Bulge Defects in Intramolecular Pyrimidine•Purine•Pyrimidine DNA Triplexes in Solution†

Yong Wang‡ and Dinshaw J. Patel*,‡,§

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

Received October 27, 1994; Revised Manuscript Received February 23, 1995⊗

ABSTRACT: We report below on NMR studies of single base bulges in intramolecular pyrimidine•purine•pyrimidine (Y•RY) DNA triplexes in aqueous solution at acidic pH. The structural studies were undertaken with the goal of elucidating the dependence of the bulge site conformation on the nature of the base (adenine or thymine) and the location of the defect site (Watson−Crick pyrimidine and purine strands and Hoogsteen pyrimidine strand). The NMR parameters establish that an extra adenine loops out of the Y•RY triplex when it is positioned either on the Watson−Crick pyrimidine strand I (designated A_I bulge triplex) or the Hoogsteen pyrimidine strand III (designated A_{III} bulge triplex) with the associated destabilization greater for the A_{III} bulge triplex relative to the A_I bulge triplex. This observation that single adenine bulges loop out of Y•RY DNA triplexes contrasts with previous NMR structural studies, which established that single adenine bulges stack into DNA duplexes in solution. We also establish that an extra thymine on the Watson−Crick purine strand II (designated T_{II} bulge triplex) loops out of a Y•RY DNA triplex. The single base bulges do not disrupt the pairing alignments of the flanking triples in all three bulge Y•RY triplexes. It therefore appears that structural constraints energetically disfavor stacking of extra bases into any of the three strands of Y•RY DNA triplexes in solution. Our NMR studies also establish that while intramolecular Y•RY DNA triplexes at low pH can accommodate single base bulges on each of the three strands, the triplex is disrupted following insertion of an A-G bulge in Hoogsteen strand III.

Bulge defects containing one or more extra bases are a common feature of RNA and serve as recognition motifs for interaction with proteins and other RNAs (Woese & Gutell, 1989; Gluick & Draper, 1992). Bulges have been implicated as intermediates in DNA replication errors resulting in frame shift mutagenesis (Streisinger et al., 1966; Kunkel, 1990; Ripley, 1990). There has been a considerable effort toward the structural characterization of nucleic acid bulges [reviewed in Turner (1992)] with the goal of understanding their role in biological processes.

Solution NMR studies have established that single purine bases stack into duplex DNA (Patel et al., 1982; Hare et al., 1986; Woodson & Crothers, 1987, 1988; Nikonowicz et al., 1989) independent of flanking sequence (Kalnik et al., 1989a). These solution experimental studies and related computational calculations (Keepers et al., 1984; Hirshberg et al., 1988) establish that the inserted purines can readily be accommodated into duplex DNA with localized conformational realignments.

The X-ray structure of a self-complementary duplex containing adenine bulges established that one adenine stacks into the duplex while the other symmetry-related adenine loops out of the helix (Joshua-Tor et al., 1992). However, the looped out adenine inserts into the cavity opposite the stacked adenine in an adjacent duplex, and hence all adenines are stacked into the helix in the crystalline state.

The earliest NMR research established that single pyrimidine bulges loop out of duplex DNA (Morden et al., 1983, 1990; Woodson & Crothers, 1987; Morden & Maskos, 1993). However, the barrier between looped-out and stacked-in conformations of the bulged pyrimidines is low at the oligomer duplex level and appears to depend on helix type, flanking sequence, and temperature. Thus, an extra thymine in a DNA duplex was stacked into the helix (Van den Hoogen et al., 1988a) while an extra uracil in a RNA duplex in an identical sequence context was looped out of the helix (Van den Hoogen et al., 1988b). Related studies established that an extra pyrimidine loops out of the helix at low temperature, but the equilibrium shifts toward a stacked alignment with increasing temperature when the bulged pyrimidine site is flanked by purines (Kalnik et al., 1989b, 1990).

This structural information on single base bulges in duplex DNA in solution have been complemented by thermodynamic (Patel et al., 1982; LeBlanc & Morden, 1991; Zieba et al., 1991) and kinetic (Chu & Tinoco, 1983) characterization of order−disorder transitions in these systems.

Gel mobility and electron microscopy studies on single adenine (Rice & Crothers, 1989) and multiple base (Hsieh & Griffith, 1989) bulges have demonstrated DNA bending at the bulge site with the magnitude of the bend increasing with the number of bases in the bulge loop (Hsieh & Griffith, 1989). Further, the extent of bending is dependent on the sequence flanking the bulge site (Wang & Griffith, 1991). The same bending phenomena has also been observed at bulge base sites in RNA duplexes and RNA−DNA hybrid

† This research was supported by NIH Grant GM34504 to D.J.P.
‡ Columbia University.
§ Memorial Sloan-Kettering Cancer Center.
⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1995.

0006-2960/95/0434-5696$09.00/0

duplexes (Bhattacharya et al., 1990; Riordan et al., 1992). Multiple adenine bulge sites have also been characterized by chemical footprinting (Bhattacharyya & Lilley, 1989) and NMR (Rosen, et al., 1992a; Aboul-ela et al., 1993) studies with the structural studies establishing that up to three adenines can be stacked into the DNA duplex without disruption of the flanking base pairs (Rosen et al., 1992b).

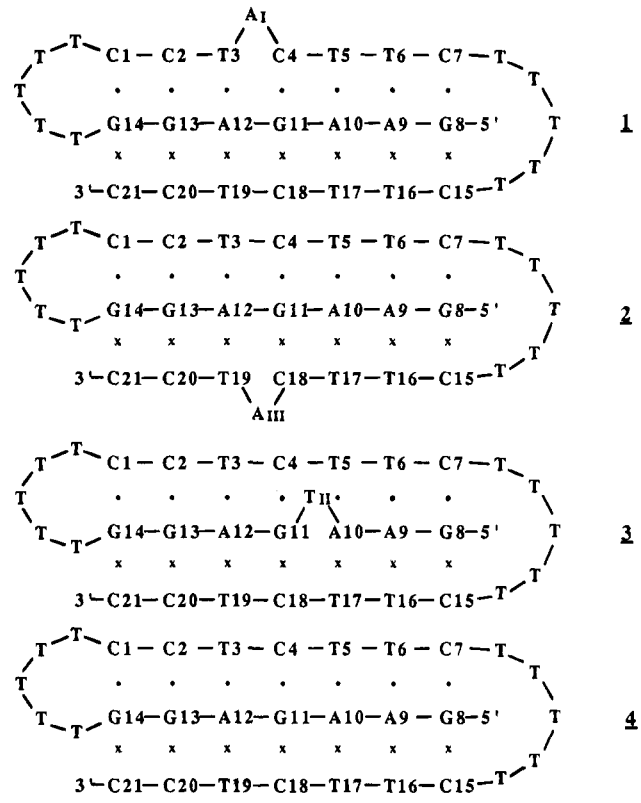
The recent interest in higher order DNA structures has led us to investigate the structure of bulges in triplexes, G-quadruplexes, and helical junctions. This paper focuses on single base bulges in pyrimidine·purine·pyrimidine (Y·RY) DNA triplexes. Such triplexes are composed of a Watson–Crick purine·pyrimidine duplex, which accommodates a pyrimidine third strand in the major groove in parallel with the purine strand (Felsenfeld et al., 1957; Arnott & Selsing, 1974; Moser & Dervan, 1987; Le Doan et al., 1987; Wells et al., 1988; Htun & Dahlberg, 1989; Thuong & Helene, 1993; Sun & Helene, 1993; Radhakrishnan & Patel, 1994c).

