Unique Pair-to-loop Dynamism due to G–G Mismatch at the N site in d-(GTTAATNATTAAC): NMR Characterization

P. K. Radha, Anup Madan and R. V. Hosur*

Chemical Physics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay-400005, India

A unique G–G mismatch-induced, pair-to-loop equilibrium has been observed in the AT rich DNA sequence recognised by homoeodomain protein LFB1.

d-TTAATNATTAA is a unique sequence recognised by the homeodomain of LFB1. The sequence is self-complementary but for the N, which can be any nucleotide T, A, G or C. The role of N, whose identity is apparently unimportant, in the otherwise symmetric duplex is not yet clear. Here, we describe one effect of structural variations introduced by the variations at the N site: a unique pair-to-loop equilibrium for a G–G mismatch.

Fig. 1 shows a comparison of exchangeable imino ^{1}H NMR spectra of (*a*) GTTAATCATTAAC (C-13) and (*b*) GTTAATGATTAAC (G-13) under identical conditions. The former has a C–C mismatch and the latter has a G–G mismatch at the N site discussed above. While the number of lines in the C-13 spectrum is indicative of a single symmetric duplex, the G-13 spectrum has more lines (see also Fig. 2) suggesting the existence of two different base-paired structures. Here, the possibility of a single completely asymmetric duplex can be eliminated on the basis that: (*i*) the C-13 spectrum, which represents a symmetric duplex, has equivalent signals in the G-13 spectrum; (*ii*) 2D ROESY[1] spectrum in D_2O solution shows exchange cross peaks (data not shown), which is not to be expected for a single rigid structure.

PAGE (poly acrylamide gel electrophoresis) analysis of G-13 under non-denaturing conditions (data not shown) showed a single band at the molecular mass of a 13 *mer* duplex indicating that the two base paired structures are both duplexes. The presence of the C-13 signals in the G-13 spectrum characterises one of the G-13 duplexes as being similar to the C-13 duplex. The remaining imino peaks in the G-13 spectrum in Fig. 1 characterise, therefore, the other duplex of the molecule. Among these additional resonances, the lone distant signal at δ 12.4, which is in the typical range of G imino protons in duplexes[2] can be readily attributed to G-7 and indicates G–G base pair formation; several hydrogen-bonding schemes are known for such pair formation. The two peaks at δ 13.3 and 13.4 (9 °C) which are not seen in the C-13 spectrum can be assigned to the two thymidine imino protons on opposite strands (T-6, T-6′) adjacent to the G-7 pair site, on the ground that structural distortions due to G–G pair formation are expected to be greatest in the near vicinity of the pair. The observation of two separate signals indicates that the G–G pairing has introduced a certain asymmetry at the centre. Further, the absence of these signals in the C-13 spectrum implies that in the C-13 duplex A–T pairing is not maintained at 6 and 8 positions and a loop is formed at the centre. Thus, the two duplexes of G-13 designated as G_L and G_P in Scheme 1, originate from the looping out or pairing of the G nucleotide at the centre of the sequence. The two structures have small stacking differences in the TTAA main stem resulting in separate signals for some of the thymidines.

Fig. 2 shows the temperature dependence of the G-13 imino proton spectrum, which permits partial assignment of the resonances and also allows investigation of the dynamism and characterization of the two duplexes. As the duplexes melt with base pairs opening progressively from the ends, it is expected that the terminal imino protons broaden first, followed by the interior ones as the temperature is increased. Thus, the terminal G-1 and T-2 imino resonances can be readily identified as indicated in Fig. 2. The intensities of the signals at the lowest temperature are indicative of two protons each, implying that for the two duplexes the G-1 and T-2 signals are equivalent; this must be expected as they are far

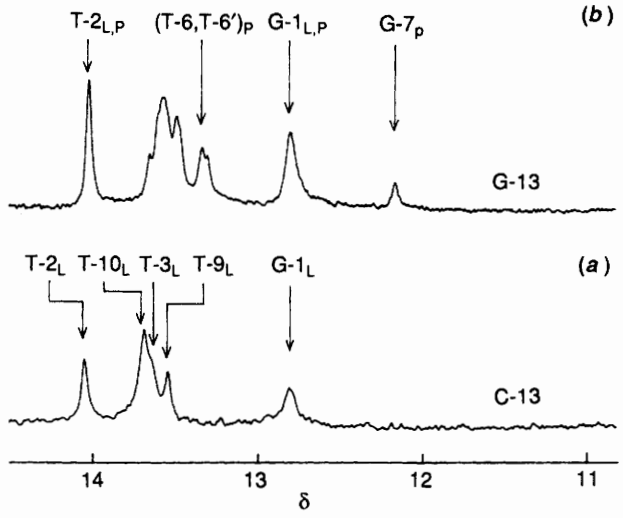

Fig. 1 Comparison of the imino proton regions of the 1D spectra of the C-13 and G-13 molecules in H_2O, at 13 °C. Sequence specific assignments are indicated. Specific assignment of T-9 and T-10 in C-13 spectrum are obtained from 2D NOESY[8] spectrum on the basis of T-imino to AH_2 and ANH_2 NOE cross peaks, not described in this paper. Subscript L stands for the looped conformation and P for the paired conformation. The oligonucleotides were synthesised on the Applied Biosystems Synthesizer Model 381 and purified using standard methods. The sample was prepared in a mixture of 90 H_2O and 10% D_2O which was 10 mmol dm^{-3} in phosphate buffer, 0.1 mol dm^{-3} in NaCl, 1 mmol dm^{-3} in EDTA, pH 7.0. The spectrum was recorded on Bruker AMX-500 spectrometer by using the jump and return sequence[7] and processed on X32 workstation using exponential line broadening to enhance S/N ratio.

