2022–2035 *Nucleic Acids Research, 1996, Vol. 24, No. 11*

The NMR structure of 31mer RNA domain of *Escherichia coli* RNase P RNA using its non-uniformly deuterium labelled counterpart [the 'NMR-window' concept]

C. Glemarec, J. Kufel[1], A. Földesi, T. Maltseva, A. Sandström, L. A. Kirsebom[1] and J. Chattopadhyaya*

Departments of Bioorganic Chemistry and [1]Microbiology, Box 581, Biomedical Centre, University of Uppsala, S-751 23 Uppsala, Sweden

Received March 7, 1996; Revised and Accepted April 12, 1996

ABSTRACT

The NMR structure of a 31mer RNA constituting a functionally important domain of the catalytic RNase P RNA from *Escherichia coli* is reported. Severe spectral overlaps of the proton resonances in the natural 31mer RNA (1) were successfully tackled by unique spectral simplifications found in the partially-deuterated 31mer RNA analogue (2) incorporating deuterated cytidines [C5 (>95 atom % ^{2}H), C2′ (>97 atom % ^{2}H), C3′ (>97 atom % ^{2}H), C4′ (>65 atom % ^{2}H) and C5′ (>97 atom % ^{2}H)] [for the 'NMR-window' concept see: Földesi,A. *et al.* (1992) *Tetrahedron*, 48, 9033; Földesi,A. *et al.* (1993) *J. Biochem. Biophys. Methods*, 26, 1; Yamakage,S.-I. *et al.* (1993) *Nucleic Acids Res.*, 21, 5005; Agback,P. *et al.* (1994) *Nucleic Acids Res.*, 22, 1404; Földesi,A. *et al.* (1995) *Tetrahedron*, 51, 10065; Földesi,A. *et al.* (1996) *Nucleic Acids Res.*, 24, 1187–1194]. 175 resonances have been assigned out of total of 235 non-exchangeable proton resonances in (1) in an unprecedented manner in the absence of ^{13}C and ^{15}N labelling. 41 out of 175 assigned resonances could be accomplished with the help of the deuterated analogue (2). The two stems in 31mer RNA adopt an A-type RNA conformation and the base-stacking continues from stem I into the beginning of the loop I. Long distance cross-strand NOEs showed a structured conformation at the junction between stem I and loop I. The loop I–stem II junction is less ordered and shows structural perturbation at and around the G11·C22 base pair.

INTRODUCTION

RNase P is responsible for the maturation of the 5′-termini of almost all tRNAs in the cell. This ubiquitous ribonucleoprotein complex consists of an ~400 nucleotide (nt) RNA, which is the catalytic subunit of RNase P RNA, and a protein subunit (1,2). Bacterial RNase P RNA cleaves various tRNA precursors *in vitro* at the correct positions in the absence of any protein (2). Thus, RNase P RNA is a true ribozyme. In the enzyme–substrate complex the RNA interacts with the two cytosine residues in the conserved 3′-terminal 'RCCA' sequence of a tRNA precursor. The 'GGU-motif' in RNase P RNA that is involved in this interaction is part of an internal loop. This loop structure is also part of a divalent metal ion(s) binding site as revealed by the fact that both Pb^{2+} and Mg^{2+} induce cleavage within this structure. Binding of Mg^{2+} to this region has been suggested to be of functional importance (3,4–6). Hence, knowledge of the three dimensional structure of this domain of RNase P RNA is important to understand the function of this ribozyme. Herein, we have investigated the solution conformation of a 31mer RNA **1** (Fig. 1) representing this domain of *Escherichia coli* RNase P RNA by NMR spectroscopy. In order to stabilize the secondary structure of this domain of RNase P RNA, i.e. nucleotides 250–299, we have excised the nucleotide residues 263–287 and have transplanted a stable C·G base pair with a tetraloop (UUCG) (7) as a link between the residues 262 and 288 to give finally our 31mer RNA **1** using T7 RNA polymerase. This study is a part of our goal to identify the conformational features of the internal loop in 31mer RNA (i.e. Loop I, see Fig. 1). Owing to the relatively large size of the natural 31mer RNA **1**, severe spectral overlap precluded an unambiguous assignment of the proton resonances from the NOESY and DQF-COSY spectra, thereby restricting the number of structural information. We have tackled this problem by using a partially-deuterated 31mer RNA analogue **2** (Fig. 1) in which deuterated cytidine residues [C5 (>95 atom % ^{2}H), C2′ (>97 atom % ^{2}H), C3′ (>97 atom % ^{2}H), C4′ (>65 atom % ^{2}H) and C5′ (>97 atom % ^{2}H)] (39) were incorporated by T7 RNA polymerase reaction. The choice of cytidine was dictated by two factors: (i) The H5-H6 NOE crosspeaks are very strong and often obscure the anomeric-aromatic region of the NOESY spectra, and (ii) the C residues are found in both stems I and II of the 31mer RNA, yet they are the only nucleotide units which are not part of the loop I. Since divalent metal ions cleave the RNase P RNA, we have also

* To whom correspondence should be addressed

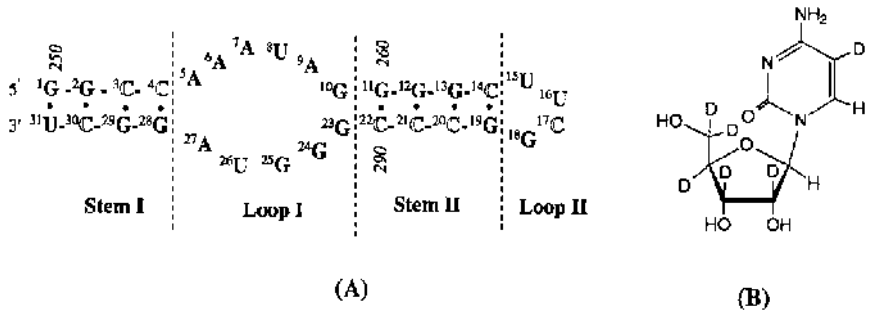

Figure 1. (**A**) Sequence of the natural 31mer RNA **1** and its partially-deuterated analogue **2** [deuterated cytidine residues (B) are shown in the 'outline' font]. The italic numbers are derived from ref. 6 and show the location of the nucleotides in the RNase P RNA subunit. All nucleotides are numbered by superscripts and refer to the notation used in the NMR study. The vertical dotted lines show the different regions of the 31mer RNA: stem I (^{1}G to ^{4}C and ^{28}G to ^{31}U), loop I (^{5}A to ^{10}G and ^{23}G to ^{27}A), stem II (^{11}G to ^{14}C and ^{19}G to ^{22}C), loop II (^{15}U to ^{18}G). (**B**) 5,2′,3′,4′,5′/5″-Hexadeuterated cytidine block [cytidine-d_6] has been incorporated in the deuterated 31mer RNA analogue **2** and shown in the 'outline' font in (A).

studied the effect of Mg^{2+} on this 31mer RNA by NMR spectroscopy.

MATERIALS AND METHODS

Synthesis of 31mer RNA and its partially-deuterated analog by T7 RNA polymerase

C5-Deuterated (20) cytidine-d_6 block was prepared from 2′,3′-*O*-isopropylidine uridine-d_5 using our procedure (19), which was subsequently converted to its 5′-triphosphate (21). The natural 31mer RNA **1** and its deuterated counterpart **2** were synthesized using T7 DNA dependent RNA polymerase as described elsewhere (22,23) and the concentration of DNA template was optimized to 0.15 μM. In the preparation of **2**, cytidine 5′-triphosphate was substituted with cytidine-d_6-5′-triphosphate (39).

NMR spectroscopy

Sample preparation. The natural 31mer RNA **1** and the partially-deuterated 31mer RNA analogue **2** were dissolved in 0.4 ml of a phosphate buffer consisting of 50 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA at pH 6.2. For measurements in $^{2}H_2O$, the sample was lyophilized three times from 99.9% $^{2}H_2O$ and dissolved in 0.5 ml of 99.98% $^{2}H_2O$. The final RNA concentration was 1 mM. All NMR spectra were recorded on a BRUKER AMX 500 MHz instrument.

Exchangeable proton spectra. The 1D spectra in 1:9, v/v, $^{2}H_2O/H_2O$ were recorded using a 1331 binomial solvent suppression pulse sequence (24).

Rate of exchange. A combination of NOESY and ROESY experiments (9) was used to calculate the rates of exchange of the imino protons. The mixing times used were 8, 10, 20, 25, 30 40 and 55 ms for NOESY and 8, 12, 16, 20 and 24 ms for ROESY. The experiments were performed at 5, 8, 11, 14, 17, 20 and 23°C (128 FIDs with 2 K data points, 16 scans per FID with a sweep width of 10200 Hz). A recycle delay of 5 s was used, which gave an optimal intensity of the NOE crosspeaks. The volumes of NOE crosspeaks and diagonal were measured using the program AURELIA (25). The activation energy (E_a) of the exchange process was obtained from Arrhenius plot of the temperature dependent exchange rate.

Spectra in $^{2}H_2O$. 2D NMR spectra of the RNA in 99.98% $^{2}H_2O$ were recorded using the TPPI method (29) and pre-irradiation at low decoupler power of the residual HDO peak. NOESY spectra (26) were recorded at 100, 300 and 600 ms (512 FIDs with 4K data points, spectral width of 5050 Hz, 32–64 scans for each FID). The data were zero-filled to 2 K × 2 K or 2 K × 1 K before applying a 2 Hz line-broadening factor in the F1 and F2 dimensions. Relaxation delays of 3 s were used. The DQF-COSY spectra (27) were acquired with 4K data points in t2 and 512 points in t1 (32–64 scans for each FID with a sweep width of 5050 Hz). The data were zero filled to give a 4K × 2K matrix, and a π/4 shifted sine-square bell window was applied in both dimensions before Fourier transformation. The clean-TOCSY spectra (28) were acquired with 512 spectra of 4K data points. The data were zero filled to give a 4K × 2K matrix, and a π/2 shifted sine-square window was applied in both dimensions before Fourier transformation with the MLEV-17 sequence applied for mixing time using an extra delay of 65 μs for compensation of NOE. Total mixing times of 25, 50 and 100 ms were used.

XPLOR structure refinement

In the distance-geometry [XPLOR v.3.1 (15)], 655 NMR constraints were used: we performed usual substructure embedding followed by regularization [(i) two energy minimizations, (ii) 625 MD steps at 2000 K, (iii) cooling to 100 K over 1000 MD steps] and seven cycles of simulated annealing [(i) from 1000 K to 100 K cooling over 2000 MD steps, (ii) 200 steps of energy minimization. Force constants for the distance constraints were set to 50 kcal $Å^{-2}$ in the first four cycles, followed by three cycles with force constants of 100, 200 and 400 kcal $Å^{-2}$]. The MD step length was set to 3 fs. Extra planarity constraint of 2.0 kcal was used in the first cycle, which was raised gradually to 20 kcal in the fourth cycle and then kept constant. For all dihedral restraints a force constant of 200 kcal rad^{-2} was used.

