NMR Structural Studies on a Nonnatural Deoxyribonucleoside Which Mediates Recognition of GC Base Pairs in Pyrimidine·Purine·Pyrimidine DNA Triplexes†

Ishwar Radhakrishnan‡ and Dinshaw J. Patel*,‡,§

Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032, and Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021

E. Scott Priestley,‖ Huw M. Nash,‖ and Peter B. Dervan*,‖

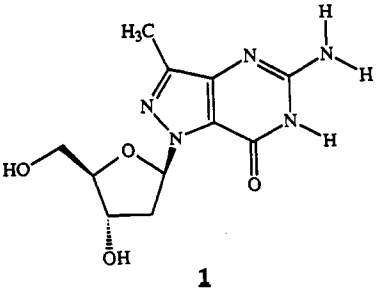
Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, California 91125

Received June 4, 1993; Revised Manuscript Received July 30, 1993®

ABSTRACT: As a part of our ongoing efforts to define the structural aspects of unusual pairing alignments in DNA triplexes by nuclear magnetic resonance spectroscopy, we have examined the structural role of a nonnatural deoxyribonucleoside, P1, that has been shown to mediate the recognition of GC base pairs in pyrimidine·purine·pyrimidine DNA triplexes [Koh, J. S., & Dervan, P. B. (1992) *J. Am. Chem Soc. 114*, 1470]. A qualitative interpretation of the NMR data indicates that this analog of protonated cytosine is readily accommodated in the third strand segment of an intramolecular triplex system. Furthermore, the observed NOE patterns position the imino and amino protons of P1 opposite the N^7 and O^6 atoms of guanine, respectively, consistent with the previously proposed pairing scheme.

Oligonucleotide-directed triple helix formation is a powerful method for sequence-specific recognition of double helical DNA (Moser & Dervan, 1987; LeDoan et al., 1987; Cooney et al., 1988; Beal & Dervan, 1991). Although triple helix formation by nucleic acids (Felsenfeld et al., 1957) was discovered soon after the elucidation of the structure of double helical DNA (Watson & Crick, 1953), it has only recently been exploited as a generalizable approach to molecular recognition of double helical DNA. This approach has now been used to mediate the single site-specific cleavage of a human chromosome (Strobel et al., 1991) and to inhibit gene transcription *in vitro* (Cooney et al., 1988; Maher et al., 1992).

Pyrimidine oligonucleotides bind within the major groove of homopurine sequences of duplex DNA, parallel to the Watson–Crick (WC) purine strand, by forming specific Hoogsteen hydrogen bonds to the WC purine bases (Moser & Dervan, 1987; Praseuth et al., 1988; de los Santos et al., 1989; Rajagopal & Feigon, 1989). Specificity derives from thymine (T) recognition of adenine–thymine (AT) base pairs (T·AT base triples) and N^3-protonated cytosine (C^+) recognition of guanine–cytosine (GC) base pairs (C^+GC base triples). The stability of triple helical nucleic acid complexes containing C^+GC base triples decreases as the solution pH increases (Moser & Dervan, 1987; Maher & Dervan, 1990; Plum et al., 1990; Singleton & Dervan, 1992), presumably due to the requirement for protonation of cytosines in the third strand in order to form two Hoogsteen hydrogen bonds to each GC base pair. Replacement of third strand cytosines by 5-methylcytosine results in increased triple helix stability at any given pH, but triple helix stability remains pH dependent (Povsic & Dervan, 1989; Singleton & Dervan, 1992). Several nucleosides have been studied which are capable of recognizing GC base pairs without protonation (Ono et al., 1991; Krawcyzk et al., 1992; Miller et al., 1992), including the nonnatural deoxyribonucleoside P1 (Koh & Dervan, 1992). The nucleoside P1 **1**, which mimics the Hoogsteen base-pairing edge of



1

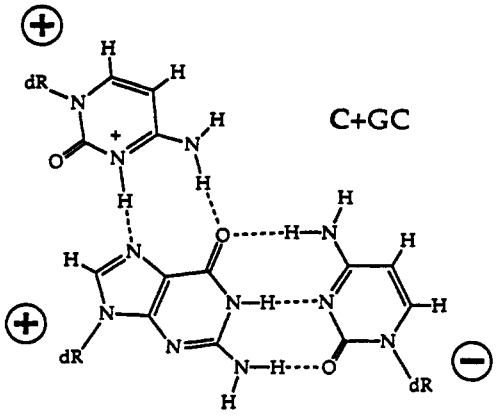
protonated cytosines in C^+GC triples **2**, is capable of sequence-specific binding of target sites containing multiple GC base pairs over an extended pH range relative to both cytosine and 5-methylcytosine (Koh & Dervan, 1992). We report here an NMR study that was undertaken to verify the proposed base-pairing scheme for the P1·GC base triple **3** (Koh & Dervan, 1992) and examine the structural implications of introducing

† We are grateful to the National Institutes of Health (GM 34504) for research support to D.J.P., the Office of Naval Research for research support to P.B.D., and the National Science Foundation for a predoctoral fellowship to E.S.P.

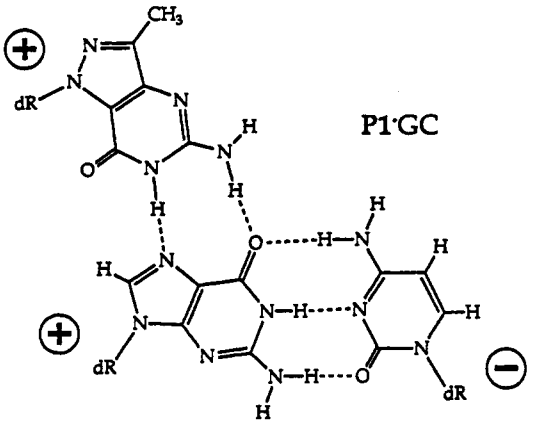
* Authors to whom correspondence should be addressed.
‡ Columbia University.
§ Memorial Sloan-Kettering Cancer Center.
‖ California Institute of Technology.

