NMR Studies of a Deoxyribodecanucleotide Containing an Extrahelical Thymidine Surrounded by an Oligo(dA)·Oligo(dT) Tract†

Kathleen M. Morden,* Bonnie M. Gunn, and Karol Maskos
Department of Biochemistry, Louisiana State University, Baton Rouge, Louisiana 70803
Received February 21, 1990; Revised Manuscript Received June 14, 1990

ABSTRACT: One- and two-dimensional NMR experiments were carried out on a decamer, d(CGCTTTTCGC)·d(GCGAAAAGCG), and on the same sequence with the addition of an unpaired thymidine, d(CGCTTTTCGC)·d(GCGAATAAGCG), which will be referred to as the T-bulge decamer. Evidence from one-dimensional NOE experiments on the exchangeable protons indicates that the unpaired thymidine is extrahelical. This conclusion is also supported by numerous cross-peaks in the two-dimensional NOESY spectrum of the nonexchangeable protons. Assignments for all of the resonances, with the exception of the H5′ and H5″ resonances, have been made for both oligonucleotide duplexes through the use of 2D NOESY, COSY, and relayed COSY experiments. Temperature dependence of the methyl resonance chemical shifts indicates that the unpaired thymidine shows unusual behavior compared to other thymidines in the duplex. Two-dimensional NOESY experiments carried out from 5 to 35 °C indicate the unpaired thymidine remains extrahelical throughout this temperature range. A similar temperature dependence for the methyl chemical shift is found in the corresponding single-strand d(GCGAATAAGCG). The oligo(dA)·oligo(dT) tracts in both the decamer and the T-bulge decamer have structures different from B-form DNA and exhibit NOEs similar to those observed in other oligonucleotides containing A·T tracts. The formation of this unusual A·T tract structure may induce the extrahelical conformation of the unpaired thymidine.

In the past 10 years there have been numerous investigations on the structural effects of unpaired nucleotides. These studies have been motivated primarily by an interest in understanding the molecular basis for frameshift mutations. Initial studies used polymeric DNA and gave general results as to the effects of an unpaired base on double-strand stability (Fresco & Alberts, 1960; Lomant & Fresco, 1974). More recently there have been studies on bulge-containing oligonucleotides using techniques such as X-ray crystallography and NMR spectroscopy. The oligonucleotides studied can be put into two categories: oligonucleotides containing an unpaired purine and those with an unpaired pyrimidine.

Solution studies on duplexes containing an unpaired purine showed that the unpaired purine, regardless of whether it was an adenine or a guanine, stacks with the adjacent base pairs, resulting in an intrahelical conformation (Patel et al., 1982; Hare et al., 1986; Woodson & Crothers, 1987, 1989; Roy et al., 1987b; Kalnik et al., 1989a; Nikonowicz et al., 1989). The duplexes investigated contain a wide variety of base pairs adjacent to the unpaired purine and yet they all adopt an intrahelical conformation; thus the neighboring base pairs do not seem to influence the conformation of the unpaired purine. Recently several of the molecules containing an unpaired adenosine, whose conformations were determined in solution, have been studied by using X-ray crystallography (Joshua-Tor et al., 1988; Miller et al., 1988). In these studies, the unpaired base was found to be excluded from the helix, allowing the adjacent base pairs to stack and resulting in an extrahelical conformation of the unpaired base. It was hypothesized that forces present in the condensed phase of the crystallographic studies were responsible for forcing the molecule into a structure which was contrary to that found in solution.

The studies on unpaired pyrimidines have been done exclusively in solution and have shown that both the extrahelical and intrahelical conformations can exist (Haasnoot et al., 1980; Morden et al., 1983; van den Hoogen et al., 1988; Kalnik et al., 1989b, 1990). The unpaired thymidine in the work presented here is in the center of an oligo(dA)·oligo(dT) region. The intent of this sequence was to model the theory of Streisinger et al. (1966), in which it is predicted that frameshift mutations will occur in regions of high base-pair repeat. Oligo(dA)·oligo(dT) tracts have also been shown to have a structure different from a standard B-form DNA. This unusual structure has been studied by X-ray crystallography (Nelson et al., 1987; Coll et al., 1987; Yoon et al., 1988) and NMR spectroscopy (Kintanar et al., 1987; Roy et al., 1987a; Katahira et al., 1988; Sarma et al., 1988; Gupta et al., 1988; Celda et al., 1989; Nadeau & Crothers, 1989; Katahira et al., 1990). The structure is characterized by a high propeller twist (≥15°), which results in NOEs[1] in the NMR spectra that are not usually observed for a random sequence oligonucleotide. It has been proposed that the structure is stabilized by a network of interstrand bifurcated hydrogen bonds, which may be unique to the A·T tract (Nelson et al., 1987; Yoon et al., 1988). Kinetic data, obtained from imino proton exchange (Leroy et al., 1988), indicates that lifetimes of A·T base pairs inside the A·T tract have unusual properties not observed for imino protons in G·C-rich, alternating, or random sequences. These investigations also indicate that this unusual structure will only be stable in repeats of four or more A·T base pairs

† This work was supported by National Institutes of Health Grant GM 38137 and Louisiana Education Quality Support Fund [LEQSF(86-89)-RD-A-12]. The NMR instrument was purchased with Grant RR02459 from the Division of Research Resources of the NIH.

[1] Abbreviations: NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; 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; ppm, parts per million.

0006-2960/90/0429-8835$02.50/0

(Leroy et al., 1988; Nadeau & Crothers, 1989).

In this paper we present NMR data for a decamer containing an unpaired thymidine in an A·T tract, d(CGCTTTTCGC)·d(GCGAATAAGCG), and compare it to the decamer d(CGCTTTTCGC)·d(GCGAAAAGCG). Evidence is given for the unpaired thymidine in the T-bulge decamer being bulged outside of the helix and for an altered structure in the surrounding A·T tract.

Materials and Methods

Oligonucleotide Samples. The oligonucleotide dCGCTTTTCGC was synthesized on a Model 380A Applied Biosystems DNA synthesizer using a solid support and methyl phosphoramidite chemistry. Purification by reverse-phase HPLC was performed on a Zorbax C18 column (Dupont) with a 0–30% acetonitrile gradient in 0.1 M ammonium acetate buffer (pH 7) over 30 min. Purity was confirmed by using strong anion-exchange HPLC on a Zorbax column (Dupont) with an ammonium acetate buffer (pH 7) in 20% ethanol with a 0.01–2.5 M ammonium acetate gradient in 60 min. The oligonucleotide was lyophilized and then dissolved in 2 M NaCl to convert from ammonium to sodium salt. The sample was then desalted over a Sep-Pak C18 cartridge (Waters Associates) and passed several times over a P-2 column (Bio-Rad), with water used to elute. dGCGAAAAGCG and dGCGAATAAGCG were purchased from Midland Certified Reagent Co. (Midland, TX) in purified form, and purity was checked for these oligonucleotides with both reverse-phase and strong anion-exchange HPLC as described above. The purchased oligonucleotides were used without further purification.

NMR Samples. The concentrations of the oligonucleotides were determined optically at 25 °C by using the following extinction coefficients per mole of strand at 260 nm: dCGCTTTTCGC, 80.4×10^3 M^{-1} cm^{-1}; dGCGAAAAGCG, 104.2×10^3 M^{-1} cm^{-1}; and dGCGAATAAGCG, 114.3×10^3 M^{-1} cm^{-1}. The final concentration for all the NMR samples was 2 mM in each single strand in 10 mM phosphate buffer containing 0.1 mM EDTA and 0.1 M NaCl. For observation of the nonexchangeable protons, samples were lyophilized three times from 99.8% D_2O (Wilmad), the pD was adjusted to 6.6 (uncorrected pH meter reading), and the sample was lyophilized and dissolved in 99.96% D_2O (Wilmad). The samples for observing the exchangeable protons were prepared in the buffer described above and dissolved in 90% H_2O/10% D_2O.

NMR Experiments. All of the NMR data were obtained on a 400-MHz Bruker AM spectrometer equipped with an Aspect 3000 computer. The chemical shifts are relative to internal sodium 3-(trimethylsilyl)[2,2,3,3-$^2H_4$]propionate (TSP).

The one-dimensional temperature profiles for the nonexchangeable protons were collected with 8192 complex data points over a 4600-Hz sweep width with 128 accumulations. The residual HDO in the sample was suppressed by low-power presaturation during all periods except acquisition. A line broadening of 0.2 Hz was applied before the spectra were transformed. The exchangeable proton spectra were obtained by using a 1331 pulse sequence for water suppression (Hore, 1983a,b) with the carrier frequency set on the H_2O resonance. The data were collected with 8192 complex data points over a sweep width of 12 000 Hz, and a line broadening of 2 Hz was applied prior to transformation of the data. Imino spectra, as a function of temperature, were the sum of 256 accumulations. The imino NOE difference spectra were collected with a 500-ms irradiation pulse in groups of 32 accumulations cycled between irradiation on resonance and irradiation off resonance for a total of 4000 accumulations each. The difference spectra were generated by subtraction after transformation.

