Evidence for $A^+(anti)$–G(*syn*) mismatched base-pairing in d-GGTAAGCGTACC

Apurba K. Sau[a], Kandala V.R. Chary[a], Girjesh Govil[a,*], Chang-qing Chen[b], Frank B. Howard[b], H. Todd Miles[b]

[a]*Chemical Physics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India*
[b]*National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA*

Received 30 October 1995

Abstract Two-dimensional NMR spectroscopy has been used to study the structure and hydrogen bonding scheme of A:G mismatched base pairing in d-GGTAAGCGTACC at pH 5.8. Under the conditions of our study, the molecule forms a B-DNA helix, with the mismatched bases in the $A^+(anti)$–G(*syn*) conformation. The adenosine exists in the protonated form. The NOESY spectrum in 90% H_2O + 10% ^2H_2O has been used to assign all observable imino and amino protons including those involved in the $A^+(anti)$–G(*syn*) base pair. Both the proton donors in the A:G mismatched inter-base hydrogen bonding are situated on adenosine.

Key words: A:G base pairing; Mismatch base pair; DNA; Nuclear Overhauser enhancement spectroscopy; Clean total correlation spectroscopy

1. Introduction

Non-Watson-Crick (mismatch) base-pairs are occasionally formed during DNA replication and genetic recombination. A:G mispairs are particularly important as they are removed from genomic DNA less efficiently than other mismatches, possibly because such a base pair does not significantly disturb the conformation of the DNA double helix [1]. Structural studies by X-ray diffraction and NMR methods on A:G mismatched DNA have been reported. Isolated A:G mismatches have been shown to adopt A(*anti*)–G(*anti*) or $A^+(anti)$–G(*syn*) conformations [2,3].

We have undertaken a study on the effect of perturbations due to base pair mismatch using the basic deoxyribose oligonucleotide I. This oligomer contains recognition and cleavage sites for restriction endonucleases *Rna*I and *Fnu*DII and adopts a uniform B-DNA conformation [4]. Perturbations were made at positions 6, 7 and 5, 8 of the base sequence thus introducing AI (II, abbreviated as IA67) and AG (III, abbreviated as AG58) mispairs, respectively.

```
1  4    8    12
GGTACGCGTACC
CCATGCGCATGG

     I

GGTACIAGTACC
CCATGAICATGG

     II

GGTAAGCGTACC
CCATGCGAATGG

     III
```

Through the use of UV, IR, NMR, gel electrophoresis, and equilibrium ultracentrifugation [5–9] techniques, it has been established that IA67 exists predominantly as a monomeric hairpin in dilute solutions, which is in equilibrium with a duplex of twice the molecular weight. The loop size was shown to be two for the hairpin. Higher concentration shifts the equilibrium from hairpin to duplex [8,9]. The IA mispair in the duplex adopts I(*anti*)–A(*anti*) conformation [8]. In this paper, we report NMR studies on AG58. UV and CD spectroscopy, more convenient for small incremental measurements, are used to illuminate specific chemical points and to show the context in which the NMR observations are made.

2. Materials and methods

AG58 was synthesized with Applied Biosystems Model 380 B DNA synthesizer using a solid-phase cyanoethylphosphoramidite chemistry. Molar extinction coefficients were measured by phosphate analysis, as described previously [10]. About 8 mg of material was dissolved in 0.4 ml of appropriate solvent (~5 mM strand concentration or 60 mM in nucleoside residues), buffered with 0.05 M sodium phosphate buffer (pH 5.8) and to which 0.1 M NaCl was added. Concentration of the oligomer (in nucleoside residues) was 10^{-4} M for CD and 4×10^{-4} M for UV experiments.

Proton NMR experiments were carried out on a Bruker AMX 500 spectrometer with a proton frequency of 500 MHz. The spectra in mixed water solvent (90% H_2O + 10% ^2H_2O) include 1D ^1H recorded with P1$\bar{1}$ pulse sequence [11] and 2D nuclear Overhauser enhancement spectroscopy (NOESY) [12] with 1$\bar{1}$ detection pulse and a mixing time of 200 ms. The 2D experiments in ^2H_2O include clean total correlation spectroscopy (clean TOCSY) [13] with a mixing time of 100 ms, NOESY with a mixing time of 120 ms and rotating frame nuclear Overhauser enhancement spectroscopy (ROESY) [14] with a mixing time of 100 ms. Temperature of 10°C was used in most NMR experiments, though 1D experiments were carried out in the range 5–40°C.

