Formation of sheared G:A base pairs in an RNA duplex modelled after ribozymes, as revealed by NMR

Masato Katahira, Mayumi Kanagawa, Hajime Sato, Seiichi Uesugi*, Satoshi Fujii¹

Toshiyuki Kohno² and Tadakazu Maeda^{2,+}

Department of Bioengineering, Faculty of Engineering, Yokohama National University, Tokiwadai 156, Hodogaya-ku, Yokohama 240, ¹Faculty of Pharmaceutical Science, Osaka University, Suita, Osaka 565 and ²Mitsubishi Kasei Institute of Life Sciences, Minamiooya 11, Machida-shi, Tokyo 194, Japan

Received April 27, 1994; Revised and Accepted June 9, 1994

ABSTRACT

The thermal stability and structure of an RNA duplex, r(GGACGAGUCC)₂, the base sequence of which was modelled after both a hammerhead ribozyme and a lead ribozyme, were studied by CD and NMR. We previously demonstrated that the corresponding DNA duplex, d(GGACGAGTCC)₂, formed unique 'sheared' G:A base pairs, where an amino proton, instead of an imino proton, of G is involved in the hydrogen bonding, and G and A bases are arranged 'side by side' instead of 'head to head' (*Nucleic Acids Res.* (1993) 21, 5418–5424). CD melting profiles showed that the RNA duplex is thermally more stable than the corresponding DNA duplex. NMR studies revealed that sheared G:A base pairs are formed in the RNA duplex, too, although the overall structure of the RNA is the A form, which differs from the B form taken on by the corresponding DNA. A model building study confirmed that sheared G:A base pairs can be accommodated in the double helical structure of the A form. A difference between the RNA and DNA duplexes in the stacking interaction involving G:A mismatch bases is also suggested. The demonstration that sheared G:A base pairs can be formed not only in DNA but also in RNA suggests that this base pairing plays an important role regarding the RNA structure.

INTRODUCTION

Since the discovery of ribozymes, studies to elucidate how ribozymes exert their enzymatic activities on the basis of their structures have been carried out [1–11]. It has been suggested that a G:A mismatch base pair is formed in some ribozymes and that the structure derived from this base pairing could play a crucial role regarding their enzymatic activities [12, 13]. Several different types of G:A base pairs have been observed in DNA; 'head to head' G(*anti*)–A(*anti*) [14–16], G(*anti*):A(*syn*) [17], G(*syn*):A(*anti*) [16, 18, 19], and 'sheared' G(*anti*):A(*anti*) [12,

20–28]. The type of G:A base pairing in DNA depends on the neighbouring base sequences and pH [16, 18, 22, 28–30].

As the first step to elucidate the structure–activity relationship for ribozymes, we studied the structures of DNA duplexes, d(GGACGAGTCC)₂ [24], d(GGACGACATC):d(GATGGAGTCC) [24] and d(GGACGCATC):d(GATGAGTCC) [28], the base sequences of which were modelled after both a hammerhead ribozyme [31] and a lead ribozyme [32]. The first six or five bases of these duplexes are identical to those of helix 2 of the hammerhead ribozyme, where the formation of G:A base pairs is suggested [13], and these bases are also very similar to those of the half part of the lead ribozyme, where the formation of G:A base pairs can be expected from the base sequence [24]. It was concluded that unique 'sheared' G:A base pairs, where an amino proton, instead of an imino proton, of G is involved in the hydrogen bonding, and G and A bases are arranged 'side by side', are formed in d(GGACGAGTCC)₂ and d(GGACGACATC):d(GATGGAGTCC) [24]. And it is suggested that a YGA:GAR sequence is an essential element for the formation of the sheared G:A base pairs [24]. In the case of d(GGACGCATC):d(GATGAGTCC), where the central two consecutive G:A mismatches of d(GGACGACATC):d(GATGGAGTCC) are replaced by a single G:A mismatch, on the other hand, another kind of G:A base pairing, presumably a 'head to head' G:A base pair, occurred [28]. Thus, sequence dependent polymorphism of the structure of the G:A base pair was shown.

The next step was to elucidate the structure of RNA with G:A mismatches modelled after ribozymes. Here we report studies on the thermal stability and structure of an RNA duplex, r(GGACGAGUCC)₂. This RNA duplex is a complete RNA version of the DNA duplex, d(GGACGAGTCC)₂, mentioned above. Therefore its first six bases are identical to those of part of the hammerhead ribozyme, and they are also very similar to bases of part of the lead ribozyme. The thermal stabilities of r(GGACGAGUCC)₂ and d(GGACGAGTCC)₂ were compared on the basis of CD melting profiles. The RNA duplex turned out to be thermally more stable than the DNA duplex. NMR studies

*To whom correspondence should be addressed

⁺Present address: Department of Physics, School of Science, Kitasato University, 1-15-1 Kitasato, Sagami-hara, Kanagawa 228, Japan

demonstrate that sheared G:A base pairs are formed in the RNA duplex, in spite of the difference in the overall structure, i.e., the RNA duplex assumes the A form while the DNA duplex assumes the B form. The demonstration that sheared G:A base pairs can be formed not only in DNA but also in RNA suggests that this base pairing plays an important role regarding the RNA structure.