T-bulge Decamer

```
   1  2  3  4  5      6  7  8  9 10
d  G - C - G - A - A - Tb- A - A - G - C - G
   C - G - C - T - T - -  - T - T - C - G - C d
  20 19 18 17 16      15 14 13 12 11
```

Decamer

```
   1  2  3  4  5  6  7  8  9 10
d  G - C - G - A - A - A - A - G - C - G
   C - G - C - T - T - T - T - C - G - C d
  20 19 18 17 16 15 14 13 12 11
```

FIGURE 1: Numbering scheme for the bases in the T-bulge decamer and the decamer. The unpaired thymidine will be referred to as T_b.

The phase-sensitive two-dimensional NOE experiments (NOESY) of the duplexes in D_2O were collected by using the TPPI method (Bodenhausen et al., 1984) with a 4500-Hz sweep width, a 200-ms mixing time, and a 1.8-s recycle delay. The residual HDO signal was suppressed by presaturation. The NOESY spectra were collected with 1024–2048 complex data points in t_2 and 300–350 points in t_1, each being the sum of 128 accumulations. The two-dimensional NOESY data sets were processed by using FTNMR software (Hare Research, Inc.). The data were apodized with a shifted sine-bell curve, shifted by 45° for both the t_1 and t_2 domains, zero-filled, and transformed. The final data sets were 2K × 2K for the decamer and 1K × 1K for the T-bulge decamer.

The two-dimensional correlated spectroscopy experiments (COSY) were obtained in the magnitude mode (Aue et al., 1976; Nagayama et al., 1980) with 1024–2048 complex points in t_2 and 300–350 points in t_1, each point being the sum of 48 accumulations. Zero filling was to 1024–2048 points for only the t_1 dimension. Both dimensions were apodized with an unshifted sine-bell curve before transformation. The magnitude relayed COSY spectra (Eich et al., 1982; Wagner, 1983) were obtained with a relayed transfer delay of 60 ms, with 1024 complex points in t_2 and 300–350 t_1 points consisting of 128 accumulations each. Processing parameters were identical with those for the magnitude COSY experiment.

Results

The decamer is a duplex of complementary strands, each consisting of 10 nucleotides. The T-bulge decamer consists of a 10-nucleotide strand combined with an 11-nucleotide strand to form 10 base pairs and an unpaired thymidine. The exact sequence and the numbering schemes for the decamer and the T-bulge decamer are shown in Figure 1. For the T-bulge decamer, the thymidine adjacent to A5 and A6 will be referred to as T_b.

NMR of Imino Resonances. The results of the one-dimensional nuclear Overhauser effect experiment for both the decamer and the T-bulge decamer are shown in Figure 2. The expected NOE interactions for the imino protons in a standard B-form helix are shown schematically in Figure 3. When a T(H3) proton is irradiated, there will be a strong NOE to A(H2) from the same base pair and NOEs will be observed to the imino protons from base pairs on the 5′ and 3′ sides of the irradiated proton. Due to the oligo(A)·oligo(T) tract in our molecules, the excited T(H3) will also give an NOE to A(H2) from the A·T base pair on the 5′ side. Figure 2 shows the NOE difference spectra resulting from excitation of imino protons in the central A·T tract. In figure 2, d and h are the spectra obtained with off-resonance irradiation for the T-bulge

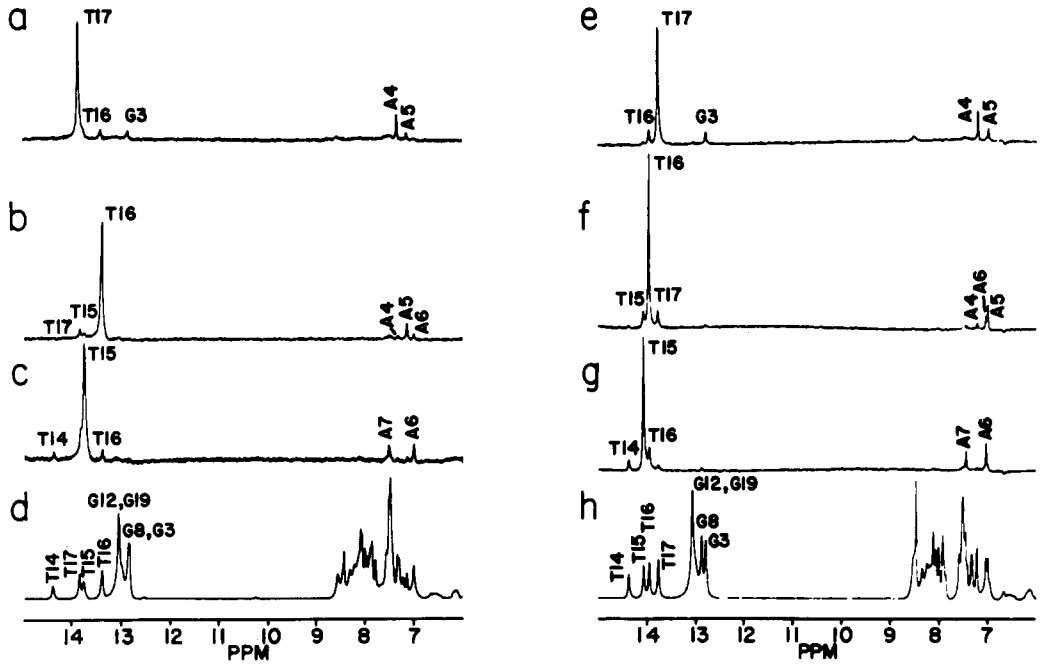

FIGURE 2: ^{1}H NMR NOE difference spectra for the T-bulge decamer (spectra a–c) and the decamer (spectra e–g). Spectra d and h are the control spectra, without irradiation of any of the resonances, for the T-bulge decamer and decamer, respectively. The spectra are the difference between 4000 accumulations without and with irradiation of a resonance. All spectra were taken at 15 °C with a 2 mM duplex sample.

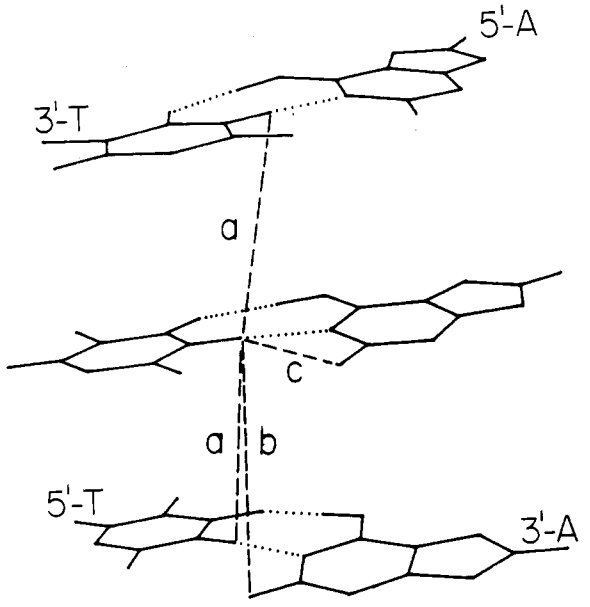

FIGURE 3: Schematic of a trinucleotide stack in a B-form DNA helix. The stack consists of $d(A_3)\cdot d(T_3)$. The 5′ and 3′ ends on each strand are as labeled. The labeled dashed lines connect the following protons: (a) T(H3) protons on adjacent base pairs, (b) T(H3) and A(H2) on adjacent base pairs, and (c) T(H3) and A(H2) on the same base pair.

decamer and the decamer, respectively. Spectra a–c and e–g are the NOE difference spectra between off-resonance and on-resonance excitation for the T-bulge decamer and the decamer, respectively. In general, these spectra show the same pattern as has been observed previously for A·T tracts (Kintanar et al., 1987; Sarma et al., 1988; Gupta et al., 1988; Katahira et al., 1988), and only unusual observations will be presented here. Upon excitation of T16(H3), as shown in Figure 2f, the expected NOE is observed to the adjacent imino protons, T15(H3) and T17(H3), and to A5(H2) and A6(H2). However, an NOE is also observed to A4(H2), which is on the A·T base pair on the 3′ side. Upon excitation of T15(H3), a similar interaction can be observed to the A5(H2) proton, shown in spectrum f as a shoulder on the A6(H2) resonance. These NOEs in the decamer reflect an unusual structure due to the oligo(A)·oligo(T) sequence in this molecule.

The NOEs observed upon excitation of the T(H3) protons in the T-bulge decamer (spectra a-c) are qualitatively the same as those observed for the decamer (spectra e–g). This is a particularly significant result for the excitation of T15(H3) and T16(H3), as these protons are on base pairs flanking the unpaired thymidine. NOEs between protons on A5·T16 and A6·T15 (Figure 2, spectra b and c) provide evidence that these two base pairs are less than 5 Å apart and that the unpaired thymidine is extrahelical. This is a qualitative analysis of distances between these protons, as the NOEs were obtained with a 500-ms mixing time and consequently include the effects of spin diffusion.

