NMR Study of G·A and A·A Pairing in $(dGCGAATAAGCG)_2$[†]

Karol Maskos, Bonnie M. Gunn, Darryl A. LeBlanc, and Kathleen M. Morden*

Department of Biochemistry, Louisiana State University, Baton Rouge, Louisiana 70803

Received June 11, 1992; Revised Manuscript Received January 11, 1993

ABSTRACT: One- and two-dimensional NMR, UV absorption experiments, and molecular mechanics calculations were conducted on an oligonucleotide duplex $(dGCGAATAAGCG)_2$ which will be referred to as the T-11-mer. This oligonucleotide forms a duplex that is primarily B-form and contains two adjacent G·A and A·A base pairs and two 3′ unpaired guanosines. The adjacent mismatch base pairs have an unusual structure which includes *overwinding* the helix and stacking with the base from the complementary strand (A4 with A8 and G3 with A7) instead of stacking with the base which is sequential on the strand. The exchangeable and nonexchangeable proton NMR spectra of the duplex have been characterized in H_2O and D_2O solution at neutral and acidic pH. The duplex is stabilized upon protonation; however, no additional hydrogen bonds are formed. We have observed the amino protons of adenosines A4 and A8 and guanosine G3 as a function of temperature and pH. These amino protons are involved in hydrogen bonds with the purine N3 or N7 acting as acceptors. Through the observation of a variety of NOE signals, the structure of the G·A and A·A mismatch base pairs has been defined.

The formation of stable mismatched base pairs has been proposed in many nucleic acid systems of biological significance and has been confirmed and characterized through a variety of techniques. Purine·purine mismatches have been observed in tRNAs (Jack et al., 1976; Rich, 1977) and proposed in the secondary structures of *Escherichia coli* 16S rRNA (Gutell et al., 1985), 23S rRNA (Gutell & Fox, 1988), the telomeres of chromosomes (Lyamichev et al., 1990; Panyutin et al., 1990), and the consensus active region of hammerhead ribozymes (Forster & Symons, 1987a,b). The structure of deoxyoligonucleotides containing non-Watson–Crick or mismatched base pairs have been reviewed for solid-state (Kennard, 1987; Kannard & Hunter, 1990) and solution structures (Patel et al., 1987). Evaluating this collection of sequences reveals that the sequence environment influences the structure of the mismatch.

The G·A mismatch is the most extensively studied of the purine·purine mismatches and exhibits many structural variations. The propensity for G·A mismatches to escape detection by DNA polymerase III has been attributed to this structural variation (Fersht et al., 1982). In crystals of deoxyoligonucleotides, the G·A mismatch has been found in three different geometries. A G(anti)·A(anti) pair was observed in $(dCCAAGATTGG)_2$ (Privé et al., 1987, 1988), a G(anti)·A(syn) pair was observed in $(dCGCGAATTACGC)_2$ (Brown et al., 1986; Hunter et al., 1986) and $(dCGCAAGCTGGCG)_2$ (Webster et al., 1990) for crystals grown at pH >7, and a G(syn)·A$^+$(anti) pair was observed in $(dCGCAAATTGGCG)_2$ (Brown et al., 1989; Leonard et al., 1990) for crystals grown at pH 6.6. In solution the G·A mismatch has been found as the G(anti)·A(anti) pair in a variety of sequences under neutral or slightly basic conditions: $(dCCAAGATTGG)_2$ (Kan et al., 1983; Nikonowicz & Gorenstein, 1990; Nikonowicz et al., 1991), $(dCGAGAATTCGCG)_2$ (Patel et al., 1984), and (dGCCACAAGCTC)·(dGAGCTGGTGGC) (Carbonnaux et al., 1991). The oligomers $(dCGGGAATTCACG)_2$, $(dCGATAATTCGCG)_2$ (Gao & Patel, 1988), and (dGCCACAAGCTC)·(dGCTGGTGGC) (Carbonnaux et al., 1991) contain (G(anti)·A(anti) base pairs at neutral pH and convert to G(syn)·A$^+$(anti) base pairs at more acidic pH of 4.0–5.5. Carbonnaux et al. (1991) postulate the formation of an additional hydrogen bond under the acidic conditions and have detected the resonance from the protonation site at 16 ppm. This observation demonstrates a slow exchange rate on the NMR time scale, suggesting a greater stability of this hydrogen bond in this sequence. The G(anti)·A(syn) pair observed in the X-ray structure has not been reported by NMR studies in solution. This is somewhat surprising as theoretical studies indicate that the G(anti)·A(anti) pair and the G(anti)·A(syn) should have similar stability when incorporated into the helix (Chuprina & Poltev, 1983; Keepers et al., 1984). When all of these investigations of G·A base pair structure are compared, it is clear that the mismatch structure is influenced by the surrounding sequence and environmental factors such as pH. This is supported by the recent proposal of yet another type of G(anti)·A(anti) observed in RNA (Heus & Pardi, 1991) and DNA oligonucleotides (Li et al., 1991; Maskos et al., 1991). This base pair will be referred to as a type II G(anti)·A(anti) base pair (see Figure 1) and will be discussed in more detail later.

Several types of A·A mismatches can be conceived from the chemical structure alone. Self-pairing has been observed in crystals of adenine, adenine derivatives, and adenosine (Voet & Rich, 1970). In poly(A), A·A base-pair formation has been suggested to involve hydrogen bonds between the amino group on one adenine and the N7 and free phosphate oxygen on the other adenine for the acidic form (Rich et al., 1961; Lerner & Kearns, 1981). In this model, the N1 site is protonated but does not participate in base-pair formation. The duplex $d(CCCAGGG)_2$ was investigated in solution using NMR; however, no evidence was found for hydrogen bonding (Arnold et al., 1987). To date, the existence of a hydrogen-bonded A·A mismatch has not been experimentally demonstrated in oligonucleotides.

[†] This work was supported by National Institutes of Health Grant GM 38137 and the Louisiana Education Quality Support Fund [LEQSF-(86-89)-RD-A-12]. D.A.L. was supported by the United States Department of Agriculture (87-GRAD-9-0091). The NMR instruments were purchased with Grants RR02459 and RR04904 from the Division of Research Resources of the NIH.

0006-2960/93/0432-3583$04.00/0

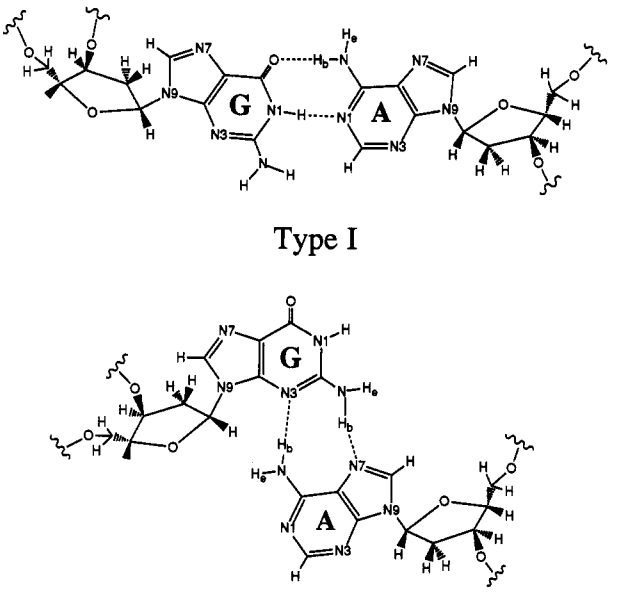

FIGURE 1: Schematic of the hydrogen bonding in two anti·anti G·A base pairs.

The studies on G·A and A·A mismatches demonstrate that a single purine·purine mismatch can exist in nucleic acids without drastically altering the nucleic acid structure or stability. Numerous examples indicate that contiguous mismatched base pairs are also a quite common structural component of natural nucleic acids (Gutell et al., 1985; Gutell & Fox, 1988; Panyutin et al., 1990; Forster & Symons, 1987a,b). Recently, several studies have addressed the question of structure and stability of tandem mismatches. A study of d(CCAAGATTGG)$_2$ containing tandem G·A mismatches found a type I G(anti)·A(anti) base pairing (Nikonowicz et al., 1990), whereas the results of an investigation of d(ATGAGCGAATA)$_2$ suggest type II G(anti)·A(anti) base pairing (Li et al., 1991). Thus for contiguous mismatches there is also an effect of surrounding sequence on the structure.

This paper reports an NMR study of (dGCGAATAAGCG)$_2$ (designated T-11-mer, Scheme I) which contains adjacent G·A and A·A mismatches. Using a combination of 1D and 2D NMR experiments, we have addressed the following questions: (1) How are the GA and AA mismatches incorporated into the helix? (2) Do they form base pairs and with what hydrogen-bonding scheme? (3) How does pH affect the stability and structure of the duplex?

MATERIALS AND METHODS

NMR Samples. The oligonucleotide dGCGAATAAGCG was purchased from Midland Certified Reagent Co. (Midland, TX) in purified form, and the purity was checked with both reverse-phase and strong anion-exchange HPLC[1] as described previously (Morden et al., 1990).

The concentration of the oligonucleotide was determined optically at 25 °C using 114.3×10^3 M^{-1} cm^{-1} as the extinction coefficient per mole of strand at 260 nm. The NMR samples were prepared as described previously (Morden et al., 1990).

NMR Experiments. All of the NMR spectra in D_2O were taken at 400 MHz on a Bruker AM400 NMR spectrometer. All of the samples were 2 mM in single strand, 10 mM phosphate buffer, 0.1 mM EDTA, and 0.1 M NaCl. The pH of the samples was adjusted with 1 M DCl or 1 M NaOD and monitored directly with a small-diameter pH electrode (Ingold). Chemical shifts are relative to the internal standard TSP.

The NOESY spectra in D_2O (mixing time 50 and 350 ms) were obtained at 5, 15, and 25 °C using the TPPI method (Bodenhausen et al., 1984) with 2048 points over a sweep width of 4500 Hz in the t_2 dimension and 300 points in the t_1 dimension. Each t_1 point is the sum of 144 accumulations. The t_1 dimension was zero-filled to 2048 points, and both dimensions were multiplied by a shifted sine-bell curve prior to transforming the data. The NOESY spectra in H_2O buffer at pH 4.23, 4.88, or 7.15 were recorded as magnitude spectra at 5 °C with mixing times of 125 and 250 ms on a Bruker AM400 spectrometer. A 1–1 pulse sequence was used as the final observation pulse (Hore, 1983a,b). The time domain data consisted of 2048 points in t_2 and 300 points in t_1 with a sweep width of 9600 Hz. For each t_1 increment, 256 scans were accumulated with a 1-s delay between each scan. The t_1 dimension was zero-filled to 1024 points, and both dimensions were apodized with an unshifted sine-bell function prior to transformation. The phase-sensitive NOESY spectrum at pH 4.59 was obtained on a Bruker AMX500 spectrometer at 9 °C using the TPPI method with a 100-ms mixing time, and the observation pulse was replaced by a jump and return sequence (Plateau & Gueron, 1982) with the pulse maximum placed at 15 ppm. The time domain data sets consisted of 2048 complex points in the t_2 dimension and 400 points in the t_1 dimension. For each t_1 increment 112 scans were collected with a repetition delay of 1.8 s between each scan. The t_1 dimension was zero-filled to 2048 points, and both dimensions were multiplied by a shifted sine-bell function prior to transformation. The 2D correlated spectroscopy experiments (COSY) (Aue et al., 1976; Nagayama et al., 1980) and the relayed COSY spectra (Eich et al., 1982; Wagner, 1983) were recorded and processed as described previously (Morden et al., 1990). All of the two-dimensional data were transferred to a Silicon Graphics 4D35TG computer and processed with the program FELIX (Hare Research, Inc.).