*Corresponding author. Fax: (91) (22) 215-2110 or 215-2181; E-mail: govil@tifrvax.tifr.res.in

Abbreviations: 2D, two-dimensional; NOESY, nuclear Overhauser enhancement spectroscopy; clean TOCSY, clean total correlation spectroscopy; ROESY, rotating frame nuclear Overhauser enhancement spectroscopy; nOe, nuclear Overhauser effect; sh, unresolved shoulder.

SSDI 0014-5793(95)01362-8

3. Results and discussion

CD spectra of AG58 in aqueous solutions undergo large changes as the pH is lowered from 7 to 4.56 (Fig. 1). Positive extrema at 282 and 289 nm (sh) at pH 7 are progressively replaced by maxima at 261 and 290 nm at pH 4.56, and the minimum shifts from 249 to 240 nm. Isodichroic points at 242 and 269 nm indicate that only two species are involved in the transition. Cooperative thermal dissociation curves are observed with both UV and CD at all pH values. These curves reach an invariant lower plateau by about 10°C for the pH values of interest.

The mp (T_m) for AG58 is around 35°C for pH > 6.5. It rises almost linearly with decreasing pH to a value of 40°C at pH 4.6. It then declines again at lower pH. Over the pH range 6.1–4.6, the data have a linear dependence on $1/T_m$, with $d(1/T_m)/d(pH) = (2.73 \pm 0.20) \times 10^{-5}$. These results indicate the uptake of protons with decreasing pH and the conversion of the neutral complex to a distinct and more stable structure in the pH range 4–6. However, the mp for AG58 is substantially lower than that of the canonical duplex I (55°C). The lower stability of the structure of AG58 at pH higher than 6, is also reflected in the quality of NMR spectra observed as a function of pH. Around pH 5.8 we observe good spectra indicating that the DNA adopts a distinct and ordered conformation.

Fig. 2 shows the temperature dependence of imino/amino region of ^1H spectrum of AG58 in 90% H_2O and 10% ^2H_2O at pH 5.8. Four resonances are observed in the region 12–14 ppm and two in the region 9–12 ppm. Integration of the resonance at 12.78 ppm ('3') accounts for two imino protons. Classification of these shifts suggests that '1' and '2' arise from thymidine 3NH and '3' and '4' from guanosine 1NH protons. Resonance '5' is assigned to a non-hydrogen bonded imino proton from the fact that it broadens at higher temperatures and vanishes in 1D spectrum recorded with presaturation. Resonance '6' is due to an amino proton. As evident from Fig. 2, resonance at 12.5 ppm ('4') also broadens with increase in temperature and disappears with presaturation; it is therefore assigned to 1NH of the terminal G1.

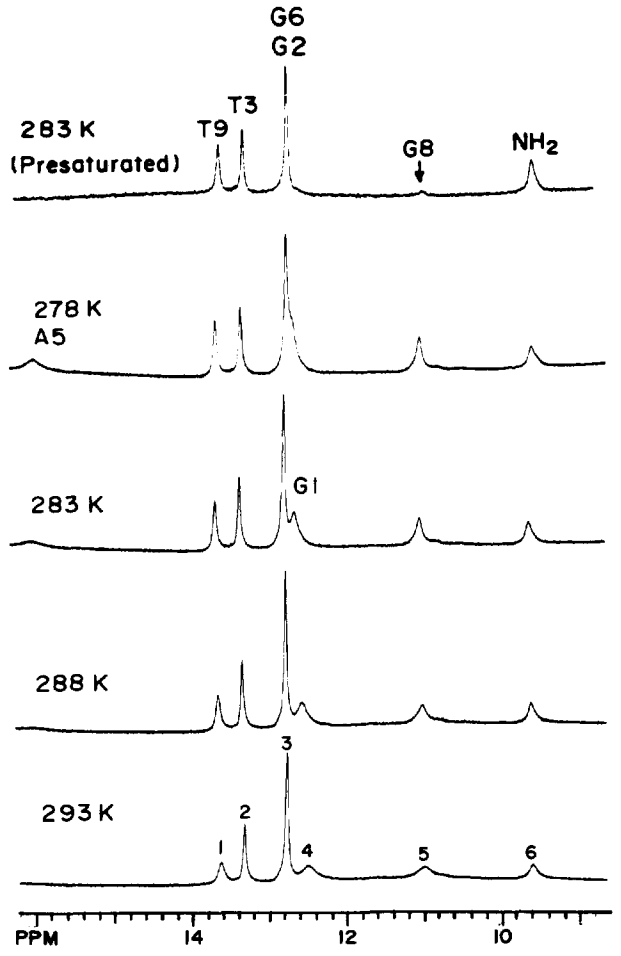

Fig. 2. Temperature dependence of imino and part of amino proton region of the 1D ^1H NMR spectrum of the AG58 recorded in a mixed solvent of 90% H_2O and 10% ^2H_2O at pH 5.8.