Temperature Dependence of Imino Resonances. Spectra of the imino resonances as a function of temperature are shown in Figure 4 for both the decamer and the T-bulge decamer. The difference in the imino resonance line widths as a function of temperature is due to the difference in melting temperatures and melting behavior for the two duplexes. Comparable line widths are observed for the T-bulge decamer at 15 °C and the decamer at 35 °C or for the T-bulge decamer at 25 °C and the decamer at 40 °C. Thus the presence of the unpaired thymidine has destabilized the duplex by approximately 15–20 °C. The order in which specific imino resonances broaden can also be compared for the two duplexes. For the decamer, the resonances from the terminal imino protons, G1 and G10, are observed at 5 °C and broaden at higher temperatures. The superimposed G12 and G19 resonances broaden next, followed by the central A·T tract resonances, T14–T17. The last imino resonances to broaden are those from G3 and G8, which are the only remaining resonances at 55 °C (data not shown).

The order of broadening for the T-bulge decamer is somewhat different. The imino resonances due to G1 and G10 are again the first to broaden. The next resonances to broaden, however, are the imino resonances for G12 and G19, which broaden simultaneously with the imino resonances for T14–T17. At a slightly higher temperature, the imino resonances

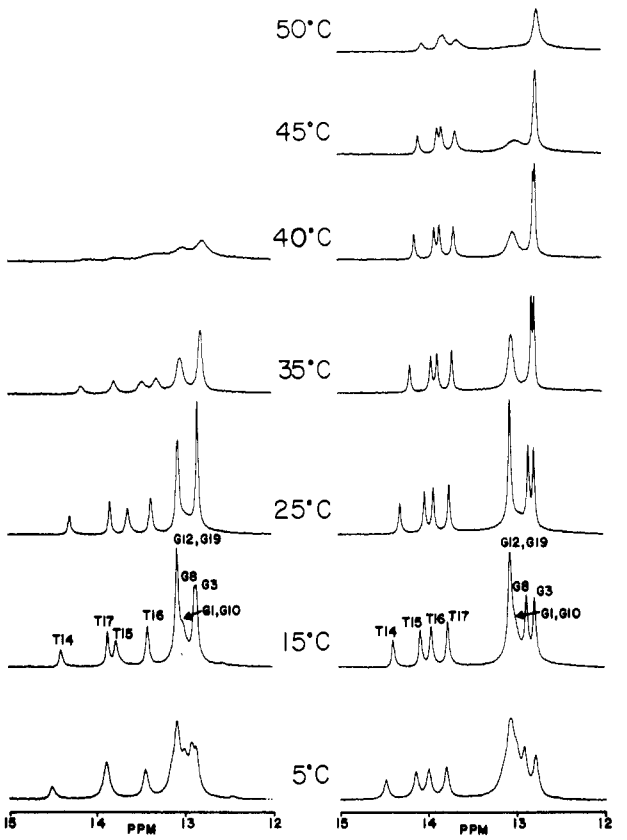

FIGURE 4: Temperature dependence of the imino proton NMR spectrum of the T-bulge decamer, left, and the decamer, right. Resonance assignments are shown for the spectrum at 15 °C.

for G3 and G8 broaden. Thus the presence of the unpaired thymidine has destabilized the central base pairs relative to the terminal base pairs.

Nonexchangeable Protons. The assignments of the nonexchangeable protons have been made by using well-established techniques (Feigon et al., 1982; Hare et al., 1983; Scheek et al., 1983; Weiss et al., 1984). Figure 5 shows an expansion of the 2D NOESY spectrum of the decamer obtained in D_2O at 25 °C. The cross-peaks in this region result from interactions between C(H6) and C(H5) protons, aromatic and H1′ protons, and aromatic and T(CH_3) protons. The cross-peaks between C(H6) and C(H5) protons are easily identified by their presence in the COSY spectrum (data not shown) and are distinguished in Figure 5 by boxed labels. The cross-peak pattern of an intranucleotide interaction between H6 or H8 and H1′ followed by a sequential interaction between H1′ and H6 or H8 from the adjacent nucleotide is observed for the decamer and is indicated by a dashed line for the dGCGAAAAGCG strand and a solid line for the dCGCTTTTCGC strand. Cross-peaks labeled with the base number are due to intranucleotide interactions. Unlabeled cross-peaks connected to labeled ones by solid or dashed lines are sequential cross-peaks. Other cross-peaks labeled with lowercase letters are identified in the figure caption.

Figure 6 shows the aromatic–H1′ region and aromatic–methyl region for the T-bulge decamer obtained in D_2O at 25 °C. The aromatic–H1′ network is indicated by a dashed line for the dGCGAATAAGCG strand and a solid line for the dCGCTTTTCGC strand. The same conventions for labeling the NOESY spectrum in Figure 5 are used for the spectrum of the T-bulge decamer in Figure 6. Also shown in Figures 5 and 6 are cross-peaks due to interactions with A(H2) protons. These are cross-peaks e and g–s for the decamer (Figure

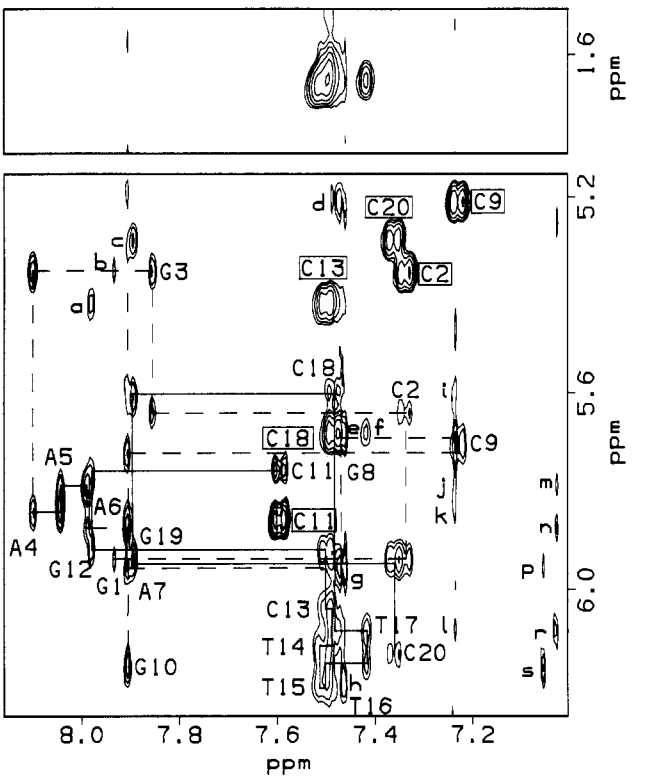

FIGURE 5: Expansion of the two-dimensional NOESY spectrum of the decamer taken at 25 °C. The region shows the H6, H8–H1′, H5 region and the H6, H8–methyl region. The dashed line follows the sequential NOE connectivities for the dGCGAAAAGCG strand in the duplex, and the solid line follows the sequential NOE connectivities for the dCGCTTTTCGC strand in the duplex. 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 can be assigned as follows: (a) G12(H8)–C13(H5), (b) G1(H8)–C2(H5), (c) G19(H8)–C20(H5), (d) G8(H8)–C9(H5), (e) A7(H2)–G8(H1′), (f) T17(H6)–C18(H5), (g) A7(H2)–A7(H1′), (h) A7(H2)–T15(H1′), (i) A4(H2)–C18(H1′), (j) A4(H2)–A5(H1′), (k) A4(H2)–A4(H1′), (l) A4(H2)–T17(H1′), (m) A5(H2)–A5(H1′), (n) A5(H2)–A6(H1′), (p) A6(H2)–A7(H1′), (r) A5(H2)–T17(H1′), and (s) A6(H2)–T16(H1′).

5) and cross-peaks f, g, and i–r for the T-bulge decamer (Figure 6). Assignments for these cross-peaks are given in the figure captions.

The H2′ and H2″ resonances have also been assigned for these two duplexes. The COSY spectrum and NOESY spectra obtained with short mixing times (50–100 ms) were used to distinguish H2′ and H2″ cross-peaks. For all of the nucleotides, the H2′ proton resonates upfield from the H2″, with the exception of the terminal nucleotide G10, where this trend is reversed, and nucleotides C20, G3, and G12, where the H2′ and H2″ resonances have identical chemical shifts. Chemical shifts for the H2′ and H2″ resonances were determined from the aromatic (H6/H8)–H2′, H2″ region of the NOESY spectrum, shown in Figure 7. 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). This network is observed for both the dGCGAATAAGCG strand (Figure 7a) and for the dCGCTTTTCGC strand (Figure 7b). The assignments of the H2′ and H2″ resonances were also independently determined from the H1′–H2′, H2″ region of the NOESY spectrum.

