NMR Structural Studies of Intramolecular $(Y+)_n\cdot(R+)_n(Y-)_n$ DNA Triplexes in Solution: Imino and Amino Proton and Nitrogen Markers of G·TA Base Triple Formation†

Ishwar Radhakrishnan,‡ Xiaolian Gao,§ Carlos de los Santos,‡ David Live,∥ and Dinshaw J. Patel*,‡

Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032, Structural and Biophysical Chemistry Department, Glaxo Research Institute, Research Triangle Park, North Carolina 27709, and Chemistry Department, California Institute of Technology, Pasadena, California 91125

Received May 21, 1991; Revised Manuscript Received July 3, 1991

ABSTRACT: We reported previously on NMR studies of $(Y+)_n\cdot(R+)_n(Y-)_n$ DNA triple helices containing one oligopurine strand $(R)_n$ and two oligopyrimidine strands $(Y)_n$ stabilized by T·AT and C^+·GC base triples [de los Santos, C., Rosen, M., & Patel, D. J. (1989) *Biochemistry 28*, 7282-7289]. Recently, it has been established that guanosine can recognize a thymidine·adenosine base pair to form a G·TA triple in an otherwise $(Y+)_n\cdot(R+)_n(Y-)_n$ triple-helix motif. [Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967-971]. The present study extends the NMR research to the characterization of structural features of a 31-mer deoxyoligonucleotide that folds intramolecularly into a 7-mer $(Y+)_n\cdot(R+)_n(Y-)_n$ triplex with the strands linked through two T_5 loops and that contains a central G·TA triple flanked by T·AT triples. The G·TA triplex exhibits an unusually well resolved and narrow imino and amino exchangeable proton and nonexchangeable proton spectrum in H_2O solution, pH 4.85, at 5 °C. We have assigned the imino protons of thymidine and amino protons of adenosine involved in Watson–Crick and Hoogsteen pairing in T·AT triples, as well as the guanosine imino and cytidine amino protons involved in Watson–Crick pairing and the protonated cytidine imino and amino protons involved in Hoogsteen pairing in C^+·GC triples in the NOESY spectrum of the G·TA triplex. The NMR data are consistent with the proposed pairing alignment for the G·TA triple where the guanosine in an anti orientation pairs through a single hydrogen bond from one of its 2-amino protons to the 4-carbonyl group of thymidine in the Watson–Crick TA pair. We detect a set of NOEs between adjacent triples that establishes that the G·TA triple stacks between flanking T·AT triples in the G·TA triplex. The imino protons of the G·TA triplex in H_2O, pH 4.85, broaden sequentially between 32 and 42 °C with the thymidine imino protons in the central G·TA triple most stable to increasing temperature. The intramolecular G·TA triplex is stable at pH 5.0 and 5 °C but converts fully to the Watson–Crick hairpin duplex on raising the pH to 6.5 at this temperature. The assigned guanosine and thymidine imino protons have been linked to their attached ring nitrogens in the G·TA triplex from an analysis of the proton-detected natural abundance nitrogen–proton two-dimensional correlation spectrum. The guanosine N1 nitrogen in the G·TA triple resonates 10 ppm to high field of the loop thymidine N3 nitrogens in contrast to their attached imino protons, which exhibit similar chemical shifts. These results demonstrate the capabilities of the NMR approach in monitoring individual base triples and their pairing alignments, as well as establishing that the G·TA triple can be readily accommodated in an otherwise intramolecular $(Y+)_n\cdot(R+)_n(Y-)_n$ triple helix in solution.

The facile formation of DNA triple-helical structures involving one deoxypurine and two deoxypyrimidine strands was established over three decades ago (Felsenfeld et al., 1957). The proposed T·AT **1** and C^+·GC **2** base triple-pairing alignments in DNA triplexes at acidic pH (see Chart I) were based on an interpretation of X-ray fiber diffraction data (Arnott & Selsing, 1974) using molecular modeling with optimized stereochemistry (Arnott et al., 1976). In this model of the $(Y+)_n\cdot(R+)_n(Y-)_n$ triplex, the $(R+)_n$ and $(Y-)_n$ strands are aligned through Watson–Crick hydrogen bonding and the third $(Y+)_n$ strand is parallel to and aligns with $(R+)_n$ through Hoogsteen hydrogen bonding in the major groove (Lee et al., 1979; Moser & Dervan, 1987). (R and Y stand for purine and pyrimidine, respectively, while the signs represent the directionality of the strands).

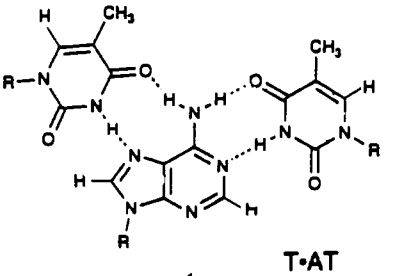
Interest in triple-stranded DNA languished until the early 1980's when it was discovered that $(R+)_n\cdot(Y-)_n$ segments in closed circular DNA were susceptible to single-stranded nucleases under conditions of superhelical stress (Larson & Weintraub, 1982). This was attributed to $(Y+)_n\cdot(R+)_n(Y-)_n$ triple-helix formation with the single-stranded nuclease sensitivity originating in the unpaired $(R-)_n$ strand (Christophe et al., 1985; Lyamichev et al., 1986; Mirkin et al., 1987). Triple-helical formation in $(R+)_n\cdot(Y-)_n$ mirror repeats under either superhelical stress and/or acidic pH can be readily monitored by two-dimensional gel electrophoresis and chemical probes [reviewed in Wells et al. (1988), Htun and Dahlberg (1989), and Broitman and Fresco (1989)]. Extensive studies have been reported on the effects of base composition, loop size, and the inclusion of interruptions on intramolecular DNA triplex formation in supercoiled plasmids (Hanvey et al., 1989;

† This research was supported by NIH Grant GM34504 to D.P. and NIH Grant DK38676 to D.L. C.d.l.S. was supported by NIH Fogarty Fellowship 5F05 TW04169. 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. Purchase of a GN-500 spectrometer at Emory University was supported in part by NSF Grant DMB8604304 to D.L.

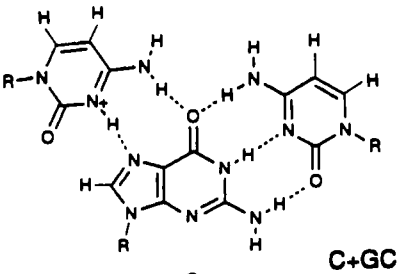
* To whom correspondence should be addressed.
‡ Columbia University.
§ Glaxo Research Institute.
∥ California Institute of Technology.

0006-2960/91/0430-9022$02.50/0

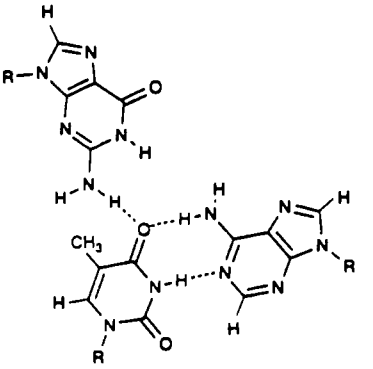
Chart I



1

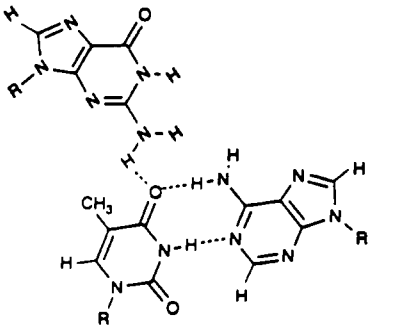


2



3a

G·TA



3b

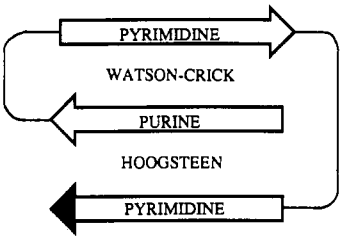
Shimuzu et al., 1989; Collier & Wells, 1990).

