Structure of a G·T·A Triplet in an Intramolecular DNA Triplex†

Edmond Wang, Shiva Malek, and Juli Feigon*

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, California 90024

Received December 2, 1991; Revised Manuscript Received February 24, 1992

ABSTRACT: A 32-base DNA oligonucleotide has been studied by one- and two-dimensional ^{1}H NMR spectroscopy and is shown to form a stable, pyr·pur·pyr, intramolecular triple helical structure, with a four C loop and a TATA loop connecting the Watson–Crick- and Hoogsteen-paired strands, respectively. This triplex contains five T·A·T base triplets, two C^+·G·C base triplets, and an unusual G·T·A base triplet which disrupts the pyr·pur·pyr motif. The G·T·A triplet consists of a Watson–Crick T·A base pair, with the T situated in the "purine strand" and the A situated in the "pyrimidine strand" and a G situated in the Hoogsteen-base-paired "pyrimidine strand" hydrogen bonded to the T. The base-pairing structure of the G·T·A triplet has been investigated and has been found to involve a single hydrogen bond from the guanine amino group to the O4 carbonyl of the thymine, leaving the guanine imino proton free. The specific amino proton involved in the hydrogen bond is the H2(2) proton. This orients the guanine such that its sugar is near the thymine methyl group. The guanine sugar adopts an N-type (C3′-endo) sugar pucker in this triplet. The stability of the G·T·A triplet within pyr·pur·pyr triplexes is discussed.

Triple-stranded DNA has been postulated to play a role in several important biological functions, such as transcriptional regulation and recombination [Weinreb et al., 1990; Hsieh et al., 1990; Htun & Dahlberg, 1989; Mirkin et al., 1987; reviewed in Wells et al. (1988)] and has been promoted as a potentially powerful biological tool, such as a chemical nuclease to map chromosomes (François et al., 1989a; Strobel et al., 1988; Strobel & Dervan, 1991; Perrouault et al., 1990) or as a highly sequence-specific repressor to block protein recognition of DNA (Cooney et al., 1988; Maher et al., 1989; Hanvey et al., 1990; Parniewski et al., 1990; François et al., 1989b). Homopyrimidine–homopurine–homopyrimidine (pyr·pur·pyr) DNA triplexes were first observed 35 years ago (Felsenfeld et al., 1957), and the low-resolution structure of triplexes has been known from X-ray fiber diffraction studies for more than 15 years (Arnott & Selsing, 1974). From these studies, a model for pyr·pur·pyr triplexes was derived and consists of a Watson–Crick purine–pyrimidine duplex with a second pyrimidine strand binding to the purine strand in a parallel orientation in the major groove of the duplex via Hoogsteen base pairing (Hoogsteen, 1959). This base-pairing scheme and parallel orientation of the third strand was confirmed by NMR work from our laboratory (Rajagopal & Feigon, 1989a,b; Sklenář & Feigon, 1990a) and other laboratories (de los Santos et al., 1989; Mooren et al., 1990; Pilch et al., 1990). From the early X-ray fiber diffraction studies, it was concluded that the DNA adopts an A′-DNA-like helical structure and, consequently, a C3′-endo sugar pucker. However, NMR studies from our laboratory on an intramolecular triplex indicate that the majority of sugars have a predominantly S-type (C2′-endo) sugar pucker, suggesting a structure closer to an underwound B-DNA than to an A-DNA helix (Macaya et al., 1992b).

Two of the four possible Watson–Crick base pairs are specifically recognized by the pyrimidine third strand. A thymine recognizes an A·T base pair, forming a T·A·T triplet, and a protonated cytosine recognizes a G·C base pair, forming a C^+·G·C triplet. Understandably, there is much interest in determining the specificity of triplex formation and in expanding the scope of triplex recognition to all four duplex base pairs within a single triple-helical motif. To address these concerns, our laboratory has been conducting an investigation of the structure of intramolecular pyr·pur·pyr triplexes (Sklenář & Feigon, 1990a; Macaya et al., 1922a,b) containing alternative base triplets by using ^{1}H nuclear magnetic resonance (NMR) on a group of closely related DNA oligonucleotides derived from the 32-base sequence d-(AGAGAGAA CCCC TTCTCTCT TATA TCTCTCTT) (Macaya et al., 1991). This DNA oligomer (CGC32) has previously been shown to fold into an intramolecular triplex consisting of eight triplets (five T·A·T and three C^+·G·C) and two loops (CCCC and TATA; shown in bold). We have created a family of triplexes by replacing the central C^+·G·C triplet (underlined) with the other X·Y·Z triplets, where X = A, C, G, or T and is in the Hoogsteen-paired strand and Y·Z is a Watson–Crick base pair. Using this system, we have recently reported on the base-pairing schemes of the X·G·C triplets, where X = A, C, G, or T, as well as preliminary results on the G·T·A triplet (Macaya et al., 1991). We have also used this system to investigate the stability and melting temperatures of all 16 possible triplets by UV spectrophotometry (Macaya et al., 1991; S. Malek and J. Feigon, unpublished results).

Dervan and co-workers (Griffin & Dervan, 1989), utilizing an Fe–EDTA DNA cleaving moiety linked to the third-strand oligonucleotides, showed that a T·A base pair could be specifically recognized by a guanine, forming a G·T·A triplet within a $d(T)_n$-$d(A)_n$-$d(T)_n$ triplex. The specific recognition of a T·A base pair by a guanine extends to three the number of duplex base pairs recognized by triple-stranded DNA. Their results also indicated that the guanine amino group was involved in recognizing the T·A base pair. Consequently, they proposed a base-pairing scheme which involves a guanine that H-bonds via its H2(1) amino proton to the thymine O4 car-

† This work was supported by grants from NIH (R01 GM 37254-01) and by NSF Presidential Young Investigator Award (DMB 89-58280) with matching funds from AmGen Inc., DuPont/Merck Pharmaceuticals, Monsanto Co., and Sterling Drug Inc. to J.F. E.W. was supported in part by an NIH Predoctoral Cell and Molecular Biology training grant (GM 07185).

* Corresponding author.

0006-2960/92/0431-4838$03.00/0

bonyl of a Watson–Crick T·A base pair. However, it is possible for the guanine H2(2) amino proton to form the H-bond instead and still remain consistent with their results. Their results also suggested that the stability of the G·T·A triplet was comparable to that of the T·A·T and C^+·G·C triplets (Griffin & Dervan, 1989; Horne & Dervan, 1991). However, Hélène and co-workers using optical melting (Mergny et al., 1991) found that the G·T·A triplet was significantly less stable than either the T·A·T or C^+·G·C triplets, but still one of the more stable mismatch triplets. Work by Frank-Kamenetskii and co-workers (Belotserkovskii et al., 1990) on the effect of X·Y·Z (mismatch) triplets on triplex formation in supercoiled plasmids (H-DNA) indicated that the G·T·A triplet was much less stable than several other X·Y·Z (mismatch) triplets as well as the T·A·T and C^+·G·C triplets.

Recently, an NMR study of an intramolecular triplex containing a G·T·A triplet was reported by Patel and co-workers (Radhakrishnan et al., 1991). They demonstrated the formation of a stable triplex but were unable to determine which guanine amino proton was involved in the base-pairing scheme although they favored the H2(1) proton. Here we present a detailed report on the base-pairing scheme and stability of the G·T·A triplet contained within a pyr·pur·pyr intramolecular triplex formed from d-(AGATAGAACCCCTTCTATCTTATATCTGTGTCTT) (GTA32).

