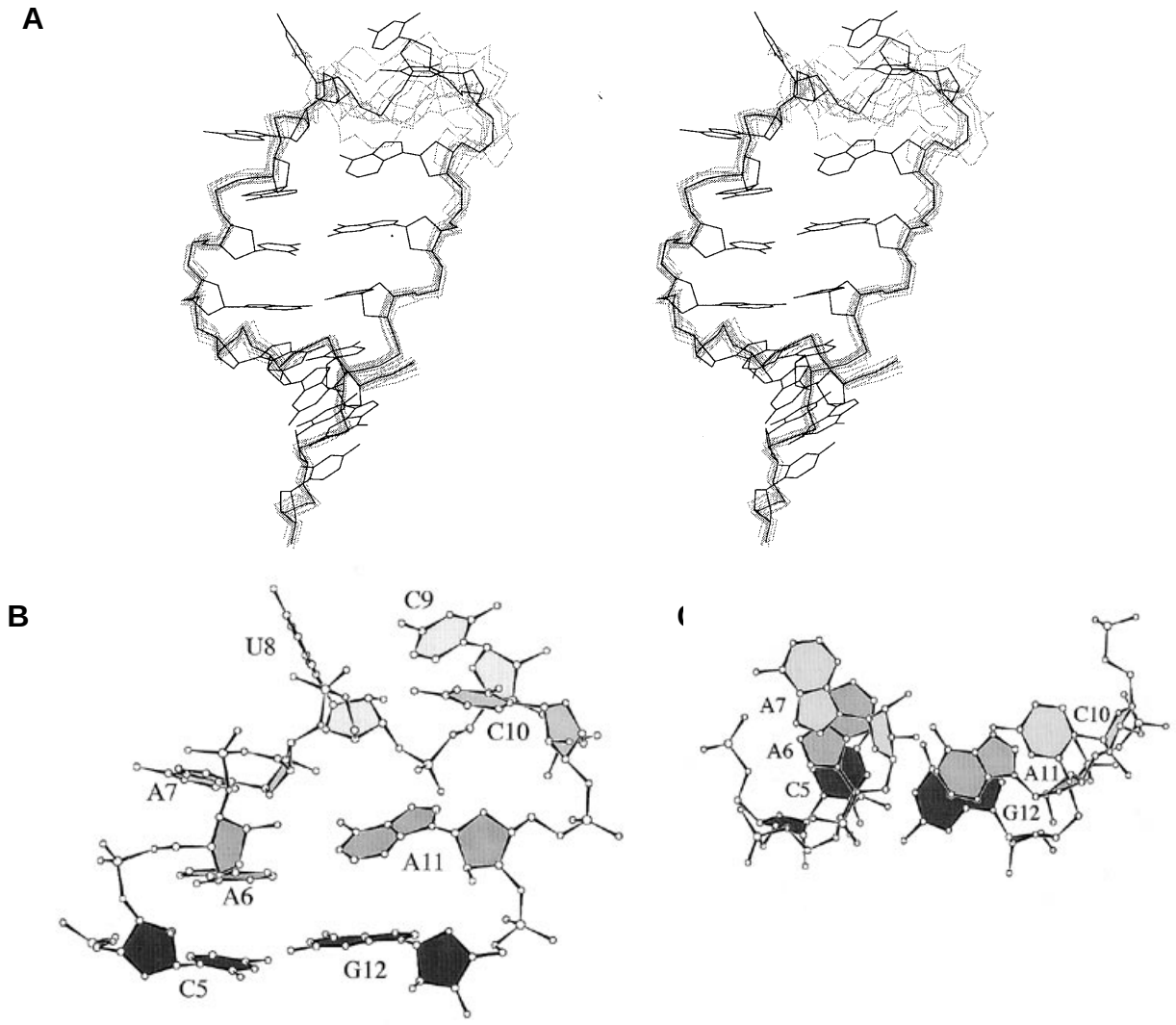

Figure 5. Calculated structures of the 16mer hairpin RLP1. (**a**) Superposition of the 10 final structures (backbones only, in grey) and their calculated average structure (in black). (**b**) Detail of the six loop nucleotides and the closing base pair of the lowest energy structure viewed from the minor groove. The unusual alpha and gamma angles of A11 are clearly evident in the RNA backbone between A11 and C10. (**c**) View of base stacking with the closing C5·G12 base pair and four of the six loop nucleotides. The base of A7 turns out of the loop and while A11 makes close contacts with the sugar protons of A6 and A7. Stacking of A6 over C5 is responsible for the large upfield shift of the C5H5 and C5H6 protons.