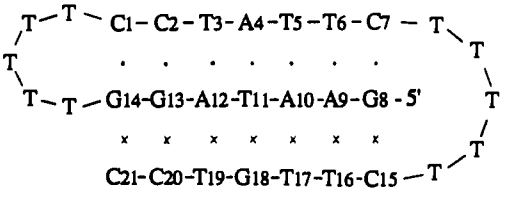
Further interest in triple-helix formation grew out of efforts directed toward sequence-specific recognition of duplex DNA by complementary third-strand probes (Moser & Dervan, 1987; Praseuth et al., 1988). There have been major developments in the design of high-affinity sequence-specific probes of the right-handed DNA helix (Helene & Thuong, 1988; Posvic & Dervan, 1989; Sun et al., 1989; Leubke & Dervan, 1989; Horne & Dervan, 1990; Perronault et al., 1990) that have culminated in oligonucleotide probes that can recognize and cleave duplex DNA at a single site at the chromosomal level (Strobel & Dervan, 1990, 1991).

In addition to the standard T·AT **1** and C^+·GC **2** triples, Griffin and Dervan (1989) recently established formation of a stable G·TA triple in an otherwise $(Y+)_n·(R+)_n(Y-)_n$ triplex. The two proposed pairing alignments of the G·TA triple are shown in Chart I, **3a** and **3b**, with the Hoogsteen alignment stabilized by a single NH_2–N hydrogen bond (Griffin & Dervan, 1989). This result has great potential since it expands on the base triple code and should permit binding of oligonucleotide probes to mixed sites on duplex DNA.

Previous NMR structural studies have focused on DNA triplexes generated from three separate strands to define the nature of T·AT and C^+·GC triple formation (de los Santos et al., 1989; Rajagopal & Feigon, 1989). These structural studies have been complemented by calorimetric measurements that define the energetics of the thermally induced order–disorder transition of DNA triplexes (Plum et al., 1990). An alternate approach results in intramolecular triplex formation from the folding of a DNA single strand where the chains in the triple helix are linked by pyrimidine hairpin loop segments (Haner & Dervan, 1990; Sklenar & Feigon, 1990). Our NMR studies on the centrally positioned G·TA triple flanked by T·AT triples were undertaken on triplex **4** (Chart II) (henceforth designated the G·TA triplex) based on a 31-mer deoxyoligonucleotide sequence. Following Haner and Dervan (1990), the linking hairpin loops were composed of $(T)_n$ segments with $n = 5$ in the current NMR study of the G·TA triplex **4**.

Chart II





4

The sequence of the helical segment of the G·TA triplex **4** generated from a single strand is the same as that of the three-stranded 7-mer triplex studied previously (de los Santos et al., 1989) except for replacement of the central C^+·GC triple by a G·TA triple. This paper focuses primarily on the NMR parameters of imino protons and attached nitrogens for the G·TA triplex and evaluates proton NOEs involving exchangeable imino and amino protons to identify structural features of the triple helix and evaluate the proposed pairing alignment at the G·TA triple site. The spectral analysis was greatly aided by the narrow proton line widths for the G·TA triplex in acidic pH at low temperature.

Experimental Procedures

Oligonucleotide Preparation. The 31-mer deoxyoligomer **4** synthesis, isolation, and purification was undertaken according to procedures described previously (de los Santos et

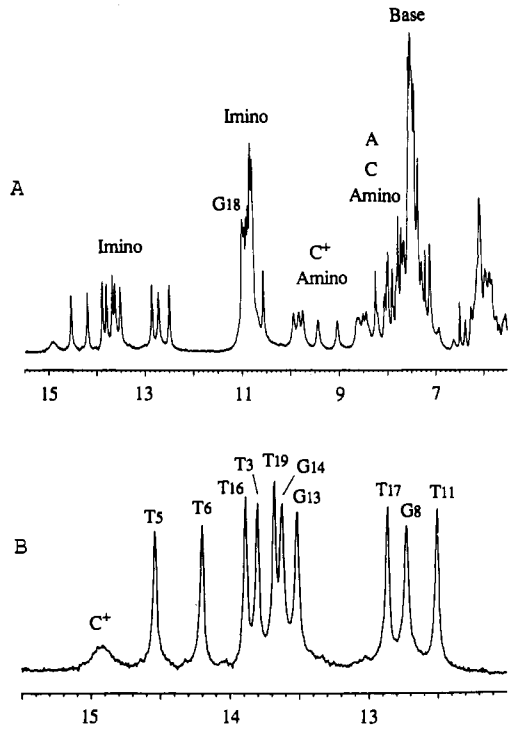


FIGURE 1: (A) Proton NMR spectrum (5.5–15.0 ppm) of the G·TA triplex in H_2O, pH 4.85, at 5 °C. The various proton classes are designated over the spectrum. (B) Expanded spectrum of the imino proton region between 12.0 and 15.0 ppm. The imino protons assignments are listed above the spectrum.

al., 1989). A single 10-μmol run using the cyanoethyl-phosphoramidite solid-phase method followed by reverse-phase HPLC purification yielded approximately 600 A_{260} units of 31-mer deoxyoligomer **4**.

NMR Experiments. Proton NMR spectra were recorded on Bruker AMX 600 (Glaxo) and AM 500 and AM 400 (Columbia University) spectrometers. The NOESY spectrum of the G·TA triplex **4** in H_2O, pH 4.85, at 5 °C was recorded on the AMX 600. Data acquisition and processing parameters were similar to those reported previously (de los Santos et al., 1989).

The proton-detected nitrogen–proton two-dimensional HMQC experiment on the G·TA triplex **4** in H_2O, pH 4.85, at 5 °C was recorded on a GN 500 spectrometer (Emory University) with the parameters listed below. The proton frequency was set at 12.3 ppm with a sweep width of 2680 Hz in 512 points. A total of 128 t_1 values were taken with 2048 scans for each increment, and the ^{15}N sweep width was 5000 Hz total centered at 141 ppm from the ammonia reference. Data in this dimension were zero-filled twice before transformation. A soft proton pulse was used. The length of the 90° pulse was the inverse of the distance in Hertz from the H_2O signal, and a composite 180° pulse at the same power level was used. The delay between scans was 800 ms. The total experiment time was about 60 h.

RESULTS

The numbering scheme for the G·TA triplex is outlined with the sequence in **4**. The purine strand that extends from G8 to G14 pairs through Watson–Crick alignment with the pyrimidine strand extending from C1 to C7 and pairs through Hoogsteen alignment with the protonated pyrimidine strand extending from C15 to C21.

Proton Spectrum in H_2O at Acidic pH. The proton NMR spectrum of the G·TA triplex in H_2O solution, pH 4.85, at 5 °C yielded unusually well-resolved exchangeable and nonexchangeable resonances. (Figure 1A). The hydrogen-bonded imino protons that resonate between 12.0 and 15.0 ppm are completely resolved and exhibit narrow line widths except for a protonated cytidine imino proton at 14.9 ppm (Figure 1B). The non-hydrogen-bonded imino protons resonate between 10.5 and 11.1 ppm and include the imino protons of the 10 loop thymidines in the G·TA triplex **4**. The protonated cytidine amino protons resonate between 8.5 and 10.0 ppm with partial resolution of the expected six amino protons from C15, C20, and C21 (Figure 1A). The nonexchangeable base protons along with the amino protons of cytidine and adenosine resonate between 7.0 and 8.5 ppm. These nonexchangeable base protons exhibit narrow resonances in the G·TA triplex even at low temperature (Figure 1A).