AMBER structure refinement

The SANDER module of the AMBER 4.0 program package (16) was used both for the MD simulations and the energy minimizations with infinite non-bonded cut-offs and flat bottom potential wells for constraining both distances [the RANDMARDI (30–33,38) assigned widths of the wells were at least ±5% or ±0.1 Å] and dihedral angles (±45° to ±90° for the backbone torsions and ±6.1° to ±21.5° for the endocyclic sugar torsions). Force

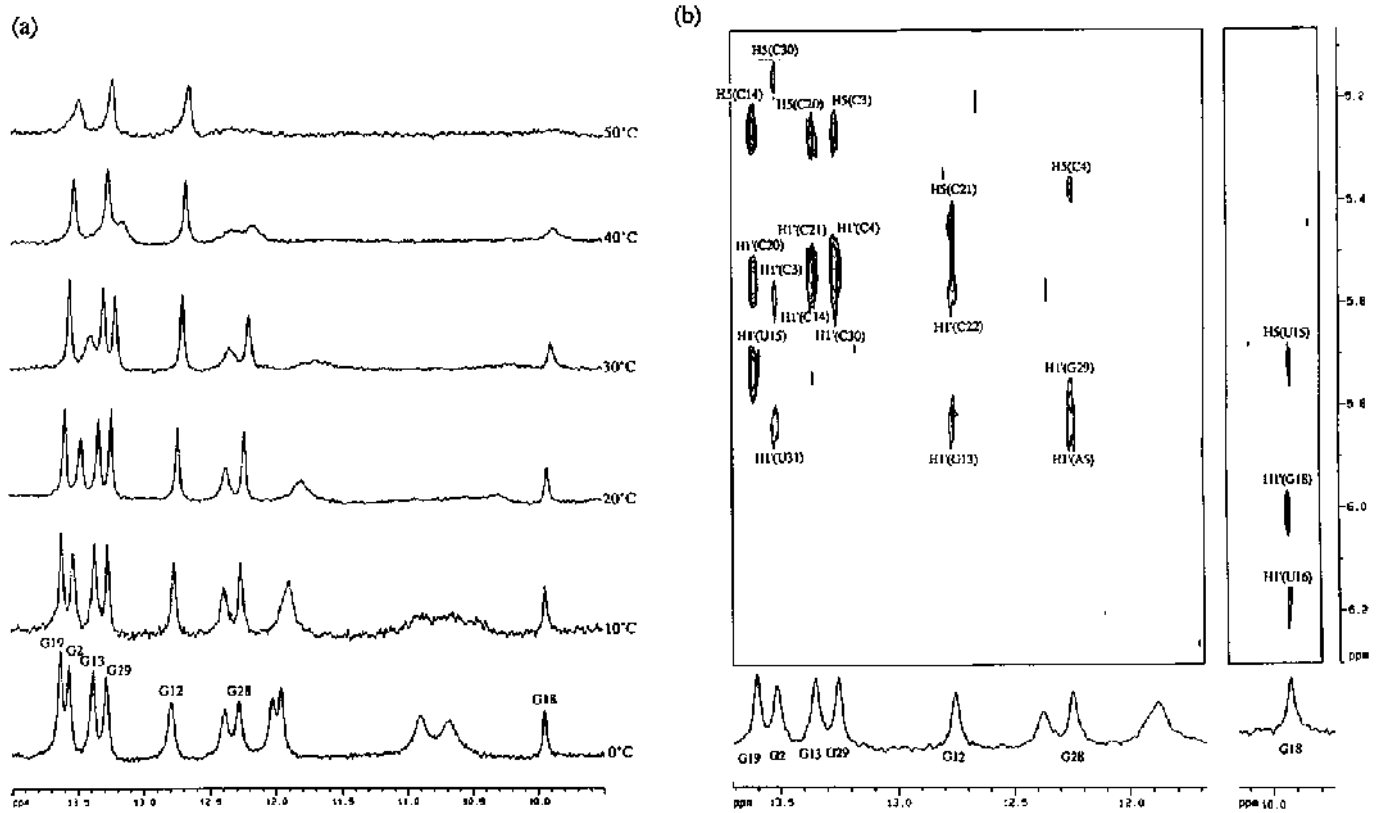

Figure 2. (**a**) ^{1}H-NMR spectra of the exchangeable imino proton of **1** as a function of temperature. The assignment of peaks was made from 2D NOESY spectra. (**b**) Contour plot of a portion of the 600 ms 2D NOESY spectrum of **1** recorded in 90%/10% $H_2O/^2H_2O$ at 10°C and showing the imino proton to H1′/H5 region. Each base pair G imino proton shows three crosspeaks in the H1′/H5 region. These NOE crosspeaks are characteristic of an A-RNA type helix (8).

constants = 40 kcal Å^{-2} for the distances and 40 kcal rad^{-2} for the backbone torsions, and 100 kcal rad^{-2} for endocyclic sugar torsions. In the MD simulations, the following settings were used: (i) time step of 1 fs, (ii) bonds involving hydrogen atoms constrained using shake (17), (iii) constant temperature mode [Berendsen algorithm (18), τ = 0.2 ps], (iv) the force constants were gradually increased from 5 kcal Å^{-2} (5 or 12.5 kcal rad^{-2} for torsions) to 40 kcal Å^{-2} (40 or 100 kcal rad^{-2} for torsions) over the first 2 ps. The AMBER refinement involved (1) Energy minimization for 500 steps. (2a) MD simulation of 7 ps at 400 K. (2b) Cooling to 300 K over a 1 ps period. (2c) MD simulation of 2 ps at 300 K. During the last 2 ps the atomic coordinates were saved at 250 fs intervals. (3) The eight collected conformers were averaged and the average structure was energy minimized.

RESULTS AND DISCUSSION

Assignment of exchangeable protons

The downfield region of the NMR spectrum of **1**, recorded in 10% H_2O–90% D_2O at 0°C, in Figure 2a shows the imino protons of the RNA. The assignments of the exchangeable imino ^{1}H resonances were obtained from the 2D NOESY spectra (data not shown). The starting point for the assignment was made using the assumption that the ^{1}H resonance at 10 p.p.m. belongs to the imino proton of G18 in the hairpin tetraloop [U15-U16-C17-G18, i.e. loop II (7)], which then was used (7a) to sequentially assign the imino protons of G19, G13 and G12 using imino to imino NOEs. Of the six remaining exchangeable resonances, one has NOEs to the resonances of G2 and G28 and hence it was attributed to G29. The two resonances from G2 or G28 could not be distinguished on the basis of imino to imino NOEs alone. The additional three resonances in the chemical shift range of base paired imino protons at 12.4, 12.0 and 11.9 p.p.m. could not be assigned due to more rapid exchange with H_2O. These are attributed to the imino protons of U15 and U16 in the loop II and to the imino proton of G11. No NOE was observed between the G18 and U15 imino protons which would indicate that a G18·U15 wobble pair is not formed in the loop II. On the other hand, the NOE from the G18 imino proton to the H5(U15) together with the *syn* conformation of G18 (see below) would support a reverse-wobble G18·U15 base pair with two hydrogen bonds (7a). The fast rate of exchange with the solvent at 0°C of the U15 imino proton could be responsible for the absence of NOE between the imino protons of U15 and G18. The C-amino protons were assigned from the NOEs from the corresponding base paired G imino protons (data not shown). The downfield resonances (7.9–9.0 p.p.m.) were assigned to the hydrogen bonded amino protons and the upfield resonances (6.4–7.0 p.p.m.) to the non-hydrogen bonded amino protons. Each G imino resonance (G2, G12, G13, G18, G19, G28 and G29) showed three NOE crosspeaks in the H1′/H5 region (Fig. 2B). One of these crosspeaks arises from the G imino proton to H5(C) interaction. The other two arise from interactions between the G imino and the H1′ of the 3′-end sugar on the same strand and from interaction between the G imino and the H1′ of the 3′-end sugar on the opposite strand (8). These two latter NOE crosspeaks are characteristic of an A-RNA type helix (8). Once the non-exchangeable spectrum was assigned (*vide infra*), these NOEs provided a

Table 1. ^{1}H-NMR chemical shifts (p.p.m.) of the natural 31mer RNA **1** at 26°C

residue	H8 / H6	H2 / H5	H1'	H2'	H3'	H4'	H5', H5''	imino	amino
G1	8.20	-	5.91	4.99	4.72	4.62	4.48,4.32	-	-
G2	7.75	-	5.89	4.63[a]	4.56[a]	4.26[a]	c	13.51[d]	-
C3	7.75	5.28	5.61	4.46[b]	c	c	c	-	8.63[d] - 6.96[d]
C4	7.59	5.39	5.51	4.62[b]	4.43[b]	4.11[b]	c	-	7.91[d] - 6.70[d]
A5	8.03	7.52	5.83	4.48	4.18	c	c	-	-
A6	7.73	7.64	5.55	4.25[a]	4.31[a]	c	c	-	-
A7	7.91	-	5.64	4.47	4.61	4.42	4.08, c	-	-
U8	7.55	5.49	5.68	4.25	4.56[a]	4.33[a]	c	-	-
A9	8.12	8.12	5.79	4.71	4.62	4.43	4.12,c	-	-
G10	7.86	-	5.86	4.79	4.59[a]	4.35[a]	4.26[a],c	-	-
G11	7.86	-	5.62	4.73	4.43[a]	4.53[a]	c	-	-
G12	7.38	-	5.82	4.59	4.52	4.18	c	12.75[d]	-
G13	7.31	-	5.81	4.57	4.49	4.09	c	13.34[d]	-
C14	7.42	5.26	5.55	4.56	4.24	4.06	c	-	8.93[d] - 6.48[d]
U15	7.83	5.76	5.72	3.85	4.41[a]	4.56[a]	c	-	-
U16	8.07	5.90	6.14	4.72	4.27	4.06	c	-	-
C17	7.74	6.18	6.01	4.14	4.53	3.83	3.65, 3.83	-	-
G18	7.89	-	6.00	4.89	5.65	4.44	4.22, 3.83	9.92[d]	-
G19	8.33	-	4.44	4.44	4.31	c	c	13.60[d]	-
C20	7.76	5.31	5.57[a]	4.49[b]	4.45[b]	4.43[b]	c	-	8.91[d] - 6.98[d]
C21	7.82a	5.58	5.54	4.42[b]	-	4.09[b]	c	-	8.57[d] - 6.98[d]
C22	7.64	5.45	5.49	4.52[b]	4.39[b]	4.11[b]	c	-	-
G23	-	-	5.61	4.37	4.13[a]	c	c	-	-
G24	7.63[a]	-	5.45[a]	4.43	4.35[a]	c	c	-	-
G25	7.81	-	5.62	4.65	4.74[a]	c	c	-	-
U26	7.72	5.66	5.83	4.48	4.69[a]	c	c	-	-
A27	8.22	7.88	6.07	4.86	4.70	4.43	4.31, 4.31	-	-
G28	7.70	-	5.20	4.63	4.48	4.40[a]	4.20[a], 4.20[a]	12.23[d]	-
G29	7.34	-	5.78	4.50	4.63	4.09	4.21	13.25[d]	-
C30	7.55	5.18	5.59[a]	4.44[b]	4.33[b]	4.51[b]	c	-	8.61[d] - 7.00[d]
U31	7.81	5.67	5.84	4.20	4.05[a]	c	c	-	-

[a] Could only be assigned unambiguously on the basis of the data from the partially-deuterated 31-mer RNA analogue **2**)
[b]Invisible in the partially-deuterated 31mer RNA **2** analogue due to substitution of these protons by deuterium, but could be assigned unambiguously by comparison of the NOESY spectra of the natural 31mer RNA **1** with that of the partially-deuterated 31mer RNA analogue **2**.
[c]Could not be extracted from the NMR spectra of the natural 31mer RNA **1** and from the NMR spectra of the partially-deuterated 31mer RNA **2** because of the crowding or a lack of the sequential connectivity.
[d]Have been observed in both the natural 31mer RNA **1** and partially-deuterated 31mer RNA **2** dissolved in 90% H_2O–10% $^{2}H_2O$ at 10°C.

further support for the assignment of the G imino protons, and for the distinction between G2 and G28 imino protons. The chemical shifts of the imino and amino protons are listed in Table 1.