® Abstract published in *Advance ACS Abstracts*, September 15, 1993.

0006-2960/93/0432-11228$04.00/0

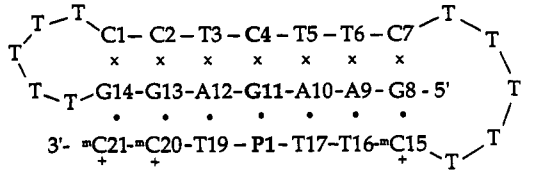


2



3

such a moiety in an intramolecular pyrimidine·purine·pyrimidine (Y·RY) triple helix **4**.



4

MATERIALS AND METHODS

Oligodeoxyribonucleotide Preparation. Oligodeoxyribonucleotide **4** was synthesized on a 10-μmol scale using solid-phase β-cyanoethyl phosphoramidite chemistry on an Applied Biosystems Model 380B DNA synthesizer. 2′-Deoxy-5-methylcytidine-derivatized controlled pore glass was prepared according to the standard method (Gait, 1984) and used as the solid support in this synthesis. The 5′-dimethoxytrityl-β-cyanoethylphosphoramidite of P1 was prepared as described (Koh & Dervan, 1992) and coupled as a 0.15 M solution in CH_3CN with >99% efficiency. The oligonucleotide was deprotected by treatment of the controlled pore glass support with saturated aqueous NH_4OH for 42 h at 55 °C. The crude 5′-dimethoxytrityl protected products were purified by FPLC on a ProRPC 16/10 reverse-phase column (Pharmacia LKB) with a linear gradient of 0–40% CH_3CN in 100 mM triethylammonium acetate, pH 7.0, detritylated in 80% acetic acid, and chromatographed a second time under the same conditions. Fractions containing pure oligonucleotide were combined and lyophilized to dryness, yielding 0.88 μmol of **4** as determined by UV absorbance (ϵ_{260} = 283 000 M^{-1} cm^{-1}). Analysis of the purified oligonucleotide by reverse-phase HPLC on a C-18 Aquapore OD-300 column (220 × 4.6 mm, Applied Biosystems) using a linear gradient of 0–20% CH_3CN in 100 mM ammonium acetate, pH 4.7, at 54 °C revealed a single UV-absorbing peak. The nucleoside composition of **4** was confirmed by digestion of a 5-nmol aliquot of the oligonucleotide with 0.01 unit of snake venom phosphodiesterase (Sigma) and 4 units of calf alkaline phosphatase (Boehringer Mannheim) in 30 μL of 30 mM Tris acetate and 70 mM $MgCl_2$, pH 8.1, at 37 °C for 2 h. The digest products were separated by reverse-phase HPLC on a C-18 Aquapore OD-300 column (220 × 4.6 mm, Applied Biosystems) using a linear gradient of 0–2.5% CH_3CN in 200 mM triethylammonium acetate, pH 6.0. Base ratios calculated from the chromatogram peak areas were 16:4.2:4.0:2.6:2.9:0.91 T:dC:dG:dA:^{me}dC:P1, using the following extinction coefficients at 260 nm: T, 8800; dC, 7300; dG, 11 700; dA, 15 400; ^{me}dC, 5700; P1, 2500.

NMR Sample Preparation. Prior to the NMR studies, the oligonucleotide **4** was desalted and converted to the Na^+ form by sequentially passing through Sephadex G-25 and Dowex ion-exchange columns.

NMR Experiments. Proton NMR spectra were recorded on a Bruker AMX500 spectrometer on approximately 260 A_{260} units of the oligonucleotide **4** in the presence of 0.1 mM EDTA and 10 mM phosphate buffer at 5 °C, pH 6.01, in H_2O and at 15 °C, pH 6.02 (uncorrected for isotope effect) in D_2O. The raw data were transferred to a Silicon Graphics Indigo workstation and processed with the Felix software package (Hare Research Inc.). The data acquisition and processing parameters are given in the figure legends.

RESULTS

The P1 base was incorporated in the third strand segment of a model intramolecular triplex system that has been the focus of several studies in our laboratory (Radhakrishnan et al., 1991; Lin & Patel, 1992; Wang and Patel, unpublished results). The sequence and numbering system for the oligonucleotide used in this study is shown in **4**. The P1·GC triple is located at the central position of the triplex, which differs from the previously studied sequences in that cytosines in the third strand segment were replaced by 5-methylcytosines in order to facilitate studies at higher pH.

Exchangeable Proton Spectra. The exchangeable proton spectrum (6.5–15.5 ppm) of the intramolecular triplex **4** at 5 °C, pH 6.01, in H_2O buffer (Figure 1A), exhibits narrow, well-resolved resonances in the hydrogen-bonded imino proton region (12.5–14.5 ppm, Figure 1B). The resonances between 8.5 and 10.2 ppm are diagnostic of a pyrimidine·purine·pyrimidine triplex species (de los Santos et al., 1989; Rajagopal & Feigon, 1989); these resonances correspond to the amino protons of the N^3-protonated 5-methylcytosines. A new resonance at 11.63 ppm can be detected under these conditions which presumably belongs to the imino proton of P1 (see below). The minor resonances that can be observed in the hydrogen-bonded imino proton region (Figure 1B) belong to the hairpin duplex with a dangling third strand that is in slow equilibrium with the predominant species—the intramolecular triplex that is the focus of the current study.

