Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. (1980) *Nature 287*, 755–758.
Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.
Zimmer, C., & Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.

# NMR Studies of Triple-Strand Formation from the Homopurine–Homopyrimidine Deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$†

Ponni Rajagopal and Juli Feigon*

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

*Received March 9, 1989; Revised Manuscript Received May 25, 1989*

ABSTRACT: The complexes formed by the homopurine and homopyrimidine deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$ have been investigated by one- and two-dimensional $^1H$ NMR. Under appropriate conditions [low pH, excess $d(TC)_4$ strand] the oligonucleotides form a triplex containing one $d(GA)_4$ and two $d(TC)_4$ strands. The homopurine and one of the homopyrimidine strands are Watson–Crick base paired, and the second homopyrimidine strand is Hoogsteen base paired in the major groove to the $d(GA)_4$ strand. Hoogsteen base pairing in GC base pairs requires hemiprotonation of C; we report direct observation of the $C^+$ imino proton in these base pairs. Both homopyrimidine strands have C3′-endo sugar conformations, but the purine strand does not. The major triplex formed appears to have four TAT and three $CGC^+$ triplets formed by binding of the second $d(TC)_4$ strand parallel to the $d(GA)_4$ strand with a 3′ dangling end. In addition to the triplexes formed, at least one other heterocomplex is observed under some conditions.

Stretches of homopurine–homopyrimidine sequences are a widely dispersed component of eukaryotic genomes. Several lines of evidence have indicated that such sequences may adopt alternative (noncanonical B-DNA) structures and that these structures may act as regulatory signals in gene expression [for a review, see Wells et al. (1988)]. It has also been suggested that homopurine–homopyrimidine sequences may play a role in telomer formation (Blackburn, 1984), chromosome folding (Johnson & Morgan, 1978), and recombination (Hoffman-Liebermann et al., 1986; Konopka et al., 1988; Wohlrab et al., 1987). Homopurine–homopyrimidine sequences frequently flank transcribed eukaryotic genes (Wells et al. 1988). Evidence for an unusual structure for these sequences includes a pH and supercoil density dependent hypersensitivity to DNase I and single-strand-specific nucleases (Elgin, 1984; Cantor & Efstratiadis, 1984). On the basis of this and other evidence, Pulleyblank and co-workers proposed a model for $d(TC)_n$·$d(GA)_n$ in supercoiled plasmids in which Watson–Crick AT base pairs alternate with Hoogsteen $^{syn}GC^+$ base pairs (Pulleyblank et al., 1985). A currently more widely accepted model is that of Frank-Kamenetskii and co-workers, in which mirror repeat homopurine–homopyrimidine sequences loop out to form a pyr–pur–pyr (homopyrimidine–homopurine–homopyrimidine) triplex and a single-stranded homopurine stretch, termed H-DNA (Mirkin et al., 1987). Several recent studies support this model (Johnson, 1988; Htun & Dahlberg, 1988). Christophe et al. (1985) have also proposed an intramolecular triplex model.

Triplex formation has also been demonstrated in vitro by Moser and Dervan (1987) and by Praseuth et al. (1988). The former synthesized homopyrimidine oligodeoxyribonucleotides with attached EDTA–Fe cleaving agents and found binding to corresponding homopurine–homopyrimidine tracts in DNA oligonucleotides and a restriction fragment. This work has generated great interest because of the potential applications of such sequences as highly sequence-specific probes for use in chromosome mapping (Strobel et al., 1988) and possible medical applications as antisense DNA.

Evidence for triplex formation in RNA and DNA homopolymers has been known since the 1950s (Felsenfeld et al., 1957), but such structures were generally assumed to be biologically irrelevant (Wells, 1988). Models of triplexes were proposed in which the second homopyrimidine strand was bound in the major groove parallel to the homopurine strand by Hoogsteen (1959) base pairing, and those models were supported by fiber diffraction data on RNA and DNA triplexes composed of two T and one A strand (Arnott & Selsing, 1974). The more recent chemical studies of Moser and Dervan (1987) and Praseuth et al. (1988) have also supported the early models and confirmed the parallel orientation of the second homopyrimidine strand relative to the purine strand which had been proposed on the basis of steric considerations. More recent physical chemical studies (CD, melting) of $d(GA)_n$ and $d(TC)_n$ polymers have indicated that pyr–pur–pyr triplexes as well as some other hetero- and homocomplexes are formed under various conditions (Lee et al., 1979; Antao & Gray, 1988).

In addition to pyr–pur–pyr triplexes, triplexes composed of two *homopurine* and one homopyrimidine strand have also been reported in polymers (Broitman et al., 1987; Letai et al., 1988). Recently, an in vitro triplex composed of a 27-base

† This research was supported by National Institutes of Health Grant R01 GM 37254-01 and Office of Naval Research Contract N00014-88-K-0180.

* Author to whom correspondence should be addressed.

0006-2960/89/0428-7859$01.50/0 

purine-rich sequence bound to the complementary DNA region 115 base pairs upstream from the transcription origin of the c-myc oncogene was reported (Cooney et al., 1988), and binding was shown to repress transcription in an in vitro assay. However, no model was proposed for the structure.

In this work, we have investigated the structures of the complexes formed from the homopurine $d(GA)_4$ and homopyrimidine $d(TC)_4$ DNA oligonucleotides. At low pH ($\leq$6.1) a pyr–pur–pyr triplex is observed, and addition of excess $d(TC)_4$ strand pushes the equilibrium toward the triplex form. We report conformational details of this triplex as well as conditions that stabilize the triplex. Proposed base-pairing schemes require protonation of the dC in the third strand for Hoogsteen base pairs to form, and we present direct evidence for the protonated dC imino in the triplex. A preliminary report on these results has recently been presented elsewhere (Rajagopal & Feigon, 1989).

## Materials and Methods

*Sample Preparation.* DNA oligonucleotides were synthesized by the phosphotriester method and purified in the laboratory of Dr. Jacques van Boom as previously described (van der Marel et al., 1981), or they were synthesized on an Applied Biosystems 381A DNA synthesizer using $\beta$-cynanoethyl phosphoramidites and 10-$\mu$mol columns. The latter DNA (used for most of the samples) was purified by gel filtration following the method of Kintanar et al. (1987). Purified oligonucleotides ran as single bands on denaturing polyacrylamide gels. Sample concentrations were determined by $A_{260}$ by using extinction coefficients of 9500 and 7500 L mol$^{-1}$ cm$^{-1}$ for $d(GA)_4$ and $d(TC)_4$, respectively (Lee et al., 1979; Antao et al., 1988) and were 2–6 mM in each strand depending on the sample. Unless otherwise indicated in the figure legend, duplex samples were prepared by mixing equimolar amounts of the two strands at pH 7.3 in the appropriate buffer. The sample was lyophilized, redissolved in $D_2O$, relyophilized, redissolved in $D_2O$, and transferred to a NMR tube. Samples were dried in the NMR tube under a stream of $N_{2(g)}$ and redissolved in 400 $\mu$L of 99.996% $D_2O$. Spectra of the same samples in water were obtained after redrying the samples and redissolving in 90% $H_2O$/10% $D_2O$. Samples containing triplex were prepared by adding appropriate amounts of the third strand dissolved in $H_2O$ or $D_2O$ and then lowering the pH to 6.1 or 5.5 by addition of HCl to the duplex samples. These samples were then redried and redissolved in 400 $\mu$L of $D_2O$ or 90% $H_2O$/10% $D_2O$. Samples in which the H8 protons had been replaced by deuterons were prepared by heating the $d(GA)_4$ strand at 95 °C for 12 h, resulting in complete exchange of the AH8 and GH8 protons (Schweizer et al., 1964), and then the triplex was prepared as above.

*NMR Spectroscopy.* All NMR experiments were done at 500 MHz on a General Electric GN500 spectrometer. Chemical shifts were obtained by reference to the chemical shift of water, which had been previously calibrated relative to DSS. One-dimensional spectra in $H_2O$ were acquired using a $1\bar{1}$ spin-echo pulse sequence ($90°_x$–$\tau$–$90°_{-x}$–$\Delta$–$90°_\phi$–$2\tau$–$90°_{-\phi}$–$\Delta$) (Sklenar et al., 1987). The delay $\tau$ was adjusted so that the region of maximum excitation was centered between the aromatic and the imino resonances ($\tau$ = 62–72 $\mu$s). Nuclear Overhauser effect spectroscopy (NOESY) spectra in $D_2O$ were obtained in the pure absorption mode following the method of States et al. (1982) by using the standard $90°$–$t_1$–$90°$–$\tau_m$–$90°$–acquire pulse sequence (Kumar et al., 1980) with presaturation of the residual HDO peak during the 2-s recycle delay. Spectra were acquired with 2048 complex points in $t_2$ and 260–350 $t_1$ values. The sweep width was 5000 Hz in both dimensions, and 64 scans were acquired per $t_1$ value. Phase-sensitive NOESY spectra of samples in $H_2O$ were obtained by replacing the last 90° pulse with a $1\bar{1}$ spin-echo pulse sequence and phase cycling appropriately in order to suppress the large water resonance (Sklenar et al., 1987). The carrier was centered at the water resonance, and the delay $\tau$ was adjusted so that both the imino and the aromatic resonances were at regions of maximum excitation ($\tau$ = 150 $\mu$s). Spectra were acquired with 4096 complex points in $t_2$ and 216–348 $t_1$ values. The sweep width was 14 048 Hz in both dimensions, and 128 scans were acquired per $t_1$ value. ROESY (Bothner-By et al., 1984; Bax & Davis, 1985) spectra in $D_2O$ were acquired in the same way as the NOESY spectra in $D_2O$ except that a spin-lock field was applied during the mixing time. For the spectrum shown, the spin-lock field was 4 kHz applied during a 116-ms mixing period. COSY (Aue et al., 1984) and DQF-COSY (Rance et al., 1983) experiments were acquired with standard pulse sequences and phase cycling (spectra not shown).

One-dimensional spectra were processed with the GE NMR software (GEM16). Two-dimensional spectra were transferred via magnetic tape or ethernet to a VAX 8800 or Microvax II computer and processed with the NMR processing program FTNMR (Hare Research). Processing parameters are indicated in the figure legends.

## Results

*One-Dimensional Spectra of Duplex and Triplex.* Spectra of $d(GA)_4$ and $d(TC)_4$ under two different conditions are shown in Figure 1. Figure 1C shows the spectrum obtained for an equimolar mixture of the two strands at pH 7.3 in $D_2O$. Under these conditions the predominant conformation is a B-DNA duplex (see below). The imino resonances of the same sample in $H_2O$ are shown in Figure 1A. Imino resonances expected from the eight Watson–Crick base pairs are evident in this spectrum. Figure 1B,D shows the spectra obtained with 1:2.5 $d(GA)_4$:$d(TC)_4$ at pH 5.5 Dramatic changes are observed in both the $D_2O$ and the $H_2O$ spectra upon lowering of the pH and addition of excess $(TC)_4$ strand. The $H_2O$ spectrum (Figure 1B) shows the appearance of several additional imino proton resonances. The $D_2O$ spectrum (Figure 1D) also shows many new resonances. Under these conditions the major form of the DNA is a triplex composed of two $d(TC)_4$ and one $d(GA)_4$ strand, with some single-stranded $d(TC)_4$ also present (see below).