The numbering of residues is as follows.

1	2	3	4	5	6	7	8	9	10
G	G	A	C	G	A	G	U	C	C
C	C	U	G	A	G	C	A	G	G
10	9	8	7	6	5	4	3	2	1

MATERIALS AND METHODS

Oligonucleotide r(GGACGAGUCC) was synthesized by the phosphoramidite coupling method with the 2'-hydroxyl protected as the tert-butyl dimethyl silyl ether with a model 392 DNA synthesizer (Applied Biosystems Co.). The oligonucleotide was purified and annealed as described previously [33] with additional treatments with ammonia in ethanol and tetrabutylammonium fluoride in tetrahydrofuran to remove the base protecting groups and the silyl group, respectively. D(GGACGAGTCC)₂ was previously prepared [24]. Thermal CD melting curves for both r(GGACGAGUCC)₂ and d(GGACGAGTCC)₂ were recorded with a Jasco J-720 spectropolarimeter interfaced with an NEC PC-9801 FX personal computer with 0.1 cm and 1.0 cm cells. Each duplex was dissolved in 20 mM phosphate buffer (pH 7.0) containing 0.15 M NaCl, with the strand concentration of either 5 or 50 μ M. The temperature of the solution was raised from 5°C to 80°C at the rate of 1°C/min. Changes in CD intensity were monitored at 262 nm for the RNA duplex and at 270 nm for the DNA duplex, respectively. The melting temperature was determined by use of the derivative of a melting curve.

For NMR measurement of non-exchangeable protons, a lyophilized sample was dissolved in 20 mM phosphate buffer (pH 7.0) containing 0.15 M NaCl. The solution was lyophilized and then dissolved in 0.2 ml of D₂O (99.96%). The strand concentration was 1.8 mM. For measurement of exchangeable proton spectra, an H₂O:D₂O mixture (19:1) was substituted for D₂O. DSS was used as an internal chemical shift reference.

NMR spectra were recorded at 5–50°C with Bruker AM-400 and AMX-500 NMR spectrometers. Phase-sensitive NOESY [34], HOHAHA [35], and DQF-COSY [36] spectra were recorded by the time-proportional phase-increment method [37]. The mixing times for NOESY were 100 and 300 ms, and that for HOHAHA was 40 ms. The repetition delay was 2.0 s. Two-dimensional spectra were recorded with 512 t_1 increments, 128–160 free induction decays of 2 K data points per increment being collected. The t_1 and t_2 data were apodized with a $\pi/3$ -shifted sine-bell function. The t_1 data were zero-filled to 1 K points. One-dimensional spectra in H₂O were accumulated with a 1–1 pulse sequence [38], and NOE difference spectra were obtained as described previously [33] with irradiation times of 100 and 300 ms.

RESULTS

The melting temperatures of RNA and DNA duplexes obtained from CD melting curves

As an example, the CD melting curve and its derivative of r(GGACGAGUCC)₂ with the strand concentration of 5 μ M are

Figure 1. The thermal CD melting curve of r(GGACGAGUCC)₂ with the strand concentration of 5 μ M (solid line) and its derivative on an arbitrary scale (dotted line).

shown in Figure 1. The melting temperatures of r(GGACGAGUCC)₂ and d(GGACGAGUCC)₂ obtained from CD melting curves are summarized in Table 1. For both RNA and DNA, the melting temperature is dependent on the oligomer concentration. Higher melting temperatures are obtained with higher concentrations. The results rule out unambiguously the possible formation of a hairpin structure for both RNA and DNA, because the melting temperature should not be dependent on the concentration of the strands if a hairpin structure is formed. It is thus suggested that a double stranded structure is formed.

The melting temperature of the RNA duplex was higher than that of the DNA duplex.

Assignments of exchangeable protons

Figure 2 shows the one-dimensional ¹H NMR spectrum and NOE difference spectra on irradiation of the imino protons of the RNA duplex in H₂O. The resonances of imino protons were assigned sequentially on the basis of observed NOEs, as described previously [23, 30]. Irradiation of U8H3 gives NOE's to G7H1 and G2H1, in addition to a strong NOE to A3H2 (7.70 ppm). Irradiation of G5H1 gives an NOE to G7H1. It is notable that the imino proton of the G5 residue resonates at a high field, 9.89 ppm, i.e., well separated from the usual region, ~13 ppm, for a resonance of a hydrogen bonded imino proton of a G residue. A similar high field shift was observed for the imino proton of the G5 residue of the DNA duplex [23].

Assignments of non-exchangeable protons

The resonances of non-exchangeable protons of the RNA duplex were assigned sequentially by analysis of two-dimensional NOESY, HOHAHA and DQF-COSY spectra in D₂O in the same way as reported for other duplexes [24, 33, 39–41], using previously established methods [39, 42–50]. Figure 3(a) shows the expansion of the NOESY spectrum, indicating the sequential assignments of H1' and H6/H8 through the H1'(i–1)–H6/H8(i)–H1'(i) connectivities. The intraresidue H5–H6 cross peaks of pyrimidines are also labelled for reference. It is

Figure 2. ^1H NMR spectrum with the assignments of the imino proton signals indicated by the residue numbers (a), and NOE difference spectra of $r(\text{GGACGAGUCC})_2$ at 15°C in the region of imino and aromatic proton resonances on 300 ms irradiation of the U8 imino (b) and G5 imino (c) resonances, respectively. The spectrum (c) is scaled up twice vertically for clarity.