Materials and Methods

Sample Preparation. DNA samples were synthesized on an Applied Biosystems 381A DNA synthesizer and purified as previously described (Sklenář & Feigon 1990a). The DNA samples were dissolved in 400 μL of 99.996% D_2O or 400 μL of 90% H_2O/10% D_2O. The final sample conditions were 100 mM NaCl, 5 mM $MgCl_2$, and 2.3 mM DNA strand (GTA32) or 2.2 mM DNA strand (CGC32). The samples were titrated to pH 5.2 (1D spectra) or pH 5.5 (2D spectra) with NaOH and/or HCl.

NMR Spectroscopy. All NMR spectroscopy was performed on a GE GN500 500-MHz NMR spectrometer. Data was transferred to a Silicon Graphics 4D/25 and processed with FTNMR/FELIX (Hare Research) except the 1D spectra, which were processed with GEM16 (GE NMR). The one-dimensional spectra were collected at 1 °C using a $1\bar{1}$ spin-echo pulse sequence (Sklenář & Bax, 1987) with τ = 50 (CGC32) or 54 (GTA32) μs to suppress the water peak and maximize the imino and amino resonances. In the two-dimensional nuclear Overhauser effect spectroscopy (NOESY) (Macura & Ernst, 1980; Kumar et al., 1980) spectra, quadrature detection was obtained by the method of States et al. (1982). The NOESY spectra in water have a $1\bar{1}$ spin-echo read pulse with τ = 50 μs. Additionally, the residual water was subtracted with a program written by Edmond Wang utilizing a time-domain convolution routine (Marion et al., 1989). In the D_2O NOESY, the residual HDO peak was suppressed by irradiation during the recycle delay. Other NMR parameters are given in the figure captions. To examine pH and temperature effects, NOESY spectra in D_2O and H_2O were obtained at several different temperatures and pHs. In addition, a series of fast D_2O NOESYs were obtained by reducing the phase cycle (no CYCLOPS) to four and the number of scans per t_1 block to eight and by reducing the recycle delay by inserting two perpendicular spin locks and a homospoil pulse. t_2 was also reduced to 512 complex points. Mixing times of the fast NOESYs ranged from 175 to 225 ms depending on the temperature.

UV Spectrophotometry. The melting behavior of the oligonucleotides was monitored optically by absorbance at 264 nm as a function of temperature. The UV spectrophotometry was performed on a Hewlett-Packard HP8452 diode array spectrophotometer as previously described (Macaya et al., 1991). DNA samples were 1 mL with 0.5–0.7 OD (264 nm) of DNA, 50 mM sodium acetate, pH 5.2, 100 mM NaCl, and 5 mM $MgCl_2$.

Results

Imino Spectra. Figure 1 shows the one-dimensional imino region of CGC32 and GTA32 in water at pH 5.2 and 1 °C. The strong similarity of the two imino spectra indicates that GTA32 forms a triplex and that the substitution of a G·T·A triplet (GTA32) for a C^+·G·C triplet (CGC32) does not grossly alter the triplex structure. This was confirmed by analysis of the NOESY spectra. Assignments of the resonances are indicated on the spectra. The greatest difference between the two spectra occurs in the chemical shifts of the G_{28} (GTA32) and the C^+_{28} (CGC32) imino resonances. The G_{28} imino resonates upfield in the region of the non-hydrogen-bonded T iminos of the loop. Other differences include minor changes in the chemical shifts of the iminos neighboring the G·T·A triplet and the absence of a protonated cytosine amino pair (C^+_{28}) in the GTA32 oligomer.

Although GTA32 readily forms a triplex at pH 5.2, results from both NMR (imino proton spectra as a function of temperature, not shown) and optical melting studies (discussed below) indicate that GTA32 is significantly less stable than CGC32.

NOESY Spectra in Water. The 10 °C NOESY spectrum in water of GTA32 is shown in Figure 2. Assignments were obtained using techniques previously developed in our laboratory (Rajagopal & Feigon, 1989b; Sklenář & Feigon, 1990a; Macaya et al., 1991). The imino region is shown in box A and the expanded plot of this imino region at 1 °C is shown in Figure 3A. Sequential imino connectivities for the Watson–Crick strands are indicated above the diagonal, and the sequential imino connectivities for the Hoogsteen strand are indicated below the diagonal. There is a break in the Hoogsteen imino sequential connectivities at G_{28}; neither the T_{27} nor the T_{29} iminos have NOEs to the G_{28} imino. The T_4 to T_{16} Watson–Crick imino sequential connectivity has a very weak cross-peak which is only visible at a lower contour level.

Intratriplet NOE connectivities characteristic of T·A·T and C^+·G·C triplets can be seen in region B of Figure 2 and expanded in Figure 3B. These intratriplet NOEs are from Watson–Crick imino protons to adenine H2 and adenine or cytosine amino protons, as well as from Hoogsteen imino protons to adenine or guanine H8 and protonated cytosine amino protons (Sklenář & Feigon, 1990a). For brevity and clarity, they are not individually labeled except where indicated below.

The G·T·A triplet also has several intratriplet connectivities including cross-peaks from the T_4 imino proton to the A_{17} H2 proton and from the T_4 imino proton to the A_{17} amino protons (Figure 3B). The G_{28} imino proton has a cross-peak to its own amino protons (Figure 2, box D) which, in turn, have a cross-peak to the T_4 methyl protons (Figure 2, box E). The two G_{28} amino protons appear as one broad peak due to exchange broadening.

Figure 3C shows an expansion of region C of Figure 2, which contains the thymine imino to thymine methyl NOE cross-peaks. All the cross-peaks are intrabase NOEs (from a thymine imino to its own methyl) except two of the largest cross-peaks, which are from the T_{27} and T_{29} iminos to the T_4 methyl, and cross-peaks to the G_{28} imino, which has a