Table I: Exchangeable and Nonexchangeable Proton Chemical Shifts in the G·TA Triplex in H_2O, pH 4.85, at 5 °C

	chemical shifts (ppm)						
	NH3/NH1	NH_2	H2	H8	H5	H6	CH_3
C1		7.12, 8.44			5.70		
C2		7.29, 8.50			5.67		
T3	13.81						1.75
A4		6.62, 7.41	7.19				
T5	14.55						1.26
T6	14.20						1.57
C7		6.94, 8.20			5.51		
G8	12.71						
A9		7.44, 7.69	7.50	7.78			
A10		7.53, 7.76	7.75	8.25			
T11	12.49					7.29	1.20
A12		7.39, 7.63	6.46				
G13	13.51						
G14	13.62						
C15		9.45, 9.83			6.12	8.05	
T16	13.89						1.71
T17	12.86						1.55
G18	11.01			8.00			
T19	13.68					7.49	0.81
C20		8.64, 9.94			5.57	7.53	
C21		9.05, 9.74			5.86	7.81	

Imino Proton Assignments. Previous NMR studies (de los Santos et al., 1989; Rajagopal & Feigon, 1989; Mooren et al., 1990) confirmed the proposed model for T·AT **1** and C^+·GC **2** pairing alignments in the DNA triplex (Arnott et al., 1976) where the purine strand pairs with the unprotonated pyrimidine strand in an antiparallel alignment through Watson–Crick pairing and with the protonated pyrimidine strand in a parallel alignment through Hoogsteen pairing. Thus, we detect NOEs between adjacent Watson–Crick imino protons starting at G8 at one end and proceeding to G14 at the other end in the G·TA triplex **4** via the path G8 to T6 to T11 to T3 to G13 to G14 (Figure 2A). The inter-base-pair imino–imino NOEs are of comparable intensity except for the NOE between the imino proton of T11 of the G·TA triple and the imino proton of T3 of the flanking T·AT triple (Figure 2A). We can also follow NOEs between adjacent Hoogsteen imino protons along the T16-T17-G18-T19 segment in the G·TA triplex (Figure 2B). This tracing establishes that the imino proton of G18 resonates upfield at 11.01 ppm amid the overlapping resonances of the loop thymidine imino protons. Further, the NOEs from the imino proton of G18 of the G·TA triple to the imino protons of T19 and T17 of flanking T·AT triples are weak in the G·TA triplex (Figure 2B). This completes the assignment of the Watson–Crick and Hoogsteen imino protons in the G·TA triplex and their chemical shifts are listed in Table I and assigned over the imino proton spectrum in Figure 1B. These assignments are confirmed by the very weak NOEs detected between the Watson–Crick and Hoogsteen thymidine imino protons in individual T·AT triples (peaks A, B, and C in Figure

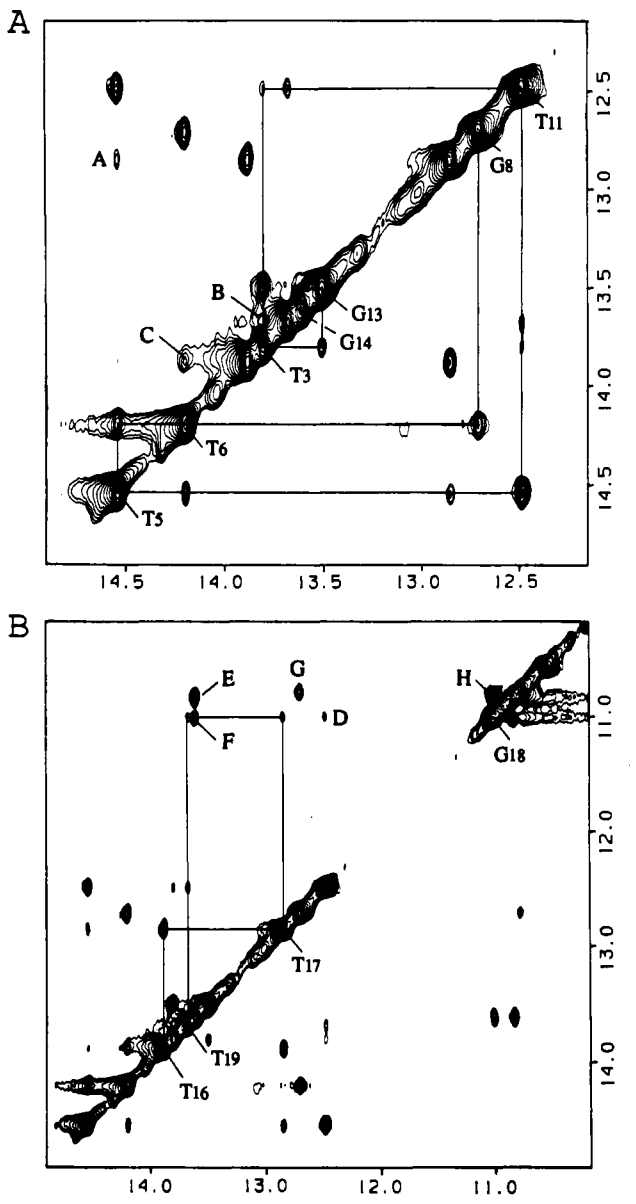


FIGURE 2: Expanded NOESY contour plots (170-ms mixing time) of the G·TA triplex in H_2O, pH 4.85, at 5 °C. (A) NOE cross peaks in the symmetrical 12.0–15.0 ppm region. The lines trace the connectivities between imino protons on adjacent Watson–Crick base pairs. The tracing starts at G8 and continues through T6, T5, T11, T3, and G13 to G14. Note that the cross peak between the imino protons of T3 and T11 is weaker than that between imino protons of T5 and T11. We also detect NOEs between thymidine imino protons within the same T·AT triple. The cross-peak assignments are A, T5-(NH3)–T17(NH3); B, T3(NH3)–T19(NH3); C, T6(NH3)–T16-(NH3). (B) NOE cross peaks in the symmetrical 10.0–15.0 ppm region. The lines trace the connectivities between imino protons on adjacent Hoogsteen base pairs. The tracing starts at T16 and continues through T17 and G18 to T19. Note that the NOEs between the T17 and G18 imino protons and between the G18 and T19 imino protons are weaker than the NOE between the imino protons of T16 and T17. Similarly, the NOE cross peak between the imino protons of T11 and G18 in the G·TA triple (peak labeled D) is weak. We also detect NOEs between the imino proton of G14 and the imino protons of loop thymidines (peaks labeled E and F) and between the imino proton of G8 and the imino proton of a loop thymidine (peak labeled G). The two loop thymidines adjacent to G14 also exhibit an NOE to each other (peak labeled H).

2A) as assigned in the legend to Figure 2A. We detect only a single broad protonated cytidine imino proton at 14.9 ppm, which is assigned to internal protonated cytidine C20 on the basis of observed NOEs to its own cytidine amino protons and to the imino proton of T19 in one-dimensional NOE experiments on the G·TA triplex at −5 °C.

Amino and Base Proton Assignments for Watson–Crick Pairs in a Triplex. The purine bases in T·AT **1** and C^+·GC **2** triples are involved in both Watson–Crick and Hoogsteen pairing alignments. The thymidine imino proton exhibits NOEs to the adenosine H2 proton and the resolved adenosine amino protons across the Watson–Crick AT pair. These cross peaks are observed for the T3, T5, T6, and T11 thymidines participating in T·AT and G·TA triples in the G·TA triplex (Figure 3A) with the assignments summarized in the legend to Figure 3. The guanosine imino proton exhibits NOEs to the hydrogen-bonded and exposed cytidine amino protons across the Watson–Crick GC pair. These cross peaks are observed for the G8, G13, and G14 guanosines participating in C^+·GC triples in the G·TA triplex (Figure 3A) with the assignments summarized in the legend to Figure 3A. The chemical shifts of the adenosine H2 and amino protons and cytidine amino protons are listed in Table I.

