Proton NMR Studies of 5′-d-$(TC)_3$ $(CT)_3$ $(AG)_3$-3′—A Paperclip Triplex: The Structural Relevance of Turns

Laura B. Pasternack,* Shwu-Bin Lin,† Tsung-Mei Chin,‡ Wei-Chen Lin,§ Dee-Hua Huang,* and Lou-Sing Kan§

*Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037 USA; †The Graduate Institute of Medical Technology, National Taiwan University, Taipei, Taiwan 10002; ‡Institute of Applied Chemistry, Chinese Culture University, Taipei, Taiwan 11114; and §Institute of Chemistry, Academia Sinica, Taipei, Taiwan 11529

ABSTRACT In this study, we present the results of structural analysis of an 18-mer DNA 5′-$T_1C_2T_3C_4T_5C_6C_7T_8C_9T_{10}C_{11}T_{12}A_{13}G_{14}A_{15}G_{16}A_{17}G_{18}$-3′ by proton nuclear magnetic resonance (NMR) spectroscopy and molecular modeling. The NMR data are consistent with characteristics for triple helical structures of DNA: downfield shifting of resonance signals, typical for the $H3^+$ resonances of Hoogsteen-paired cytosines; pH dependence of these $H3^+$ resonance; and observed nuclear Overhauser effects consistent with Hoogsteen and Watson-Crick basepairing. A three-dimensional model for the triplex is developed based on data obtained from two-dimensional NMR studies and molecular modeling. We find that this DNA forms an intramolecular "paperclip" pyrimidine-purine-pyrimidine triple helix. The central triads resemble typical Hoogsteen and Watson-Crick basepairing. The triads at each end region can be viewed as hairpin turns stabilized by a third base. One of these turns is comprised of a hairpin turn in the Watson-Crick basepairing portion of the 18-mer with the third base coming from the Hoogsteen pairing strand. The other turn is comprised of two bases from the continuous pyrimidine portion of the 18-mer, stabilized by a hydrogen-bond from a purine. This "triad" has well defined structure as indicated by the number of nuclear Overhauser effects and is shown to play a critical role in stabilizing triplex formation of the internal triads.

INTRODUCTION

It is known that DNA adopts various configurations depending on environmental conditions and base sequences (Wells et al., 1988; Radhakrishnan and Patel, 1994a,b; Sklenar and Feigon, 1990; Rajagopal and Feigon, 1989a, b). One of these configurations is triplex DNA, formed by the binding of a third DNA strand in the major groove of a double helix (Frank-Kamenetskii and Mirkin, 1995). If the third strand is pyrimidine-rich, it binds parallel to the purine strand of the duplex via Hoogsteen basepairing (T*AT or C^+*GC). When cytosine is present in the third strand, these triplexes tend to be stable only at low pH because of the requirement of protonation of the cytosines of the Hoogsteen pairing strand.

It has become increasingly apparent that triplex DNA structures have important biological implications: as a major structural feature in supercoiled plasmids and chromatin (Mirkin et al., 1987; Mirkin and Frank-Kamenetskii, 1994), for selectivity of protein binding (Musso et al., 1998), in gene regulation (Maher et al., 1989; Volkmann et al., 1995; Helene, 1991; Cooney et al., 1988; Degols et al., 1994; Grigoriev et al., 1992; Maher, 1992), and for genetic manipulation (Praseuth et al., 1999). In addition, there is considerable interest in the use of various triplex binding ligands to induce, enhance, or disrupt triplex formation (Cassidy et al., 1996; Xu et al., 1997; and Vigneswaran et al., 1996). The thorough understanding of the structural features of triplex DNA is a critical piece in our overall understanding of the biological functions and the potential therapeutic uses of triplex-forming DNA.

A number of triplex structures have been investigated by nuclear magnetic resonance (NMR) spectroscopy methods (Wang et al., 1996; Radhakrishnan and Patel, 1994a,b; Carbonnaux et al., 1991; Tarkoy et al., 1998; Koshlap et al., 1997; Gilbert and Feigon, 1999; Macaya et al., 1992; Radhakrishnan et al., 1991; de los Santos et al., 1989; Bartley et al., 1997). Early studies of DNA triplexes involved the triplex formation among three separate strands of DNA. In later studies the three strands were connected by linker regions. These DNAs then formed a "paperclip" configuration, with the region of interest being the triplex region in the center.

Interestingly, recent studies investigating the dependence of triplex stability on the length of the linker region revealed the surprising result that even with no linker present, a stable triplex structure could be formed (Chin et al., 2000). In this study we investigate the structural properties of this triplex formation, in which no linker regions are present. NMR data obtained for the single-stranded DNA 5′-$T_1C_2T_3C_4T_5C_6C_7T_8C_9T_{10}C_{11}T_{12}A_{13}G_{14}A_{15}G_{16}A_{17}G_{18}$-3′ (18-mer) suggests that it exists in a triplex configuration (Fig. 1 *A*). Triplex formation in this 18-mer requires sharp turns of the DNA backbone, which have been observed in several DNA hairpin structures (Chou et al., 1996, 1999a,b; Mauffrett et al., 1998; Van Dongen et al., 1997; Gallego et al., 1997; Hare and Reid, 1986; Blommers et al., 1991; Avizonis and Kearns, 1995; Mariappan et al., 1996). The triplex structure under investigation here can be thought of

Submitted July 23, 2001, and accepted for publication March 6, 2002.

Address reprint requests to Lou-sing Kan, Institute of Chemistry, Academia Sinica, Nankang, Taipei, Taiwan 11529. Tel.: 886-2-2789-8550; Fax: 886-2-2788-4184; E-mail: lskan@chem.sinica.edu.tw.

0006-3495/02/06/3170/11 $2.00

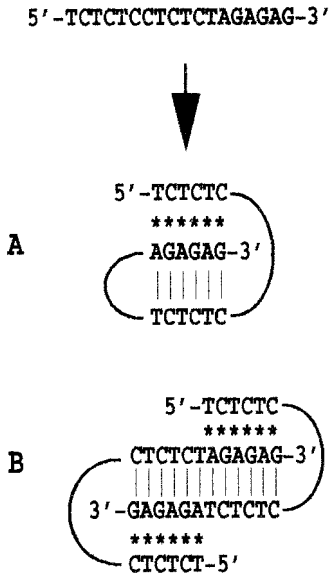


FIGURE 1 Schematic representation of two possible triad formations of the 18-mer 5′-TCTCTCCTCTCTAGAGAG-3′; (*A*) intramolecular or (*B*) intermolecular. A star denotes the potential for Hoogsteen basepairing, and a dash denotes the potential for Watson-Crick basepairing.

as being comprised of two hairpin turns, stabilized by a third strand. To date, there is little understanding of how basepairing of an additional nucleotide may stabilize the residues of a hairpin turn.

