Snyder, M., Buchman, A. R., & Davis, R. W. (1986) *Nature 324*, 87-89.
Stenzel, T. T., Patel, P., & Bastia, D. (1987) *Cell 49*, 709-717.
Tang, R. S., & Draper, D. E. (1990) *Biochemistry 29*, 5232-5237.
Trifonov, E. N. (1980) *Nucleic Acids Res. 8*, 4041-4053.
Trifonov, E. N., & Sussman, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3816-3820.
Ulanovsky, L., Bodner, M., Trifonov, E. N., & Choder, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 862-866.
Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., & Crothers, D. M. (1990) *Science 249*, 1281-1285.
Wells, R. D. (1988) *J. Biol. Chem. 263*, 1095-1098.
Wu, H.-M. & Crothers, D. M. (1984) *Nature 308*, 509-513.
Wu, H.-N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221-8227.
Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332-6336.
Zahn, K., & Blattner, F. R. (1985a) *EMBO J. 4*, 3605-3616.
Zahn, K., & Blattner, F. R. (1985b) *Nature 317*, 451-453.

Solution Structure of an Oncogenic DNA Duplex Containing a G·A Mismatch

Claire Carbonnaux,[‡] Gijs A. van der Marel,[§] Jacques H. van Boom,[§] Wilhelm Guschlbauer,[‡] and G. Victor Fazakerley*,[‡]

Service de Biochimie et de Génétique Moléculaire, Bâtiment 142, Département de Biologie Cellulaire et Moléculaire, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-sur-Yvette, France, and Gorlaeus Laboratoria, University of Leiden, 2300A Leiden, The Netherlands

Received November 13, 1990; Revised Manuscript Received March 8, 1991

ABSTRACT: The DNA duplex 5'-d(GCCACAAGCTC)·d(GAGCTGGTGGC), which contains a central G·A mismatch has been studied by one and two-dimensional NMR techniques. The duplex corresponds to the sequence 29-39 of the K-*ras* gene. The mismatch position is that of the first base of the Gly12 codon, a hot spot for mutations. The observed NOEs of the nonexchangeable protons show that both of the bases of the mismatched pair are intrahelical over a wide range of pH. However, the structure of the G·A mispair and the conformation of the central part of the duplex change with pH. This structural change shows a p*K* of 6.0. At low pH, the G·A bases are base paired with hydrogen bonds between the keto group of the G residue and the amino group of the A residue and, secondly, between the N7 of the G and a proton on N1 of A. This causes the G residue to adopt a syn conformation. On raising the pH, the N1-H proton of the protonated A residue is removed, and the base pair rearranges. In the neutral G·A base pair both residues adopt an anti conformation, and the mismatch is stabilized by hydrogen bonds. Our results on the exchangeable and A(H2) protons of the mismatched pair indicate a shift from a classical face-to-face two hydrogen-bonded structure to a slipped structure stabilized by bifurcated hydrogen bonds. This may be a particular characteristic of this oncogenic sequence in which the G·A error is poorly repaired.

The presence of enzymatic systems specialized in the repair of noncomplementary DNA base pairs (mismatches) is now well established. In *Escherichia coli*, three different pathways have been described. First, the postreplicative mismatch repair is dependent on the *mutS*, *mutL*, *mutH*, *mutU*, and *ssb* genes and acts shortly after DNA synthesis to eliminate the incorrect base (Radman & Wagner, 1986). Detection of the mismatch by the MutS and MutL proteins is followed by a series of events resulting in excision of the newly synthesized strand at nonmethylated GATC sites (Lu et al., 1984; Längle-Rouault et al., 1987). The repair efficiency depends on the type of mismatch, on the surrounding sequence (Kramer et al., 1984; Dohet et al., 1985), and correlates with the affinity of the MutS protein for the mismatch (Su et al., 1988). The second pathway characterized by excision of short sequences (very short patch repair) is independent of the *mutH* and *mutU* genes but requires MutS and MutL proteins (Lieb, 1987). This system specifically corrects G·T to G·C in the sequence TAGG, thus diminishing the mutagenic effect of 5-methylcytosine deamination to thymine (Jones et al., 1987). The third pathway requires the *mutY* gene and specifically corrects A·G to C·G (Au et al., 1988; Su et al., 1988).

Recently, mismatch-specific enzymatic activities have also been found in frog eggs (Brooks et al., 1989) and in human cells (Brown & Jiricny, 1988). In other studies, the mismatch repair specificity in rat liver extracts was tested on synthetic substrates mimicing replication errors that would lead to oncogenic mutations in the K-*ras* codon 12 sequence; the specificity thus observed resembles that of mismatch repair in *E. coli* (Dimitrijevic et al., private communication).

To better understand the mechanism of mismatch recognition by this repair system, we have studied by NMR the G·A-containing heteroduplex corresponding to the sequence of the K-*ras* gene in which the repair system was tested. This G·A-containing heteroduplex leads to a GGT → TGT mutation in the K-*ras* Gly12 codon found in colon carcinomas (Nishimura & Sekiya, 1987, and references therein). The recognition and repair of the G·A mismatch in this sequence may play an important role in the carcinogenesis.

Materials and Methods

The oligonucleotides corresponding to the sequence 29-39 of the K-*ras* gene mutated noncoding strand (Cys12, strand

* Correspondence should be addressed to G. Victor Fazakerley.
‡ Département de Biologie Cellulaire et Moléculaire, Centre d'Etudes Nucléaires de Saclay.
§ Gorlaeus Laboratoria, University of Leiden.

0006-2960/91/0430-5449$02.50/0

Chart I

5′-d(G1	C2	C3	A4	C5	**A6**	A7	G8	C9	T10	C11)-3′	strand 1
3′-d(C22	G21	G20	T19	G18	**G17**	T16	C15	G14	A13	G12)-5′	strand 2

1) or wild-type coding strand (Gly12, strand 2) were synthesized by a classical phosphotriester method (van der Marel et al., 1981; Marugg et al., 1984). The pair of oligonucleotides was heated to 80 °C followed by slow cooling to form the G·A duplex as shown in Chart I. The duplex was 4 mM in strand concentration dissolved in 10 mM phosphate buffer, 150 mM NaCl, and 0.2 mM EDTA. NMR spectra were recorded on either WM500 or AM600 Bruker spectrometers in either 99.99% D_2O or 90% H_2O/10% D_2O. Chemical shifts were measured relative to the internal reference tetramethylammonium chloride (3.18 ppm).

NOESY[1] spectra were recorded with mixing times of either 60 or 400 ms for spectra in D_2O and 250 ms for spectra in H_2O in the phase-sensitive mode (Bodenhausen et al., 1984). The time domain data sets consisted of 1024 complex points in the t_2 dimension and 256 increments in the t_1 dimension. After zero filling, a slightly shifted (10°–15°) sine bell function (but shifted by $\pi/2$ for cross-peak volume determination) was applied to the data prior to Fourier transformation. For spectra recorded in H_2O, the observation pulse was replaced by a jump and return sequence (Plateau & Guéron, 1982), and the pulse maximum was placed at 15 ppm. This gave high-quality spectra with virtually no base-plane distortion.

TOCSY spectra were recorded in the phase-sensitive mode (Davis & Bax, 1985) with a mixing time of 60 ms.

Results

We have recorded a series of 1D spectra at different pH values to determine the best experimental conditions. These spectra clearly revealed two pH-dependent structures for the G·A duplex. To follow the pH transition, we have measured the chemical shifts of A7(H8) and T16(H6) base protons as function of pH (not shown). The pK of the transition is found to be 6.00. We therefore chose to study the structures of the G·A duplex at pH 4.7 and 8.5 close to the ends of the transition.

The G·A Duplex at pH 4.7, Spectra in D_2O, and Assignment of the Nonexchangeable Proton Resonances. This was obtained from analysis of a NOESY spectrum recorded with a 400-ms mixing time and TOCSY spectra in D_2O. The strategy for assignment has been described in detail (Hare et al., 1983; Fréchet et al., 1983; Feigon et al., 1983; Scheek et al., 1984).