Assignments of the H3′ and H4′ resonances were made from the H1′–H3′ and H1′–H4′ regions of the NOESY spectrum. These assignments were independently confirmed by using a relayed COSY experiment (Eich et al., 1982; Wagner, 1983) to detect H1′–H3′ cross-peaks. Assignments were also con-

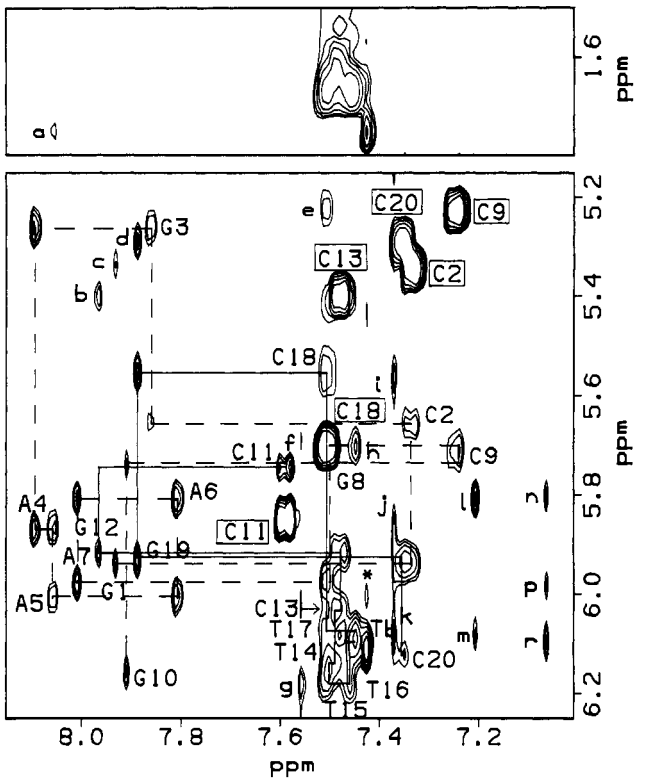

FIGURE 6: Expansion of the two-dimensional NOESY spectrum of the T-bulge decamer taken at 25 °C. The region shows the H6, H8–H1′, H5 region and the H6, H8–methyl region. The dashed line follows the sequential NOE connectivities for the dGCGAATAAGCG strand in the duplex, and the solid line follows the sequential NOE connectivities for the dCGCTTTTCGC strand in the duplex. 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 asterisk labels the cross-peak between A5(H1′)–T_b(H6). The lettered cross-peaks are assigned as follows: (a) A5(H8)–T_b(CH_3), (b) G12(H8)–C13(H5), (c) G1(H8)–C2(H5), (d) G19(H8)–C20-(H5), (e) G8(H8)–C9(H5), (f) A7(H2)–G8(H1′), (g) A7(H2)–T15(H1′), (h) T17(H6)–C18(H5), (i) A4(H2)–C18(H1′), (j) A4-(H2)–A4(H1′), (k) A4(H2)–T17(H1′), (l) A5(H2)–A6(H1′), (m) A5(H2)–T17(H1′), (n) A6(H2)–A6(H1′), (p) A6(H2)–A7(H1′), and (r) A6(H2)–T16(H1′).

firmed by the H3′–H4′ cross-peaks in both the COSY and NOESY spectra and by H2′, H2″–H3′ and H4′ cross-peaks in the NOESY spectrum.

The chemical shifts at 25 °C for all of the base proton resonances for the decamer and the T-bulge decamer are given in Table I. The chemical shifts at 25 °C for all the sugar proton resonances, excluding H5′ and H5″, are given for both duplexes in Table II. Also calculated in these tables is the difference in the chemical shifts for analogous protons in the two duplexes.

Temperature Dependence of Selected Nonexchangeable Proton Resonances. The chemical shift versus temperature plots for the T(Me) protons from the T-bulge decamer and the decamer are shown in Figure 8. The curves for the decamer, shown in Figure 8b, indicate an approximate melting temperature of 65 °C for the methyl protons, with a possible premelting transition at approximately 20–25 °C. The curves for the T-bulge decamer, shown in Figure 8a, indicate an approximate melting temperature of 50 °C for the methyl protons on the dCGCTTTTCGC strand, with a possible premelting transition at approximately 15–25 °C. The methyl resonance from the dGCGAATAAGCG strand shows non-sigmoidal temperature behavior and thus a melting temperature is not estimated from this curve. The temperature behavior of the chemical shifts from the T_b(Me) protons in the T-bulge decamer is compared to the temperature behavior in the dGCGAATAAGCG single strand and shown in Figure 9. In the single strand, two resonances are observed for the T_b methyl protons, represented by × and ◆ in Figure 9.

Figure 10 shows an expansion of the region around the A5–A6 sequential cross-peak for the T-bulge decamer at temperatures from 5 to 37 °C. The position for the sequential cross-peak between T_b(H1′) and A6(H8), even if not observed, is indicated by an arrow. The solid line designates the part of the aromatic–H1′ network that is observable in this expanded region.

Discussion

Evidence for an Extrahelical Conformation. The evidence presented under Results indicates that the unpaired thymidine

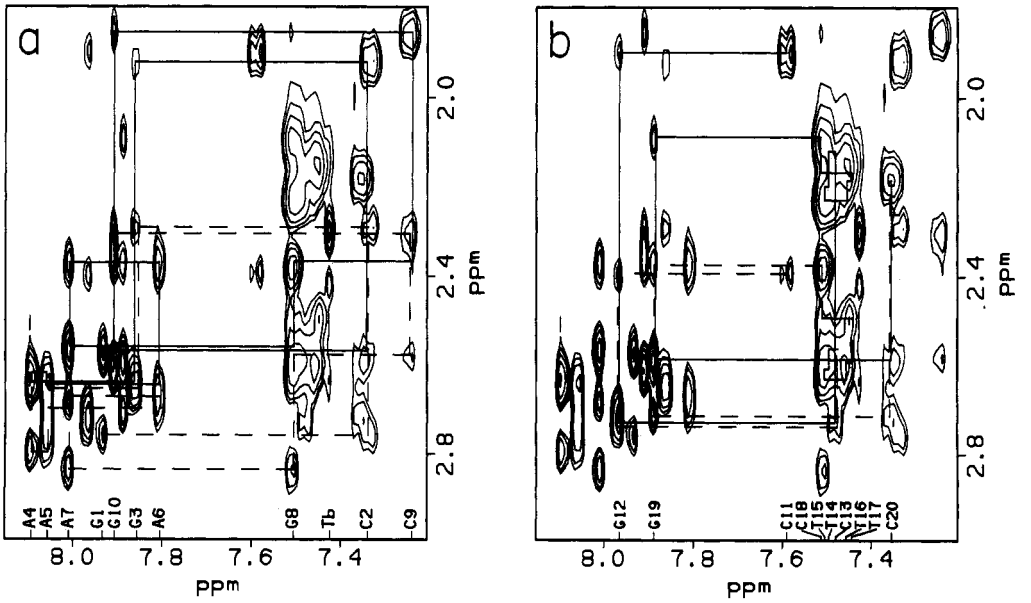

FIGURE 7: Expansion of the two-dimensional NOESY spectrum of the T-bulge decamer taken at 25 °C. The region shows the H6, H8–H2′, H2″ region. In spectrum a, the solid line follows the aromatic–H2′ NOE network and the dashed line follows the aromatic–H2″ NOE network for the dGCGAATAAGCG strand in the duplex. In spectrum b, the same convention is used for the dCGCTTTTCGC strand. The chemical shifts of the H8 and H6 protons are indicated along the *x*-axis.

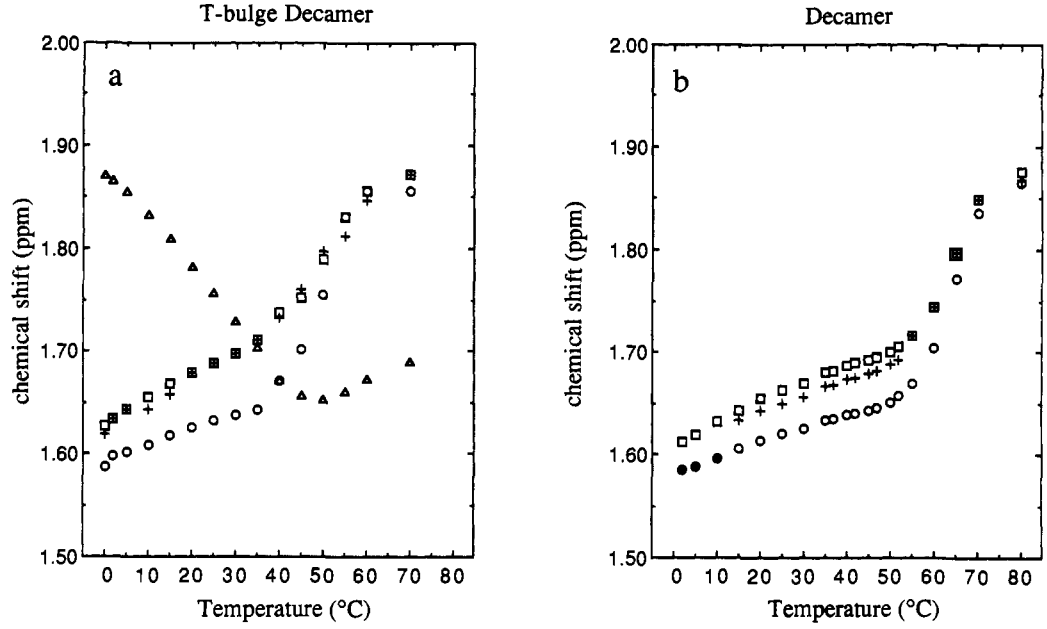

FIGURE 8: Chemical shift versus temperature curves for the thymidine methyl resonances in (a) the T-bulge decamer and (b) the decamer; (○) T14, (+) T17, (□) T15, T16, and (△) T_b.