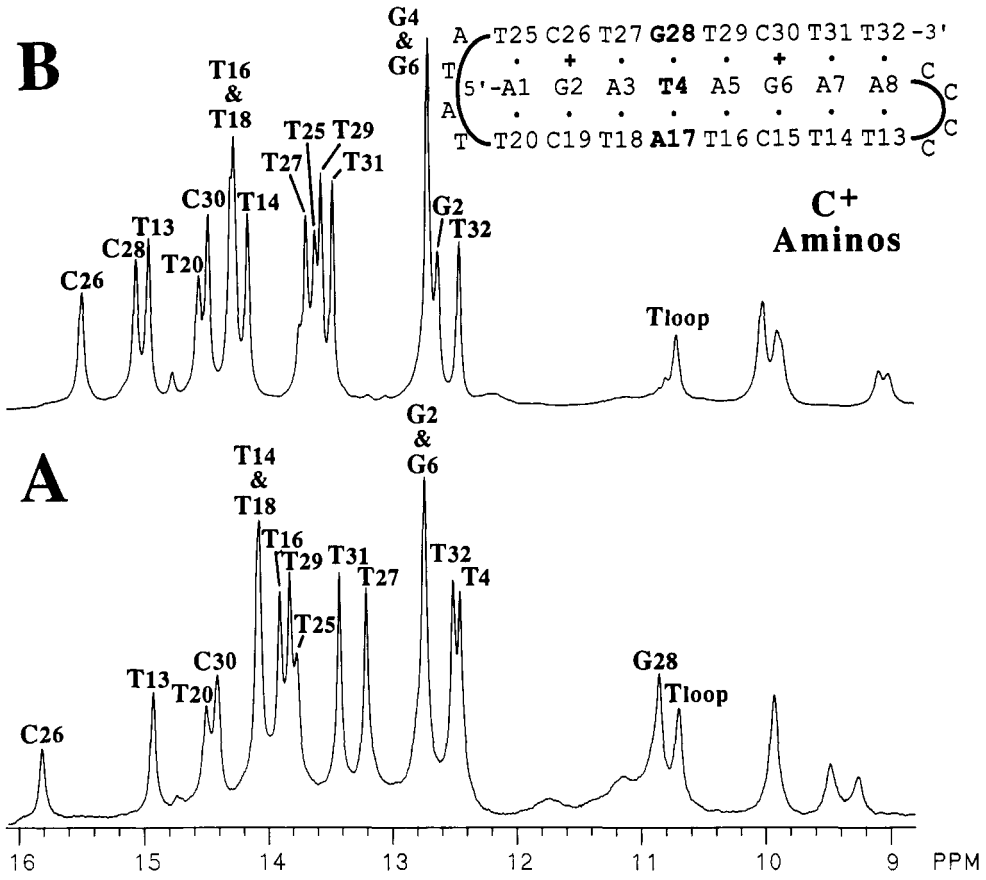

FIGURE 1: 1D 1$\bar{1}$ spin-echo ^{1}H NMR spectra of the imino and C$^+$ amino region of (A) GTA32 and (B) CGC32 in 5 mM $MgCl_2$, 100 mM NaCl, pH 5.2, 90% H_2O/10% D_2O, 2 mM strand at 1 °C. Assignments are indicated above each spectrum. A schematic of the folded triple helix with its base numbering scheme is shown at the top of the figure. The lower two strands are the Watson–Crick strands and the upper strand is the Hoogsteen third strand; 4096 complex points were collected with a sweep width of 12 048 (CGC32) or 11 364 (GTA32) Hz and 256 (CGC32) or 512 (GTA32) scans. The apodization is an exponential multiplication with a line broadening of 3 Hz.

cross-peak to the T_4 and T_{16} methyls and to the two neighboring methyls, T_{27} and T_{29}. The cross-peak from the G_{28} imino to the T_{29} methyl is very weak and can only be seen at a lower contour level than is plotted in Figure 3C.

NOESY Spectra in D_2O. Expanded regions of a 20 °C NOESY spectrum of GTA32 in D_2O at pH 5.5 are shown in Figure 4. Assignments were made as described (Sklenář & Feigon, 1990a,b; Macaya et al., 1992b) and by comparison to CGC32. The chemical shifts of the protons in the G·T·A triplet are shown along the edges of the spectrum and the cytosine H5 to H6 cross-peaks are labeled with their corresponding numbers. In Figure 4A, a strong intratriplet cross-peak from the G_{28} H1′ proton to the T_4 methyl protons is observed (peak a). This cross-peak is almost as intense as the cytosine H5–H6 cross-peaks and is important in determining the base-pairing scheme of the G·T·A triplet (see Discussion). Figure 4B shows the base to H1′ region, which has an especially strong cross-peak from G_{28} H8 to G_{28} H3′ (peak b). Also indicated is the A_{17} H2–A_5 H2 cross-peak (peak d) and an unusual cross-strand NOE, A_{17} H2–A_5 H1′ (peak c).

In order to assess the effect of pH and temperature on the structure and stability of the G·T·A triplet, a series of fast D_2O NOESYs were obtained at pH 5.5 (35, 40, and 45 °C), pH 6.0 (25, 35, and 40 °C), and pH 6.5 (20, 25, and 30 °C) (spectra not shown). The intensities of the G_{28} H1′–T_4 methyl cross-peak (Figure 4C, peak a) and the G_{28} H8–G_{28} H3′ cross-peak (Figure 4B, peak b) were compared to those obtained at 20 °C, pH 5.5. At pH 5.5, both cross-peaks are strong up to 45 °C, at which point the G_{28} H1′–T_4 methyl cross-peak decreases slightly. At pH 6.0, both cross-peaks are strong at 25 °C but decrease slightly in intensity at 35 °C. The G_{28} H1′–T_4 methyl cross-peak is no longer observable at 40 °C and the G_{28} H8–G_{28} H3′ cross-peak is broadened but still observable. At pH 6.5, there is a significant amount of the partial duplex structure. At 20 °C, both cross-peaks are fairly sharp with strong intensity, but they are broad and weak at 25 °C and above.

Table I: UV Melting Temperatures of CGC32, TAT32 and XTA32 Series Triplexes

triplet	T_{m1}[a]	T_{m2}[b]
CGC	68	
TAT	64	
GTA	55	
ATA	39[c]	54
TTA	35	56
CTA	35	53

[a] Melting transition of the third strand. CGC32, TAT32 and GTA32 melt in a single transition. The error is approximately ±2 °C. [b] Melting transition of the partial duplex. [c] Broad transition. The error is approximately ±5 °C.

UV Melting Temperatures. The melting temperature of CGC32 and related triplexes in which the central C$^+$·G·C triplet was replaced with T·A·T (TAT32), G·T·A (GTA32), or X·T·A (X = C, T, or A) triplets were determined by UV spectrophotometry as previously described (Macaya et al., 1991). In the conditions used (pH 5.2), the triplexes CGC32, TAT32, and GTA32 all melt in a single cooperative transition (Figure 5). The other X·T·A triplexes melt in separate, broad triplex and duplex transitions indicating that the Hoogsteen-base-paired strand melts before the Watson–Crick-paired duplex. The melting temperatures are given in Table I.