The H8/H6–H1′/H5 region of the NOESY spectrum recorded at 30 °C, pH 4.7, is shown in Figure 1. Seven strong cross-peaks (marked X) corresponding to the H6–H5 interactions are observed and identify the seven cytosines. The H8 resonances of the 5′-terminal residues (G1 on the first strand and G12 on the second one) are characterized by only one cross-peak, that with their own H1′. Starting from G1(H6) G1(H8) 7.92 ppm, we can follow the classical chain of connectivities for a right-handed B-DNA helix through A6 of the mismatch pair to C11(H6). From G12, we can follow the connectivities to the T16(H1′) intraresidue cross-peak, but here the chain is broken. Similarly, from the other end of the strand, C22, we can follow the chain to the G17(H1′) intraresidue cross-peak. Even at lower levels, we do not observe

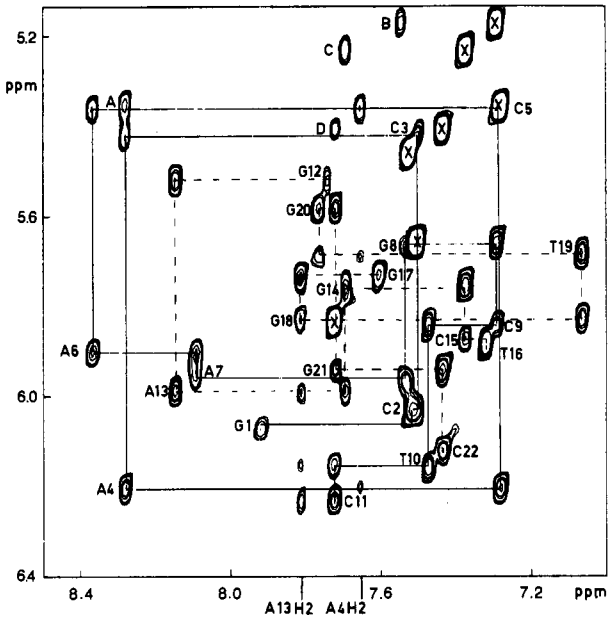


FIGURE 1: Expanded contour plot of the H6/H8–H1′/H5 region of the NOESY spectrum (400-ms mixing time) of the G·A duplex in D_2O, at 30 °C and pH 4.7. This region shows the connectivities between the base protons and the H1′/H5 protons for strand 1 (solid line) and strand 2 (broken line). Cross-peaks marked with an X correspond to C(H6)–C(H5) interactions. Peaks A–D are described in the text.

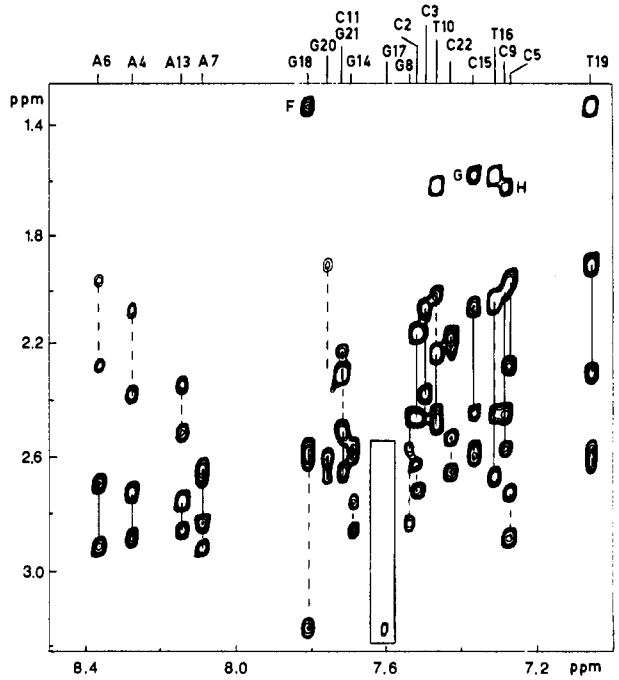


FIGURE 2: Expanded contour plot of the H6/H8–H2′/H2″/CH_3 region of the NOESY spectrum (400-ms mixing time) of the G·A duplex in D_2O, at 30 °C and pH 4.7. This region shows intraresidue (solid line) and interresidue (broken line) interactions between the base protons and the H2′/H2″ protons. The insert is plotted with 2-fold lower contour levels. Peaks F–H are described in the text.

a T16(H6)–G17(H1′) cross-peak. Four interbase cross-peaks (A–D in Figure 1) corresponding to the H8/H6–H5 interactions at the A4–C5, G8–C9, G14–C15, and G21–C22 steps are observed. The A13(H2) resonance at 7.82 ppm exhibits cross-peaks with its own H1′, the T10(H1′) on the partner

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2D nuclear Overhauser effect enhancement spectroscopy; TOCSY, total correlation spectroscopy.

Table I: Chemical Shifts of Nonexchangeable Protons at 30 °C and of Exchangeable Protons at 1 °C for the G·A Duplex at pH 4.7[a]

	H8/H6	H5/H2 CH_3	H1′	H2′	H2″	H3′	H4′	NH	NH2
G1	7.92		6.08	2.62	2.72	n.o.	n.o.	n.o.	
C2	7.51	5.40	6.02	2.17	2.46	4.84	4.24		
C3	7.51	5.64	5.43	2.06	2.38	4.83	4.07		8.47, 6.80
A4	8.28	7.66	6.21	2.73	2.89	5.04	4.42		7.64, 5.65
C5	7.30	5.35	5.37	1.97	2.27	4.80	4.12		8.17, 6.73
A6	8.37	7.56	5.91	2.70	2.92	5.06	4.37	16.00	10.03, 8.24
A7	8.10	7.46	5.97	2.64	2.83	5.06	4.39		7.65, 6.25
G8	7.56		5.67	2.45	2.57	4.91	4.19	12.76	
C9	7.30	5.15	5.85	2.00	2.45	4.92	4.19		8.00, 6.50
T10	7.46	1.62	6.16	2.23	2.47	4.85	4.19	13.97	
C11	7.66	5.70	5.84	2.29	2.30	4.54	4.01		
G12	7.76		5.53	2.33	2.50	4.77	4.14	12.51	
A13	8.14	7.82	5.99	2.75	2.86	5.04	4.39		7.64, 6.20
G14	7.70		5.76	2.57	2.60	4.98	4.38	12.79	
C15	7.38	5.22	5.88	2.06	2.44	4.82	4.23		7.95, 6.50
T16	7.33	1.60	5.91	2.03	2.67	4.85	4.07	13.55	
G17	7.60		5.74	3.22	2.60	4.94	4.27	11.00	
G18	7.84		5.84	2.57	2.61	4.88	4.32	12.51	
T19	7.08	1.35	5.69	1.91	2.30	4.99	4.17	13.70	
G20	7.78		5.59	2.60	2.67	4.95	4.15	12.90	
G21	7.72		5.95	2.51	2.66	4.94	4.35	n.o.	
C22	7.36	5.26	6.13	2.16	2.22	4.48	4.03		

[a] n.o., not observed.

strand, and the adjacent C11(H1′). Similarly, the A4(H2) at 7.66 ppm gives a weak cross-peak with its own H1′, the T19H1′ on the partner strand, and a stronger one with the adjacent C5H1′. These interbase interactions are characteristic of right-handed B-DNA. At this stage, we are not able to identify the A6- and A7(H2) resonances.

The H8/H6–H2′/H2″ region of the NOESY spectrum, Figure 2, confirms the assignment previously shown. As in Figure 1, we observe normal intra- and interresidue interactions for the first strand through A6. For the second strand, the G18(H8) gives intraresidue cross-peaks to its H2′ and H2″ resonances at 2.57 and 2.61 ppm and to an unusually downfield-shifted resonance at 3.22 ppm. This same proton is also observed, but very weakly, at the chemical shift of G17(H8). Examination of the matrix column shows another cross-peak at 2.62 ppm but no cross-peaks with the T16 residue. Three interbase cross-peaks, F–H in Figure 2, corresponding to the H8/H6–CH_3 interactions and characteristic of right-handed B-DNA are seen. The intraresidue cross-peaks for the two 5′-terminal G residues are very weak, presumably due to fraying. Interresidue cross-peaks from A13 and G12 are seen, and the assignment is confirmed in the H1′ to H2′/H2″ region (not shown).

In order to determine the stacking of the central base pairs, we have examined the aromatic region of the NOESY spectrum (not shown). We observe a cross-peak arising from the A6(H8)–C5(H6) interaction, which shows the stacking of A6 inside the helix. By contrast, we do not observe any cross-peak from G17(H8), suggesting a different orientation of the G17 base in comparison with a guanosine stacked inside a classical B-DNA.