Imino Proton Assignments. The symmetrical imino proton region in the NOESY spectrum (mixing time 150 ms) of triplex **4** is shown in Figure 2. Sequential NOE connectivities

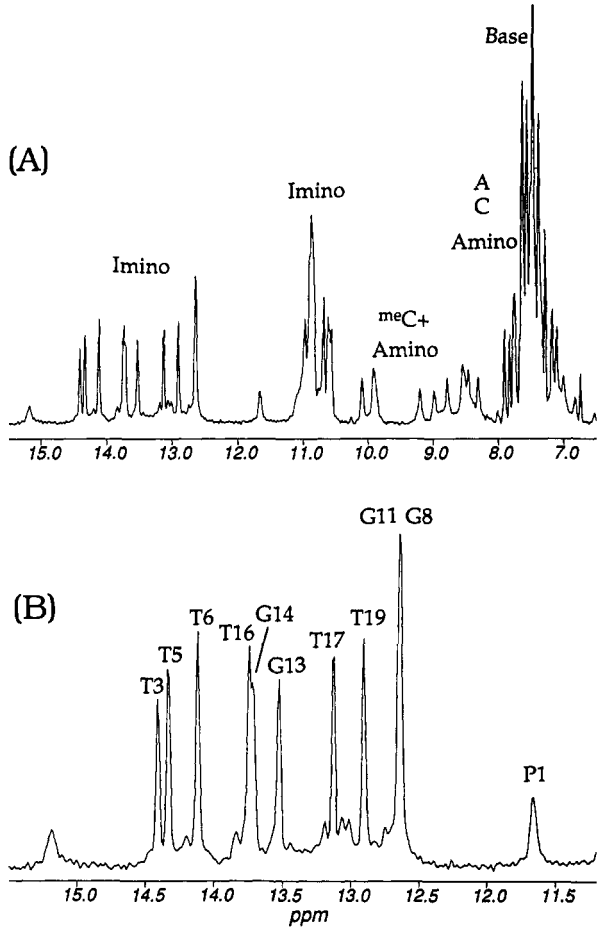


FIGURE 1: One-dimensional proton NMR spectrum of triplex **4** recorded in H_2O buffer at pH 6.01, 5 °C. (A) Only the region downfield to the H_2O resonance (6.5–15.5 ppm) is shown. The "jump-and-return" sequence (Plateau & Gueron, 1982) was used to suppress the H_2O signal with the delay between the two pulses set to 56 μs for maximum excitation of the hydrogen-bonded imino proton resonances. (B) Expanded plot of the hydrogen-bonded imino proton region (11.2–15.5 ppm). The assignments for the various imino proton resonances are labeled.

between imino protons on adjacent base pairs can be traced in the duplex portion of the molecule (Figure 2A). The tracing starts from the G8 imino proton and proceeds via T6, T5, G11, T3, and G13 imino protons to the G14 imino proton. Sequential NOEs between imino protons on adjacent bases in the third strand can also be observed starting from T16 and proceeding via T17 and P1 imino protons to the T19 imino proton. The latter NOEs involving the P1 imino proton with its flanking thymine imino protons establishes a stacked-in conformation for the P1 base within the third strand of the helix. The P1 imino proton exchanges more rapidly with the solvent than the other interior hydrogen-bonded imino proton resonances thus accounting for the weak intensity of these cross peaks.

AT and GC Watson–Crick Interactions. The bases in the purine and the first pyrimidine strand segments participate in Watson–Crick-type hydrogen-bonding interactions. This could be verified by the chemical shifts of the imino and amino protons as well as by the presence of NOEs that characterize these alignments. In particular, NOEs could be detected correlating the imino protons of T3, T5, and T6 with the H2 and hydrogen-bonded amino protons of A12, A10, and A9, respectively, in AT base pairs, while the imino protons of G8, G11, G13, and G14 show NOEs to the hydrogen-bonded and exposed amino protons of C7, C4, C2, and C1, respectively,