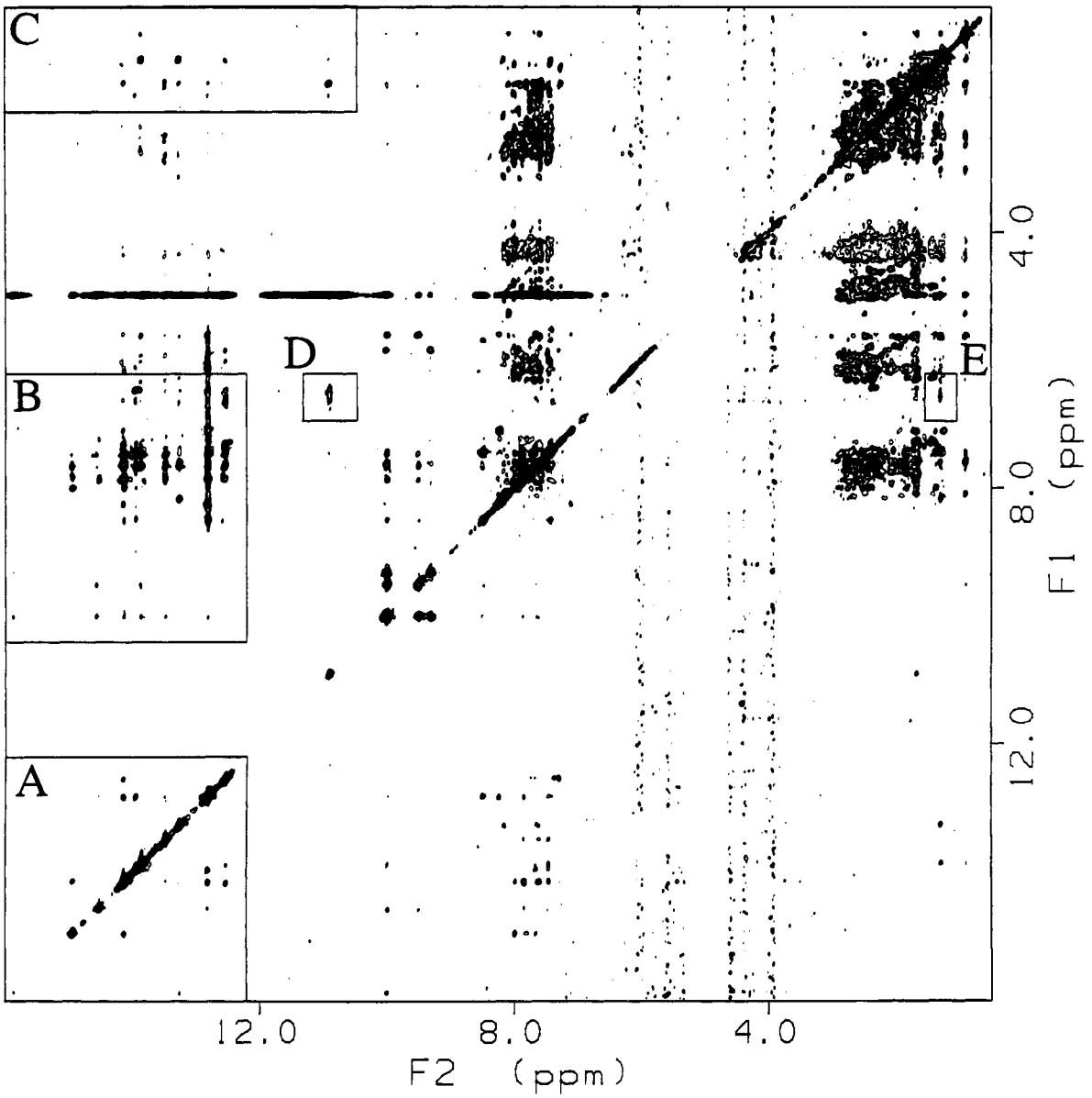

FIGURE 2: Full 2D NOESY of the GTA32 triplex in 90% H_2O/10% D_2O at 10 °C with a mixing time of 150 ms. The sample is at pH 5.5, otherwise conditions are the same as in Figure 1. Regions A, B, and C show the imino to imino, imino to aromatic, and imino to methyl regions, respectively, and are expanded in Figure 3. Region D: the G_{28} imino to G_{28} amino NOE. Region E: the G_{28} amino to T_4 methyl NOE. A total of 2048 and 445 complex points were collected in t_2 and t_1, respectively, with 32 scans per t_1 block and a sweep width of 11 364 Hz in both dimensions. The t_2 apodization was a 65°-phase-shifted sine-bell squared, and the baseline was flattened with a sixth-order polynomial. t_1 was apodized with a 75°-phase-shifted sine-bell squared, zero-filled to 2048 complex points and baseline flattened with a second-order polynomial.

DISCUSSION

Formation of a Triple-Stranded DNA Helix. Previous work in our laboratory on the CGC32 oligomer (Macaya et al., 1991, 1992b) and other DNA triplexes (Rajagopal & Feigon, 1989a,b; Sklenář & Feigon, 1990b) has demonstrated the formation of triple-stranded DNA composed of T·A·T and C^+·G·C triplets. Comparison of the 1D imino spectrum of the CGC32 triplex with that of the GTA32 oligomer (Figure 1A,B) reveals a strong resemblance, indicating that, like the CGC32 triplex, the GTA32 oligomer also forms a stable triplex. This was confirmed by analysis of NOESY spectra in D_2O. In addition to Watson–Crick imino resonances, the GTA32 spectrum shows Hoogsteen imino and protonated cytosine amino resonances which are characteristic of triplex formation. The iminos of GTA32 resonate at frequencies similar to those of the CGC32 triplex with the most notable exceptions being the iminos within the G·T·A triplet (C^+·G·C triplet, in CGC32) and its two flanking triplets. The largest difference in chemical shifts occurs between the G_{28} (GTA32) and the C^+_{28} (CGC32) imino resonances. The G_{28} imino resonance is shifted upfield into the region of the non-H-bonded T iminos in the loop, indicating that it is not H-bonded. The GTA32 T_4 imino chemical shift does not differ substantially from the CGC32 G_4 imino chemical shift but is unusual because it resonates much further upfield than the other Watson–Crick thymine iminos. This may indicate that the T_4 base has a substantially different stacking environment and that its position relative to the helical axis may be much different than that of the other Watson–Crick thymines. Other significant changes in chemical shift involve the T_{16}, T_{18}, T_{27}, and T_{29} imino resonances which are all neighboring the G·T·A triplet and are expected to have at least some change in chemical shift.

Further evidence for triplex formation is observed in Figure 3A, which shows the imino to imino NOE connectivities between both the Hoogsteen iminos and the Watson–Crick iminos. The presence of these imino connectivities confirms the overall folded structure of the triplex (Sklenář & Feigon,

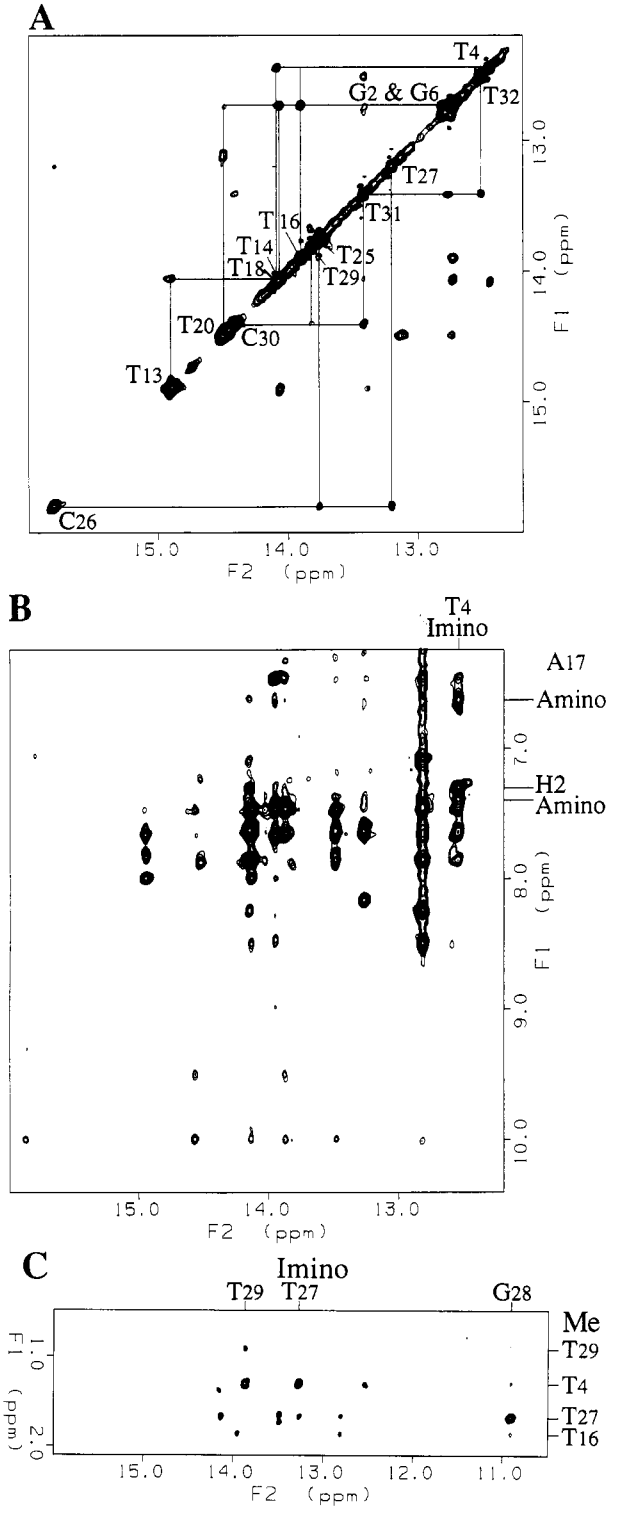

FIGURE 3: Expansion of the 2D NOESY in water. Conditions are the same as in Figure 2 except for panel A. (A) Imino to imino region at 1 °C with a mixing time of 100 ms. The Watson–Crick imino sequential connectivities are indicated above the diagonal and the Hoogsteen imino sequential connectivities are indicated below the diagonal. A total of 2048 and 456 complex points were collected in t_2 and t_1, respectively, with 64 scans per t_1 block. The t_2 apodization was a 40°-phase-shifted sine-bell, and the baseline was flattened with a sixth-order polynomial. t_1 was apodized with a 45°-phase-shifted sine-bell, zero-filled to 2048 complex points, and the baseline was flattened with a second-order polynomial. (B) Imino to aromatic/amino region. The cross-peaks between the T_4 imino proton and the A_{17} H2 and amino protons are indicated which confirm the Watson–Crick base pairing of A_{17} and T_4. Note the large chemical shift separation of the A_{17} amino protons. (C) Imino to methyl region. Selected imino to methyl cross-peaks are indicated, including two large cross-peaks from the T_4 methyl protons to the T_{27} and T_{29} imino protons and the cross-peaks from the G_{28} imino proton to the T_4, T_{16}, T_{27}, and T_{29} methyl protons.