Temperature dependence of imino protons. Figure 2a shows the temperature-dependent NMR spectra of imino protons. At 0°C, all resonances are relatively sharp except for the broad peaks around 10.5–11 p.p.m. These broad resonances are most likely due to non-hydrogen-bonded imino protons of U and/or G residues located in the internal loop and/or at the 3′ and 5′-termini, which are not well protected from solvent. The three non-assigned imino resonances at 11.9, 12.0 and 12.4 p.p.m. broaden first at 10°C. The G2 and G18 imino resonances start broadening at 20°C. Above 30°C, the G28 and G29 resonances begin to broaden and vanish at 50°C. The G19, G12 and G13 resonances corresponding to the base pairs of the stem II only start to broaden at this temperature. These data show that the base pairs located in the stem II near the loop II are more stable than the base pairs constituting the stem I of the 31mer RNA. The behaviour of G19 also shows that the loop II protects the H-bonded imino proton from a rapid solvent exchange and increase the stability of the base pair.

Rate of exchange of the imino protons with water. A combination of NOESY and ROESY experiments (9) has been used to evaluate the rate of exchange (k_{ex}) of the imino protons of **1** at pH 6.2 with water at 3°C intervals from 5 to 23°C (Table 2). The k_{ex} of resonances at 11.9 and 12.0 p.p.m. could not be measured even at 5°C because of rapid exchange. The imino proton at 12.4 p.p.m. exchanges fastest, followed by G2 and G18. The rates of exchange of G28, G12 and G13 imino protons are quite similar while G19 and G29 exchange the slowest at 23°C. The resonance at 12.4 p.p.m. has a chemical shift close to G28 imino proton and is in the chemical shift range of a G·C base pair, thereby speculating that it corresponds to the G11·C22 base pair. The fact that G11 and G28, located at the two ends of loop I, and yet the k_{ex} of G11 is much faster than that of G28 at all temperatures suggests that the access of water to the G28·C4 base paired imino proton is much reduced compared with the G11·C22 base pair. The reduced water activity around G28·C4 is most probably due to stronger stacking interactions (see below). The slow rate of exchange of G19 imino proton shows that the loop II strengthens (7a) the stability of hydrogen bonding of G19·C14 base pair. The estimation of the E_a (Table 2) for exchange of imino protons with

Table 2. Rate constants k_{ex} (s^{-1}) and E_a (kcal.mol^{-1}) determined for the exchange process of the imino protons of **1** by NOESY[#] experiments without Mg^{2+} (a) and with Mg^{2+} (b)

T		G2-C30	G11-C22	G12-C21	G13-C20	G18-U15	G19-C14	G28 - C4	G29 - C3
5°C	a	3.9±0.3	15.8± 1.2	2.4 ± 0.3	0.9 ± 0.1	1.5 ± 0.3	*	5.1 ± 0.6	*
8°C	a	6.1 ± 0.7	20.6 ±2.6	1.8 ± 0.5	1.5 ± 0.3	2.4 ± 0.1	*	1.2 ± 0.2	*
11°C	a	8.3 ± 0.2	28.9 ±1.0	2.9 ± 0.2	2.1 ± 0.1	3.5 ± 0.2	0.5 ± 0.2	2.2 ± 0.2	*
	b	7.1 ± 0.2	30.8 ± 5.4	4.5 ± 0.1	2.3 ± 0.5	4.7 ± 0.7	0.8 ± 0.3	2.8 ± 0.9	*
14°C	a	14.4± 0.6	44.9± 1.9	4.4 ± 1.2	3.1 ± 0.3	5.4 ± 0.4	0.9 ± 0.2	4.1 ± 0.4	*
	b	12.3 ± 0.8	46.1 ± 2.8	5.7 ± 0.3	3.5 ± 0.4	8.5 ± 0.7	0.9 ± 0.1	3.6 ± 0.9	*
17°C	a	21.9± 1.2	57.3	5.5 ± 0.5	3.7 ± 0.2	7.8 ± 0.5	1.2 ± 0.1	6.2 ± 0.4	0.7 ± 0.1
	b	19.7 ± 0.6	77 ± 13	7.0 ± 0.2	4.8 ± 0.4	16.7 ± 1.8	1.6 ± 0.1	5.4 ± 0.1	*
20°C	a	34.9± 2.5	92.1 ± 19	7.1 ± 0.2	5.4 ± 0.8	12.9 ±0.6	2.1 ± 0.2	8.3 ± 0.2	1.6 ± 0.2
	b	35.0 ± 2.2	114.9 ± 26	10.6 ± 1.1	7.2 ± 0.9	26.5 ± 3.6	2.5 ± 0.1	7.8 ± 0.7	0.8 ± 0.2
23°C	a	58.2± 5.3	**	9.4 ± 0.3	7.1 ± 0.2	19.3 ± 1	3.0 ± 0.2	10.8± 0.4	2.8 ± 0.4
	b	57.9 ± 6.1	**	14.1 ± 0.9	8.2 ± 0.5	41.4 ± 5.1	4.3 ± 0.3	9.8 ± 0.4	1.5 ± 0.1
E_a	a	26.8 ± 1.0	20.6 ± 2.8	15.8 ± 1.6	16.8 ± 1.6	24.0± 1.0	24.8± 2.2	22.0 ± 3.6	39.6 ± 6.4
E_a	b	29.4 ± 0.8	24.8 ± 1.4	16.2 ± 2.0	18.2 ± 2.4	30.8 ± 2.2	24.6 ± 4.2	18.4 ± 1.6	36.4[¤]

[#]The k_{ex} calculated from NOESY and ROESY were very similar, hence only the k_{ex} from NOESY are reported.
*Cannot be calculated due to slow exchange.
**Cannot be calculated due to fast exchange.
[¤]Only two points could be measured at 20 and 23°C.

the bulk water showed that they can be distinguished in two groups: the G29 proton has a high E_a (≈39 kcal mol^{-1}) while all other protons have E_a between 17 and 25 kcal mol^{-1}.

Assignment of non-exchangeable protons

Aromatic and anomeric protons. All 13 H5-H6 crosspeaks were assigned to each of the 13 pyrimidine residues in the DQF-COSY spectra of **1**, showing that a single conformation predominates under our NMR measurement conditions. The U residues could be easily distinguished (10) from the C residues in the DQF-COSY spectra of **2** in which all C residues are C5 deuterated, resulting (10) in a disappearance of the H5-H6 crosspeaks.

The assignment of the non-exchangeable protons was made using the connectivities H8/H6(i)-H1′(i)-H8/H6(i+1) in the 2D NOESY spectra of **1** at 26°C (Fig. 3A). The weak H8/H6(i) to H5(i+1) NOEs were also used together with the sequence information to determine as well as to confirm the assignment. Most of the ambiguities encountered in the resonance assignment procedure owing to the spectral overlap could be nicely resolved from the NOESY spectrum recorded at 40°C.

We found typical A-RNA features (11,12) such as two crosspeaks for each H8/H6 except for the one at 8.2 p.p.m. which was then assigned to the H8(G1). Similarly, we found two crosspeaks for each H1′ except for the one at 5.85 p.p.m. which was then assigned as the H1′(U31). Figure 3a shows the anomeric-aromatic walk from G1 to C17. At 26°C, the H8(G2) and H6(C3) have the isochronous chemical shifts, but these two resonances are well separated at 40°C. The NOE crosspeaks between H8(G13)-H1′(G13) and H8(G13)-H1′(G12) are not resolved at the temperatures investigated (10, 17, 26 and 40°C). The sequence specific connectivity was confirmed when Mg^{2+} was added. In this case, the NOESY spectra at 26 and 40°C showed two slightly resolved H8(G13)-H1′(G13) and H8(G13)-H1′(G12) crosspeaks (data not shown). The U16 to C17 NOE is very weak but an NOE from H6(U16) to H5(C17) confirms the sequential assignment. The sequence connectivity stops at C17. No NOE was observed between C17 and G18. However, the chemical shifts for the ^{1}H resonances of the C14-(U15-U16-C17-G18)-G19 hairpin loop II part of **1** and **2** were nearly identical to those reported by Varani *et al.* (7a). Thus, U16 and C17 were easily assigned from their appearance as a doublet in the H1′ dimension of the NOESY spectrum which indicates that the ribose of U16 and C17 is in the S-type conformation. Similarly G18 was easily assigned in the NOESY spectrum from its very strong H8-H1′ crosspeak which indicates a *syn* conformation. The very unusual downfield shift of H3′(G18) at 5.5 p.p.m. is characteristic of a purine nucleotide in a *syn*-North conformation and is again similar to what has been described by Varani *et al* (7a).

The anomeric-walk from C20 to U31 was complicated due to the strong H5-H6 NOE crosspeaks which obscured some of the weaker aromatic/anomeric NOEs that are critical for a full assignment. Thus, the H6(C21)-H1′(C20), H1′-H8(G24) and H6(U31)-H1′(C30) NOE crosspeaks could not be assigned unambiguously in the natural 31mer. The problem was neatly solved by examining the NOESY spectrum of the partially-deuterated analogue **2** (Fig. 3b) where the H5-H6 crosspeaks of the C residues are absent. In this spectrum, it has been possible to assign those important NOE crosspeaks which are hidden in the NOESY spectrum of the natural RNA **1** due to spectral overlap with the H5-H6 crosspeaks of C22 with H8-H1′(G24) and of C21 with H6(U31)-H1′(C30) and H6(C21)-H1′(C20) (compare Fig. 3a with b). This ability of unambiguous assignment with the help of partially-deuterated 31mer RNA analogue **2** is particularly important for the loop I protons where the assumption that the NOEs are sequential as in the helical region cannot be made. Thus, all aromatic and anomeric resonances in natural 31mer RNA **1** have been assigned unambiguously with the help of the partially-deuterated analogue **2** and their chemical shifts are shown in Table 1.