Amino and Base Proton Assignments for Hoogsteen Pairs in a Triplex. The thymidine imino protons of T16, T17, and T19 of the protonated third strand of the G·TA triplex **4** exhibit NOEs to the adenosine H8 and resolved adenosine amino protons across the Hoogsteen AT pair. These cross peaks are labeled in Figure 3B, their assignments listed in the accompanying figure legend, and their chemical shifts outlined in Table I. In addition, the imino proton of G18 of the G·TA triple resonates at 11.01 ppm in the G·TA triplex. This imino proton exhibits NOEs to the amino protons of A4 that are separated by 1.21 ppm (peaks Y and Z in Figure 3B).

The additional but unlabeled cross peaks for the G·TA triplex observed in Figure 3A,B reflect NOEs from an imino proton on a given triple to the base and amino protons of flanking triples. These NOEs were very useful for the independent confirmation of imino, amino, and base proton assignments of the G·TA triplex listed in Table I.

Protonated and Unprotonated Cytidine Amino Protons. The hydrogen-bonded and exposed cytidine amino protons exhibit strong NOEs to each other and the adjacent cytidine H5 protons. These NOEs are observed for both the protonated cytidines C15, C20, and C21, as well as the unprotonated cytidines C1, C2, and C7 in the G·TA triplex (Figure 4). It is possible to go to lower contour levels than those shown for the NOESY contour plot of the G·TA triplex in Figure 4 and in the process detect cross peaks between the hydrogen-bonded protons of protonated and unprotonated cytidines within a given C^+·GC triple in the G·TA triplex. Such an analysis provides an independent check of the cytidine amino proton assignments in the G·TA triplex. Also note that the hydrogen-bonded and exposed amino protons of A4 are well resolved from each other since their cross peak is positioned well off the diagonal axis (peak A4 in Figure 4).

Methyl Assignments. A range of NOEs are detected between the imino protons and CH_3 groups for the segment involving the G·TA triple and its flanking T·AT triples in the G·TA triplex. Thus, the imino protons of T17 and T19 on the protonated third strand exhibit NOEs to their own CH_3 protons, as well as strong NOEs to the CH_3 protons of adjacent T11 on the purine-containing strand in the G·TA triplex (boxed peaks G and H in Figure 5A). In addition, the T16, T17, and T19 imino protons on the protonated third strand each exhibit NOEs to two sugar H2′,2″ protons in the G·TA triplex (peaks A–F in Figure 5A). Similarly, the imino protons of G18 on the protonated third strand exhibit an NOE to the CH_3 protons of T11 within the G·TA triple (boxed peak I, left panel, in Figure 5B) and to the CH_3 protons of T17 and T19

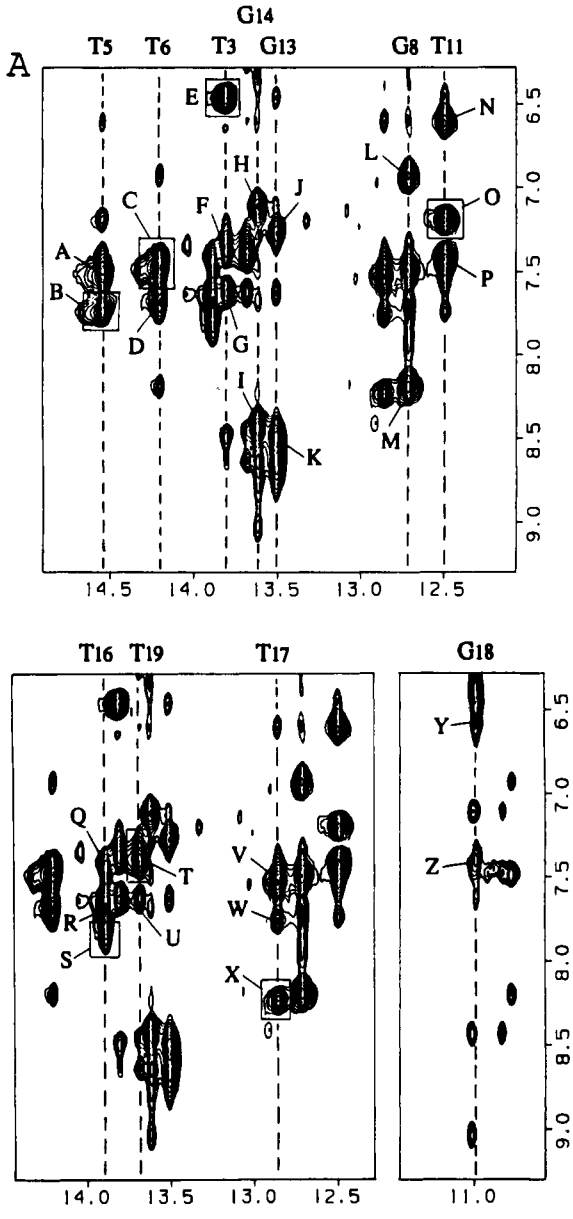


FIGURE 3: Expanded NOESY contour plots (170-ms mixing time) of the G·TA triplex in H_2O, pH 4.85, at 5 °C. (A) NOE cross peaks between imino protons (12.0–15.0 ppm) and the amino and base protons (6.5–9.0 ppm) of the Watson–Crick pairs in the triplex. The NOEs between the thymidine imino protons and adenosine H2 protons are enclosed in boxes. The labeled cross peaks are assigned as follows: A, T5(NH3)–A10(NH_2-6e); B, T5(NH3)–A10(NH_2-6h)/A10(H2); C, T6(NH3)–A9(NH_2-6e)/A9(H2); D, T6(NH3)–A9(NH_2-6h); E, T3(NH3)–A12(H2); F, T3(NH3)–A12(NH_2-6e); G, T3(NH3)–A12(NH_2-6h); H, G14(NH1)–C1(NH_2-4e); I, G14(NH1)–C1(NH_2-4h); J, G13(NH1)–C2(NH_2-4e); K, G13(NH1)–C2(NH_2-4h); L, G8(NH1)–C7(NH_2-4e); M, G8(NH1)–C7(NH_2-4h); N, T11(NH3)–A4(NH_2-6e); O, T11(NH3)–A4(H2); P, T11(NH3)–A4(NH_2-6h). (B) NOE cross peaks between imino protons (10.5–14.5 ppm) and the amino and base protons (6.5–9.0 ppm) of the Hoogsteen pairs in the triplex. The NOEs between the thymidine imino protons and adenosine H8 protons are enclosed in boxes. The labeled cross peaks are assigned as follows: Q, T16(NH3)–A9(NH_2-6e); R, T16(NH3)–A9(NH_2-6h); S, T16(NH3)–A9(H8); T, T19(NH3)–A12(H8)/A12(NH_2-6e); U, T19(NH3)–A12(NH_2-6h); V, T17(NH3)–A10(NH_2-6e); W, T17(NH3)–A10(NH_2-6h); X, T17(NH3)–A10(H8); Y, G18(NH3)–A4(NH_2-6e); Z, G18(NH3)–A4(NH_2-6h).

on adjacent T·AT triples (left panel in Figure 5B) in the G·TA triplex.