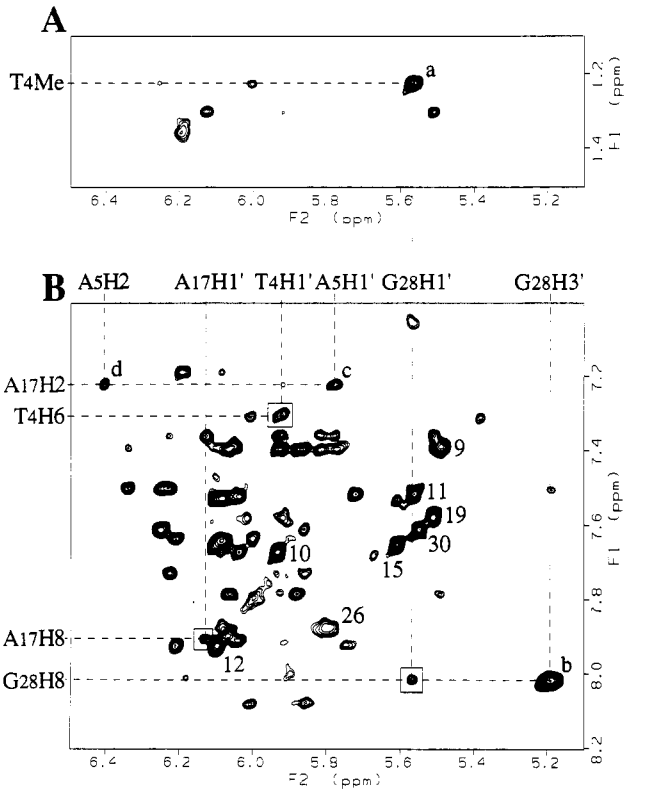

FIGURE 4: Expanded regions of the 2D NOESY of the GTA32 triplex in D_2O at 20 °C showing (A) H1′,H5-methyl cross-peaks and (B) H1′,H5–aromatic cross-peaks. The chemical shifts of the protons in the G·T·A triplet are shown along the edges of the spectrum (also A_5 H1′ and A_5 H2). The CH5–CH6 cross-peaks are labeled with their corresponding numbers. Boxes surround intranucleotide base (H6,H8)–H1′ cross-peaks in the G·T·A triplet. The spectrum was collected with a mixing time of 200 ms and a sweep width of 5714 Hz in both dimensions. A total of 1024 and 450 complex points were collected in t_2 and t_1, respectively, with 32 scans per t_1 block. The t_2 apodization was a 30°-phase-shifted sine-bell, zero-filled to 2048 complex points, and the baseline was flattened with a first-order polynomial. t_1 was apodized with a 30°-phase-shifted sine-bell, zero-filled to 2048 complex points, and the baseline was flattened with a second-order polynomial.

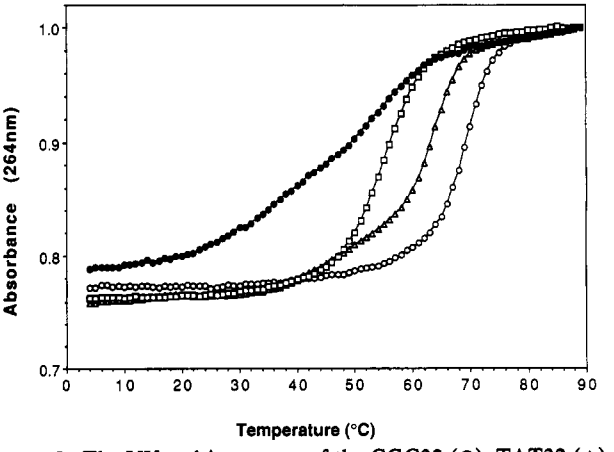

FIGURE 5: The UV melting curves of the CGC32 (○), TAT32 (△), GTA32 (□), and CTA32 (●) triplexes monitored at 264 nm are shown. The melting curves were normalized to 1 at 89 °C. In the CGC32, TAT32, and GTA32 triplexes, the binding of the third strand stabilizes the duplex resulting in a single, cooperative transition. In contrast, the CTA32 triplex melts in two broad transitions; the first transition is the third strand melting from the duplex and the second transition is the duplex melting. The TTA32 and the ATA32 triplexes have similar melting curves as the CTA32 triplex (data not shown).

1990a). Additionally, intratriplet NOE connectivities beween the imino protons and the amino and aromatic protons within

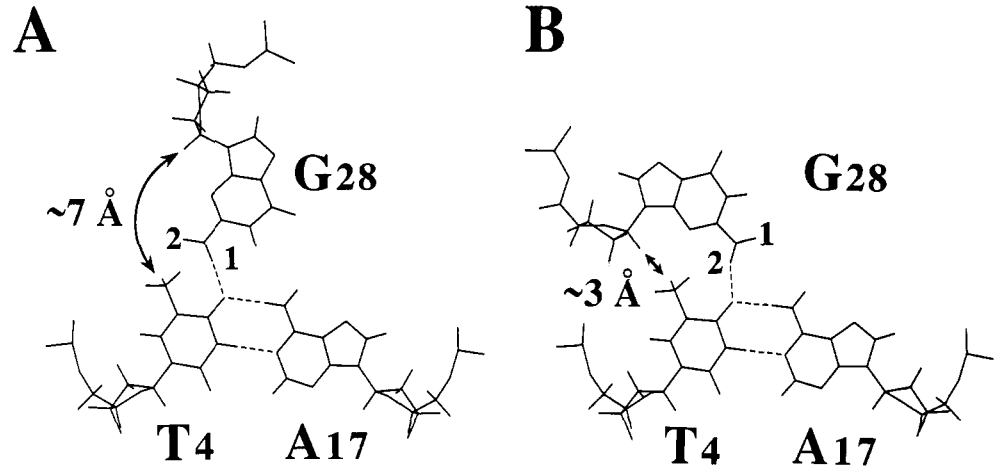

FIGURE 6: Two possible G·T·A base-pairing schemes. (A) G_{28} forms a H-bond to the T_4 O4 carbonyl with its H2(1) amino proton. The G_{28} H1′ to T_4 methyl distance is ~7 Å. (B) G_{28} forms a H-bond to the T_4 O4 carbonyl with its H2(2) amino proton. The G_{28} H1′ to T_4 methyl distance is ~3 Å.

the Watson–Crick and Hoogsteen base pairs of all the T·A·T and C^+·G·C triplets in the GTA32 triplex are observed (Figure 3B), confirming the formation of the T·A·T and C^+·G·C triplets (Sklenář & Feigon, 1990a). Furthermore, cross-peaks from the imino protons of the pyrimidine third strand to the H2′,2″ protons of the purine base of the 5′ neighboring triplet are also observed (Figure 2). Such cross-peaks are never seen in duplex DNA and are unique to triple-stranded DNA (Macaya et al., 1991, 1992b). Two strong imino to methyl NOEs are observed which are absent in the CGC32 triplex (Figure 3C). These cross-peaks are from the T_{27} and T_{29} iminos to the T_4 methyl and indicate that the T_{27} and T_{29} bases are binding in the major groove near the T_4 base, as expected. Together, these observations clearly establish a triple-helical structure for the GTA32 oligomer and suggest that this structure is similar to the CGC32 triplex except in the region of the G·T·A triplet and its two neighboring triplets.