The observation of sequential anomeric-aromatic connectivities within the loop I indicates that the stacking is an important conformational feature. Some NOE connectivities are however weak at some points in the loop I suggesting specific structural perturbations as in A7-U8 and U8-A9, C22-G23 and G25-U26

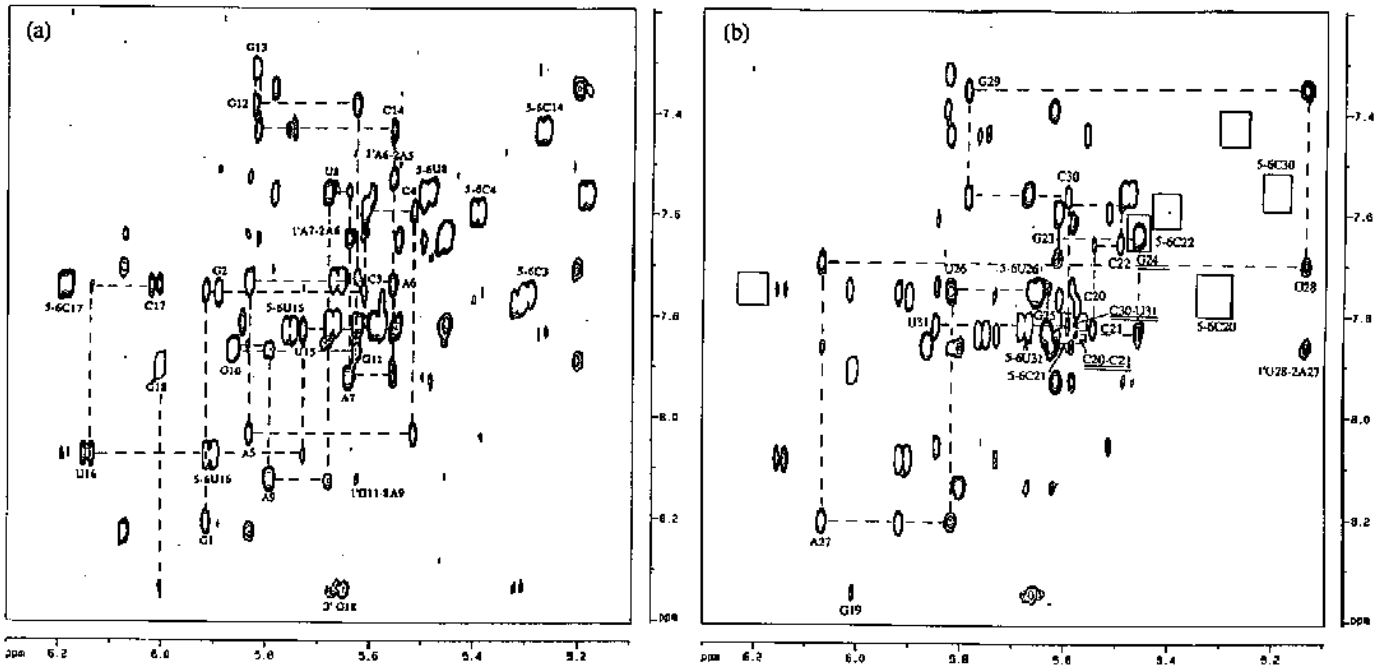

Figure 3. (**a**) The anomeric-aromatic region of the 300 ms 2D NOESY spectrum of the natural 31mer RNA**1** recorded in ^2H_2O at 26°C. The anomeric-aromatic NOE walk from G1 (crosspeak at 8.20–5.91 p.p.m.) to C17 is indicated by dotted lines. The H5-H6 crosspeaks for the cytidine and uridine residues are also indicated. (**b**) The aromatic-anomeric region of the 300 ms 2D NOESY spectrum of the partially deuterated 31mer RNA**2** recorded in ^2H_2O at 26°C. The NOE walk from C20 (crosspeak at 7.76–5.57 p.p.m.) to U31 is indicated by dotted lines. The H5-H6 crosspeaks of the cytidine residues have vanished due to deuteration at the C5 position in**2**. Only the H5-H6 crosspeaks of the uridine residues are present. Some important NOE crosspeaks hidden due to spectral overlap in the natural 31mer RNA**1** are now easily assigned: H8-H1′(G24), H6(U31)-H1′(C30), H6(C21)-H1′(C20). The peak denoted G19 represents the H8(G19)-H1′(G18) crosspeak.

junctions. This is also corroborated by the fact that at 40°C, no NOE connectivity was found between C22-G23 and between G25-U26 residues. Although sequential connectivities are observed from A9 to G10 to G11, weak non-sequential NOEs between H8(A9) and H1′(G11) and between H8(G11) and H1′(A9) are also observed suggesting that G10 might be bulged out while A9 and G11 are stacked on each other.

Assignment of the H2 resonances of adenosine.
Intrastrand NOEs. In A-RNA, H2(A), in the minor groove, is close to both the H1′ of its 3′-neighbour on the same strand and to the H1′ at the 3′-end of its base pair on the opposite strand. At both 100 and 300 ms mixing times at 26 and 40°C, three of the five H2(A) protons give NOEs to the H1′ of their respective 3′-neighbouring residue: A5, A6 and A27. The H2(A7) was identified only at 40°C by the NOE observed at 100 and 300 ms with the H1′ of U8. The assignments of H2(A5), H2(A6) and H2(A27) were also confirmed by the weak NOEs at 300 ms to their own H1′. The H2(A9) did not show any NOE with its H1′ or with the H1′ of its 3′-neighbour, but a weak NOE with H1′(G11) and H1′(U8) was found at 26°C.

Interstrand NOEs. Eight interstrand NOEs have been observed at 26 and 40°C involving H2(A) and H1′. At 300 ms, the H2(A5) shows an NOE with H1′(G28) and H2(A27). At 100 and 300 ms, an NOE is observed between H2(A6)-H1′(A27) and between H2(A6)-H1′(U26). The aromatic/aromatic region of the 300 ms NOESY spectrum at 26°C also shows an NOE between H2(A6)-H2(A27). No interstrand NOE was observed for H2(A7). At 26°C, an NOE is observed between H2(A9)-H1′(G25), both at 100 and 300 ms. Finally, at 100 and 300 ms at 26 and 40°C, the H2(A27) shows an NOE with the H1′(A6). These interstrand NOEs together with the extra imino resonance observed in a region where hydrogen bonded imino protons appear provide evidence that the loop I is indeed structured.

Non-anomeric sugar protons. For a molecule of such a size, the assignment of the non-anomeric sugar protons is a considerable challenge due to severe spectral overlap. This problem was tackled by analyzing the NMR spectra of **2**. The NOESY spectra of **2** are strongly simplified compared to that of **1** since all crosspeaks involving the H5, H2′ to H5′/H5″ of all C residues do not appear in the NMR spectra (compare Fig. 4a and b, and Fig. 5a and b). In this way, it has been possible to assign 22 additional non-anomeric sugar protons from the A, G and U residues of loop I by comparing the NMR spectra of **1** and **2**.

The H2s were identified by the strong NOE to H1′ at short mixing time (100 ms) using high contour levels as well as from the crosspeaks in the DQF-COSY spectra. The H1′/H5 to H2′ region of the NOESY spectrum of **1** is crowded in the 5.4–5.7 p.p.m. and 4.35–4.65 p.p.m. range, and the assignments of H1′-H2′ crosspeaks of A6, A7, G23, G24, G25, C3, C20, C21 and C30 were achieved by comparing the NOESY spectra of **1** and **2** where all H1′-H2′ crosspeaks involving the C residues are masked (compare Fig. 5a and b).

In the DQF-COSY spectrum of **1**, only 14 crosspeaks are present indicating that the corresponding sugars have an appreciable amount of S-type conformation (Table 3). A6, G23, G24, A27 of loop I and the terminal G1 and U31 residues have coupling

Table 3. $^{3}J_{1'2'}$ (Hz) coupling constants and sugar conformation[a] of **1** at 26°C in $^{2}H_2O$

Nucleotide	G1	G2	C3	C4	A5	A6	A7	U8	A9	G10	G11
$^{3}J_{1'2'}$	4	< 2	< 2	< 2	< 2	5	7	8	8	7	< 2
% North	56	100	100	100	100	42	13	0	0	13	100
Nucleotide	**G12**	**G13**	**C14**	**U15**	**U16**	**C17**	**G18**	**G19**	**C20**	**C21**	**C22**
$^{3}J_{1'2'}$	< 2	< 2	< 2	< 2	8	8	< 2	< 2	< 2	< 2	< 2
% North	100	100	100	100	0	0	100	100	100	100	100
Nucleotide	**G23**	**G24**	**G25**	**U26**	**A27**	**G28**	**G29**	**C30**	**U31**		
$^{3}J_{1'2'}$	6	6	8	8	5	< 2	< 2	< 2	4		
% North	27	27	0	0	42	100	100	100	56		

[a]The sugar conformation was estimated from the H1′-H2′ coupling constants measured in the DQF-COSY spectrum using the equation: % N = 100 × (7.9 – $J_{1'2'}$)/6.9 (45).

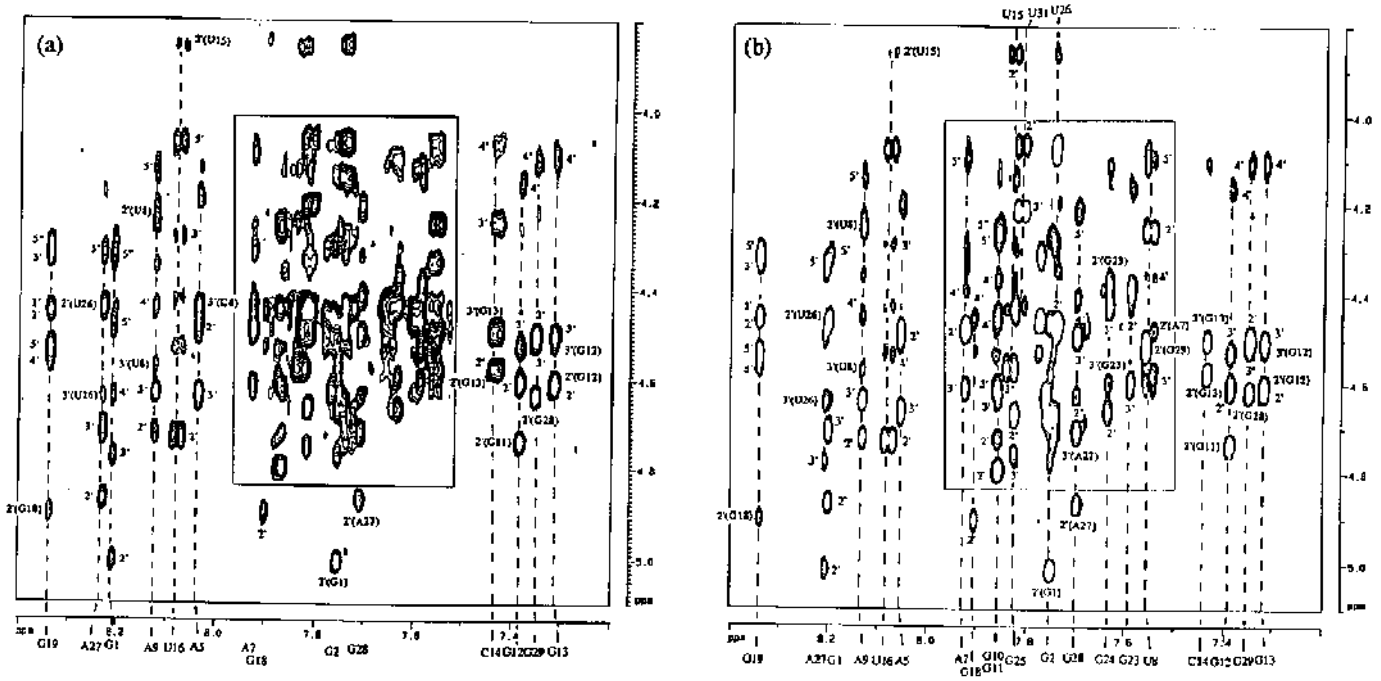

Figure 4. Contour plots of the aromatic to H2′, H3′, H4′, H5′ and H5″ region of the 300 ms 2D NOESY spectrum. (**a**) The natural 31mer RNA **1** at 26°C. The region between 8 and 7.5 p.p.m. is crowded and an unambiguous assignment or accurate determination of NOE volumes is impossible. (**b**) The partially deuterated 31mer RNA **2**. The region between 8 and 7.5 p.p.m. is relatively much less crowded compared with the natural counterpart due to the banishment of all crosspeaks involving the H2′, H3′, H4′ and H5′/H5″ of the cytidine residues. The residual crosspeaks are much sharper allowing an easier and more accurate extraction of the NOE volumes used for molecular modelling.