We also detect NOEs between purine H8 and pyrimidine CH_3 protons in purine–pyrimidine (3′–5′) steps in a given

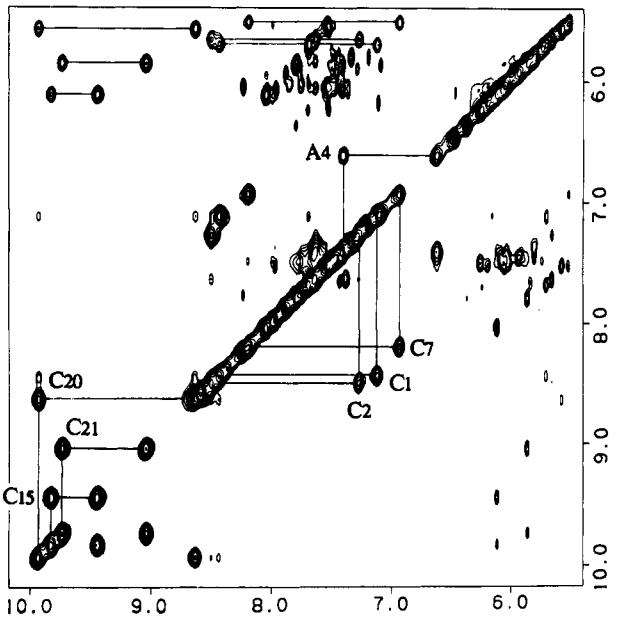


FIGURE 4: Expanded NOESY contour plot (170-ms mixing time) of the G·TA triplex in H_2O, pH 4.85, at 5 °C. NOE cross peaks in the symmetrical 5.5–10.0 ppm region. The NOEs between the geminal amino protons of protonated cytidines C15, C20, and C21 and betwen the geminal amino protons of cytidines C1, C2, and C7 are designated in figure. Each of these amino protons also exhibits NOEs to its linked cytidine H5 proton resonating between 5.5 and 6.2 ppm. An NOE is also detected between the hydrogen-bonded and exposed protons of A4.

strand as shown for the A10–T11 step and the G18–T19 step (boxed peaks J and K, right panel, in Figure 5B) in the G·TA triplex. These NOEs are characteristic of right-handed duplexes and provide additional proton assignments for base protons on the G·TA triple and its flanking T·AT triples in the G·TA triplex (Table I).

Temperature Dependence of Exchangeable Proton Spectra. The temperature dependence of the imino protons (12.0–15.5 ppm) in the G·TA triplex in H_2O, pH 4.85, between −5 and 42 °C are plotted in Figure 6. The imino proton of protonated C20 that resonates at 14.9 ppm at −5 °C broadens out on raising the temperature to 17 °C. The thymidine and guanosine imino protons broaden in a sequential manner on raising the temperature further to 42 °C. The imino proton of T11 in the G·TA triple can still be detected in the G·TA triplex spectrum at 42 °C in contrast to significant broadening of all the other imino protons at this temperature (Figure 6). The protonated cytidine amino protons of C15, C20, and C21 (8.5–10.5 ppm) exhibit similar broadening patterns with temperature on proceeding from 17 to 42 °C.

pH Dependence of Exchangeable Proton Spectra. The imino proton (12.0–15.5 ppm) and protonated cytidine amino proton (8.5–10.5 ppm) spectra of the G·TA triplex in H_2O between pH 4.85 and 6.53 at 5 °C are plotted in Figure 7. The intramolecular G·TA triplex **4** formed at low pH is in slow equilibrium with the corresponding Watson–Crick antiparallel duplex involving pairing between C1–C7 and G8–G14 segments formed at higher pH. (Figure 7). The midpoint of the transition occurs at pH 6.0 in the G·TA triplex at 5 °C. It should be noted that we do not detect the $(Y+)_n{\cdot}(R+)_n(Y-)_n$ G·TA triplex in the pH 6.5 spectrum (Figure 7).

Imino Nitrogen Assignments. The excellent resolution of the imino protons in the G·TA triplex (Figure 1B) permits the correlation of the deoxyguanosine N1 and thymidine N3 base ring nitrogens with the assigned one-bond-coupled imino proton resonances following an analysis of the proton-detected ni-

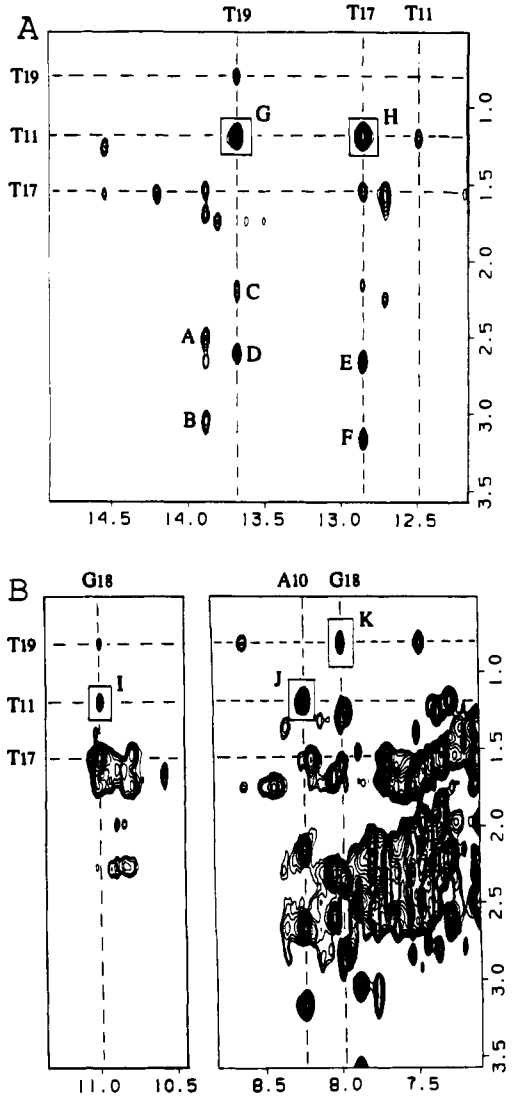


FIGURE 5: Expanded NOESY contour plots (170-ms mixing time) of the G·TA triplex in H_2O, pH 4.85, at 5 °C. (A) NOE cross peaks between imino protons (12.0–15.0 ppm) and the CH_3 and H2′,2″ protons (0.5–3.5 ppm). The boxed cross peaks correspond to NOEs between the imino proton of T19 and the CH_3 of T11 (peak labeled G), as well as between the imino proton of T17 and the CH_3 of T11 (peak labeled H). The imino protons of T19, T16, and T17, which are involved in Hoogsteen pairing in T·AT triples, exhibit NOEs to sugar H2′,2″ protons (peaks labeled A–F). (B) NOE cross peaks between imino and base protons (7.0–11.5 ppm) and the CH_3 and H2′,2″ protons (0.5–3.5 ppm). The boxed cross peak labeled I in the left panel is between the imino proton of G18 and the CH_3 of T11. Note that cross peak between the H8 of G18 and the CH_3 of T19 is weak. The boxed cross peaks J and K in the right panel are between the H8 of A10 and the CH_3 of T11, as well as between the H8 of G18 and the CH_3 of T19, respectively.

trogen–proton HMQC data set. An expanded contour plot establishing one-bond nitrogen–proton correlations for the guanosine N1–H1 and thymidine N3–H3 positions in the G·TA triplex in H_2O, pH 4.85, at 5 °C is plotted in Figure 8. The nitrogen–proton heteronuclear coupling cross peaks are well resolved except for the loop thymidines, which fall in a cluster. The observed doublet pattern of individual cross peaks yield the one-bond N–H coupling constant. The cross-peak assignments are listed in Figure 8, and the chemical shifts and coupling constants are tabulated in Table II.