*NOESY Spectra of the Nonexchangeable Resonances of the Duplex and Triplex.* A NOESY spectrum of the sample under the conditions where it is predominantly in the duplex form [with some excess single-stranded $d(TC)_4$ present] is shown in Figure 2. For the most part the sample under these conditions [pH 8.3, 1:1.3 $d(GA)_4$:$d(TC)_4$] gives a pattern of cross-peaks that would be expected for a B-DNA helix (Feigon et al., 1983; Scheek et al., 1983; Hare et al., 1983). This allowed straightforward assignments of the resonances of the DNA in the duplex form, and those assignments are discussed below. In addition to the expected cross-peaks, several additional cross-peaks are observed in this spectrum, which arise from exchange between the duplex form and a small amount of another DNA conformation that is present under these conditions. These exchange cross-peaks are observed close to the diagonal, most obviously in the aromatic, H1′, and methyl regions of the spectrum. This third DNA conformation (hereafter referred to as the heterocomplex) is not the same as the triplex observed in the presence of excess $d(TC)_4$ strand and low pH, does not correspond to single strands, and has

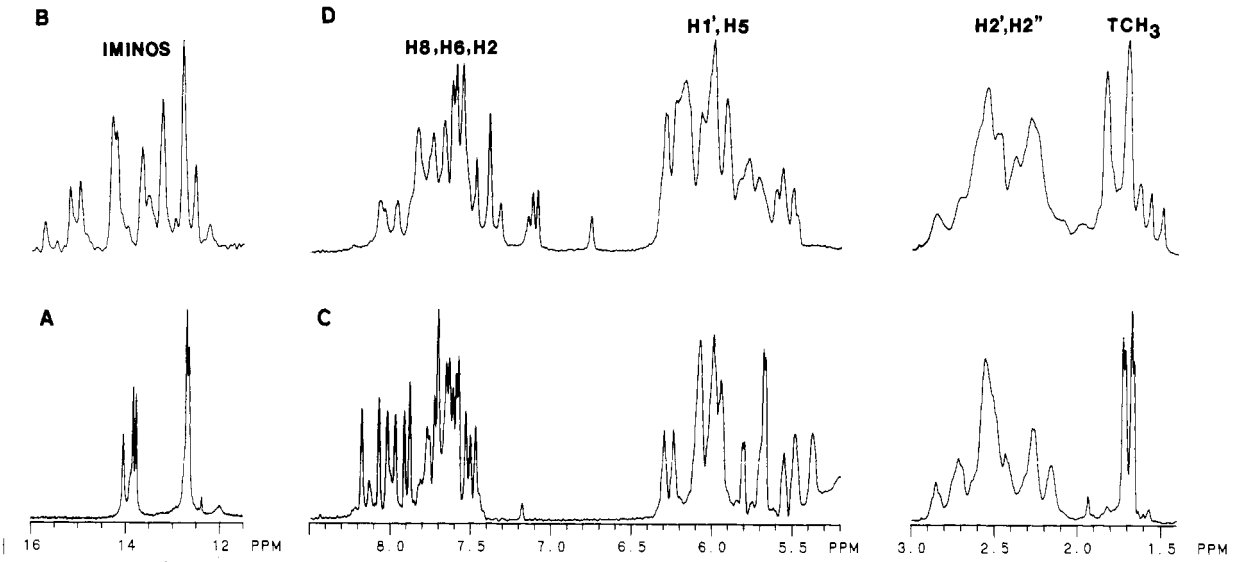



FIGURE 1: 500-MHz $^1$H NMR spectra of $d(GA)_4 + d(TC)_4$ in 90% $H_2O$ (A and B) and $D_2O$ (C and D). (A) 1:1 $d(GA)_4$:$d(TC)_4$, 20 mM phosphate, pH 7.3, 50 mM NaCl, 5 mM $MgCl_2$, 5 °C, 2 mM in each strand. (B) 1:2.5 $d(GA)_4$:$d(TC)_4$, 10 mM phosphate, pH 5.5, 100 mM NaCl, 5 mM $MgCl_2$, 10 °C. For (A) and (B), 256 acquisitions were acquired by using a $1\bar{1}$ spin-echo pulse sequence as described, and spectra were line-broadened by 3 Hz. (C) 1:1 $d(GA)_4$:$d(TC)_4$, 20 mM phosphate, pH 7.3, 50 mM NaCl, 5 mM $MgCl_2$, 15 °C [same sample as (A) in $D_2O$]. (D) 1:2.5 $d(GA)_4$:$d(TC)_4$, 10 mM phosphate, pH 5.5, 100 mM NaCl, 5 mM $MgCl_2$, 35 °C [same sample as (B) in $D_2O$]. Spectra in (C) and (D) were acquired with a sweep width of 5000 Hz, 8K complex points, and 128 and 64 acquisitions, respectively, and were resolution enhanced by double-exponential multiplication.

Table I: Assignments of the Nonexchangeable (15 °C) and Exchangeable (1 °C) Resonances of the Duplex $d(GA)_4$·$d(TC)_4$

| base | AH8, GH8, CH6 | AH2, TMe, CH5 | H1′ | H2′ | H2″ | H3′ | H4′ | imino H1, H3 | amino AH4, CH6(2) | amino AH4, CH6(1) |
|---|---|---|---|---|---|---|---|---|---|---|
| G1 | 7.88 | | 5.53 | 2.48 | 2.67 | 4.81 | 4.14 | | | |
| A2 | 8.18 | 7.64 | 5.94 | 2.73 | 2.85 | 5.03 | 4.37 | | 7.81 | 6.13 |
| G3 | 7.70 | | 5.47 | 2.54 | 2.69 | 4.98 | 4.35 | 12.61 | | |
| A4 | 8.01 | 7.51 | 5.99 | 2.82 | 2.56 | 5.01 | | | 7.76 | 5.80 |
| G5 | 7.62 | | 5.37 | 2.47 | 2.60 | 4.95 | 4.30 | 12.66 | | |
| A6 | 7.96 | 7.61 | 5.94 | 2.45 | 2.72 | 4.98 | | | 7.81 | 5.02 |
| G7 | 7.61 | | 5.49 | 2.42 | 2.55 | 4.92 | 4.29 | 12.64 | | |
| A8 | 8.05 | 7.78 | 6.29 | 2.54 | 2.40 | 4.63 | 4.22 | | | |
| T9 | 7.58 | 1.72 | 6.11 | 2.23 | 2.54 | 4.75 | | 13.90 | | |
| C10 | 7.77 | 5.80 | 6.08 | 2.25 | 2.54 | 4.81 | | | 8.4 | 7.11 |
| T11 | 7.56 | 1.67 | 6.06 | 2.27 | 2.58 | 4.89 | | 13.82 | | |
| C12 | 7.65 | 5.66 | 6.00 | 2.23 | 2.30 | 4.82 | 4.00 | | 8.4 | 7.11 |
| T13 | 7.50 | 1.65 | 5.98 | 2.15 | 2.48 | 4.87 | | 13.75 | | |
| C14 | 7.64 | 5.66 | 5.96 | 2.14 | 2.30 | 4.76 | | | 8.4 | 7.11 |
| T15 | 7.49 | 1.71 | 6.08 | 2.26 | 2.57 | 4.86 | | 14.04 | | |
| C16 | 7.64 | 5.78 | 6.25 | 2.24 | 2.36 | 4.56 | 3.99 | | 8.26 | 7.13 |

only been observed as a minor component of the sample under conditions where the duplex predominates (see Discussion).

A NOESY spectrum of the sample under the conditions shown in Figure 1D [excess $d(TC)_4$ strand and low pH] is given in Figure 3. Close examination of this spectrum shows a pattern of cross-peaks that is considerably different from those observed for the duplex. Assignments of the DNA conformation under these conditions cannot be obtained directly from this spectrum without making assumptions about the conformation. In addition to the NOE cross-peaks observed in this spectrum, some exchange cross-peaks between excess single-strand $d(TC)_4$ and a $d(TC)_4$ strand in the triplex are observed, most obviously near the diagonal in the aromatic and H1′ sugar regions.

*Assignments of the Nonexchangeable Proton Resonances of the Duplex.* Some expanded regions of the NOESY spectrum of the duplex shown in Figure 2 are given in Figure 4A,B. Figure 4A shows the region of cross-peaks between the thymine methyl and the aromatic resonances. Seven methyl aromatic resonances are observed, consisting of four intrabase TMe–TH6 cross-peaks (confirmed by DQF-COSY spectra, not shown) and three interbase TMe–CH6 cross-peaks. These latter are expected between the thymine methyl protons and their 5′ neighboring cytosines for a B-DNA-type helix (Feigon et al., 1982). Two additional methyl aromatic cross-peaks of lower intensity arise from some single-stranded $d(TC)_4$ in the sample. We note that although intrastrand cross-peaks are observed for the slight excess single-stranded $d(TC)_4$ in this sample, no exchange cross-peaks between the single strand and the duplex are observed in this spectrum. Figure 4B shows the region of cross-peaks between the aromatic and the H1′ sugar resonances. Sequential assignments of the base H8, H6, and H1′ sugar resonances were made following established procedures and are illustrated for the $d(GA)_4$ strand on the spectrum. Assignments of the CH5 and CH6 resonances were confirmed by cross-peaks in the COSY spectra (not shown). Sugar H2′, H2″, and H3′ resonances were assigned by NOESY from the H1′, H8, and H6 resonances and were confirmed by scalar connectivity in the DQF-COSY spectra. These assignments are summarized in Table I.