constants between 4–6 Hz. On the other hand, the A7, U8, A9, G10, G25 and U26 sugar residues of the loop I and U16 and C17 sugar residues of the loop II have coupling constants of 7–8 Hz, suggesting that these sugars are more locked in the preferred South conformation. The other sugars (for G2, C3, C4, G28, G29, C30 of stem I and G11, G12, G13, C14, G19, C20, C21, C22 of stem II and U15, G18 of the loop II) do not show any H1′-H2′ crosspeaks indicating that the $^{3}J_{1'2'}$ coupling constants are <2 Hz and that the corresponding sugars are in the N-type conformation. Interestingly, only the A5 residue of the loop I does not show a H1′-H2′ crosspeak indicating that its sugar is in the North conformation. The perusal of all of these $J_{1'2'}$ couplings points to three structural features of 31mer RNA in a qualitative manner: (i) stems I and II have features of A-type RNA conformation (11,12). (ii) Loop II has a conformation very similar to the one reported by Varani *et al.* (7a). (iii) The conformation of loop I is more flexible than that of the loop II, as expected.

The above assignments were then confirmed and extended by using the H8/H6 to H2′/H3′/H4′/H5′/5″ and the H1′/H5 to H2′/H3′/H4′ regions of the NOESY spectra of **1** and **2** (Fig. 4a and b, and Fig. 5a and b). In A-type RNA, the H2′(i)-H6/H8(i+1) distance is very short (≈2 Å). Thus, at 100 ms, the base to sugar region showed strong H8/H6(i) to H2′(i–1) NOE crosspeak connectivities between G1 and U16. The H8(i)-H2′(i+1) crosspeaks were weaker between A6 and G10, which is consistent with the fact that these nucleotides are not in pure N-type conformation. No H2′(i)-H6/H8(i+1) connectivities were observed between U16, C17 and G18. A weak NOE was observed between H8(G19) and

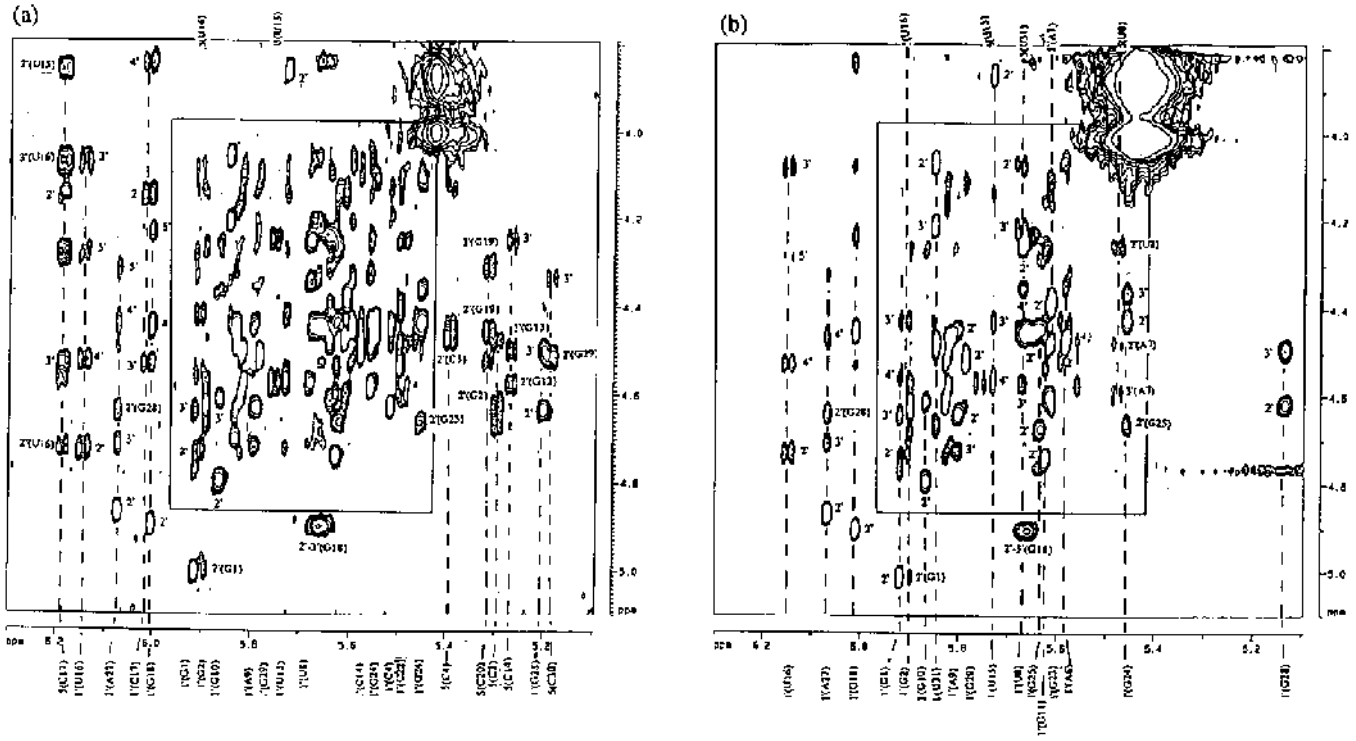

Figure 5. Contour plots of the H1′/H5 to H2′, H3′, H4′, H5’ and H5″ region of the 300 ms 2D NOESY spectrum: (**a**) The natural 31mer RNA **1** at 26°C. The region between 5.9 and 5.5 p.p.m. is crowded and an unambiguous assignment or accurate determination of NOE volumes is difficult. (**b**) The partially deuterated 31mer RNA **2**. The region between 5.9 and 5.5 p.p.m. is relatively much less crowded compared with the natural counterpart due to the banishment of all crosspeaks involving the H2′, H3′, H4′ and H5′/H5″ of the cytidine residues. The residual crosspeaks are much sharper allowing an easier and more accurate extraction of the NOE volumes used for molecular modelling.

H2′(G18). The sequence specific connectivities were then further identified from G19 to U31. The H3′ and H4′ protons are normally assigned from DQF-COSY spectrum, but for a molecule of this size, the spectra were too crowded for any reliable analysis. Instead, the H3′ and H4′ were assigned by using the intensities of NOESY crosspeaks at 100 and 300 ms in the H1′ to sugar and base to sugar regions. In the 300 ms NOESY spectra, the H1′ to sugar region shows crosspeaks from the H1′ to H3′ and H4′ via spin diffusion from H2′. The distinction between H3′ and H4′ was based on the fact that nucleotides with a N-type sugar and an *anti* conformation of the glycosidic bond have their H3′ close (2.5–3 Å) to their own H8/H6 proton and to the H8/H6 of their neighbouring nucleotides, whereas the H4′ is relatively far away (~4.5 Å) from the aromatic protons. Few H5′/H5″ were tentatively assigned from the weak NOE crosspeaks observed at long mixing times (300 ms) with H1′ or by the appearance of the weak crosspeak in the base to sugar region in the 300 ms NOESY spectra (H5′ and H5″ were not distinguished in these cases). Since in typical A-RNA, H2′(i)-H1′(i+1) is 4.0 Å, whereas it is >5 Å in B-DNA, our observation of weak H1′(i+1)-H2′(i) NOEs for both the stem I and II nucleotides [(G1-C4), (G11-U16), (G28-C30)] confirms the A-type conformation for both stem I and II. NOEs were also found between H2′(C4)-H1′(A5) and between H2′(A27)-H1′(G28) indicating that the A-type conformation of stem I continues even in the beginning of the loop I.

Effect of Mg^{2+}

Exchangeable imino protons. Upon addition of Mg^{2+} ions (up to 10 equiv.), no new imino resonance was observed, and the imino resonances at 11.9, 12.0 and 12.4 p.p.m. were not stabilized by addition of Mg^{2+} ions, suggesting that Mg^{2+} ion does not induce a significant conformational change in the base pair scheme.

Rate of exchange of the imino protons. Upon addition of $MgCl_2$, only the rate of exchange of the imino proton of G18 increases (Table 2), and its E_a also changes significantly, suggesting that the first Mg^{2+} binding site is in the vicinity of G18. The increase of exchange rate observed upon Mg^{2+} binding might be due to its hydrated nature, which promotes exchange rate with the imino proton of G18 at the G18·U15 base pair. The fact that E_a increases despite the fact that k_{ex} increases also indicates that the entropy of activation has a large contribution to the observed rate. It is noteworthy that a G·U base pair also defines a divalent ion binding site in a tRNA acceptor stem (13) and also at the 5′-splice site of Group I introns (14).

Non-exchangeable aromatic and anomeric protons. Upon addition of Mg^{2+} ion, H8(A27) is shielded by 0.06 p.p.m. and H1′(G28) by 0.15 p.p.m., while the protons of the other nucleotides have chemical shifts changes between 0 and 0.04 p.p.m., suggesting that the second Mg^{2+} binds in the vicinity of the stem I–loop I junction which is in spatially proximity of U26-G25-G24 sequence [corresponding to residues U294, G293 and G292 of RNase P RNA (3,4–6)].