In this study, we find that: 1) the sequence 5′-TCTCTC-CTCTCTAGAGAG-3′ forms an intramolecular triplex; 2) the internal triads form typical Hoogsteen*Watson/Crick (H*WC) pairing; 3) the 3′ end C_6*G_{18}-C_7 turn triad forms a modified H*WC structure; and 4) this 3′ end C_6*G_{18}-C_7 turn triad is critical in stabilizing triplex formation. In addition, we have developed a model of the triplex conformation of the 18-mer including a model of a prymidine "hairpin" turn stabilized by purine hydrogen bonding. The model represents the most probable conformation of the 18-mer that is in good agreement with the experimental data.

MATERIALS AND METHODS

Oligonucleotide synthesis and purification

The 18-mer was synthesized on a DNA synthesizer (Applied Biosystems Model 391, foster City, CA) using solid-support phosphoramidite chemistry (Yuhasz et al., 1987; Atkinson and Smith, 1984). After cleavage from the solid support these deoxyoligonucleotides were purified by reverse-phase cartridges (Poly-Pak, Glen Research Corporation, Sterling, VA), using a procedure recommended by the manufacturer. The chain length and purity was verified by gel electrophoresis. The concentration of single-stranded 18-mer was determined at 260 nm, using the molar extinction coefficient of 165,680 $cm^{-1}M^{-1}$ based on the calculation of Fasman method (Fasman, 1975).

Ultraviolet (UV) thermal melting and circular dichroism (CD) spectroscopy

UV absorbance versus temperature profiles and CD spectra were performed as described previously (Chin et al., 2000).

NMR spectroscopy

The 18-mer was dissolved in 0.5 ml of H_2O (with 10% D_2O) or 100% D_2O containing 0.15 M NaCl and 0.01 M acetate buffer. The samples' pH was adjusted to 4.5. Phase-sensitive two-dimensional (2-D) nuclear Overhauser and exchange spectroscopy (NOESY), double quantum-filtered correlated spectroscopy (DQF-COSY), and total correlation spectroscopy experiments were performed on either a Bruker DRX-700 or DRX-600 spectrometer (Billerica, MA) at 1°C. For NOESY experiments the mixing time was set to 200 ms. Additional 2-D DQF-COSY experiments were also performed at 25°C. Solvent resonance signal (H_2O) was suppressed with the water suppression by gradient tailored excitation pulse sequence.

Spectral assignment was carried out by standard sequential analysis procedures (Wijmenga et al., 1993). Examination of the COSY 2-D NMR spectrum provides assignment of H5 and H6 resonances of cytosine residues and leads to the identification of their H1′ proton in NOESY spectra. Analysis of the entire H1′-H6/H8 region of the NOESY spectrum provides the sequential assignment of H1′, H6, and H8 resonances for all residues. Along the H6/H8 chemical shift nuclear Overhauser effects (NOEs) identify H2′, H2″, H3′, H4′, and thymine methyl resonances. These assignments are then verified by analysis of H3′-H1′, H3′-H4′, H2′-H2″, and H2′/H2″-H3′ NOEs. Inspection of NOESY spectra acquired with the DNA in H_2O provides assignment of the exchangeable imino and amino protons. For thymine, the imino has an intrabase NOE to the methyl protons. For cytosine, there is a characteristic NOE between the amino protons and the intrabase H5 and H6 protons. In addition, for Hoogsteen-paired cytosines, the imino can be identified by its intrabase NOE to the amino protons. For guanine, the imino proton is identified by its NOE to the Watson-Crick basepaired amino group. For adenine, the amino protons have a strong NOE to the Watson-Crick–paired imino resonance. The H2 resonance of adenine is assigned by its small intranucleotide NOE to H1′ and to the very strong NOE to the imino proton of the Watson-Crick—basepaired nucleotide.

Distance restraints were determined by analysis of NOESY spectra acquired in both D_2O and H_2O. For nonexchangeable protons, peak intensities were converted to distances by comparison to the H2′-H2″ cross-peak (1.8 Å) and grouped as small (1.8–2.4 Å), medium (1.8–3.5 Å), and large (1.8–5.0 Å). For exchangeable protons, a uniform restraint of 1.8–5.0 Å was used, because exchange with water makes peak integrals independent of proton-proton distance. Glycosidic angle torsional restraints were determined based on comparison of the H5-H6 (of cytosine) NOE intensity with the H6-H1′ and H8-H1′ intensities.

Molecular modeling

The DNA was built as a single continuous strand using MSI software and the Insight II program (Acelrys Inc., San Diego, CA). Na^+ ions were placed 4 Å from each phosphorous atom. The strand was loosely folded into a right-handed parallel triplex with pyrimidine-purine-pyrimidine pairs as indicated by CD experiments (Chin et al., 2000), minimized and counterions added. This was used as the starting structure for all calculations. Other globally folded structures were evaluated for compliance with the NOE data. However, the right-handed parallel pyrimidine-purine-pyrimidine described above complied the best. All calculations were performed on a Silicon Graphics Power Challenge with the Discover program using the Amber forcefield. Calculations occur in a two-step process: 1) simulated annealing without explicit water using a distant, dependent dielectric constant of 4*r and 2) minimization. Simulated annealing includes tem-

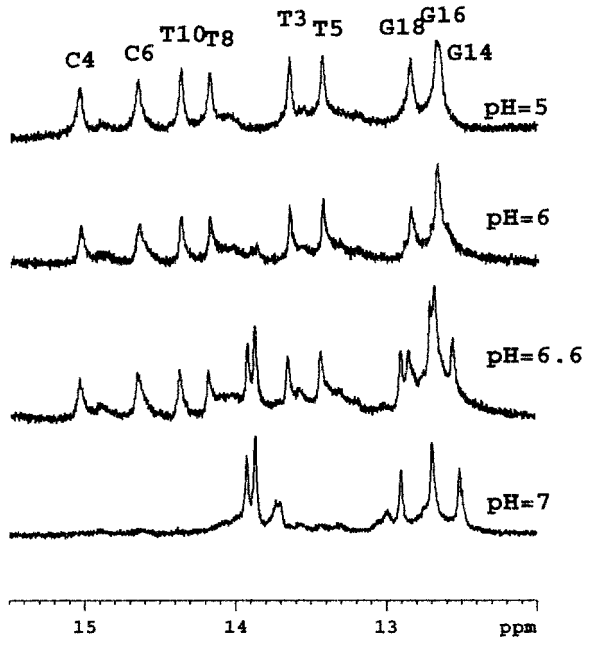


FIGURE 2 The imino proton region of the ^{1}H NMR, acquired at 600 MHz, of the 18-mer in H_2O at various pH values. Resonance assignments at pH 5 are indicated above the peaks. Resonances associated with the $H3^+$ of Hoogsteen-bonding cytosine residues disappear at higher pH values.

perature ramp to 800 K, high-temperature equilibration to generate random starting structures, ramping on of restraints and force field parameters to establish backbone conformation and base orientation, and slow cooling to 300 K. Minimization includes 1000 iterations of restrained conjugate gradient and 5000 iterations of steepest descent. The program Suppose was used for root mean square deviation (RMSD) calculations. All RMSD values are reported as all atom RMSDs to the mean structure.