Figure 3. Expansion of the NOESY spectra of $r(\text{GGACGAGUCC})_2$ in a D_2O solution obtained with mixing times of 300 ms (a) and 100 ms (b) at 33°C , and with a mixing time of 300 ms at 37°C (c). The lines show the $\text{H1}'(i-1)-\text{H6}/\text{H8}(i)-\text{H1}'(i)$ connectivities. The intraresidue $\text{H6}/\text{H8}-\text{H1}'$ cross peaks are labelled with their residue numbers. The intraresidue $\text{H5}-\text{H6}$ cross peaks of pyrimidines are labelled with underlined residue numbers. The $\text{A3H2}-\text{C4H1}'$ and $\text{A6H2}-\text{G7H1}'$ cross peaks are indicated by a and b, respectively. Arrows indicate the interstrand $\text{A6H2}-\text{A6H1}'$ NOEs.

remarkable that $\text{G7H1}'$ resonates at a very high field, 4.14 ppm. An extreme up field shift of the $\text{G7H1}'$ resonance was not observed for the DNA duplex [24]. Although the position of the $\text{G7H1}'$ resonance was difficult to identify at the early stage of the assignment procedure, it was definitely identified on

Figure 4. Expansion of the NOESY spectrum of $r(\text{GGACGAGUCC})_2$ obtained with a mixing time of 100 ms in a D_2O solution at 33°C . The intraresidue cross peaks between $\text{H1}'$ and either $\text{H2}'$ (circles) or $\text{H3}'$ (triangles) or $\text{H4}'$ (rectangles) are connected by lines, with numbering of residues being labelled. For G7, exceptionally, cross peaks between U8H5 and either $\text{H1}'$ (cross) or $\text{H2}'$ (circle) or $\text{H3}'$ (triangle) or $\text{H4}'$ (rectangle) are connected with the residue number 7 underlined (See text).

completion of the assignment of $\text{H1}'$, $\text{H2}'$, $\text{H3}'$ and $\text{H4}'$ resonances at the late stage.

It should be noted that a very strong interstrand $\text{A6H2}-\text{A6H1}'$ cross peak was observed (arrows in Figure 3). The $\text{A6H2}-\text{A6H1}'$ and $\text{A6H8}-\text{A6H1}'$ cross peaks overlap each other in Figure 3(a), but these two peaks are separated in Figure 3(b) which shows the area symmetrical to Figure 3(a) with respect to the diagonal line due to the higher resolution with the aromatic resonances. The separation of the two cross peaks is further better in Figure 3(c) recorded at slightly different temperature, 37°C . In these figures, it was confirmed that the $\text{A6H2}-\text{A6H1}'$ cross peak is very strong. The $\text{A6H2}-\text{A6H1}'$ cross peak should be the interstrand cross peak, because the intraresidue $\text{AH2}-\text{AH1}'$ distance is too long for any conformation to give a solid cross peak [51].

Most of the $\text{H1}'$ resonances give no cross peak in either HOHAHA or DQF-COSY spectra (data not shown). This is a typical phenomenon for the A form structure, where sugar pucker is C3'-endo , and thus the coupling constant between $\text{H1}'$ and $\text{H2}'$ is almost zero. The $\text{H1}'-\text{H2}'$ distance is the shortest among the $\text{H1}'-\text{H2}'$, $\text{H1}'-\text{H3}'$ and $\text{H1}'-\text{H4}'$ distances for virtually all pucker geometries of the sugar ring [51, 52] (Only for pucker around O4'-endo , is the $\text{H1}'-\text{H4}'$ distance shorter than the $\text{H1}'-\text{H2}'$ distance [51]). Particularly for the C3'-endo pucker suggested above, the $\text{H1}'-\text{H2}'$ distance is much shorter than the other two distances [51]. Thus, $\text{H2}'$ resonances were assigned on the basis of the strong cross peaks in Figure 4. The assignments for $\text{H2}'$ are further confirmed by the successful tracing of the sequential $\text{H2}'(i-1)-\text{H6}/\text{H8}(i)-\text{H2}'(i)$ connectivities in Figure 5. Generally, very strong interresidue $\text{H2}'(i-1)-\text{H6}/\text{H8}(i)$ cross peaks and weaker intraresidue $\text{H2}'(i)-\text{H6}/\text{H8}(i)$ cross peaks are observed. This is again a

characteristics typical of the A form structure with C3'-endo sugar puckering [42, 47, 49, 51, 52].