*Assignments of the Nonexchangeable Proton Resonances of the Triplex.* Expanded regions of the NOESY spectrum of the triplex in Figure 3, showing the region of cross-peaks between the aromatic and methyl resonances and between the aromatic and the sugar H1′ resonances, are given in Figure

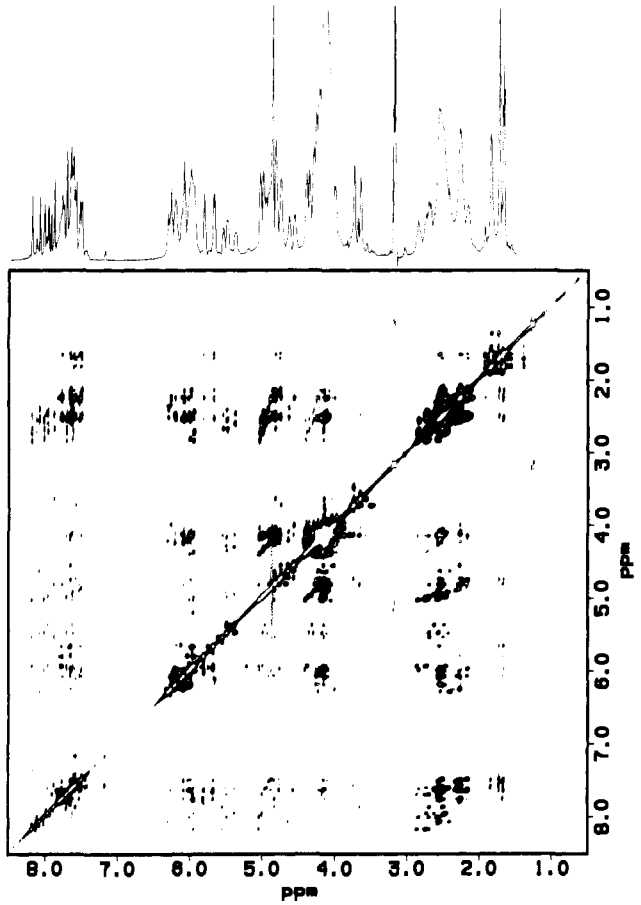



FIGURE 2: NOESY spectrum of 1:1.3 $d(GA)_4$:$d(TC)_4$ in 20 mM phosphate, pH 8.3, 50 mM NaCl, 15 °C. Sample contains 1 mM $d(GA)_4$ and 1.3 mM $d(TC)_4$ and is primarily duplex under these conditions, with the excess $d(TC)_4$ present as single strand. The spectrum was acquired with a sweep width of 5000 Hz in both dimensions, 2K complex points in $t_2$, and 64 scans per $t_1$ value, and 282 $t_1$ values were collected. Residual HDO was irradiated during the recycle delay of 2 s, and $\tau_m$ = 350 ms. The $t_1$ domain was zero-filled to 1K complex points prior to Fourier transformation, and the data in both dimensions were apodized by a sine-bell squared function phase shifted by 65°. The one-dimensional spectrum at the top was acquired with 8K complex points, 64 acquisitions, and irradiation of the residual HDO during the 2-s recycle delay. The impurity that gives rise to sharp lines at 3.2 ppm appeared in some samples after adjusting the pH and can be removed by gel filtration.

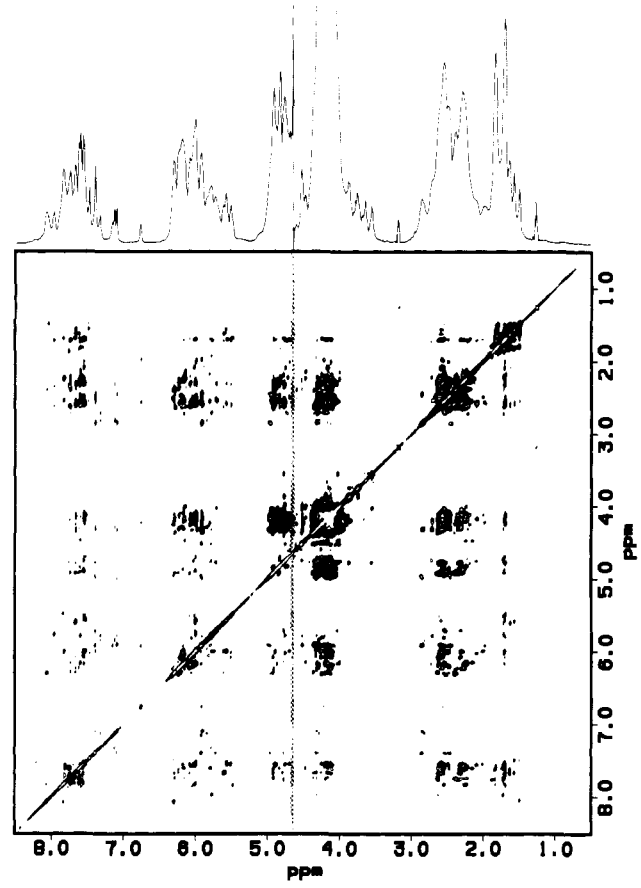



FIGURE 3: NOESY spectrum of 1:2.5 $d(GA)_4$:$d(TC)_4$ in 10 mM phosphate, pH 5.5, 100 mM NaCl, 5 mM $MgCl_2$, 35 °C. Sample contains 2 mM $d(GA)_4$ and 5 mM $d(TC)_4$ and is primarily triplex under these conditions with the excess $d(TC)_4$ present as single strand. The spectrum was acquired and processed as in Figure 2 except that 322 $t_1$ values were collected and a squared sine-bell function phase shifted by 40° was used in $t_2$ and 60° in $t_1$. The one-dimensional spectrum at the top was acquired as in Figure 2.

4C,D. Many more cross-peaks are observed in both regions of the spectrum shown compared to the duplex. These cross-peaks cannot be readily assigned or interpreted from these spectra alone. Fortunately, we were able to establish conditions where exchange cross-peaks between the putative triplex and duplex as well as single-stranded $d(TC)_4$ were observed in ROESY and NOESY spectra (Jeener et al., 1979). This allowed assignment of the triplex form without making any assumptions about the structure of the DNA (Feigon et al., 1984), provided that assignments for the duplex and single strand could be obtained. Assignments of the duplex were obtained as described above. Assignments of the single strand were made as to proton type (i.e., TH6, CH6, etc.) but could not be made sequentially along the strand because of the extreme spectral overlap in the single-strand resonances (spectra not shown). Portions of a ROESY spectrum of the sample under conditions where a mixture of duplex, triplex, and single strand $d(TC)_4$ are present are shown in Figure 5. Positive contours only are shown, which arise from the diagonal peaks and exchange cross-peaks. In the aromatic region, cross-peaks are observed between the Watson–Crick base paired $d(GA)_4$ and $d(TC)_4$ strands in the duplex and their corresponding strands in the triplex and between single-stranded $d(TC)_4$ and the second $d(TC)_4$ strand in the triplex. Sequence-specific assignments for the former $d(TC)_4$ and $d(GA)_4$ strands in the triplex were obtained from this spectrum, and these were confirmed by analysis of the exchange cross-peaks which also appear in NOESY spectra of this sample (not shown). The resonances for this $d(TC)_4$ strand in the triplex have chemical shifts that are very close to those in the duplex, and therefore the assignments made by exchange were not as straightforward as for the $d(GA)_4$ strand. These assignments were obtained both directly from aromatic–aromatic exchange cross-peaks and indirectly from assignments obtained from H1′–H1′ and H5–H5 exchange cross-peaks followed by NOEs to the H6 and H8. The aromatic assignments for the second $d(TC)_4$ strand in the triplex were made from the exchange cross-peaks observed between this strand and the single-stranded $d(TC)_4$ in this sample. Since only the assignments to base type rather than sequential assignments could be made for the single strand, the assignments of the aromatic protons of the second $d(TC)_4$ strand obtained from this spectrum are also only to base type. Sequence-specific assignments for the thymines in this strand were subsequently obtained from indirect NOEs observed in spectra of the triplex in water (see below). Assignments of the sugar H1′ resonances in the triplex were also made from exchange cross-peaks between the triplex and the duplex and single $d(TC)_4$ strand (Figure 5B). Since these resonances are also scalar coupled, some of the exchange cross-peaks could be obscured by

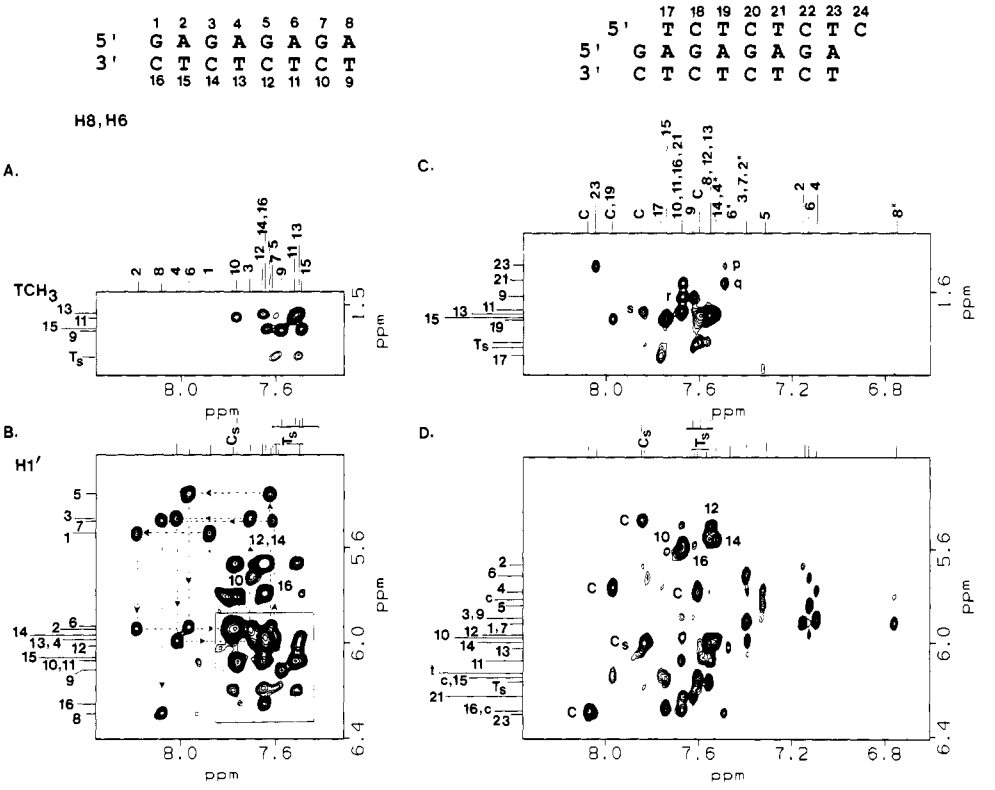



FIGURE 4: Expanded regions of the NOESY spectra of the duplex (A and B) and triplex (C and D) shown in Figures 2 and 3, respectively. The numbering systems used for the duplex and triplex are indicated at the top of the figure. (A) Region of cross-peaks between the thymine methyl and the aromatic proton resonances (Figure 2). (B) Region of cross-peaks between the H1′, CH5, and the aromatic resonances (Figure 2). Cross-peaks between CH5–CH6 are labeled. Sequential assignments of the purine strand are indicated by connecting lines. The cross-peaks arising from the pyrimidine strand are boxed. (C) Region of cross-peaks between the thymine methyl and the aromatic proton resonances (Figure 3). Cross-peaks arising from a second form of the triplex (see text) are labeled p–s. (D) Region of cross-peaks between the H1′, CH5, and the aromatic resonances (Figure 3). Cross-peaks between CH5–CH6 are labeled. Assignments of the TMe, H8, H6, and H1′ resonances are indicated along the sides of the spectra. Numbers with asterisks (*) are AH2 resonances.