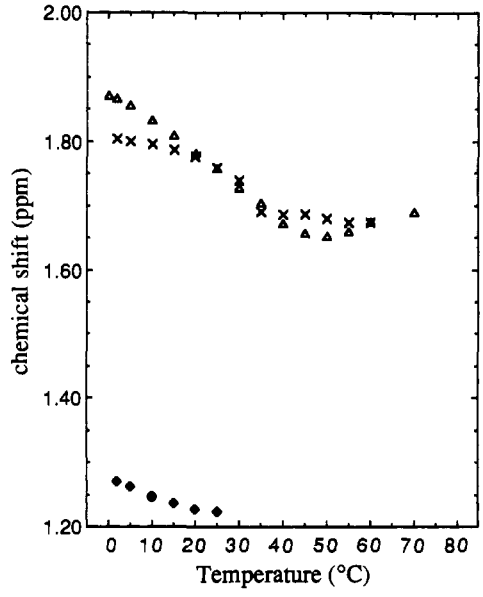

FIGURE 9: Chemical shift versus temperature curves for the thymidine methyl resonances in the T-bulge decamer (△) compared to those from the different conformational forms of the single strand dGCGAATAAGCG (× and ◆).

is extrahelical and remains extrahelical from 5 °C until the duplex melts to single strands. The observation of NOE interactions between protons from base pairs A5·T16 and A6·T15, which flank the unpaired base, indicates that these two base pairs must stack upon each other, thereby excluding the unpaired thymidine from the helix. For the T-bulge decamer at 15 °C, NOE interactions were observed between T16(H3) and T15(H3), T16(H3) and A6(H2), and T15(H3) and A5(H2) (Figure 2). Partial evidence for the unpaired thymidine being extrahelical at 25 °C is given by the sequential cross-peak between A5(H1′) and A6(H8) shown in Figure 6. This cross-peak is of comparable intensity to other sequential cross-peaks, such as the sequential cross-peak between A4(H1′) and A5(H8). Cross-peaks in the aromatic–H2′, H2″ network also provide evidence for the thymidine being extrahelical. There are cross-peaks from A5(H2′ and H2″) to A6(H8) (Figure 7). There are no cross-peaks observed between A5-

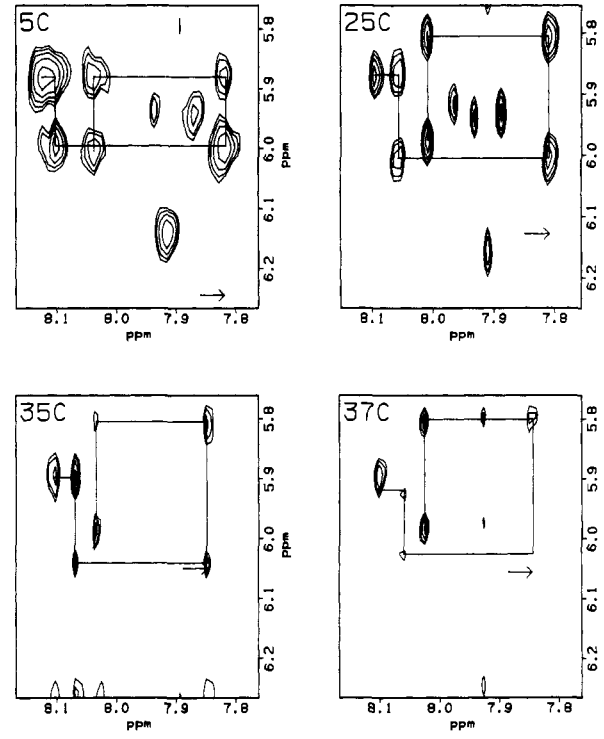

FIGURE 10: Expansion of the two-dimensional NOESY spectrum of the T-bulge decamer taken at different temperatures. The solid line shows the sequential NOE connectivities for A4–A7. The arrow indicates the position for the T_b(H1′)–A6(H8) cross-peak.

(H2″) and T_b(H6) or between T_b(H2′ or H2″) and A6(H8), and only a weak cross-peak is observed between A5(H2′) and T_b(H6). These cross-peaks should be observed if the thymidine were intrahelical. Two cross-peaks involving A(H2) protons also provide evidence for the unpaired thymidine being extrahelical, the cross-peak between A5(H2) and A6(H1′) (labeled l in Figure 6) and the cross-peak between A6(H2) and T16(H1′) (labeled r in Figure 6). Additional evidence confirming the extrahelical conformation is also obtained from cross-peaks in the aromatic–aromatic region of the spectrum (data not shown). A cross-peak is observed from A5(H2) to A6(H2) and from A5(H8) to A6(H8). All of the cross-peaks

Table I: Comparison of the Base Proton Chemical Shifts for the Decamer and the T-bulge Decamer[a]

	H8			H2			H6			H5/CH_3			H1/H3		
residue	decamer	T-bulge decamer	$\Delta\delta$[b]	decamer	T-bulge decamer	$\Delta\delta$	decamer	T-bulge decamer	$\Delta\delta$	decamer	T-bulge decamer	$\Delta\delta$	decamer	T-bulge decamer	$\Delta\delta$
G1	7.93	7.94	−0.01										13.00	13.00	0.00
C2							7.33	7.34	−0.01	5.35	5.33	0.02			
G3	7.85	7.87	−0.02										12.79	12.86	−0.07
A4	8.10	8.11	−0.01	7.23	7.39	*−0.16*									
A5	8.04	8.07	−0.03	7.02	7.23	*−0.21*									
T_b								7.44			1.77				
A6	7.99	7.81	*0.18*	7.05	7.08	−0.03									
A7	7.91	8.02	*−0.11*	7.46	7.57	*−0.11*									
G8	7.47	7.52	−0.05										12.87	12.88	−0.01
C9							7.22	7.25	−0.03	5.21	5.22	−0.01			
G10	7.91	7.91	0.00										12.97	12.97	0.00
C11							7.59	7.61	−0.02	5.86	5.86	0.00			
G12	7.97	7.97	0.00										13.06	13.08	−0.02
C13							7.50	7.50	0.00	5.42	5.40	0.02			
T14							7.49	7.51	−0.02	1.63	1.64	−0.01	14.39	14.41	−0.02
T15							7.51	7.52	−0.01	1.66	1.71	−0.05	14.08	13.77	*0.31*
T16							7.50	7.48	0.02	1.65	1.68	−0.03	13.96	13.42	*0.54*
T17							7.41	7.47	−0.06	1.65	1.69	−0.04	13.77	13.87	*−0.10*
C18							7.47	7.53	−0.06	5.68	5.71	−0.03			
G19	7.89	7.89	0.00										13.05	13.09	−0.04
C20							7.35	7.37	−0.02	5.29	5.29	0.00			

[a] All of the chemical shifts are reported at 25 °C with the exception of the H1/H3 shifts, which are at 15 °C. All are reported relative to TSP. [b] $\Delta\delta$ = δ(decamer) − δ(T-bulge decamer), $\Delta\delta \geq \pm 0.1$ ppm are shown in italic type.

Table II: Comparison of the Deoxyribose Proton Chemical Shifts for the Decamer and the T-bulge Decamer at 25 °C[a]

	H1′			H2′			H2″			H3′			H4′		
residue	decamer	T-bulge decamer	$\Delta\delta$[b]	decamer	T-bulge decamer	$\Delta\delta$	decamer	T-bulge decamer	$\Delta\delta$	decamer	T-bulge decamer	$\Delta\delta$	decamer	T-bulge decamer	$\Delta\delta$
G1	5.94	5.96	−0.02	2.58	2.61	−0.03	2.76	2.78	−0.02	4.85	4.86	−0.01	4.25	4.27	−0.02
C2	5.64	5.67	−0.03	1.93	1.93	0.00	2.30	2.31	−0.01	4.84	4.86	−0.02	4.13	4.15	−0.02
G3	5.35	5.28	0.07	2.68	2.67	0.01	2.68	2.67	0.01	4.99	4.98	0.01	4.29	4.29	0.00
A4	5.85	5.89	−0.04	2.64	2.65	−0.01	2.85	2.81	0.04	5.05	5.05	0.00	4.41	4.41	0.00
A5	5.79	6.02	*−0.23*	2.57	2.67	*−0.10*	2.83	2.73	*0.10*	5.04	5.09	−0.05	4.43	4.42	0.01
T_b		6.13			2.31			2.43			4.85			4.25	
A6	5.87	5.83	0.04	2.53	2.39	*0.14*	2.87	2.70	*0.17*	5.03	4.91	*0.12*	4.43	4.44	−0.01
A7	5.95	5.99	−0.04	2.51	2.60	−0.09	2.83	2.86	−0.03	5.01	5.01	0.00	4.41	4.43	−0.02
G8	5.69	5.71	−0.02	2.37	2.39	−0.02	2.58	2.61	−0.03	4.92	4.93	−0.01	4.36	4.37	−0.01
C9	5.72	5.76	−0.04	1.86	1.86	0.00	2.31	2.32	−0.01	4.80	4.83	−0.03	4.15	4.17	−0.02
G10	6.16	6.16	0.00	2.62	2.64	−0.02	2.39	2.39	0.00	4.68	4.68	0.00	4.19	4.20	−0.01
C11	5.75	5.77	−0.02	1.90	1.91	−0.01	2.41	2.41	0.00	4.70	4.71	−0.01	4.08	4.09	−0.01
G12	5.92	5.93	−0.01	2.73	2.75	−0.02	2.73	2.75	−0.02	5.01	4.99	0.02	4.39	4.38	0.01
C13	6.04	6.05	−0.01	2.17	2.16	0.01	2.57	2.58	−0.01	4.79	4.78	0.01	4.31	4.31	0.00
T14	6.12	6.18	−0.06	2.21	2.22	−0.01	2.65	2.65	0.00	4.91	4.93	−0.02	4.27	4.21	0.06
T15	6.19	6.20	−0.01	2.25	2.24	0.01	2.69	2.66	0.03	4.93	4.92	0.01	4.25	4.27	−0.02
T16	6.14	6.12	0.02	2.21	2.21	0.00	2.67	2.63	0.04	4.93	4.90	0.03	4.25	4.28	−0.03
T17	6.09	6.09	0.00	2.17	2.19	−0.02	2.55	2.51	0.04	4.92	4.93	−0.01	4.20	4.21	−0.01
C18	5.61	5.57	0.04	2.07	2.12	−0.05	2.39	2.43	−0.04	4.87	4.87	0.00	4.14	4.16	−0.02
G19	5.93	5.94	−0.01	2.61	2.60	0.01	2.71	2.75	−0.04	4.99	4.99	0.00	4.36	4.37	−0.01
C20	6.14	6.15	−0.01	2.20	2.20	0.00	2.20	2.19	0.01	4.49	4.50	−0.01	4.03	4.03	0.00