Structure determination

Distance constraints. NOESY volumes were obtained from four different NOESY spectra, 100 and 300 ms mixing times for both

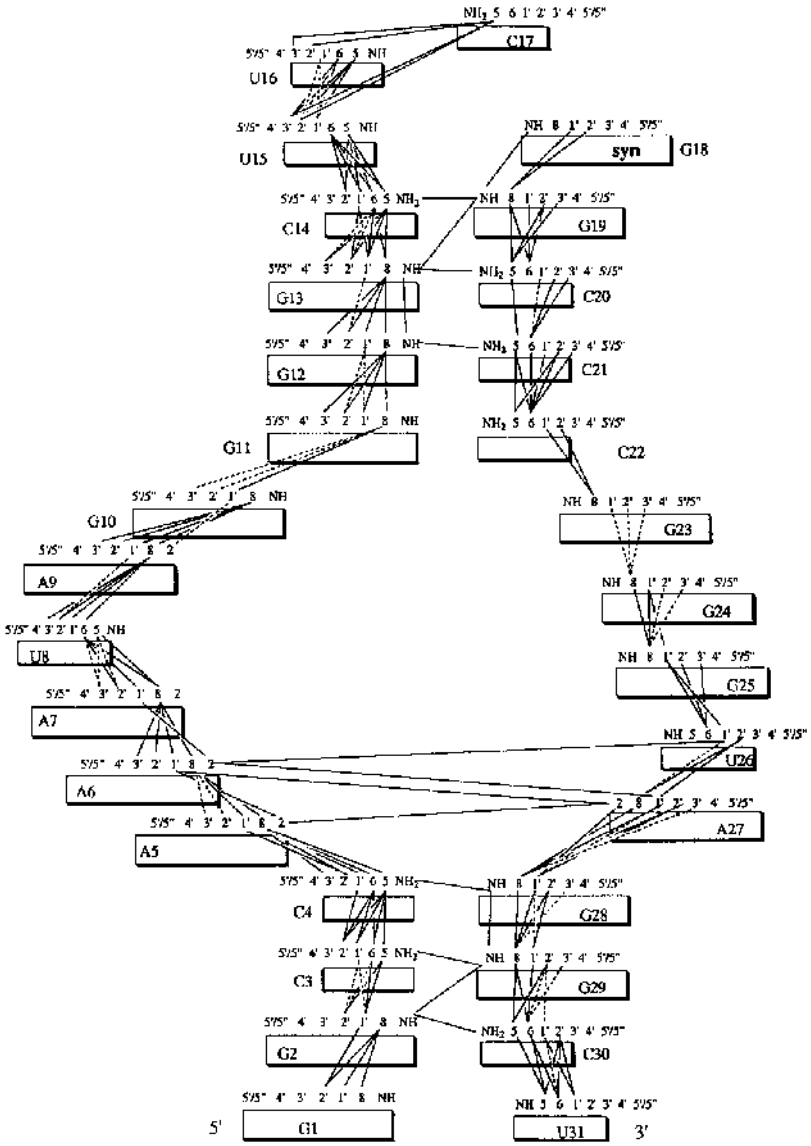

Figure 6. Schematic representation of sequential and long-range interesidual NOEs observed in the 31mer RNA. The solid lines represent the NOEs obtained from the natural 31mer RNA **1** and the dotted lines represent the NOEs obtained from the C-d_6 partially deuterated 31mer RNA **2**.

the natural 31mer and its partially-deuterated counterpart. Schematic representation of sequential and long-range interresidual NOEs observed in the 31mer RNA is given in Figure 6. The distance constraints in XPLOR were obtained by the classification of NOESY crosspeak intensities as strong (1.8–3.0 Å), medium (1.8–4.0 Å), weak (1.8–5.0 Å) or very weak (1.8–7.0 Å). The distance constraints for AMBER were derived from the experimental NOESY volumes using the iterative relaxation rate matrix method of MARDIGRAS and the RANDMARDI modification of MARDIGRAS by James *et al.* (30–33,38) starting from many different distance-geometry derived conformers. In addition, data from a 600 ms NOESY experiment in H_2O was used to determine the spatial proximities between bases. Due to the decrease in the problem of signal overlap, the partially-deuterated 31mer allowed us to extract an additional 28 distances (8 intra- and 20 interresidue). From the four NOESY experiments run in D_2O a total of 375 distance constraints were generated, 207 intraresidue distances and 168 interresidue distances. In addition, we used an extra 52 distance constraints arising from both (i) the NOESY spectra in D_2O (which were not used directly due to spin diffusion) as well as imino-imino, imino of residue (i) to H1′(i+1) and imino(i) to H1′ of the residue base pairing with residue (i–1) (i.e. typical A-RNA features) crosspeaks in H_2O were also included, where precise distances could not be estimated. In this context, the distances found in models of the UUCG hairpin RNA (7a) were also used for constraints. Other extra constraints included the distances between heteroatoms in the H-bonded base pairs.

In the final MD simulations and energy minimizations, distance constraints calculated by averaging the results from a series of RANDMARDI (30–33,38) distance calculations on 109 conformers using four different NOESY spectra (100 and 300 ms mixing times for both the natural and the partially-deuterated 31mer RNA, yielding 436 individual RANDMARDI results) were used in conjunction with the same 52 extra distance constraints described above. The RANDMARDI program calculated distance constraints from NOESY volumes by averaging the results from 30 individual

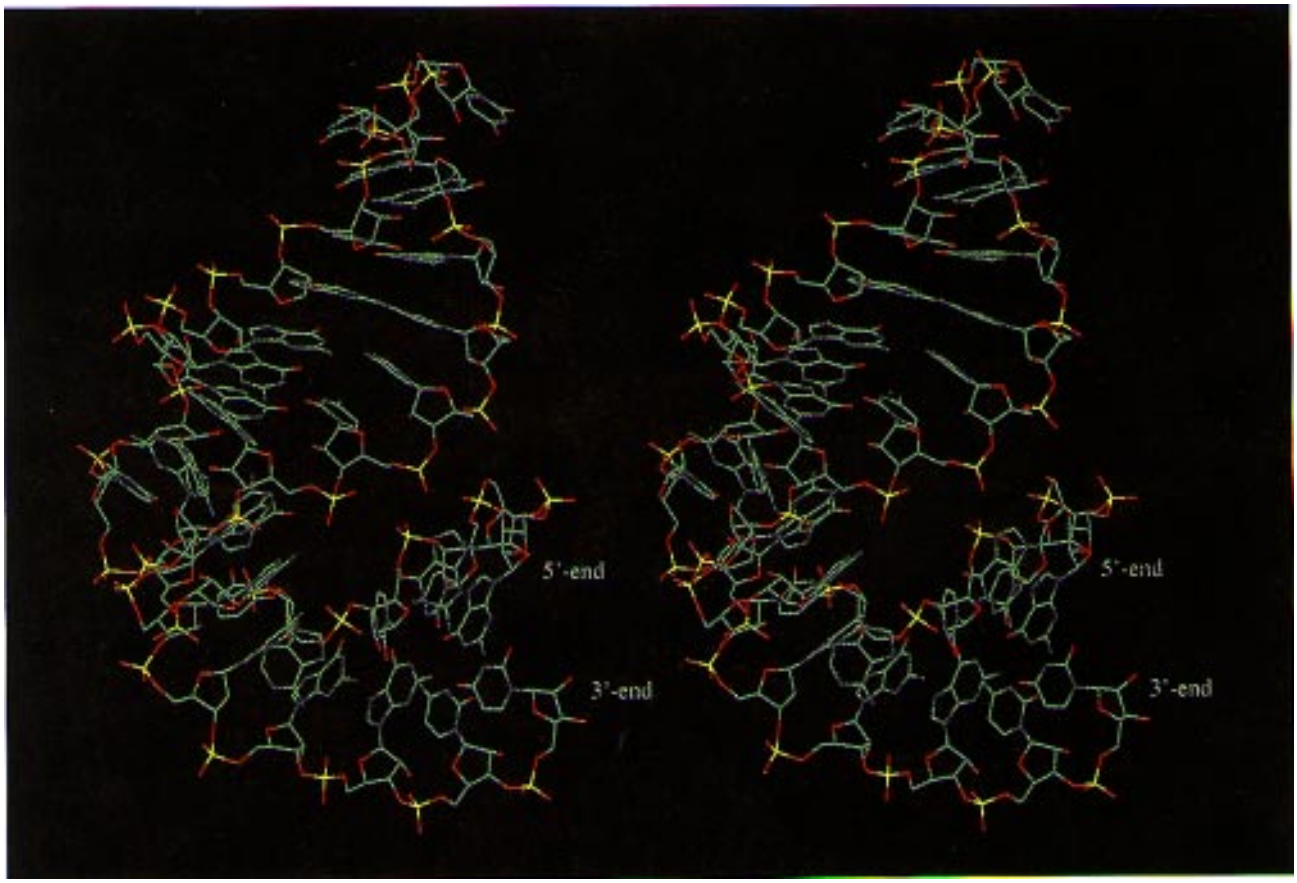

Figure 7. Stereoview of the best of the 109 structures after NMR constrained AMBER MD and energy minimization, as judged by R-factor analysis by MARDIGRAS. Only heavy atoms are shown (see Fig. 1 for the numbering scheme and Fig. 8 for the zooms of various parts of this structure).

MARDIGRAS calculations in which randomized noise with a noise level corresponding to the smallest volume in the entire spectrum was added (34) to the experimental NOESY volumes. The MARDIGRAS program (30–33) is based on the analysis of the complete relaxation rate matrix and, given a particular starting conformer, tries to find the inter-proton distances which best fit the experimental NOESY volumes in an iterative fashion. There were a total of 362 different distance constraints generated by the 436 RANDMARDI calculations, 13 distance constraints were rejected by the RANDMARDI program.

Dihedral constraints. A total of 228 dihedral constraints were used in the following manner. The stem II and loop II residues (^{11}G^{12}G^{13}**G**^{14}C^{15}U^{16}U^{17}C^{18}G^{19}G^{20}**C**^{21}C^{22}C) in our 31mer RNA have the corresponding residues (except for the base paired ^{13}G and ^{20}C) in the 12mer hairpin loop RNA, GG**A**CUUCG-G**U**CC, studied by Varani *et al.* (7). The chemical shifts of the residues within the loop ^{14}C^{15}U^{16}U^{17}C^{18}G^{19}G in 31mer RNA and in Varani's tetraloop (7) are within ±0.15 p.p.m., suggesting their structural similarity. Hence, 79 of the backbone torsions of stem II and loop II residues were constrained to the values (±45°) found in the 12mer RNA: α, β and γ for residues G12–C22, δ and χ for residues G11–C22 and ε and ζ for residues G11–C21. The sugar moieties of residues G11–C22 were similarly constrained, as found in the 12mer RNA loop, by applying constraints on ν_0, ν_1, ν_2 and ν_4. The center value and the allowed variation of each torsion was calculated in such a way that the phaseangle of pseudorotation of a particular sugar would be within 20° and the puckering amplitude would be within 5° of the value for the corresponding sugar in the 12mer RNA. Typically, torsions were allowed to differ between ±9 and ±15° from the center value. 48 dihedral constraints of this type were used. 26 dihedral constraints, typical of A-RNA (as evident from NOESY spectra), were applied to backbone torsions of the double-stranded part of the 31mer (stem I): the α, β, γ, ε, ζ and χ torsions were constrained to –60° ± 60°, 180° ± 60°, 60° ± 60°, 180° ± 60°, –60° ± 60° and 180° ± 90°, respectively. All residues for which the $J_{1',2'}$ indicated a clear preference for either the S-type or the N-type sugar were constrained (P = 10° ± 30° for N-type sugars and 165° ± 30° for S-type sugars with a puckering amplitude of 38° ± 5°). In total, we used 655 experimental distance and torsional constraints for distance-geometry and simulated annealing of 31mer RNA, which makes an average of 21.1 constraints/nucleotide residue. However, it should be noted that for constrained MD and energy minimization by AMBER, only 642 constraints were used (average of 20.7 constraints/nucleotide residue).