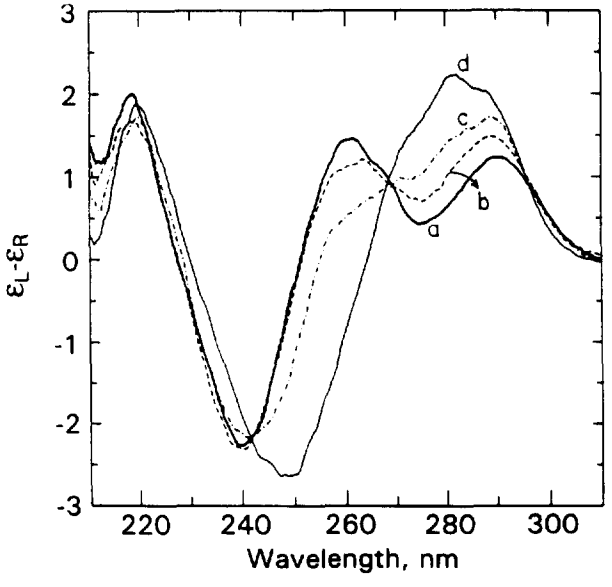

Fig. 1. CD spectra of AG58 at 4°C in 0.5 M NaCl and either (a) 0.002 sodium acetate pH 4.6 (heavy solid line), or 0.002 M sodium cacodylate (b) pH 5.1 (dashed line), (c) pH 5.7 (dot-dashed line), (d) pH 7.0 (light solid line). Isodichroic points occur at 242 and 268 nm.

Sequential resonance assignments of non-exchangeable protons were achieved through established procedures [15–19]. As an illustrative example of the resonance assignment procedure, Fig. 3 shows a selected region (H1'/H5 – H2/H6/H8) of pure absorption NOESY spectrum of AG58 with the sequential connectivities.

The assignments of exchangeable imino and amino protons have been carried out from the NOESY spectrum in 90% H_2O + 10% ^2H_2O. Fig. 4 depicts regions showing nOes between the non-exchangeable and exchangeable protons (ω_1 axis) and the hydrogen-bonded imino protons (ω_2 axis). Protons G1 (1NH) and C12 ($4NH_2$) are not seen, presumably because of broadening arising from the fraying of terminal base pairs. Two of the four imino resonances could be assigned to the thymines from the observation of nOes to their respective methyl groups. From the knowledge of the sequential assignment of CH_3 resonances, peak '1' is assigned to T9 (3NH) and '2' to T3 (3NH). Observations of the strong nOes to A4 (H2) and A10 (H2) in the opposite strands (resulting from the short T (3NH)-A (H2) inter-proton distances in an AT base pair), are in conformity