RESULTS AND DISCUSSION

Identification of triplex formation by 1-D NMR

1-D ^{1}H NMR data of the 18-mer in H_2O were acquired on a 600-MHz spectrometer. Fig. 2 shows the imino proton region of the data at various pH values. At pH 5 resonance signals appear between 14 and 15 ppm, typical of $H3^+$ resonances of hydrogen-bonding cytosine residues of Hoogsteen-basepairing (Wang et al., 1996; Tarkoy et al., 1998; Koshlap et al., 1997). These resonances disappear at higher pH, typical of the pH dependence of triplex formation with cytosine in the Hoogsteen-binding strand. Thus, the 1-D ^{1}H NMR indicates the 18-mer forms a triplex structure.

Assignment

Assignment of proton resonances was accomplished by typical 2-D sequential analysis as described in the Materials and Methods section. Fig. 3 shows the H6/H8-H1′ and H8/H6-H3′ regions of the NOESY spectra acquired in D_2O which provided sequential assignments for the H1′, H3′, H6, and H8 protons. The assignment is more complicated for triplex structures because intratriad NOEs between H1′ of the Hoogsteen-paired bases and the H8 of the purine is also expected. These NOEs are considered to constitute proof of triplex formation. The majority of nonexchangable protons were assigned and listed in Table 1. Assignment of exchangeable protons was accomplished using the NOESY spectra of the 18-mer in H_2O (Fig. 4). The imino to imino connectivities can be traced sequentially from residue T_3 to residue C_6 of the Hoogsteen-paired section, and cross-strand from C_6 to T_{10} of the Watson-Crick–paired section (Fig. 4 *B*). Resonances associated with the imino protons of all but the T_1, C_2, and T_{12} bases have been assigned. Assignment of $H3^+$ resonances of Hoogsteen-bonded cytosines was made by their characteristic downfield chemical shifts and their characteristic NOE pattern to their own NH_2 (Fig. 4 *A*). The two $H3^+$ resonances belong to C_4 and C_6 cytosine residues. No resonance signal could be located for the $H3^+$ of the C_2 residue. This indicates that considerable fraying occurs at the 5′ end, through the C_2 residue, and that the C_6 residue, suspected to be involved in a tight turn, is strongly hydrogen bonded. This would be atypical of a hairpin turn. Usually the imino resonance is not seen. The structural features leading to this are discussed later. Several resonances of the of T_{12} and A_{13} residues show upfield shifting. This indicates that these protons are more shielded than in standard B-DNA. This shielding is typical of residues in hairpin turn regions because of severe kinking of the backbone that places the sugar residue protons and backbone H4′ and H5′/H5″ protons closer to the neighboring or their own base. The H5′/H5″ assignments were not made because of extreme spectral overlap. Chemical shifts are listed in Table 1.

NOE analysis

NOE data also indicate the formation of triplex DNA by the 18-mer. Typical connectivities indicating the presence of Watson-Crick—basepairing were observed: NOE between the cytosine amino and guanine imino for G:C basepairs and between thymine imino and adenine amino and H2 proton for A:T basepairs (Fig. 4 *A*). In similar fashion, NOEs indicative of Hoogsteen-basepairing are present: NOE between the $H3^+$ of the pymimidine and H6/H8 of the purine (Fig. 4 *A*). In addition, numerous imino-to-imino NOEs are present (Fig. 4 *B*). The complete set of NOEs determined from NOESY spectra is available as Supporting Information. Individual NOEs are discussed below. Torsional restraints for glycosidic angles were determined based on the comparison of H5-H6 NOE strength with H6-H1′ and H8-H1′ intensities, found to be significantly weaker than the H5-H6 NOE intensities in all cases. Thus the glycosidic angles were restricted to between −90 and −175°.

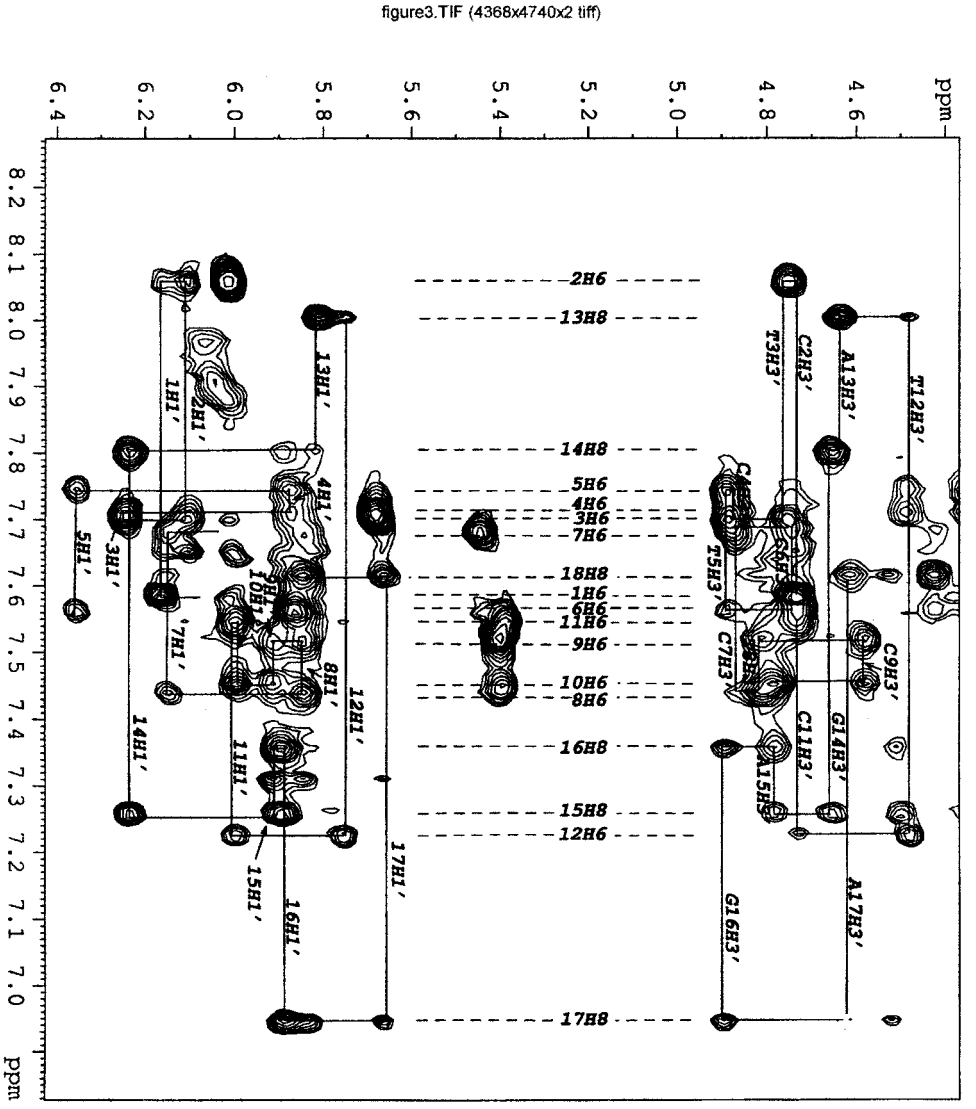


FIGURE 3 Region of NOESY spectrum acquired at 600 MHz of the 18-mer in D_2O depicting the H1′-H8/H6 and H3′-H8/H6 walk used for initial resonance assignment.