[a] All chemical shifts are relative to TSP. [b] $\Delta\delta$ = δ(decamer) − δ(T-bulge decamer), $\Delta\delta \geq \pm 0.1$ ppm are shown in italic type.

discussed above are due to interactions between the base pairs on either side of the unpaired thymidine and thus confirm that the flanking base pairs must be stacked and the unpaired thymidine is extrahelical.

Temperature Behavior of Extrahelical Conformation. In solution there is always an equilibrium between the duplex (or multiple duplex forms) and the single strands. This equilibrium is shifted as a function of temperature. This equilibrium has been monitored by observing chemical shifts as a function of temperature for the methyl protons on the thymidines in the two duplexes (Figure 8). The melting temperature for these protons is approximately 65 °C for the decamer and 50 °C for the T-bulge decamer. Thus the unpaired thymidine has lowered the melting temperature by approximately 15 °C, indicating a destabilized duplex. Both duplexes have a premelting transition in approximately the same temperature range. It is not clear what causes this premelting transition. However, the transition is probably not due to a differential melting of regions of the duplex because all of the imino resonances, with the exception of those from the terminal G·C base pairs, are observed up to the melting transition (see Figure 4).

Also evident in Figure 8a is the unique behavior of the methyl protons from the unpaired thymidine. The chemical shift at low temperature is similar to the shift for the other methyl resonances at high temperature, and thus it can be implied that this unpaired thymidine is in a single-stranded environment at low temperature. As the temperature increases, the chemical shift moves upfield until it reaches a minimum at 45–50 °C. The change in chemical shift from 0 to 45 °C is greater than 0.2 ppm. Similar melting curves have been observed for the H5 resonance from an unpaired cytosine (Kalnik at al., 1989b) and for the methyl resonances from an unpaired thymidine (Kalnik et al., 1990). The unusual

behavior was attributed to the unpaired base being extrahelical at low temperature and then stacking between two adjacent purines at higher temperatures. To test this explanation for our duplex, we have monitored the temperature dependence of the chemical shifts for the single strands. The melting behavior of the chemical shifts from the methyl protons in the duplex compared to the single strand is shown in Figure 9. In the single strand, two methyl resonances are observed whose relative intensities change as a function of temperature. The methyl resonance that is furthest upfield is the most intense at 2 °C and gradually decreases in intensity while the downfield resonance increases in intensity. The upfield methyl resonance completely disappears by 30 °C. Preliminary studies on the dGCGAATAAGCG single strand in 90% H_2O show four imino resonances (two G·C, one A·T, and one G·A) at temperatures below 27 °C, indicating that the single strand exists in equilibrium with a homoduplex form. UV absorption studies show a very strong concentration dependence of the secondary structure formation, thus indicating that the complex is bimolecular in nature and therefore is not a hairpin. From these studies in H_2O and studies by Roy et al. (1986), we conclude that the upfield methyl resonance is due to a homoduplex form of the single strand. The resonance represented by × in Figure 9 is most likely due to the single strand, and thus the unpaired thymidine in the duplex (Δ) has a chemical shift very close to that in the single strand for the entire temperature range. This would imply that the unpaired thymidine remains unstacked at all temperatures.

Additional evidence that the thymidine is extrahelical is the observed sequential cross-peak between A5(H1′) and A6(H8) and the lack of a sequential cross-peak between T_b(H1′) and A6(H8). These two interactions were monitored from 5 to 45 °C (Figure 10). The sequential cross-peak between A5 and A6 is observed at 5–35 °C, whereas the sequential cross-peak between T_b and A6 is not significantly above the noise at any temperature monitored. The relative intensity of the A5 intranucleotide and the A5–A6 sequential cross-peaks remains constant for all of these temperatures, although both cross-peaks decrease in intensity with increasing temperature. At temperatures above 35 °C, the sequential cross-peak between A5(H1′) and A6(H8) disappears. This temperature is not far from the melting temperature, and thus it is not surprising that the intensity of cross-peaks in the NOESY spectrum would decrease. Other cross-peaks that were previously discussed as being evidence for the unpaired thymidine being extrahelical were also observed in the temperature-dependent NOESY spectra. The cross-peak from A5(H2) to A6(H2) was observed at 5–35 °C. The cross-peak from A5(H8) to A6(H8) was observed at 15–25 °C, is a weak cross-peak at 35 °C, and is difficult to assign at 5 °C because of overlapping resonances. Thus the evidence supports the unpaired thymidine being extrahelical from 5 to 35 °C.

The imino resonances broaden and disappear at a temperature below the melting temperature (Figure 4). A comparison of line widths can provide insight into the relative stabilities of base pairs in different molecules. The T-bulge decamer at 15 and 25 °C has comparable line widths to those found in the decamer at 35 and 40 °C. Thus the T-bulge decamer double strand is destabilized by 15–20 °C relative to the decamer double strand. This confirms the results from the melting curves of the methyl chemical shifts that were discussed previously. For many sequences, this line-broadening phenomenon starts from the ends of the duplex and progresses toward the center of the molecule as the temperature is increased. However, it has also been shown that the broadening can be sequence dependent and that the order of broadening is not necessarily indicative of location in the molecule (Patel et al., 1983). The relative order of broadening between the two duplexes is slightly different. For both duplexes, the imino resonances from the ends of the molecule broaden first, the oligo A·T core broadens next, and the G3 and G8 imino resonances are the last to broaden. However, in the T-bulge decamer the oligo A·T core is destabilized relative to the rest of the molecule, as shown by the broadening of resonances from the oligo A·T core simultaneously with the G18 and G12 imino resonances.

Evidence for Unusual A·T Tract Structure. Tracts of A·T base pairs have been shown, among other properties, to have a high (>15°) propeller twist (Nelson et al., 1987; Coll et al., 1987; Yoon et al., 1988), to have NOESY cross-peaks not found in other B-form sequences (Kintanar et al., 1987; Roy et al., 1987a; Katahira et al., 1988; Sarma et al., 1988; Gupta et al., 1988; Celda et al., 1989; Nadeau & Crothers, 1989), and to exhibit unusual hydrogen exchange phenomena (Leroy et al., 1988). Both Leroy et al. (1988) and Nadeau and Crothers (1989) conclude that these unusual structures exist for runs of oligo(A)·oligo(T) that contain four or more A·T base pairs.

The NOESY cross-peaks that characterize the A·T tract involve interactions with the A(H2) protons. These cross-peaks are shown for the decamer in Figure 5 and can be classified in three categories: (1) cross-peaks between the H2 and H1′ protons on the same nucleotide, as demonstrated by cross-peak g for the A7 nucleotide, (2) cross-peaks between H2 protons and the H1′ on the same strand but on the nucleotide in the 3′ direction, such as cross-peak e in Figure 5 between A7(H2) and G8(H1′), and (3) cross-peaks due to a diagonal interaction between A(H2) and the H1′ belonging to the nucleotide adjacent (in the 3′ direction) to the complementary T, as demonstrated by cross-peak h between A7(H2) and T15(H1′). Similar cross-peaks involving interactions with A4(H2), A5-(H2), and A6(H2) were observed and are shown in Figure 5.