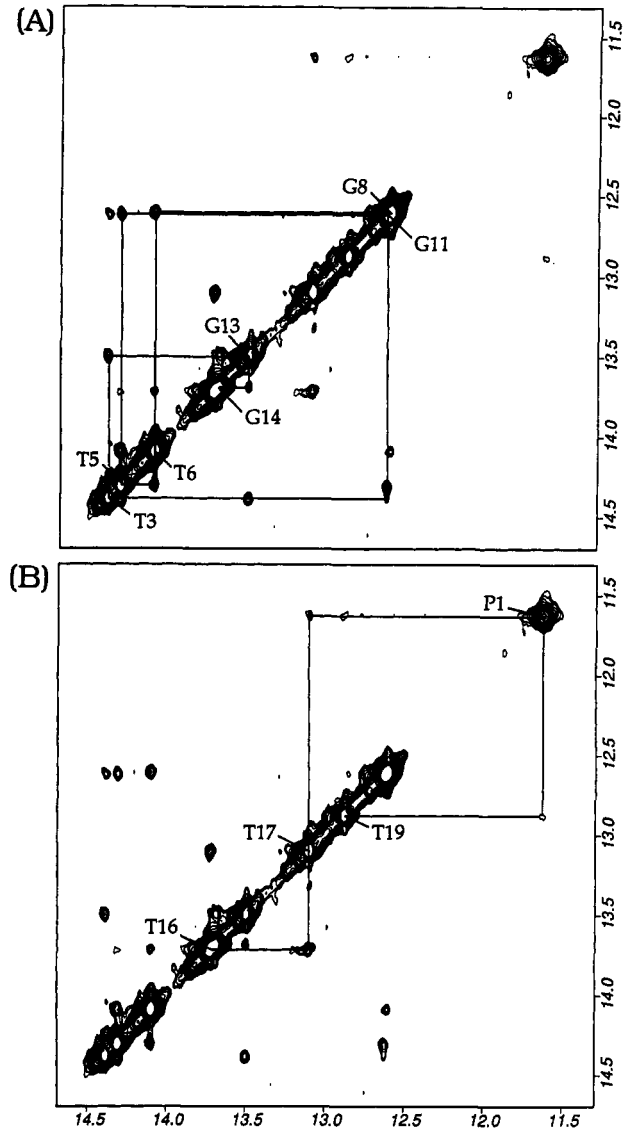


FIGURE 2: Duplicate, expanded contour plots of the symmetrical imino proton region (11.3–14.7 ppm) in the NOESY (mixing time 150 ms) spectrum of triplex **4** recorded in H_2O buffer, at pH 6.01, 5 °C. Hard 70° preparation and mixing pulses were used with the detection pulse replaced by a "jump-and-return" sequence (Plateau & Gueron, 1982). The carrier frequency was placed on the solvent signal and the spectral width was set to 11.1 kHz in both dimensions. The data were acquired with 1024 complex data points in t_2 and 400 complex data points in t_1 in the TPPI–States' mode (Marion et al., 1989b). Prior to Fourier transformation in t_2, the residual solvent signal in each increment was deconvoluted in the time domain (Marion et al., 1989a), apodized with a skewed 30° phase-shifted sine bell window function, and zero-filled to 2048 points. Prior to Fourier transformation in t_1, baseline correction was applied in ω_2 and the interferograms apodized with a 90° phase-shifted sine bell squared function and zero-filled to yield a final processed matrix size of 2048 × 2048 in real points. (A) The lines trace the sequential connectivities between imino protons in adjacent base pairs for the duplex portion of triplex **4**. The tracing starts at G8 and proceeds via T6, T5, G11, T3, and G13 to G14. (B) The tracing of sequential connectivities between adjacent bases in the third strand of triplex **4** starts at T16 and proceeds via T17 and P1 to T19. The N^3-protonated imino proton resonances of 5-methylcytosines in this strand exchange too rapidly with the solvent to show detectable NOEs.

in GC base pairs. The NOE cross peak assignments for the central trinucleotide segment of the duplex portion of the molecule are shown in Figure 3A.

T·AT and $^{me}C^+$GC Hoogsteen Interactions. In addition to the Watson–Crick-type interactions, the purine bases in the second strand segment participate in Hoogsteen interactions

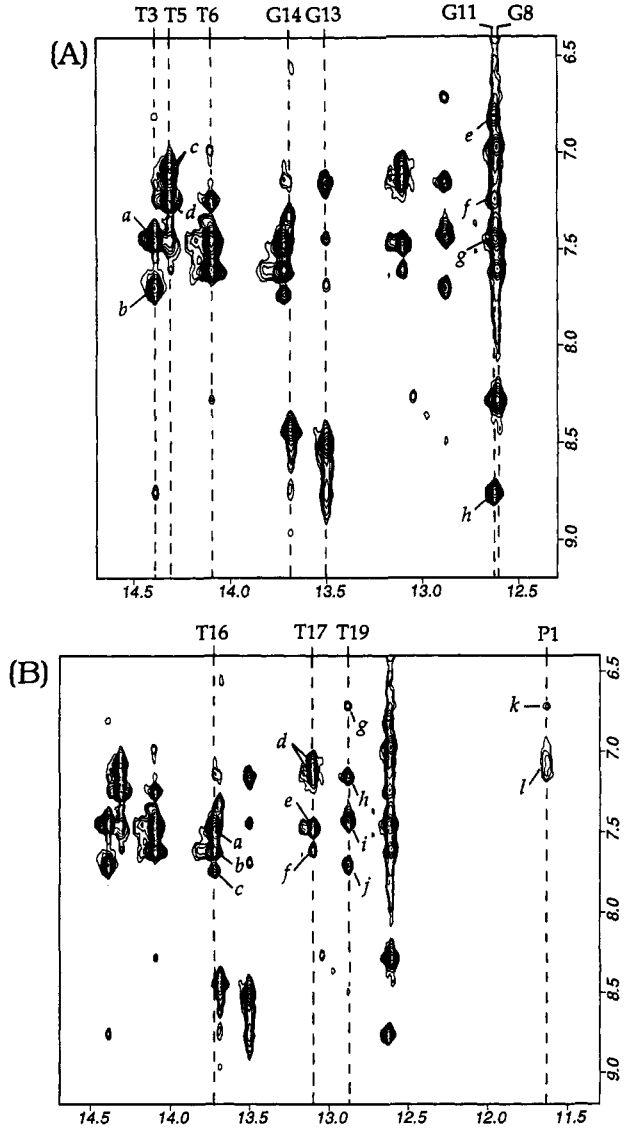


FIGURE 3: Expanded contour plots of the same NOESY spectrum of triplex **4** shown in Figure 2. (A) NOE cross peaks correlating the imino protons (12.3–14.7 ppm) and the amino and base protons (6.4–9.2 ppm) in the Watson–Crick pairs in the triplex. The cross peak assignments for the protons in the central trinucleotide segment are as follows: a, T3(NH3)–A12(H2)/A12(NH_2-h); b, T3(NH3)–A12(NH_2-wc); c, T5(NH3)–A10(NH_2-wc)/A10(NH_2-h); d, T5(NH3)–A10(H2); e, G11(NH1)–C4(NH_2-e); f, G11(NH1)–A10(H2); g, G11(NH1)–A12(H2); h, G11(NH1)–C4(NH_2-b). (B) NOE cross peaks between the imino protons (11.3–14.7 ppm) and the amino and base protons (6.4–9.2 ppm) of the Hoogsteen pairs in the triples. The cross peak assignments are as follows: a, T16(NH3)–A9(NH_2-wc)/A9(NH_2-h); b, T16(NH3)–A9(H8); c, T16(NH3)–G8(H8); d, T17-(NH3)–A10(NH_2-wc)/A10(NH_2-h); e, T17(NH3)–A10(H8); f, T17(NH3)–A9(H8); g, T19(NH3)–G11(H8); h, T19(NH3)–A12-(H8); i, T19(NH3)–A12(NH_2-h); j, T19(NH3)–A12(NH_2-wc); k, P1(NH1)–G11(H8); l, P1(NH1)–(NH_2). In adenines, amino protons involved in Watson–Crick and Hoogsteen pairing are denoted by "wc" and "h", respectively, while in cytosines, the hydrogen-bonded and exposed amino protons are denoted by "b" and "e", respectively.

with the third pyrimidine strand segment. NOEs that establish these alignments in T·AT triples could be detected between the imino protons of T16, T17, and T19 and the H8 and amino protons of A9, A10, and A12, respectively (peaks a and b, d and e, and h, i and j, Figure 3B). Unlike these thymine imino protons, the protonated 5-methylcytosine imino protons are too broad to show analogous NOE patterns that would characterize the Hoogsteen alignments in $^{me}C^{+}GC$ triples. However, the chemical shifts of protonated 5-methylcytosine amino protons (8.5–10.2 ppm) taken along with the NOEs observed between these protons and the amino protons of cytosines in the Watson–Crick duplex within the same triple (data not shown) implicate $^{me}C15$, $^{me}C20$, and $^{me}C21$ in Hoogsteen interactions with G8, G13, and G14 bases, respectively, in the purine strand.