Several trends are observed for the nitrogen chemical shifts in the G·TA triplex. The thymidine N3 nitrogens on the $(Y-)_n$ strand (161.8–162.4 ppm) that participate in Watson–Crick NH–N pairing (T3, T5, and T6) are to lower field compared

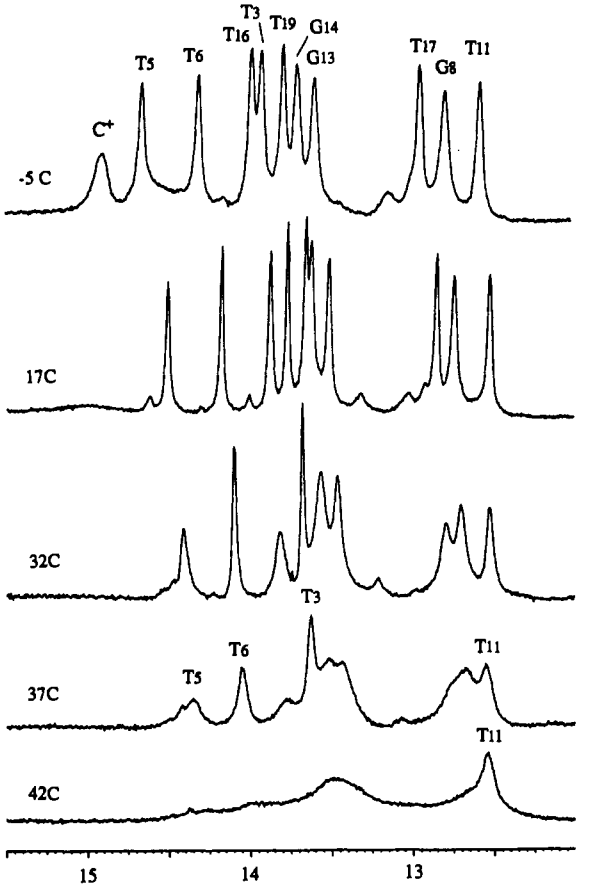


FIGURE 6: Imino proton NMR spectra (12.0–15.5 ppm) of the G·TA triplex in H_2O, pH 4.85, between −5 and 42 °C. The imino proton assignments are listed above the spectra.

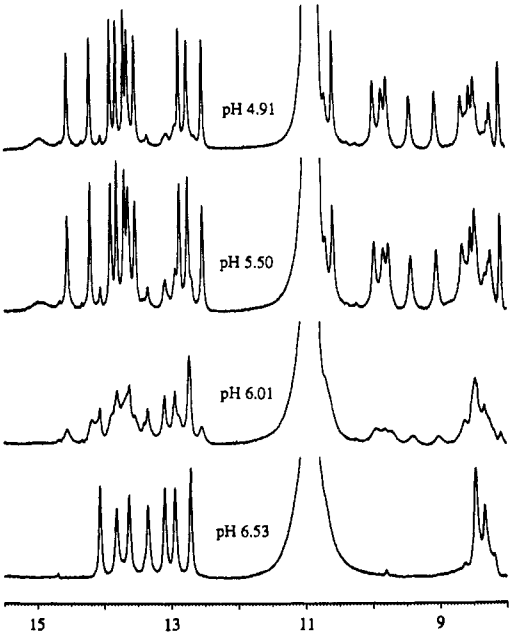


FIGURE 7: Imino proton NMR spectra (12.0–15.5 ppm) of the G·TA triplex in H_2O, at 5 °C between pH 4.85 and 6.5.

to the thymidine N3 nitrogens on the $(Y+)_n$ strand (160.1–161.4 ppm), which participate in Hoogsteen NH–N pairing (T16, T17, and T19) in the G·TA triplex (Figure 8). The Watson–Crick paired thymidine N3 of T11 on the $(R+)_n$ strand, which is flanked by adenosines, resonates at highest field (159.0 ppm). The loop thymidine N3 nitrogens resonate at still higher field (156.9–158.5 ppm) except for cross peak(s)

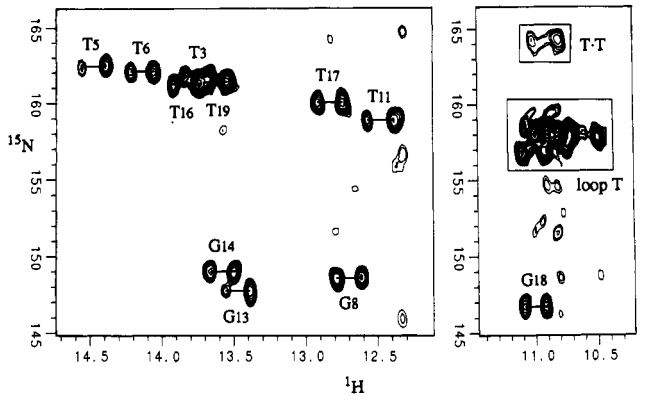


FIGURE 8: Expanded contour plot of the proton-detected nitrogen–proton heteronuclear HMQC experiment on the G·TA triplex in H_2O, pH 4.85, at 5 °C. Each cross peak is split by a one-bond nitrogen–proton coupling. The known imino proton assignments yield the guanosine N1 and thymidine N3 assignments.

Table II: Imino Nitrogen Chemical Shifts and One-Bond Imino Proton–Nitrogen Coupling Constants in the G·TA Triplex in H_2O, pH 4.85, at 5 °C

	chemical shifts (ppm)		coupling (Hz)
	δ (^{15}N)	δ (^{1}H)	J (^{15}N–^{1}H)
T3	161.8	13.82	78
T5	162.4	14.55	86
T6	162.1	14.20	82
G8	148.7	12.73	90
T11	159.0	12.50	95
G13	147.9	13.51	86
G14	149.1	13.63	83
T16	161.2	13.88	94
T17	160.1	12.86	87
G18	146.9	10.99	86
T19	161.4	13.67	70
T (loop)	156.9	10.99	
T (loop)	158.1	10.60	
T (loop)	158.5	10.90	89
T (loop)	164.1	10.93	76

at 164.1 ppm that may correspond to loop thymidines involved in T·T wobble pair (NH–O hydrogen bond) formation.

The guanosine N1 nitrogens of G8, G13, and G14 on the $(R+)_n$ strand that participate in Watson–Crick NH–N pairing resonate at 147.9–149.1 ppm (Figure 7). Of greatest interest is the large nitrogen chemical shift difference between the N1 nitrogen of G18 (146.9 ppm) and the N3 nitrogens of the loop thymidines (156.9–158.5 ppm) (Figure 7) despite their imino protons exhibiting similar chemical shifts in the G·TA triplex (Figure 1A).

Discussion

Pyr·Pur·Pyr Triplex Formation. The exceptionally well-resolved proton NMR spectrum of the G·TA triplex **4** in H_2O, pH 4.85, at 5 °C (Figure 1) highlights the advantage of single-strand DNA triplex formation (Haner & Dervan, 1990; Sklenar & Feigon, 1990). The choice of sequence and the positioning of the T_5 loop segments define the relative alignment of the three strands in the G·TA triplex with the protonated pyrimidine-rich strand, which extends from C15 to C21, aligning parallel to the purine-rich strand, which extends from G8 to G14.

Standard T·AT and C$^+$·GC Triples. Our group (de los Santos et al., 1989) and others (Rajagopal & Feigon, 1989; Mooren et al., 1990) have previously reported on proton NMR studies of $(Y+)_n·(R+)_n(Y-)_n$ triplexes at acidic pH in aqueous solution. The NOE parameters characteristic of T·AT and C$^+$·GC pairing alignments observed in these studies were also found for the same triples in the G·TA triplex **4** reported in the present study.