Examination of the curves generated from Karplus equations (25,34) reveals that a H4′–H5′ coupling constant of 8–10 Hz would be expected for this conformation. The measured coupling constant of 5.2 Hz for RLP1 suggests that conformational averaging is taking place. Averaging within the gamma angle range of ~240–300 degrees could account for the smaller than expected $^3J_{H4'–H5'}$, while remaining consistent with the small H4′–H5″ coupling constant and the NOEs from A11H8.

Strong intranucleotide P-C4′ cross peaks could be observed in the HCP spectrum of RLP1 for the loop nucleotides and the closing C5·G12 base pair. The absence of strong intramolecular P-H5′ or P-H5″ cross peaks in the HP-HETCOR spectrum suggested a coupling pattern ($^3J_{P\text{-}C4'} > 5$ Hz $> {}^3J_{P\text{-}H5'/H5''}$) indicative of *trans* dihedral β angles. A similar approach was used to evaluate the ε backbone angles. In this case, inspection of curves derived from the Karplus equations relating $^3J_{C4'\text{-}P}$, $^3J_{C2'\text{-}P}$ and $^3J_{H3'\text{-}P}$ to epsilon (25) implied a *trans* dihedral angle conformation for these nucleotides.

The phosphorus chemical shifts were limited to a range of approximately 1 p.p.m. for all three hairpins. As a *trans* conformation of the α or ζ dihedral angles is known to be accompanied by a large phosphorus shift (26), we conclude that all the α and ζ dihedral angles are either *gauche* (+) or *gauche* (–).

Structure calculations for RLP1

Using the extensive set of dihedral angles and 149 NOE constraints, an ensemble of structures were calculated for the RLP1 hairpin. Of the forty starting structures with randomized loop backbone angles, ten converged from the rMD protocol. Figure 5a shows a superimposition of the backbones of these ten structures, together with an average structure calculated and energy minimized using X-PLOR. The pairwise average r.m.s.d. for the ten structures was 1.68 Å for the loop nucleotides and 1.49 Å for the entire molecule based on the heavy atoms. The ensemble had 16 NOE violations of >0.1 Å, eleven of which were in the 0.10–0.15 Å range, and none over 0.3 Å. No dihedral angle constraint violations of >5 degrees were detected in any of the ten structures. Examination of the 10 structures confirmed that no short distances between non-exchangeable protons were present which were not part of the NOE constraint set. The lowest energy structure of the ensemble had no NOE violations >0.1 Å and no dihedral angle violations >5 degrees.

The inclusion of a comprehensive set of backbone dihedral angle constraints did little to improve the efficiency of convergence. The most probable explanation for this is that the generous ranges accompanying each constraint still permit wide conformational sampling. However, given the current methods of RNA structure determination by NMR, it would seem unreasonable to apply more stringent angle constraints based on experimental data such as ^{31}P chemical shifts and semi-quantitative ^{31}P-^{1}H /^{13}C coupling constants.

Details of the loop and closing base-pair nucleotides of the lowest energy structure are shown in Figure 5b and c. The hairpin appears to be divisible in two parts of three nucleotides each. The first three nucleotides, A6–U8, are progressively turned away from the longitudinal axis of the molecule in the direction of the major groove of the helix. These bases are exposed to the exterior of the molecule. All three nucleotides show characteristic signs of sequential base stacking, although U8 appears to be somewhat isolated. A turn in the backbone occurs between nucleotides U8 and C9. This turn is accomplished without great deviation from standard A-form RNA backbone geometry, in particular no *trans* conformations of the α or ζ dihedral angles were observed. The second side of the loop, comprised nucleotides C9 to A11, approximates the continuation of the helical stem in the 3′ direction.

The *gauche*(–) gamma angle conformation of A11 is accompanied by a shift of the alpha angle from *gauche*(–) to *gauche*(+). These two conformations of α and γ together avoid the awkward backbone twist that would result from one or the other conformation alone. Surprisingly, this conformation differs very little from standard A-type with respect to the phosphodiester backbone. However, the lateral position of the base is modified. In RLP1, A11 adopts this conformation to permit it to complete the loop and maintain the C9 to G12 base stacking seen in the NOE data. This gamma angle conformation appears to be quite rare. A search of the Protein Data Bank revealed that there are only three examples of this nucleotide conformation in deposited RNA structures (F. Major, personal communication). Two were from a crystal structure of transfer RNA with *syn* glycosidic bases (35), and the third was G28 (C3′-endo, anti) from a recent pseudoknot structure determined by NMR (36).