The 1D exchangeable proton spectra were obtained using a 1331 pulse sequence (Hore, 1983a,b) for observation. The NOE difference spectra were collected with a 100–500-ms irradiation pulse in groups of 64 accumulations cycled between irradiation on resonance and irradiation off resonance. Spectra were collected at different pH (4.23, 4.59, 4.88, 6.14, 6.72, and 7.15) and temperatures (–5–15 °C).

UV Melting Experiments. UV melting experiments were performed using a Gilford Response II spectrophotometer and were analyzed as described previously (LeBlanc & Morden, 1991).

Circular Dichroism Spectra. CD spectra were recorded on the same samples used in the UV melting experiments using a JASCO J-500C spectropolarimeter interfaced to an OKI if 800 model 30 computer.

Molecular Modeling. All of the modeling calculations were conducted on the duplex d(GCGAA)·d(TAAGC), which represents half of the base-paired region of the T-11-mer where the two halves are related through a *C*2 rotation. The starting structure for the dynamics calculations was generated through

[1] Abbreviations: NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; CD, circular dichroism; COSY, two-dimensional correlated spectroscopy; relayed COSY, two-dimensional relayed coherence transfer spectroscopy; EDTA, ethylenediaminetetraacetic acid; TSP, sodium 3-(trimethylsilyl)[2,2,3,3-$^2H_4$]propionate; HPLC, high-performance liquid chromatography; TPPI, time-proportional phase increment; Pu, purine; Py, pyrimidine.

an embedding process using DSPACE (version 4.0, Hare Research, Inc.). The structure was constrained using 61 constraints from NMR data; 44 of these constraints were extracted from a NOESY spectrum obtained in D_2O with a 50-ms mixing time, and 17 constraints were extracted from a NOESY spectrum obtained in H_2O with a 100-ms mixing time. Although more distances could be determined from the NOESY spectrum in H_2O, only those involving exchangeable protons were used in the structure determination. Median distances were derived using a two-spin approximation and comparing all volume integrals to that of the C10(H5)–C10-(H6) cross-peak where the distance was fixed at 2.48 Å. Distance constraints were entered as upper and lower limits determined by the median distance ± a range. The ranges for the distance constraints were set to ±0.5 Å for all of the distances derived from the H_2O NOESY. For distances derived from the D_2O NOESY, the ranges were ±0.3 Å for distances ≤2.5 Å, ±0.4 Å for distances ≤3.5 Å, and ±0.5 Å for distances >3.5 Å. An additional eight constraints were used for the hydrogen bonds in the Watson–Crick base pairs (G1·C10, C2·G9, and A5·T6). Ten structures were generated by DSPACE, and the one with the least number of violations was used as the starting structure for the simulated annealing procedure described below.

The simulated annealing calculations were conducted using DISCOVER (version 2.8) run from within INSIGHT (version 2.12, BIOSYM, Inc.). The calculations used an AMBER forcefield from within the BIOSYM software. The 61 NMR constraints and eight hydrogen-bond constraints used in the distance geometry procedure were also used during the simulated annealing along with an additional 11 base pair constraints that were applied to the Watson–Crick base pairs according to the procedure of Baleja et al. (1990). The simulated annealing procedure consists of nine steps; throughout all nine steps the constraints described above were applied. The nine steps were as follows: (step 1) steepest descent minimization for 100 iterations; (step 2) conjugate gradient minimization for 500 iterations; (step 3) initialization of the temperature to 600 K; (step 4) molecular dynamics for 3 ps while increasing the force applied to distance constraints from 1× to 6×; (step 5) molecular dynamics for 15 ps at 600 K with the distance constraints at 6×; (step 6) molecular dynamics for 3 ps while lowering the temperature from 600 to 300 K; (step 7) molecular dynamics for 2.5 ps while lowering the temperature from 300 to 50 K and simultaneously lowering the distance constraint force from 6× to 1×; (step 8) steepest descent minimization for 500 iteractions; (step 9) conjugate gradient minimization for 1000 iterations. The simulated annealing procedure was repeated 10 times. The structure presented in this paper contains the least violations to the distance constraints.

Helical parameters were analyzed for the final structure using NEWHEL92 (obtained from R. E. Dickerson, UCLA Molecular Biology Institute).

RESULTS

The temperature behavior of the methyl resonance from the dGCGAATAAGCG single strand was briefly discussed in a previous paper (Morden et al., 1990). These results indicated at low temperature the presence of two conformations in slow exchange on the NMR time scale. A similar behavior is also observed for the aromatic resonances. At 5 °C, 15 resonances are observed between 6.6 and 8.8 ppm, and at 20 °C there are 26 sharp resonances, indicating two conformations in slow exchange. At 30 °C, the two conformations are in

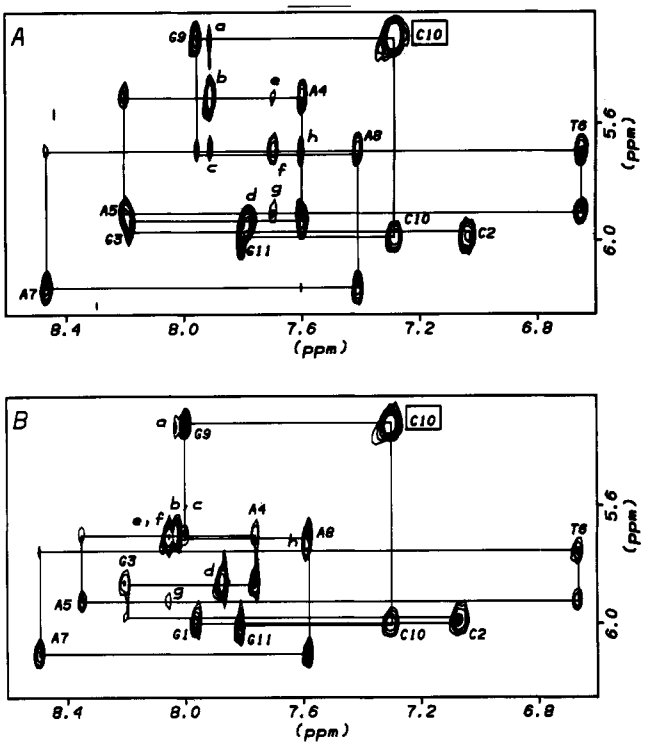

FIGURE 2: Expansion of the 2D NOESY spectrum of the T-11-mer at 5 °C and (A) pH 7.15 or (B) pH 5.1. The solid line indicates the sequential NOE connectivities where the cross-peaks due to intrabase NOEs are labeled by base and sequence position. The boxed labels are adjacent to C(H6)–C(H5) cross-peaks. The lettered cross-peaks are assigned as follows: (a) A8(H2)–G9(H1′), (b) A8(H2)–A4-(H1′), (c) A8(H2)–A8(H1′), (d) A7(H2)–G3(H1′), (e) A4(H2)–A4(H1′), (f) A4(H2)–A8(H1′), (g) A4(H2)–A5(H1′), and (h) A5(H2)–T6(H1′).

Scheme I

```
          1 2 3 4 5
5'      G C G A A T A A G C G     3'
3'    G C G A A T A A G C G       5'
      11 10 9 8 7 6
```

intermediate exchange as evidenced by the broadening of all the resonances. At higher temperatures, the resonances begin to sharpen, and by 45 °C the aromatic region again shows 15 sharp well-resolved resonances.

The present study has focused on the structural characterization of the self-complex as shown in Scheme I. The melting curves obtained from UV absorption studies show a very strong concentration dependence, indicating a multimolecular complex rather than a monomolecular species (unpublished results of LeBlanc and Morden). The CD spectra show a positive band at 273 nm and a negative band at 245 nm (data not shown) and thus are consistent with a right-handed structure. 2D NMR was used to obtain a more detailed view of the structure of this duplex.

Resonance Assignment of the Nonexchangeable Protons. The assignments of the nonexchangeable protons were made using well-established techniques (Feigon et al., 1982; Hare et al., 1983; Scheek, et al., 1983; Weiss et al., 1984). Figure 2, panels A and B, shows an expansion of the 2D NOESY spectra of the duplex obtained in D_2O at 5 °C at pH 7.15 and 5.1, respectively. The cross-peak pattern of an intranucleotide interactions between H6 or H8 and H1′ followed by a sequential interaction between H1′ and H6 or H8 from the adjacent nucleotide that is characteristic of a B-form helix is observed. Although these cross-peaks suggest that the bases from G1 to G11 stack without interruption, the following

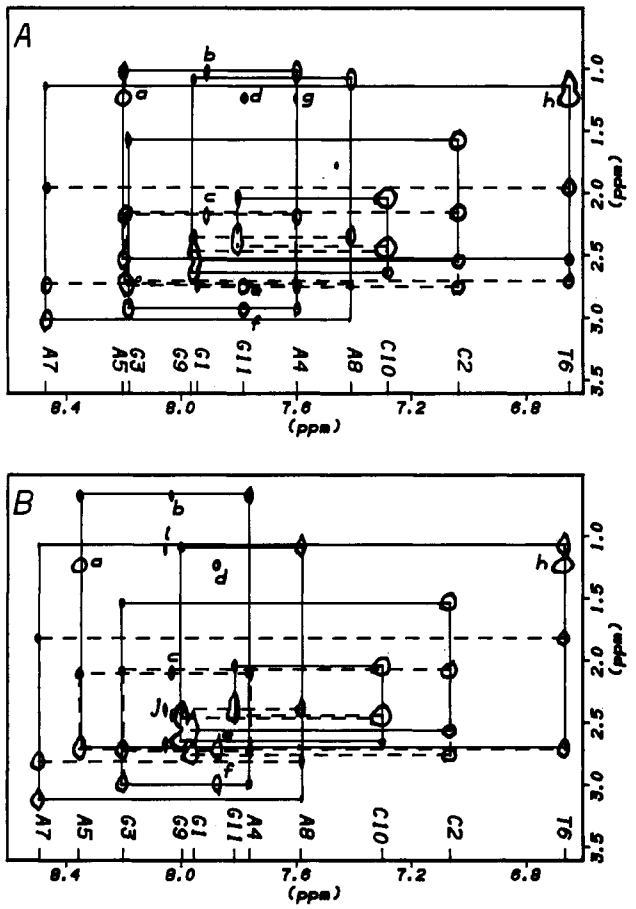

FIGURE 3: Expansion of the 2D NOESY spectrum of the T-11-mer at 5 °C and (A) pH 7.15 or (B) pH 5.1. The solid line follows the aromatic to H2′ NOE network, and the dashed line follows the aromatic to H2″ NOE network. The chemical shifts of the aromatic protons are indicated along the *x*-axis. The lettered cross-peaks can be assigned as follows: (a) A5(H8)–T6(Me), (b) A8(H2)–A4(H2′), (c) A8(H2)–A4(H2″), (d) A7(H2)–T6(Me), (e) A7(H2)–G3(H2′), (f) A7(H2)–G3(H2″), (g) A4(H8)–T6(Me), (h) T6(H6)–T6(Me), (i) A4(H2)–A8(H2′), and (j) A4(H2)–A8(H2″).

very strong interstrand interactions indicate considerable overlap between G3 and A7, as well as A4 and A8: A4-(H2)–A8(H1′) (cross-peak f in Figure 2A,B), A8(H2)–A4-(H1′) (cross-peak b in Figures 2A,B), and A7(H2)–G3(H1′) (cross-peak d in Figure 2A,B). These cross-peaks indicate that rather than an intrastrand stacking interaction of the bases in the mismatches there is an unusual interstrand stacking.