In the normal B-DNA structure, all the glycosidic torsion angles are anti. In this case, the interproton distance between the base H8/H6 and its own H1′ proton is much longer than the interproton distance between H5 and H6 protons of cytidines (3.7 and 2.3 Å, respectively). In the case of a syn configuration, the two distances are approximately equal. To investigate the configuration of the glycosidic bond of the G17 and A6 residues, we have compared the cross-peak intensities of H5–H6 relative to H8/H6–H1′ in a NOESY spectrum recorded with a 60-ms mixing time to avoid spin diffusion (not shown). With one exception, all the H8/H6–H1′ intraresidue cross-peaks, including that for A6, are very weak compared to the C(H6)–C(H5) interactions, consistent with an anti orientation about the glycosidic bond. The exception is that of the G17 residue. The intraresidue G17(H8)–(H1′) cross-peak has a volume of ca. 50% of that of the average of the nonterminal C(H6)–C(H5) cross-peaks. This corresponds to a distance of ca. 2.5 Å and a conformation in the syn range. Examination of the H8/H6–H2′/H2″ region of this NOESY spectrum shows, for the first strand, only one unusual interaction, that between A6(H8) and C5(H2″), which is significantly weak. At this stage, our results are only qualitative; however, this weak interresidue interaction could result from a lateral displacement of A6 into the major groove or a greater propeller twist for A6. For the second strand, we do not observe a G17(H8)–(H2′) intraresidue cross-peak, which was weak even at 400 ms. This confirms the syn orientation about the glycosidic bond.

The relative assignment of the H2′ and H2″ resonances was obtained from analysis of the H1′–H2′/H2″ region of the NOESY spectrum recorded with a 60-ms mixing time. The complete assignment of the H3′ and H4′ was obtained from analysis of the TOCSY spectrum (not shown). The observed chemical shifts are given in Table I. The downfield-shifted resonance observed at 3.22 ppm corresponds to G17(H2)′, which shows an inversion of the usual relative chemical shift of H2′ and H2″ resonances.

Spectra in H_2O. Assignment of the Exchangeable Proton Resonances. The 1D spectrum of the imino, amino, and aromatic region of the duplex recorded at 1 °C and pH 4.7 is shown in Figure 3A. Usually T imino protons resonate between 13.5 and 14.5 ppm, and we observe three resonances in this region. The G imino resonances are normally found between 12.5 and 13.5 ppm, and in a resolution-enhanced spectrum six resonances, one of which probably corresponds to two protons, are seen. The remaining three resonances at 16.00, 11.00, and 10.03 ppm may arise from the mismatched pair.

The assignment of the exchangeable proton resonances was obtained from analysis of the NOESY spectrum recorded with a mixing time of 250 ms. Two regions are shown in Figure 4. In this spectrum, the resonances at 16 and 11 ppm and also the resonances of the terminal G residues are strongly attenuated by exchange with the solvent.

The lower part shows the imino–imino interactions. The

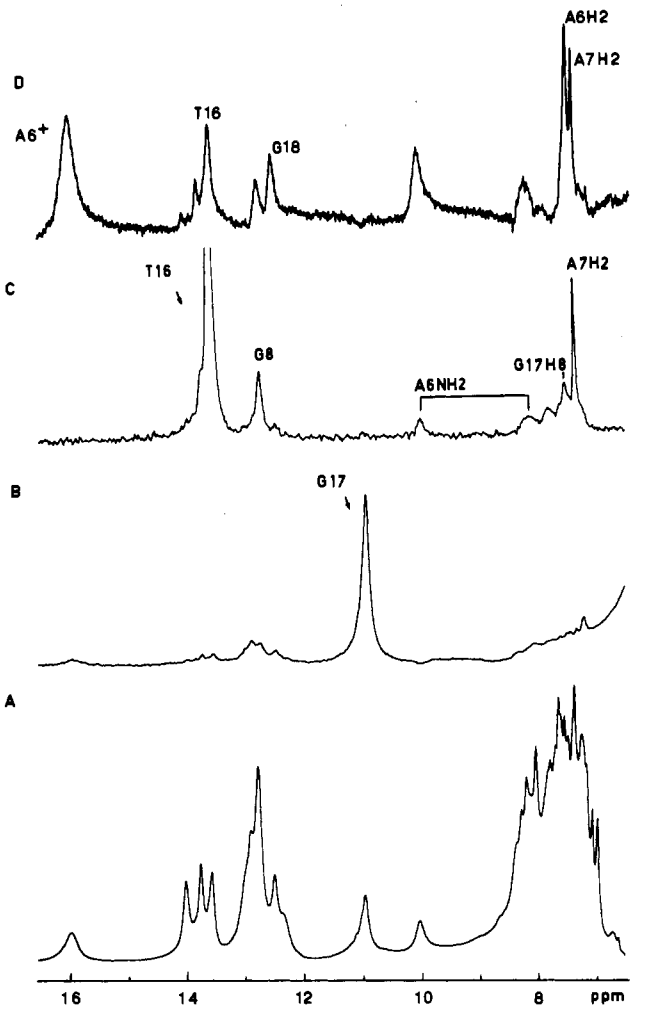


FIGURE 3: (A) One-dimensional spectrum (6.5–16.5 ppm) in H_2O, at 1 °C and pH 4.7. (B–D) Difference spectra after presaturation for 0.25 s of the resonances at 11.00, 13.55, and 16.00 ppm, respectively.

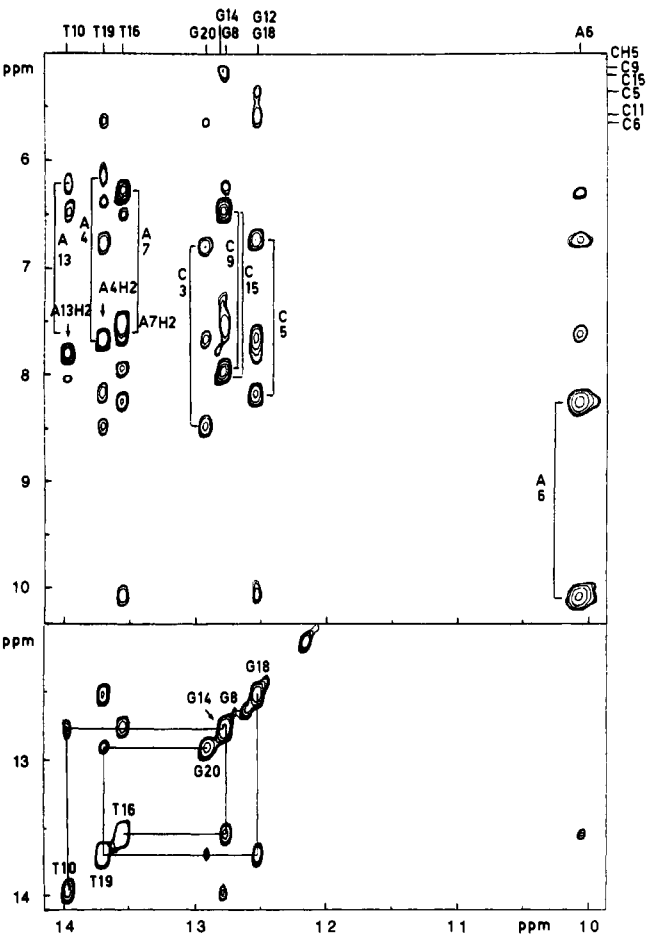


FIGURE 4: Expanded contour plots of the NOESY spectrum (250-ms mixing time) of the G·A duplex in H_2O, at 1 °C and pH 4.7. (Lower part) The region showing the interactions between the imino protons. (Upper part) The region showing the interactions between the imino protons and the amino/H2 protons. The two protons of the amino group of cytosine (C) or adenosine (A) are connected by a solid line. Unlabeled cross-peaks arise in the upper part from the spin diffusion to C(H5) protons.

resonance at 13.70 ppm shows interactions with two G imino protons. In the upper part of Figure 4, this imino proton shows a strong cross-peak with an A(H2) resonance at 7.66 ppm, which has already been identified as A4(H2). The imino resonance must correspond to T19. Similarly the T imino resonance at 13.97 ppm can be assigned to T10 from the NOE with A13(H2). The cross-peak observed with a G imino resonance at 12.79 ppm must be that with G14 as none would be expected with a terminal G residue. The remaining T imino resonance is that of T16 as shown by the strong cross-peak to A7(H2); it also gives a cross-peak with the G8 imino proton.