We note that, for T·AT triples in the G·TA triplex, the imino proton chemical shifts of thymidines in Watson–Crick base pairs (T3, T5, and T6) are generally downfield from those of the thymidine imino protons in Hoogsteen base pairs (T16, T17, and T19) as shown in Figure 1B. This trend in thymidine imino proton chemical shifts is matched by a similar trend for the thymidine N3 nitrogen chemical shifts where Watson–Crick paired nitrogens resonate to lower field relative to Hoogsteen paired nitrogens in the G·TA triplex (Figure 8). The amino protons of A9, A10, and A12 are separated by less than 0.25 ppm (Table I) with this small chemical shift difference indicating that both amino protons are involved in hydrogen-bond formation (de los Santos et al., 1989), consistent with the T·AT triple pairing alignment shown in **1**. The most upfield-shifted thymidine CH_3 protons are those that occur in purine–thymidine (3′–5′) A4–T5, A10–T11, and G18–T19 steps, consistent with stacking patterns for a right-handed helical orientation (Van de Ven & Hilbers, 1988) at these steps in all three strands of the triplex. The H2 proton of A12 exhibits an unusually upfield chemical shift of 6.46 ppm in the G·TA triplex (peak E in Figure 3A and Table I). A possible explanation is that the H2 proton of A12 of the T·AT triple is stacked over the purine ring of A4 of the adjacent G·TA triple in the G·TA triplex **4**.

The imino protons of G8, G13, and G14 involved in Watson–Crick G·C pairing resonate between 12.7 and 13.7 ppm, while the imino proton of protonated C20 involved in Hoogsteen G·C$^+$ pairing resonates further downfield at 14.9 ppm in the G·TA triplex (Figure 1B). The broad line width of the imino proton of protonated C20 and our inability to detect the imino protons of terminal protonated C15 and C21 residues, which are presumably broadened out in the G·TA triplex, suggests that protonated cytidine imino protons exhibit the fastest exchange rates with solvent H_2O. We observe the Watson–Crick paired guanosine N1 nitrogens of G8, G13, and G14 between 147.9 and 149.1 ppm but were unable to detect the protonated cytidine N3 nitrogen of C20 due to its broad proton line width in the G·TA triplex in H_2O, pH 4.85, at 5 °C (Figure 8). We have noted in a separate study on a longer DNA triplex containing alternating T·AT and C$^+$·GC triples that the protonated cytidine N3 nitrogen involved in Hoogsteen pairing resonates 2 ppm to low field of the guanosine N1 nitrogen involved in Watson–Crick pairing in C$^+$·GC triples (Live et al., 1991). The chemical shift difference between hydrogen-bonded and exposed cytidine amino protons remains essentially constant at approximately 1.25 ppm for C1, C2, and C7 involved in Watson–Crick G·C pairing for C$^+$·GC triples in the G·TA triplex (Table I and Figure 4). By contrast, the chemical shift difference between hydrogen-bonded and exposed amino protons of protonated cytidines for C$^+$·GC triples varies between 0.38 ppm for C15 and 1.50 ppm for C20 in the G·TA triplex (Table I and Figure 4).

G·TA Triple. The demonstration of the formation of a stable G·TA base triple in an otherwise $(Y+)_n·(R+)_n(Y-)_n$ triplex by Griffin and Dervan (1989) extended the triple pairing code to mixed sequence DNA. The proposed pairing alignment for the G·TA triple shown in **3a** and **3b** involves formation of a single hydrogen bond between one of the 2-amino protons of G on the pyrimidine-rich Hoogsteen strand and the 4-carbonyl group of T on the purine-rich strand. Our NMR data are consistent with such a pairing alignment of G18 with the T11·A4 Watson–Crick pair on the basis of the

observations outlined below: The imino proton of G18 resonates at 11.01 ppm with its upfield chemical shift, consistent with this imino proton not hydrogen bonded to an acceptor group as postulated for the G·TA pairing alignments **3a** and **3b**. It should be noted that the 146.9 ppm exposed nitrogen of G18 in the G·TA triple is high field of the N1 nitrogens of Watson–Crick hydrogen-bonded G8, G13, and G14, which resonate at 147.9–149.1 ppm in the G·TA triplex (Figure 8). By contrast, the 1.2 ppm chemical shift difference between the amino protons of A4 (Figure 4 and Table I) establishes that one amino proton is hydrogen bonded and the other is exposed as postulated for the G·TA pairing alignments **3a** and **3b**. We are unable to unambiguously identify the amino protons of G18 in the G·TA triplex at this time, and this limits our ability to distinguish between pairing alignments **3a** and **3b** proposed as alternate arrangements for the G·TA triple. The very weak NOEs detected between the imino proton of G18 and the imino protons of flanking T17 and T19 in the G·TA triplex (Figure 2B) may tend to favor pairing alignment **3a** over **3b** since the imino proton of G18 is displaced further away from the pseudoaxis that defines the position of the Hoogsteen imino protons of flanking T·AT triples in **3b**.

The observed NOE between the imino proton of G18 and both amino protons of A4 (peaks Y and Z in Figure 3A, right panel), as well as a weak NOE to the CH_3 of T11 (peak I in Figure 5B, left panel) establish an anti glycosidic torsion angle at G18 in the G·TA triple. The observed strong NOEs between the CH_3 groups of T11 of the G·TA triple and the imino protons of T17 and T19 of flanking T·AT triples (peaks H and G in Figure 5A) are consistent with the relative position of these protons in a stacked alignment of the G·TA triple between T·AT triples in a right-handed triple helix. The T11 imino proton at 12.49 ppm is the highest field imino proton in Figure 1B, and this upfield shift must result from stacking of T11 with adjacent purine bases A10 and A12 in the purine-rich strand of the G·TA triplex. Similarly, the Watson–Crick paired thymidine N3 nitrogens of T11 in the G·TA triplex at 159.0 ppm resonate at higher field compared to the remaining Watson–Crick and Hoogsteen paired thymidine N3 nitrogens in the G·TA triplex (Figure 8).

Loop Thymidines. Previous NMR studies have established that T_4 and T_5 hairpin loops are stabilized through wobble pairing of the first and last thymidines in the loop with the resulting T·T mismatch pair stacking on the adjacent Watson–Crick stem base pair (Hare & Reid, 1986; Van de Ven & Hilbers, 1988). The G14 imino proton exhibits NOEs to loop thymidine imino protons at 10.84 and 11.02 ppm (peaks E and F in Figure 2B) with a strong NOE between the latter thymidine imino protons (peak H in Figure 2B). Thus, these thymidine imino protons must originate in the T·T mismatch pair that stacks on the C1·G14 Watson–Crick base pair and belong to the T_5 loop that connects the $(R+)_n$ and $(Y-)_n$ strands in the G·TA triplex.

We also detect an NOE between the loop thymidine imino proton at 10.79 ppm and the G8 imino proton (peak G in Figure 2B), establishing that this thymidine belongs to the T_5 loop, which connects the $(Y-)_n$ and protonated $(Y+)_n$ strands. Our NMR data do not provide evidence of T·T mismatch formation for this T_5 loop in the G·TA triplex.

Stacking of the loop thymidine bases on the terminal base triples may contribute to the enthalpic stabilization of the triplex on the basis of earlier research establishing the contribution of dangling ends to the stability of duplexes (Senior et al., 1988).

Stability of the G·TA Triplex. The G·TA triplex is generated at acidic pH (Figure 8), consistent with the requirement for protonation of the $(Y+)_n$ strand in $(Y+)_n\cdot(R+)_n(Y-)_n$ triplexes due to formation of $C^+\cdot GC$ pairs (De los Santos et al., 1989; Rajagopal & Feigon, 1989). We do note that intramolecular formation of the G·TA triplex is not observed at pH 6.5 (Figure 7). By contrast, it has been reported previously that an intramolecular $(Y-)_n\cdot(R+)_n(Y-)_n$ triplex containing standard T·AT and $C^+\cdot GC$ triples can be detected in the pH range 7–8. (Sklenar & Feigon, 1990)

Structural Perturbations Centered about the G·TA Triple Site. An interesting question relates to the potential structural disruption of the triplex at the G·TA site flanked by T·AT triples in the intramolecular G·TA triplex. We pointed out earlier the unusally weak NOE detected between the Watson–Crick thymidine imino protons of T11 in the G·TA triple and of T3 in the flanking T·AT triple (Figure 2A). This suggests that replacement of a standard T·AT or $C^+\cdot GC$ triple by a G·TA triple does require structural readjustments. Clearly, this question can only be put on a quantitative basis following a complete analysis of NOE connectivities involving nonexchangeable protons in the G·TA triplex. We are encouraged by the partial resolution of the nonexchangeable base proton region (6.8–7.8 ppm) of the G·TA triplex in H_2O, pH 4.85, at 5 °C. Current efforts are directed toward a three-dimensional analysis of the NOESY–TOCSY proton spectra of the G·TA triplex in D_2O at 25 °C. The analysis of such a complete experimental data set should shed light on potential structural perturbations at the G·TA triple site in an otherwise $(Y+)_n\cdot(R+)_n(Y-)_n$ triplex.