Considerable progress has been made on the NMR characterization of base triple pairing alignments in such triplexes (De los Santos et al., 1989; Rajagopal & Feigon, 1989; Mooren et al., 1990), and these studies have been complemented by parallel investigation of the energetics of the triplex–duplex transition (Plum et al., 1990; Roberts & Crothers, 1991; Singleton & Dervan, 1992). Such approaches at the oligomer level are facilitated by studies on intramolecular pyrimidine·purine·pyrimidine DNA triplexes (Haner & Dervan, 1990; Sklenar & Feigon, 1990; Radhakrishnan et al., 1991a) and have culminated in the solution structure of two intramolecular Y·RY DNA triplexes (Radhakrishnan & Patel, 1994a,b).

The regulation of gene expression can be modulated by the sequence-specific binding of single-stranded nucleic acids. Thus, transcriptional control can be achieved through the antisense strategy where a third strand oligonucleotide probe targets the major groove of the DNA duplex. An improvement in the ability of oligonucleotide probes to target their primary binding sites and to discriminate against secondary sites requires a detailed understanding of the structural and energetic characteristics associated with DNA triplex formation, including the consequence of mismatch alignment and bulge formation.

The available research on bulges in Y·RY DNA triplexes (Hanvey et al., 1989; Roberts & Crothers, 1991; Mergny et al., 1991; DiStefano et al., 1991) have clearly established destabilization of the triplex by the bulge site (Mergny et al., 1991; Roberts & Crothers, 1991). The efforts to date have not directly addressed the conformation of the bulge site as to whether the bulged base stacks into the triple helix or loops out into solution.

The present paper reports on NMR studies of single adenine bulges positioned either in the pyrimidine Watson–Crick strand I (designated A_I bulge triplex **1**) or the pyrimidine Hoogsteen strand III (designated A_{III} bulge triplex, **2**) of the triplex. In addition, we have studied a single thymine bulge positioned in purine Watson–Crick strand II (designated T_{II} bulge triplex, **3**) of the triplex. The NMR structural characterizations of these bulge-containing triplexes were undertaken by monitoring two-dimensional NOESY data sets in H_2O solution (Radhakrishnan et al., 1991b) and three-dimensional NOESY-TOCSY data sets in D_2O solution (Radhakrishnan et al., 1992) as recently reported from our laboratory for structure determination of G·TA-containing



(Radhakrishnan & Patel, 1994a) and T·CG-containing (Radhakrishnan & Patel, 1994b) intramolecular Y·RY DNA triplexes.

The structural studies were undertaken with the goal of elucidating the dependence of the bulge site conformation on the nature of the base (purine or pyrimidine) and the location of the defect site (Watson–Crick pyrimidine strand, Watson–Crick purine strand, and Hoogsteen pyrimidine strand) on the Y·RY triplex. These studies also permit a comparison of single base bulges in duplexes and triplexes.

EXPERIMENTAL PROCEDURES

Oligonucleotide Synthesis and Purification. The 31-mer control DNA oligomer and the 32-mer single base bulge DNA oligomers were synthesized at a 10-μmol level on an Applied Biosystems 391 DNA synthesizer using standard solid-phase cyanoethyl phosphoamidite chemistry. The crude 5′-dimethoxytritylated oligonucleotides were cleaved from the support following treatment with concentrated ammonia for 15 h at 55 °C and then purified by HPLC on a preparative C18 column. The 5′-dimethoxytrityl group was removed by treatment with 80% acetic acid for 30 min at room temperature followed by further purification by reverse-phase HPLC on a C18 reverse-phase column. The purified DNA oligomers were then desalted on a Sephadex G25 column and converted to the sodium form on a Dowex sodium form cation-exchange column. The DNA oligomers (approximately 600 A_{260} units) were dissolved in 0.4 mL in 10 mM sodium phosphate/0.1 mM EDTA, aqueous solution, and the pH was adjusted to 4.8.

NMR Experiments. Proton NMR spectra were recorded on 600, 500, and 400 MHz NMR spectrometers. The NOESY spectra in H_2O buffer, pH 4.8, at 5 °C were recorded with a mixing time of 170 ms. Data acquisition and processing parameters were similar to those reported previously (De los Santos et al., 1989).

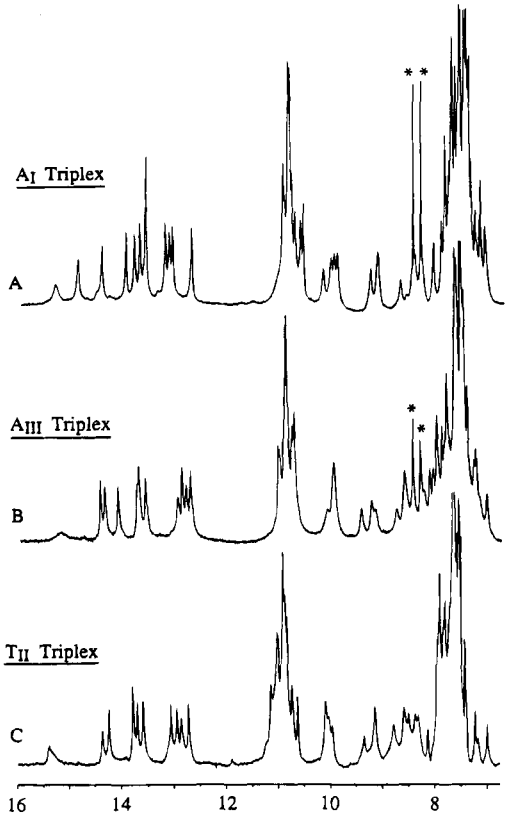


FIGURE 1: Proton NMR spectra (7.0−16.0 ppm) of (A) the A_I bulge triplex, (B) the A_{III} bulge triplex, and (C) the T_{II} bulge triplex in H_2O buffer, pH 4.8, at 5 °C. The H8 and H2 protons of the bulged adenine are designated by asterisks.

The three-dimensional NOESY-TOCSY data set in D_2O buffer, pH 4.8, at 30 °C was recorded at 500 MHz with the mixing time of 200 ms for the NOESY component and 30 ms for the TOCSY component. The spectral width was 9.7 ppm in each dimension with 512 × 60 × 110 complex data points in the $\omega_3/\omega_2/\omega_1$ dimensions, respectively. Eight transients were accumulated for each free induction decay. Data processing was carried out as described previously (Radhakrishnan et al., 1992). The final matrix consisted of 512 × 256 × 256 real data points.

RESULTS

Exchangeable Proton Spectra and Assignments in A_I Bulge Triplex **1**. The exchangeable proton spectrum of the A_I bulge triplex in H_2O buffer, pH 4.8, at 5 °C is plotted in Figure 1A. The spectrum is unusually well resolved in the hydrogen-bonded imino proton (12.5−15.5 ppm), the protonated cytidine amino proton (9.0−10.5 ppm), and the nonprotonated cytidine amino and nonexchangeable base proton (7.0−8.5 ppm) regions. The temperature dependence (15−48 °C) of the hydrogen-bonded imino protons of the A_I bulge triplex are plotted in supplementary Figure S1. The hydrogen-bonded imino proton assignments listed in Figure S1 were deduced following analysis of NOESY spectra of the A_I bulge triplex in H_2O buffer, pH 4.8, at 5 °C. Expanded NOESY contour plots establishing NOE connectivities involving imino protons in the Watson−Crick duplex segment of the triplex are shown in Figure 2 while those involving imino protons in the Hoogsteen third strand are shown in Figure 3. The NOE cross peak assignments are listed in the captions to Figures 2 and 3, and the chemical shifts are tabulated in Table 1.