The pairs of amino resonances (C and A) shown in Figure 4 are confirmed by examination of the amino–amino region (not shown). The resonance at 10.03 ppm gives a strong cross-peak with an exchangeable proton resonance at 8.24 ppm, showing that these two protons are those of an amino group. It also shows a weak cross-peak with the T16 imino proton, demonstrating that the amino resonances belong to the mismatch site. Cross-peaks are observed to the A7 amino protons and one of the C5 amino protons. The other overlaps with the peak at 8.24 ppm. Having identified two amino protons of the mismatch, we note that cross-peaks with these protons are observed with the T16 imino proton and also one of the G residues, which in turn shows a cross-peak with T19. This identifies G18 and thus G20. We are unable to identify G21. In the upper part of Figure 4, we observe cross-peaks to C(H5) protons via spin diffusion, confirming the assignment, but no cross-peak is observed with C2(H5).

We have recorded 1D difference spectra to further probe the mismatch structure. Presaturation of the peak at 11.00 ppm (Figure 3B) does not result in any significant NOEs. This exchangeable proton must lie on the exterior of the helix. We have also carried out a 1D experiment with presaturation of the T16 imino proton (Figure 3C) in order to better resolve the overlapping cross-peaks in the range of 7.5–7.7 ppm observed in the NOESY spectrum. In the 1D difference spectra, we clearly see an NOE to the G17(H8) resonance, which in the NOESY spectrum overlaps with that of A7(H2). Lastly, presaturation of the downfield-shifted proton at 16.00 ppm (Figure 3D) results in NOEs to the imino protons on either side of the mismatch (also, via spin diffusion, to the imino protons of the base pairs farther out in the helix), to the amino protons of A6, and another to A7(H2). Additionally, we observe an NOE to a nonexchangeable proton at 7.56 ppm. All the nonexchangeable base proton resonances, except that of A6(H2), have already been assigned. We have found G8(H8) at 7.56 ppm, but such an NOE from the mismatch pair would be impossible. The NOE must be that with A6-(H2). We note that there is a shoulder on the low-field side of this peak at a chemical shift corresponding to that of G17(H8). Referring back to Figure 3C, we note that if the T16 imino proton gives an NOE to the A6(H2) proton that this NOE must be weak.

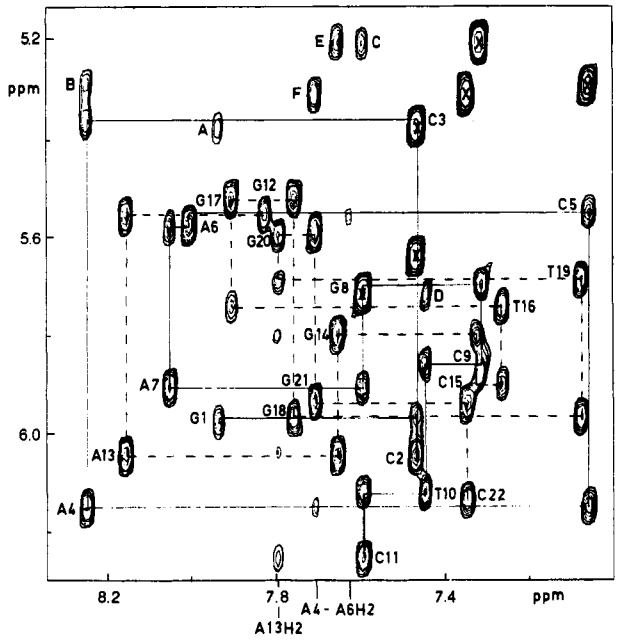


FIGURE 5: Expanded contour plot of the H6/H8–H1′/H5 region of the NOESY spectrum (400-ms mixing time) of the G·A duplex in D_2O, at 30 °C and pH 8.5. This region shows the connectivities between the base protons and the H1′/H5 protons for strand 1 (solid line) and strand 2 (broken line). Cross-peaks marked with an X correspond to C(H6)–C(H5) interactions. Peaks A–F are described in the text.

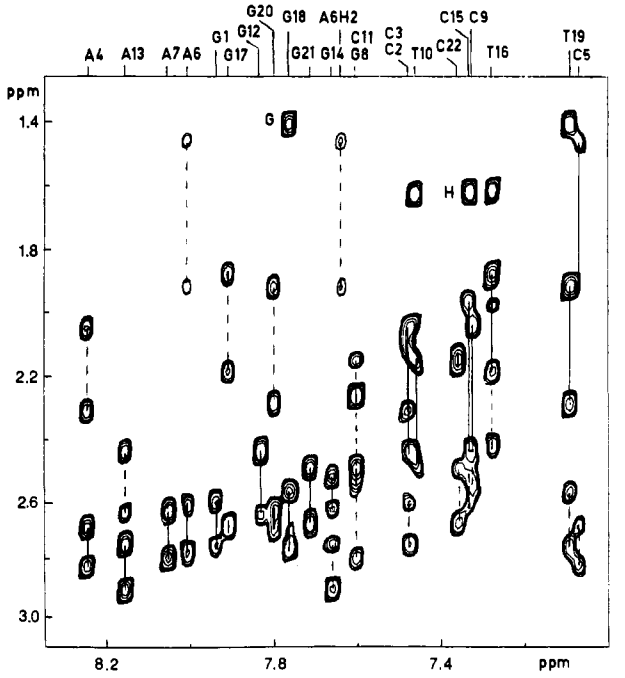


FIGURE 6: Expanded contour plot of the H6/H8–H2′/H2″ region of the NOESY spectrum (400-ms mixing time) of the G·A duplex in D_2O, at 30 °C and pH 8.5. This region shows intraresidue (solid line) and interresidue (broken line) interactions between the base protons and the H2′/H2″ protons. Peaks G and H are described in the text.

The imino resonance at 11.00 ppm corresponds to a chemical shift of a non-hydrogen-bonded imino proton as observed for loop structures (Haasnoot et al., 1980, 1983; Patel et al., 1985; Woodson & Crothers, 1988) or exposed imino protons in syn structures (Gao & Patel, 1988; Kouchakdjian et al., 1989). We have shown that G17 has a syn conformation and this imino resonance can be attributed to G17.

Amino resonances shifted downfield to ca. 10 ppm have recently been observed in triple-stranded protonated structures (de los Santos et al., 1989; Sklenăr & Feigon, 1990). Similarly, protonation would explain the downfield shift of the imino proton at 16.00 ppm. Thus the amino protons at 10.03 and 8.24 ppm and the imino proton at 16.00 ppm must belong to the A6 residue. Further evidence that A6 is protonated is found in the strong downfield shift of A6(H8) relative to the neutral structure (see below) as predicted for base protonation (Sowers et al., 1986). The exchangeable proton chemical shifts are given in Table I.

The G·A Duplex at Neutral and Basic pH. In order to characterize the second G·A duplex structure, the same analysis was carried out at pH 8.5 in D_2O.

Spectra in D_2O. Assignment of the Nonexchangeable Proton Resonances. The H8/H6–H1′/H5 region of the NOESY spectrum recorded at 30 °C, pH 8.5, is shown in Figure 5. Starting from G1(H8) at 7.94 ppm, the classical chain of connectivities for a right-handed B-DNA can be followed without ambiguity to C11(H6). Similarly, on the other strand the connectivities can be followed from G12(H8) at 7.83 ppm to C22(H6). Six interbase cross-peaks (A–F in Figure 5) corresponding to the H8/H6–H5 interactions for the G1–G2, A4–C5, G8–C9, T10–C11, G14–C15, and G21–C22 steps are observed. The A13(H2) at 7.80 ppm gives rise to a cross-peak with C11(H1′), a weak one with its own H1′, and also with that of G14. Similarly the A4(H2) at 7.70 ppm gives a weak cross-peak with its own H1′. All the observed interbase cross-peaks are characteristic of right-handed B-DNA. In addition, a resonance at 7.64 ppm gives rise to a weak cross-peak with C5(H1′). The A4(H2) proton could give rise to an NOE with C5(H1′), but it has already been assigned and this is confirmed below. The only other possible candidate for the resonance at 7.64 ppm is that of A6(H2). This is an interaction in the opposite sense to that normally observed, suggesting a significant conformational deformation relative to Watson–Crick pairing.