The expected sequential interactions between H6 or H8 and H2′, H2″ protons are shown in Figure 3,panels A and B. A network similar to that observed for H1′ protons is observed for H2′ protons (designated by a solid line) and for H2″ protons (designated by a dashed line). The NOESY spectrum obtained with short mixing time (50 ms) and the COSY spectrum were used to distinguish H2′ and H2″ cross-peaks. The NOESY spectrum (Figure 3) also shows some intriguing cross-peaks affiliated with the unusual conformation of this duplex; interstrand cross-peaks are observed for A7(H2)–G3(H2′) (peak e), A7(H2)–G3(H2″) (peak f), A8(H2)–A4(H2′) (peak b), and A8(H2)–A4(H2″) (peak c).

Assignments of the H3′ and H4′ resonances were made from the H1′–H3′ and H1′–H4′ connectivities of the NOESY spectrum. These assignments were independently confirmed using a relayed COSY experiment to detect H1′–H3′ cross-peaks. The assignments can also be independently corroborated by H3′ and H4′ cross-peaks to the aromatic region, where a complete H3′- and H4′-base proton sequential walk is observed (data not shown). Also observed in this region of the NOESY spectrum are unusual cross-peaks between adenosine H2 protons and the H3′ or H4′ protons from the 3′ neighbor on the complementary strand. There are at least 11 H2–H3′ or H2–H4′ cross-peaks, of which A4(H2)–A8-(H4′), A7(H2)–G3(H3′), and A8(H2)–A4(H4′) are very strong. The A4(H2)–A8(H4′) and A8(H2)–A4(H4′) cross-peaks are visible in a NOESY spectrum acquired with a 50-ms mixing time indicating reasonably short distances between H2 and H4′ atoms. These cross-peaks are unusual in that they are not expected if the mismatches have a classical B-DNA structure. The observation of these cross-peaks supports the model that the helix is overwound at the G3·A8/A4·A7 step leading to stacking of G3 on A7 and A4 on A8.

The striking similarities in the NOESY cross-peaks and connectivities observed at pH 7.15 and 5.1 (Figures 2 and 3) indicate that the neutral and the low-pH structures are similar. The assignments of the nonexchangeable proton resonances for the T-11-mer duplex at 5 °C and pH 7.15 are given in Table I.

Resonance Assignment of the Exchangeable Protons. The G1, G9, and G11 imino resonances are assigned using the strong NOEs to the base-paired cytidine amino protons, the temperature dependence of the line width, and the chemical shift. The resonance at 12.96 ppm (Figure 4A) is assigned to G1(H1) and G9(H1). The G11(H1) resonance at 10.62 ppm rapidly broadens and disappears on raising the temperature above −5 °C as might be expected for the dangling terminal residue. An NOE was not observed upon saturation of the G11(H1). The resonance at 12.71 ppm (Figure 5B) was assigned to T6(H3) on the basis of observed NOEs to A5(H2) (strong), A7(H2) (weaker), and an exchangeable proton resonance at 11.22 ppm [A4($NH_{2,b}$)]. In the 2D NOESY spectrum in H_2O (Figure 6), we observe cross-peaks between the T6(H3) and the A5(H2) (peak j), A7(H2) (peak h), and two amino protons of A5 located at 7.72 and 6.76 ppm (peaks i and k). The T6(H3) resonance is upfield from the chemical shift typically observed (13.5–14.5 ppm) for thymidine imino resonances, probably due to unusual stacking with the adjacent A·A mismatch.

The G3(H1) resonance at 10.51 ppm is assigned by NOEs with G9(H1), A7(H8), and the three hydrogen-bonded amino protons of C2, G3, and A4 (Figure 5D). The G3(H1) resonance is far upfield from the typically observed chemical shift in a base pair (12.5–13.5 ppm), indicating that the G3-(H1) proton is most likely not involved in a hydrogen bond to another base but is shielded by stacking with adjacent bases.

The amino group of cytidine differs from those of guanosine and adenosine in that it shows restricted rotation about the carbon–nitrogen bond (McConnell & Seawell, 1972, 1973; Raszka & Kaplan, 1972; Raszka, 1974; Mcconnell, 1984). Using NMR, well-resolved proton resonances have been observed for the hydrogen-bonded (8.0–8.5 ppm) and exposed amino protons (6.4–6.9 ppm) of cytidine in oligonucleotide duplexes (Patel, 1976; Boelens et al., 1985; Sklenar et al., 1987). In contrast, broad resonances are observed for the amino protons of adenosine (6–7 ppm) and guanosine (5.5–6.0 ppm).

In the 2D NOESY spectrum (Figure 6), strong cross-peaks are observed between ($NH_{2,e}$) and ($NH_{2,b}$) resonances for C2 and C10 (peaks aa and jj). The assignment as cytidines is also confirmed by NOEs from the complementary G(H1) to the hydrogen-bonded amino proton resonance observed in the 2D spectrum (Figure 6, peaks e and b) and the 1D NOE

Table I: Chemical Shifts of Nonexchangeable Protons at 5 °C and pH 7.15[a]

residue	H8/H6	H2	H5/CH$_3$	H1′	H2′	H2″	H3′	H4′
G1	7.96			6.01	2.56	2.77	4.84	4.25
C2	7.06		5.03	6.00	1.60	2.18 (–0.10)	4.83	4.23
G3	8.20			5.96 (–0.11)	2.96	2.77	5.16	4.55
A4	7.62 (0.14)	7.72 (0.33)		5.53 (0.15)	1.05 (–0.38)	2.21 (–0.11)	4.78	4.37
A5	8.21 (0.13)	7.62		5.93	2.55 (0.17)	2.73	4.93	4.42
T6	6.67		1.26	5.71	1.16	1.98 (–0.15)	4.73	4.13
A7	8.49	7.80		6.18	3.04	2.76	5.00	4.57
A8	7.43 (0.15)	7.93 (0.10)		5.71	1.12	2.37	4.82	4.38
G9	7.97			5.33	2.66	2.49	4.93	4.35
C10	7.30		5.33	6.00	2.06	2.47	4.77	4.27
G11	7.82			6.04	2.42	2.33	4.63	4.06

[a] All of the chemical shifts are reported relative to TSP. Values of δ(pH 5.10) – δ(pH 7.15) ≥ 0.1 ppm are shown in parentheses.

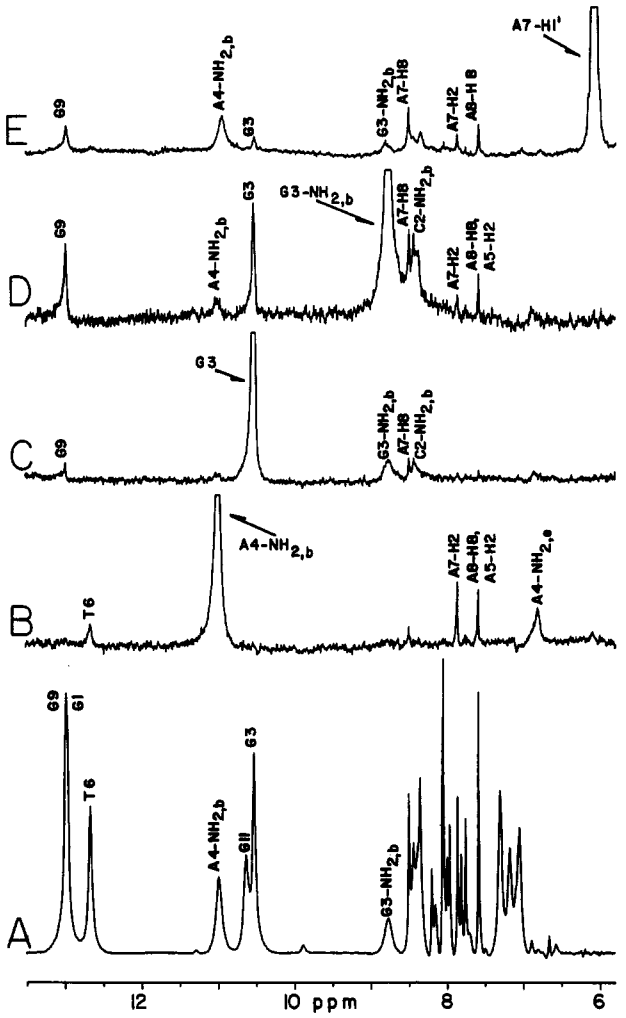

FIGURE 4: (A) NMR spectrum of the T-11-mer at pH 4.88 and 5 °C and NOE difference spectra using 500-ms saturation of (B) A4-($NH_{2,b}$), 10.98 ppm, (C) G3(H1), 10.51 ppm, (D) G3($NH_{2,b}$), 8.77 ppm, and (E) A7(H1′), 6.08 ppm. The saturated resonance is designated by an arrow.

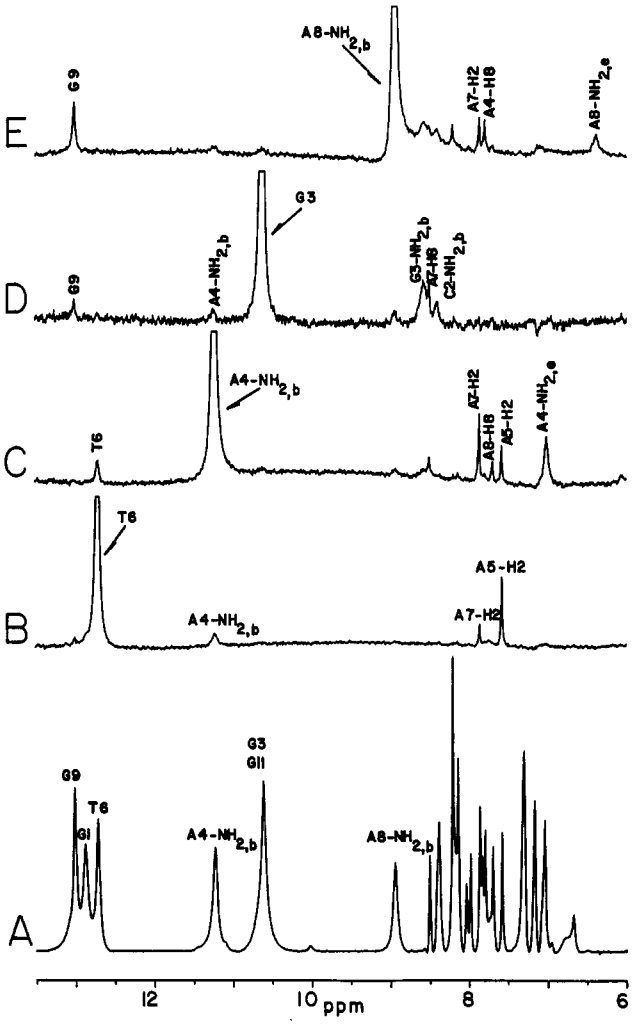

FIGURE 5: (A) NMR spectrum of the T-11-mer at pH 4.23 and 5 °C and NOE difference spectra using 500-ms saturation of (B) T6-(H3), 12.70 ppm, (C) A4($NH_{2,b}$), 11.22 ppm, (D) G3(H1), 10.61 ppm, and (E) A8($NH_{2,b}$), 8.95 ppm. The saturated resonance is designated by an arrow.

studies at different pH values and temperatures (data not shown).