Formation of the G·T·A Triplet. The G·T·A base-pairing scheme proposed by Griffin and Dervan (1989) is illustrated in Figure 6A. There are several important components of the G·T·A triplet structure, one of which is a standard Watson–Crick T·A base pair. The experimental evidence for a Watson–Crick T·A base pair within the G·T·A triplet includes NOEs from the T_4 imino proton to the A_{17} amino protons and an NOE from the T_4 imino proton to the A_{17} H2 proton (Figure 3B). These NOEs indicate that the T_4 imino proton is within 5 Å of the A_{17} amino and H2 protons and is probably H-bonded to the A_{17} N3 creating a Watson–Crick T·A base pair. An important difference between the T·A base pair in the G·T·A triplet and the A·T base pairs in the T·A·T triplets is that the adenine amino group of the G·T·A triplet has only a single H-bond while both adenine amino protons in a T·A·T triplet are H-bonded. This difference is apparent in the large chemical shift separation of the two A_{17} amino protons (0.77 ppm at 10 °C), indicating that only the downfield amino proton is involved in a H-bond. In contrast, the other adenine amino groups, with the exception of the terminal adenine, have a small chemical shift separation (0.09–0.22 ppm) and are downfield shifted indicating that both amino protons are involved in H-bonds.

Another feature of the G·T·A base-pairing scheme is the binding of the guanine to the T·A base pair in the major groove by a single H-bond involving the guanine amino group and the thymine O4 carbonyl, leaving the guanine imino free. Consistent with this scheme is the observation that the G_{28} imino resonates in the region of the non-H-bonded T iminos in the loop, upfield of the H-bonded iminos (Figures 1A and 2). Additionally, the G_{28} imino has a cross-peak to its own aminos (Figure 2, region D). Since the guanine aminos normally exchange quickly with each other or with solvent, they are usually not observable when free. Therefore, the fact that we can observe a broad cross-peak from the G_{28} aminos to the G_{28} imino strongly suggests the aminos are involved in a H-bond. Finally, there is an NOE cross-peak from the G_{28} aminos to the T_4 methyl (Figure 2, region E) indicating that the G_{28} aminos are H-bonded near the T_4 methyl, most probably to the O4 carbonyl. Together, these observations confirm that T_4 and A_{17} form a Watson–Crick base pair and that G_{28} H-bonds, via its amino group, to the T_4 O4 carbonyl. We are unable to observe an expected cross-peak from the G_{28} aminos to the A_{17} aminos because it would appear in a region close to the excitation null in the $1\bar{1}$ spin-echo water NOESY.

Base-Pairing Scheme of the G·T·A Triplet. The results presented so far are consistent with the G·T·A base-pairing scheme proposed by Griffin and Dervan (1989) (Figure 6A). However, the results are also consistent with the alternate base-pairing scheme shown in Figure 6B. These two schemes differ only by the guanine amino proton involved in the H-bond, but this difference radically alters the position and orientation of the guanine nucleotide within the helical structure and affects the stability of the G·T·A triplet. Results presented recently by Patel and co-workers (Radhakrishnan et al., 1991) on a different G·T·A-containing triplex are consistent with the data discussed above. However, they could not unambiguously distinguish between the two the G·T·A base-pairing schemes, but favored scheme A. We present conclusive evidence below in support of scheme B for the G·T·A base triplet in GTA32.

One of the consequences of the base-pairing structure in Figure 6B is to position the G_{28} sugar close to the T_4 methyl. Experimentally, we observe a strong cross-peak from the G_{28} H1′ to the T_4 methyl (Figure 4A, peak a). By comparison of the integrated intensity of this cross-peak to that of the CH5–CH6 cross-peaks (Figure 4B) which have a fixed distance of 2.44 Å, we can estimate this distance to be approximately 3–4 Å. The distance from the G_{28} H1′ to the T_4 methyl in scheme A would be approximately 7 Å, which is too long to give rise to an NOE. However, the interproton distance in scheme B would be about 3 Å, approximately the distance estimated from the NOE cross-peak intensity. Weaker cross-peaks from other G_{28} sugar protons (H2′,2″) to the T_4 methyl protons are also observed (not shown), which further

supports scheme B. Scheme B would also place the G_{28} imino much closer to the T_{27} methyl and further from the T_{29} methyl. This is confirmed by the large difference in NOE intensity between the G_{28} imino–T_{27} methyl cross-peak and the G_{28} imino–T_{29} methyl cross-peak (Figure 3C). Scheme A would be expected to result in two cross-peaks of more similar intensity. A structural advantage of scheme B is that it places the G_{28} sugar closer to the equivalent position of the C^+_{28} sugar of the CGC32 triplex, likely causing less distortion in the sugar–phosphate backbone, leading to a greater stability. Interestingly, Hélène and co-workers using energy minimization calculations on model triplexes (Mergny et al., 1991) also obtained the G·T·A base pairing of scheme B, although it is not clear whether their calculations ever sampled the conformational space of scheme A.

It should be noted that the G·T·A base pairing illustrated in Figure 6B is depicted with a less than optimal angle for the G_{28} amino H-bond. An optimal H-bond angle would require a greater rotation of the G_{28} base plane and would place the G_{28} H1′ approximately 2.5 Å from the T_4 methyl, which is closer than we observe. This would also place the G_{28} H1′ near to the T_4 H6, but this cross-peak is absent, indicating either that the G_{28} H-bond is not at an ideal angle or, possibly, that motional averaging of NOEs from the G_{28} base is occurring.

In the D_2O NOESY spectrum there is an unusually strong H3′ cross-peak (Figure 4B, peak b). This cross-peak is an NOE from the G_{28} H3′ to its own H8 proton. The intensity of this cross-peak is comparable in intensity to the cytosine H5–H6 cross-peaks which means the distance between these protons is only 2–3 Å. Furthermore, the G_{28} H8 to G_{28} H2′ and H2″ cross-peaks (not shown) are much weaker than normally observed for S-type sugar puckers. The only explanation for these distances is if the G_{28} sugar adopts a C3′-endo pucker (Wüthrich, 1986). Additional support for this conclusion is the observation of C3′-endo cross-peak patterns for the G_{28} H1′–H2′,2″ and H2′,2″–H3′ cross-peaks in a P.COSY (Marion & Bax, 1988) spectrum (spectrum not shown). Model building also correlates the C3′-endo sugar pucker with a rotation of the G_{28} base plane into the same orientation as that of scheme B. Alternatively, a C2′-endo sugar pucker favors the base pairing of scheme A. In other triplexes studied in our laboratory, we found that all the sugars except the sugars of the protonated cytosines in the third strand are predominantly S-type (near C2′-endo) (Macaya et al., 1992a). Since the G_{28} in GTA32 replaces a C^+, we cannot determine whether the G_{28} N-type sugar pucker is a result of the stress from a purine in an otherwise pyrimidine backbone or if it is a result of its position in the sequence of the triple helix.