Figure 6 shows the H8/H6–H2′/H2″ interactions. The chain of connectivities can be followed on each strand, and the H8/H6 resonance assignment shown in Figure 5 is confirmed. The two cross-peaks G and H correspond to the interbase H8/H6–CH_3 interactions for the residues G18–T19 (G), C15–T16, and C9–T10 (H), which overlap and characterize a right-handed B-DNA structure. The A6/H8–H2′/H2″ interresidue interactions are very weak compared to other interresidue interactions. Moreover, two weak cross-peaks from the resonance at 7.64 ppm to C5(H2′)/(H2″) are observed. All the H8/H6 resonances have already been assigned, as have the H2 resonances of A13 and A4. Those of A6 and A7 remain to be unambiguously assigned. As A6·G17 stacks into the helix, an NOE from A7(H2) to C5(H2′)/(H2″) is impossible. The resonance at 7.64 ppm can only be assigned to that of A6(H2), which displays unusual interactions. These features must be due to the specific orientation of the A6 residue relative to the C5 sugar, which positions the H8 and the H2 aromatic protons of the adenosine at approximately equal distances from the C5(H2′)/(H2″).

From the H8/H6–H8/H6 region (not shown), we observe a cross-peak arising from the G17(H8)–T16(H6) interaction that confirms the stacking of G17 inside the helix. By contrast, we do not observe any cross-peak arising from the A6-(H8)–C5(H6) interaction. This may be due to the displacement of the A6 base relative to the C5 base.

The relative assignment of the H2′ and H2″ resonances was obtained from analysis of H1′–H2′/H2″ region of the NOESY spectrum recorded with a 60-ms mixing time. The complete assignment of H3′ and H4′ resonances was obtained from

Table II: Chemical Shifts of Nonexchangeable Protons at 30 °C, pH 8.5, and of Exchangeable Protons at 1 °C, pH 7.6, for the G·A Duplex[a]

	H8/H6	H5/H2 CH3	H1′	H2′	H2″	H3′	H4′	NH	NH2
G1	7.94		5.97	2.61	2.75	4.83	4.23	12.93	
C2	7.48	5.38	6.05	2.13	2.44	4.84	4.23		
C3	7.48	5.64	5.36	2.06	2.32	4.83	4.08		8.50, 6.80
A4	8.25	7.70	6.16	2.69	2.83	5.01	4.40		7.75, 6.25
C5	7.08	5.29	5.55	1.45	1.90	4.81	4.05		8.05, 6.60
A6	8.01	7.64	5.58	2.77	2.63	4.93	4.25		
A7	8.05	7.46	5.91	2.65	2.80	5.00	4.38		7.67, 6.20
G8	7.61		5.70	2.50	2.57	4.92	4.16	12.70	
C9	7.33	5.21	5.87	2.05	2.45	4.65	4.16		7.95, 6.48
T10	7.46	1.62	6.12	2.15	2.49	4.86	4.17	14.00	
C11	7.61	5.72	6.25	2.25	2.29	4.55	4.01		8.35, n.o.
G12	7.83		5.52	2.45	2.65	4.82	4.16	12.70	
A13	8.16	7.80	6.05	2.75	2.90	5.05	4.42		7.82, 6.05
G14	7.66		5.80	2.53	2.64	4.96	4.41	12.80	
C15	7.34	5.21	5.90	1.97	2.44	4.65	4.06		7.95, 6.50
T16	7.29	1.61	5.74	1.88	2.20	4.81	4.06	13.70	
G17	7.91		5.52	2.68	2.70	4.95	4.33	11.70	
G18	7.77		5.97	2.58	2.76	4.95	4.41	12.58	
T19	7.10	1.40	5.69	1.92	2.30	4.82	4.09	13.77	
G20	7.80		5.60	2.63	2.70	4.95	4.32	13.00	
G21	7.72		5.94	2.50	2.68	4.93	4.33	n.o.	
C22	7.36	5.31	6.13	2.15	2.17	4.47	4.02		8.15, n.o.

[a] n.o., not observed.

analysis of the TOCSY spectrum (not shown).

To determine the anti or syn configuration of the G17·A6 mismatched base pair, we have measured in the NOESY spectrum recorded with a 60-ms mixing time the ratio of the cross-peak volumes of H8–H1′ relative to those H5–H6. The ratios are small and very similar to those observed for all the other intraresidue NOEs of this type. These results show that both G17 and A6 adopt the anti configuration.

The observed chemical shifts are shown in Table II. The relative assignment of the A6(H2′)/(H2″) protons gives the H2′ upfield shifted relative to the H2″. We have measured, in the NOESY spectrum recorded with a 60-ms mixing time, the ratio of the cross-peak volumes for the intraresidue interactions H8/H6–H2′ and H8/H6–H3′ to determine the predominant sugar conformation (Cuniasse et al., 1989). The results show that the A6 and G17 sugars are predominantly in the C2′-endo conformation as are all other nonterminal residues.

Spectra in H_2O. The 1D spectrum of the duplex recorded at 1 °C and pH 8.5 is shown in Figure 7A. We observe only one resonance shifted out of the normal imino region, a peak at 11.70 ppm corresponding to one proton.

At pH 8.5, only one species is present in solution, but attempts to record NOESY spectra gave poor results due to exchange with solvent. We therefore lowered the pH to 7.0 where ca. 10% of the duplex exists in the protonated form as can be seen in 1D spectra. Although certain resonances are broadened by exchange, the NOESY spectra are of better quality. Cross-peaks arising from the protonated species are generally not observed due to their low intensity. Two regions are shown in Figure 8.

The lower part shows the imino–imino interactions. As previously, we can follow the chain of connectivities from T10 to T16 and from G18 to G20. The imino resonance assignment was confirmed by 1D spectra recorded as a function of pH starting from the well-defined spectrum in acidic solution. A comparison of Tables I and II shows that for the Watson–Crick base pairs the shifts are very small.

The upper part of Figure 8 shows the interactions between imino and amino, H2 or H5, protons. Three strong cross-peaks corresponding to T imino–A(H2) interactions are observed. The T16 imino proton clearly shows NOEs to the A7(H2), the A7 amino protons, and the adjacent C15 amino protons,

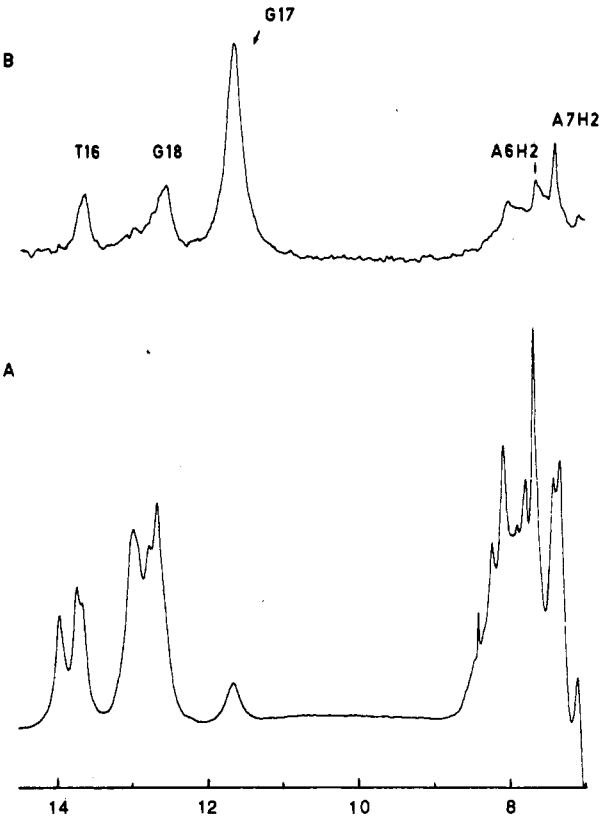


FIGURE 7: (A) One-dimensional spectrum (7–14.5 ppm) in H_2O, at 1 °C and pH 7.6. (B) Difference spectrum after presaturation for 0.25 s of the resonance at 11.70 ppm.

but no cross-peak is observed to the A6(H2) resonance. The G8 imino proton shows NOEs to the C15 amino protons and to the adjacent A7(H2). The remaining cross-peaks correspond to the interaction between A13(H2) and the G12 imino proton resonance that overlaps with the G8 imino proton and also one of the C11 amino protons at 8.35 ppm. This assignment can be confirmed in the lower part of Figure 8 by a weak NOE between T10 and G12 imino protons (cross-peak marked with an arrow).