The adenosine amino protons have been assigned through a combination of 2D and 1D NOE experiments. The A4-($NH_{2,b}$) resonance is at 11.22 ppm (Figure 5C, 10.98 ppm in Figure 4B) and exhibits NOEs to T6(H3), A5(H2), A7(H2), A8(H8), and A4($NH_{2,e}$). Saturation of the A7(H1′) proton (Figure 4E) results in a strong NOE ato A4($NH_{2,b}$). In the NOESY spectrum in H_2O (Figure 6), we observe cross-peaks between the hydrogen-bonded and exposed amino protons of A4 (peak n), and from the A4($NH_{2,b}$) to the A7(H1′) (peak o) and A7(H2) (peak m). The A5(NH_2) resonances have been assigned from observation in the NOESY spectrum (Figure 6) of cross-peaks to T6(H3) (peaks i and k) and the cross-peak between A5($NH_{2,b}$) and A5($NH_{2,e}$) (peak ee). The A8($NH_{2,b}$) is identified by the 1D NOE (Figure 5E) to G9-(H1), A7(H2), A4(H8), and A8($NH_{2,e}$). In the 2D NOESY spectrum (Figure 6), cross-peaks are observed from A8($NH_{2,b}$) to G3($NH_{2,e}$) (peak z), to A8($NH_{2,e}$) (peak x), and to G3-(H1′) (peak y).

The guanosine amino proton resonances were the most difficult to assign; however, at pH 4.88 the G3($NH_{2,b}$) has

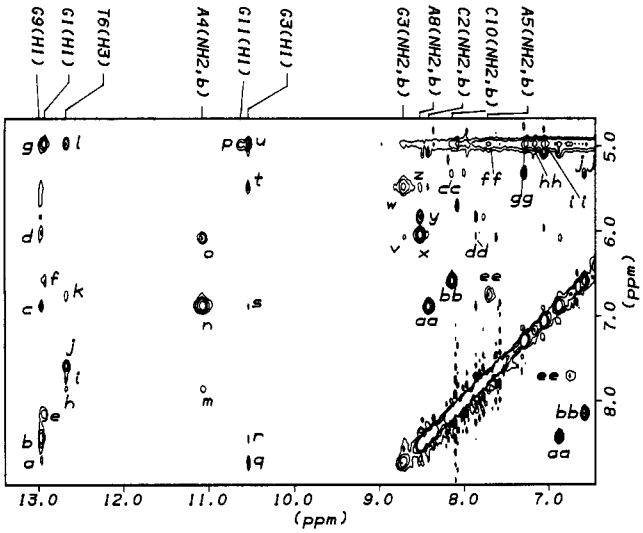

FIGURE 6: Expansion of the 500-MHz NOESY spectrum of the T-11-mer recorded in H_2O at pH 4.6 and 9 °C. Cross-peaks are assigned as follows: (a) G9(H1)–G3($NH_{2,b}$), (b) G9(H1)–C2($NH_{2,b}$), (c) G9(H1)–C2($NH_{2,e}$), (d) G9(H1)–A8($NH_{2,e}$), (e) G1(H1)–C10($NH_{2,b}$), (f) G1(H1)–C10($NH_{2,e}$), (g) G1(H1) exchange with H_2O, (h) T6(H3)–A7(H2), (i) T6(H3)–A5($NH_{2,b}$), (j) T6(H3)–A5(H2), (k) T6(H3)–A5($NH_{2,e}$), (l) T6(H3) exchange with H_2O, (m) A4($NH_{2,b}$)–A7(H2), (n) A4($NH_{2,b}$)–A4($NH_{2,e}$), (o) A4($NH_{2,b}$)–A7(H1′), (p) G11(H1) exchange with H_2O, (q) G3(H1)–G3($NH_{2,b}$), (r) G3(H1)–C2($NH_{2,b}$), (s) G3(H1)–C2($NH_{2,e}$), (t) G3(H1)–G3($NH_{2,e}$), (u) G3(H1) exchange with H_2O, (v) G3($NH_{2,b}$)–A8($NH_{2,e}$), (w) G3($NH_{2,b}$)–G3($NH_{2,e}$), (x) A8($NH_{2,b}$)–A8($NH_{2,e}$), (y) A8($NH_{2,b}$)–G3(H1′), (z) A8($NH_{2,b}$)–G3($NH_{2,e}$), (aa) C2($NH_{2,b}$)–C2($NH_{2,e}$), (bb) C10($NH_{2,b}$)–C10($NH_{2,e}$), (cc) C10($NH_{2,b}$)–C10(H5), (dd) A7(H2)–A8($NH_{2,e}$), (ee) A5($NH_{2,b}$)–A5($NH_{2,e}$), (ff) A5($NH_{2,b}$) exchange with H_2O, (gg), (hh), and (ii) G1(NH_2), G9(NH_2), and G11(NH_2) exchange with H_2O and (jj) C10($NH_{2,e}$)–C10(H5).

been assigned to the resonance at 8.77 ppm. Upon excitation of this resonance, NOEs are observed to G3(H1), G9(H1), C2($NH_{2,b}$), A4($NH_{2,b}$), A7(H8), and A8(H8) (Figure 4D). In the NOESY spectrum in H_2O (Figure 6), a cross-peak is also observed between G3($NH_{2,b}$) and G3($NH_{2,e}$) (peak w). Assignment of the amino proton resonances for G1, G9, and G11 has proved elusive. Irradiation of the G1(H1), G9(H1), or G11(H1) at temperatures between –5 and 15 °C did not produce NOEs that could be attributed to guanosine amino protons. However, in the NOESY spectrum we have observed three cross-peaks involving protons that are in exchange with water (peaks gg, hh, and ii in Figure 6). These resonances, located at 7.04, 7.17, and 7.30 ppm, have been tentatively assigned to the amino proton(s) of G1, G9, and G11 although NOEs were not detected upon saturation of these resonances.

Observation of the relatively sharp amino resonances from adenosines A4 and A8 and guanosine G3 is unusual, and our assignments indicate these resonances are shifted downfield by 3–4 ppm from the positions expected for Watson–Crick base pairs. The magnitude of this downfield shift may reflect, in part, a change in the hydrogen-bond acceptor from a carbonyl oxygen to a ring nitrogen. The assignments of all the exchangeable proton resonances for the duplex at 5 °C are given in Table II.

Effect of pH on the Nonexchangeable Protons. The pK_a values for dC and dA have been determined using proton chemical shifts to be 4.3 and 3.8, respectively (Sowers et al., 1986). These studies demonstrate that the A(H8), A(H2), C(H5), and C(H6) resonances show a significant downfield shift upon protonation. In the T-11-mer there is no significant shift in the C(H5) or C(H6) resonances as a function of pH,

Table II: Chemical Shifts of Exchangeable Protons[a]

base pair	T(H3)	G(H1)	$C(H4)_{b,e}$	$A(H6)_{b,e}$	$G(H2)_{b,e}$
G1·C10		12.96	8.15, 6.58		
C2·G9		12.96	8.44, 6.88		
G3·A8		10.51		8.37, 6.38	8.77, 5.50
A4·A7				10.98, 6.80	
A5·T6	12.64			7.70, 6.75	
G11		10.62			

[a] All of the chemical shifts are reported in parts per million relative to TSP at 5 °C and pH 4.88.

indicating that the cytidines are not protonated. Figure 7 shows the chemical shift of the A(H8) and A(H2) resonances from the T-11-mer as a function of pH. The largest shift is observed for A4(H2) which changes from 7.72 ppm at pH 7.15 to 8.22 ppm at pH 4.16. Significant shifts are also observed for the H2 and H8 resonances of A8 upon changing the pH. These chemical shift changes are in agreement with those expected upon protonation of the N1 ring nitrogen (Sowers et al., 1986) and thus indicate that the adenosines A4 and A8 in the T-11-mer are protonated at low pH. The changes induced upon lowering the pH cannot be attributed to decomposition of the sample because the original spectrum is reproduced upon adjusting the pH back to a neutral value.

Effect of pH on the Exchangeable Proton Resonances. The 1D exchangeable proton spectra (8.6–13.5 ppm) of the duplex at 5 °C and different pH are plotted in Figure 8. The line width of the G1(H1), T6(H1), and G9(H1) resonances change very little with pH (Figure 9A). The line width of the 3′-terminal G11(H1) resonance does not change in the pH range 7.15–6.3, narrows significantly from ~84 Hz at pH 6.3 to ~30 Hz at pH 5, and broadens 10 Hz from pH 5 to 4.7. G11 is not involved in base-pair hydrogen bonds and thus would be more susceptible to exchange with the solvent. The line narrowing can therefore be explained by diminished exchange with solvent at the lower pH. The G3(H1) resonance also displays narrowing as the pH is decreased; however, the reduction in the line width is not as drastic as for G11. Due to resonance overlap, the line widths of the G3(H1) and G11(H1) resonances cannot be measured accurately below pH 4.7.

Figure 9B shows the effect of pH on the chemical shift of the resonances for the imino protons and selected amino protons. The chemical shift of the G1(H1) and G9(H1) resonances do not change in the pH range 7.15–5.0. Below pH 5 the G1(H1) resonance moves slightly upfield and that of G9(H1) downfield. The chemical shift of G11(H1) does not change appreciably over the pH range ($\Delta\delta$ = –0.01 ppm). The T6(H3) signal is constant to pH 6.5 and then moves 0.2 ppm downfield upon further reduction in the pH. Upon lowering the pH, the G3(H1) resonance moves downfield ($\Delta\delta$ = 0.3 ppm). At neutral pH, G3($NH_{2,b}$) and A4($NH_{2,b}$) resonances are observed as one very broad signal which becomes asymmetric at pH 6.51 and splits into two broad separated resonances at pH 6.14. Lowering the pH results in downfield shifts of A4($NH_{2,b}$) ($\Delta\delta$ = 2.3 ppm) and A8($NH_{2,b}$) ($\Delta\delta$ = 1 ppm); in contrast, an upfield shift is observed for G3($NH_{2,b}$) ($\Delta\delta$ = –0.4 ppm). These shifts may be partially due to changes in the strength of the hydrogen bonds in which these amino protons are involved. The C2($NH_{2,b}$) resonance moves upfield ($\Delta\delta$ = –0.1 ppm) with decreasing pH, in contrast to C2($NH_{2,e}$), which moves downfield ($\Delta\delta$ = 0.17 ppm). The C10(NH_2) resonances do not shift significantly until below pH 5 where the C10($NH_{2,e}$) resonance shifts downfield.