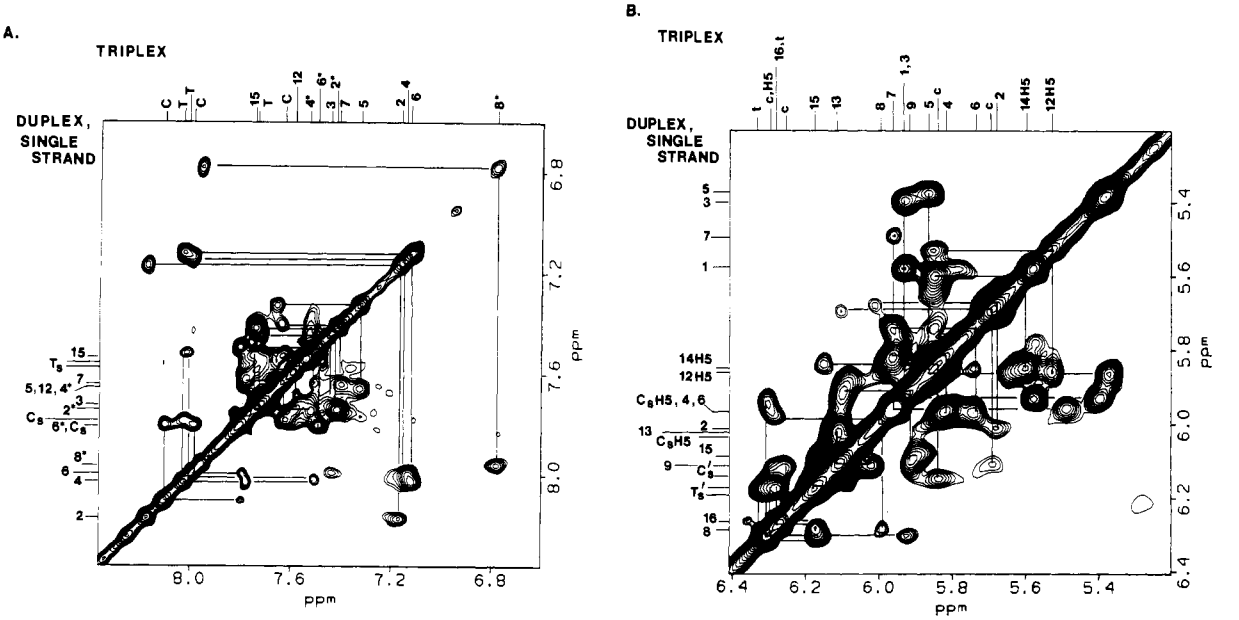



FIGURE 5: Portions of a ROESY spectrum of 1:1.25 $d(GA)_4$:$d(TC)_4$ in 20 mM phosphate, pH 5.5, 50 mM NaCl, 5 mM $MgCl_2$, 42 °C. Sample is 2 mM in $d(GA)_4$ and 2.5 mM in $d(TC)_4$. (A) The aromatic region of the spectrum, showing exchange cross-peaks between duplex and triplex and between triplex and single $d(TC)_4$ strand. (B) The H1′ region of the spectrum. Assignments of triplex and duplex resonances are indicated on the tops and sides of the spectra, respectively. $T_s$ indicates resonances from $d(TC)_4$ single strand, numbers with (*) are AH2 resonances, c and t are the Hoogsteen base paired pyrimidine strand H1′ resonances, and C and T are the Hoogsteen base paired pyrimidine strand H6 resonances. The spectrum was acquired with a 4-kHz spin-lock field applied during the mixing time of 116 ms. The sweep width was 5000 Hz in both dimensions, and 2K complex points were collected in $t_2$. A total of 259 $t_1$ values of 64 scans each were collected. Spectra were zero-filled to 1K in the $t_1$ dimension and apodized with a sine-bell squared function phase shifted by 65°.

overlapping TOCSY signals, although we did not find this to be a problem here. Assignments of the base and H1′ proton resonances of the triplex are summarized in Table II.

*Assignments of the Exchangeable and AH2 Resonances of the Duplex in $H_2O$.* A NOESY spectrum of the duplex under the conditions shown in Figure 2 in $H_2O$ at 1 °C was obtained

Table II: Assignments of the Base and H1′ Resonances in the Triplex $d(GA)_4 \cdot d(TC)_4 \cdot d(TC)_4$

| base | H8 | H2 | H1′ | base | H6 | CH5, TMe | H1′ | base | H6 | TMe |
|---|---|---|---|---|---|---|---|---|---|---|
| G1 | 7.73 | | 5.91 | T9 | 7.61 | 1.62 | 5.89 | T17 | 7.75 | 1.86 |
| A2 | 7.15 | 7.40 | 5.67 | C10 | 7.67 | 5.57 | 5.97 | C | (8.06)[a] | |
| G3 | 7.39 | | 5.91 | T11 | 7.67 | 1.69 | 6.07 | T19 | 7.97 | 1.61 |
| A4 | 7.09 | 7.52 | 5.78 | C12 | 7.55 | 5.50 | 5.50 | C | (7.97)[a] | |
| G5 | 7.32 | | 5.84 | T13 | 7.55 | 1.70 | 5.99 | T21 | 7.67 | 1.56 |
| A6 | 7.12 | 7.48 | 5.71 | C14 | 7.53 | 5.56 | 6.00 | C | (7.84)[a] | |
| G7 | 7.38 | | 5.99 | T15 | 7.74 | 1.72 | 6.14 | T23 | 8.04 | 1.47 |
| A8 | 7.56 | 6.77 | | C16 | 7.67 | 5.61 | 6.23 | C | (7.60)[a] | |

[a] Sequential assignment not determined.

(not shown). At this temperature all eight of the imino proton resonances expected from four Watson–Crick AT (13.6–14.1 ppm) and four Watson–Crick GC (12.2–12.8 ppm) base pairs were observed. These were assigned by sequential connectivities between the AT and GC imino resonances along the strand for the internal base pairs and by melting behavior for the terminal base pairs. Assignments to the AH2, A amino, and C amino resonances were made by NOE cross-peaks observed to the iminos and are given in Table I.

*Assignments of the Exchangeable Resonances of the Triplex in $H_2O$ and Sequence-Specific Assignments of the Thymines in the Second $d(TC)_4$ Strand.* Portions of NOESY spectra of the sample under conditions where the predominant conformation is a triplex are shown in Figure 6 at 1 and 30 °C. As illustrated in the one-dimensional spectra in Figure 1, several additional imino resonances are present in the triplex compared to the duplex, and these give rise to a number of additional imino–imino cross-peaks (Figure 6B). The imino resonances also show cross-peaks to the amino and the AH2, AH8, and GH8 resonances (Figure 6A,C). AT imino cross-peaks to AH2 resonances are expected for Watson–Crick base pairs (Sânchez et al., 1980), and these were identified for four AT base pairs in the triplex. Imino–imino cross-peaks between these identified AT Watson–Crick base pairs and neighboring GC base pairs were used to assign imino resonances sequentially along the Watson–Crick strands (boxed region in Figure 6B). These are the base pairs that show exchange cross-peaks to the Watson–Crick base pairs in the duplex in samples containing a mixture of duplex and triplex (Figure 5). A second set of imino resonances are identified as arising from Hoogsteen base pairs on the basis of the strong cross-peaks observed between these imino resonances and the GH8 and AH8 resonances (Figure 6A,C) (Gilbert et al., 1989). These cross-peaks arise from Hoogsteen base pairs between the second $d(TC)_4$ strand (identified by exchange spectroscopy with the single strand in the $D_2O$ spectra) and the $d(GA)_4$ strand.

As the temperature is raised, imino resonances arising from the Hoogsteen $GC^+$ base pairs begin to broaden and disappear while the Hoogsteen AT imino resonances remain relatively sharp. In the NOESY spectra in water many of the cross-peaks arising from Hoogsteen $GC^+$ base pairs have disappeared by 30 °C (Figure 6C), thus simplifying the analysis of the spectrum and providing confirmation of the assignments made in the less well resolved 1 °C NOESY spectrum.

For these assignments, and for identification of Watson–Crick versus Hoogsteen base pairs, it was essential that the AH2 resonances be unambiguously distinguished from the AH8 and GH8 resonances. For standard B-DNA, AH2 resonances are normally assigned on the basis of their strong cross-peaks to AT imino resonances and only occasionally show other cross-peaks to neighboring AH2 resonances or H1′ resonances. However, if the base pairs adopt the Hoogsteen conformation, then a strong cross-peak between the imino proton of the base pair and the corresponding AH8 or GH8 resonance is expected instead (Gilbert et al., 1989). The AH8, GH8, and AH2 resonances were assigned in the triplex by exchange spectroscopy with the duplex, for which straightforward assignments could be obtained as discussed above. However, because some spectral overlap is present, especially in the water spectra, the assignments to proton type were confirmed by preparing a sample in which the AH8 and GH8 resonances had been replaced by deuterons (see Materials and Methods). Spectra of the deuterated triplex in $D_2O$ and $H_2O$ were obtained (not shown), and they confirmed that these assignments were correct. Therefore, although there is some overlap between H8 and AH2 resonances in the NOESY spectra in $H_2O$, there is no ambiguity in the assignment of imino–H8 versus imino–AH2 (or CH4) cross-peaks.

Table III: Assignments of the Exchangeable Resonances from the Watson–Crick and Hoogsteen Base Pairs in the Triplex $d(GA)_4 \cdot d(TC)_4 \cdot d(TC)_4$[a]

| | iminos | | iminos | aminos | |
|---|---|---|---|---|---|
| base pair | GH1, TH3 | base pair | $C^+$H1, TH3 | CH4(2) | CH4(1) |
| $G_1 \cdot C_{16}$ | 12.43 | | | | |
| $A_2 \cdot T_{15}$ | 14.16 | $A_2 \cdot T_{17}$ | 13.12 | | |
| $G_3 \cdot C_{14}$ | 12.67 | $G_3 \cdot C_{18}$ | 15.66 | 9.97 | 9.01 |
| $A_4 \cdot T_{13}$ | 14.22 | $A_4 \cdot T_{19}$ | 13.69 | | |
| $G_5 \cdot C_{12}$ | 12.67 | $G_5 \cdot C_{20}$ | 14.88 | 9.89 | 8.88 |
| $A_6 \cdot T_{11}$ | 14.29 | $A_6 \cdot T_{21}$ | 13.21 | | |
| $G_7 \cdot C_{10}$ | 12.71 | $G_7 \cdot C_{22}$ | 15.08 | 9.66 | 9.43 |
| $A_8 \cdot T_9$ | 13.15 | $A_8 \cdot T_{23}$ | 13.58 | | |

[a] A aminos resonate in two groups at 7.29 and 7.72 ppm. Watson–Crick base paired C aminos are at 8.31, 8.18, and 8.00 ppm (hydrogen bonded) and 7.13, 7.0, and 6.89 (non hydrogen bonded).

AT Hoogsteen base pairs were distinguished from GC Hoogsteen base pairs on the basis of their cross-peaks to previously assigned AH8 and GH8 resonances, respectively. In addition, the imino protons in $GC^+$ base pairs show unusually low-field chemical shifts. A set of imino–imino cross-peaks connecting the Hoogsteen AT and GC base pairs was observed and was used to assign these sequentially along the strand (Figure 6B). Additional amino resonances are also observed in the spectra of the triplex in $H_2O$ (Rajagopal & Feigon, 1989). Hoogsteen base pairing of the second $d(TC)_4$ strand to the Watson–Crick base paired $d(GA)_4$ strand results in hydrogen bonding of both of the A amino and an additional C amino proton per base triplet (see Discussion). Cross-peaks are observed between the Hoogsteen AT imino and the A amino resonances and between the Hoogsteen $GC^+$ imino and the hydrogen- and non-hydrogen-bonded $C^+$ amino resonances. The $C^+$ amino resonances appear between 8.5 and 10 ppm and show cross-peaks to the Hoogsteen $GC^+$ imino resonances between 15.2 and 15.6 ppm, confirming their assignment to $GC^+$ base pairs. Assignments of the exchangeable resonances of the triplex are given in Table III.