Relaxation measurements

NMR relaxation measurements of RLP1 were carried out to monitor the dynamics of the hairpin. Both the H1′ longitudinal relaxation times (T_1) and the ^{1}H-^{13}C heteronuclear NOEs showed increased mobility in the loop nucleotides and at the end of the hairpin stem. The ^{1}H-^{13}C heteronuclear NOE experiment was acquired twice with and without the addition of 1 mM EDTA. In the absence of EDTA, the heteronuclear NOE was decreased at residues G1 and G2 which we interpret as quenching by a contaminating metal ion complexed to the 5′ triphosphate. Spectral overlap prevented measurements for the H1′/C1′ atoms of residues G3, C9, C10 and G16; however, the heteronuclear NOEs of other carbons (data not shown) confirmed the local mobility of the loop and terminal nucleotides.

Comparison of the the average r.m.s.d. of the C1′ atoms in the calculated ensemble of structures and the NMR relaxation parameters shows a clear correlation between molecular motion and the degree of divergence in the structures (Fig. 6). Residue U8 shows both the largest heteronuclear NOE (shortest H1′ T_1) and the greatest disorder in the calculated structures.

DISCUSSION

Loop base pairs

The long range NOEs between A11H2 and A6H2/A6H1′ led us to investigate the possibility of a mismatched A·A base pair adjacent to the loop closing C·G pair. However, no NOEs suggesting hydrogen bonding were detected in NOESY spectra recorded in H_2O, nor were base–base interactions evident in any of the converged structures. Examination of the final structures did reveal at least two potential hydrogen bonds across the loop: between the A11N6 amino proton and A7O2′ (~2.4 Å) and between a C10N4 amino proton and one of the U8 phosphate oxygens (~1.8 Å). We were unable to obtain independent confirmation of the existence of these hydrogen bonds; however, they could be important in stabilizing the loop structure. An A7–A11 hydrogen bond might contribute stability in the absence of A6–A11 base–base interactions, and a U8–C10 interaction could stabilize the turn in the phosphodiester backbone.

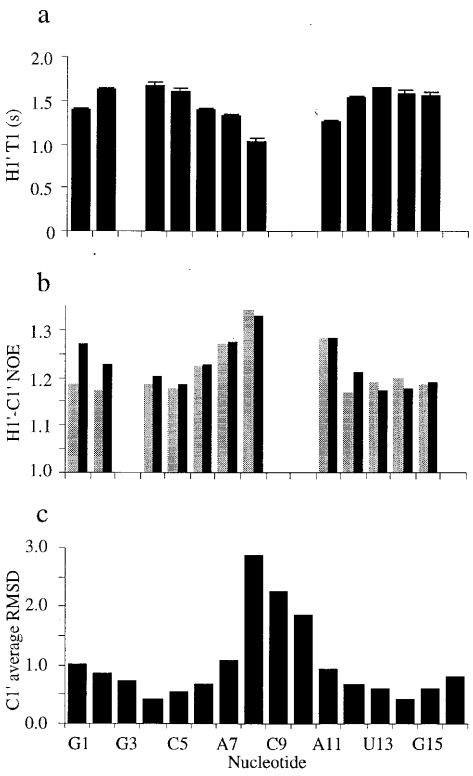

Figure 6. Comparison of measured relaxation parameters of RLP1 with calculated average r.m.s.d. of C1′ atoms of the 10 final structures. (**a**) Apparent T_1 values for the H1′ protons. (**b**) Heteronuclear NOEs measured on the C1′ carbons. Identical experiments were run with (black bars), and without (grey bars) 1 mM disodium EDTA. Overlap of the C1′–H1′ crosspeaks of G3 and C10 and those of C9 and C16 in the $^{1}H/^{13}C$ correlation spectra prevented the extraction of proton T_1 and heteronuclear NOE values for these nucleotides. (**c**) Calculated average r.m.s.d. for the C1′ positions in the structures presented in Figure 5a.