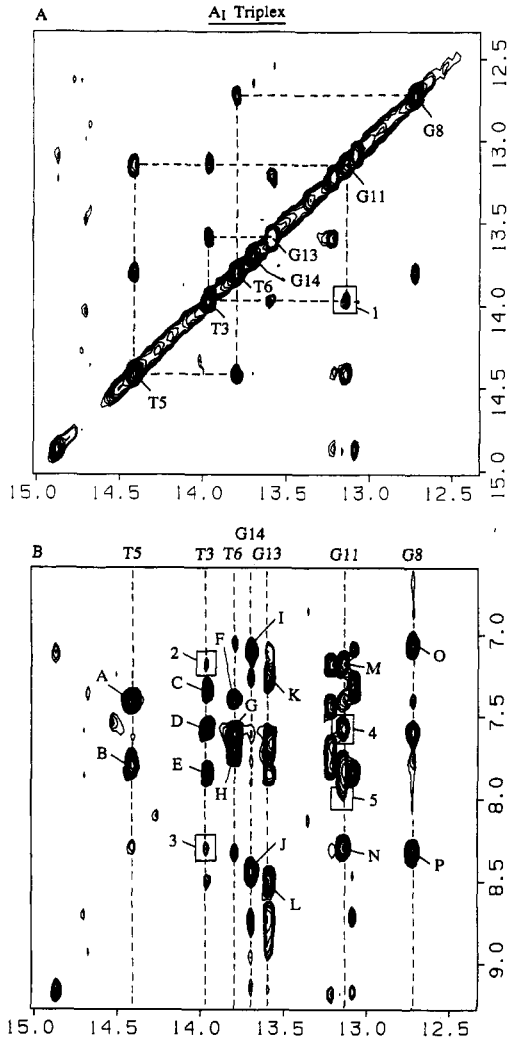


FIGURE 2: Expanded NOESY contour plots (170 ms mixing time) of the A_I bulge triplex in H_2O buffer, pH 4.8, at 5 °C. These plots correlate NOE cross peaks involving the Watson−Crick strands in the triplex. (A) NOE cross peaks in the symmetrical 12.5−15.0 ppm imino proton region. The imino proton assignments are listed along the diagonal. The lines trace the NOEs between adjacent imino protons along the Watson−Crick duplex segment of the triplex. Peak 1 corresponds to an NOE between the imino protons of T3 and G11 at the bulge site. (B) NOE cross peaks between the 12.5−15.0 ppm imino proton region and the 7.0−9.5 ppm amino and base proton region. The cross peaks A−P are assigned as follows: A, T5(NH3)-A10(H2)/A10(NH_2-6e); B, T5(NH3)-A10(NH_2-6h); C, T3(NH3)-A12(NH_2-6e); D, T3(NH3)-A12(H2); E, T3-(NH3)-A12(NH_2-6h); F, T6(NH3)-A9(H2); G, T6(NH3)-A9(NH_2-6e); H, T6(NH3)-A9(NH_2-6h); I, G14(NH1)-C1(NH_2-4e); J, G14(NH1)-C1(NH_2-4H); K, G13(NH1)-C2(NH_2-4e); L, G13(NH1)-C2(NH_2-4h); M, G11(NH1)-C4(NH_2-4e); N, G11(NH1)-C4(NH_2-4h); O, G8(N1)-C7(NH_2-4e); P, G8(NH1)-C7(NH_2-4h). The labels h and e stand for hydrogen-bonded and exposed amino protons. The peaks 2−5 correspond to NOEs between T3•A12•T19 and C4•G11•C18$^+$ base triples flanking the bulge site. They are assigned as follows: 2, T3(NH3)-C4(NH_2-4e); T3(NH3)-C4(NH_2-4h); 4, G11(NH1)-A12(H2); 5, G11(NH1)-A12(NH_2-6h).

Of special interest are NOE cross peaks involving protons on the T3•A12•T19 and C4•G11•C18$^+$ base triplets that flank the adenine bulge site in triplex **1**. We detect NOEs between the imino protons of T3 and G11 (peak 1, Figure 2A) and between the imino protons of C18$^+$ and T19 (peak 1, Figure 3A) as well as between the imino protons of T3 and both amino protons of C4 (peaks 2 and 3, Figure 2B) and between the imino proton of G11 and the H2 and amino protons of

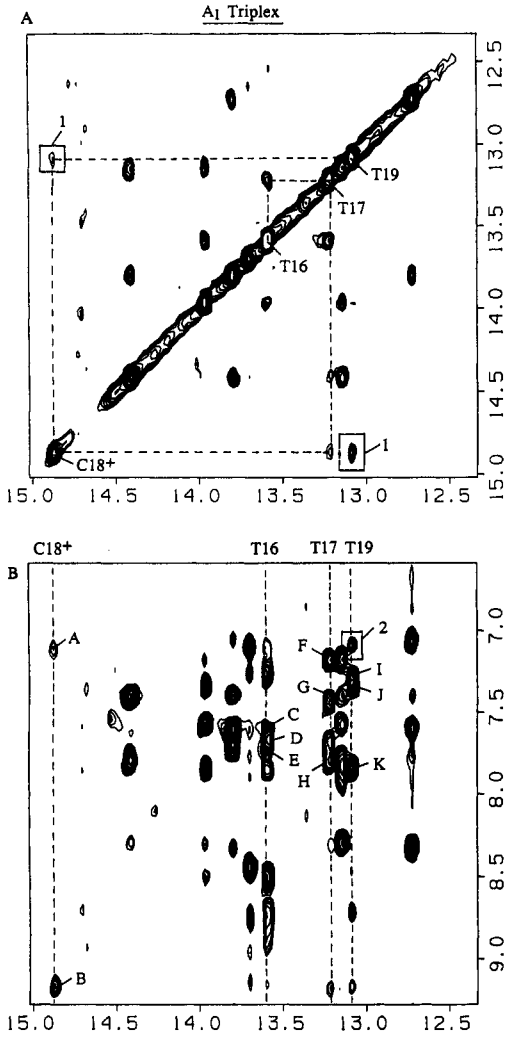


FIGURE 3: Expanded NOESY contour plots (170 ms mixing time) of the A_I bulge triplex in H_2O buffer, pH 4.8, at 5 °C. These plots correlate NOE cross peaks involving the Hoogsteen strand in the triplex. (A) NOE cross peaks in the symmetrical 12.5−15.0 ppm imino proton region. The imino proton assignments are listed along the diagonal. The lines trace the NOEs between adjacent imino protons along the Hoogsteen strand of the triplex. Peak 1 corresponds to an NOE between the imino protons of $C18^+$ and T19 at the bulge sites. (B) NOE cross peaks between the 12.5−15.0 ppm imino proton region and the 7.0−9.5 ppm amino and base proton region. The cross peaks A−K are assigned as follows: A, $C18^+$(NH3)-G11(H8); B, $C18^+$(NH3)-$C18^+$(NH_2-4e); C, T16(NH3)-A9(NH_2-6e); D, T16(NH3)-A9(H8); E, T16(NH3)-A9(NH_2-6h); F, T17(NH3)-A10(H8); G, T17(NH3)-A10(NH_2-6e); H, T17(NH3)-A10(NH_2-6h); I, T19(NH3)-A12(H8); J, T19(NH3)-A12(NH_2-6e); K, T19(NH3)-A12(NH_2-6h). The labels h and e stand for hydrogen-bonded and exposed amino protons. Peak 2 corresponds to an NOE between T19(NH3) and G11(H8) at the bulge site.

A12 (peaks 4 and 5, Figure 2B). We also detect NOEs between the amino protons of $C18^+$ and A12 on the third strand for the base triples that flank the bulge site (peaks 1−4, supplementary Figure S2A).