Up to this point, base resonance assignments in the second $d(TC)_4$ strand had only been made to proton type. However,

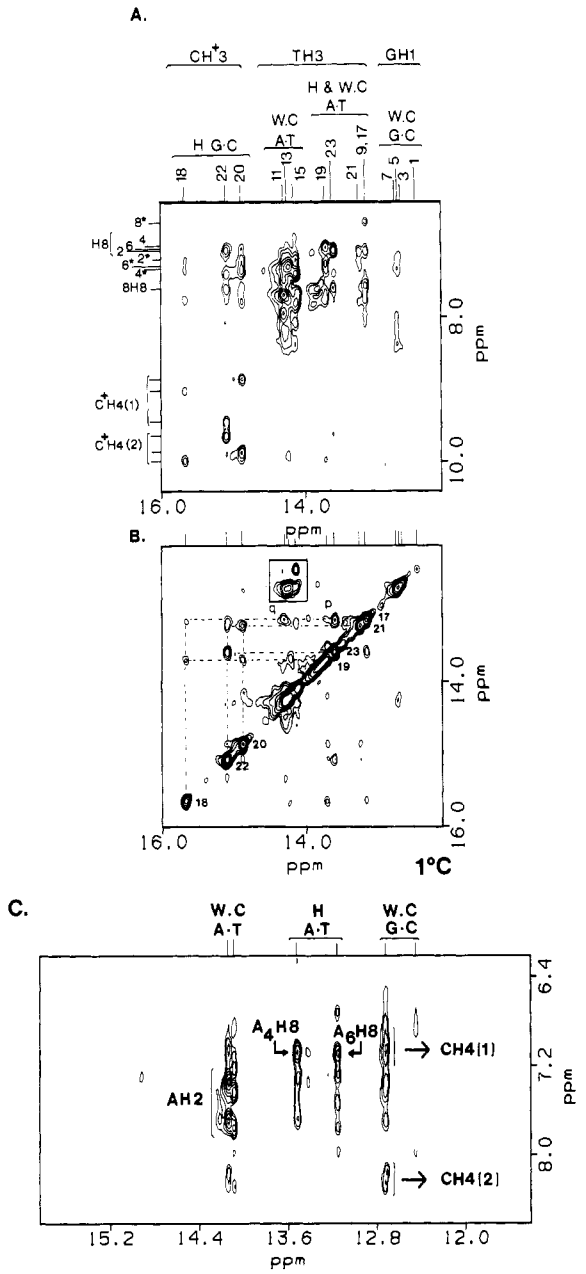



FIGURE 6: Portions of NOESY spectra of the triplex in $H_2O$ (90% $H_2O$:10% $D_2O$) at 1 °C (A and B) and 30 °C (C). (A) Region containing cross-peaks between the imino and the aromatic and amino resonances. Assignments of the imino, H8, AH2 [indicated by (*)], and C amino resonances are indicated. (B) Region containing the imino proton resonances and their cross-peaks. Sequential assignments for the Hoogsteen base paired iminos are indicated by the dotted lines. Cross-peaks between Watson–Crick base paired iminos are boxed. Some exchange cross-peaks between the two triplex forms (see text) are labeled p and q. (C) Same region of the spectrum as (A) at 30 °C. Cross-peaks between the Watson–Crick AT imino and AH2, the Hoogsteen AT imino and AH8, and the Watson–Crick GC imino and C amino resonances are indicated. Cross-peaks from Hoogsteen $GC^+$ resonances are absent at this temperature. The sample for (A) and (B) is 1:2.5 $d(GA)_4$:$d(TC)_4$, 10 mM phosphate, pH 5.5, 25 mM NaCl. The sample for (C) is the same with 100 mM NaCl, 5 mM $MgCl_2$. Spectra were acquired as described with $\tau_m$ = 150 ms, and 215 $t_1$ values were collected. The spectrum in (A) and (B) was apodized by a squared sine-bell function (skew = 1.6) and phase shifted by 85° in both dimensions. The spectrum in (C) was apodized in $t_1$ and $t_2$ by 60° and 45° phase-shifted sine-bell squared functions, respectively. The $t_1$ dimensions were zero-filled to 2K complex points prior to Fourier transformation.

assignment of the Hoogsteen AT imino protons allowed assignment of the thymine methyl protons in the second $d(TC)_4$ strand by indirect NOEs from imino to the methyl protons via the A amino protons. The thymine H6 resonances were then assigned by observed TMe–TH6 cross-peaks. Unfortunately, spectral overlap precluded similar sequence-specific assignment of the CH6 and CH5 resonances in the second $d(TC)_4$ strand.

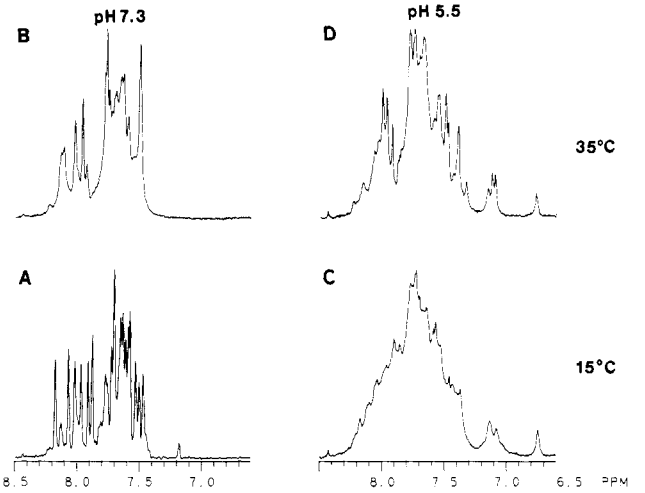



FIGURE 7: Aromatic proton spectra of 1:1 $d(GA)_4$:$d(TC)_4$ as a function of pH and temperature. (A) pH 7.3, 15 °C; (B) pH 7.3, 35 °C; (C) pH 5.5, 15 °C; (D) pH 5.5, 35 °C. Samples contained 20 mM phosphate, 50 mM NaCl, and 5 mM $MgCl_2$. Spectra were resolution enhanced with double-exponential multiplication (8 Hz). Spectra in (A) and (B) are primarily duplex, and those in (C) and (D), primarily triplex plus single-stranded $d(GA)_4$.

*Effect of pH and Temperature on Duplex and Triplex Formation.* Figure 7 shows the aromatic region of the spectra obtained for 1:1 $d(GA)_4$:$d(TC)_4$ at pH 7.3 and pH 5.5 at 15 and 35 °C. Figure 7A,B is the same sample as Figure 1C. At 15 °C and pH 7.3, the sample is primarily duplex. As the temperature is raised to 35 °C, the DNA begins to melt, and the sample is partially single stranded at this temperature. The pH of this sample was adjusted to pH 5.5 by addition of HCl, and the spectra obtained are shown in Figure 7C,D. New resonances corresponding to the triplex are evident in these spectra. At low temperature, the resonance lines are not well resolved, probably due to sample aggregation. In contrast to the duplex, the triplex is stable at 35 °C and exhibits much better resolved resonances than at 15 °C. Our analysis of this sample (NOESY, ROESY, and COSY spectra not shown) showed that under these conditions the sample contains a mixture of duplex and triplex as well as some single-stranded $d(GA)_4$. Triplex formation is also observed in samples at pH 6.1, but analysis of the spectra is complicated by additional exchange cross-peaks between a small amount of the unknown heterocomplex and duplex DNA (see Discussion). Spectra for 1:1 $d(GA)_4$:$d(TC)_4$ at pH 8.2 (Figure 9D) are essentially the same as those observed at pH 7.3.

*Effect of Strand Concentration on Triplex Formation at Low and High pH.* Although triplex formation is favored by low pH, the equilibrium can be pushed toward the duplex form in the presence of excess $d(GA)_4$ strand. Figure 8 shows the aromatic and methyl regions of the spectra obtained at pH 5.5 at 50 °C for samples containing 40 mM $MgCl_2$. The spectrum at bottom has 3:1 $d(GA)_4$:$d(TC)_4$ and is predominantly duplex plus single-stranded $d(GA)_4$. At lower temperatures, some triplex is also present (spectrum not shown). The high concentration of $Mg^{2+}$ in this sample is necessary to stabilize the duplex at 50 °C; in the absence of or at low levels of $Mg^{2+}$ the duplex melts below this temperature (Figure 7B). At higher relative $d(TC)_4$:$d(GA)_4$ ratios, the triplex predominates. Figure 8B shows the spectrum obtained at 1:1.4

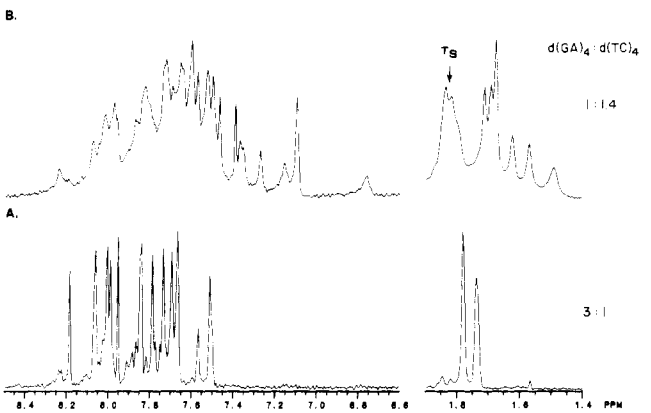



FIGURE 8: Aromatic and methyl proton spectra of $d(GA)_4 + d(TC)_4$ as a function of relative strand concentration at pH 5.5 and 50 °C. (A) 3:1 $d(GA)_4$:$d(TC)_4$; (B) 1:1.4 $d(GA)_4$:$d(TC)_4$. Samples contain 20 mM phosphate, pH 5.5, 225 mM NaCl, and 40 mM $MgCl_2$. Spectra are line-broadened by 1 Hz.

$d(GA)_4$:$d(TC)_4$ under the same conditions. Triplex, $d(TC)_4$ single strand, duplex, and $d(GA)_4$ single strand are present, with the former two predominating.

At pH 8.2, no triplex is observed at any relative strand concentration studied (Figure 9). Titration of $d(TC)_4$ to a sample containing $d(GA)_4$ at 15 °C and pH 8.2 leads to formation of duplex DNA, plus $d(TC)_4$ single strand at d-$(TC)_4$:$d(GA)_4$ ratios of greater than 1 (Figure 9C–F). Reference spectra of samples containing only $d(TC)_4$ and $d(GA)_4$ are also shown (Figure 9A–C). The temperature dependence of the $d(GA)_4$ spectra is discussed below (see Discussion).