Figure 10 compares the pH dependence of the melting transition of the duplex determined from UV absorption

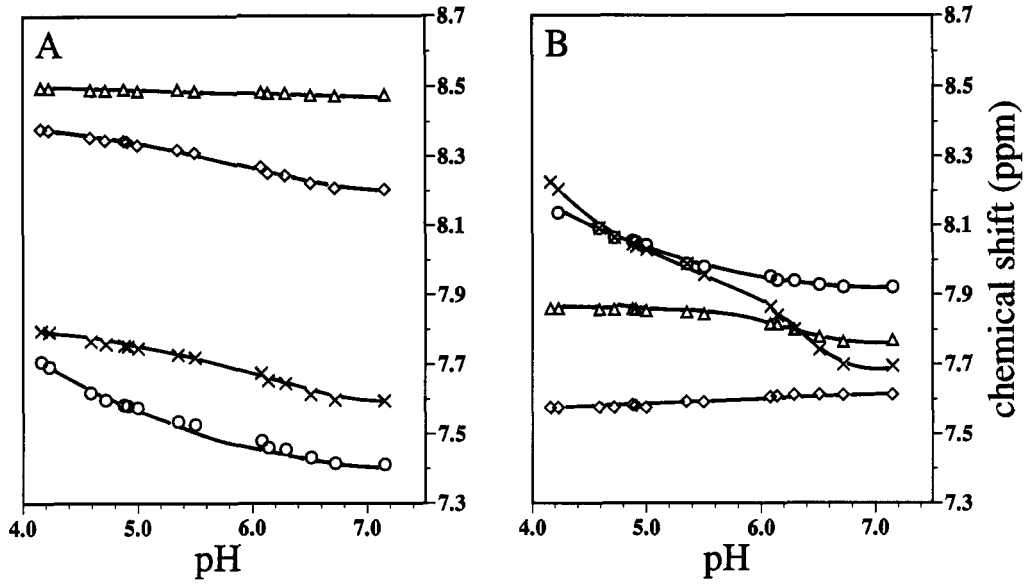

FIGURE 7: pH dependence of the adenosine (A) H8 and (B) H2 proton chemical shifts of the duplex at 5 °C. (×) A4, (◇) A5, (△) A7, (○) A8.

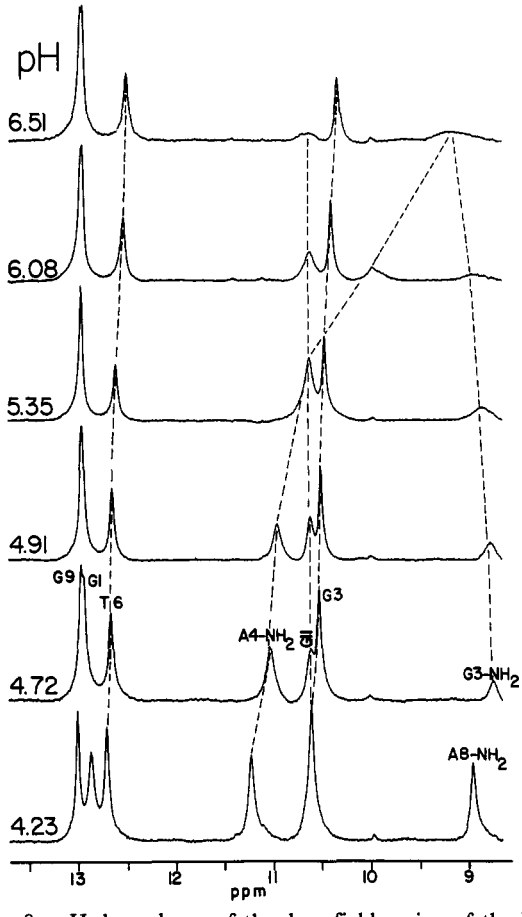

FIGURE 8: pH dependence of the downfield region of the NMR spectrum at 5 °C.

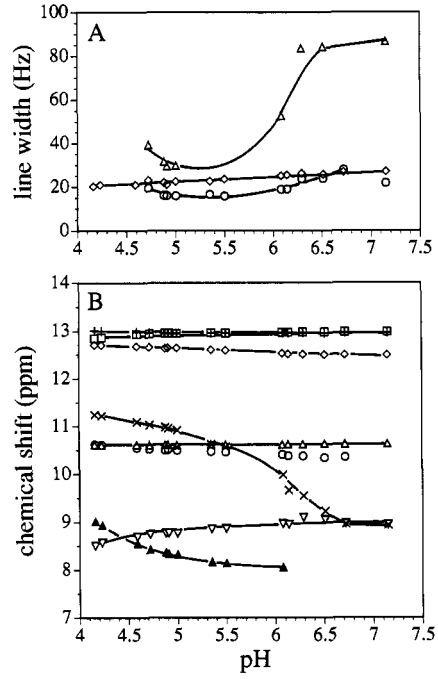

FIGURE 9: pH dependence of (A) the line widths of the imino proton resonances: (○) G3, (◇) T6, and (△) G11 of the T-11-mer duplex at 5 °C. (B) Chemical shifts of exchangeable protons at 5 °C: (+) G9(H1), (□) G1(H1), (◇) T6(H3), (×) A4($NH_{2,b}$), (○) G3(H1), (△) G11(H1), (▽) G3($NH_{2,b}$), and (▲) A8($NH_{2,b}$).

experiments with the pH dependence of the A4($NH_{2,b}$) chemical shift. Decreasing pH from 7 to 4 results in a T_m increase from 28 to 41 °C. The apparent pK_a of protonation in the duplex is estimated from these curves to be pH 5.1–5.9, which is more than one pH unit higher than the pK_a of free adenosine. Thus the local envirnment in the oligonucleotide has greatly influenced the protonation of the adenosine.

Temperature Dependence of Exchangeable Protons. Spectra of the downfield region recorded at pH 4.88 and 7.15 as a function of temperature are shown in Figure 11, panels A and B. At pH 4.88 six resonances are observed in the spectrum between 8.5 and 13.5 ppm (Figure 11A). The broadening observed below 15 °C is due in part to the greater viscosity of the solution. When the temperature is raised to 15 °C, the resonances of the G1, G9, and T6 imino protons get narrower. However, peaks corresponding to the A4($NH_{2,b}$), G11(H1), and G3($NH_{2,b}$) broaden. The resonance of G3(H1) does not change in width up to 15 °C at pH 4.88 and up to 10 °C at pH 7.15. At 15 °C the line widths of the G3(H1) and T6(H3) resonances are approximately the same (~16 Hz); however, above 15 °C the broadening of the G3(H1) resonance is larger than that of the T6(H3) resonance, especially at pH 7.15. Above 15 °C the relaxation rates (the changes in line width) rise due to an increasing contribution from exchange

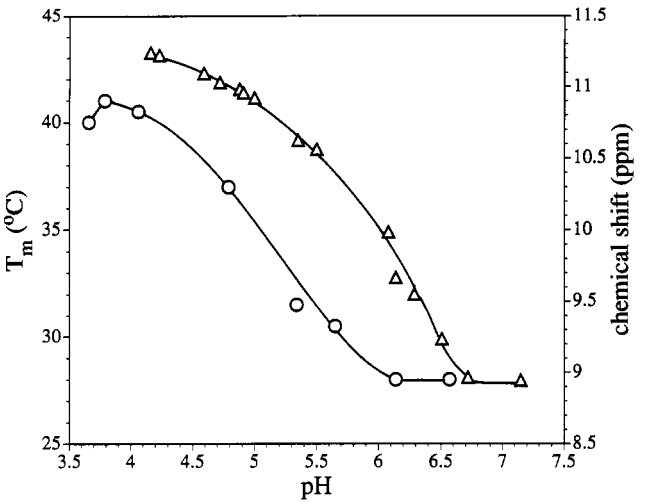

FIGURE 10: pH dependence of melting transitions (O) and the A4-($NH_{2,b}$) chemical shift (Δ) of the T-11-mer. For the melting temperatures the concentration was 8×10^{-5} M strand in 1 M NaCl.

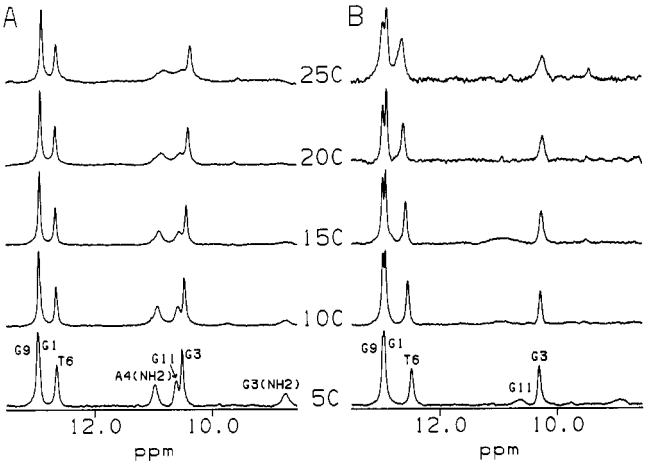

FIGURE 11: Temperature dependence of the downfield region of the NMR spectra (8.5–13.5 ppm) at (A) pH 4.88 and (B) pH 7.15. Resonances labeled only with sequence position refer to either G(H1) or T(H3) resonances.

with solvent. With increase in temperature, all exchangeable proton resonances, except T6(H3), shift upfield. The downfield shift of the T6(H3) resonance, which is negligible at pH 4.88 (0.03 ppm), is much greater at pH 7.15 (~0.3 ppm). Above 35 °C the relaxation rates increase very rapidly, and the exchangeable resonances are no longer observed.

DISCUSSION

Structure of the Mismatch Base Pairs. The structure of the G3·A8 mismatch can be characterized from analysis of the experiments presented in Figures 4–6. The 3–4 ppm downfield shift of the amino protons of G3 and A8 reflects a change in the hydrogen-bond acceptor from a carbonyl oxygen to a ring nitrogen (Gao & Patel, 1987). The observed NOEs from G3($NH_{2,b}$) to A8(H8) and from A8($NH_{2,b}$) to G3(H1′) unambiguously determine the spatial arrangement of the G·A base pair. The proposed structure corresponds to the mismatched type II G(anti)·A(anti) base pair as shown in Figure 1. One of the unusual features of this base pair is the involvement of N3 or N7 as an acceptor and an amino group as a donor in the base-pairing hydrogen bonds. Such a base pair is predicted from base geometries (Saenger, 1984), and its existence has been suggested by several authors (Li et al., 1991; Heus & Pardi, 1991). However, to the best of

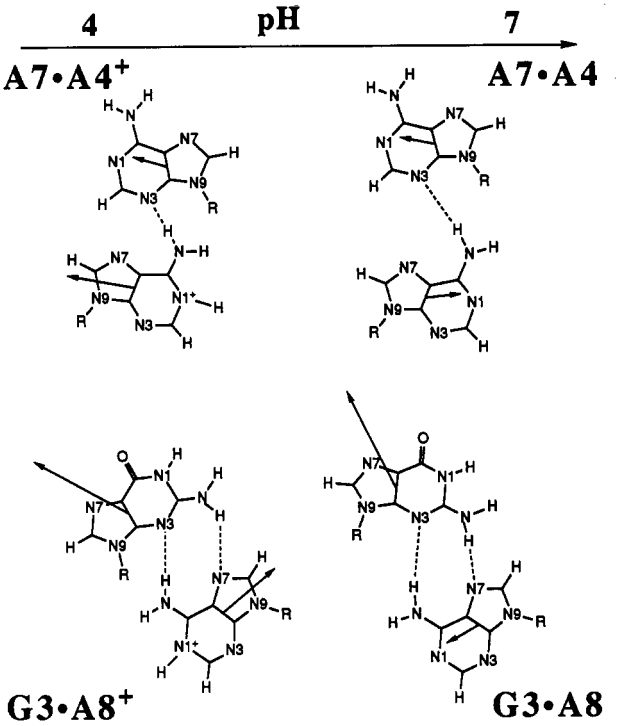

FIGURE 12: Schematic of the dipole moments and the structure of the G·A and A·A base pairs as the function of pH.

our knowledge, this work represents the first observation of experimental NOE evidence for this type of hydrogen bonding between G and A.