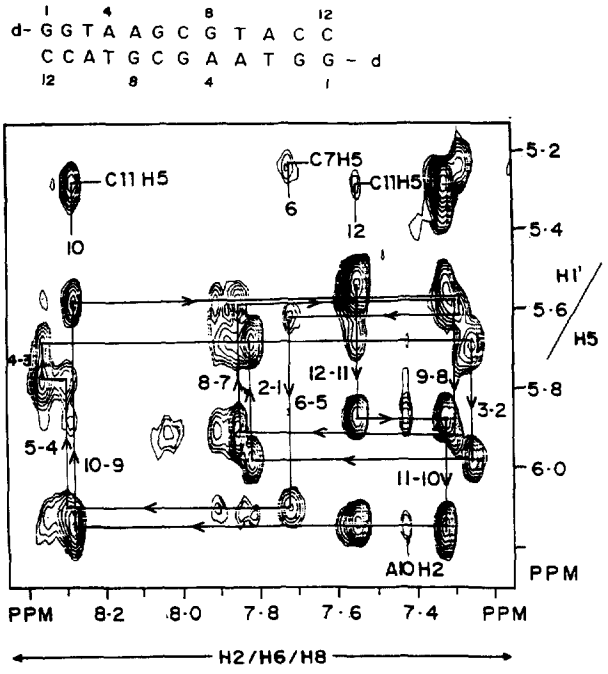

Fig. 3. Selected region of pure-absorption NOESY spectrum of AG58 showing self and sequential nOe connectivities H1'/H5-H2/H6/H8 (solid lines). Experimental parameters: t_{1max} = 62.4 ms, $t_2$max = 540 ms, recycle delay = 1 s, 96 scans/t_1 increment, time-domain data points were 512 and 4096 along t_1 and t_2 dimensions, respectively. The data were multiplied with sine bell window functions shifted by $\pi/4$ and $\pi/8$ along t_1 and t_2 axes, respectively and zero-filled to 2048 data points along t_1 dimension prior to 2D-FT. Digital resolution along ω_1 and ω_2 corresponds to 1.85 Hz/pt.

with the assignments. Peak '2' shows inter-imino proton nOe to '3' (Fig. 4), indicating close contact between T3 (3NH) and the imino proton belonging to the neighbouring G2. The assignment of G2 (1NH) is again consistent with the observation of other nOes: (i) G2 (1NH)–C11 ($4NH_2$); and (ii) G2 (1NH)-A10 (H2). The second resonance under peak '3' is assigned to G6 (1NH) from the observation of nOes from the exposed and the hydrogen bonded amino protons of the base paired cytosine in the opposite strand, i.e. G6 (1NH)-C7 ($4NH_2$). By elimination, peak '5' of Fig. 2 in the non-hydrogen bonded imino proton region is assigned to G8 (1NH). This assignment indicates that G8 (1NH) in the A5G8 base pair does not participate in hydrogen bond.

The imino proton and amino proton resonances, indicate hydrogen bonded base pairs along the length of AG58. Intricate networks of interstrand NOESY cross peaks establish the hydrogen bonded base pairing and thus the secondary structure of AG58. These are: G2 (1NH) to C11 ($4NH_2$); G2 (1NH) to A10 (H2); T3 (3NH) to A10 (H2); A4 (H2) to T9 (3NH); A4 (H2) to G8 (1NH); G6 (1NH) to C7 ($4NH_2$), and the corresponding pairs arising from the two-fold symmetry in the DNA duplex. Qualitative analysis of the relative NOESY cross-peak intensities establishes that the molecule adopts a right handed B-DNA duplex. The nOe data further confirms that the GC and AT base pairs are associated through the usual Watson and Crick base pairing schemes with all these bases are in *anti* conformation. There is no evidence for the formation of a hairpin structure from NMR at 5 mM concentration or CD and UV studies which were performed at a much lower concentrations.

It is known that amino protons of adenosine belonging to AT base pairs exchange fast and are broad, while the amino protons of C belonging to GC base pairs show sharp peaks in the range 6–8.4 ppm. The peak at 9.6 ppm ('6') arising from an amino proton shows a strong nOe to another peak at 8.28 ppm (Fig. 5). This indicates that the two resonances belong to the same amino group. A large chemical shift difference of 1.32 ppm between the two peaks indicates that the proton at 9.6 ppm is hydrogen bonded while the one at 8.28 ppm is not. This is also evident from the water presaturated 1D spectrum (not