## Discussion

*Duplex and Triplex Formation.* The homopurine $d(GA)_4$ and homopyrimidine $d(TC)_4$ strands combine to form at least two major conformations depending on relative strand concentration and pH. The major heterocomplex present at neutral pH and above, under all conditions studied, is a DNA duplex which is readily identifiable as a B-DNA-type duplex on the basis of the NOESY spectra in $D_2O$ and $H_2O$. At lower pH a second major complex is formed which dominates in the presence of excess $d(TC)_4$ strand. This second form is a triplex composed of two homopyrimidine and one homopurine strands. The relevant equilibrium is as follows: $2\ d(GA)_4 \cdot d(TC)_4 \leftrightarrow d(TC)_4 \cdot d(GA)_4 \cdot d(TC)_4 + d(GA)_4$.

*Exchange between Duplex and Triplex.* Assignments of the resonances in the triplex were made by exchange spectroscopy (Jeener et al., 1979) under conditions where the sample contained a mixture of duplex, triplex, and single $d(TC)_4$ strand. Two separate sets of exchange cross-peaks were observed. The duplex containing one $d(TC)_4$ and one $d(GA)_4$ strand showed exchange cross-peaks to corresponding strands in the triplex. The second $d(TC)_4$ strand in the triplex showed exchange cross-peaks only to the single-stranded d-$(TC)_4$. This indicates that triplex formation on addition of $d(TC)_4$ to a solution containing duplex DNA occurs by binding of the second $d(TC)_4$ strand to the duplex without dissociation of the duplex.

*Conformation of the Duplex.* Qualitative analysis of the NOESY spectra obtained on $d(GA)_4 \cdot d(TC)_4$ indicates that the duplex forms a right-handed B-DNA-type structure. The unusually low melting temperature for this sequence as compared to other non-homopurine–homopyrimidine oligonucleotides that we have looked at (Feigon et al., 1984 and unpublished results) indicates, however, that there may be something unusual about the structure of this molecule.

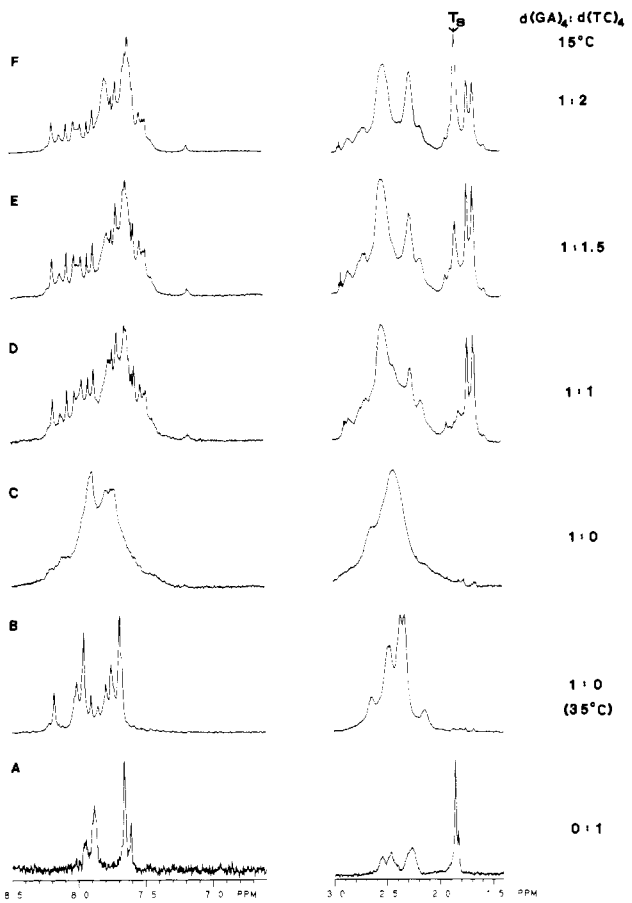



FIGURE 9: Aromatic and methyl proton spectra of $d(GA)_4$ and $d(TC)_4$ as a function of relative strand concentration at pH 8.2 at 15 °C [except (A) and (B)]. (A) $d(TC)_4$ at 10 °C; (B) $d(GA)_4$ at 35 °C; (C) $d(GA)_4$; (D) 1:1 $d(GA)_4$:$d(TC)_4$; (E) 1:1.5 $d(GA)_4$:$d(TC)_4$; (F) 1:2 $d(GA)_4$:$d(TC)_4$. Samples (B)–(F) contain 2 mM $d(GA)_4$, 20 mM phosphate, pH 8.2, 50 mM NaCl, 5 mM $MgCl_2$. Appropriate amounts of $d(TC)_4$ were titrated into sample (B) to obtain the conditions of (D)–(F).

*Watson–Crick Base Paired Strands in the Triplex.* The $d(GA)_4$ and $d(TC)_4$ strands in the triplex that show exchange cross-peaks in the ROESY spectra in $D_2O$ to the strands in the duplex (Figure 5) also show cross-peaks in the NOESY spectra in water that indicate that these strands are Watson–Crick base paired (Figure 6). These strands also give a pattern of cross-peaks in the NOESY spectra in $D_2O$ that are indicative of a right-handed A-DNA- or B-DNA-type helix. Cross-peaks are observed in the methyl region between each of the thymine methyl resonances from the $d(TC)_4$ strand to the cytosines on their 5′ side, similar to the pattern observed for the B-DNA duplex (Figure 4C). However, sequential connectivities from base to H1′ sugar to base could not be made from NOESY spectra of the triplex since not all bases showed sequential base–H1′ NOEs even at long mixing times.

*Hoogsteen Base Pairs from the Second $d(TC)_4$ Strand in the Triplex.* The second $d(TC)_4$ strand in the triplex, which shows exchange cross-peaks to the single-stranded $d(TC)_4$, gives rise to exchangeable imino and amino resonances which for the most part appear at somewhat different chemical shifts from those normally observed in Watson–Crick base paired DNA. These imino resonances show cross-peaks in the NOESY spectra in $H_2O$ that indicate that this strand is Hoogsteen base paired to the $d(GA)_4$ strand. This places the second $d(TC)_4$ strand in what would be the major groove of A- or B-DNA and results in the formation of base triplets (Chart I). This creates a second series of imino protons which

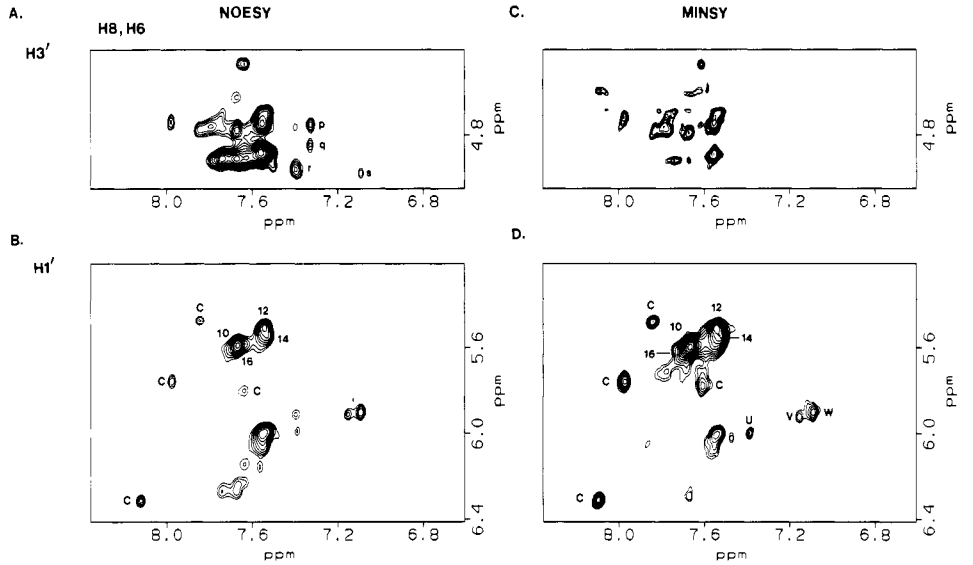
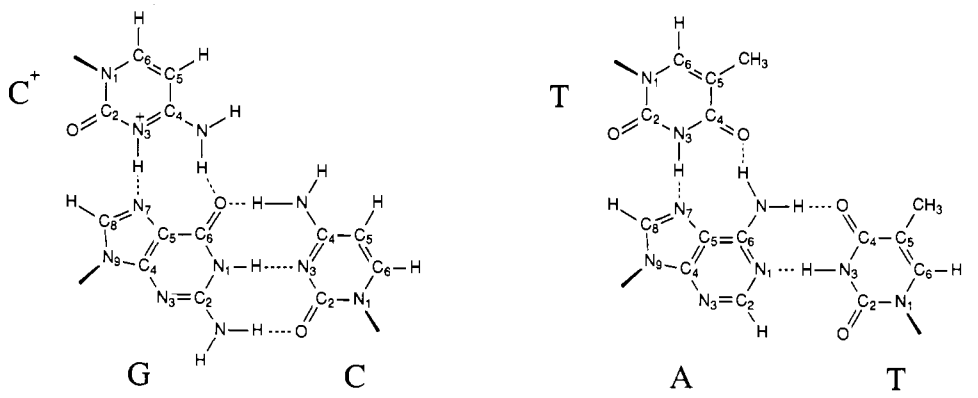



FIGURE 10: Portions of NOESY (A and B) and MINSY (C and D) spectra of the triplex at 35 °C. Regions of cross-peaks between the aromatic and H3′ protons are shown in (A) and (C); regions of cross-peaks between the aromatic and H1′, CH5 protons are shown in (B) and (D). In (A) GH8–H3′ cross-peaks are labeled p–r and AH8–H3′ cross-peaks is labeled s. In (B) and (D) CH5–CH6 cross-peaks from the Watson–Crick base paired strands are indicated by number and CH5–CH6 cross-peaks for the Hoogsteen base paired $d(TC)_4$ strand are indicated by C. In (D) additional cross-peaks are labeled as follows: (u) $A_2H2–C_{14}H1'$, (v) $A_6H8–G_5H1'$, and (w) $A_4H8–G_3H1'$. The sample contained 1:2 $d(GA)_4$:$d(TC)_4$, 20 mM phosphate, pH 5.5, 50 mM NaCl, and 5 mM $MgCl_2$. The control NOESY spectrum was acquired with a $\tau_m$ = 200 ms. The MINSY spectrum was acquired under the same conditions as the NOESY except that the region of the spectrum containing the H2′,H2″ resonances was completely saturated during the mixing time and residual HDO was not irradiated during the recycle delay. 2K complex points were collected with 64 scans per $t_1$ value, and 261 $t_1$ values were collected. The spectra were zero-filled in $t_1$ and apodized with a squared sine bell phase shifted by 90° in both dimensions.

Chart I




should give rise to sequential connectivities in the NOESY spectra in water. These were observed and were used to confirm the assignments of the Hoogsteen base pairs in the triplex.