Nonexchangeable Proton Spectra and Assignments in A_I Bulge Triplex **1**. The nonexchangeable protons of the A_I bulge triplex are narrow and amenable to characterization using an assignment strategy involving analysis of a three-dimensional NOESY-TOCSY data set as reported previously for a related intramolecular DNA triplex (Radhakrishnan et al., 1992). The strategy utilizes the H1′ proton ω_3 planes and relies on the recognition of cross peak patterns for

Table 1: Exchangeable and Nonexchangeable Proton Chemical Shifts in the A_I Bulge Triplex in H_2O, pH 4.8, at 5 °C

	chemical shifts (ppm)				
	NH3/NH1	NH_2	H2	H8/H6	H5/CH_3
C1		7.09, 8.43		7.75	5.82
C2		7.26, 8.51		7.62	5.58
T3	13.96				1.68
C4		7.18, 8.29		7.68	5.64
T5	14.40				1.64
T6	13.79				1.64
C7		7.05, 8.34		7.62	5.59
G8	12.72			7.84	
A9		7.60, 7.73	7.42	7.68	
A10		7.44, 7.80	7.37	7.18	
G11	13.14			7.08	
A12		7.34, 7.84	7.57	7.28	
G13	13.58				
G14	13.70				
C15		9.29, 10.06		8.08	6.23
T16	13.58				1.69
T17	13.21				1.69
C18	14.87	9.17, 10.20		7.72	5.73
T19	13.08				1.68
C20		8.71, 9.98		7.86	5.81
C21		9.13, 9.93		7.94	5.94

Table 2: Nonexchangeable Proton Chemical Shifts in the A_I Bulge Triplex in D_2O, pH 4.8, at 30 °C

	chemical shifts (ppm)						
	H8/H6	H2	H1′	H2′,2″	H3′	H4′	CH_3
C1	7.76		6.00	2.33, 2.61	4.87	4.27	
C2	7.63		6.16	2.11, 2.66	4.89	4.23	
T3	7.48		6.27	2.48, 2.55	5.02	4.34	1.70
C4	7.67		5.80	2.38, 2.66	4.63	4.36	
T5	7.70		6.02	2.25, 2.64	4.87	4.27	1.66
T6	7.51		6.16	2.11, 2.57	4.86	4.26	1.66
C7	7.59		6.12	2.17, 2.38	4.83	4.15	
G8	7.82		5.95	2.55, 3.00	4.95	4.19	
A9	7.67	7.40	6.16	2.59, 3.00	5.14	4.38	
A10	7.21	7.44	6.00	2.21, 2.91	4.81	4.60	
G11	7.02		5.83	2.33, 2.83	4.85	4.38	
A12	7.31	7.55	5.91	2.23, 2.76	4.86	4.40	
G13	7.48		5.84	2.53, 2.73	4.93	4.38	
G14	7.40		5.91	2.40, 2.61	4.72	4.34	
C15	8.08		6.04	2.40, 2.68	4.76	4.34	
T16	7.73		6.23	2.38, 2.76	4.89	4.32	1.74
T17	7.67		6.27	2.45, 2.72	4.91	4.42	1.72
C18	7.75		5.91	2.38, 2.68	4.61	4.43	
T19	7.78		6.08	2.38, 2.68	4.87	4.34	1.68
C20	7.93		6.08	2.21, 2.72	4.80	4.34	
C21	7.97		6.40	2.27, 2.42	4.57	4.14	
A_I	8.46	8.38	6.57	2.85, 3.01	5.12	4.44	

determining both intraresidue and sequential assignments. Two-dimensional contour plots of the NOESY plane and the ω_3 plane corresponding to sugar H1′ protons at 6.00 ppm are plotted in supplementary Figure S3A,B, respectively, for data recorded on the A_I bulge triplex in D_2O buffer, pH 4.8, at 30 °C. The resulting simplification permits assignments of the cross peaks as demonstrated for the tracing of base−sugar H2′,2″connectivities in the H1′ proton ω_3 planes for the G8-A9-A10-G11-A12-G13-G14 purine strand in the A_I bulge triplex (Figure 4). Such as approach has permitted assignment of the nonexchangeable base and sugar protons in the A_I bulge triplex at 30 °C, and these are listed in Table 2.

A striking observation is the downfield chemical shifts for the H8 (8.46 ppm) and H2 (8.38 ppm) protons of the bulged adenine (Table 2) which are readily detected as

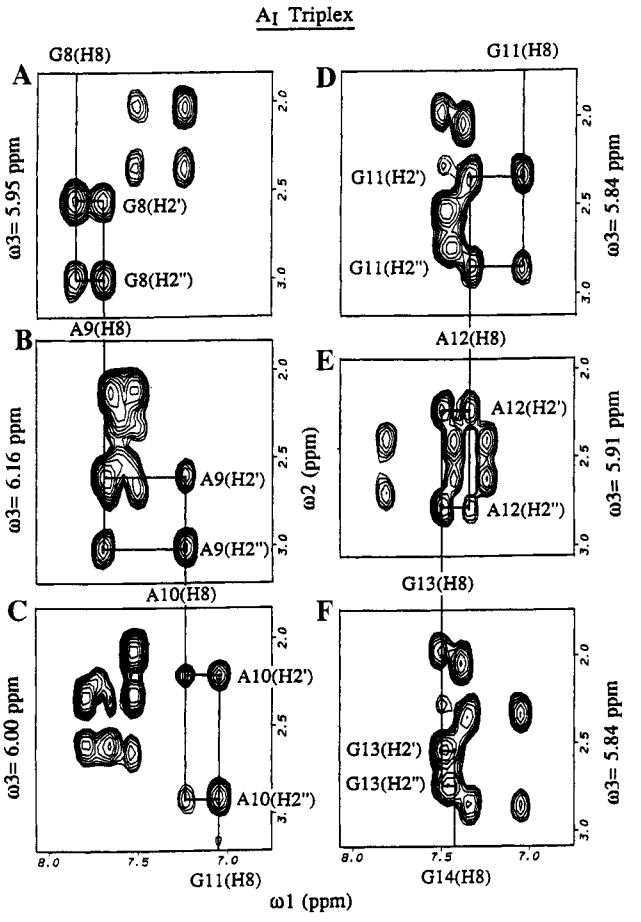


FIGURE 4: Sequential tracing of connectivities in the NOESY-TOCSY spectrum of the A_I bulge triplex in D_2O buffer, pH 4.8, at 30 °C. The NOESY mixing time was 200 ms, and the TOCSY mixing time was 30 ms. Expanded two-dimensional contour plots of the base−H2′,2″ region viewed in the H1′ ω_3 planes in residues G8(A), A9(B), A10(C), G11(D), A12(E), and G13(F).

narrow resonances in the proton spectrum of the A_I bulge triplex (peaks designated by asterisks in Figure 1A). We did not detect any NOEs between these resolved nonexchangeable protons of the bulged adenine with the exchangeable and nonexchangeable protons of the flanking T3•A12•T19 and C4•G11•C18$^+$ triples in the A_I bulge triplex.

Exchangeable Proton Spectra and Assignments in A_{III} Bulge Triplex **2**. The proton NMR spectrum (6.5−16.0 ppm) of the A_{III} bulge triplex **2** in H_2O buffer, pH 4.8, at 5 °C is plotted in Figure 1B. The temperature dependence (5−48 °C) of the hydrogen-bonded imino protons of the A_{III} bulge triplex are plotted in supplementary Figure S4. The hydrogen-bonded imino protons have been assigned (Figure S4) following analysis of NOESY data sets on the A_{III} bulge triplex in H_2O buffer, pH 4.8, at 5 °C. The expanded NOESY contour plots are shown for NOEs involving the Watson−Crick imino protons in Figure 5 and for NOEs involving the Hoogsteen imino protons in Figure 6. The NOE peak assignments are listed in the figure captions, and the chemical shifts are listed in Table 3.