Structural calculation

A total of 279 experimentally derived distance constraints and 13 torsional angles were calculated from two different 200-ms mixing time NOESY experiments (600 and 700 MHz), and were used during all calculations. Conformation of sugar residues was left unrestrained. Calculations were performed as described in Materials and Methods. During the first phase of the calculation, high temperature structures were generated to ensure that sufficiently random starting structures are used. Ramping on of NOE constraints at high temperature followed by the ramping on of force field parameters during slow cooling results in all conformations that satisfy the experimental NMR data. Analysis of the results from initial calculations indicated correct base orientation for hydrogen bonding in a typical Hoogsteen triplex hydrogen bond formation for most of the central basepairs. Therefore, a total of 16 hydrogen bonds and their associated dihedral constraints were added to the calculation. Dihedral angle constraints for the backbone in the H*WC forming central region were restricted to all allowable angles for DNA. No backbone dihedrals were used in the turn regions. In addition, dihedral constraints were applied to maintain planarity of the bases. An iterative process of evaluation of the resulting model for protons residing within 6 Å of one another and reanalysis of the NOESY data for those NOEs was used to detect additional NOEs. The final number of NOE restraints is listed in Table 2 along with calculation results. A total of 41 structures were generated, 78% of the structures converged to a single structure with an average total atom RMSD to the mean structure of 0.646 Å.

Overall conformation

The superposition of the ten lowest energy structures (of the converging 32 final structures) is presented in Fig. 5 and a

TABLE 1 Proton chemical shifts of the 18mer

	H1′	H2′	H2″	H3′	H4′	H2/H5	H6/H8	Methyl	NH_2	NH
T1	6.162	2.333	2.490	4.733	4.128		7.580	1.719		UA
C2	6.103	2.250	2.719	4.750	4.280	6.017	8.065		8.935/10.070	UA
T3	6.249	2.390	2.554	4.888	4.277		7.702	1.652		13.634
C4	5.875	2.134	2.667	4.898	UA	5.683	7.714		9.145/9.910	15.038
T5	6.356	2.269	2.616	4.896	4.091		7.747	1.647		13.416
C6	6.000	2.112	2.403	4.742	4.206	5.865	7.570		8.128/9.768	14.639
C7	6.147	2.142	2.483	4.872	4.076	5.448	7.686		8.259/7.391	
T8	5.841	2.216	2.503	4.813	4.078		7.444	1.625		14.176
C9	5.911	1.910	2.435	4.580	4.059	5.402	7.524		8.179/7.096	
T10	5.998	2.014	2.445	4.776	4.077		7.457	1.625		14.360
C11	6.000	2.511	2.021	4.735	4.077	5.380	7.554		7.066/8.148	
T12	5.745	1.448	1.943	4.475	4.004		7.231	1.702		UA
A13	5.804	2.157	2.291	4.626	3.807	UA	8.004		7.288/7.811	
G14	6.237	2.559	2.946	4.648	4.288		7.803			12.599
A15	5.890	2.039	2.791	4.778	4.505	7.661	7.259		7.372/7.710	
G16	5.884	2.364	2.748	4.892	4.236		7.362			12.665
A17	5.662	1.834	2.614	4.614	4.519	7.313	6.946		7.732/8.205	
G18	5.838	2.300	2.487	4.429	UA		7.620			12.832

UA = unassigned

stereoview of a single structure in Fig. 6. The backbone forms an intramolecular paperclip. Residues T_8 and A_{17}, C_9 and G_{16}, T_{10} and A_{15}, and C_{11} and G_{14} form Watson-Crick basepairing similar to B-DNA. Residues T_1 through T_5 lie in the major groove of the Watson-Crick double helix and residues C_2 through T_5 form typical Hoogsteen basepairing with residues T_{14} through A_{17}.

It should be noted that the 18-mer has the potential to form an intermolecular triplex with Watson-Crick pairing of the self-complementary C_7 through G_{18} region with residues T_1 through C_6 binding to the duplex major groove from both ends (Fig. 1 *B*). The major structural difference between the intramolecular binding (Fig. 1 *A*) and intermolecular binding (Fig. 1 *B*) is that, in the intermolecular triplex, the residues T_{12} and A_{13} would be part of a continuous B-DNA–like duplex, whereas, in the intramolecular triplex, they would be part of a sharp turn. Three lines of evidence indicate that triplex formation is indeed intramolecular: 1) chemical shift analysis displays the presence of upfield shifting of resonances associated with T_{12} and A_{13} sugar protons, typical of the sugar residues of hairpin turn regions (Van Dongen et al., 1997); 2) intratriad NOEs of T_{12} and A_{13} do not form the pattern typically found in duplex DNA; and 3) lack of resonance associated with the imino proton of residue T_{12} is typical of hairpin formations. The detail discussion will be given in a later section (Conformation of the T_1*A_{13}-T_{12} turn). In addition, our CD and UV studies indicate that triplex formation is concentration-independent (Chin et al., 2000). This also suggests that the triplex formed unimolecularly.

Conformation of the central T_3*A_{15}-T_{10}, C_4*G_{16}-C_9 H*WC forming region

In our model the central H*WC binding region is comprised of the T_3*A_{15}-T_{10} and C_4*G_{16}-C_9 residues (Fig. 7). As discussed above, this section forms a well defined region with an all atom RMSD to the mean structure of 0.523 Å for T_3*A_{15}-T_{10} and 0.500 Å for C_4*G_{16}-C_9.

For the T_3*A_{15}-T_{10} triad there are clear NOEs between T_{10}:H3 and A_{15}:H2 along with A_{15}:NH_2 indicating Watson-Crick basepairing. In addition, the NOE between T_3:H3 and A_{15}:H8 indicates Hoogsteen pairing. All three residues are planar in the final model.

For the C_4*G_{16}-C_9 triad, an NOE was observed between C_4:NH and G_{16}:H8 (Fig. 4 *A*) and between C_4:NH_2 and C_9:NH_2, typical for Hoogsteen basepairings. NOE was observed between C_9:NH_2 and G_{16}:H1, typical for Watson-Crick basepairing. All three residues are planar in the final model.

Conformation of the C_2*G_{14}-C_{11}, and T_5*A_{17}-T_8 H*WC forming base triads

In the final model, C_2*G_{14}-C_{11} and T_5*A_{17}-T_8 also form typical H*WC triads (Fig. 7). This section forms a well defined region with all atom RMSD values to the mean structure of 0.569 Å for T_5*A_{17}-T_8 and 0.534 Å for C_2*G_{14}-C_{11}.