Hoogsteen base pairing of CG bases requires protonation of the cytosine, as indicated in Chart I. Although this protonation has been assumed in models for triplexes containing $CGC^+$ base triplets, the three $C^+$ imino resonances observed in the spectra of the triplex represent the first direct observation of the protonated dC in DNA triplexes (Rajagopal & Feigon, 1989). These imino resonances occur about 3 ppm downfield from the usual Watson–Crick base paired G·C iminos (Kearns, 1979). Kallenbach et al. (1976) also observed resonance intensity at ~16 ppm in complexes of oligo-$C_{17}$ with GMP at pH 5.8 which they attributed to a protonated C imino in $CGC^+$ triplets. The $C^+$ aminos also show large (~2 ppm) downfield shifts from their resonance positions in Watson–Crick base pairs and appear between 8.5 and 10 ppm.

*Conformation of the Purine and Pyrimidine Sugars and Bases in the Triplex.* The NOESY spectra shown in Figures 3 and 4 are useful for assignment purposes and for qualitative information about the DNA structure, but interproton distance information derived from cross-peak intensities will not be accurate due to spin diffusion at the mixing times used. Therefore, NOESY spectra were obtained at several shorter mixing times (75, 150, and 200 ms). In addition, MINSY spectra were obtained (Massefski & Redfield, 1988). Figure 10 shows portions of a MINSY spectrum of the triplex in which the H2′,H2″ resonances have been selectively saturated during the mixing time of 200 ms. A control NOESY spectrum at the same mixing time is shown at left. Weak cross-peaks between the base and the H1′ sugar resonances are normally observed for DNAs in which the bases adopt an anti conformation (A- and B-DNA) (Feigon et al., 1982). These can arise partially from spin diffusion via H2′,H2″ protons which are much closer to the base H8 and H6 protons than are the H1′ when the bases are in the anti conformation. Observation of weak or no base–H1′ cross-peaks in the NOESY and MINSY spectra indicates that all of the bases in the triplex are in the anti conformation.

Saturation of the H2′,H2″ resonances during the mixing time of the MINSY experiment results in observation of direct rather than indirect (spin diffusion) NOEs between the base and the H1′ and H3′ resonances. Figure 10A shows cross-

peaks between the aromatic and H3′ resonances that appear in the control NOESY spectrum. Cross-peaks are observed between all of the H6 and some of the H8 resonances and one or more H3′ resonances. For B-DNA with average C2′-endo (S-type) sugar conformation, intranucleotide H6,H8–H3′ distances are expected to be greater than 4.0 Å and internucleotide H6,H8–H3′ distances are about 4.6 Å. In contrast, for A-DNA with C3′-endo (N-type) sugar conformation, the intranucleotide distances are about 2.6 Å and the internucleotide distances are about 3.0 Å (Wüthrich, 1986). These distances depend on both sugar pucker and glycosidic angle. In the MINSY experiment (Figure 10C), which shows only direct NOEs between the base H6,H8 and the H3′ resonances, the cross-peaks from the AH8 and GH8 resonances have disappeared (labeled p–r in Figure 10A) while those from the CH6 and TH6 in the two $d(TC)_4$ strands remain. Since all of the bases in the triplex are anti, this suggests that the pyrimidine sugars in both $d(TC)_4$ strands adopt an A-DNA-like (N-type) sugar conformation, while the purine sugars remain in or close to an S-type conformation. We note that we were unable to obtain coupling constants, which would also give information on the sugar pucker, from the DQF-COSY spectra due to spectral overlap.

Panels B and D of Figure 10 show the region of cross-peaks between the aromatic and the H1′ sugar and CH5 resonances in the control NOESY and the MINSY spectrum, respectively. In both the NOESY and the MINSY spectra, four strong cross-peaks are observed from the four CH5–CH6 in the Watson–Crick base paired $d(TC)_4$ strand, and four weaker CH5–CH6 cross-peaks are observed for these resonances in the Hoogsteen base paired $d(TC)_4$ strand. Since these NOEs arise from protons that have a fixed interproton distance, the difference in CH5–CH6 cross-peaks intensities for the two strands must be a result of strand exchange between the Hoogsteen base paired $d(TC)_4$ strand and free $d(TC)_4$ strand during the mixing time of the experiments. The remaining cross-peaks in the NOESY spectrum in Figure 10B,D almost all arise from H6,H8–H1′ interactions. Most of these occur as overlapping peaks in the region between 7.5 and 7.8 ppm and cannot be completely resolved and assigned. Again, the observed H6,H8–H1′ cross-peaks in the MINSY spectrum can only arise from direct NOEs. Intrabase distances between H8–H1′ and H6–H1′ are not very different in A- and B-DNA and fall in the range of 3.6–3.8 and 3.4–3.6 Å, respectively. In contrast, interbase H8–H1′ and H6–H1′ distances are expected to be significantly different for A- and B-DNA. For A-DNA the $H8_n$–$H1'_{n-1}$ and $H6_n$–$H1'_{n-1}$ are approximately 4.6 Å, while the corresponding distances for B-DNA are 3.6 and 3.5 Å, respectively (Wüthrich, 1986). Resolved, cross-peaks are seen for $A_6$ and $A_4H8_n$–$H1'_{n-1}$, and these cross-peaks remain in the MINSY spectrum (labeled v and w in Figure 10D). This is consistent with the sugars in the purine strand having a B-DNA-like (S-type) conformation.

Models for triplexes formed from poly(U)·poly(A)·poly(U) and poly(dT)·poly(dA)·poly(dT) have been proposed based on X-ray fiber diffraction (Arnott & Selsing, 1974). The diffraction data suggested an A′-DNA conformation of the two Watson–Crick base paired strands with the third strand Hoogsteen base paired in the major groove in the same conformation. A′-RNA has a pitch of 36°, resulting in 12 nucleotides per turn, a 10° base pair tilt angle, a 3.0-Å rise per residue, C3′-endo sugars, and the base pairs being displaced 4.4 Å from the helix axis. This structure has a very deep major groove which should be able to accommodate a third strand. Our NMR data indicate that both of the homopyrimidine strands have C3′-endo sugars, but the purine strand appears to be in a different conformation from that predicted by the fiber diffraction data. It may be that triplexes formed from mixed-sequence homopurine–homopyrimidine strands have a different conformation than those containing only dA and dT. Interestingly, Moser and Dervan (1988) found that ethylene glycol, which should help induce an A-DNA conformation, caused enhanced cleavage rates in a dT·dA·dT triplex with EDTA–Fe cleaving moieties but did not increase the cleavage rates for a triplex containing mixed C and T in the complementary third strand.

Chart II

```
      5'    17 18 19 20 21 22 23 24
            T  C  T  C  T  C  T  C
         1  2  3  4  5  6  7  8
5'       G  A  G  A  G  A  G  A
3'       C  T  C  T  C  T  C  T
        16 15 14 13 12 11 10  9
```

```
      5'    17 18 19 20 21 22 23 24
            T  C  T  C  T  C  T  C
5'       G  A  G  A  G  A  G  A
3'       C  T  C  T  C  T  C  T
```

*Orientation of the Second $d(TC)_4$ Strand.* Models for pyr–pur–pyr triplexes proposed on the basis of steric considerations and the fiber diffraction data have the Hoogsteen base paired homopyrimidine strand oriented parallel to the homopurine strand. This orientation has been supported by the studies of Moser and Dervan (1988) in which several homopyrimidine strands of different sequence with an attached EDTA–Fe DNA cleaving moiety were shown to bind to the complementary double-stranded DNA and gave cleavage patterns that could only result from a parallel orientation to the homopurine sequence in the DNA duplex.

For the DNA oligomers used in this study, an antiparallel orientation of the Hoogsteen base paired homopyrimidine strand relative to the homopurine strand would result in eight base triplets, while a parallel orientation would result in only seven base triplets. For the latter case, the triplex would contain either four TAT and three $CGC^+$ (Chart II) or three TAT and four $CGC^+$ triplets. Analysis of the cross-peaks from the imino proton resonances in the NOESY spectra of the triplex (Figure 6B) shows a set of sequential connectivities between Hoogsteen AT and $GC^+$ base pairs along the helix which begin and end with AT. Thus, the predominant triplex contains four TAT and three $CGC^+$ base pairs. Examination of the imino proton spectra as a function of temperature may give some insight into why the triplex with a 3′ dangling end is the preferred conformation. At 1 °C imino proton resonances from all 15 hydrogen-bonded imino protons in the triplex can be seen. However, as the temperature is raised, the Hoogsteen base paired $GC^+$ imino resonances begin to broaden and disappear long before the Hoogsteen AT, Watson–Crick AT (except terminal AT), and Watson–Crick GC base paired imino resonances do (not shown). Thus, imino protons from Hoogsteen $GC^+$ base pairs in the triplex exchange much more rapidly with water than do Hoogsteen AT or Watson–Crick AT and GC imino protons. This faster exchange for Hoogsteen $GC^+$ base pairs may reflect an inherently lower stability for these base pairs versus Hoogsteen AT base pairs due to the requirement for protonation of the C for Hoogsteen $GC^+$ base pairs to form. While this explanation is consistent with predominant formation of the triplex with only three $GC^+$ Hoogsteen base pairs (3′ dangling end), further experiments need to be done to see if the faster exchange is really due to a shorter lifetime for these base pairs or only a faster intrinsic catalysis rate (Gueron et al., 1987).

While the predominant triplex formed has four TAT and three $CGC^+$ base triplets, there is evidence in the spectra for a minor conformer of the triplex with three TAT and four $CGC^+$ base triplets. The imino resonances in the NOESY spectrum of the triplex (Figure 6B) show some cross-peaks

which are attributed to exchange between the two triplex forms. These are identified as Hoogsteen $A_2 \cdot T_{17}$ (p) and $A_4 \cdot T_{19}$ (p) and Watson–Crick $A_8 \cdot T_9$ (q) base pairs in the two conformers of the triplex. Small cross-peaks from the minor conformer of the triplex can be seen in the methyl aromatic region of the NOESY spectrum of the triplex in Figure 4C. Resonances from the minor conformer of the triplex appear mostly as shoulders on the major resonances in resolution-enhanced spectra of the triplex. We estimate that the minor conformer is about 10% of the total triplex. Thus, two triplexes of the type illustrated in Chart II appear in equilibrium with each other, with the triplex with four TAT and three $CGC^+$ base triplets predominating and "slippage" occurring between the two forms during the mixing time of the NOESY experiments.