The observed NOE cross peaks between the T3•A12•T19 and C4•G11•C18$^+$ base triples flanking the bulged adenine on the Hoogsteen strand in the A_{III} bulge triplex are designated by numbers in Figures 6 and 7 and supplementary Figure S2B. We detect NOEs on the Watson−Crick strand between the imino protons of T3 and G11 (peak 1, Figure

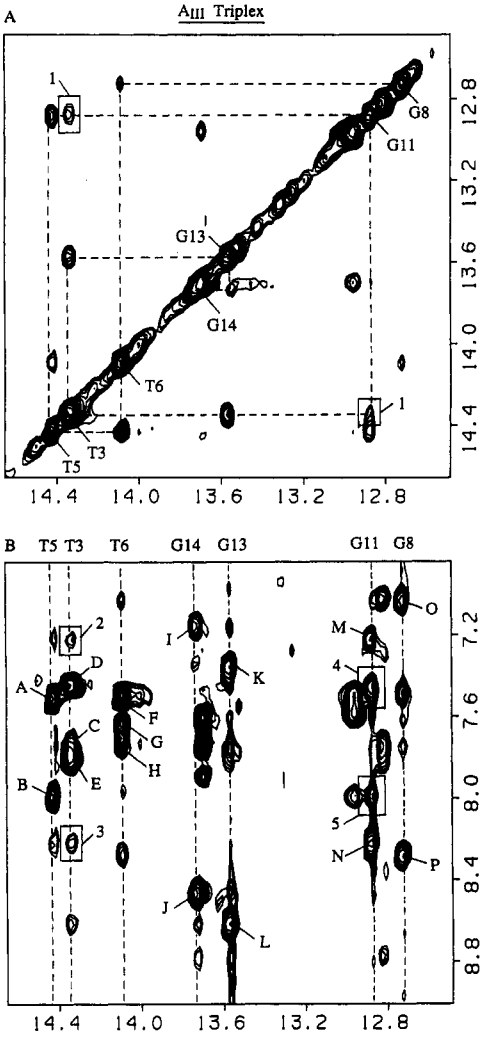


FIGURE 5: Expanded NOESY contour plots (170 ms mixing time) of the A_{III} bulge triplex in H_2O buffer, pH 4.8, at 5 °C. These plots correlate NOE cross peaks involving the Watson−Crick strands in the triplex. (A) NOE cross peaks in the symmetrical 12.5−15.0 ppm imino proton region. The imino proton assignments are listed along the diagonal. The lines trace the NOEs between adjacent imino protons along the Watson−Crick duplex segment of the triplex. Peak 1 corresponds to an NOE between the imino protons of T3 and G11 at the bulge site. (B) NOE cross peaks between the 12.5−15.0 ppm imino proton region and the 6.8−9.0 ppm amino and base proton region. The cross peaks A−P are assigned as follows: A, T5(NH3)-A10(H2)/A10(NH_2-6e); B, T5-(NH3)-A10(NH_2-6h); C, T3(NH3)-A12(NH_2-6e); D, T3(NH3)-A12-(H2); E, T3(NH3)-A12(NH_2-6h); F, T6(NH3)-A9(H2); G, T6-(NH3)-A9(NH_2-6e); H, T6(NH3)-A9(NH_2-6h); I, G14(NH1)-C1(NH_2-4e); J, G14(NH1)-C1(NH_2-4h); K, G13(NH1)-C2(NH_2-4e); L, G13(NH1)-C2(NH_2-4h); M, G11(NH1)-C4(NH_2-4e); N, G11(NH1)-C4(NH_2-4h); O, G8(NH1)-C7(NH_2-4e); P, G8(NH1)-C7(NH_2-4h). The labels h and e strand for hydrogen-bonded and exposed amino protons. The peaks 2−5 correspond to NOEs between T3•A12•T19 and C4•G11•C18$^+$ base triples flanking the bulge site. They are assigned as followed: 2, T3(NH3)-C4(NH_2-4e); 3, T3(NH3)-C4(NH_2-4h); 4, G11(NH1)-A12(H2); 5, G11-(NH1)-A12(NH_2-6h).

5A), between the imino proton of T3 and the amino protons of C4 (peaks 2 and 3, Figure 5B), and between the imino proton of G11 and H2 and amino protons of A12 (peaks 4 and 5, Figure 5B). We also detect NOEs involving the Hoogsteen third strand between the imino proton of T19 and the H8 proton of G11 (peak 2, Figure 6B) and between the amino protons of C18$^+$ and A12 (peaks 1−4, supplementary Figure S2B).

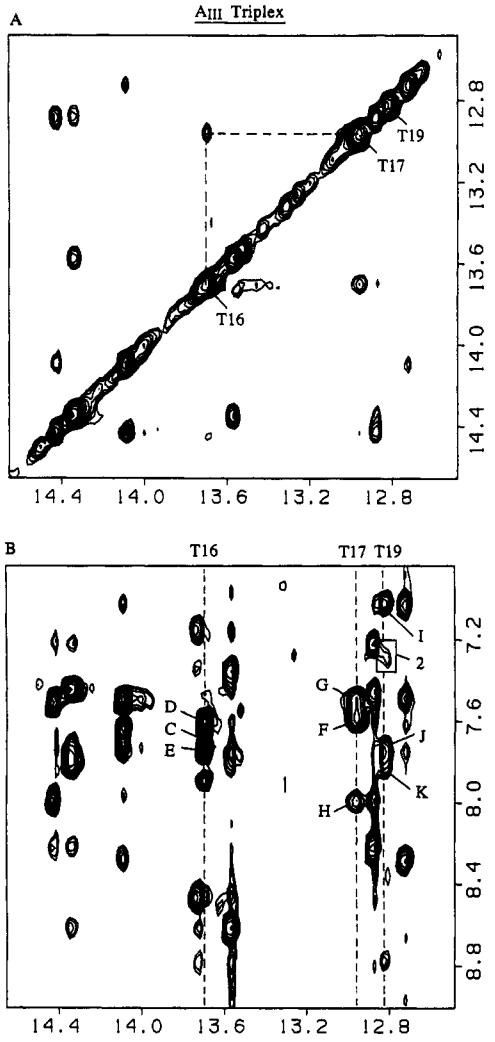


FIGURE 6: Expanded NOESY contour plots (170 ms mixing time) of the A_{III} bulge triplex in H_2O buffer, pH 4.8, at 5 °C. These plots correlate NOE cross peaks involving the Hoogsteen strand in the triplex. (A) NOE cross peaks in the symmetrical 12.5−15.0 ppm imino proton region. The imino proton assignments are listed along the diagonal. The lines trace the NOEs between adjacent imino protons of T16 and T17 along the Hoogsteen strand of the triplex. (B) NOE cross peaks between the 12.5−15.0 ppm imino proton region and the 6.8−9.0 ppm amino and base proton region. The cross peaks C−K are assigned as follows: C, T16(NH3)-A9-(NH_2-6e); D, T16(NH3)-A9(H8); E, T16(NH3)-A9(NH_2-6h); F, T17(NH3)-A10/A9(H8); G, T17(NH3)-A10(NH_2-6e); H, T17-(NH3)-A10(NH_2-6h); I, T19(NH3)-A12(H8); J, T19(NH3)-A12-(NH_2-6e); K, T19(NH3)-A12(NH_2-6h). The labels h and e stand for hydrogen-bonded and exposed amino protons. Peak 2 corresponds to an NOE between T19(NH3) and G11(H8) at the bulge site.