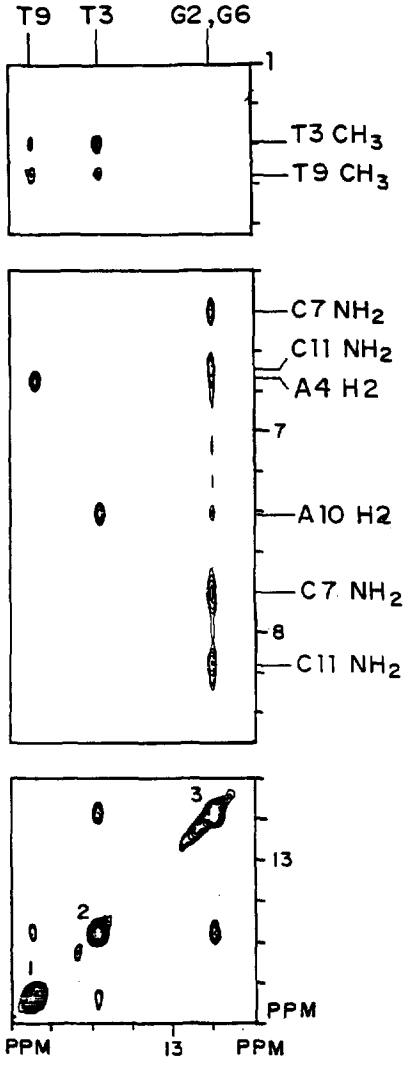

Fig. 4. Selected regions of pure-absorption NOESY spectrum of the AG58 taken in a mixed solvent of 90% H_2O + 10% ^2H_2O at 20°C and pH 5.8. Experimental parameters were as follows: $t_1$max = 22.5 ms, $t_2$max = 205.0 ms, recycle delay = 1 s, 96 scans/t_1 increment, time-domain data points were 450 and 2048 along t_1 and t_2 dimensions, respectively. The ^{1}H-carrier frequency was set on the water resonance. The data were multiplied with sine bell window functions shifted by $\pi/4$ and $\pi/8$ along t_1 and t_2 axes, respectively and zero-filled to 1024 data points along t_1 dimension prior to stripped 2D-FT. This region shows the nOe connectivities from T (CH_3)/A (H2)/C ($4NH_2$)/A5 ($6NH_2$) protons to the imino protons (T (3NH) and G (1NH). In addition, inter imino-proton nOes are also shown. The numbering of digonal peaks 1, 2 and 3 is consistent with the numbering in Fig. 2.

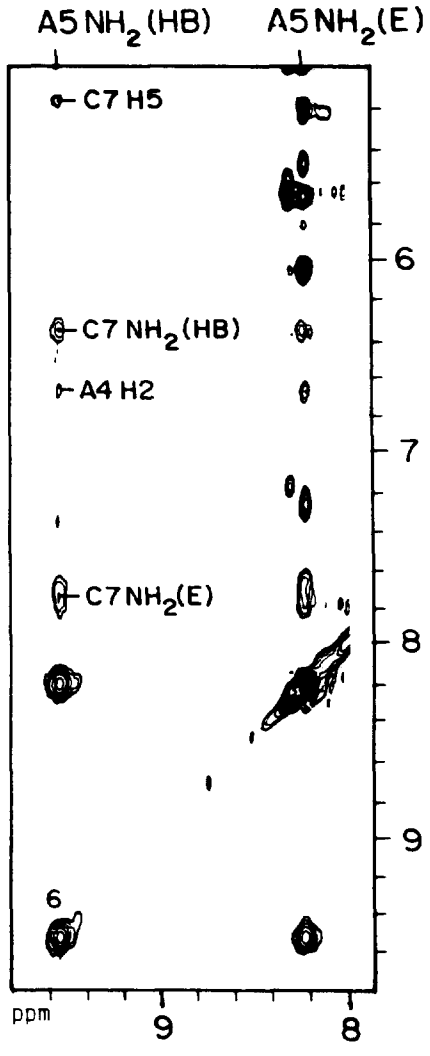

Fig. 5. Selected region of pure-absorption NOESY spectrum of AG58 in a mixed solvent of 90% H_2O + 10% ^2H_2O at 20°C and pH 5.8. Experimental parameters are same as Fig. 3. This region shows the intra amino proton nOe (within A5 ($6NH_2$) protons). In addition, interstrand nOe connectivities from C7 $4NH_2$ to A5 $6NH_2$ protons are also shown.

shown here) where the peak at 8.28 ppm vanishes. The two protons are unambiguously assigned to $6NH_2$ of A5 on the basis of the following observed nOes: (i) the protons at 9.6 ppm shows intrastrand sequential nOe to G6 (1NH); and (ii) both amino protons show nOes to the already assigned amino protons of C7 of the second strand (Fig. 5).