For the C_2*G_{14}-C_{11} triad an NOE was observed between C_2:NH_2 and C_{11}:NH_2, typical of Hoogsteen basepairing. NOE was observed between C_{11}:NH_2 and G_{14}:H1, typical of Watson-Crick basepairing. In the final model, residues G_{14} and C_{11} are planar with residue C_2 slightly out of plane. This out of plane is most likely attributable to fraying of the 5′ end as discussed below.

For the T_5*A_{17}-T_8 triad there are clear NOEs between T_8:H3 and A_{17}:H2 and between T_8:H3 and A_{17}:NH_2, indicating Watson-Crick basepairing. An NOE observed between T_5:H3 and A_{17}:H8 indicates Hoogsteen pairing. In addition, the T_5:H1′ to A_{17}:H8 NOE can clearly be ob-

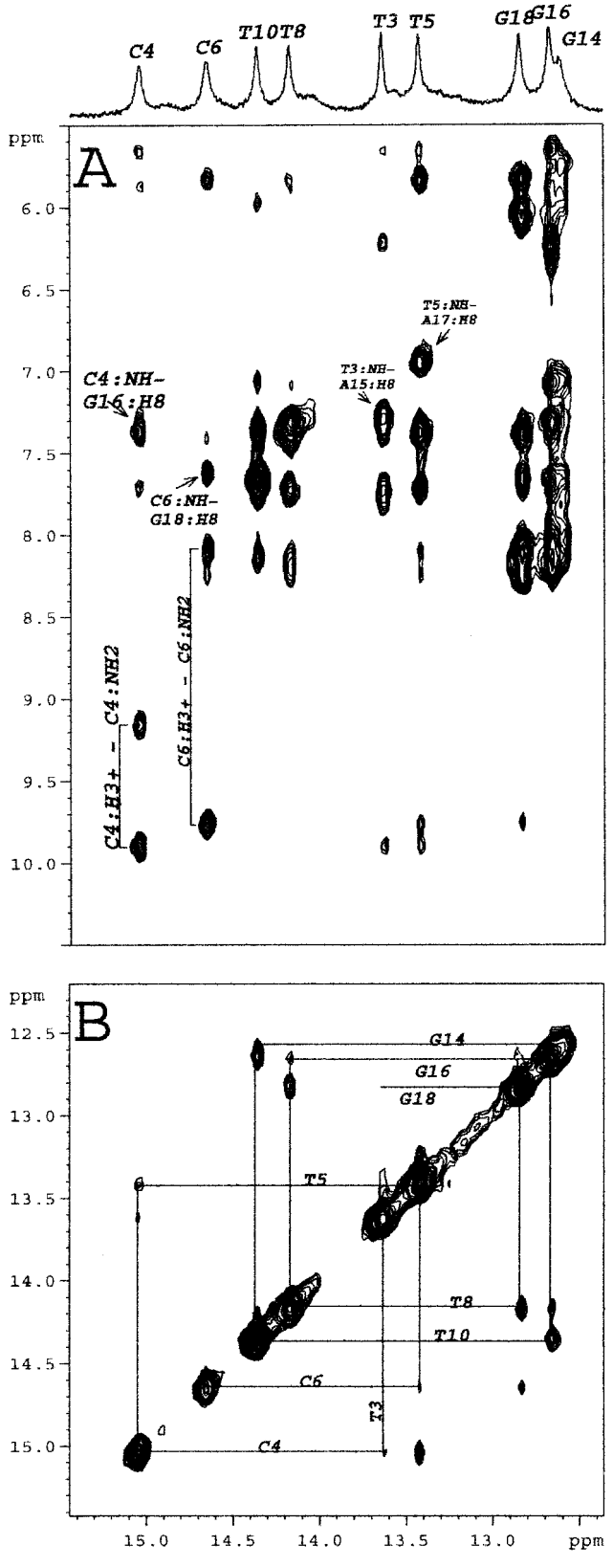


FIGURE 4 Region of NOESY spectrum acquired at 600 MHz of the 18-mer in H_2O depicting (*A*) the $H3^+$-NH_2 connectivity and assignment of cytosines of the Hoogsteen pairing bases, and (*B*) imino-imino connectivities.

TABLE 2 Input and results of structure calculations

Initial calculations included:	
Total NOE restraints	283
Intranucleotide*	85
Sequential	133
Nonsequential internucleotide	65
Hydrogen bonds	16
Glycosidic angle constraints	13
All atom RMSD to the mean structure	
Violations of experimentally determined distance restraints (>0.8Å)	2
Violations of experimentally determined dihedral restraints (>10)	2

*Note: trivial intranucleotide sugar NOE distances were not included

served. The presence of the NOE between H1′ of the Hoogsteen-binding strand and H8 of the purine strand is considered proof of triplex formation. Residues T_5 and A_{17} are planar with T_8 slightly out of plane. This seems to be attributable to the proximity of the turn region. Numerous additional intratriad and intertriad NOEs are identified and summarized in the Supporting Information section.

Conformation of the T_1*A_{13}-T_{12} turn

In the final model the residues T_{12}, A_{13}, and T_1 appear to form a turn involving residues T_{12} and A_{13} (turn associated with the Watson-Crick basepairing hairpin section) with additional loose association of a fraying residue T_1 (indicated by the narrow line width of T_1 resonances). Fig. 7 depicts a cross-section through this triad. However, a stereoview highlighting the orientation of these residues relative to the whole model can also be seen in Fig. 6. The three residues form a fairly well defined turn with an all atom RMSD of 0.796 Å. These residues do not have the typical hydrogen-bonded proton resonance of the imino protons of T_{12} and T_1 to N1 and N7 of A_{13}, respectively, in a standard H*WC pairing pattern. In consequence, the NOE between the imino proton of T_1 to H8 of A_{13} in Hoogsteen pairing is also missing. Furthermore, NOEs typically seen in nonturn regions of Watson-Crick pairing in Fig. 1 *B* such as 1) T_{12}:H3 to A_{13}:H2, 2) G_{14}:H8 to A_{13}:H2′/H2″, 3) C_{11}:NH_2 to A_{13}:NH_2, 4) T_{12}:CH_3 to A_{13}:NH_2, 5) C_{11}:H2′ to T_{12}:H6, and 6) C_{11}:H2″ to T_{12}:CH_3, are not presented in our study. These NOEs are observed for all other residues. In fact, the NH proton of residue T_{12} was not observed, indicating that this proton is in fast exchange with the solvent. In the final model, this NH proton is fully exposed to solvent with no hydrogen-bonding possibility. The A_{13}:H4′, A_{13}:H5′/H5″, and T_{12}:H2′/H2″ protons show highly unusual upfield shifts, indicating the influence of ring currents from a base directly above or below the proton. This is typical of residues in a hairpin turn (Van Dongen et al., 1997). The orientation of residues T_{12}, A_{13}, and G_{14} in the final model account for the upfield shift of these resonances. In standard