Distance-geometry and simulated annealing using the XPLOR program. Using the constraints described above, 200 structures were generated using distance-geometry, regularization and simulated annealing calculations (see Materials and Methods). After the seven cycles of simulated annealing, the 200 structures

Table 4. Number of distance and dihedral constraints used in XPLOR. Distance constraints generated by RANDMARDI approach, shown in parenthesis, were used in AMBER

Residue	G1	G2	C3	C4	A5	A6	A7	U8	A9	G10	G11
Intrares. dist.[a]	7 (7)	6 (6)	6 (5)	7 (6)	6 (6)	5 (5)	7 (7)	7 (6)	7 (7)	7 (7)	7 (7)
Interres. dist.[b,c]	4 (4)	10 (10)	13 (13)	12 (12)	10 (10)	10 (10)	12 (12)	12 (12)	11 (11)	7 (7)	11 (11)
Interres. dist.[b,d]	-	-	-	-	1 (1)	5 (5)	-	-	1 (1)	-	-
Extra[e,c]	1	3	2	-	-	-	-	-	-	-	1
Extra[e,d]	-	5	3	4	1	-	-	-	-	-	3
Dihedral	-	8	11	9	5	-	5	5	5	5	8
Total constr.	12 (12)	32 (32)	35 (34)	32 (31)	23 (23)	20 (20)	24 (24)	24 (23)	24 (24)	19 (19)	30 (30)
Residue	**G12**	**G13**	**C14**	**U15**	**U16**	**C17**	**G18**	**G19**	**C20**	**C21**	**C22**
Intrares. dist.[a]	7 (7)	7 (7)	8 (7)	7 (6)	8 (7)	7 (6)	8 (8)	4 (4)	8 (7)	7 (6)	8 (7)
Interres. dist.[b,c]	11 (11)	15 (15)	19 (19)	16 (16)	12 (12)	5 (5)	3 (3)	9 (9)	9 (9)	8 (8)	7 (7)
Interres. dist.[b,d]	-	-	-	2 (2)	-	2 (2)	-	-	-	-	-
Extra[e,c]	3	3	1	1	2	1	2	4	2	1	1
Extra[e,d]	4	5	3	3	1	-	3	5	3	4	4
Dihedral	11	11	11	11	11	11	11	11	11	11	9
Total constr.	36 (36)	41 (41)	42 (41)	40 (39)	34 (33)	26 (25)	27 (27)	33 (33)	33 (32)	31 (30)	29 (28)
Residue	**G23**	**G24**	**G25**	**U26**	**A27**	**G28**	**G29**	**C30**	**U31**		
Intrares. dist.[a]	6 (6)	6 (6)	4 (4)	7 (6)	8 (8)	7 (7)	6 (6)	5 (4)	7 (6)		
Interres. dist.[b,c]	5 (5)	8 (8)	9 (9)	7 (7)	11 (11)	13 (13)	13 (13)	17 (17)	9 (9)		
Interres. dist.[b,d]	-	-	1 (1)	1 (1)	5 (5)	-	-	-	-		
Extra[e,c]	-	1	1	-	1	3	3	2	1		
Extra[e,d]	-	-	-	-	-	4	5	3	1		
Dihedral	5	5	5	5	-	8	11	9	-		
Total constr.	16 (16)	20 (20)	20 (20)	20 (19)	25 (25)	35 (35)	38 (38)	36 (35)	18 (17)		

[a]The number of intraresidue distance constraints calculated directly from the NOESY volumes by the 'strong-medium-weak' approach or by RANDMARDI (in parenthesis).
[b]The number of interresidual constraints calculated directly from the NOESY volumes by the 'strong-medium-weak' approach or by RANDMARDI (in parenthesis). The number of interresidual distance constraints is shown twice because of involvement of two residues in a NOE crosspeak.
[c]Intrastrand distance constraints.
[d]Interstrand distance constraints.
[e]Number of distance constraints derived indirectly from NOESY data such as the hydrogen bond constraints in a Watson–Crick base pair or distances approximated from volumes with a high degree of spin diffusion as well as all data from NOESY experiments in H_2O. The number of interresidue distance constraints is shown twice because of involvement of two residues in a NOE crosspeak.

were evaluated in terms of number of violations of the distance constraints. 91 Structures had more than two 0.2 Å distance violations and they were rejected. The remaining 109 structures were analyzed by the MARDIGRAS program and the agreement between their theoretically calculated NOE volumes with the experimental ones were measured. The R-, R^2-, R_x- and R_x^2-factors were found to be in the range 0.543–0.871, 0.658–1.106, 0.099–0.159 and 0.126–0.194 respectively. The conformational similarity between the 109 structures was assessed by calculating an average structure and comparing all 109 structures to the average (rmsd): 5.97 Å (σ = 1.04 Å) for all heavy atoms of the structure, 1.87 Å (σ = 0.84 Å) for stem I, 1.93 Å (σ = 0.79 Å) for stem II, 4.4 Å (σ = 0.62 Å) for loop I and 2.26 Å (σ = 0.5 Å) for loop II.

Molecular dynamics simulations and energy minimizations using the AMBER program. The 109 selected structures from the distance-geometry simulated annealing structure refinement procedure were used as starting structures for 109 separate runs using the NMR constrained AMBER procedure (see Materials and Methods).

Precision of the structure. Nine conformers out of the 109 structures were selected because they showed the lowest combination of R-factors after the constrained MD and minimization by AMBER in the previous step, the structure showing the best agreement with the NMR data, as evident by R-factor analysis, is shown in Figures 7 and 8. The ranges of the R-, R^2-, R_x- and R_x^2-factors for these nine best conformers were 0.330–0.429, 0.369–0.522, 0.066–0.083 and 0.082–0.110 respectively. It should be noted here that such low R-factors in the MARDIGRAS were earlier obtained only for a much smaller oligo-DNA/RNA (35–37), comprising approximately two-thirds of the number of nucleotides of the present 31mer RNA. The nine individual structures were then compared with the average of these nine structures with respect to all heavy atom, showing an

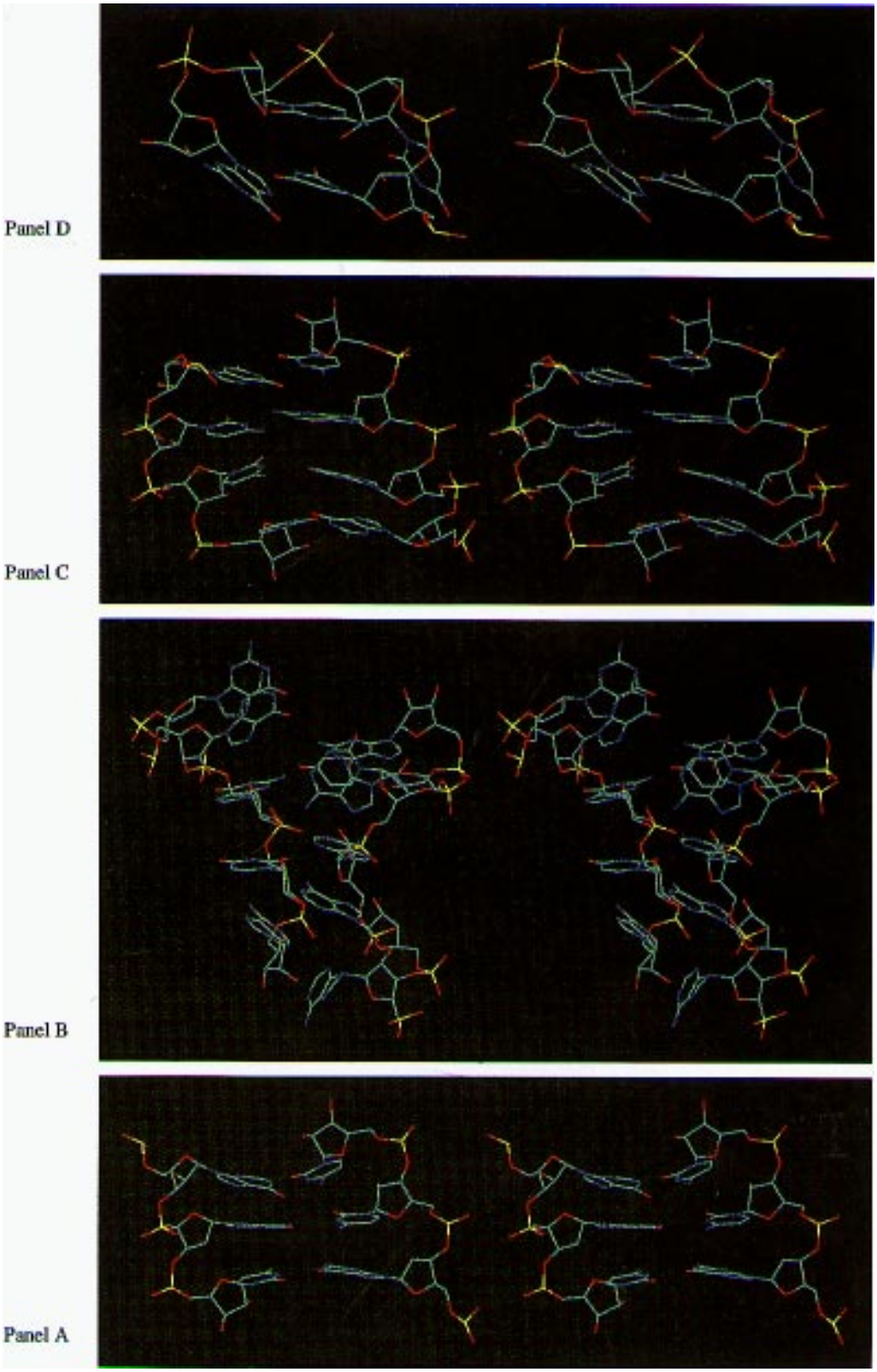

Figure 8. Stereoviews of the zooms of the various parts of the best structure shown in Figure 7 (see Fig. 1 for the numbering scheme as well as for the nomenclature). (**A**) Watson–Crick base paired A-RNA type helix for the ^{30}C to ^{28}G residues (left strand) and ^{2}G to ^{4}C residues (right strand) in stem I. (**B**) A partial continuity of the base-stacked A-RNA type helix for the ^{27}A to ^{23}G residues (left strand) and ^{5}A to ^{10}G residues (right strand) in the flexible loop I. (**C**) Watson–Crick base paired A-RNA type helix for the ^{22}C to ^{19}G residues (left strand) and ^{11}G to ^{14}C residues (right strand) in stem II. (**D**) The UUCG tetra loop for ^{15}U (lower right) to ^{18}G (lower left) residues in loop II.

average rmsd of 6.02 Å. When only a specific region of the 31mer was considered, the average rmsd was 0.86, 0.77, 4.08 and 1.02 Å for stem I and II, loop I and II, respectively. A simple comparison of these rmsds reveals that most of the structural differences in all heavy atom comparison come from the large loop I region, but in contrast the conformation of the two double-stranded stems I and II regions and loop II region show significant conformational homogeneity. One of the main reasons for the observed rmsd differences between stem I, stem II, loop II, in one hand, and loop I on the other, is owing to the fact that relatively larger number of experimental NMR constraints per nucleotide residue is available in the former group (22.8 distance constraint/nucleotide) compared with the latter (17.5 distance constraint/nucleotide) (Table 4). A comparison of the types of constraints for various nucleotides in various parts of the 31mer RNA in Table 4 also shows an absence of imino to H1′(i+1) and imino to H1′(i–1) of the opposite strand, Watson–Crick hydrogen-bonds, as well as of imino to imino cross- and intra-strand distance constraints for loop I nucleotides, which makes it very difficult to define its conformation, thereby suggesting either a highly flexible nature or simply that the residues are more than~5 Å away from each other.