Nonexchangeable Proton Spectra and Assignments in A_{III} Bulge Triplex. We detect two narrow nonexchangeable protons at 8.42 and 8.30 ppm that are assigned to the H8 and H2 protons of the bulged adenine in the A_{III} bulge triplex. We do not detect NOEs between the nonexchangeable protons of the bulged adenine and the imino and amino protons of the flanking T3·A12·T19 and C4·G11·C18$^+$ triples in the A_{III} bulge triplex.

Exchangeable Proton Spectra and Assignments in T_{II} Bulge Triplex **3**. The imino protons in the proton spectrum of the T_{II} bulge triplex **3** in H_2O buffer, pH 4.8, at 5 °C are sufficiently well resolved (Figure 1C) to monitor their temperature dependence (supplementary Figure S6). The

Table 3: Exchangeable and Nonexchangeable Proton Chemical Shifts in the A_{III} Bulge Triplex in H_2O, pH 4.8, at 5 °C

	chemical shifts (ppm)				
	NH3/NH1	NH_2	H2	H8/H6	H5/CH_3
C1		7.14, 8.44		7.77	5.84
C2		7.34, 8.60		7.65	5.62
T3	14.32				1.71
C4		7.20, 8.21		7.60	5.57
T5	14.41				1.71
T6	14.07				1.64
C7		7.01, 8.28		7.56	5.57
G8	12.70			7.89	
A9		7.62, 7.73	7.49	7.58	
A10		7.48, 7.98	7.51	7.58	
G11	12.87			7.01	
A12		7.71, 7.78	7.44	7.01	
G13	13.55				
G14	13.72				
C15		9.42, 9.95		8.12	6.22
T16	13.68				1.67
T17	12.95				1.66
C18		9.15, 10.10		7.82	5.72
T19	12.78				1.67
C20		8.76, 9.93		7.95	5.95
C21		9.23, 9.97		8.04	5.95

Table 4: Exchangeable and Nonexchangeable Proton Chemical Shifts in the T_{II} Bluge Triplex in H_2O, pH 4.8, at 5 °C

	chemical shifts (ppm)				
	NH3/NH1	NH_2	H2	H8/H6	H5/CH_3
C1		7.20, 8.48		7.77	5.84
C2		7.16, 8.58		7.66	5.64
T3	14.21				1.70
C4		7.13, 8.35		7.62	5.60
T5	14.34				1.68
T6	13.77				1.69
C7		6.98, 8.29		7.57	5.60
G8	12.71			7.87	
A9		7.63, 7.72	7.61	7.73	
A10		7.61, 7.81	7.47	7.37	
G11	12.92			6.98	
A12		7.56, 7.90	7.56	7.39	
G13	13.56				
G14	13.67				
C15		9.32, 10.00		8.11	6.23
T16	13.75				1.70
T17	12.84				1.68
C18		9.12, 10.08		7.81	5.77
T19	13.04				1.74
C20		8.77, 10.08		7.86	5.84
C21		9.12, 9.94		7.92	5.93

observed NOEs involving the Watson−Crick imino protons (Figure 7) and the Hoogsteen imino protons (Figure 8) can be assigned in the expanded NOESY contour plots and are listed in the figure captions, and the chemical shifts are tabulated in Table 4.

We detect NOEs between the imino protons of T5 and G11 (peak 1, Figure 7A) that flank the thymine bulge site in the T_{II} bulge triplex. An additional NOE is detected between the imino proton of G11 and the amino proton of A10, establishing distance connectivities between the C4·G11·C18$^+$ and T5·A10·T17 triples flanking the bulge site. The imino proton of C18$^+$ exchanges rapidly in the T_{II} bulge triplex and could not be used as a marker for monitoring the adjacent thymine bulge in the triplex.

Nonexchangeable Proton Spectra and Assignments in T_{II} Bulge Triplex **3**. We have been unable to identify the H6

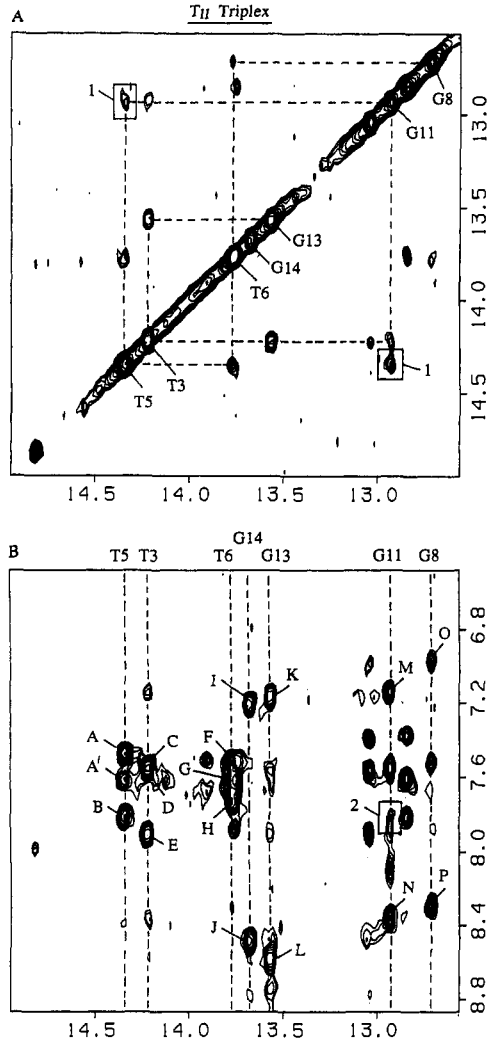


FIGURE 7: Expanded NOESY contour plots (170 ms mixing time) of the T_{II} bulge triplex in H_2O buffer, pH 4.8, at 5 °C. These plots correlate NOE cross peaks involving the Watson–Crick strands in the triplex. (A) NOE cross peaks in the symmetrical 12.5–15.0 ppm imino proton region. The imino proton assignments are listed along the diagonal. The lines trace the NOEs between adjacent imino protons along the Watson–Crick duplex segment of the triplex. Peak 1 corresponds to an NOE between the imino protons of T5 and G11 at the bulge site. (B) NOE cross peaks between the 12.5–15.0 ppm imino proton region and the 6.8–8.8 ppm amino and base proton region. The cross peaks A–P are assigned as follows: A, T5(NH3)-A10(H2); A′, T5(NH3)-A10(NH_2-6e); B, T5-(NH3)-A10(NH_2-6h); C, T3(NH3)-A12(H2); D, T3(NH3)-A12-(NH_2-6e); E, T3(NH3)-A12(NH_2-6h); F, T6(NH3)-A9(H2); G, T6(NH3)-A9(NH_2-6e); H, T6(NH3)-A9(NH_2-6h); I, G14(NH1)-C1-(NH_2-4e); J, G14(NH1)-C1(NH_2-4h); K, G13(NH1)-C2(NH_2-4e); L, G13(NH1)-C2(NH_2-4h); M, G11(NH1)-C4(NH_2-4e); N, G11-(NH1)-C4(NH_2-4h); O, G8(NH1)-C7(NH_2-4e); P, G8(NH1)-C7-(NH_2-4h). The labels h and e stand for the hydrogen-bonded and exposed amino protons. Peak 2 corresponds to an NOE between the G11 and A10 residues flanking the bulge site and is assigned as G11(NH1)-A10(NH_2-6h).

and CH_3 protons of the thymine bulge in the nonexchangeable proton spectra of the T_{II} bulge triplex.

Control Triplex **4**. We have undertaken a detailed NMR characterization of the control triplex **4** in H_2O buffer, pH 4.8, at 5 °C. The temperature dependence of the imino proton spectra are plotted in supplementary Figure S6, expanded NOESY contour plots are in supplementary Figures S7 and S8, and the exchangeable proton chemical shifts are summarized in supplementary Table S1.