The H3' resonances were assigned on the basis of H3'(i-1) H6/H8(i)-H3'(i) connectivities (not drawn in Figure 5). Generally, medium to strong cross peaks are observed for both interresidue H3'(i-1)-H6/H8(i) and intrasidue H3'(i)-H6/H8(i) NOEs, which is consistent with the A form structure [51]. The H4' resonances were assigned on the basis of the through-bond connectivities with the H2' and H3'

Figure 5. Expansion of the NOESY spectrum of r(GGACGAGUCC)₂ obtained with a mixing time of 100 ms in a D₂O solution at 33°C. Intraresidue cross peaks between H6/H8 and either H2' (circles) or H3' (triangles) or H4' (rectangles) are indicated. The lines show the H2'(i-1)-H6/H8(i)-H2'(i) connectivities, with the intraresidue cross peaks being labeled. G7H8-A6H3' and C4H6-A3H3' are indicated by solid and dotted arrows, respectively.

resonances in each sugar observed in HOHAHA and DQF-COSY spectra. As an example, the H3'-H4' cross peaks in the DQF-COSY spectrum are shown in Figure 6.

The assignments of the H3' and H4' resonances are confirmed by the H1'-H3' and H1'-H4' cross peaks in Figure 4, and the intraresidue H6/H8-H4' cross peaks in Figure 5. As demonstrated in Figure 3, the examination of a symmetric area and spectra recorded at different temperatures was very helpful to resolve overlapping of closely located resonances.

In Figure 4, the interresidue U8H5-G7H2', H3' and H4' cross peaks, together with the U8H5-G7H1' cross peak, are shown instead of the G7H1'-G7H2', H3' and H4' ones. G7H1' (4.14 ppm) resonates in the region for H2', H3' and H4' resonances. Thus the G7H1'-G7H2', H3' and H4' cross peaks overlap H2'-H3', H2'-H4' and H3'-H4' cross peaks of other residues. Therefore the cross peaks involving G7H1' are not so useful for assignment. So the cross peaks involving U8H5 are shown alternatively. Similar interresidue cross peaks were also observed for C4H5, C9H5 and C10H5, and they were useful for confirming the assignments. The assignments are summarized in Table 2.

DISCUSSION

Comparison of the thermal stabilities of r(GGACGAGUCC)₂ and d(GGACGAGTCC)₂

The observation of concentration-dependency of the melting temperature for both RNA and DNA (Table 1) clearly excludes the possible formation of a hairpin structure. These analyses were necessary, because the hairpin structure was reported for r(CUCAGAGAUGAG) [53], the base sequence of which is also 'self-complementary' except for G:A mismatches in the central region, as well as our sequences. Thus, it turns out that a double stranded structure is formed for our RNA and DNA.

The melting temperature of the RNA duplex is higher than that of the DNA duplex (Table 1), for which it has already been concluded that sheared G:A base pairs are formed, resulting in

Table 1. Melting temperatures (°C) of r(GGACGAGUCC)₂ and d(GGACGAGTCC)₂ obtained from CD melting curves.

	Oligomer strand concentration	
	50 μM	5 μM
r(GGACGAGUCC) ₂	49.3	55.9
d(GGACGAGTCC) ₂	42.7	50.7

Table 2. ¹H chemical shifts (ppm) for r(GGACGAGUCC)₂ at 33°C^a.

Residue	H6/H8	H2/H5	H1'	H2'	H3'	H4'	imino
G1	7.82	—	5.54	4.71	4.44	4.16	12.28
G2	7.36	—	5.79	4.57	4.48	4.48	12.48
A3	7.72	7.71	5.96	4.59	4.48	4.53	—
C4	6.97	4.98	5.18	4.43	4.40	4.13	—
G5	7.89	—	5.66	4.27	4.84	4.49	9.89
A6	7.87	7.88	5.87	4.88	4.45	4.57	—
G7	7.41	—	4.14	4.26	4.29	4.09	13.28
U8	7.68	5.00	5.51	4.46	4.42	4.42	14.19
C9	7.84	5.61	5.56	4.25	4.43	4.43	—
C10	7.64	5.50	5.64	4.03	4.16	4.13	—

^aAt 5°C for imino protons.

an intensive stacking interaction between the two G:A base pairs [24]. This implies that some kind of G:A base pairs is also formed for the RNA duplex, because its melting temperature must be lower than that of the DNA duplex if G:A base pairs are not formed, and thus G and A bases are flipped out of the duplex.

Overall structure of the RNA duplex with G:A mismatches

The overall structure of the RNA duplex can be assessed on the basis of qualitative information on coupling constants and NOEs. The absence of H1'–H2' cross peaks in both the HOHAHA and DQF-COSY spectra for most residues (data not shown) indicates C3'-endo sugar puckering characteristic of the A form structure. The appearance of H3'–H4' cross peaks in the DQF-COSY spectrum (Figure 6) is consistent with the C3'-endo puckering. Exceptionally, G5 gives a H1'–H2' cross peak in the HOHAHA spectrum, although it is very weak. This suggests a subtle deviation from the C3'-endo puckering for the G5 sugar involved in the G:A mismatch. The very strong H1'–H2' cross peak in the NOESY spectrum for each residue (Figure 4) is another indication of the C3'-endo puckering.

The observation of strong interresidue H2'(i–1)–H6/H8(i) and weaker intraresidue H2'(i)–H6/H8(i) cross peaks in the NOESY spectrum (Figure 5) indicates that the overall structure of the RNA duplex with G:A mismatches is the A form. Observation of the medium to strong interresidue H3'(i–1)–H6/H8(i) and intraresidue H3'(i)–H6/H8(i) cross peaks in the NOESY spectrum (Figure 5) is another indication of the A form structure. It should be noted that the C4H6–A3H3' cross peak (indicated by a dotted arrow in Figure 5) is exceptionally weak and that the G7H8–A6H3' cross peak (indicated by a solid arrow in Figure 5) is particularly strong. This implies some minor perturbation of the A form structure around the G:A mismatch region. The position of the A3H3' resonance was identified on the basis of the A3H8–A3H3' and A3H1'–A3H3' cross peaks.