The T-bulge decamer also shows NOEs involving A(H2) protons that are characteristic of A·T tracts. These cross-peaks are shown in Figure 6; examples are cross-peak n, due to an interaction within the same nucleotide, cross-peak p, due to interaction within the same strand, and cross-peak r, due to diagonal interstrand interaction. Also present in Figure 6 is cross-peak k, which is between A4(H2) and T17(H1′); this is an interaction across the duplex but within the same base pair. An analogous cross-peak is observed for the decamer (cross-peak l, Figure 5), and a similar cross-peak has been observed by Katahira et al. (1988) for the terminal A·T base pair on the 3′ side of the tract. This "unusual" cross-peak has been associated with a structural discontinuity at the junction of the A·T tract. For the decamer and the T-bulge decamer, this cross-peak is observed at the 5′ end of the tract.

We have obtained NOE data for the T-bulge decamer at several mixing times: 25, 50, 75, 100, and 200 ms. These data have not yet been analyzed quantitatively to obtain distances. The following qualitative analysis was made on the basis of the relative intensities of cross-peaks at the different mixing times, with the exclusion of the 25-ms data, which were too noisy to provide any significant data. The strongest of the interstrand A(H2) cross-peaks in the T-bulge decamer is between A7(H2) and T15(H′) (cross-peak g in Figure 6), which is visible at mixing times of 50–200 ms. The cross-peak between A5(H2) and T17(H1′) is visible at 75–200 ms. The cross-peak between A4(H2) and T18(H1′) is visible at 100 and 200 ms, and the weakest cross-peak, between A6(H2) and

T16(H1′), is only visible at 200 ms. With the exception of the interaction for A6(H2), this is in agreement with previous results supporting a narrowing of the minor groove from the 5′ to 3′ direction along the dA tract (Katahira et al., 1990). The anomaly at the A6 position is probably due to the adjacent unpaired thymidine.

Structural Consequences of the Unpaired Thymidine. From the imino resonance broadening and the melting temperatures derived from the methyl resonances, it is evident that the unpaired thymidine has caused an overall destabilization of the duplex. The imino resonances indicate more specifically that this destabilization is localized in the oligo(A·T) core of the molecule. Evidence for localized structural changes in the T-bulge decamer is also seen in a comparison of the chemical shifts for the two duplexes (Table I). Chemical shift changes greater than 0.1 ppm occur for protons that are on adenines adjacent to or one base pair removed from the unpaired thymidine, for example, the H8 protons on A6 and A7 and the H2 protons on A4, A5, and A7. There are also large changes in the shifts for H1′, H2′, and H2″ of A5 and H2′, H2″, and H3′ of A6 (Table II). There are sizable changes in the chemical shifts of the exchangeable H3 protons (Table I), with the most significant differences observed for T15(H3) and T16(H3), the thymidines adjacent to the perturbation. In contrast, the chemical shift changes for the nonexchangeable protons on the dCGCTTTTCGC strand are minimal. This observation is difficult to interpret, as it could indicate that there are no structural changes for the dCGCTTTTCGC strand or that they are not reflected in the chemical shifts, perhaps due to minimal ring-current shifts from the run of thymidines.

Evidence for a perturbed A·T tract in the T-bulge decamer is obtained from the chemical shifts of the imino resonances. Nadeau and Crothers (1989) observed a trend for the chemical shifts of the imino resonances in the A·T tracts where the imino resonance from the 5′-most thymidine is the furthest downfield and the chemical shifts progressively move upfield, going from the 5′-most to the 3′-most thymidine imino resonance. This trend in chemical shifts is observed for the decamer (Figures 2h and 4); however, it is disrupted in the T-bulge decamer (Figures 2d and 4).

Several cross-peaks involve interactions with the T_b protons and are consistent with the unpaired thymidine being intrahelical. This would imply that there is a rapid equilibrium between the extrahelical and intrahelical conformation. An estimate of the relative populations of these conformations can be made by comparing intensities from slices of the NOESY spectrum shown in Figure 6. For example, comparing the intensity of the T_b(H1′)–A6(H8) cross-peak, which is just above the noise level, with the A5(H1′)–A6(H8) cross-peak results in a minimum of 85% in the extrahelical conformation. Comparing the intensity of the A5(H1′)–T_b(H6) cross-peak (designated with an asterisk in Figure 6) with the A5-(H1′)–A6(H8) cross-peak results in a minimum of 65% in the extrahelical conformation. Similar estimates are obtained by analyzing data from the aromatic–H2′, H2″ region of the NOESY spectrum. We believe the value of 65% is a lower limit for the extrahelical conformation because preliminary modeling studies in our laboratory indicate it is possible to build a helix with the extrahelical thymidine situated in the minor groove, allowing for an NOE between A5(H1′) and T_b(H6). That an extrahelical base could lie in the groove of the DNA and thus in close proximity to one of the neighboring nucleotides has been demonstrated by Joshua-Tor et al. (1988) and Miller et al. (1988) by use of X-ray crystallography.

For the duplexes containing an unpaired pyrimidine that were studied by Kalnik et al. (1989b, 1990), the intrahelical population was proposed to increase with temperature. If this occurs in the T-bulge decamer, then cross-peaks attributed to the intrahelical conformation should increase in intensity at higher temperatures. This is observed for the cross-peak between T_b(H6) and A5(H2″), which is observed only at 35 and 37 °C. However, the cross-peak observed between T_b(Me) and A5(H8) (peak a in Figure 6) is observed at 15–32 °C and disappears at temperatures above 32 °C. In addition, a weak cross-peak is observed between A5(H1′) and T_b(H2′) at 15 °C; however, it is not observed at any other temperature.

Comparison with Other Oligonucleotides Containing Unpaired Bases. The crystallographic studies (Joshua-Tor et al., 1988; Miller et al., 1988) are the only results thus far to show that an unpaired purine can exist in an extrahelical conformation. There have been several solution studies on unpaired adenines (Patel et al., 1982; Hare et al., 1986; Woodson & Crothers, 1987, 1989; Roy et al., 1987b; Kalnik et al., 1989a; Nikonowicz et al., 1989) and guanines (Woodson & Crothers, 1988a,b) where the base was found to be intrahelical. It has been hypothesized that the difference in conformation may be due to forces that are present in the condensed phase from which the crystallography results are derived.

There have also been several studies on duplexes containing unpaired pyrimidines. These investigations have all been in solution, and both the extrahelical and intrahelical conformations have been observed. Van den Hoogen et al. (1988) found for the sequence d(CTGGTGCGG)·d(CCGCCCAG) that the unpaired thymidine was intrahelical. One of their most compelling pieces of evidence is the observation of the imino resonance from the unpaired thymidine, which suggests a protected environment for this otherwise labile proton. We do not see an analogous resonance for the T-bulge decamer; however, observation of this resonance is probably pH dependent, and the experiments presented here were restricted to pH 7. In another study by Morden et al. (1983) on the sequence d(CAAACAAAG)·d(CTTTTTTG), the extra cytosine was found to be extrahelical. Results from Kalnik et al. (1989b) on the sequence $d(CCGCGAATTCCGG)_2$ found the unpaired cytosines to be extrahelical at low temperature and intrahelical closer to the melting temperature. Similar results were found for an unpaired thymidine in the sequence of $d(CCGTGAATTCCGG)_2$ (Kalnik et al., 1990). Thus, it has been shown in solution that an unpaired pyrimidine can exist in either an intrahelical or extrahelical conformation. What are the factors that determine the dominant conformation in solution? One of the factors that influences the structure of an unpaired base is the sequence of the flanking base pairs. In the T-bulge decamer, the unpaired thymidine is in the center of an A·T tract. This A·T tract has an unusual structure in both the decamer and the T-bulge decamer. It is interesting to note that the other exclusively extrahelical base was also in an A·T tract (Morden et al., 1983).

What is the balance of forces that decides whether the unpaired base is extrahelical or intrahelical? The intrahelical conformation of the unpaired base would be stabilized by the hydrophobic effect and by the two stacking interactions of the unpaired base with each of the flanking base pairs. The extrahelical conformation would be stabilized by the stacking interaction between the flanking base pairs and perhaps by a favorable entropy term due to the conformational freedom of the unstacked thymidine. For the base sequence in the oligonucleotide studied here, the intrahelical conformation has an A-T and a T-A stacking interaction, compared to an A-A

stack in the extrahelical conformation. The extrahelical conformation also has the additional stabilization of the oligo(A·T) stack, which only occurs in runs of at least four A·T base pairs. There will also be a contribution from the solvent and counterion binding; however, it is not clear how these act on the two conformations. Our evidence indicates that for the T-bulge decamer the extrahelical conformation dominates. Perhaps this can be understood in that several of the factors that stabilize the intrahelical conformation are very weak. Thymine is the least hydrophobic of the four bases, thus being the least destabilizing in the extrahelical conformation. The A-T and the A-A stack are moderately good; however, the T-A stack is the poorest of all the base-stacking interactions (Orenstein et al., 1978; Gotoh & Takashira, 1981), thus gaining little for the intrahelical conformation.

CONCLUSIONS

The unpaired thymidine in the T-bulge decamer has been shown to be extrahelical at all temperatures below the melting temperature. The exact nature of how the extrahelical thymidine is accommodated in the duplex is not known; however, the flanking A·T base pairs form a B-form-like stack. Several additional NOEs are observed between the unpaired base and the flanking base pairs that may be useful in determining the position of the thymidine.