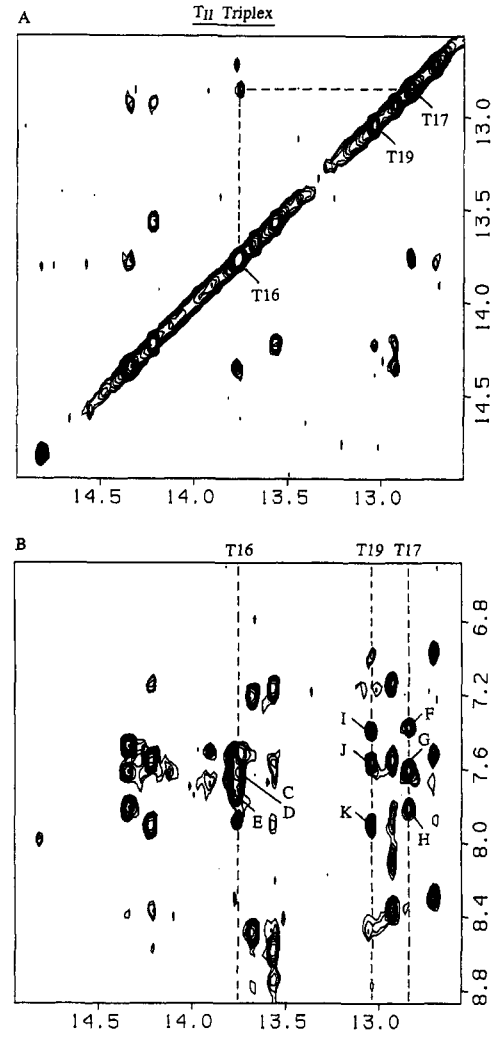


FIGURE 8: Expanded NOESY contour plots (170 ms mixing time) of the T_{II} bulge triplex in H_2O buffer, pH 4.8, at 5 °C. These plots correlate NOE cross peaks involving the Hoogsteen strand in the triplex. (A) NOE cross peaks in the symmetrical 12.5–15.0 ppm imino proton region. The imino proton assignments are listed along the diagonal. The lines trace the NOEs between adjacent imino protons of T16 and T17 along the Hoogsteen strand of the triplex. (B) NOE cross peaks between the 12.5–15.0 ppm imino proton region and the 6.8–8.8 ppm amino and base proton region. The cross peaks C–K are assigned as follows: C, T16(NH3)-A9(NH_2-6e); D, T16(NH3)-A9(NH_2-6h); E, T16(NH3)-A9(H8); F, T17-(NH3)-A10(H8); G, T17(NH3)-A10(NH_2-6e); H, T17(NH3)-A10-(NH_2-6h); I, T19(NH3)-A12(H8); J, T19(NH3)-A12(NH_2-6e); K, T19(NH3)-A12(NH_2-6h). The labels h and e strand for the hydrogen-bonded and exposed amino protons.

DISCUSSION

Triplex Formation. Two NMR spectral features establish the formation of intramolecular Y·RY triplexes at acidic pH despite the incorporation of single base bulge defects in the A_I bulge, A_{III} bulge, and T_{II} bulge triplexes. These are the observation of hydrogen-bonded imino protons from the Hoogsteen third strand between 12.7 and 13.7 ppm (Figure 1 and supplementary Figures S1, S4, and S5) and the observation of the amino protons of protonated cytosines in the Hoogsteen strand between 9.7 and 10.2 ppm (hydrogen-bonded amino protons) and between 8.8 and 9.5 ppm (exposed amino protons) (Figure 1). Thus, the intramolecular Y·RY triplexes at low pH can accommodate single base bulges on each of the three strands of the triplex.

Base Triples Flanking the Bulge Site. We observe the imino protons of the T·AT and C^+·GC base triples flanking the bulge sites in the A_I bulge, A_{III} bulge, and T_{II} bulge triplexes. Further, the NOE connectivities between imino, amino, and nonexchangeable purine base protons establish that these flanking triples adopt the T·AT and C^+·GC pairing alignments initially proposed from fiber diffraction data (Arnott & Selsing, 1974) and verified from NMR distance connectivities (De los Santos et al., 1989; Rajagopal & Feigon, 1989). Thus, the single base bulges do not disrupt the pairing alignments of the flanking triples in all three bulge triplexes studied in this contribution.

Conformation at Adenine Bulges. The observation of intrastrand NOEs at the T3−C4 step in Watson−Crick pyrimidine strand I, at the G11-A12 step in the Watson−Crick purine strand II, and at the $C18^+$-T19 step in Hoogsteen strand III in the A_I bulge triplex (Figures 2 and 3) and in the A_{III} bulge triplex (Figures 5 and 6) strongly establishes that the T3·A12·T19 and the C4·G11·$C18^+$ triples are stacked on each other. This conclusion is further supported by the interstrand NOEs between the imino protons of T3 and G11 in both triplexes (Figures 2A and 5A). These observations require that the A_I bulge in the T3-A_I-C4 segment of triplex **1** and the A_{III} bulge in the $C18^+$-A_{III}-T19 segment of triplex **2** loop out of the triple helix. In support of this conclusion are the downfield H8 and H2 proton resonances of the adenine bulges in the A_I bulge triplex (Figure 1A) and the A_{III} bulge triplex (Figure 1B) and the absence of NOEs between these adenine base protons and the protons of flanking triples.

This observation that single adenine bulges in pyrimidine strands loop out of the helix in Y·RY triplexes contrasts strikingly with the earlier demonstration that single adenine bulges stack into duplex DNA (Patel et al., 1982). Clearly, structural constraints energetically disfavor stacking of extra adenines into pyrimidines strands of DNA triplexes.

Triplex Destabilization by Adenine Bulges. The temperature dependence of the imino proton spectra of the control triplex **4** (Figure S6) can be compared with the corresponding data for the A_I bulge triplex **1** (Figure S1) and the A_{III} bulge triplex **2** (Figure S4). We note a decrease in the triplex-unfolding transition midpoint on proceeding from the control triplex to the A_I bulge triplex to the A_{III} bulge triplex.

Conformation at Thymine Bulges. The NMR parameters of the thymine bulge in the T_{II} bulge triplex **3** are consistent with the thymine looping out of the purine strand into solution. This conclusion is primarily supported by the observed interstrand NOEs between the imino protons of T5 and G11 on base triples flanking the thymine bulge site (Figure 7A). Looping out of extra pyrimidine bulges has been reported at the DNA duplex level (Morden et al., 1983) similar to our observations at the triplex level. The pyrimidine bulge in the purine strand destabilizes the T_{II} bulge triplex **3** (Figure S5) relative to the control triplex **4** (supplementary Figure S6). The thermal stability of the T_{II} bulge triplex (Figure S5) is comparable to that observed for the A_I bulge triplex (Table S1), and both are more thermally stable than the A_{III} bulge triplex.

General Comments. Our laboratory has previously reported on the high-resolution solution structures of G·TA and T·CG containing Y·RY DNA triplexes in solution [reviewed in Radhakrishnan and Patel (1994c)]. We have not attempted to determine the solution structures of bulge-containing DNA triplexes since we have no restraints between the looped out bases and the rest of the helix. We therefore cannot define the alignment of the looped out base, which would be the purpose for a detailed three-dimensional structure determination.

Comparative studies by others have established that single base bulges in the pyrimidine-rich third strand of Y·RY DNA triplexes resulted in a decrease in transition midpoint for the triplex to duplex transition (Hanvey et al., 1989; Mergny et al., 1991; Sun et al., 1991). Our NMR studies support this conclusion with the additional observation that the destabilization is more pronounced when the extra adenine is on the pyrimidine-rich Hoogsteen strand (A_{III} bulge triplex **2**) compared to the pyrimidine-rich Watson−Crick strand (A_I bulge triplex **1**).