The structure of the A·A base pair is defined by the following planar interactions observed in the 1D and 2D NMR experiments: the NOEs from A4($NH_{2,b}$) to A7(H1′), from A4($NH_{2,b}$) to A7(H2), and from A7(H2) to A4(H8). The proposed structure of the A(anti)·A(anti) base pair with a single hydrogen bond between A4($NH_{2,b}$) and A7(N3) is shown in Figure 12. This is also the first direct NMR evidence to indicate the formation of a hydrogen-bonded A·A mismatch pair.

An anti conformation for all of the glycosidic bonds in the G·A and the A·A mismatches is a critical feature of our structural model. One way to distinguish between the anti and the syn conformation of the glycosidic bond is to compare the intensities of the NOE cross-peaks for the intranucleotide cross-peak (H8 or H6)–(H1′) for the nucleotide of interest with the intensity of a cytidine (H5)–H6) cross-peak from the same molecule. The (H5)–(H6) interproton distance is fixed at ~2.5 Å, while the (H8 or H6)–(H1′) interproton distance varies from ~3.7 Å for an anti conformation to ~2.5 Å for a syn conformation of the glycosidic bond. For the T-11-mer, we compared intensities for these cross-peaks in a NOESY spectrum obtained with a 50-ms mixing time. All the (H8 or H6)–(H1′) intranucleotide cross-peaks were absent or very weak compared to the C(H5)–C(H6) cross-peaks. This suggests that all of the nucleotides in the T-11-mer, including those involved in the mismatched base pairs, are in an anti conformation. For all of the purines involved in the mismatched base pairs, the order of the relative intensities of intranucleotide cross-peaks is H8-H2′ > H8-H2″ > H8-H1′, which is also characteristic of an anti conformation. Finally, if one of the bases in G3·A8 or A4·A7 adopted a syn conformation, we would expect to observe NOE cross-peaks for A8(H2)–G3(H8), or A4(H2)–A7(H8), or A7(H2)–A4-(H8); however, none of these cross-peaks are observed in the NOESY spectra of the T-11-mer. Hence we conclude that

Scheme II

the mismatched base pairs are $G3_{anti}$·$A8_{anti}$ and $A4_{anti}$·$A7_{anti}$.

Many of the NOEs presented in Figures 4–6 which are crucial to the structure determination have only been observed upon lowering the pH. The lower pH is used because the resonances of amino protons are better resolved and sharper under these conditions. The same results were obtained for the samples at neutral pH and lower temperatures (below 0 °C); however, much higher decoupler power was required to saturate the broader resonances, and excitation of multiple resonances became a serious problem. The similarities in the NOE connectivities between base and sugar protons at pH 7.15 and 5.1 (Figures 2 and 3) indicate that the neutral and the low-pH forms of the duplex are structurally similar.

Base Pair Stacking. One of the more interesting features of the molecule is the stacking that occurs between G3·A8 and A4·A7. These adjacent base pairs exhibit interstrand stacking, stacking with the sequential base from the opposite strand, rather than stacking with the base which is sequential in the same strand. A schematic of the stacking for the entire duplex is shown in Scheme II. The three-dimensional characteristics of this interstrand stack can be seen in the stereoviews in Figure 13 and enlarged and viewed along the helix axis as base-pair stacks in Figure 14. Evidence for this stacking comes from several observations. (1) There are very strong interstrand interactions: A7(H2)–G3(H1′), A8(H2)–A4(H1′), and A4(H2)–A8(H1′) (Figure 2A,B). These cross-peaks rule out a standard B-form stack between G3, A4, A7, and A8 and instead indicate that there is considerable overlap between G3 and A7, as well as A4 and A8. This interrupts the intrastrand stacking interaction which is substituted by an unusual interstrand stacking of the bases. (2) There are interstrand cross-peaks between the A(H2) and H2′,H2″ protons: A7(H2)–G3(H2″) is observed as a weak cross-peak at 50 ms; A7(H2)–G3(H2′) and A8(H2)–A4(H2″) are observed as very weak cross-peaks at 50 ms. The cross-peaks for A7(H2)–G3(H1′) and A8(H2)–A4(H1′) are very strong, and thus the cross-peaks observed to the H2′ and H2″ protons could be due to spin diffusion. (3) There are unusual cross-peaks between adenosines H2 protons and the H3′ and H4′ protons, the strongest of these being A4(H2)–A8(H4′), A7-(H2)–G3(H3′), and A8(H2)–A4(H4′). (4) There are interstrand NOE interactions between the exchangeable protons and the aromatic or sugar protons: G3(H1)–A7(H8), G3-(H1)–A7(H1′), G3(H1)–A4($NH_{2,b}$), G3($NH_{2,b}$)–A4(NH2,b), G3($NH_{2,b}$)–A7(H8), G3($NH_{2,b}$)–A7(H2), G3($NH_{2,b}$)–A7-(H1′), A4($NH_{2,b}$)–A8(H8), A8($NH_{2,b}$)–A4(H8), and A8-($NH_{2,b}$)–A7(H2). Such cross-peaks are not usually observed for B-form DNA and thus indicate a deviation from the classical B-form structure.

Chemical Shift Considerations. In a conventionally stacked B-DNA helix the H8 signal usually appears downfield from the H2 resonance as observed for the A5 and A7 resonances of the T-11-mer. The A5(H2) and A7(H2) exhibit similar chemical shifts in acidic and neutral conditions (Table I). The resonances are upfield of the single-strand unstacked value of ~8.0 ppm, reflecting the shielding of these H2 protons by ring current contributions from flanking base-pairs. By contrast, the A4(H2) and A8(H2) at pH 5.1 (8.05 and 8.03 ppm, respectively) are close to the single-strand values demonstrating that these H2 protons experience little ring current contributions from flanking base pairs. This is compatible with the proposed model of the G3·A8 and A4·A7 mismatch base pairs (Figure 14) that exhibits poor overlap between the A4(H2) or A8(H2) and flanking base pairs. The A5(H2) and A7(H2) are located in the center of the base pair and thus have extensive overlap whereas the A4(H2) and A8(H2) are in the groove and thus experience less overlap.

The H2′ and H2″ resonances of A4 and A8 are shifted upfield considerably, in comparison to the position of the analogous resonances of the other nucleotides in the T-11-mer. This is in agreement with the modeling studies that show the H2′ and H2″ protons of A4 are inside the shielding zone of A5 and those of A8 are inside the shielding zone of G9. Upfield shifts of similar magnitude have also been observed for the H2′ and H2″ resonances of the A residue in other G·A mismatches (Orbons et al., 1987; Li et al., 1991). The other bases in the mismatches, G3 and A7, have H2′ and H2″ resonances that are shifted downfield in comparison to the position of the analogous resonances of the Watson–Crick base pairs in the T-11-mer.

The distance between H1′ and H2″ is always less than that between H1′ and H2′ independent of the pseudorotation phase angle of the sugar ring (van de Ven & Hilbers, 1988). Therefore, in the NOESY spectrum at 50-ms mixing time, the cross-peak intensity was used to differentiate these two cross-peaks. In the COSY spectrum the H1′–H2′ cross-peaks usually exhibit four broad, widely spaced lobes, whereas the H1′–H2″ cross-peaks exhibit from eight to 16 (depending on the digital resolution) narrow, closely spaced lobes (Chazin et al., 1986). The cross-peak fine-structure is a function of the coupling constants and ultimately the sugar conformation. For most B-form duplexes the H2′ resonance of a given nucleotide is upfield of the H2″ resonance for that nucleotide. Thus, in the COSY spectrum of a B-form duplex, the four-lobe cross-peak for a particular nucleotide will be observed upfield from the 8- or 16-lobe cross-peak. However, in the T-11-mer the more complex cross-peak is observed upfield for G3, A7, G9, and G11. G11 is the 3′ terminus, and this chemical shift change has been observed for other 3′-terminal residues (Wemmer et al., 1984; Weiss et al., 1984). The chemical shift change has also been observed for the cytidine residue in A·C mismatch base pairs (Gao & Patel, 1987).

Protonation of Adenosines Results in Stabilization of the Duplex. The chemical shifts of the A4(H2) and A8(H2), compared to the other H2 protons, are sensitive to pH. For example, the A4(H2) resonance moves 0.5 ppm downfield upon lowering the pH. There is also a downfield change in the chemical shift of the A4(H8) and A8(H8) resonances upon lowering the pH. A downfield shift of the adenosine H2 and H8 resonances upon protonation at the N1 position has been observed previously (Sowers et al., 1986). The results for the T-11-mer strongly indicate that A4 is protonated at N1 upon lowering the pH. The data also suggest that the A8(N1) position is protonated although the chemical shift changes are not as large as those observed for A4. A downfield change is observed in the chemical shift of the A5(H8) resonance; however, there is no concurrent downfield shift observed for the A5(H2) resonance; in fact, the A5(H2) resonance moves slightly upfield upon lowering the pH (Figure 7).

The change in chemical shift is accompanied by an increase in the stability of the duplex demonstrated by an increase in the optically determined T_m from 28 to 41 °C (Figure 10). The CD spectrum observed at pH 7 is characteristic of B-form