It is evident, from comparing all the duplexes that contain unpaired pyrimidines, that the neighboring sequence has major influence on the conformation of the unpaired base. The duplex presented in this paper has the unpaired base in an oligo(A·T) tract, and the ability of this sequence to form a unique structure is part of the inducement for the unpaired thymidine being extrahelical. However, the structure of any duplex is determined by a delicate balance of many forces, and the type of base that is in the unpaired location will also be a factor. Preliminary results from our laboratory on the analogous sequences containing an unpaired cytosine and an unpaired guanine indicate that these bases are extrahelical and intrahelical, respectively. Clearly, more sequences need to be investigated in order to determine how these sequence-dependent forces contribute to the final structure of the duplex.

ACKNOWLEDGMENTS

We thank Marty Beasley for the synthesis and purification of the dCGCTTTTCGC oligonucleotide.

Registry No. d(CGCTTTTCGC)·d(GCGAAAAGCG), 128683-58-9; d(CGCTTTTCGC)·d(GCGAATAAGCG), 128683-60-3; thymidine, 50-89-5; cytosine, 71-30-7; guanine, 73-40-5.

REFERENCES

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

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

Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., & Wüthrich, K. (1989) *Biochemistry 28*, 1462–1471.

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

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

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

Fresco, J. R., & Alberts, B. M. (1960) *Biochemistry 46*, 311–321.

Gotoh, O., & Takashira, Y. (1981) *Biopolymers 20*, 1033–1042.

Gupta, G., Sarma, M. H., & Sarma, R. H. (1988) *Biochemistry 27*, 7909–7919.

Haasnoot, C. A. G., den Hartog, J. H. J., de Rooij, J. F. M., van Boom, J. H., & Altona, C. (1980) *Nucleic Acids Res. 8*, 169–181.

Hare, D. R., Wemmer, D. E., Chou, S., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7456–7464.

Hore, P. J. (1983a) *J. Magn. Reson. 54*, 539–542.

Hore, P. J. (1983b) *J. Magn. Reson. 55*, 283–300.

Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 82–84.

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.

Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R., & Tomita, K. (1988) *Nucleic Acids Res. 16*, 8619–8632.

Katahira, M., Sugeta, H., & Kyogoku, Y. (1990) *Nucleic Acids Res. 18*, 613–618.

Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acid Res. 15*, 5845–5862.

Leroy, J., Charretier, E., Kochoyan, M., & Guêron, M. (1988) *Biochemistry 27*, 8894–8898.

Lomant, A. J., & Fresco, J. R. (1974) *Prog. Nucleic Acid Res. Mol. Biol. 14*, 185–218.

Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 85–86.

Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 22*, 5557–5563.

Nadeau, J. G., & Crothers, D. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2622–2626.

Nagayama, K., Kumar, A., Wüthrich, K., & Ernst, R. R. (1980) *J. Magn. Reson. 40*, 321–334.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714–8725.

Ornstein, R. L., Rein, R., Breen, D. L., & MacElroy, R. D. (1978) *Biopolymers 17*, 2341–2360.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 445–451.

Patel, D. J., Kozlowski, S. A., Ikuta, S., Itakura, K., Bhatt, R., & Hare, D. R. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 197–206.

Roy, S., Weinstein, S., Borah, B., Nickol, J., Appella, E., Sussman, J. L., Miller, M., Shindo, H., & Cohen, J. S. (1986) *Biochemistry 25*, 7417–7423.

Roy, S., Borah, B., Zon, G., & Cohen, J. S. (1987a) *Biopolymers 26*, 525–536.

Roy, S., Sklenar, V., Appella, E., & Cohen, J. S. (1987b) *Biopolymers 26*, 2041–2052.

Sarma, M. H., Gupta, G., & Sarma, R. H. (1988) *Biochemistry 27*, 3423–3432.

Scheek, R. M., Russo, N., Boelens, R., & Kaptein, R. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77–84.

van den Hoogen, Y. Th., van Beuzekom, A. A., van den Elst, H., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988) *Nucleic Acids Res. 16*, 2971–2986.
Wagner, G., (1983) *J. Magn. Reson. 55*, 151–156.
Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 130–134.
Woodson, S. A., & Crothers, D. M. (1987) *Biochemistry 26*, 904–912.
Woodson, S. A., & Crothers, D. M. (1988a) *Biochemistry 27*, 436–445.
Woodson, S. A., & Crothers, D. M. (1988b) *Biochemistry 27*, 3130–3141.
Woodson, S. A., & Crothers, D. M. (1989) *Biopolymers 28*, 1149–1177.
Yoon, C., Privé, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.

Two-Dimensional ^1H and ^{31}P NMR Spectra and Restrained Molecular Dynamics Structure of a Mismatched GA Decamer Oligodeoxyribonucleotide Duplex†

Edward P. Nikonowicz and David G. Gorenstein*
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
Received March 29, 1990; Revised Manuscript Received June 12, 1990

ABSTRACT: Assignment of the ^1H and ^{31}P NMR spectra of a tandem G·A mismatched base pair decamer oligodeoxyribonucleotide duplex, d(CCAA*GA*TTGG)$_2$, has been made by two-dimensional ^1H–^1H and heteronuclear ^{31}P–^1H correlated spectroscopy. Unusual downfield ^{31}P resonances have been assigned by a pure absorption phase constant-time heteronuclear ^{31}P–^1H correlated spectrum to be associated with the phosphates on the 5′- and 3′-sides of the mismatched guanosine residue. $J_{H3'-P}$ coupling constants for each of the phosphates of the decamer were obtained from the ^1H–^{31}P *J*-resolved selective proton-flip 2D spectrum. The two most downfield-shifted ^{31}P resonances each appear to consist of two overlapping signals that can be resolved into two distinct doublets with different coupling constants in the *J*-resolved spectrum. This as well as the temperature dependence of the ^{31}P spectra demonstrates that two distinct conformations exist at lower temperatures. By use of a modified Karplus relationship, the C4′–C3′–O3′–P torsional angles (ϵ) were obtained. A linear correlation between ^{31}P chemical shifts and the measured coupling constants is quite good (only when the larger set of coupling constants of the two most downfield ^{31}P signals is included). The ^{31}P chemical shifts as well as the measured coupling constants tend to follow the positional variation seen in other duplexes of interior phosphates resonating more upfield than terminal residues and of interior phosphates exhibiting smaller coupling constants; however, this pattern is disrupted at the site of the mismatch. Modeling and initial NOESY distance restrained molecular mechanics energy minimization and restrained molecular dynamics support previous observations that the mismatched guanine and adenine bases are both in anti conformations. Most significantly, the ϵ backbone torsional angle variaions calculated from the NOESY distance restrained structures are in agreement with both the crystal structure values and the measured $J_{H3'-P}$ coupling constants.

The formation of base pair mismatches in DNA and the associated problems during replication, transcription, and gene activation if left unrepaired are current topics of concern (Modrich, 1987; Radman & Wagner, 1986). Their recognition and efficiencies of repair have been studied in bacteria (Fazakerley et al., 1986; Lu et al., 1984; Radman & Wagner, 1984), the latter having been found to be sequence dependent. Both X-ray crystallography and NMR spectroscopy have been used to probe the structural anomalies of mismatched base pairs (Brown et al., 1986; Dodgson & Wells, 1977a,b; Hare et al., 1986a; Kan et al., 1983; Kouchakdjian et al., 1988; Patel et al., 1984a,b; Prive et al., 1987). One of the first nonstandard types of base pairs to be structurally investigated in solution was the naturally occurring purine–pyrimidine base pair G·U found in tRNA (Hare & Reid, 1979; Johnston & Redfield, 1978). This G·U base pair was shown to be in a "wobble"-type pairing scheme. Later, Patel et al. examined the dynamics of the imino protons of four pseudo-self-complementary dodecamers containing pairs of well-separated mismatches of the types G·T, G·A, A·C, and T·C derived from d(CGX$_1$GAATTCX$_2$CG)$_2$, where X$_1$ and X$_2$ are the mismatched pair (Patel et al., 1984c, 1982a,b). The G·T was demonstrated to be wobble base paired; the G·A formed an anti–anti base-paired scheme through the adenine NH$_2$/N3 and guanine O6/NH1.

X-ray crystallography has also proved to be a useful tool for the elucidation of the structure of deoxyoligonucleotides, although crystal results are not always in good agreement with solution-state results (Joshua-Tor et al., 1988; Nikonowicz et al. 1990). The purine–purine mismatch G·A has been crystallized in two different sequences: the dodecamer d(CGCGAATTAGCG)$_2$ (Brown et al., 1986) and the decamer d(CCAAGATTGG)$_2$ (Prive et al., 1987). In the former case, the mismatch was determined to be in a G_{anti}–A_{syn} conformation and in the latter molecule the conformation was found

† Supported by the NIH (AI27744), the Purdue University Biochemical Magnetic Resonance Laboratory which is supported by the NIH (Grant RR01077 from the Biotechnology Resources Program of the Division of Research Resources), the NSF National Biological Facilities Center on Biomolecular NMR, Structure and Design at Purdue (Grants BBS 8614177 and 8714258 from the Division of Biological Instrumentation), and the National AIDS Research Center at Purdue (AI727713).

0006-2960/90/0429-8845$02.50/0