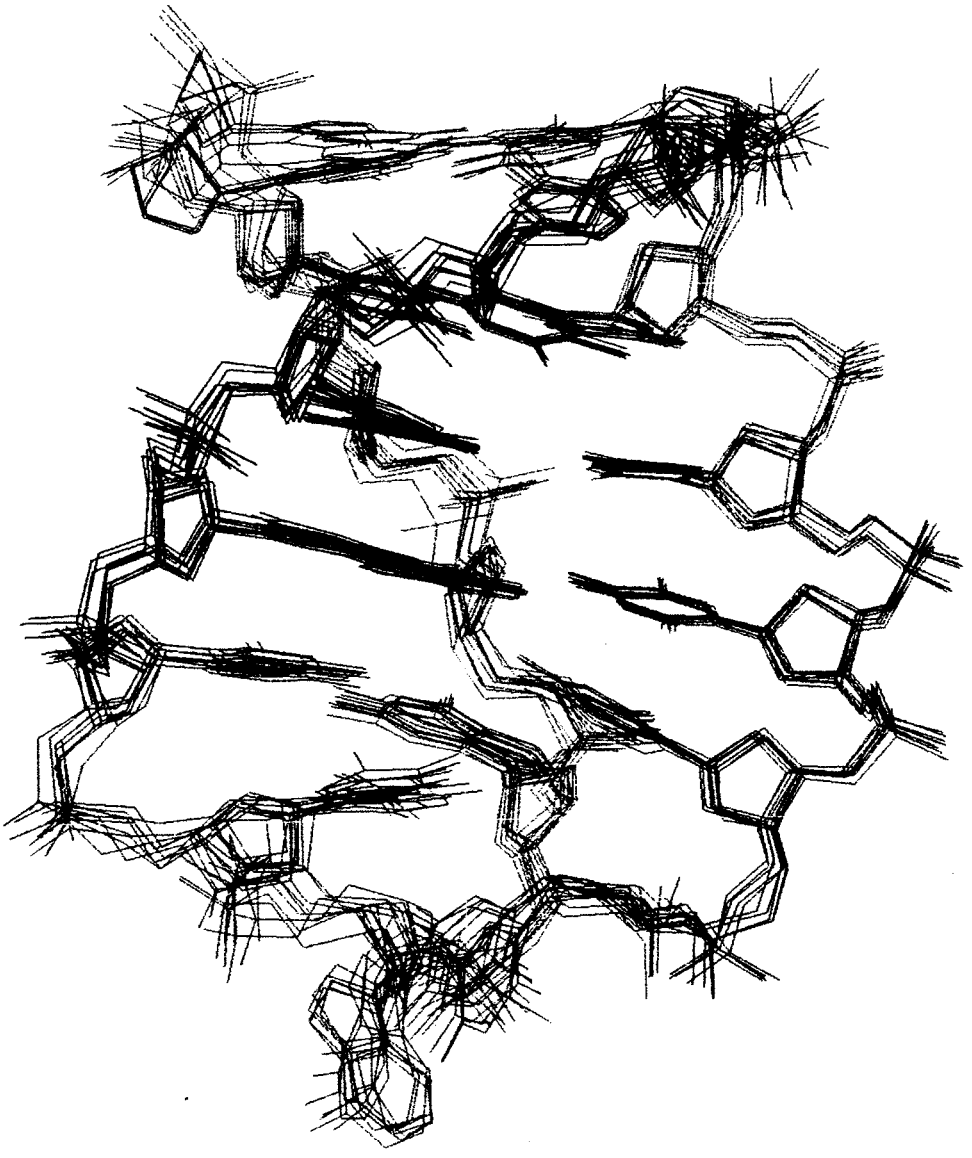


FIGURE 5 Superimposition of the 10 low energy structures. Hydrogens have been removed to simplify the figure.

B-DNA the H5′/H5″ and H4′ protons are oriented toward the exterior of the DNA and away from the bases. This is not the case in a tight turn were the backbone kinks and comes fairly close to the internal bases (Fig. 6). Indeed, in the final model, the base of residue A_{13} is near the H5′/H5″ of residue A_{13} and the base of residue G_{14} is near A_{13}:H4′. The base orientation of the T_{12} through A_{13} turn of the 18-mer is remarkably similar to a previously reported DNA hairpin structure with the 5′ bases of the turn stacking in a continuous fashion and the 3′-loop base folded out of the plane of base stacking and into the major groove (Van Dongen et al., 1997). Thus, this end appears to be a typical hairpin turn with base fraying of the 5′-T_1 residue. The RMSD of residues T_{12} and A_{13} is 0.742 and 0.888 Å, respectively, compared with an RMSD of 0.531 Å for the internal triplets, indicating that this region is slightly less well defined than the rest of the model.

Conformation of the C_6*G_{18}-C_7 triad

Residues C_6, G_{18}, and C_7 comprise those residues responsible for the other turn region. These three basepairs form a well defined turn with an RMSD value of 0.860 Å (Fig. 6). The turn associated with the residues C_6*G_{18}-C_7 is quite different from that of the T_1*A_{13}-T_{12} turn. Residue G_{18} shows linewidths typical of a nonfraying basepair and numerous NOEs to its adjacent residue A_{17} and to residues C_6