It is clear that in the A:G base pair, a hydrogen bond is formed involving one of the $6NH_2$ protons of A5. The fact that G8 (1NH) is not hydrogen bonded, rules out the possibility of A(*anti*)–G(*anti*) and A(*syn*)–G(*anti*) conformations for the A5G8 base pair. Further, G8 shows strong intranucleotide nOe between H8 and H1′ in comparison the intranucleotide nOe between H8 and H2′/H2″ protons are weak. The intranucleotide H8-H1′ nOe for all other nucleotides including A5, are weaker (Fig. 3). In order to substantiate this, a ROESY spectrum has been recorded with a lower mixing time (100 ms). Compared to NOESY, ROESY has an additional advantage of disinguishing nOe cross-peaks from exchange cross-peaks. In the ROESY spectrum (Fig. 6C), the G8 (H8)-G8 (H1′) cross-peak is as strong as the H5-H6 cross-peaks of cytosines. In contrast, self and sequential H8-H1′ peaks from all other nucleotides are either weak or absent (Fig. 6C). These observations establish that the G8 is in *syn* and all other nucleotides including A5 are in *anti* conformation. One observes a substantial downfield shift of the G8 (H2′) and a weak G8 (H8)-T9 (CH_3) sequential nOe, characteristic of the syn conformation.

The influence of pH on the stability of the AG58 duplex is attributed to protonation at N1 of A5, providing A with two adjacent H bond donors for pairing with the N7 and O6 acceptor positions of G8. The displacement of the pK of A from its monomer value of 3.7 [20] can be explained by the general principles outlined by Record [21]. We estimate a pK of about 6.1 for the duplex from UV titration. The observed shift of 2.4 pK units is a measure of stability and is consistent with values often observed in polynucleotide chemistry. We looked for the N1 resonance of $A5^+$ and observed a broad resonance at 16 ppm that could be assigned to this proton. The broad features

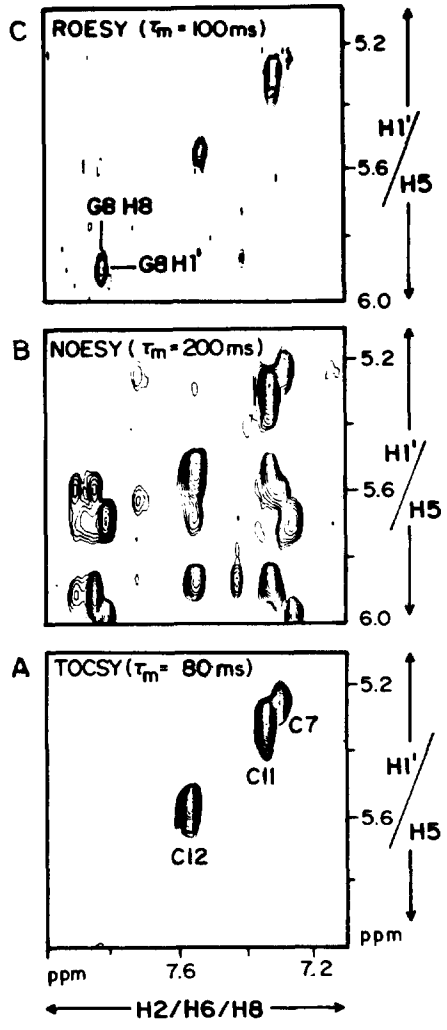

Fig. 6. Selected regions of (A) TOCSY, (B) Pure absorption NOESY and (C) ROESY spectra showing H1′/H5-H2/H6/H8 connectivities. All these spectra were recorded under identical experimental conditions (legend of Fig. 3), except the mixing times which are shown in the upper left corner of each region.

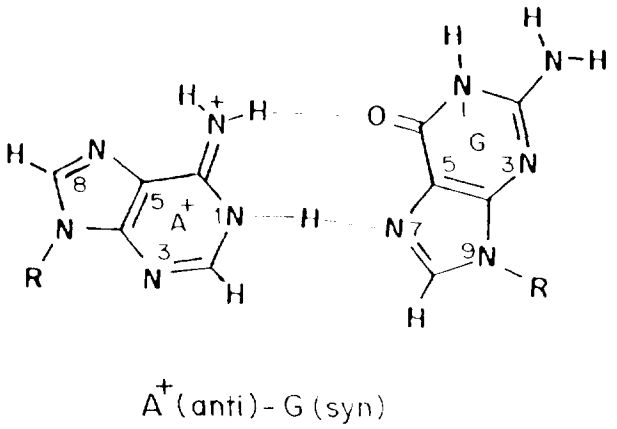

Fig. 7. Schematic diagram for A$^+$(*anti*)·G(*syn*) base pairing.