Pairing Alignment in the P1·GC Triple. For comparison, the proposed pairing alignment for the P1·GC triple **3** (Koh & Dervan, 1992) is shown along with the pairing alignment for a $C^{+}GC$ triple **2**. In both cases, the imino proton is involved in hydrogen bonding with the N^7 atom of guanine while one of the amino protons is hydrogen-bonded to the O^6 atom of the same guanine. The NOEs shown by these imino and amino protons together with their chemical shifts are critical in defining the pairing alignment in the P1·GC triple. Two NOEs involving the imino proton of P1 are clearly discernible in the imino to amino and base proton region of the NOESY spectrum of triplex **4** (Figure 3B). The imino proton of P1 shows an NOE to the H8 proton of G11 (peak k, Figure 3B) within the same triple. The other NOE (peak l, Figure 3B) which is comparatively stronger is between the imino and amino protons of P1. It appears that, in the time scale of chemical shift differences, rotation around the C–N bond of the amino group of P1 is in the intermediate regime giving rise to a single broad resonance for these protons. The P1 imino proton also shows much weaker NOEs to the amino protons of neighboring A12 (data not shown). Most importantly, however, is the *absence* of any detectable NOE between the imino proton of P1 and the amino protons of C4 in the same triple. Since the P1 imino proton exchanges rapidly with the solvent, special care was taken to arrive at this conclusion. In the $\omega_2 = 11.63$ ppm slice corresponding to the P1 imino proton resonance in the NOESY spectrum, we do not detect any peaks above the noise level at the resonance positions ($\omega_1 = 8.75, 7.98$ ppm) corresponding to the C4 amino protons. The noise in this slice is sufficiently uniform to arrive at the aforementioned conclusion. If the imino proton were hydrogen-bonded to the O^6 carbonyl oxygen of G11, an NOE would be expected between this proton and the amino protons of C4 that would be much stronger than the NOE between the P1 imino and G11 H8 protons. These observations place the imino proton within hydrogen-bonding distance of the N^7 atom of G11 and provide definitive evidence for the alignment shown in **3**. Corroborating evidence in the form of NOE patterns shown by the amino proton of P1 could not be obtained because of resonance overlap with the amino protons of A10. Although the resonances at these chemical shift positions show NOEs to the C4 amino protons, the NOEs could not be unambiguously assigned to the interaction between the P1 amino and C4 amino protons since the A10 amino protons located on the adjacent triple could also participate in such interactions with the C4 amino protons.

The chemical shift of the P1 imino proton resonance shows a dependence on the pH of the solution. A systematic pH study on an oligonucleotide sequence related to **4**, where 5-methylcytosines were replaced with cytosines, revealed that the chemical shift of this resonance moves downfield when the pH of the solution is lowered from pH 6.2 to 5 (Figure 4). By contrast, the chemical shifts of the methyl protons in the same ring system as well as the other imino proton resonances in the triplex undergo only subtle changes in the same pH interval. Although an analogous study was not undertaken for triplex **4**, a similar behaviour for the P1 imino proton resonance in this sequence was noticed when the pH was lowered. We suspect that the downfield shift associated with this resonance in a P1·GC triple is due either to pro-

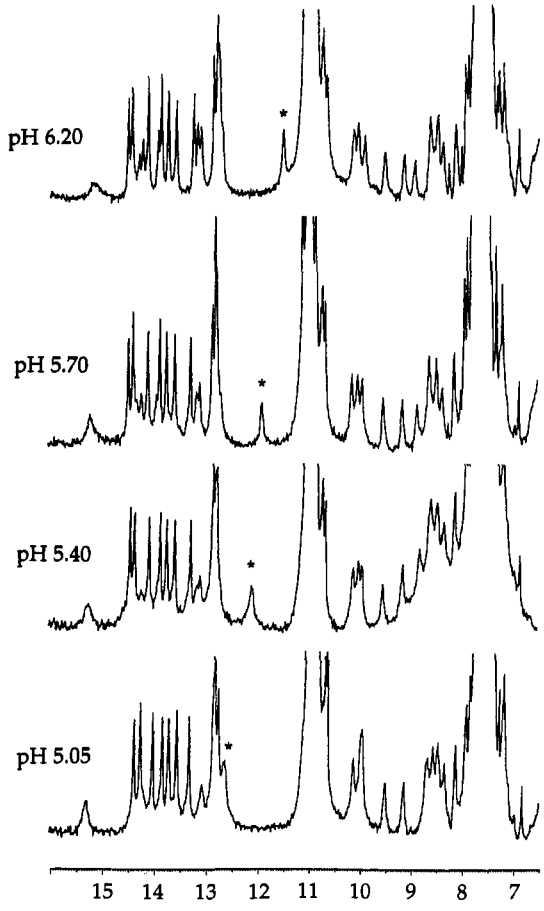


FIGURE 4: One-dimensional proton NMR spectra of a triplex 4 analog where cytosines have replaced 5-methylcytosines, recorded at 1 °C in H_2O in the presence of 0.1 mM EDTA and 10 mM phosphate buffer at different pH values. Only the region downfield from the water resonance (6.5–16.0 ppm) has been plotted. The position of the P1 imino proton resonance in each spectrum is indicated by an asterisk.