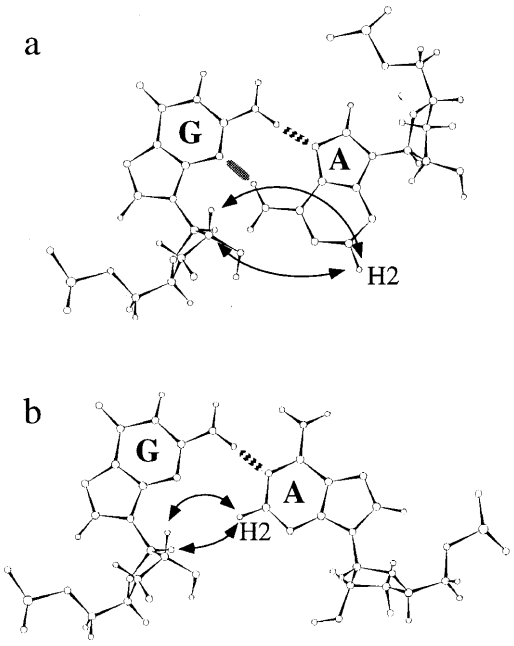

Figure 7. Schematic drawing of planar G·A base pairs. (**a**) Configuration of the commonly observed two hydrogen bond pair, (**b**) single hydrogen bond base pair proposed for G6·A11 of RLP3. Dotted lines denote putative hydrogen bonds which were set to 1.7 Å. Arrows indicate the relative AH2–GH1′/H2′ proximities for each configuration (see Figure 2). The drawing was generated with *anti*-glycosidic angle C3′-endo nucleotides, using Insight II (Biosym Technologies, Inc.).

In RLP3, the first loop base is guanine which could allow formation of a G·A base pair. Turner and co-workers have reported the presence of G·A base pairs in positions 1 and 6 of 45% of the hexanucleotide loops in ribosomal RNA and that this base pairing contributes to overall loop stability (37). A study on the effects of the flanking sequences of G·A mismatches in duplex DNA concluded that the most common conformation comprised GN2H amino-AN7 and AN6H amino-GN3 hydrogen bonds (38) (Fig. 7a). Recent structural studies of a GUAAUA hexanucleotide loop reported this type of G·A pairing adjacent to the loop closing C·G base pair (39,40). The non-hydrogen bonded G imino proton resonance was reported to resonate between 10.5 and 11 p.p.m.

The chemical shift of the G6 imino proton of RLP3 (9.89 p.p.m.) is close to the reported value for G·A base pairs. No informative NOEs could be observed with this imino proton as it exchanged rapidly with H_2O. Typical hydrogen bonding for a G·A base pair (Fig. 7a) was not consistent with the relatively intense A11H2–G6H1′/H2′ NOEs observed at 60 ms NOESY mixing time in D_2O. Rather, the NOEs are more consistent with the formation of a single hydrogen bond G·A base pair, involving G6N2H and A11N1, as shown in Figure 7b. This base pairing mode would situate A11H2 closer to G6H1′ and G6H2′ (a distance of <4 Å was estimated from the NOE intensities). The single hydrogen bond might also account for the slight upfield shift of the G6 imino resonance (from ~10.5 to 9.89 p.p.m.).

Loop mobility

With less than six internucleotide constraints per nucleotide for the loop and closing base pair region, it is perhaps not surprising that a highly converged ensemble of structures was not generated. Two factors may account for the lack of NOEs in the loop: sample concentration and loop dynamics. At a sample concentration of 0.6 mM, the vast majority of the *detected* NOEs were assigned, but the possibility that other NOEs would be detectable in a more concentrated sample cannot be ruled out. Attempts at 3D ^{13}C edited NOESY experiments were not fruitful, due to an unfavorable signal to noise ratio. Nonetheless, it is unlikely that additional medium to strong NOEs would be detected in a more concentrated sample. Exchange of the amino resonances with water impeded the detection of any NOEs involving exchangeable protons in the loop, but this phenomenon should not show concentration dependence.

An alternative explication for the small number of loop NOEs is the intrinsic dynamic nature of the loop nucleotides. Figure 6 shows the increase in heteronuclear NOE and the decrease in H1′ T_1 which reflect increased molecular motion in the loop. Fast internal dynamics may well reduce the number of observable NOEs. In fact, the internal motions are such that the actual structure of the loop may not be describable by a single conformation. The intermediate sizes of the U8(H1′–H2′) and A11(H4′–H5′) coupling constants are more likely the result of conformational averaging rather than the result of a static structure with energetically unfavorable sugar puckers and eclipsed γ angle geometries. Relaxation measurements of the type performed here are straightforward experimentally and provide

an excellent means of monitoring dynamics which otherwise can only be inferred from NOE and coupling constant data.