Acknowledgments

We thank Karen Greene for technical assistance in the heteronuclear proton–nitrogen HMQC two-dimensional experiments.

References

Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 88*, 509.
Arnott, S., Bond, P. J., Selsing, E., & Smith, P. J. (1976) *Nucleic Acids Res. 3*, 2459–2470.
Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, P., & Vassart, G. (1985) *Nucleid Acids Res. 13*, 5127.
Collier, D. A., & Wells, R. D. (1990) *J. Biol. Chem. 265*, 10652–10658.
De los Santos, C., Rosen, M., & Patel, D. J. (1989) *Biochemistry 28*, 7282–7289.
Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.
Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967–971.
Haner, R., & Dervan, P. B. (1990) *Biochemistry 29*, 9761–9765.
Hanvey, J. C., Shimizu, M., & Wells, R. D. (1989) *J. Biol. Chem. 264*, 5950–5956.
Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.
Helene, C., & Thuong, N. T. (1988) *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) Vol. 2, pp 105–123, Springer-Verlag, New York.
Horne, D. A., & Dervan, P. B. (1990) *J. Am. Chem. Soc. 112*, 2435–2437.
Htun, H., & Dahlborg, J. E. (1989) *Science 243*, 1571–1576.
Larson, A., & Weintraub, H. (1982) *Cell 29*, 609–622.
Lee, J. S., Johnson, D. A., & Morgan, A. R. (1979) *Nucleic Acids Res. 6*, 3073–3091.
Live, D. H., Radhakrishnan, I., Misra, V., & Patel, D. J. (1991) *J. Am. Chem. Soc. 113*, 4687–4688.

Luebke, K. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc. 111*, 8733–8735.
Lyamichev, V. I., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1986) *J. Biomol. Struct. Dyn. 3*, 667–669.
Mirkin, S. N., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Fillipov, S. A., & Frank-Kamenetskii, M. D. (1987) *Nature 330*, 495–497.
Mooren, M. M., Pulleyblank, D. E., Wijmenga, S. S., Bloomers, M. J., & Hilbers, C. W. (1990) *Nucleic Acids Res. 18*, 6523–6529.
Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.
Perronault, L., Asseline, U., Rivalle, C., Thuong, N. T., Bisagni, E., Giovannangeli, C., Le Doan, T., & Helene, C. (1990) *Nature 344*, 358–360.
Posvic, T. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc. 111*, 3059–3061.
Plum, G. E., Park, Y. W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436–9440.
Praseuth, D., Perronault, L., LeDoan, T., Chassignol, M., Thuong, N., & Helence, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1349–1353.
Rajagopal, P., & Feigon, J. (1989) *Biochemistry 28*, 7859–7870.
Senior, M., Jones, R. A., & Breslauer, K. J. (1988) *Biochemistry 27*, 3879–3885.
Shimizu, M., Hanvey, J. C., & Wells, R. D. (1989) *J. Biol. Chem. 264*, 5944–5949.
Sklenar, V., & Feigon, J. (1990) *Nature 345*, 836–838.
Strobel, S. A., & Dervan, P. B. (1990) *Science 249*, 73.
Strobel, S. A., & Dervan, P. B. (1991) *Nature 350*, 172–174.
Sun, J. S., Francois, J. C., Montenay-Garestier, T., Saison-Behmoras, T., Roig, V., Thuong, N. T., & Helene, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9198–9202.
Van de Ven, F. J., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlrab, F. (1988) *FASEB J. 2*, 2939–2949.

DNA-Induced Increase in the α-Helical Content of C/EBP and GCN4†

Karyn T. O'Neil,‡ Jon D. Shuman,§ Christophe Ampe,‖ and William F. DeGrado*,‡

Central Research and Development Department, The DuPont Company, Wilmington, Delaware 19880-0328, Department of Embryology, Carnegie Institution of Washington, Baltimore, Maryland 21210, and Department of Biochemistry and Biophysics, Yale University, New Haven, Connecticut 06511

Received March 25, 1991; Revised Manuscript Received June 7, 1991

ABSTRACT: Leucine zipper proteins comprise a recently identified class of DNA binding proteins that contain a bipartite structural motif consisting of a "leucine zipper" dimerization domain and a segment rich in basic residues responsible for DNA interaction. Protein fragments encompassing the zipper plus basic region domains (bZip) have previously been used to determine the conformational and dynamic properties of this motif. In the absence of DNA, the coiled-coil portion is α-helical and dimeric, whereas the basic region is flexible and partially disordered. Addition of DNA containing a specific recognition sequence induces a fully helical conformation in the basic regions of these fragments. However, the question remained whether the same conformational change would be observed in native bZip proteins where the basic regions might be stabilized in an α-helical conformation even in the absence of DNA, through interactions with portions of the protein not included in the bZip motif. We have now examined the DNA-induced conformational transition for an intact bZip protein, GCN4, and for the bZip fragment of C/EBP with two enhancers that are differentially symmetric. Our results are consistent with the induced helical fork model wherein the basic regions are largely flexible in the absence of DNA and become fully helical in the presence of the specific DNA recognition sequence.

The leucine zipper class of DNA binding proteins comprises a group of structurally related proteins including numerous factors involved in transcriptional modulation and oncogenesis (Johnson & McKnight, 1989; Lanschultz et al., 1988). These proteins bind specifically to DNA as dimers and frequently recognize dyad-symmetric DNA sequences using a bipartite "bZip" motif (Vinson et al., 1989) consisting of a basic region and a leucine zipper. Genetic and biochemical evidence (Lanschultz et al., 1988; Hope & Struhl, 1987; Kouzarides & Ziff, 1989; Sassone-Corsi et al., 1988; Gentz et al., 1989; Turner & Tjian, 1989) indicates that the leucine zipper domain is responsible for forming homo- and heterodimers while the DNA binding activity resides in the basic region, which is immediately N-terminal to the leucine zipper and approximately 25 residues in length. This sequence contains numerous positively charged residues as well as neutral residues that appear essential for sequence-specific interactions with the base pairs (O'Neil et al., 1990; Nakabeppu & Nathans, 1989; Sellers & Struhl, 1989; Nueberg et al., 1989; Shuman et al., 1990). Biophysical studies (O'Shea et al., 1989a,b) indicate that the leucine zipper forms a two-stranded coiled-coil, a 2-fold symmetric dimer of parallel α-helices (Crick, 1953) often found in fibrous proteins (Cohen & Parry, 1990). This structure shows an exact seven-residue repeat, matching the

† C.A. is supported by an Anna Fuller Fund fellowship; additional funding was provided by National Institutes of Health Grant GM39546. J.D.S. is supported by a fellowship stipend from the Leukemia Society; additional funding was provided by the Howard Hughes Medical Research Institute and the Carnegie Institution of Washington.
‡ The Dupont Company.
§ Carnegie Institution of Washington.
‖ Yale University.

0006-2960/91/0430-9030$02.50/0