The cross peak a in Figure 3(a) indicates the A3H2–C4H1' NOE. Appearance of the AH2(i)–H1'(i+1) cross peak is also

consistent with A form structure, but not with B form structure [51]. Similarly the A6H2–G7H1' cross peak is observed (cross peak b in Figure 3(a)), although it is overlapping another cross peak, G5H8–C4H4'.

In summary, it is concluded that the overall structure of the RNA duplex with the G:A mismatches is the A form on the basis of five lines of experimental evidence: (1) the absence of H1'–H2' and presence of H3'–H4' cross peaks in the HOHAHA and DQF-COSY spectra, (2) very strong H1'–H2' cross peaks in the NOESY spectra, (3) strong H2'(i–1)–H6/H8(i) and weaker H2'(i)–H6/H8(i) cross peaks in the NOESY spectra, (4) medium to strong H3'(i–1)–H6/H8(i) and H3'(i)–H6/H8(i) cross peaks in the NOESY spectra, and (5) appearance of AH2(i)–H1'(i+1) cross peaks in the NOESY spectra. This conclusion of the A form structure should apply to the G:A mismatch region, too. Although H1'–H2' cross peak is observed for G5 in the HOHAHA spectra, it is still very weak. G5 gives a very strong H1'–H2' cross peak in the NOESY spectrum like other residues. Therefore, the deviation of the G5 sugar from C3'-endo puckering should be subtle. And nearly all the intra- and interresidue H2'–H6/H8 and H3'–H6/H8 cross peaks of each residue, including the G:A mismatch residues, show the intensities expected for the A form.

The local structure of the G:A mismatches in the RNA duplex

As mentioned under RESULTS, intra- and interresidue H1'–H6/H8, H2'–H6/H8 and H3'–H6/H8 connectivities were all traced completely without interruption. This suggests once again that the G and A bases of the G:A mismatches are not flipped out of the duplex, but accommodated in the duplex, forming base pairs. The G7H1' resonance was found at an extremely high field position (4.14 ppm). A similar extreme high field shift was reported for the hairpin structures of r(CUCAG-AGAUGAG) (U9H1', 3.92 ppm) [53] and r(GGGCGCAA-GCCUUAU) (G9H1', 3.66 ppm) [54]. This result also made it necessary to check the concentration-dependency of the melting temperature in order to examine the possible formation of a hairpin structure in our RNA. It was shown that a hairpin structure is not formed for our RNA. A ring current effect may cause the extreme up-field shift of the G7H1' resonance, as discussed below.

There are four different types of G:A base pairing; (a) 'head to head' G(anti):A(anti), (b) G(anti):A(syn), (c) G(syn):A(anti), and (d) 'sheared' or 'side by side' G(anti):A(anti). The imino proton of G5 of the RNA duplex appears in the high field region at 9.89 ppm (Figure 2a), as observed for the DNA duplex (10.29 ppm) [24]. This indicates that the imino proton of G5 is not involved in the hydrogen bonding, because a hydrogen-bonded imino proton signal of G usually appears at 12.0–13.5 ppm. Thus (a) and (b), where the imino proton of G is the donor for the hydrogen bonding, are excluded as the G:A base pair of the RNA duplex. The absence of a strong NOE to the aromatic proton on irradiation of the imino proton of G5 (Figure 2c) also excludes (a) and (b), because a strong NOE to either A6H2 or A6H8 is expected for (a) and (b), respectively. A weak NOE (7.83 ppm) could be the NOE to A6H8 caused by spin diffusion via amino protons of G5.

Before analyzing NOESY spectra in detail, it was checked whether or not the factor, ρ^*/τ_c , can be assumed to be constant for each residue in the RNA duplex [55]. The intensities of CH5–CH6 and UH5–UH6 cross peaks in the NOESY spectrum

Figure 6. Expansion of the DQF-COSY spectrum of r(GGACGAGUCC)₂ obtained in a D₂O solution at 33°C. Intraresidue H3'–H4' crosspeaks are labelled with their residue numbers. Positive and negative levels are indicated by solid and dotted lines, respectively.

with the mixing time of 100 ms were very similar for each residue, which suggests the ρ^*/τ_c can be assumed to be constant in the duplex.

In the *syn* conformation the H8(i)–H1'(i) distance is ~ 2.6 Å, while in the *anti* conformation this distance is ~ 3.8 Å [51]. The H8(i)–H1'(i) distances of G5 and A6 were calculated on the basis of the intensities of the cross peaks in the NOESY spectrum by using geometrically fixed CH5–CH6 and UH5–UH6 distances as the internal reference distances [40]. The NOESY spectrum with the mixing time of 100 ms was used in order to avoid the effect of spin diffusion to a reasonable extent. And the area shown in Figure 3(b) was used, because discrimination of the closely located G5H8 (7.89 ppm), A6H8 (7.87 ppm) and A6H2 (7.88 ppm) resonances is possible, and crowded cross peaks can be evaluated separately due to the better resolution with aromatic resonances in this area. The calculated H8–H1' distance is 3.6 Å for both G5 and A6. When it is taken into account that the obtained distances tend to be underestimates due to spin diffusion [56], the results clearly indicate the *anti* conformation for both G5 and A6. Thus (b) and (c) are excluded as the G:A base pair of the RNA duplex. Thus (d), the sheared type, turns out to be the only base pairing structure which is consistent with the experimental data.