At the solvent frequency, exchange cross-peaks are observed for some of the imino protons and for the proton at 11.70 ppm. In figure 7A this resonance is visibly much broader than those

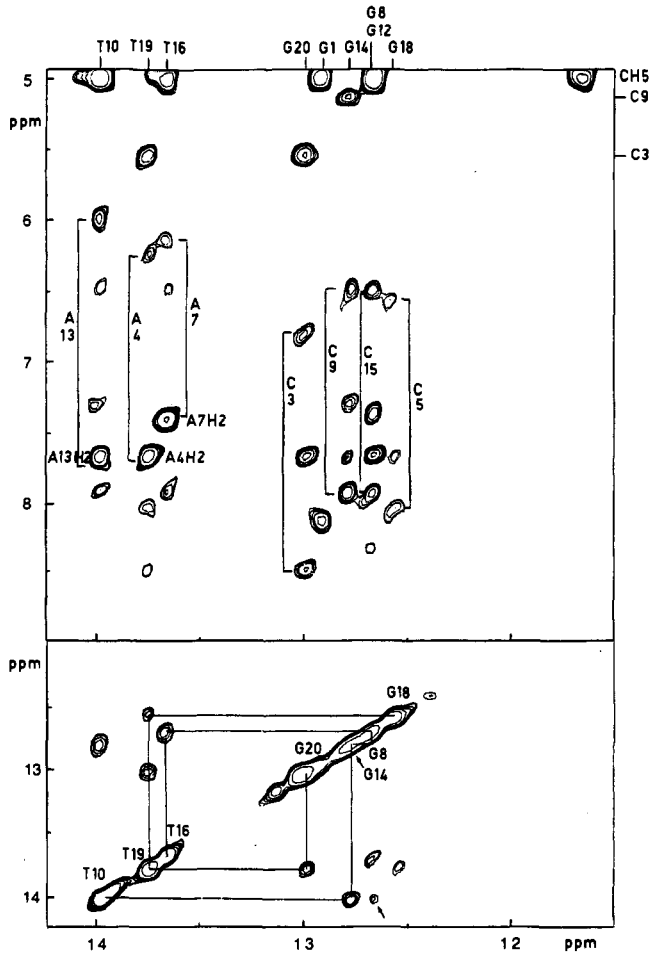


FIGURE 8: Expanded contour plots of the NOESY spectrum (250-ms mixing time) in H_2O, at 1 °C and pH 7.0. (Lower part) The region showing the interactions between the imino protons. The cross-peak marked with an arrow is described in the text. (Upper part) The region showing the interactions between the imino protons and the amino/H2 protons. The two protons of the amino group of cytosine (C) or adenosine (A) are connected by a solid line. Unlabeled cross-peaks in the upper part arise from the spin diffusion to C(H5) protons.

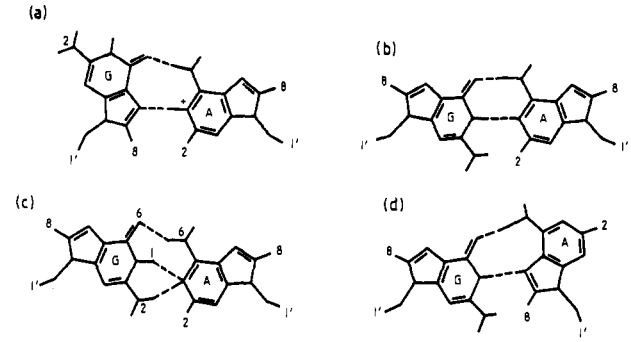
of the A·T and G·C imino protons. The difference spectrum following presaturation for 0.25 s is shown in Figure 7B. NOEs are observed to the imino protons of T16 and G18, showing that this proton belongs to the A6·G17 pair. No strong NOE is observed to an exchangeable proton to higher field, which shows that this resonance is an imino and not an amino resonance and must be that of G17. Two NOEs are observed to nonexchangeable protons, that to A7(H2) of the adjacent base pair and the intra-base-pair NOE to A6(H2), which is unusually weak.

Clearly both bases of the mismatch pair lie within the helix and, as shown above, adopt an anti–anti conformation. However, the chemical shift and line width of the G17 resonance are unusual. The large line width suggests greater solvent accessibility for this proton. Its chemical shift corresponds to one approximately midway between that of a normal G·C imino proton and that of a non-hydrogen-bonded or NH-carbonyl imino proton. The latter can be excluded in this case.

Discussion

Our results demonstrate that the G·A-containing oncogenic sequence exists in a pH-dependent equilibrium between two structures of globally B DNA, characterized by different configurations of the G17·A6 base pair. In the pH range of 4–5, the mismatch adopts a G(syn)·A^+(anti) structure in which

Chart II: Possible Pairing Structures for the G·A Duplex[a]



[a] (a) G(syn)·A^+(anti); (b) G(anti)·A(anti) with two hydrogen bonds; (c) G(anti)·A(anti) with bifurcated hydrogen bonds as proposed in the text; (d) G(anti)·A(syn).

the N1 of A6 is protonated. In previous reports of protonated mismatch structures (Sowers et al., 1986; Gao & Patel, 1988), the extra proton was not directly observed and probably exchanged rapidly with solvent. Here we do observe this proton at the unusual chemical shift of 16.00 ppm. There is no evidence of significant destabilization of the surrounding structure. Above pH 7, a neutral species in which the G·A mismatch adopts an anti–anti conformation is observed. The results clearly show that both configurations of the mismatched base pair are integrated in a B-DNA right-handed helix. Comparison of the chemical shifts between the two structures reveals important modifications for the three central base pairs (C5A6A7)·(T16G17G18), but no significant differences for the other residues, suggesting structural perturbations localized only in the central nucleotides.

G(syn)·A^+(anti) Pairing at Acidic pH. At acidic pH, the observed NOE between the G17(H8) and the T16 imino proton demonstrates that the G17 residue is stacked into the helix with its H8 in the minor groove. A syn configuration for G17 is required to account for this orientation. This is confirmed by the fast NOE buildup between G17(H8) and G17(H1′) and the extremely weak intraresdue NOEs with the H2′ and H2″ protons in the 400-ms NOESY spectrum and supported by the absence of NOEs to the neighboring T16 sugar protons. The cross-peak corresponding to the G17-(H8)–A6(H2) interaction that should be seen in the G-(syn)·A^+(anti) structure could not be detected because the resonances overlap.

The formation of a G(syn)·A^+(anti) structure results in an upfield shift of G17(H8) to 7.60 ppm and a downfield shift of G17(H2)′ to 3.22 ppm. These modifications seem to be characteristic of the G(syn) configuration (Patel et al., 1986; Gao & Patel, 1988; Kouchakdjian et al., 1989) and may reflect changes in the ring current contributions. On the partner strand, the observed structural perturbation at the C5–A6 step consists of a shift of the base A6 that increases the A6-(H8)–C5(H2′)(H2″) distances in order to better accommodate the purine–purine mismatch in the helix.

In the G(syn)·A^+(anti) pairing (Chart IIa), the hydrogen bond between N7 of guanosine and H-N1 of the protonated adenosine stabilizes the base pair, but this proton had not been observed in previous studies. For the first time, we detect the $A6^+$ H-N1 at 16.00 ppm, and the oberved NOEs confirm this assignment. This observation demonstrates a slow exchange rate on the NMR time scale, suggesting a greater stability of this hydrogen bond in the G·A-containing oncogenic sequence. The protonation of A6 results in a downfield shift of its own H8, as we have previously shown (Sowers et al., 1986) and of the amino protons. Similar effects have been reported by

Kouchakdjian et al. (1989) in the X(syn)·A^+(anti) configuration where X is a modified guanosine.

The pK_a estimated from the transition curve therefore represents the adenosine protonation. Its value, pK_a = 6.0, ca. 2 pH units higher than in the free base, demonstrates the influence of the surrounding bases and that of base pairing. It is comparable to that of poly(A) (Guschlbauer & Vetterl, 1969) but higher than that of poly (dA) (Ts'o et al., 1966).

The results reported here on the G·A mismatch in acidic solution are in agreement with a previously reported NMR study (Gao & Patel, 1988), but we have been able to make more proton assignments and thus study in more detail the environment of the mismatch site. The proposed structure is also in agreement with an X-ray study of a G·A duplex for which the crystals were grown at pH 6.6 (Brown et al., 1989).

G(anti)·A(anti) Pairing at Neutral and Basic pH. At pH 8.5, the intensity of the intraresidue NOEs between the H8 and the H1′ for G17 and A6 shows that the mismatched base pair is in the anti–anti configuration. In addition, the G17 imino proton gives rise to NOEs to A6(H2) and A7(H2), G18, and T16 imino protons, which indicates G(anti)·A(anti) pairing (Chart II). In such a configuration, the C1′–C1′ distance is greater than in Watson–Crick pairing (12.5 Å instead of 10.5 Å). This must create some perturbation of the local geometry. Analysis of the observed unusual interactions and chemical shifts reveals in fact significant structural perturbations at the mismatch site. We detect the most important modifications in the C5–A6 step, where we observed unusual weak interactions between A6(H8) and C5(H2′)/(H2″) and do not observe the A6(H8)–C5(H6) interaction. In addition, weak interactions between A6(H2) and C5(H1′)/(H2′)/(H2″) are detected. These results suggest an important longitudinal displacement of the A6 base. On the other strand, a strong G17(H8)–T16(H2′) interresidue interaction relative to G17-(H8)–T16(H2″) is observed (not shown) in the NOESY spectrum recorded with a 60-ms mixing time and suggests that the G17 is shifted toward the minor groove.