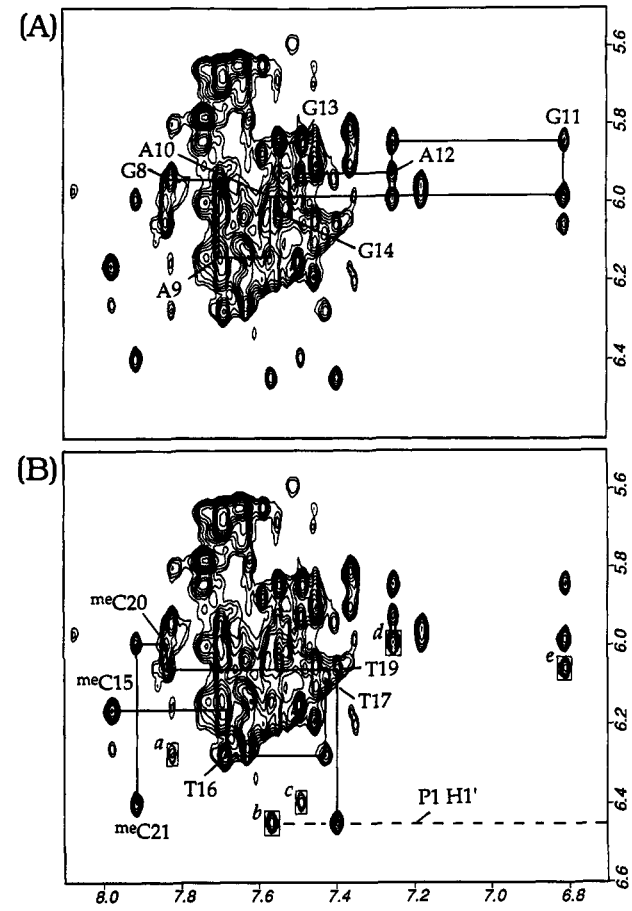


FIGURE 5: Duplicate, expanded contour plots correlating the base (6.7–8.1 ppm) and sugar H1′ protons (5.5–6.6 ppm) in the NOESY spectrum (mixing time 250 ms) of triplex 4 recorded in D_2O buffer, at pH 6.02, 15 °C. The data were acquired with 1024 complex data points in t_2 and 512 real data points in t_1 in the TPPI mode (Marion & Wüthrich, 1983) with the spectral width set to 5.05 kHz in both dimensions and a repetition delay of 1.6 s. 90° phase-shifted, sine bell squared, and sine bell window functions were used for apodization in t_2 and t_1 dimensions, respectively. The data were zero-filled to 2048 points in both dimensions prior to Fourier transformation. Baseline correction was applied to the frequency domain data in both dimensions. (A) The lines trace sequential NOE connectivities between the base and sugar H1′ protons in the purine strand of triplex 4. The tracing starts at G8 and traverses the strand up to G14. The intraresidue base to sugar H1′ cross peaks in this strand are labeled by residue number. (B) The solid lines trace sequential NOE connectivities between the base and sugar H1′ protons in the third strand of triplex 4 starting from ^{me}C15. The tracing continues up to the intraresidue base to sugar H1′ cross peak of T17 where it is interrupted due to the absence of a base proton on P1 but resumes at the T19(H6)–P1(H1′) cross peak and continues to ^{me}C21 at the end of the strand. The intraresidue cross peaks involving the base and sugar H1′ protons in this strand are labeled by residue number. The dashed line corresponds to the position of the P1 H1′ resonance. The cross-strand NOEs between the base protons in the purine strand and the H1′ protons in the third strand are boxed and these are assigned as follows: a, G8(H8)–T16(H1′); b, A10(H8)–P1(H1′); c, G13(H8)–^{me}C21(H1′); d, A12(H8)–^{me}C20(H1′); e, G11(H8)–T19-(H1′).

tonation of the P1 ring and/or to the formation of a stronger hydrogen bond with the N^7 atom of guanine.

Structural Perturbations Introduced by the P1 Base in the Third Strand. The third strand segment in triplex 4 consists exclusively of pyrimidines with the exception of the P1 base at the central position. To assess possible structural perturbations induced by the introduction of this P1 ring system, studies were extended to D_2O. Figure 5A traces the sequential connectivities for the purine strand of triplex 4 in the base to sugar H1′ region of a NOESY spectrum (mixing time 250 ms) recorded in D_2O buffer at pH 6.02, 15 °C. Despite severe overlap, some of the cross peaks are well-resolved, particularly those in the central trinucleotide segment of the molecule. Continuity of sequential connectivities is maintained in the A10–G11–A12 segment not only in this region but in the base to H2′, H2″ region as well. In the third strand of the molecule, we can trace sequential connectivities starting from ^{me}C15 and proceeding via T16 until the intraresidue base to H1′ cross peak of T17 (Figure 5B). The tracing is interrupted in this region because of the absence of a suitable base proton on P1 that is sufficiently close to show cross peaks with any of its sugar protons. However, the tracing is resumed with the T19(H6)–P1(H1′) cross peak and proceeds to the 3′-end of the strand via ^{me}C20 to ^{me}C21 (Figure 5B). In the symmetrical methyl and sugar H2′, H2″ proton region, however, we can trace sequential connectivities between methyl protons in adjacent bases in the third strand starting from T17 and continuing via P1 and T19 to ^{me}C20 (Figure 6). This is consistent with the stacked-in conformation for P1 deduced from the H_2O data. An NOE is also detected between the P1 methyl proton and the H2″ proton of T17 (boxed peak, Figure 6). Other sequential contacts involving P1 including T17(H6)–P1(CH_3), T19(H6)–P1(H2′), and T19(H6)–P1-(H2″) are also observed and suggest no obvious structural perturbation at this site.