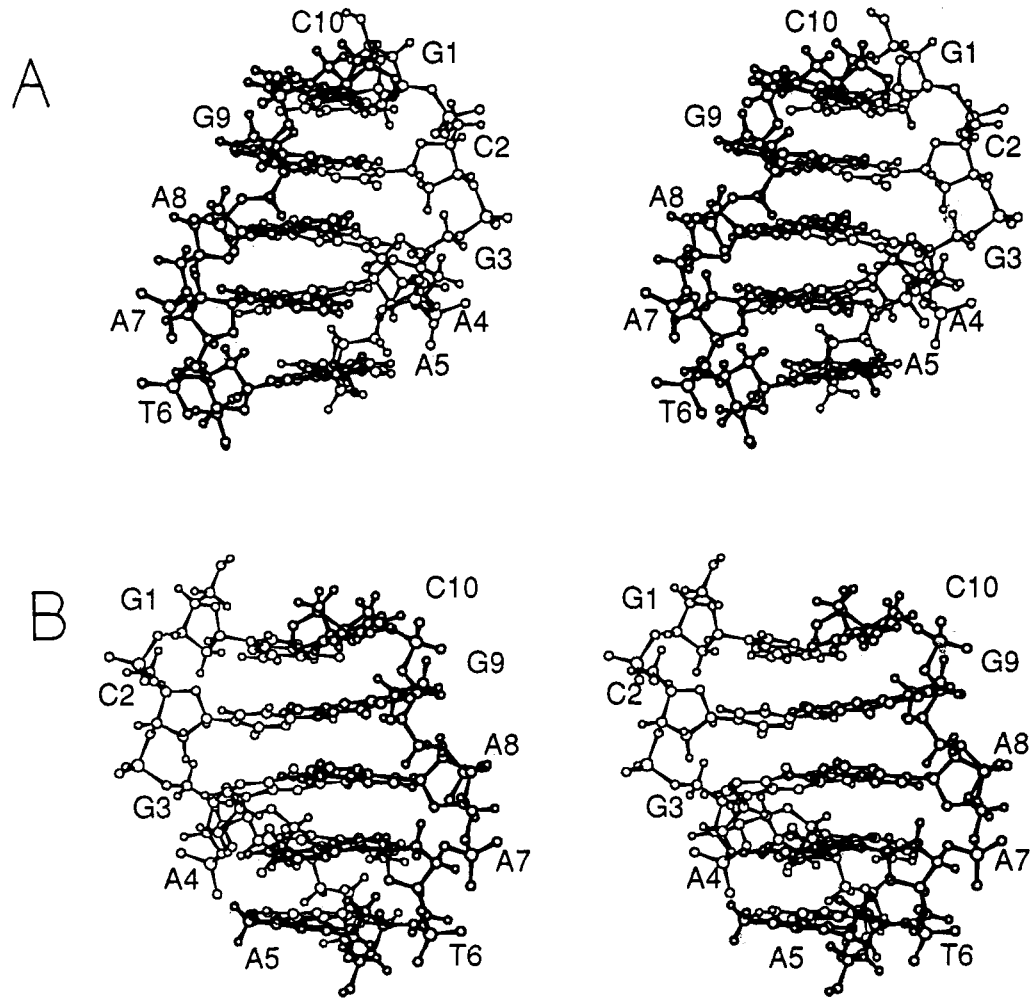

FIGURE 13: Stereoviews of the T-11-mer structure determined from NOE constraints. The d(GCGAA) strand is shown with thinner bonds, whereas the d(TAAGC) strand is shown with thicker bonds. (A) View looking into the minor groove of the duplex; (B) view looking into the major groove of the duplex.

DNA. Upon protonation, the negative and the positive CD bands shift slightly to the blue by ~3 nm. The magnitude of the positive band increases by less than 20% while the negative band does not change significantly. These changes in the CD spectra as a function of pH indicate there is not a drastic conformational change upon protonation. One simple explanation for the observed increase in stability of the duplex would be the formation of additional hydrogen bonds upon protonation. However, in the proposed base-pairing scheme (Figure 12), the protonation site is not involved in the base-pair hydrogen bonding. The increased stability of the duplex at acidic pH can be explained by the following: (1) The adenine amino hydrogens are more acidic upon protonation at the nitrogen N1 and thus the hydrogen bonds involving these hydrogens would be shorter and stronger. This is indicated by molecular orbital calculations using AM1 (Dewar et al., 1985) on protonated and unprotonated 9-methyladenine (Maskos and Sygula, unpublished results) and by the observed downfield shift for the ($NH_{2,b}$) resonances expected for shortening these hydrogen bonds (Wagner et al., 1983). (2) According to the calculations of Jordan and Sostman (1973), the magnitude of the stacking interactions is increased upon protonation of one member of a stacking pair. In the T-11-mer, stacking would improve for the protonated A4 with a neutral A5 and the protonated A8 with a neutral G9. However, these calculations do not predict the effect on stacking interactions upon protonation of both members of the stack as is the case for the $A4^+/A8^+$ stack. Solely considering the dipole contributions to this stacking interaction we note that the calculated dipole moment of the protonated adenine is nearly twice (4.5 D) and approximately in the opposite direction (arrows in Figure 12) to that of the neutral form (2.3 D) (Maskos and Sygula, unpublished results). The $A7{\cdot}A4^+$ base pair has parallel dipole moments and thus has a destabilizing dipole interaction compared to the unprotonated base pair where the dipole moments are antiparallel. For the $G3{\cdot}A8^+$ base pair, the G3 dipole moment is approximately perpendicular to the A8 dipole moment and thus will be minimally affected by the protonation. For the A4/A8 stack the dipoles are approximately antiparallel and thus favorable. Upon protonation of both bases, both dipoles will increase in magnitude and change direction but will still remain antiparallel. Thus, the $A4^+/A8^+$ stack will be more stable than the A4/A8 stack, considering only the dipole interaction. The base-pair interactions will most likely contribute less than the stacking interaction due to the $1/r^3$ dependence of the interaction and the greater distance between the dipoles. (3) The positive charge present in the base pair upon protonation would decrease the net negative charge and thus reduce the electrostatic repulsion in the duplex.

The Molecular Model of the T-11-mer. The NOE data have been used to provide distance constraints for molecular mechanics calculations resulting in a three-dimensional model of the T-11-mer shown in Figure 13. Of the 61 distance

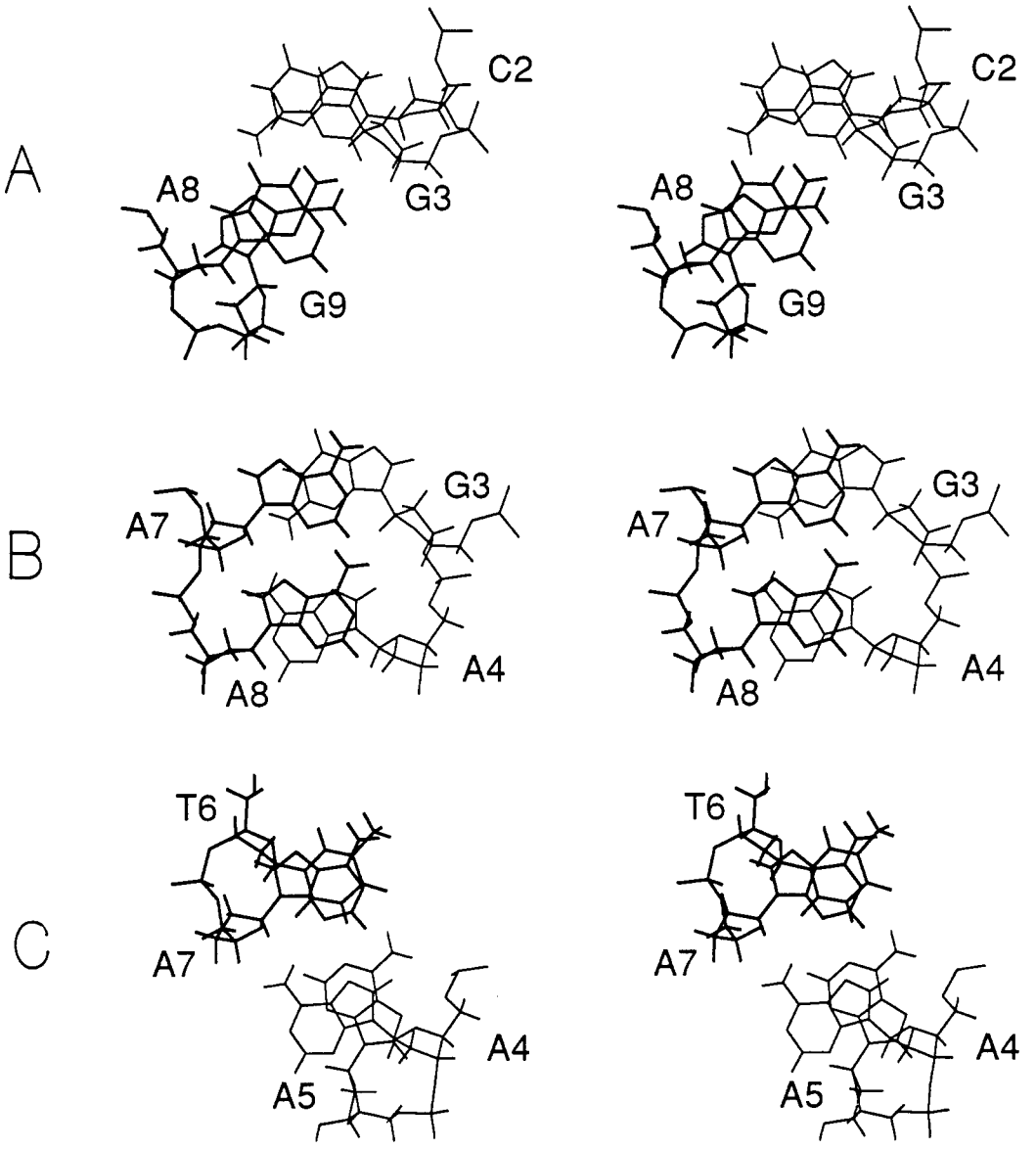

FIGURE 14: Stereoviews of the stacking of bases, looking down the helix axis at the mismatch sites from the calculated model of d(GCGAA)·d(TAAGC). (A) (5′–GC–3′)·(5′–AG–3′); (B) (5′–GA–3′)·(5′–AA–3′); (C) (5′–AA–3′)·(5′–TA–3′).

constraints determined from the NMR data 38 involve at least one proton from within the mismatches. All of the structures generated using these constraints displayed the unusual base pairing scheme of the mismatches (Figure 12) and the unusual interstrand stacking (Figures 13 and 14). We are currently obtaining data from a more concentrated NMR sample in order to refine the structure presented here by defining more distances and analyzing coupling constants to obtain torsion angle constraints.

The base-pair stacks that involve the mismatch base-pairs are shown in Figure 14. The excellent stacking interactions between the mismatched purines (Figure 14B) provide a major contribution to stabilizing the duplex. The stacking between the mismatched purines and their neighboring Watson–Crick base pairs (Figure 14A,C) is also comparable to the stacking found in a regular B-form DNA. In the mismatched base pair, the distance from A8($NH_{2,b}$) to the mismatched G3-(N3) is ~2 Å, a reasonable hydrogen bond distance. The distance between A4($NH_{2,b}$) and A7(N3) is longer (~2.7 Å), and there are two possible hydrogen bonds, between A4-(NH_2) and A7(O3′) (~2.7 Å) or A4(NH_2) and A7(O4′) (~3 Å), that could stabilize the A4·A7 base pair. There is also a possible hydrogen bond between A4(NH_2) and A8-(O5′) (~3 Å) that could stabilize the A4/A8 stack.

It has been proposed that the repair efficiency of a mismatch is correlated to the separation between C1′ atoms and the asymmetry of the λ angles within the base pair (Kennard, 1987; Kennard & Hunter, 1990), where λ is defined as the angle between the C1′–C1′ vector and the glycosyl bond. In the Watson–Crick base pairs the λ values are symmetric; for G·C, λ is 52° and 54°, and for A·T, λ is 50° and 51° (Seeman et al., 1976). The C1′–C1′ separation is 10.8 Å for G·C and 11.1 Å for A·T. These values can be compared with those obtained from the X-ray studies for the G(anti)·A(syn) (10.7 Å; λ = 58°, 40°), type I G(anti)·A(anti) (12.5 Å; λ = 53°, 52°), and G(syn)·A^+(anti) (10.8 Å; λ = 35°, 47°) [for a review, see Kennard and Hunter (1990)]. The G(anti)·A-(anti) mismatch has the largest distortion of the C1′–C1′ distance but the most symmetric λ values. In the minimized structure of the T-11-mer these parameters are measured as G3·A8 (9.5 Å; λ = 99°, 15°) and A4·A7 (8.7 Å; λ = 7°, 93°). Because of the unusual stacking, it is also interesting to compare the helical twist for the stacks surrounding the mismatches. The twist for C2·G9/G3·A8 is ~7° and for

A4·A7/A5·T6 is ~27°, indicating unwinding at these steps. However, at the G3·A8/A4·A7 step the helix is highly overwound, measured as ~87°.