Biological significance

The objective of this study was to explore the structural features of the 5′-AANCCA loops that might account for their strong conservation in the IRES of polioviruses and members of the E/R viral families (Klinck *et al*., submitted). Our results show that 5′-AAUCCA adopts a distinct 'two sided' structure in solution, the bases from the 5′ side being exposed in the major groove while those on the 3′ side are more aligned with the stem.

Residue 8 appears to be both the most isolated nucleotide and at the 'turning point' of the loop. This is interesting in light of the fact that this nucleotide shows no conservation in sequence alignments of loop III and VI in E/R IRESes. In fact, based on the similarity of the NMR spectra of RLP1 and RLP2, it appears that there are no significant structural differences when U8 is replaced by A8. By virtue of their exposure to the exterior of the molecule the first three loop bases may serve as a site for ribosomal subunit and/or *trans* acting factor recognition in the course of internal initiation of translation. Another possible interpretation is that these bases are positioned to allow access for tertiary RNA interactions within the IRES involving C9–A11. Le and co-workers (6) suggested that the bases ANCCA in loop VI (Fig. 1) may make base-pairing contacts with a complementary sequence 50 nt downstream. They reported the possibility of pseudoknot formation involving Watson–Crick pairing of nucleotides A7–A11. The continuous base stacking from C9 to G12 observed in our structures suggests that base pairing to form a pseudoknot should only involve nucleotides C9 to A11. Pseudoknot formation would be stabilized by the coaxial stacking of the existing and newly formed helical stems. Indeed, the absolute conservation of the CCA loop VI sequence in all E/R IRESes reported to date might be an indication of a common site of tertiary RNA interaction. The existence of CCA sequences in loops III and VI opens the possibility of a long-range competition between between loops III and VI for base pairing to the downstream sequence. We are currently investigating the potential of AANCCA loops to form pseudoknots in model oligonucleotides.

A number of recent NMR studies of hexanucleotide loops have been reported, describing the structural features of the iron responsive element, IRE (41), the *trans* activation response element, TAR (42–44), and large subunit ribosomal RNA (39,40). At present, there appears to be no common structural motif linking all of these sequences. A wide range of structural features are proposed: non-canonical base pairing, U-turn motifs (45) and varying degrees of base stacking and loop dynamics. These results together with those presented here underscore the large conformational diversity of hexanucleotide loops.

Supplementary material

See supplementary material available in NAR Online.

ACKNOWLEDGEMENTS

We thank Nahum Sonenberg for introducing us to the poliovirus IRES and for critical reading of the manuscript. This research was supported by a Natural Sciences and Engineering Research Council Fellowship to R.K., a McGill Majors Student Fellowship to T.S., and a Medical Research Council operating grant and salary award to K.G.