The observation of cross-strand NOEs between residues in adjacent triples provides an indication of the size of the groove formed by the purine and pyrimidine third strands

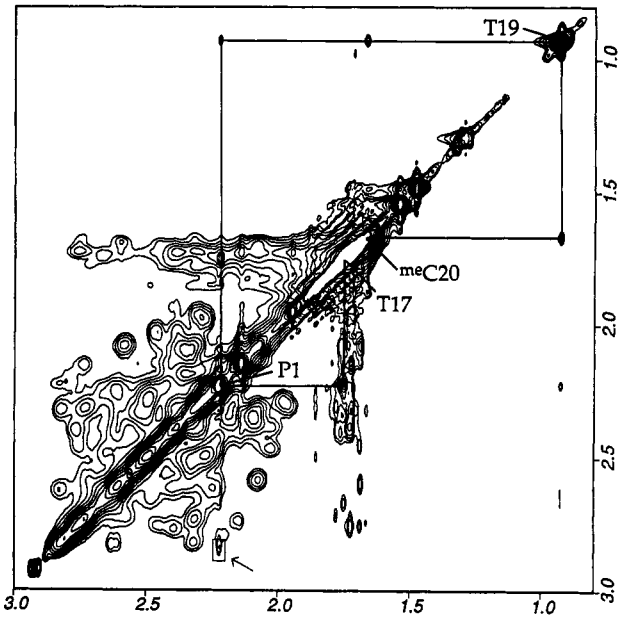


FIGURE 6: Expanded contour plot of the symmetrical methyl and H2′,2″ proton region in the same NOESY spectrum of triplex **4** shown in Figure 5. The lines trace the sequential connectivities between methyl protons on adjacent bases in the third strand of the molecule starting from T17 and proceeding via P1 and T19 to ^{me}C20. The boxed peak (indicated by an arrow) corresponds to the NOE observed between the sugar H2″ proton of T17 and the methyl protons of P1.

(Radhakrishnan et al., 1992). Three classes of NOEs are useful in this regard. The imino protons of thymines show NOEs not only to the H8 protons of the bases with which they hydrogen bond but also to the H8 protons of the bases on the 5′-side of their Hoogsteen partners. A similar trend is observed for the triplex **4** where the T16, T17, and T19 imino protons show such NOEs to the H8 protons of G8, A9, and G11 bases in the purine strand (peaks c, f, and g, Figure 3B). NOEs are also observed between these thymine imino protons and the sugar H2′, H2″ protons in the same pairs of residues (data not shown). The third class of NOEs that are observed involve the base protons in the purine strand and the H1′ protons of the bases in the third strand. In triplex **4**, we detect these types of interactions clearly, and they are shown as boxed peaks in Figure 5B. Around the P1·GC triple, we detect NOEs between the H1′ protons of P1 and T19 and the H8 protons of A10 and G11, respectively (peaks b and e, Figure 5B). The intensities of the cross peaks arising from these types of interactions were compared in a NOESY spectrum acquired with a short mixing time (50 ms). The intensity of the cross peaks b, d, and e were somewhat stronger than those of cross peaks a and c shown in Figure 5B. These observations suggest that the groove width is not significantly altered by the introduction of P1 in the pyrimidine third strand helix.

Chemical Shifts. Several unusual shifts are observed in triplex **4** especially in and around the P1·GC triple (Table I). These include the rather upfield shifts for G11 H8 (6.81 ppm) and T19 CH_3 (0.93 ppm) protons, as well as the downfield shifts associated with the H1′ (6.45 ppm) and CH_3 (2.22 ppm) protons of P1. While the cause for the upfield shift for the G11 H8 proton is not obvious, the upfield shift for the methyl protons of T19 is most likely due to the positioning of the methyl group on top of the P1 ring system. The latter effect was observed in an earlier study where T19 was flanked by a guanine on the 5′-side (Radhakrishnan et al., 1992). In the instances involving P1 protons, it should be noted that the ring system of P1 might also contribute to the unusual chemical shifts.

DISCUSSION

As noted earlier, the requirement for protonation at the N^3 positions of third strand cytosines in Y·RY triplexes has several drawbacks. The stability of these structures is sensitive to the solution pH, and the addition of charge arising from protonation has a destabilizing effect on the structure, particularly in sequences that contain several contiguous C^+GC triples (Moser & Dervan, 1987; Maher & Dervan, 1990; Plum et al., 1990; Singleton & Dervan, 1992). In order to obviate this requirement for protonation, several synthetic approaches have been undertaken in designing suitable protonated cytosine analogs (Ono et al., 1991; Krawcyzk et al., 1992; Miller et al., 1992; Koh & Dervan, 1992; Kiessling et al., 1992). The nonnatural nucleoside P1 not only mimics protonated cytosines in sequence-specific interactions with guanines in GC pairs but can also bind to several contiguous GC base pairs over an extended pH range. In this study, the structural consequences of introducing this base at a single site opposite a guanine in an intramolecular Y·RY triplex **4** have been examined by NMR spectroscopy.

From our NMR data, it appears that the P1 base is readily accommodated in an otherwise pyrimidine third strand segment. Sequential NOE contacts between P1 and its flanking thymidine residues involving imino protons (Figure 2B) and methyl protons (Figure 6) attest to a stacked-in conformation for this base. Sequential NOEs between base and sugar protons consistent with a regular, right-handed helix are observed (Figures 4B and 5) for the segment T17–P1–T19. In the 5′-P1–T19-3′ step, an NOE between the H6 proton of T19 and the H1′ proton of P1 is detected (Figure 5B), while in the 5′-T17–P1-3′ step, an NOE between the methyl proton of P1 and the H2″ proton of T17 is observed (Figure 6). Other NOEs including T19(H6)–P1(H2′), T19-(H6)–P1(H2″), and T17(H6)–P1(CH_3) provide further