Model building of the RNA structure

It was concluded in the previous sections that the overall structure of the RNA duplex is the A form and that the type of G:A base pairing is the sheared type. Therefore a model was built by incorporating the two consecutive sheared G:A base pairs in the A form structure (Figure 7). Then the energy was minimized by the use of X-PLOR [57] to avoid steric hindrance.

As a consequence of unique base pairs of the sheared type, the mode of base stacking becomes extremely different from the standard one. One A6 base stacks on another A6 base of the other strand, while an A6 base does not stack on the G5 base of the same strand at all. This structure is supported by the observation of the very strong unique interstrand A6H2–A6H1' cross peak (Figure 3). The corresponding interstrand distance is very short in the structure shown in Figure 7a.

The unique stacking mode also gives a hint as to why the imino proton of G5 is visible, although it is not involved in the hydrogen bonding. Probably, it is due to reduced accessibility of the imino proton to the solvent water (Figure 7a). The structure also explains well the extreme up-field shift of the resonance of G7H1' which is located close to base A6 (Figure 7b). Its ring current effect on G7H1' could be the reason for the large up-field shift.

Similarities and differences between the RNA and DNA duplexes with G:A mismatches

Here we have reported a structural study on $r(\text{GGACG-AGUCC})_2$, and we previously reported a study on the corresponding DNA, $d(\text{GGAC GAGTCC})_2$ [24]. Thus, similarities and differences between the two structures can be addressed.

In both the RNA and DNA duplexes, the two consecutive G:A mismatches form unique sheared G:A base pairs, where an amino proton instead of an imino proton of G is involved in the hydrogen bonding, and G and A bases are arranged 'side by side' instead of 'head to head'. Because of the 'side by side' arrangement of the G and A bases, unique interstrand stacking is achieved in the central G5A6:G5A6 segment at the cost of ordinary

Figure 7. Mode of base stacking model built on the basis of the experimental observations for the G5A6:G5A6 segment (a) and the A6G7:C4G5 segment (b) of the RNA duplex. The upper bases are shown as thick lines. The proximity of interstrand A6H2–A6H1' is indicated by double-headed arrows in (a), and G7H1' is indicated by a single-headed arrow in (b).

intrastrand stacking. In particular, the interstrand stacking between the two A6 bases is very good in both the RNA and DNA duplexes.

The overall structures of the RNA and DNA duplexes are different as expected. The RNA duplex assumes the A form, while the DNA duplex assumes the B form. As a consequence of this difference, the stacking interaction in the central G5A6:G5A6 segment is not the same in the RNA and DNA duplexes. Although one A6 base stacks on another A6 base of the other strand very well in both the RNA and DNA duplexes, interstrand stacking of the two G5 bases is poor in the RNA duplex (Figure 7a), while it is very good in the DNA duplex (Figure 6 in reference [24]).

The difference in the overall structure makes the stacking interaction in the A6G7:C4G5 segment also different. The stacking is less in the RNA duplex. Additionally, G7H1' is located close to the A6 base in the RNA duplex, but not in the DNA duplex. This could explain the extreme up-field shift of the G7H1' resonance observed exclusively for the RNA duplex.

Our RNA duplex is thermally more stable than the corresponding DNA duplex. The higher stability of RNA over DNA is a general phenomenon. In the case of our RNA and DNA, the type of base pairing is the same, and the stacking interaction seems to be better in the DNA duplex. The higher stability of the RNA duplex might be due to the formation of extra hydrogen-bonds involving 2'OH. For instance, the 2'OH

of G5 could be close to the amino group of A6 of the opposite strand and also to O5' of A6 of the same strand. Concerning the extra hydrogen bonds involving the 2'OH, the water bridge between the 2'OH and the 3'phosphate group was suggested [58, 59]. However this bridge is not likely in our case, because a characteristic resonance at ca. 6.5 ppm is not observed.

The structure of the G:A base pair in RNA

To our knowledge, only four studies have been reported on the structure of G:A mismatches in RNA [53, 54, 60, 61], although many studies have been reported on that in DNA [12–28]. A hairpin structure was found in three cases [53, 54, 60], and a duplex structure with a bulged residue and an uncommon reverse-Hoogsteen A:U base pair in another case [61]. Knowledge on the structure of the G:A mismatches in a complete duplex has not been available. Our study on r(GGACGAGUCC)₂ provides the first knowledge on the G:A mismatch structure in the complete RNA duplex, together with the very recent report by Turner's group [62]. Both our and their results indicate the formation of sheared G:A base pairs in the double-stranded A form structure. It is meaningful to demonstrate that sheared G:A base pairs can be formed not only in the B form structure of DNA but also in the A form structure of RNA. Our result for r(GGACGAGUCC)₂ is of particular interest, because it was modelled after both a hammerhead ribozyme and a lead ribozyme, which contain a base sequence identical or very similar to that of the first six residues of our RNA.