Fast NOESY spectra of GTA32 were also obtained in order to examine the structure of the G·T·A triplet as a function of pH and temperature. The observation of the G_{28} H1′–T_4 methyl and G_{28} H3′–G_{28} H8 cross-peaks in these spectra (discussed above) confirm that the G·T·A triplet base pairs as in scheme B and that the G_{28} sugar remains C3′-endo over a wide range of pH and temperature. As expected, the third strand becomes less stable as pH is increased due to the requirement of protonation of the cytosines. However, the structure of the G·T·A triplet maintains the base pairing of scheme B up to the point where the entire third strand begins to melt.

Stacking and Stability of the GTA32 Triplex. An important question concerns what we can conclude about the effect on the stability and structure of the triple helix caused by the inclusion of a G·T·A triplet. A purine guanine, in an otherwise pyrimidine third strand, is likely to cause a substantial distortion in the helical structure. As discussed above, the G_{28} sugar possibly relieves some of the strain by adopting an N-type (near C3′-endo) sugar pucker which positions its sugar in a more favorable location within the phosphate backbone. This positions the G_{28} nucleotide so that the H2(2) amino proton will form an H-bond instead of the H2(1) amino proton. Although the *position* of the G_{28} sugar is similar to that of the CGC32 C^+_{28} sugar, the *orientation* of the G_{28} sugar is significantly different. If we visualize the triple helix as a cylinder, then the C^+_{28} glycosidic bond is roughly perpendicular to the cylinder's surface. However, the G_{28} glycosidic bond is almost tangent to the cylinder's surface.

This alternate orientation of the G_{28} sugar could be expected to cause a local distortion in the helix such as a slide or an unusual helical twist at the T_{27}–G_{28} and G_{28}–T_{29} steps. There are several NOE cross-peaks that suggest that the helix parameters in the vicinity of the G·T·A triplet may, indeed, be unusual. As previously noted, the C_{28} imino has cross-peaks to several methyl groups (Figure 3C) including the T_{27} and T_{29} methyl groups. The T_{27} methyl cross-peak is much larger than the T_{29} methyl cross-peak, which has been discussed above as evidence for the G·T·A base pairing of scheme B. However, the T_{27} methyl cross-peak is so much larger than the T_{29} methyl cross-peak that it also suggests that the G·T·A triplet has an unusual twist or slide. The T_4–T_{16} imino sequential NOE (Figure 3A) is extremely weak when it is compared to the other imino sequential connectivities and also suggests unusual helix geometry between the G_{28} and T_{29} triplets. Furthermore, cross-peaks are observed from the A_5 H2 and A_5 H1′ protons to the A17 H2 proton (Figure 4B, peaks c and d) indicating that these protons are closer than normally observed in A- or B-DNA (Wüthrich, 1986). The A_5 H1′–A_{17} H2 cross-peaks is especially unusual considering that cross-strand H2–H1′ cross-peaks in A- or B-DNA are normally weak or absent. These observations suggests that the G_{28}·T_4·A_{17} triplet is shifted within the helix such that the A_{17} base stacks more directly over the A_5 base of the neighboring T_{29}·A_5·T_{16} triplet. From model building, this is the expected direction of a slide or twist in order to relieve the strain caused by the unusual orientation of the G_{28} sugar. Energy minimization calculations performed by Hélène and co-workers (Mergny et al., 1991) suggest that the G in the G·T·A triplet may tilt and form a second H-bond from its H2(1) amino proton to a Watson–Crick thymine O4 carbonyl of a neighboring T·A·T triplet. Their model predicts that we should observe strong NOEs from the G_{28} imino and/or amino protons to the T_{16} or T_{18} methyl protons. However, we do not observe these cross-peaks except for a weak G_{28} imino to T_{16} methyl cross-peak (Figures 2 and 3C) which may exist even if the G is not tilted.

Another important concern is how the G·T·A triplet affects the stability of a pyr·pur·pyr triplex. From optical melting studies on CGC32 and the related triplexes, we found that CGC32, TAT32, and GTA32 melt at 68, 64, and 55 °C, respectively, at *pH 5.2* and have only a single melting transition (Figure 5 and Table I), while the melting temperature of triplexes containing the other X·Y·Z triplets range from 35 to 45 °C and have separate triplex and duplex melting transitions (Macaya et al., 1991; S. Malek and J. Feigon, unpublished results). These results indicate that within a pyr·pur·pyr triplex, the G·T·A triplet is significantly more stable than other noncanonical triplets but is also significantly less stable than either T·A·T or C^+·G·C. Dervan and coworkers (Griffin & Dervan, 1989; Horne & Dervan, 1991)

also concluded that a triplex containing the G·T·A triplet was significantly more stable than those with any other noncanonical X·Y·Z triplet, resulting in sequence-specific recognition of T·A base pairs by G within a predominantly pyr·pur·pyr triple helix, although their results indicated that the G·T·A triplet was comparable in stability to the T·A·T and C^+·G·C triplets.

The sequences of the neighboring triplets, as well as other factors, apparently affect the stability of triplexes containing a G·T·A triplet. In both this study and that of Griffin and Dervan (1989), the triplets flanking the G·T·A were both T·A·T. In a related triplex where the G·T·A triplet was flanked by C^+·G·C triplets, we obtained a mixture of both duplex and triplex conformations and were unable to push the equilibrium to the triplex conformation only (E. Wang and J. Feigon, unpublished results). Preliminary studies on the UV melting of intramolecular triplexes containing a G·T·A triplet flanked by one or two C^+·G·C triplets indicate that the stability of the triplex decreases as the number of C^+·G·C neighbors increases (S. Malek and J. Feigon, unpublished results). In the study by Frank-Kamenetskii and co-workers (Belotserkovskii et al., 1990) where the formation of H-DNA was assayed by two-dimensional gel electrophoresis, the X·Y·Z triplet was flanked by one C^+·G·C (5′G) and one T·A·T (3′A) triplet. Under their conditions, which were also at lower pH (4.2) than this study and that of Griffin and Dervan (1989), the triplex (H-DNA) containing the G·T·A was one of the most difficult to form (required higher superhelical density). Roberts and Crothers (1991) found that the destabilization caused by a G·T·A triplet (flanked by a T·A·T (5′A) and a C^+·G·C (3′G) triplet) within a pyr·pur·pyr triplex was about the same as other mismatches (3.2–4.0 kcal/mol). Similarly, work by Hélène and co-workers (Mergny et al., 1991) on the UV melting of triplexes containing mismatches indicated that, although the G·T·A triplet (flanked by a T·A·T (5′A) and a C^+·G·C (3′G) triplet) was one of the most stable mismatches, it was not *the most* stable mismatch and was significantly less stable than the T·A·T and C^+·G·C triplets. Interestingly, they found that the G·G·C triplet was more stable than the G·T·A triplet, which suggests that a G in the third strand might preferentially recognize a G·C base-pair depending on the flanking triplets. Thus, it appears that the G·T·A triplet is significantly more stable than other X·Y·Z triplets within a pyr·pur·pyr triplex only when it is flanked by T·A·T triplets, as in our GTA32 triplex. However, it is important to note that other factors, such as pH and ionic strength, may be contributing to the relative stabilities of the triplexes discussed above. The effect of pH on the overall melting temperature of the triplexes is expected to vary depending on the C^+·G·C content, since low pH will stabilize the C^+·G Hoogsteen pairing but will destabilize Watson–Crick pairing.