The chemical shift of G17 imino proton at 11.70 ppm suggests that the G(anti)·A(anti) structure in our sequence is different from that observed in previous studies. When the G·A mismatch was studied embedded between two Watson–Crick base pairs, the chemical shifts of the G imino resonance was found at ca. 13.6 ppm for the sequence GGG (Gao & Patel, 1988) and at 13.2 ppm for the sequence CGC (Patel et al., 1984) but at 12.82 ppm for the GGA sequence (Fazakerley et al., 1986). In terms of purine/pyrimidine content our sequence, TGG, is intermediate. It is difficult, however, to explain the G17 imino proton chemical shift in terms of the oligonucleotide sequence.

Imino resonances at 10–12 ppm have been observed either for NH···O=C pairing in wobble structures (Patel et al., 1982; Tibanyenda et al., 1984; Carbonnaux et al., 1990) or for imino protons non-hydrogen-bonded in loop and other structures (Haasnoot et al., 1980, 1983; Patel et al., 1985; Woodson & Crothers, 1988). Formation of a hydrogen bond between the G17 imino proton and a keto group is not possible in this sequence.

Base-paired and unpaired imino resonances have usually been described in the range of 12–14 and 10–11 ppm, respectively. Resonances between 11 and 12 ppm have been found in delocalized bulged guanosine in homopolymeric sequences (Woodson & Crothers, 1988). In that study, the chemical shift of an isolated G imino proton was observed at 10.5 ppm. For a frame-shift sequence with a central G3·C2, where the central G residue forms a G·C pair part of the time, it was found at 11.00 ppm; and as the sequence G_nC_{n-1} is lengthened, the resonances of the central base pairs move downfield as they spend a greater proportion of time hydrogen bonded. While our sequence cannot behave in this way, it is interesting to compare the percentage time spent non-hydrogen-bonded with the chemical shift observed by these authors. A chemical shift of 11.70 ppm observed for the G17 imino resonance would correspond to ca. 35% non-hydrogen-bonded.

We do not wish to overinterpret chemical shift data, but that of G17 is highly unusual. At pH 8.5, we know that A6 and G17 are stacked into the helix, and we observe only one form for the mismatch. From the above arguments, we could propose a rapid equilibrium between an intrahelical hydrogen-bonded and a non-hydrogen-bonded form. However, it is very difficult to incorporate a G·A pair inside the helix in which the respective donors and acceptors would be sufficiently far apart. Energetically, such a structure seems highly unlikely. The only explanation that we are able to propose consists of modifying the hydrogen-bonding pattern as shown in Chart IIc. Such bifurcated hydrogen bonds have never been observed between mismatched pairs by X-ray crystallography, although they have been between a G·A mismatched pair and an adjacent nucleotide (Privé et al., 1987). It should be noted that the NOEs observed in Figure 7B are in agreement with this structure. For adjacent A·T base pairs, the intraresidue imino–A(H2) distance is shorter than the interresidue imino–A(H2) distance; thus the former NOE is significantly larger than the latter. The same would be true for a G-(anti)·A(anti) base pair adjacent to an A·T base pair. In Figure 7B we observe the reverse situation, the intraresidue NOE, G17 imino–A6(H2), is smaller than that of G17 imino–A7(H2), suggesting that the distance corresponding to the former interaction is long. The proposed model in Chart IIc is in accord with the observation; the G17 imino–A6(H2) distance would be ca. 4 Å. This structure would also tend to push the A6(H2) farther out into the minor groove, and this is consistent with the absence of observed cross-peaks from the G18 and T16 imino protons with A6(H2). At this stage of the study, the base pairing proposed in Chart IIc is very tentative. We cannot exclude the normal hydrogen bonding as shown in Chart IIb. Few interactions are sensitive to the lateral displacement involved between the two structures.

The G·A mismatch in neutral or slightly basic solution has been studied many times and shows a certain polymorphicity. In one case when the mismatch is embedded in a highly A·T rich oligonucleotide, the DNA adopted a bulged out structure (Fazakerley et al., 1986). Comparison of NMR and X-ray results shows only one case of agreement (Kan et al., 1983; Nikonowicz & Gorenstein, 1990; Privé et al., 1988) where both showed an anti–anti structure, though the sequence is somewhat special as two G·A mismatches are adjacent. Otherwise solution studies have shown for isolated G·A mismatches an anti–anti structure (Patel et al., 1984; Gao & Patel, 1988, and this work), whereas X-ray studies show a G(anti)·A(syn) structure (Brown et al., 1986; Webster et al., 1990) (Chart IId). This is somewhat surprising and cannot be explained by sequence effects. Although the precise sequence was no doubt chosen for different reasons, the sequence that we have studied here contains only one modification in eight pairs relative to that of an X-ray study (Webster et al., 1990). In our sequence 5′-CYCAAGCT, the second residue is an A, whereas in an X-ray study it was a G. We cannot exclude the presence in solution of a G(anti)·A(syn) structure in fast equilibrium with the anti–anti structure, but (a) we observe no evidence for such an equilibrium, and (b) fast exchange

between anti and syn structures is not likely. Alternatively, the anti–syn structure is present but at a relative concentration that we are unable to detect. Theoretical studies (Chuprina & Poltev, 1983; Keepers et al., 1984) have shown that G(anti)·A(anti) and G(anti)·A(syn) structures display similar stability when incorporated in the interior of a helix. The different results obtained by NMR and two of the X-ray studies may be due to the different solvent conditions or preferred crystallization of the G(anti)·A(syn) structure.

The conclusions described here are only qualitative but provide the principle features of the base-pairing scheme. These results serve as a basis for a more detailed study involving distance determination and restrained molecular dynamics calculations, which are currently in progress.

Our preliminary results on the protooncogene sequence containing a central C6–G17 base pair with all the other 10 base pairs being the same as studied here show that this sequence adopts a normal B-type structure.

Biological Implications. The in vitro studies on the specificity for mismatch recognition in the K-*ras* gene show that the G·A mismatch is not well repaired by the eukaryotic repair system. (Dimitrijevic et al. private communication). This result is similar to that reported for the prokaryotic postreplicative repair system (Dohet et al., 1985; Su et al., 1988). These observations suggest that the oncogenic replication error G·A may escape the postreplicative correction in vivo and thus contribute to a high incidence of G·C → T·A mutations in the K-*ras* Gly12 codon among human tumors. Our structural studies demonstrate that the G·A mismatch in the K-*ras* Gly12 codon exists in a pH-dependent equilibrium between the G(anti)·A(anti) and G(syn)·A^+(anti) structures. In the cell, we can suppose that the neutral species will predominate since the p*K* is 6.0. Both structures are accommodated in the B-DNA helix with structural perturbations that are slight but sufficient to modify the recognition specificity of the binding protein. One feature that would influence the interaction with the protein could be the larger C1′–C1′ distance of the base pair G(anti)·A(anti), which significantly modifies the stacking of the adenosine and may create a distorsion in the sugar–phosphate backbone. Such a distorsion could prevent the diffusion of the protein on the mismatched DNA and thus decrease the affinity for binding. Another important factor that determines the specificity is the position of the functional groups as described generally for protein–DNA interactions (von Hippel & McGhee, 1972; Seeman et al., 1976). An interesting feature of all four structures illustrated in Chart II is the absence of a carbonyl group in the minor groove. This could determine the low affinity of the recognition protein by preventing specific hydrogen-bonding interactions with the mismatched DNA. Structural studies on other heteroduplexes on which the repair system was tested are in progress and will provide more information on the correlation between the recognition specificity and the structural characteristics of the mismatches.

References

Au, K. G., Cabrera, M., Miller, J. M., & Modrich, P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9163–9166.

Bodenhausen, G., Kogler, H., & Ernst, E. E. (1984) *J. Magn. Reson. 58*, 370–388.

Brooks, P., Dohet, C., Almouzni, G., Mêchali, M., & Radman M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4425–4429.