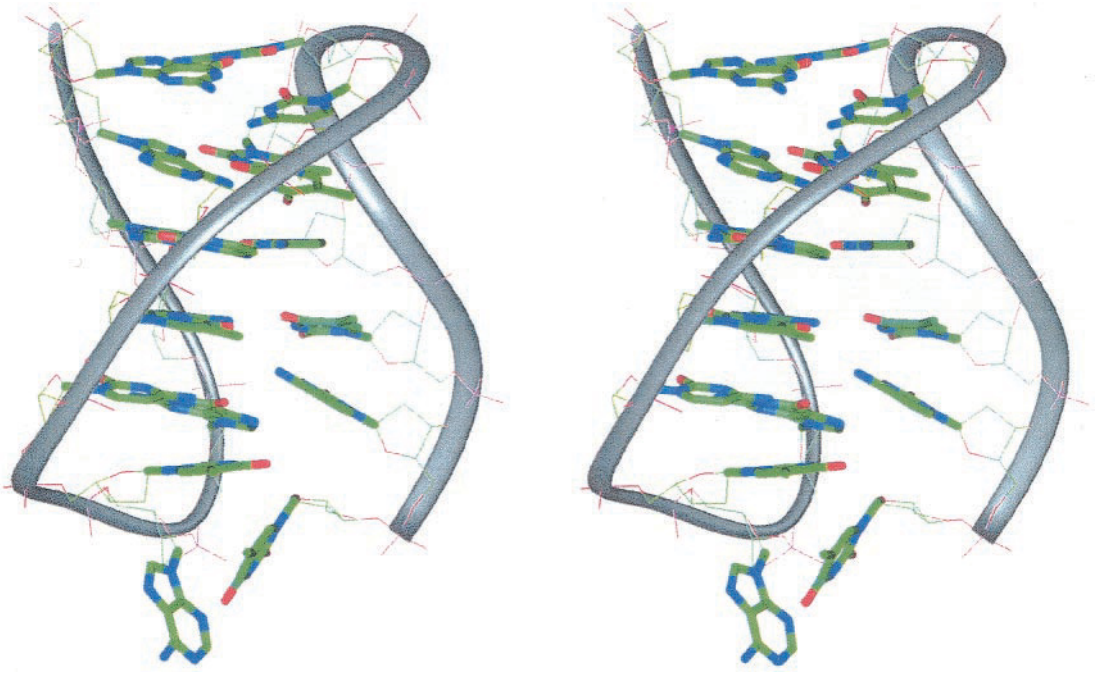


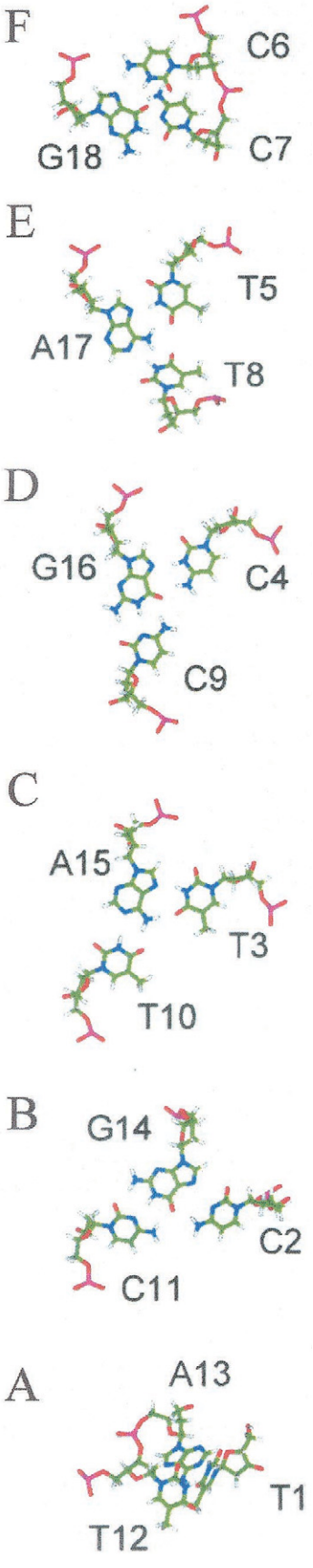
FIGURE 6 Stereoview of the triplex. Hydrogens have been removed to simplify the figure.

and C_7. This indicates that there is little base fraying at the 3′ end of the triplex. Indeed, NOEs typical of a Watson-Crick and Hoogsteen triad are present between these bases, indicating that they share some similarity to a standard Watson-Crick and Hoogsteen pairing. However, the NOE profile between C_6 and G_{18} is clearly different from that of typical Hoogsteen paired bases. NOE was observed between C_6:NH_2 and G_{18}:H8, not typical of Hoogsteen pairing. Presence of the $H3^+$ resonance for residue C_6 indicates its involvement in hydrogen bonding. Based on the experimentally observed data, initial molecular modeling results indicated correct base orientation and distance for hydrogen bonding between C_6:$H3^+$ and G_{18}:O6. This hydrogen bond was included in subsequent calculations. In the final model the G_{18} and C_6 bases create a planar pairing with the C_7 base displaced below the plane (Figs. 6 and 7). This creates a very energetically favorable stacking arrangement of the 3′ bases where C_6 stacks on C_7, which is stacking on T_8. This arrangement shows great similarity to previously published hairpin turns in which basepair stacking is seen on the 5′ side of the hairpin turn. It is important to note, however, that in this case the T_1 through T_{12} turn being discussed is not a hydrogen bonding hairpin but rather, a turn comprised of consecutive pyrimidine residues (Fig. 6). Thus this kind of hairpin turn may only exist when additional bases are present to stabilize it.

It is interesting, therefore, to consider just how the two stacking bases of the turn (C_6 and C_7) interact with the third base (G_{18}) to form a stable triad. As discussed above, structural analysis of the NMR data indicates hydrogen bonding between the cytosine C_6 (at 5′ end of the turn) and the third-base guanine G_{18}. In addition, the final model indicates hydrogen-bonding potential between the C_7 and G_{18} bases in typical Watson-Crick fashion, even though C_7 is displaced out of the plane of the C_6 to G_{18} basepair. It should be noted that no hydrogen bonding between C_7 and G_{18} was included in the calculation. There is also some hydrogen bonding possible between the C_7 base and the A_{17} base. In addition, no glycosidic angle restraints were applied for residues C_6 and C_7 during the calculation to allow orientation of the bases by NOE distance only. Final structures, however, have glycosidic angles of $-141.89°$ for C_6 and $-112.36°$ for C_7, well within the range of angles indicted by the intensity of the H6-H1′ NOE. The fairly high RMSD of 1.165 Å for the G_{18} residue is attributable to variation of the backbone region of the residue, whereas the position of the base is fairly well defined.

CD and UV analysis of the stabilizing influence of end basepairs

To evaluate the contribution of the end basepairs to the stability of triplex formation, two 17-mer sequences were studied by CD and UV. One of them lacks the T base at 5′-end, $C(TC)_2(CT)_3(AG)_3$, (5′-$C_2T_3C_4T_5C_6C_7T_8C_9T_{10}C_{11}T_{12}A_{13}$-$G_{14}A_{15}G_{16}A_{17}G_{18}$-3′ or 18-mer-$T_1$) and the other has no G base at 3′-end, $(TC)_3(CT)_3(AG)_2A$, (5′-$T_1C_2T_3C_4T_5C_6C_7T_8$-$C_9T_{10}C_{11}T_{12}A_{13}G_{14}A_{15}G_{16}A_{17}$-3′ or18-mer-$G_{18}$). The expected effect of the former one is the elimination of stacking



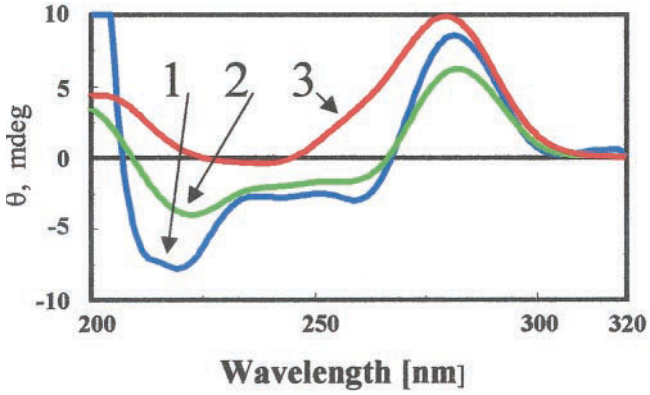


FIGURE 8 CD analysis of the 18-mer $(TC)_3(CT)_3(AG)_3$ (1), 17mer $C(TC)_2(CT)_3(AG)_3$ (2), and 17-mer $(TC)_3(CT)_3(AG)_2A$ (3) showing stabilizing effect of the 3′ end G_{18} residue on formation of triplex DNA of the internal residues.

effect of T_1 to A_{13} in the type IV A_{13}-T_{12} turn and the latter one destroys the C_6*G_{18}-C_7 base triad by exclusion of the center G base. As shown in Fig. 8, the CD spectrum of 18-mer-T_1 resembles that of the 18-mer under the same condition (150 mM NaCl, 10 mM phosphate buffer, pH 4.5) and suggests that there is significant triplex formation by 18-mer-T_1. In contrast, the CD spectrum of 18-mer-G_{18} at pH 4.5 lacks triplex formation. Similar results are obtained by UV melting curves. The T_m of the 18-mer is 54.0 and the 18-mer-T_1 is 59.0, although the T_m of the 18-mer-G_{18} is 41.9. These results showed that the triad C_6*G_{18}-C_7 is a critical factor in stabilization of triplex formation of the 18-mer.