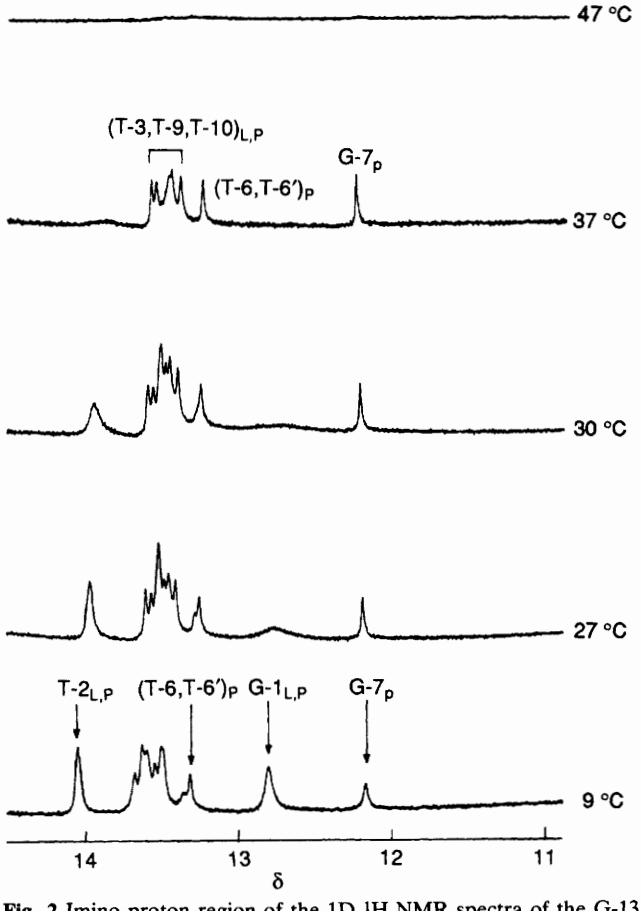

Fig. 2 Imino proton region of the 1D ^1H NMR spectra of the G-13 molecules recorded at different temperatures. The experimental conditions were same as indicated in the previous figure.

removed from the mismatch site. This leaves six signals in the δ 13.4–13.8 range, which can be attributed to thymidine imino protons in the TAA segments. Some of the specific assignments of these have been obtained from comparison of 2D NOESY spectra of C-13 and G-13 samples at 13 °C (data not shown).

In the temperature profiles in Fig. 2 it is observed that all the thymidine signals in the TAA stem and also the G-7 and (T-6, T-6') signals have similar line-broadening features. The signals of both the duplexes are stable up to 37 °C and completely disappear on increasing the temperature to 47 °C. This implies that both the duplexes have nearly equal stabilty. Besides, it is also seen that T-6, T-6' signals become equivalent as the temperature is increased. This indicates that a conformational flip-flop occurs at the G-7 site on the two strands, removing the asymmetry at G–G pair and rendering the two strands equivalent.

The observed dynamism in the G-13 is shown in Scheme 1.

Scheme 1

The results reported here are important on the following counts: (*i*) In the other investigations on G–G mismatches which were on G–C rich sequences,[3–5] only the G–G paired conformation was observed under similar experimental conditions as in the present study. This suggests sequence and composition dependence of structural variations caused by G–G mismatches, (*ii*) G–G mismatch introduces a greater dynamism than a C–C mismatch in the LFB1 recognition sequence. This can perhaps be generalised to purine–purine and pyrimidine–pyrimidine mismatches. (*iii*) While LFB1 may be transparent to A–T or G–C pairing at the N site, it is unlikely to be transparent to the G–G or C–C mismatches. (*iv*) Designing of chemical reagents for recognising G mismatch sites[6] has to take into consideration conformational dynamism.

The facilities provided by 500 MHz FTNMR National Facility supported by the Department of Science and Technology are gratefully acknowledged.

Received, 28th May 1993; Com. 3/03062H

References

1 A. A. Bothner-By, R. L. Stephens, J. Lee, C. D. Warren and R. W. Jeanholz, *J. Am. Chem. Soc.*, 1984, **106**, 811.
2 K. Wuthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
3 V. A. Roongta, C. R. Jones and D. G. Gorenstein, *Biochemistry*, 1990, **29**, 5245.
4 J. A. H. Cognet, J. Gabarro-Arpa, M. Le Bret, G. A. van der Marel, J. H. van Boom and G. V. Fazakerley, *Nucl. Acids Res.*, 1991, **19**, 24, 6771.
5 Y. Oda, S. Uesugi, M. Ikehara, Y. Kawase and E. Ohtsuka, *Nucl. Acids. Res.*, 1991, **19**, 19, 5263.
6 X. Chen, J. Burrows and S. E. Rotika, *J. Am. Chem. Soc.*, 1992, **114**, 322.
7 P. Plateau and M. Gueron, *J. Am. Chem. Soc.*, 1982, **104**, 7310.
8 A. Kumar, R. R. Ernst and K. Wuthrich, *Biochem. Biophys. Res. Commun.*, 1980, **95**, 1.