Brown, T. C., & Jiricny, J. (1988) *Cell 54*, 705–711.

Brown, T., Hunter, W. M., Kneale, G., & Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2402–2406.

Brown, T., Leonard, G. A., Booth, E. D., & Chambers, J. (1989) *J. Mol. Biol. 207*, 455–457.

Carbonnaux, C., Fazakerley, G. V., & Sowers, L. C. (1990) *Nucleic Acids Res. 18*, 4075–4081.

Chuprina, V. P., & Poltev, V. I. (1983) *Nucleic Acids Res. 11*, 5205–5222.

Cuniasse, P., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Le Bret, M., Guschlbauer, W., & Fazakerley, G. V. (1989) *Biochemistry 28*, 2018–2026.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.

de los Santos, C., Rosen, M., & Patel, D. (1989) *Biochemistry 28*, 7282–7289.

Dimitrijevic, B., Sunjevarici, I., Cerovic, G., Bozin, D., & Radman, M. (1991) *Cancer Res.* (submitted).

Dohet, C., Wagner, R., & Radman, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 503–505.

Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H., Jones, M., & Radman, M. (1986) *EMBO J. 5*, 3697–3703.

Feigon, J., Denny, W. A., Leupin, W., & Kearns, D. R. (1983) *Biochemistry 22*, 5930–5042.

Frêchet, D., Cheng, D. M., Kan, L. S., & Ts'o, P.O.P. (1984) *Biochemistry 22*, 1371–1376.

Gao, X., & Patel, D. J. (1988) *J. Am. Chem. Soc. 110*, 5178–5182.

Guschlbauer, W., & Vetterl, V. L. (1969) *FEBS Lett. 4*, 57–60.

Haasnoot, C. A. G., den Hartog, J. H. J. de Rooij, J. F. M., van Boom, J. H., & Altona, C. (1980) *Nucleic Acids Res. 8*, 169–181.

Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnedijk, T. J. J., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1983) *J. Biomol. Struct. Dyn. 1*, 115–129.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, D., & Reid, B. R. (1983) *J. Mol. Biol. 181*, 319–336.

Jones, M., Wagner, R., & Radman, M. (1987) *Cell 50*, 621–626.

Kan, L.-S., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.

Keepers, J. W., Schmidt, P., James, T. L., & Kollman, P. A. (1984) *Biopolymers 23*, 2901–2929.

Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. (1989) *Biochemistry 28*, 5647–5657.

Kramer, B., Kramer, W., & Fritz, H. J. (1984) *Cell 38*, 879–887.

Längle-Rouault, F., Maenhaut-Michel, G., & Radman, M. (1987) *EMBO J. 6*, 1121–1127.

Leonard, G. A., Booth, E. D., & Brown, T. (1990) *Nucleic Acids Res. 18*, 5617–5623.

Lieb, M. (1987) *J. Bacteriol. 169*, 118–125.

Lu, A. L., Welsh, K., Clark, S., Su, S.-S., & Modrich, P. (1984) *Cold Spring Harbor Symp. Quant. Biol. 49*, 589–593.

Marugg, J. E., Tromp, M., Ihurani, P., Hoyng, C. F., van der Marel, G. A., & van Boom, J. H. (1984) *Tetrahedron 40*, 73–78.

Nishimura, S., & Sekija, T. (1987) *Biochem. J. 243*, 313–327.

Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.

Patel, D. J., & Shapiro, L. (1986) *J. Mol. Biol. 188*, 677–692.

Patel, D. J., Kozlowski, S. A., Markly, L. A., Rice, J. A., Broka, C., Dallas, J., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 437–444.

Patel, D. J., Kozlowski, S. A., Ikuta, A., & Itakura, K. (1984) *Biochemistry 23*, 3207-3217.
Plateau, P., & Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310-7311.
Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498-504.
Radman, M., & Wagner, R. (1986) *Annu. Rev. Genet. 20*, 523-538.
Scheek, R. R., Boelens, R., Russo, N., van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371-1376.
Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804-808.
Sklenář, V., & Feigon, J. (1990) *Nature 345*, 836-838.
Sowers, L. C., Fazakerley, G. V., Kim H., Dalton L., & Goodman M. F. (1986) *Biochemistry 25*, 3983-3988.
Su, S.-S., Lahue, R. S., Au, K. G., & Modrich, P. (1988) *J. Biol. Chem. 263*, 6829-6835.
Tibanyenda, N., de Bruin, S. H., Haasnoot, C. A. G., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1984) *Eur. J. Biochem. 139*, 19-27.
Ts'o, P. O. P., Rappaport, S. A., & Bollum, F. J. (1966) *Biochemistry 5*, 4153-4160.
van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & van Boom, J. H. (1981) *Tetrahedron Lett. 22*, 3887-3890.
von Hippel, P. H., & McGhee, J. D. (1972) *Annu. Rev. Biochem. 41*, 231-236.
Webster, G. D., Sanderson, M. R., Skelly, J. V., Neidle, S., Swann, P. F., Li, B. F., & Tickle, I. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6693-6697.
Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 436-445.

Predicting the Three-Dimensional Folding of Transfer RNA with a Computer Modeling Protocol†

John M. Hubbard‡ and John E. Hearst*
Department of Chemistry, University of California, Berkeley, California 94720
Received July 10, 1990; Revised Manuscript Received March 11, 1991

ABSTRACT: We have developed a computer modeling protocol that can be used to predict the three-dimensional folding of a ribonucleic acid on the basis of limited amounts of secondary and tertiary data. This protocol extends the use of distance geometry beyond the domain of NMR data in which it is usually applied. The use of this algorithm to fold the molecule eliminates operator subjectivity and reproducibly predicts the overall dimensions and shape of the transfer RNA molecule. By use of a replacement pseudoatom set based on helical substructures, a series of transfer RNA foldings have been completed that utilize only the primary structure, the phylogenetically deduced secondary structure, and five long-range interactions that were determined without reference to the crystal structure. In a control set of foldings, all the interactions suspected to exist in 1969 have been included. In all cases, the modeling process consistently predicts the global arrangement of the helical domains and to a lesser extent the general path of the backbone of transfer RNA.

The ability of single-stranded RNA[1] to form intramolecular hydrogen bonds gives it a much greater conformational variability than double-stranded DNA. This versatility combined with the large number of nucleotides in most cellular RNAs presents us with a formidable problem as we attempt to probe the form/function relationships of RNA. The sequence (primary structure), Watson-Crick base-pairing pattern (secondary structure), and compact folded conformation (tertiary structure) of RNA have been the focus of extensive research. With the vast improvements in DNA sequencing technology, the determination of the primary sequences of RNAs from the analysis of genomic DNA has become routine. With use of the thermodynamics of base stacking (Tinoco et al., 1973), it is possible to evaluate the possible secondary structures that an RNA may form. Improved empirical parameters and computer programs now make it possible to produce RNA foldings that correspond to the native hydrogen-bonding patterns with some accuracy (Jaeger et al., 1989). But the phylogenetic determination of secondary structure by comparison of the sequences of RNA molecules with common functions from different species is still the most productive technique. Phylogeny may also indicate that some bases are involved in tertiary or noncanonical base pairing (Gutell & Woese, 1990). Folded RNA molecules can also be probed for three-dimensional relationships with a variety of chemical and enzymatic techniques. Still the determination of the fully folded conformation of ribonucleic acids remains a difficult problem in spite of the rapidly increasing amount of structural information.

There are very few well-established tertiary RNA structures. The average conformational RNA A-form helix is known from fiber diffraction data (Arnott et al., 1973). Recently, the structures of two RNA oligomer duplexes have been determined (Dock-Bregeon et al., 1989; Happ et al., 1988). They generally conform to the A-form helix with local variations in stacking similar to those seen in the structures of DNA oligomers. Most significantly, the structures of phenylalanine,

† This work was supported in part by the Director, Office of Energy Research, Office of General Life Sciences, Molecular Biology Division of the U.S. Department of Energy under contract No. DE AC03-76SF00098.
* To whom correspondence should be addressed.
‡ Present address: Department of Biochemistry and Biophysics, College of Physicians and Surgeons, Columbia University, New York, NY 10032.

[1] Abbreviations: RNA, ribonucleic acid; NMR, nuclear magnetic resonance; cgr, conjugate gradient refinement; rms, root mean square; DHU, dihydrouridine; TΨC, ribothymidine-pseudouridine-cytosine; vdW, van der Waals; R_g, radius of gyration.

0006-2960/91/0430-5458$02.50/0