Interestingly, in the oligonucleotides we studied, the G·T·A triplet is more stable than the A^+·G·C triplet (T_m = 44 °C), although A^+·G·C has one more Hoogsteen hydrogen bond than G·T·A (Macaya et al., 1991). The A^+·G·C triplet is possibly less stable both because of the requirement for protonation at N_3 and because the *position* of the sugar would cause a distortion in the sugar–phosphate backbone. We also found that the G·T·A triplet is the most stable of all the X·T·A triplets, where X = A, C, G, or T (Table I). This is perhaps surprising since a cytosine could potentially H-bond to a T·A base pair in the same orientation as a guanine except that the cytosine glycosidic bond would be perpendicular to the "cylinder" surface and would presumably be more stable. However, the C·T·A triplet is one of the more unstable triplets both in our system and in the assay system used by Dervan and co-workers (Griffin & Dervan, 1989; Horne & Dervan, 1991), where the flanking triplets are T·A·T, and in the system used by Hélène and co-workers (Mergny et al., 1991), where the flanking triplets are a T·A·T (5′G) and a C^+·G·C (3′A). In contrast, the C·T·A triplet is substantially more stable than the G·T·A triplet in the assay used by Frank-Kamenetskii and co-workers (Belotserkovskii et al, 1990), where the flanking triplets are a C^+·G·C (5′G) and a T·A·T (3′A). This further suggests that stacking interactions play an important role in the stability of noncanonical triplets in DNA triplexes. This, as well as pH, base content, and ionic conditions, will contribute to the stability of triplexes containing alternative triplets (Macaya et al., 1991) and needs to be considered in assaying the effectiveness of nucleotide analogues designed to form stable DNA triplexes.

Added in Proof

A more recent paper by Radhakrishnan et al. (1992) reports the same base-pairing scheme for the G·T·A triplet as that shown here.

Acknowledgments

We thank Karl Koshlap for synthesizing and purifying the DNA oligonucleotides used in this study.

References

Arnott, S., & Selsing, E (1974) *J. Mol. Biol. 88*, 509–521.
Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobrynin, V. N., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1990) *Nucleic Acids Res. 18*, 6621–6624.
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. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.
François, J.-C., Saison-Behmoaras, T., Barbier, C., Chassignol, M., Thuong, N. T., & Hélène, C. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9702–9706.
François, J.-C., Saison-Behmoaras, T., Thuong, N. T., & Hélène, C. (1989b) *Biochemistry 28*, 9617–9619.
Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967–971.
Hanvey, J. C., Shimizu, M., & Wells, R. D. (1990) *Nucleic Acids Res. 18*, 157–161.
Hoogsteen, K. (1959) *Acta Crystallogr. 12*, 822–823.
Horne, D. A., & Dervan, P. B. (1991) *Nucleic Acids Res. 19*, 4963–4965.
Hsieh, P., Camerini-Otero, C. S., & Camerini-Otero, R. D. (1990) *Genes Dev. 4*, 1951–1963.
Htun, H., & Dahlberg, J. E. (1989) *Science 243*, 1571–1576.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.
Macaya, R. F., Gilbert, D. E., Malek, S., Sinsheimer, J., & Feigon, J. (1991) *Science 254*, 270–274.
Macaya, R. F., Schultze, P., & Feigon, J. (1992a) *J. Am. Chem. Soc. 114*, 781–783.
Macaya, R. F., Wang, E., Schultze, P., Sklenář, V., & Feigon, J. (1992b) *J. Mol. Biol.* (in press).
Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95.
Maher, L. J., III, Wold, B., & Dervan, P. B. (1989) *Science 245*, 725–730.
Marion, D., & Bax, A. (1988) *J. Magn. Reson. 80*, 528–533.
Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425–430.

Mergny, J.-L., Sun, J.-S., Rougee, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J., & Hélène, C. (1991) *Biochemistry 30*, 9791-9798.
Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1987) *Nature 330*, 495-497.
Mooren, M. M. W., Pulleyblank, D. E., Wijmenga, S. S., Blommers, M. J. J., & Hilbers, C. W. (1990) *Nucleic Acids Res. 18*, 6523-6529.
Parniewski, P., Kwinkowski, M., Wilk, A., & Klysik, J. (1990) *Nucleic Acids Res. 18*, 605-611.
Perrouault, L., Asseline, U., Rivalle, C., Thuong, N. T., Bisagni, E., Giovannangeli, C., Le Doan, T., & Hélène, C. (1990) *Nature 344*, 358-360.
Pilch, D. S., Levenson, C., & Shafer, R. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1942-1946.
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. (1989a) *Nature 339*, 637-640.
Rajagopal, P., & Feigon, J. (1989b) *Biochemistry 28*, 7859-7870.
Roberts, R. W., & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397-9401.
Sklenář, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469-479.
Sklenář, V., & Feigon, J. (1990a) *Nature 345*, 836-838.
Sklenář, V., & Feigon, J. (1990b) *J. Am. Chem. Soc. 112*, 5644-5645.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286-292.
Strobel, S. A., & Dervan, P. B. (1991) *Nature 350*, 172-174.
Strobel, S. A., Moser, H. E., & Dervan, P. B. (1988) *J. Am. Chem. Soc. 110*, 7927-7929.
Weinreb, A., Collier, D. A., Birshtein, B. K., & Wells, R. D. (1990) *J. Biol. Chem. 265*, 1352-1359.
Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlrab, F. (1988) *FASEB J. 2*, 2939-2949.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.

Fluorescence Resonance Energy Transfer Analysis of the Structure of the Four-Way DNA Junction†

Robert M. Clegg,*,‡ Alastair I. H. Murchie,§ Annelies Zechel,‡ Carsten Carlberg,‡ Stephan Diekmann,‡ and David M. J. Lilley§

Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Postfach 2841, D-W-3400 Göttingen, Federal Republic of Germany, and Department of Biochemistry, The University, Dundee DD1 4HN, U.K.

Received May 14, 1991; Revised Manuscript Received December 9, 1991

ABSTRACT: We have carried out fluorescence resonance energy transfer (FRET) measurements on four-way DNA junctions in order to analyze the global structure and its dependence on the concentration of several types of ions. A knowledge of the structure and its sensitivity to the solution environment is important for a full understanding of recombination events in DNA. The stereochemical arrangement of the four DNA helices that make up the four-way junction was established by a global comparison of the efficiency of FRET between donor and acceptor molecules attached pairwise in all possible permutations to the 5′ termini of the duplex arms of the four-way structure. The conclusions are based upon a comparison between a series of many identical DNA molecules which have been labeled on different positions, rather than a determination of a few absolute distances. Details of the FRET analysis are presented; features of the analysis with particular relevance to DNA structures are emphasized. Three methods were employed to determine the efficiency of FRET: (1) enhancement of the acceptor fluorescence, (2) decrease of the donor quantum yield, and (3) shortening of the donor fluorescence lifetime. The FRET results indicate that the arms of the four-way junction are arranged in an antiparallel stacked X-structure when salt is added to the solution. The ion-related conformational change upon addition of salt to a solution originally at low ionic strength progresses in a continuous noncooperative manner as the ionic strength of the solution increases. The mode of ion interaction at the strand exchange site of the junction is discussed.

Recombination between DNA molecules is of great biological importance. Homologous recombination in chromosomes of diploid organisms generates reassortment of genes that allows the best combinations to be selected and is thus of enormous significance in evolution. Site-specific recombination systems are involved in many DNA rearrangements from the integration of bacteriophage genomes to the generation of antibody diversity.

Most models for homologous genetic recombination postulate a four-way (Holliday) junction in DNA as the central intermediate (Holliday, 1964; Meselson & Radding, 1975; Orr-Weaver et al., 1981) that must undergo subsequent cleavage (resolution) to recreate unconnected duplex molecules. Early attempts to model the molecular geometry (Calascibetta et al., 1984; Sigal & Alberts, 1972; Sobell, 1972) all involved

† We are indebted to NATO for a travel grant and the Wellcome Trust and the C.R.C. for financial support.
* Author to whom correspondence should be addressed.
‡ Max Planck Institute for Biophysical Chemistry.
§ The University, Dundee.

0006-2960/92/0431-4846$03.00/0