It is interesting that Turner's group also observed an extreme up field shift of H1' for the residue corresponding to G7 in our case (G6H1', 4.32 ppm in their case), which suggests a similar stacking interaction in the two RNA duplexes. Moreover, the G7H8–A6H3' cross peak is very strong in our case, as pointed out above, and the corresponding cross peak in their case (G6H8–A5H3') is also strong. On the other hand, the C4H6–A3H3' crosspeak is weak in our case, as described above, while such a weak H6/H8(i)–H3'(i–1) cross peak was not found in their case. Moreover, A6H8–G5H2' and G7H8–A6H2' cross peaks exhibit medium to strong intensities in our case, as mentioned above, while the corresponding cross peaks in their case (A5H8–G4H2' and G6H8–A5H2') are weak. These detailed differences suggest some minor difference between the two structures, although the main features of the RNA duplex with the G:A mismatches were solidly established in the two studies consistently.

In the case of DNA, the type of G:A base pairing depends on neighbouring base sequences [22, 28–30]. It must be interesting to examine the sequence-dependency of G:A base pairing in RNA.

ACKNOWLEDGEMENTS

We wish to thank Professor Akutsu (Yokohama Natl. Univ.) for giving us the opportunity to use his NMR apparatus. This research was supported in part by a Grant-in-Aid for Scientific Research on Priority Area (#03242101) from the Ministry of Education, Science and Culture of Japan. M.K. was supported by grants from the Ministry of Education, Science and Culture of Japan (#05780473), the CIBA-GEIGY Foundation for the Promotion of Science, Takarazuka, Japan, the Kihara Yokohama Life Science Foundation, and the Kanagawa Science Technology Academy.

REFERENCES

1. Cech, T.R. (1987) *Science*, **236**, 1532–1539.
2. Uhlenbeck, O.C. (1987) *Nature*, **328**, 596–600.
3. Koizumi, M., Iwai, S. and Ohtsuka, E. (1988) *FEBS Lett.*, **228**, 228–230.
4. Ruffner, D.E., Stormo, G.D. and Uhlenbeck, O.C. (1990) *Biochemistry*, **29**, 10695–10702.
5. Perreault, J.P., Wu, T.F., Cousineau, B., Ogilvie, K.K. and Cedergren, R. (1990) *Nature*, **344**, 565–567.
6. Heus, H.A., Uhlenbeck, O.C. and Pardi, A. (1990) *Nucleic Acids Res.*, **18**, 1103–1108.
7. Pieken, W.A., Olsen, D.B., Benseler, F., Aurup, H. and Eckstein, F. (1991) *Nature*, **253**, 314–317.
8. Fu, D.J. and McLaughlin, L.W. (1992) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 3985–3989.
9. Williams, D.M., Pieken, W.A. and Eckstein, F. (1992) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 918–921.
10. Taylor, N.R., Kaplan, B.E., Swiderski, P., Li, H. and Rossi, J.J. (1992) *Nucleic Acids Res.*, **20**, 4559–4565.
11. McCall, M.J., Hendry, P. and Jennings, A. (1992) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 5710–5714.
12. Li, Y., Zon, G. and Wilson, W.D. (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 26–30.
13. Uesugi, S., Odai, O., Kodama, H., Hiroaki, H., Sakata, T. and Tanaka, T. (1992) In Sarma, R.H. and Sarma, M.H. (eds.), *Structure & Function, Adenine Press*, New York, Vol. II, pp. 143–158.
14. Kan, L.S., Chandrasegaran, S., Pulford, S.M. and Miller, P.S. (1983) *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 4263–4265.
15. Prive, V.I., Heinemann, U., Chandrasegaran, S., Kan, L.S., Kopka, M.L. and Dickerson, R.E. (1987) *Science*, **238**, 498–504.
16. Carbonnaux, C., van der Marel, G.A., van Boom J.H., Guschlbauer, W. and Fazakerley, G.V. (1991) *Biochemistry*, **30**, 5449–5458.
17. Brown, T., Hunter, W.N., Kneale, G. and Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 2402–2406.
18. Gao, X. and Patel, D.J. (1988) *J. Am. Chem. Soc.*, **110**, 5178–5182.
19. Brown, T., Leonard, G.A., Booth, E.D. and Chambers, J. (1989) *J. Mol. Biol.*, **207**, 455–457.
20. Li, Y., Zon, G. and Wilson, W.D. (1991) *Biochemistry*, **30**, 7566–7572.
21. Chou, S.H., Cheng, J.W. and Reid, B.R. (1992) *J. Mol. Biol.*, **228**, 138–155.
22. Cheng, J.W., Chou, S.H. and Reid, B.R. (1992) *J. Mol. Biol.*, **228**, 1037–1041.
23. Maltseva, T., Sandstrom, A., Ivanova, I.M., Sergeyev, D.S., Zarytova, V.F. and Chattopadhyaya, L. (1993) *J. Biochem. Biophys. Methods*, **26**, 173–236.
24. Katahira, M., Sato, H., Mishima, K., Uesugi, S. and Fujii, S. (1993) *Nucleic Acids Res.*, **21**, 5418–5424.
25. Maskos, K., Gunn, B.M., LeBlanc, D.A. and Morden, K.M. (1993) *Biochemistry*, **32**, 3583–3595.
26. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K. and Miura, K. (1994) *Nucleic Acid Res.*, **22**, 576–582.
27. Greene, K.L., Jones, R.L., Li, Y., Robinson, H., Wang, A.H.J., Zon, G. and Wilson, W.D. (1994) *Biochemistry*, **33**, 1053–1062.
28. Katahira, M., Kanagawa, M. and Uesugi, S. (1994) *Nucleosides & Nucleotides*, in press.
29. Fazakerley, G.V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G.A., van Boom, J.H., Jones, M. and Radman, M. (1986) *EMBO J.*, **5**, 3697–3703.
30. Ebel, S., Lane, A.N. and Brown, T. (1992) *Biochemistry*, **31**, 12083–12086.
31. Odai, O., Kodama, H., Hiroaki, H., Sakata, T., Tanaka, T. and Uesugi, S. (1990) *Nucleic Acids Res.*, **18**, 5955–5960.
32. Pan, T. and Uhlenbeck, O.C. (1992) *Nature*, **358**, 560–563.
33. Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R. and Ken-ichi Tomita. (1988) *Nucleic Acids Res.*, **17**, 8619–8632.
34. Jeener, L., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.
35. Davis, D.G. and Bax, A. (1985) *J. Am. Chem. Soc.*, **107**, 2821–2822.
36. Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.
37. Marion, D. and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
38. Clore, G.M., Kimber, B.J. and Gronenborn, A.M. (1983) *J. Magn. Reson.*, **54**, 170–173.
39. Katahira, M., Lee, S.J., Kobayashi, Y., Sugeta, H., Kyogoku, Y., Iwai, S., Ohtsuka, E., Benevides, J.M. and Thomas, G.J. (1990) *J. Am. Chem.*