Table I: ^{1}H Chemical Shifts for the Exchangeable[a] and Nonexchangeable[b] Protons in the Central Trinucleotide Segment in Triplex **4**

residue	H1/H3	H2[c]	H8/H6	NH_2[d]	H5/CH_3	H1′	H2′/H2″	H3′
T3	14.39		7.63		1.78	6.04	*e*, 2.62	4.94
C4			7.69	8.75, 7.98	5.69	6.20	2.40, 2.76	4.80
T5	14.31		7.46		1.69	5.88	2.08, 2.60	4.93
A10		7.24	7.57	7.08, 7.13		5.99	2.32, 2.83	4.87
G11	12.62		6.81			5.85	2.32, 2.81	4.97
A12		7.45	7.25	7.70, 7.45		5.93	2.21, 2.80	4.84
T17	13.10		7.43		1.75	6.08	2.15, 2.84	4.97
P1	11.63[f]			7.07	2.22	6.45	2.63, 2.68	4.94
T19	12.88		7.40		0.93	6.06	2.32, 2.74	4.64

[a] Chemical shifts for the exchangeable protons at 5 °C, pH 6.01. [b] Chemical shifts for the nonexchangeable protons at 15 °C, pH 6.02. [c] Chemical shifts for the H2 protons in H_2O at 5 °C, pH 6.01. [d] For cytidine and adenine amino protons, the chemical shift of the Watson–Crick hydrogen-bonded proton is indicated first. [e] Could not be unambiguously assigned due to severe overlap. [f] Chemical shift is pH dependent.

evidence for a right-handed helical conformation in the T17–P1–T19 segment. Cross-strand NOE patterns between protons on adjacent triples suggest no apparent structural distortions manifested in an altered width for the groove formed by the purine strand and pyrimidine third strand.

As we noted earlier, in order to define the pairing alignment of the P1·GC triple, the P1 imino proton should provide an important marker. The P1 imino proton resonance could be easily assigned from the NOESY spectrum recorded in H_2O (Figure 2B). However, compared to the other interior hydrogen-bonded imino protons in the molecule, the P1 imino proton exchanges with the solvent at a faster rate, which attenuates the intensity of the NOEs shown by this proton. We do detect a critical NOE, albeit weak, between this imino proton and the H8 proton of G11 in the same NOESY spectrum (Figure 3B). In addition, an NOE of weak intensity could be observed between the imino and amino protons of P1 that allowed us to assign the amino proton resonance (7.07 ppm). Most importantly, however, we do not detect any NOE between the same imino proton and the amino protons of C4. These observations taken in the light of our findings in D_2O (discussed above) constrain the position of the P1 imino proton opposite the N^7 proton of G11 while implicitly placing the amino protons nearer to the O^6 atom of the same base, consistent with the proposed alignment 3 (Koh & Dervan, 1992).

In summary, we have structurally characterized the interaction of a nonnatural analog of protonated cytidine with a GC base pair in an intramolecular pyrimidine·purine·pyrimidine triple helix by NMR spectroscopy. The P1 base is readily accommodated in an otherwise pyrimidine third strand segment while participating in hydrogen-bonding interactions consistent with the proposed pairing alignment 3.

ACKNOWLEDGMENT

We thank Dr. Steven Woski for a supply of 2′-deoxy-5-methylcytidine-derivatized controlled pore glass. The NMR spectrometers were purchased from funds donated by the Robert Woods Johnson Trust toward setting up an NMR Center in the Basic Medical Sciences at Columbia University. We acknowledge the use of the Molecular Modeling Facility for Molecular Biology at Columbia University, supported in part by NSF Grant DIR-8720229.

REFERENCES

Beal, P. A., & Dervan, P. B. (1991) *Science 251*, 1360–1363.
Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.
de los Santos, C., Rosen, M., & Patel, D. (1989) *Biochemistry 28*, 7282–7289.
Felsenfeld, G., Davies, D., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.
Gait, M. J. (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press Limited, Oxford.
Kiessling, L. L., Griffin, L. C., & Dervan, P. B. (1992) *Biochemistry 31*, 2829–2834.
Koh, J. S., & Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 1470–1478.
Krawcyzk, S. H., Milligan, J. F., Wadwani, S., Moulds, C., Froehler, B. C., & Mateucci, M. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3761–3764.
LeDoan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thoung, N. T., Lhomme, J., & Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749–7760.
Lin, C. H., & Patel, D. J. (1992) *J. Am. Chem. Soc. 114*, 10658–10660.
Maher, L. J., & Dervan, P. B. (1990) *Biochemistry 29*, 8820–8826.
Maher, L. J., Dervan, P. B., & Wold, B. (1992) *Biochemistry 31*, 70–81.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Marion D., Ikura, M., & Bax, A. (1989a) *J. Magn. Reson. 84*, 425–430.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989b) *J. Magn. Reson. 85*, 393–399.
Miller, P. S., Bahn, P., Cushman, C. D., & Trapane, T. L. (1992) *Biochemistry 31*, 6788–6793.
Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.
Ono, A., Ts'o, P. O. P., & Kan, L. S. (1991) *J. Am. Chem. Soc. 113*, 4032–4033.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.
Plum, E. G., Park, Y.-W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436–9440.
Povsic, T. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc. 111*, 3059–3061.
Praseuth, D., Perrouault, L., Le Doan, T., Chassignol, M., Thuong, N., & Hélène, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1349–1353.
Radhakrishnan, I., Gao, X., de los Santos, C., Live, D., & Patel, D. J. (1991) *Biochemistry 30*, 9022–9030.
Radhakrishnan, I., Patel, D. J., & Gao, X. (1992) *Biochemistry 31*, 2514–2523.
Rajagopal, P., & Feigon, J. (1989) *Biochemistry 28*, 7859–7870.
Singleton, S. F., & Dervan, P. B. (1992) *Biochemistry 31*, 10995–11003.
Strobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E., & Dervan, P. B. (1991) *Science 254*, 1639–1642.
Watson, J. D., & Crick, F. H. C. (1953) *Nature 171*, 737–738.