Previous studies by others have shown that the enthalpy of Y·RY DNA triplex formation for one base bulges on the third pyrimidine-rich strand was the same as for a perfect triplex (Roberts & Crothers, 1991). This result was interpreted in terms of similar stacking interactions in perfect and single bulge-containing Y·RY triplexes with the bulged base most likely looped out from the helix (Roberts & Crothers, 1991). Our NMR data have provided experimental support for this interpretation and established further that looping out of the bulged base is independent of base and strand type in Y·RY DNA triplexes.

Combinatorial approaches have been reported by others to determine RNA sequences that will bind to a $(purine)_n$·$(pyrimidine)_n$ DNA duplex through triple helix formation in the Y·RY triplex motif (Pei et al., 1991). Interestingly, this selection approach yielded sequences that resulted not only in single base bulges but also in hairpin loop structures protruding from the third strand of an otherwise paired RNA·DNA·DNA triplex (Pei et al., 1991). We therefore attempted to study two base bulges where an A−G bulge replaced the single A_{III} bulge in Y·RY DNA triplex **2**. The NMR spectra of this A-G bulge containing Y·RY DNA triplex were of poor quality, precluding any attempt at structural characterization of the bulged residues in this triplex. It is not surprising that the insertion of the two base bulge in the pyrimidine Hoogsteen strand disrupted triplex formation since the A_{III} bulge triplex was the most destabilized single bulge triplex reported in this study.

SUPPLEMENTARY MATERIAL AVAILABLE

One table listing exchangeable and nonexchangeable proton chemical shifts in the C·GC triplex in water and eight figures showing imino proton NMR spectra and expanded NOESY contour plots (11 pages). Ordering information is given on any current masthead page.

REFERENCES

Aboul-ela, F., Murchie, A. I. H., Homans, S. W., & Lilley, D. M. J. (1993) *J. Mol. Biol. 229*, 173−188.
Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 88*, 509−521.
Bhattacharya, A., & Lilley, D. M. (1989) *Nucleic Acids Res. 17*, 6821−6840.
Bhattacharya, A., Murchie, A. I. H., & Lilley, D. M. J. (1990) *Nature 343*, 484−487.
Chu, Y. G., & Tinoco, I., Jr. (1983) *Biopolymers 22*, 1235−1246.
De los Santos, C., Rosen, M., & Patel, D. J. (1989) *Biochemistry 28*, 7282−7288.
Distefano, M. D., Shin, J. A., & Dervan, P. A. (1991) *J. Am. Chem. Soc. 113*, 5901−5902.

Felsenfeld, G., Davies, D., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023−2024.
Gluick, T. C., & Draper, D. E. (1992) *Curr. Opin. Struct. Biol. 2*, 338−344.
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., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7456−7464.
Hirshberg, M., Sharon, R., & Sussman, J. L. (1988) *J. Biomol. Struct. Dyn. 5*, 965−979.
Hsieh, C. H., & Griffith, J. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4833−4837.
Htun, H., & Dahlborg, J. E. (1989) *Science 243*, 1571−1576.
Joshua-Tor, L., Frolow, F., Appela, E., Hope, H., Rabinovich, D., & Sussman, J. L. (1992) *J. Mol. Biol. 225*, 397−431.
Kalnik, M. W., Norman, D. G., Swann, P. F., & Patel, D. J. (1989a) *J. Biol. Chem. 264*, 3702−3712.
Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989b) *Biochemistry, 28*, 294−303.
Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., & Patel, D. J. (1990) *J. Biol. Chem. 265*, 636−647.
Keepers, J. W., Schmidt, P., James, T. L., & Kollman, P. A. (1984) *Biopolymers 23*, 2901−2929.
Kunkel, T. A. (1990) *Biochemistry 29*, 8003−8011.
LeBlanc, D. A., & Morden, K. M. (1991) *Biochemistry 30*, 4042−4047.
Le Doan, T., Perronault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., & Helene, C. (1987) *Nucleic Acids Res. 15*, 7749−7760.
Mergny, J. L., Sun, J. S., Rougee, M., Montenay-Garestier, T., Barcelo, B., Chomilier, J., & Helene, C. (1991) *Biochemistry 30*, 9792−9798.
Mooren, M. M., Pulleyblank, D. E., Wijmenga, S. S., Bloomers, M. J., & Hilbers, C. W. (1990) *Nucleic Acids Res. 18*, 6523−6529.
Morden, K. M., & Maskos, K. (1993) *Biopolymers 33*, 27−36.
Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 22*, 5557−5563.
Morden, K. M., Gunn, B. M., & Maskos, K. (1990) *Biochemistry 29*, 8835−8845.
Moser, H., & Dervan, P. B. (1987) *Science 238*, 645−650.
Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714−8725.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 445−451.
Pei, D., Ulrich, H. D., & Schultz, P. G. (1991) *Science 253*, 1408−1411.
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.
Radhakrishnan, I., & Patel, D. J. (1994a) *Structure 2*, 395−405.
Radhakrishnan, I., & Patel, D. J. (1994b) *J. Mol. Biol. 241*, 600−619.
Radhakrishnan, I., & Patel, D. J. (1994c) *Biochemistry 33*, 11405−11416.
Radhakrishnan, I., & Patel, D. J., & Gao, X. (1991a) *J. Am. Chem. Soc. 113*, 8542−8544.
Radhakrishnan, I., Gao, X., De los Santos, C., Live, D., & Patel, D. J. (1991b) *Biochemistry, 30*, 9022−9030.
Radhakrishnan, I., Patel, D. J., & Gao, X. (1992) *Biochemistry 31*, 2514−2523.
Rajagopal, P., & Feigon, J. (1989) *Biochemistry 28*, 7859−7870.
Rice, J. A., & Crothers, D. M. (1989) *Biochemistry 28*, 4512−4516.
Riordan, F. A., Bhattacharya, A., McAteer, S., & Lilley, D. M. J. (1992) *J. Mol. Biol. 226*, 305−310.
Ripley, L. S. (1990) *Annu. Rev. Genet. 24*, 189−213.
Roberts, R. W., & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397−9401.
Rosen, M. A., Live, D., & Patel, D. J. (1992a) *Biochemistry 31*, 4004−4014.
Rosen, M. A., Shapiro, L., & Patel, D. J. (1992b) *Biochemistry 31*, 4015−4026.
Singleton, S. F., & Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 6957−6965.
Sklenar, V., & Feigon, J. (1990) *Nature 345*, 836−838.
Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugeta, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77−84.
Sun, J. S., & Helene, C. (1993) *Cur. Opin. Struct. Biol. 3*, 345−356.
Thuong, N. T., & Helene, C. (1993) *Angew. Chem. Int. Ed. Engl. 32*, 666−690.
Turner, D. H. (1992) *Curr. Opin. Struct. Biol. 2*, 334−337.
Van den Hoogen, Y. T., van Beuzekom, A. A., van den Elst, H., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988a) *Nucleic Acids Res. 16*, 2971−2986.
Van den Hoogen, Y. T., van Beuzekom, A. A., de Vroom, E., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988b) *Nucleic Acids Res. 16*, 5013−5030.
Wang, Y. H., & Griffith, J. (1991) *Biochemistry 30*, 1358−1363.
Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlraub, F. (1988) *FASEB J. 2*, 2939−2949.
Woese, C. R., & Gutell, R. R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3119−3122.
Woodson, S. A., & Crothers, D. M. (1987) *Biochemistry 26*, 904−912.
Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 3130−3141.
Zieba, K., Chu, T. M., Kupke, D. W., & Markey, L. A. (1991) *Biochemistry 30*, 8018−8026.

BI942518J