- Soc., **112**, 4508–4512.
40. Katahira, M., Sugeta, H., Kyogoku, Y. and Fujii, S. (1990) *Biochemistry*, **29**, 7214–7222.
41. Katahira, M., Sugeta, H. and Kyogoku, Y. (1990) *Nucleic Acids Res.*, **18**, 613–618.
42. Reid, D.G., Salisbury, S.A., Bellard, S., Shakked, Z. and Williams, D.H. (1983) *Biochemistry*, **22**, 2019–2025.
43. Scheek, R.M., Russo, N., Boelens, R., Kaptein, R. and van Boom, J.H. (1983) *J. Am. Chem. Soc.*, **105**, 2914–2916.
44. Hare, D.R., Wemmer, D.E., Chou, S.H., Drobny, G. and Reid, B.R. (1983) *J. Mol. Biol.*, **171**, 319–336.
45. Feigon, J., Denny, W.A., Leupin, W. and Kearns, D.R. (1983) *Biochemistry*, **22**, 5930–5942.
46. Clore, G.M. and Gronenborn, A.M. (1983) *EMBO J.*, **2**, 2109–2115.
47. Mellema, J.R., Haasnoot, C.A.A., van der Marel, G.A., Wille, G., van Boeckel, C.A.A., van Boom, J.H. and Altona, C. (1983) *Nucleic Acid Res.*, **11**, 5717–5738.
48. Clore, G.M., Lauble, H., Frenkiel, T.A. and Gronenborn, A.M. (1984) *Eur. J. Biochem.*, **145**, 629–636.
49. Clore, M., Gronenborn, A.M. and McLaughlin, L. (1985) *Eur. J. Biochem.*, **151**, 153–165.
50. Weiss, M.A., Patel, D.J., Sauer, R.T. and Karplus, M. (1984) *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 130–134.
51. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, Inc.
52. Chou, S.H., Flynn, P. and Reid, B. (1989) *Biochemistry*, **28**, 2422–2435.
53. Orita, M., Nishikawa, F., Shimayama, T., Taira, K., Endo, Y. and Nishikawa, S. (1993) *Nucleic Acids Res.*, **21**, 5670–5678.
54. Heus, H.A. and Pardi, A. (1991) *Science*, **253**, 191–194.
55. Derome, A.E. (1987) *Modern NMR Techniques for Chemistry Research*, Pergamon Press.
56. Nilsson, L., Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1986) *J. Mol. Biol.*, **188**, 455–475.
57. Brünger, A.T. (1990) *X-PLOR: A System for Crystallography and NMR*, Yale University Press, New Haven, CT.
58. Bolton, P.H. and Kearns, D.R. (1978) *Biochim. Biophys. Acta*, **517**, 329–337.
59. Bolton, P.H. and Kearns, D.R. (1979) *J. Am. Chem. Soc.*, **101**, 479–484.
60. SantaLucia, J., Jr., Kierzek, R. and Turner, D.H. (1992) *Science*, **256**, 217–219.
61. Wimberly, B., Varani, G. and Tinoco, I., Jr. (1993) *Biochemistry*, **32**, 1078–1087.
62. SantaLucia J., Jr., and Turner, D. (1993) *Biochemistry*, **32**, 12612–12623.