Conclusion

We have shown that the deuteration of nucleotides helps to simplify the NMR spectra of RNA. By deuteration at the C5, C2′, C3′, C4′ and C5′ positions of the cytosine nucleotides which constitute the stems, we were able to assign without ambiguity the anomeric and aromatic protons. This was not possible from the NOESY spectra of the natural 31mer RNA where the strong H5-H6 crosspeaks were overlapping with some aromatic-anomeric NOEs. This possibility of assignment without ambiguity was particularly important for the loop I region where no assumption on the conformation can be made. The use of the partially-deuterated 31mer RNA has enormously facilitated the resonance assignment of the non-anomeric sugar protons and the extraction of the NOE volumes. Out of 235 non-exchangeable proton resonances in the natural 31mer RNA **1**, 175 resonances have been unambiguously assigned. The 60 resonances that we could not assign belong to the category of H4′ and H5′/5″. Of the 175 assigned resonances, 41 resonances (15 belonging to the C residues) in the natural 31mer could only be assigned on the basis of the comparative analysis between the natural 31mer RNA and its partially-deuterated analogue, thereby showing the usefulness of our non-uniform deuterium labelling techniques (39–44). This has led us to obtain additional 8% constraints from the deuterated 31mer RNA compared to the natural counterpart. It is clear that a successive incorporation of the deuterated A, G and U residues in 31mer RNA would give us additional structural information, giving larger number of NMR constraints for further structure refinement. Alternatively, we are also trying to improve our RNA solid-phase synthesis protocol to create an ‘NMR-window’ (44) on the loop I, enabling only a selected protonated part to be visible by NMR spectroscopy.

The NMR data supporting Watson–Crick base pairing, base–base stacking interactions and the North conformation of the sugars indicates that both stems I and II of the 31mer RNA have an A-type geometry. The base–base stacking found in stem I is conserved into the loop I. The base pairs C4·G28 and G11·C22, which are at two ends of the loop II, have very different behaviour. The access of water to the G28·C4 base paired imino proton is reduced probably through stacking interactions. Thus, NOESY data have shown that G28 strongly stacks on A27 and that C4 strongly stacks on A5. The observation of interstrand NOEs in the internal loop in close proximity of the G28·C4 base pair together with the observation of NOEs between the H2 of A5, A6 and A27 to the H1′ of their 3′-neighbouring residue also indicate that the A-type conformation continues until the beginning of the stem I to loop I junction. The loop I conformation is more flexible in the loop I–stem II region. The stacking interactions between A9, G10, G11 and between C22, G23 and G24 are weaker and NOE data suggest that G10 is bulged out. These data together with the fast rate of exchange of G11 indicate a less ordered region of the loop I with structural perturbation at and around G11. Magnesium ion binding occurs most probably around the G18·U15 base pair in the loop II and around the A27 and G28 nucleotides at the stem I–loop I junction. This second site of binding is of particular interest since magnesium ion has been shown to bind in the vicinity of loop I region in the *E.coli* RNase P RNA. Work is now in progress in our lab to study the solution conformation of large functional RNA using the deuterated analogues in which deuterated residues are incorporated in the NMR-invisible part (39–44) and appropriate ^{13}C and ^{15}N labelled sugar residues are incorporated in the NMR-visible part (39–44) to extract the sugar–phosphate backbone constraint for more accurate structure determination.

ACKNOWLEDGEMENTS

We thank Swedish Natural Science Research Council (NFR), Swedish Council of Engineering and Research (TFR) and Swedish Board for Technical Development (NUTEK) for generous financial support (to JC and LK). We would also like to thank Prof. T. James for kindly supplying us with the latest versions of MARDIGRAS and RANDMARDI. Thanks are also due to the Wallenbergstiftelsen, Forskningsrådsnämnden, and University of Uppsala for funds for the purchase of a 500 MHz Bruker AMX NMR spectrometer and for the generous funding of our new NMR facility.

REFERENCES

1 (a) Deutscher,M.P. (1990) *Prog. Nucleic Acids Res. Mol. Biol.*, **39**, 209–240; (b) Altman,S., Kirsebom,L.A. and Talbot,S. (1995) In Söll,D. and RajBhandary,U.R. (eds) *tRNA: Structure, Biosynthesis and Function*, American Society for Microbiology, Washington, DC 20005, pp. 67–78; (c) Kirsebom,L.A.(1995) *Mol. Microbiol.*, **17**, 411–420.
2 Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) *Cell*, **35**, 849–857.
3 Kazakov,S. and Altman,S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9193–9197.
4 (a) Kirsebom,L.A. and Svärd,S.G., (1994) *EMBO J.*, **13**, 4870–4876; (b) Kufel,J. and Kirsebom,L.A. (1994) *J. Mol. Biol.*, **244**, 511–521; (c) Ciesiolka,J., Hardt,W.-D., Schlegl,J., Erdmann,V.A. and Hartmann,R.H. (1994) *Eur. J. Biochem.*, **219**, 49–56; (d) Zito,K., Huflenhofer,A. and Pace,N.R. (1993) *Nucleic Acids Res.*, **21**, 5916–5920.
5 Brown,J.W. and Pace,N.R. (1992) *Nucleic Acids Res.*, **20**, 1451–1456.
6 Haas,E.S., Morse,D.P., Brown,J.W., Schmidt,F.J. and Pace,N.R. (1991) *Science*, **254**, 853–856.
7 (a) Varani,G., Cheong,C. and Tinoco,I. Jr. (1991) *Biochemistry*, **30**, 3280–3289. (b) For a modified structure for UUCG loop, see Allain,F.H.-T and Varani,G. (1995) *J. Mol. Biol.*, **250**, 333–353.
8 Heus,H.A. and Pardi,A. (1991) *J. Am. Chem. Soc.*, **113**, 4360–4361.
9 Maltseva,T.V., Yamakage,S.-I., Agback,P. and Chattopadhyaya,J. (1993) *Nucleic Acids Res.*, **21**, 4288–4295.

10 White,S.A., Nilges,M., Huang,A., Brunger,A.T. and Moore,P.B. (1992) *Biochemistry*, **31**, 1610–1621.
11 Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids.* Wiley, New York, NY.
12 Varani,V. and Tinoco,I. (1991) *Quart. Rev. Biophys.*, **24**, 479–532.
13 (a) Limmer,S., Hofmann H.-P., Ott,G. and Sprinzl,M. (1993) *Proc. Nat. Acad. Sci. USA*, **90**, 6199–6202. (b) Ott,G., Arnold,L. and Limmer,S. (1993) *Nucleic Acids Res.*, **21**, 5859–5864.
14 Allain,F.H.-T. and Varani,G. (1995) *Nucleic Acids Res.*, **23**, 341–350.
15 Brünger,A.T. XPLOR v.3.1 (Yale University, New Haven, CT, 1992).
16 Pearlman,D.A., Case,D.A., Caldwell,J.C., Seibel,G.L., Singh,U.C., Weiner,P.and Kollman,P.A. AMBER v.4.0 (University of California, San Fransisco, CA, 1991).
17 Ryckaert,J.P. (1985) *Mol. Phys.*, **55**, 549–556.
18 Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.
19 Zhou,X.-X., Welch,C. J. and Chattopadhyaya,J. (1986) *Acta Chem. Scand.*, **B40**, 806–816.
20 Rabi,J.A. and Fox,J. J. (1973) *J. Am. Chem. Soc.*, **95**, 1628–1632.
21 Ludwig,J. (1981) *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131–133.
22 Milligan,J.F., Groebe,D.R., Whiterell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.
23 Wyatt,J.R., Chastain,M., Puglisi,J.D. (1991) *BioTechniques*, **11**, 764–769.
24 Hore,P.J. (1983) *J. Magn. Reson.*, **55**, 283–300.
25 The program AURELIA was supplied by BRUKER spectrospin.
26 Bodenhausen,G., Kogler,H. and Ernst,R.R. (1984) *J. Magn. Reson.*, **58**, 370–388.
27 Neuhaus,D., Wagner,G., Vasak,M., Kagi,J.H.R. and Wuthrich,K. (1985) *Eur. J. Biochem.*, **151**, 257–273.
28 Griesinger,C., Otting,G., Wuthrich,K. and Ernst,E.E. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.
29 Marion,D. and Wuthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
30 Borgias,B.A. and James,T.L. (1988) *J. Magn. Reson.*, **79**, 493–512.
31 Borgias,B.A., Gochin,M., Kerwood,D.J. and James,T.L. (1990) *Progr. NMR. Spect.*, **22**, 83–100.
32 Schmitz,U. Sethson,I., Egan,W.M. and James,T.L. (1988) *J. Mol. Biol.*, **227**, 510–531.
33 Borgias,B.A., Thomas,P.D., Liu,H., Kumar,A. and James,T.L. MARDIGRAS/RANDMARDI v.5.1 (University of California, San Fransisco, CA, 1995).
34 Spielmann,H.P., Dwyer,T.J., Hearst,J.E. and Wemmer,D.E. (1995) *Biochemistry*, **34**, 12937–12953.
35 Ulyanov,N.B., Schmitz,U. and James,T.L. (1993) *J. Biomol. NMR*, **3**, 547–568.
36 Mujeeb,A., Kerwin,S.M., Kenyon,G.L. and James,T.L. (1993) *Biochemistry*, **32**, 13419–13431.
37 Gonzalez,C., Stec,W., Reynolds,M.A. and James,T.L. (1995) *Biochemistry*, **34**, 4969–4982.
38 Liu,H., Spielmann,H.P., Ulyanov,N.B., Wemmer,D.E. and James,T.L. (1995) *J. Biomol. NMR*, **6**, 390–402.
39 Földesi,A., Nilson,F.P.R., Glemarec,C., Gioeli,C. and Chattopadhyaya,J. (1992) *Tetrahedron*, **48**, 9033–9072.
40 Földesi,A., Nilson,F.P.R., Glemarec,C., Gioeli,C. and Chattopadhyaya,J. (1993) *J. Biochem. Biophys. Methods*, **26**, 1–26.
41 Yamakage,S.-I., Maltseva,T.V., Nilson,F.P. Földesi,A. and Chattopadhyaya,J. (1993) *Nucleic Acids Res.*, **21**, 5005–5011.
42 Agback,P., Maltseva,T.V., Yamakage,S.-I., Nilson,F.P.R., Földesi,A. and Chattopadhyaya,J. (1994) *Nucleic Acids Res.*, **22**, 1404–1412.
43 Földesi,A., Yamakage,S.-I., Maltseva,T.V., Nilson,F. P., Agback,P. and Chattopadhyaya,J. (1995) *Tetrahedron*, **51**, 10065–10092.
44 Földesi,A., Yamakage,S.-I., Nilson,F.P.R., Maltseva,T. and Chattopadhyaya,J. (1996) *Nucleic Acids Res.*, **24**, 1187–1194.
45 Altona,C. (1982) *Recl. Trav. Chim. Pays-Bas*, **101**, 413–433.