Comparison with Other Mismatch Containing Oligonucleotides. It is clear that this oligomer has an unusual structure but how prevalent is this structure? Orbons et al. (1987) studied the sequence $(dCGCGAGCG)_2$ which forms adjacent G·A base pairs and observed upfield shifts for sugar protons similar to those observed in the mismatches of the T-11-mer. The authors did not propose type II G·A base pairing but this is not inconsistent with their results. Recently, Li et al. (1991) have proposed a similar *overwound* structure for the self-complementary oligomer $(dATGAGCGAATA)_2$. The proposed base-pair structure was supported by a decrease in duplex stability upon substituting guanosine with inosine in the G·A base pairs, thus establishing the involvement of the amino group.

The unusual base pairing is not limited to DNA oligomers. The type II G·A base pair has been proposed for the G·A base pair formed at the base of the loop in an RNA hairpin (Heus & Pardi, 1991). In an NMR investigation of the RNA oligomer $(GCGAGCU)_2$, no evidence for hydrogen bonding of the imino proton in the G·A mismatch was observed, and thus this duplex may also contain type II G·A base pairing (SantaLucia et al., 1990).

The DNA sequences described above suggest that the 5′–GA–3′ step is required for the formation of this unusual structure. However, X-ray and NMR studies of $(dCCAAGATTGG)_2$ have shown that both G·A base pairs are type I G(anti)·A(anti) and are accommodated by creating a bulge in the B-form duplex (Privé et al., 1987; Nikonowicz & Gorenstein, 1990; Nikonowicz et al., 1991). Thus the presence of the 5′–GA–3′ step is not sufficient for stabilizing the unusual base pairing and stacking structure. Unpublished data from our laboratory indicate that a type II G·A base pair also exists in three other oligonucleotides of similar sequence to the T-11-mer; dGCGAAXAAGCG, where X = C or G, or no nucleotide. Thus, it seems that this type of base pairing and *overwound* stack is not unique and can be readily formed by a variety of DNA or RNA sequence types.

Generalizing from the minimal data base available, we propose that the *overwound* double-stranded structure of nucleic acid will form if the sequence $(5'–PyPuPuPu–3')_2$ is present but such a structure will not form in the case of the $(5'–PuPuPuPy–3')_2$ sequence. Clearly, more mismatch sequences need to be investigated to determine what sequence is required to stabilize this unusual structure. Investigations of several oligonucleotides designed to test the sequence context of the tandem mismatch structure are currently underway (W. D. Wilson, personal communication).

CONCLUSIONS

One of the most interesting features of our model is the stacking that occurs between G3·A8 and A4·A7. Instead of stacking with the base which is sequential on the strand there is stacking with the base from the opposite strand. In the phylogenetically determined secondary structures of 21 16S rRNAs (Gutell et al., 1985) and 38 23S rRNAs (Gutell & Fox, 1988), the internal loop 5′–GA–3′/3′–AA–5′ is observed 39 times. It can be speculated that such geometrical arrangements of the mismatched G·A and A·A base pairs, as shown in Figure 12, could also be present in the internal loops of rRNAs. It has been suggested already (SantaLucia et al., 1990) that current models for internal loops in RNA are oversimplified because potential hydrogen bonding interactions within internal loops are not included.

In the proposed arrangements (Figures 1 and 12) of the mismatched G·A base pair, different functional groups are exposed, i.e., not engaged in hydrogen bonding between bases. There is evidence that in vivo recognition and repair of G·A mismatches may be dependent on the surrounding sequence (Fazakerly et al., 1986). This may well be related to the very different functional groups which are free to interact with polymerase or a repair enzyme in the different geometrical arrangements of the G·A pair. Given this variability, it is perhaps not surprising that the G·A mismatch so effectively escapes repair.

NOTE ADDED IN PROOF

Recently there have been several more investigations of the Wilson 11-mer which contains tandem mismatches (Chou et al., 1992a,b; Cheng et al., 1992). These studies further characterize the structure of the tandem G·A mismatches and confirm the sequence dependence of the type II G(anti)·A(anti) mismatch.

REFERENCES

Arnold, F. H., Wolk, S., Cruz, P., & Tinoco, I., Jr. (1987) *Biochemistry 26*, 4068–4075.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Baleja, J. D., Germann, M. W., van de Sande, J. H., & Sykes, B. D. (1990) *J. Mol. Biol. 215*, 411–428.

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

Boelens, R., Scheek, R. M., Dijistra, K., & Kaptein, R. (1985) *J. Magn. Reson. 62*, 378–386.

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

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

Carbonnaux, C., van der Marel, G. A., van Boom, J. H., Guschlbauer, W., & Fazakerley, G. V. (1991) *Biochemistry 30*, 5449–5458.

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

Chazin, W. J., Wutrich, K., Hyberts, S., Rance, M., Denny, W. A., & Leupin, W. (1986) *J. Mol. Biol. 190*, 439–453.

Cheng, J.-W., Chou, S.-H., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 1037–1041.

Chou, S.-H., Cheng, J.-W., Fedoroff, O. Y., Chuprina, V. P., & Reid, B. R. (1992a) *J. Am. Chem. Soc. 114*, 3114–3115.

Chou, S.-H., Cheng, J.-W., & Reid, B. R. (1992b) *J. Mol. Biol. 228*, 138–155.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., & Stewart, J. J. P. (1985) *J. Am. Chem. Soc. 107*, 3902–3909.

Eich, G., Bodenhausen, G., Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 3731–3732.

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

Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., & Kearns, D. R. (1982) *J. Am. Chem. Soc. 104*, 5540–5541.

Fersht, A. R., Knill-Jones, J. W., & Tsui, W. C. (1982) *J. Mol. Biol. 156*, 37–51.

Forster, A. C., & Symons, R. H. (1987a) *Cell 49*, 211–220.

Forster, A. C., & Symons, R. H. (1987b) *Cell 50*, 9–16.

Gao, X., & Patel, D. J. (1987) *J. Biol. Chem. 262*, 16973–16984.

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

Gutell, R. R., & Fox, G. E. (1988) *Nucleic Acids Res. 16* (Suppl.), r175–r269.

Gutell, R. R., Weiser, B., Woese, C. R., & Noller, H. F. (1985) *Prog. Nucleic Acids Res. Mol. Biol. 32*, 155–216.
Hare, D. R., Wemmer, D. E., Chou, S., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.
Heus, H. A., & Pardi, A. (1991) *Science 253*, 191–194.
Hore, P. J. (1983a) *J. Magn. Reson. 54*, 539–542.
Hore, P. J. (1983b) *J. Magn. Reson. 55*, 283–300.
Hunter, W. N., Brown, T., & Kennard, O. (1986) *J. Biomol. Struct. Dyn. 4*, 173–191.
Jack, A., Ladner, J. E., & Klug, A. (1976) *J. Mol. Biol. 108*, 619–649.
Jordan, F., & Sostman, H. D. (1973) *J. Am. Chem. Soc. 95*, 6544–6554.
Kan, L.-S., Chandrasegaran, S., Pulford, S. M., Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.
Keepers, J. W., Schmidt, P., James, T. L., & Kollman, P. A. (1984) *Biopolymers 23*, 2901–2929.
Kennard, O. (1987) in *Nucleic Acids and Molecular Biology* (Lilley, D., & Eckstein, E. Eds.) Vol. 1, pp 25–52, Springer-Verlag, Berlin.
Kennard, O., & Hunter, W. N. (1990) *Q. Rev. Biophys. 23*, 327–379.
LeBlanc, D. A., & Morden, K. M. (1991) *Biochemistry 30*, 4042–4047.
Leonard, G. A., Booth, E. D., & Brown, T. (1990) *Nucleic Acids Res. 18*, 5617–5623.
Lerner, D. B., & Kearns, D. R. (1981) *Biopolymers 20*, 803–816.
Li, Y., Zon, G., & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.
Lyamichev, V. I., Mirkin, S. M., Danilevskaya, O. N., Voloshin, O. N., Balatskaya, S. V., Dobrynin, V. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1990) in *Structure & Methods, DNA & RNA* (Sarma, R. H., & Sarma, M. D., Eds.) Vol. 3, pp 175–184, Adenine Press, NY.
Maskos, K., LeBlanc, D. A., & Morden, K. M. (1991) *Biophys. J. 59*, 491a.
McConnell, B. (1984) *J. Biomol. Struct. Dyn. 1*, 1407–1421.
McConnell, B., & Seawell, P. C. (1972) *Biochemistry 11*, 4382–4392.
McConnell, B., & Seawell, P. C. (1973) *Biochemistry 12*, 4426–4434.
Morden, K. M., Gunn, B. M., & Maskos, K. (1990) *Biochemistry 29*, 8835–8845.
Nagayama, K., Kumar, A., Wüthrich, K., & Ernst, R. R. (1980) *J. Magn. Reson. 40*, 321–334.
Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.
Nikonowicz, E. P., Meadows, R. P., Fagan, P., & Gorenstein, D. G. (1991) *Biochemistry 30*, 1323–1334.
Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1987) *Eur. J. Biochem. 170*, 225–239.
Panyutin, I. G., Kovalsky, O. I., Budovsky, E. I., Dickerson, R. E., Rikhirev, M. E., & Lipanov, A. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 867–890.
Patel, D. J. (1976) *Biopolymers 15*, 533–558.
Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Biochemistry 23*, 3207–3217.
Patel, D. J., Shapiro, L., & Hare, D. (1987) in *Nucleic Acids and Molecular Biology* (Lilley, D., & Eckstein, E., Eds.) Vol. 1, pp 70–84, Springer-Verlag, Berlin.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.
Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.
Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L., Kopka, M. L., & Dickerson, R. E. (1988) in *Structure and Expression* (Sarma, R. H., & Sarma, M. H., Eds.) Vol. 2, pp 27–47, Adenine Press, Schenectady, NY.
Raszka, M. (1974) *Biochemistry 13*, 4616–4622.
Raszka, M., & Kaplan, N. O. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2025–2029.
Rich, A. (1977) *Acc. Chem. Res. 10*, 338–402.
Rich, A., Davies, D. R., Crick, F. H. C., & Watson, J. D. (1961) *J. Mol. Biol. 3*, 71–86.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813–8819.
Scheek, R. M., Russo, N., Boelens, R., & Kaptein, R. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.
Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804–808.
Sklenar, V., Brooks, B. R., Zon, G., & Bax, A. (1987) *FEBS Lett. 216*, 249–252.
Sowers, L. C., Fazakerley, G. V., Kim, H., Dalton, L., & Goodman, M. F. (1986) *Biochemistry 25*, 3983–3988.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Voet, D., & Rich, A. (1970) *Prog. Nucleic Acid Res. Mol. Biol. 10*, 183–265.
Wagner, G. (1983) *J. Magn. Reson. 55*, 151–156.
Wagner, G., Pardi, A., & Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 5948–5949.
Webster, G. D., Sanderson, M. R., Skelly, J. V., Swann, P. F., Li, B. F., & Tickle, I. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6693–6697.
Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 130–134.
Wemmer, D. E., Chou, S.-H., & Reid, B. R. (1984) *J. Mol. Biol. 180*, 41–60.