REFERENCES

1 Jackson,R.J., Howell,M.T. and Kaminski,A. (1990) *Trends Biochem. Sci.*, **15**, 477–483.
2 Jackson,R.J. and Kaminski,A. (1995) *RNA*, **1**, 985–1000.
3 Pelletier,J. and Sonenberg,N. (1988) *Nature*, **334**, 320–325.
4 Jang,S.K., Kräusslich,H.-G., Nicklin,M. J.H., Duke,G.M., Palmenberg,A.C. and Wimmer,E. (1988) *J. Virol.*, **62**, 2636–2643.
5 Pilipenko,E.V., Blinov,V.M., Romanova,L.I., Sinyakov,A.N., Maslova,S.V. and Agol,V.I. (1989) *Virology*, **168**, 201–209.
6 Le,S.-Y., Chen,J.-H., Sonenberg,H. and Maizel,J.V. (1992) *Virology*, **191**, 858–866.
7 Nicholson,R., Pelletier,J., Le,S.-Y. and Sonenberg,N. (1991) *J. Virol.*, **65**, 5886–5894.
8 Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.
9 Wyatt,J.R., Chastain,M. and Puglisi,J.D. (1991) *BioTechniques*, **11**, 764–769.
10 Batey,R.T., Inada,M., Kujawinski,E., Puglisi,J.D. and Williamson,J.R. (1992) *Nucleic Acids Res.*, **20**, 4515–4523.
11 Nikonowicz,E.P., Sirr,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) *Nucleic Acids Res.*, **20**, 4507–4513.
12 Varani,G. and Tinoco,I.,Jr. (1991) *Q. Rev. Biophys.*, **24**, 479–532.
13 Heus,H.A., Wijmenga,S.S., van de Ven,F. J.M. and Hilbers,C.W. (1994) *J. Am. Chem. Soc.*, **116**, 4983–4984.
14 Marino,J.P., Schwalbe,H., Anklin,C., Bermel,W., Crothers,D.M. and Griesinger,C. (1994) *J. Am. Chem. Soc.*, **116**, 6472–6473.
15 Kay,L.E., Ikura,M. and Bax,A. (1990) *J. Am. Chem. Soc.*, **112**, 888–889.
16 Fesik,S., Eaton,H.L., Olejniczak,E.T., Zuiderweg,E., McIntosh,L.P. and Dahlquist,F.W. (1990) *J. Am. Chem. Soc.*, **112**, 886–888.
17 Nikonowicz,E.P. and Pardi,A. (1993) *J. Mol. Biol.*, **232**, 1141–1156.
18 Marino,J.P., Schwalbe,H., Glaser,S. J. and Griesinger,C. (1996) *J. Am. Chem. Soc.*, **118**, 4388–4395.
19 Hines,J.V., Varani,G., Landry,S.M. and Tinoco,I.,Jr. (1993) *J. Am. Chem. Soc.*, **115**, 11002–11003.
20 Kellogg,G.W., Szewczak,A.A. and Moore,P.B. (1992) *J. Am. Chem. Soc.*, **114**, 2727–2728.
21 Sklenar,V., Miyashiro,H., Zon,G., Miles,H.T. and Bax,A. (1986) *FEBS Lett.*, **208**, 94–104.
22 Wijmenga,S.S., Heus,H.A., Leeuw,H.A.E., Hoppe,H., van der Groat,M. and Hilbers,C.W. (1995) *J. Biol. NMR*, **5**, 82–86.
23 Ludvigsen,S., Andersen,K.V. and Poulsen,F.M. (1991) *J. Mol. Biol.*, **217**, 731–736.
24 Schwalbe,H., Marino,J.P., Glaser,S. J. and Griesinger,C. (1995) *J. Am. Chem. Soc.*, **117**, 7251–7252.
25 Varani,G., Aboul-ela,F. and Allain,F.H.-T. (1996) *Prog. NMR Spec.*, **29**, 51–127.
26 Gorenstein,D.G. (1984) In Gorenstein,D.G. (ed.) *Phosphorus-31 NMR: Principles and Applications*. Academic Press, New York.
27 Major,F., Turcotte,M., Gautheret,D., Lapalme,G., Fillion,E. and Cedergren,R. (1991) *Science*, **253**, 1255–1260.
28 Brünger,A.T. (1992) X-PLOR, 3.1 Ed., Yale University Press, New Haven.
29 Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, Berlin.
30 Johnson,C.E. and Bovey,F.A. (1958) *J. Chem. Phys.*, **29**, 1012–1014.
31 Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
32 Wimberly,B. (1994) *Struct. Biol.*, **1**, 820–827.
33 Remin,M. and Shugar,D. (1972) *Biochem. Biophys. Res Commun.*, **48**, 636–642.
34 Wijmenga,S.S., Mooren,M.M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed.) *NMR of Macromolecules, A Practical Approach*. Oxford University Press, Oxford, Vol. 134.
35 Westhof,E., Dumas,P. and Moras,D. (1988) *Biochimie*, **70**, 145–165.
36 Shen,L.X. and Tinoco,I.,Jr. (1995) *J. Mol. Biol.*, **247**, 963–978.
37 Serra,M.J., Axenson,T.J. and Turner,D.H. (1994) *Biochemistry*, **33**, 14289–14296.
38 Li,Y. and Agrawal,S. (1995) *Biochemistry*, **34**, 10056–10062.
39 Huang,S., Wang,Y.-X. and Draper,D.E. (1996) *J. Mol. Biol.*, **258**, 308–321.
40 Fountain,M.A., Serra,M.J., Krugh,T.R. and Turner,D.H. (1996) *Biochemistry*, **35**, 6539–6548.
41 Laing,L.G. and Hall,K.B. (1996) *Biochemistry*, **35**, 13586–13596.
42 Colvin,R.A., White,S.W., Garcia-Blanco,M.A. and Hoffman,D.W. (1993) *Biochemistry*, **32**, 1105–1112.
43 Jaeger,J.A. and Tinoco,I.,Jr. (1993) *Biochemistry*, **32**, 12522–12530.
44 Aboul-ela,F., Karn,J. and Varani,G. (1995) *J. Mol. Biol.*, **253**, 313–332.
45 Quigley,G. J. and Rich,A. (1976) *Science*, **194**, 796–806.