indicate a slow exchange rate of A5 (1NH) (Fig. 7), as also suggested by Carbonnaux et al. [22]. Interestingly, protonation of A5 does not influence the chemical shift of A5 (H8), in contrast to the observation of Gao and Patel [3]. In their work assignment of the A(*anti*)·G(*syn*) conformation was based on the nOe between A (H2) and G (H8). They could not detect a resonance corresponding to the second hydrogen bond for A$^+$(*anti*)·G(*syn*) base pair, and were unable to differentiate between the amino protons of G and A. The nOes A5 (6NH_2)-C7 (4NH_2) diagonally across the strands and the intrastrand sequential A5 (6NH_2)-G6 (1NH) nOe further support A$^+$(*anti*)·G(*syn*) conformation in AG58, and indicate that G6 is stacked over A5. For the A$^+$(*anti*)·G(*syn*) conformation the 2-amino group of G8 lies in the major groove and is not involved in the inter-base hydrogen bonding. This results in the rapid exchange of these protons with the solvent.

Thus, this work shows that under the conditions of our studies, AG58 exists as a stable B-DNA duplex with the A:G mismatch base pairs in A$^+$(*anti*)·G(*syn*) conformation. Both hydrogen donors in the mismatched inter-base hydrogen bonding, are situated on the same base A5 (A5 (1N) and A5 (6N)).

Acknowledgements: The facilities provided by the High Field FT NMR National Facility supported by the Department of Science and Technology, Department of Biotechnology, Council of Scientific and Industrial Research and Tata Institute of Fundamental Research, Bombay are gratefully acknowledged.

References

[1] Fersht, A.R., Knill-Jones, J.W. and Tsui, W.C. (1982) J. Mol. Biol. 156, 37–51.
[2] Brown, T., Leonard, G.A., Booth, E.D. and Chambers, J. (1989) J. Mol. Biol. 207, 455–457.
[3] Gao, Xiaolian and Patel, J. (1988) J. Am. Chem. Soc. 110, 5178–5182.
[4] Chary, K.V.R., Hosur, R.V., Govil, G., Tan, Z.K. and Miles, H.T. (1987) Biochemistry 26, 1315–1322.
[5] Howard, F.B., Chen, C., Ross, P.D. and Miles, H.T. (1991) Biochemistry 30, 779–782.
[6] Howard, F.B., Frazier, J., Ross, P.D. and Miles, H.T. (unpublished).
[7] Ross, P.D., Howard, F.B. and Lewis, M.S. (1991) Biochemistry 30, 6269–6275.
[8] Rastogi, V.K., Chary, K.V.R., Govil, G., Howard, F.B. and Miles, H.T. (1994) Appl. Magn. Reson. 4, 1–9.
[9] Chary, K.V.R., Rastogi, V.K., Govil, G., Howard, F.B. and Miles, H.T. (1994) Proc. Indian Acad Sci. (Chem. Sci.) 106, 1491–1504.
[10] Muraoka, M., Miles, H.T. and Howard, F.B., (1980) Biochemistry 19, 2429–2439.
[11] Hore, P.J. (1983) J. Magn. Reson. 55, 283–300.
[12] Anil Kumar, Wagner, G., Ernst, R.R. and Wuthrich, K. (1980) Biochem. Biophys. Res. Commun. 96, 1156–1163.
[13] Griesinger, C., Otting, G., Wuthrich, K. and Ernst, R.R. (1988) J. Am. Chem. Soc. 110, 7870–7872.
[14] Kessler, H., Griesinger, C., Kerssebaum, R., Wagner, K. and Ernst, R.R. (1987) J. Am. Chem. Soc. 109, 607–608.
[15] Feigon, J., Leupin, W., Denny, W.A. and Kearns, D.R. (1983) Biochemistry 22, 5943–5950.
[16] Reid, B.R. (1987) Q. Rev. Biophys. 20, 1–34.
[17] Scheek, R.M., Boelens, R., Russo, N., Van Boom, J.H. and Kaptein, R. (1984) Biochemistry 23, 1371–1376.
[18] Van de ven, F.J.M. and Hilbers, C.W. (1988) Eur. J. Biochem. 178, 1–38.
[19] Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, pp. 165–170, Wiley.
[20] Clauwaert, J. and Stockx, J. (1968) Natureforsch. B. 23, 25–30.
[21] Record, M.T. (1967) Bioploymers 5, 993–1008.
[22] Carbonnaux, C., Van dev Marel, G.A., Van Boon, J.H., Guschlbauer, W. and Fazakerley, G.V. (1991) Biochemistry 30, 5449–5458.