*Other Conformations.* In the aromatic region of the spectrum of the duplex shown in Figure 1C, some small resonances appear (e.g., 7.17 and 8.13 ppm) that do not arise from the duplex or either form of the triplex discussed above or from either of the single strands. These small resonances show exchange cross-peaks with both strands of the duplex in NOESY spectra of the duplex (Figure 2). These resonances appear at temperatures up to 35 °C at pHs from at least 6.1 to 8.3 and in the absence or presence of magnesium. In samples at pH 6.1 with excess $d(TC)_4$ strand, duplex, triplex, and this heterocomplex are all present and exchange cross-peaks are observed between duplex and the heterocomplex but not between the duplex and the triplex (spectra not shown). At pH 5.5 these resonances and exchange cross-peaks are no longer observed (Figures 3 and 10) regardless of strand concentration. The fact that exchange cross-peaks are observed to both strands of the duplex indicates that this heterocomplex must be composed of at least one $d(TC)_4$ and one $d(GA)_4$ strand. Fresco and colleagues have presented evidence for pyr–pur–pur triplex formation at neutral pH from poly[d(T)]·poly[d(A)]·poly[d(A)] (Broitman et al., 1987) and poly[d(C)]·poly[d(G)]·polyd(G) (Letai et al., 1988). Recently, Cooney et al. (1988) presented evidence for triplex formation at high pH from binding of a *purine*-rich 27 base long DNA sequence to the complementary DNA at a site within the 5′ end of the human c-*myc* oncogene. We therefore investigated the possibility that this unknown heterocomplex might be a pyr–pur–pur triplex. However, addition of excess $d(GA)_4$ strand did not significantly affect the intensity of the resonances from this third complex under any of the above conditions (Figure 9 and spectra not shown). We conclude that there is no evidence for a pyr–pur–pur triplex for this sequence under any of the conditions that we have studied.

Spectra of the aromatic and methyl resonances of each of the single strands and a titration of $d(TC)_4$ to $d(GA)_4$ at pH 8.2 are shown in Figure 9. At 15 °C, the $d(TC)_4$ strand shows overlapping sharp resonances expected for a single strand. At 35 °C, the $d(GA)_4$ strand also shows partially resolved sharp resonances. However, at 15 °C the spectrum of the $d(GA)_4$ becomes very broad, indicating some specific self-complexation or else aggregation. We note that one of the obvious resonances from the third complex at 7.17 ppm does not appear in this spectrum; this is expected since that resonance shows exchange cross-peaks to the $d(TC)_4$ strand of the duplex. As $d(TC)_4$ is titrated at 15 °C, resonances corresponding to the duplex appear, as well as resonances from the third complex. At $d(GA)_4$:$d(TC)_4$ ratios above 1:1, single-stranded $d(TC)_4$ is present with the duplex, as evident from the resonance intensity from the single-stranded thymine methyls. No triplex formation is observed at this pH, and the amount of the third complex remains approximately the same at all $d(TC)_4$ concentrations.

The spectrum of $d(GA)_4$ at 15 °C (Figure 9C) suggests that this oligomer may form a specific self-complex. Sen and Gilbert (1988) have recently reported that single-stranded DNA oligomers containing runs of guanines will self-associate to make four-stranded structures. Their model has the four strands in parallel orientation with the guanines Hoogsteen paired to each other. These four-stranded structures also form stable complexes with their complementary strands, making a complex containing eight strands. We speculate that the third complex which we observe in our samples may be such an eight-stranded structure. We are currently investigating this possibility, along with the structure of the homopurine self-complex.

Pulleyblank et al. (1985) have also proposed a supercoil density and pH-dependent transition of $d(GA)_n \cdot d(TC)_n$ oligomers in supercoiled plasmids in which Watson–Crick AT base pairs alternate with Hoogsteen $GC^+$ base pairs in a duplex DNA. We see no evidence for this structure under any of the conditions that we studied.

*Stability of Triplex and Duplex.* The melting behavior of the triplex versus the duplex indicates that the triplex is more thermally stable than the duplex. This is consistent with observations made for other groove binding ligands, such as minor groove binding drugs (Patel, 1979). The samples of the duplex and triplex in Figure 7 both contain 50 mM NaCl and 5 mM $MgCl_2$. Under these conditions, the duplex melts at least 20 °C lower than the triplex. Significant line broadening resulting from duplex strand exchange is evident in the spectrum of the duplex at 35 °C (Figure 7B). In contrast, the triplex has not begun to melt at this temperature and the resonances remain sharp.

Addition of $Mg^{2+}$ stabilizes both the duplex and the triplex forms but does not appear to affect the equilibrium (Figures 9 and 10). No changes were observed in spectra of a sample containing a mixture of triplex and duplex when the $Mg^{2+}$ concentration was titrated from 0 to 24 mM (spectra not shown). Thus, divalent metal ions are not required for triplex formation. Antao et al. (1988) obtained triplexes from poly[d(AG)] and poly[d(CT)] as assayed by CD in buffers containing $Na^+$ cations only. However, Letai et al. (1988) also observed a strong ionic strength dependence of the stability of DNA and RNA triplexes; we find that this is the case for the $d(GA)_4 \cdot d(TC)_4$ triplex and duplex. Both Moser and Dervan (1988) and Cooney et al. (1988) found that addition of divalent cations was essential to observe triplex formation in their systems, probably as a result of a high binding constant for the third strand needed to detect the triplex under their assay conditions.

## Summary

In this work, we have shown that the DNA octanucleotides $d(GA)_4$ and $d(TC)_4$ will form a pyr–pur–pyr triplex which is favored under conditions of low pH and excess $d(TC)_4$ strand and is stabilized, along with the duplex, by addition of $Mg^{2+}$. The triplex is composed of Watson–Crick base paired $d(GA)_4$ and $d(TC)_4$ strands, with the second $d(TC)_4$ strand Hoogsteen base paired along the major groove parallel to the $d(GA)_4$ strand. This pairing scheme will result in either a 5′ or 3′ dangling end for the Hoogsteen base paired $d(TC)_4$ strand; we find that the predominant triplex formed has four TAT and three $CGC^+$ triplets, indicating formation of a triplex with a 3′ dangling dC. Hoogsteen CG base pairs require hemiprotonation of the C, and we are able to observe these imino protons directly. All of the bases in the triplex are anti, and

the sugar conformations in the pyrimidine strands are C3′-endo, consistent with models based on fiber diffraction of poly(dA)·poly(dT)·poly(dT) (Arnott & Selsing, 1974) that indicated an A′-DNA helix for the triplex. However, the sugar conformation of the purine strand is not C3′-endo, indicating that this pyr–pur–pyr triplex containing all four bases has a conformation that may be different from $d(T)_n$–$d(A)_n$–$d(T)_n$ triplexes. In addition to the triplexes, another unidentified heterocomplex is observed in small amounts under some conditions. The single strand $d(GA)_4$ self-associates, and the heterocomplex may be formed from binding of $d(TC)_4$ to the self-associated $d(GA)_4$.

## Acknowledgments

We thank Drs. Gijs A. van der Marel and Jacques van Boom for providing the DNA oligonucleotides that were used in the early stages of this work and Dr. Andrew H.-J. Wang for helpful discussions.

## Supplementary Material Available

Expanded regions of DQF-COSY spectra of the duplex and triplex, NOESY spectra of the duplex in $H_2O$ including assignments of the imino and amino resonanes, and one-dimensional spectra of the exchangeable resonances of the triplex at 1 and 35 °C (4 pages). Ordering information is given on any current masthead page.

## References

Antao, V. P., Gray, D. M., & Ratliff, R. L. (1988) *Nucleic Acids Res. 16*, 719–739.
Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 108*, 619.
Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Blackburn, E. H. (1984) *Cell 37*, 7–8.
Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811–813.
Broitman, S. L., Im, D. D., & Fresco, J. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5120–5124.
Cantor, C. R., & Efstratiadis, A. (1984) *Nucleic Acids Res. 12*, 8059–8072.
Christophe, D., Cabier, B., Bacolle, A., Targovnik, H., Pohl, V., & Vassart, G. (1985) *Nucleic Acids Res. 13*, 5127–5144.
Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.
Elgin, S. C. R. (1984) *Nature 309*, 213–214.
Feigon, J., Wright, J. M., Denny, W. A., Leupin, W., & Kearns, D. R. (1982) *J. Am. Chem.* Soc. *104*, 5540–5541.
Feigon, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1983) *Biochemistry 22*, 5943–5951.
Feigon, J., Wang, A. H.-J., Van der Marel, G. A., Van Boom, J. H., & Rich, A. (1984) *Nucleic Acids Res. 12*, 1243–1263.
Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.
Gilbert, D. E., van der Marel, G. A., van Boom, J. H., & Feigon, J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 13*, 3006–3010.
Guéron, M., Kochoyan, M., & Leroy, J.-L. (1987) *Nature 328*, 89–92.
Haniford, D. B., & Pulleyblank, D. E. (1985) *Nucleic Acids Res. 13*, 4343–4363.
Hanvey, J. C., Klysik, J., & Wells, R. D. (1988) *J. Biol. Chem. 15*, 7386–7396.
Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319.
Hoffman-Liebermann, B., Liebermann, D., Troutt, A., Kedes, L. H., & Cohen, S. N. (1986) *Mol. Cell. Biol. 6*, 3632–3642.
Hoogsteen, K. (1959) *Acta Crystallogr. 12*, 822–823.
Htun, H., & Dahlberg, J. E. (1988) *Science 241*, 1791–1796.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546.
Johnson, D., & Morgan, A. R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1637–1641.
Kallenbach, N. R., Daniel, W. E., Jr., & Kaminker, M. A. (1076) *Biochemistry 15*, 1218–1223.
Kearns, D. R. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 477.
Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acids Res. 15*, 5845–5861.
Konopka, A. K. (1988) *Nucleic Acids Res. 16*, 1739–1758.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.
Lee, J. S., Johnson, D. A., & Morgan, A. R. (1979) *Nucleic Acids Res. 6*, 3073–3091.
Letai, A. G., Palladino, M. A., Fromm, E., Rizzo, V., & Fresco, J. R. (1988) *Biochemistry 27*, 9108–9112.
Massefski, W., Jr., & Redfield, A. G. (1988) *J. Magn. Reson. 78*, 150–155.
Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1987) *Nature 330*, 495–497.
Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.
Patel, D. J., Kozlowski, S. A., Nordheim, A., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1413–1417.
Praseuth, D., Perrouault, L., Doan, T. L., Chassignol, M., Thuong, N., & Hélène, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1349–1353.
Rajagopal, P., & Feigon, J. (1989) *Nature 339*, 637–640.
Rance, M., Sørensen, M., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Richards, J. E., Gilliam, A. C., Shen, A., Tucker, P. W., & Blattner, F. R. (1984) *Nature 306*, 483–487.
Sánchez, V., Redfield, A. G., Johnston, P. D., & Tropp, J. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5659–5662.
Scheek, R. M., Russo, N., Boelens, R., & Kaptein, R. (1983) *J. Am. Chem. Soc. 105*, 2914–2915.
Sen, D., & Gilbert, W. (1988) *Nature 334*, 364–366.
Sklenar, V., Brooks, B. R., Zon, G., & Bax, A. (1987) *FEBS Lett. 216*, 249–252.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Strobel, S. A., Moser, H. E., & Dervan, P. B. (1988) *J. Am. Chem. Soc. 110*, 7927–7929.
van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & van Boom, J. H. (1981) *Tetrahedron Lett. 22*, 38–87.
Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlrab, F. (1988) *FASEB J. 20*, 2939–2949.
Wohlrab, F., McLean, M. J., & Wells, R. D. (1987) *J. Biol. Chem. 262*, 6407–6416.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.