CONCLUSION

Triplex formation of 5′-TCTCTCCTCTCTAGAGAG-3′ was initially detected by UV and CD analyses in our laboratory (Chin et al., 2000). In this study we have pursued a more detailed investigation of the structure of this 18-mer by 1-D and 2-D NMR.

Our NMR data support the presence of triplex formation of the 18-mer. 1-D NMR spectra showing highly downfield-shifted resonances indicate the presence of $H3^+$ protons of Hoogsteen-binding cytosine residues. The pH titration data verify the pH dependence of these resonances, also typical of Hoogsteen hydrogen bonding. Imino resonance and interbase NOEs typical of triad formation are present for most of the central residues. In addition, NOEs and linewidths indicate that the end turn regions are well defined turns with some fraying of the 5′ end nucleotide.

The model derived from experimentally observed NOEs and dihedral angles is a triplex with typical Watson-Crick

FIGURE 7 Cross-section through each triad; T_1*A_{13}-T_{12}, C_2*G_{14}-C_{11}, T_3*A_{15}-T_{10}, C_4*G_{16}-C_9, T_5*A_{17}-T_8, and C_6*G_{18}-C_7.

and Hoogsteen hydrogen bonding in the central two triads. In addition, the neighboring two triads show typical H*WC basepairing, which, however, have bases which are out of plane because of proximity of the turn regions. As no linker regions exist, tight turns are apparent. These turns can be thought of as hairpin turns, each stabilized by a third nucleotide in the vicinity. The two turns are different. One is a hairpin turn created by the Watson-Crick basepairing sections with the third nucleotide (T_1) coming from the Hoogsteen-pairing portion. This turn seems to be accomplished by the displacement of the purine residue out of the plane of the adjacent bases and into the major groove. This hairpin turn shares features with one previously published DNA hairpin (Van Dongen et al., 1997). The other turn is a pyrimidine fold of the continuous pyrimidine section of the 18-mer. This turn is distinctly different from a hairpin turn. In the pyrimidine fold, there is no stabilizing hydrogen bonding between the two arms of the fold. Such a fold must be stabilized by nearby residues. In this case, it is stabilized by the purine section of the 18-mer. Interestingly, CD and UV analysis of a 17-mer lacking the 3′ G_{18} (18-mer-G_{18}) residue shows no triplex formation, indicating that triad formation occurring at the C_6*G_{18}-C_7 turn is a critical feature in stabilization of the entire triplex.

We thank Dr. Detlef Moskau of Bruker AG in Switzerland for obtaining NMR data on the 700-MHz spectrometer.

This work is partially supported by the National Science Council Executive Yuan Tiawan. The Scripps Research Institute Manuscript number is 13501-CH.

REFERENCES

Atkinson, T., and M. Smith. 1984. Oligonucleotide Synthesis: A Practical Approach. M. J. Gait, editors. IRL Press, Oxford.

Avizonis, D. Z., and D. R. Kearns. 1995. Structural characterization of d(CAACCCGTTG) and d(CAACGGGTTG) mini-hairpin loops by heteronuclear NMR: the effects of purines versus pyrimidines in DNA hairpins. *Nucleic Acids Res.* 23:1260–1268.

Bartley, J. P., T. Brown, and A. N. Lane. 1997. Solution conformations of an intramolecular DNA triplex containing a nonnucleotide linker: comparison with the DNA duplex. *Biochemistry.* 36:14502–14511.

Blommers, M. J., F. J. Van de Ven, G. A. Van der Marel, J. H. Van Bloom, and C. W. Hilbers. 1991. The three-dimensional structure of a DNA hairpin in solution. Two-dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT). *Eur. J. Biochem.* 201:33–51.

Carbonnaux, C., G. A. Van der Marel, J. H. Van Bloom, W. Guschlbauer, and G. V. Fazakerley. 1991. Solution structure of an oncogenic DNA duplex containing a G-A mismatch. *Biochemistry.* 30:5449–5458.

Cassidy, S. A., L. Strekowski, and K. R. Fox. 1996. DNA sequence specificity of a naphthylquinoline triple helix-binding ligand. *Nucleic Acids Res.* 24:4133–4138.

Chin, T. M., S. B. Lin, S. Y. Lee, M. L. Chang, A. Y. Cheng, F. C. Chang, L. Pasternack, D. H. Huang, and L. S. Kan. 2000. "Paper-clip" type triple helix formation by 5′-d-$(TC)_3Ta(CT)_3Cb(AG)_3$ (a and b = 0–4) as a function of loop size with and without the pseudoisocytosine base in the Hoogsteen strand. *Biochemistry.* 39:12457–12464.

Chou, S. H., Y. Y. Tseng, and B. Y. Chu. 1999a. Stable formation of a pyrimidine-rich loop hairpin in a cruciform promotor. *J. Mol. Biol.* 292:309–320.

Chou, S. H., Y. Y. Tseng, and S. W. Wang. 1999b. Stable sheared A-C pair in DNA hairpins. *J. Mol. Biol.* 287:301–313.

Chou, S. H., L. Zhu, Z. Gao, J. W. Cheng, and B. R. Reid. 1996. Hairpin loops consisting of single adenine residues closed by sheared A-A and G-G pairs formed by the DNA triplets AAA and GAG: solution structure of the d(GTACAAAGTAC) hairpin. *J. Mol. Biol.* 264:981–1001.

Cooney, M., G. Czernuszewicz, E. H. Postel, S. J. Flint, and M. E. Hogan. 1988. Site-specific oligonucleotide binding represses transcription of the human c-*myc* gene in vitro. *Science.* 241:456–459.

Degols, G., J. P. Clarenc, B. Lebleu, and J. P. Leonetti. 1994. Reversible inhibition of gene expression by a psoralen functionalized triple helix forming oligonucleotide in intact cells. *J. Biol. Chem.* 269: 16933–16937.

de los Santos, C., M. Rosen, and D. Patel. 1989. NMR studies of DNA $(R+)_n$-$(Y-)_n$-$(Y+)_n$ triple helices in solution: imino and amino proton markers of T-A-T and C-G-C+ base-triple formation. *Biochemistry.* 28:7282–7289.

Fasman, G. D. 1975. CRC Handbook of Biochemistry and Molecular Biology, 3rd ed, Vol. 1. CRC Press, Cleveland, OH.

Frank-Kamenetskii, M. D., and S. M. Mirkin. 1995. Triplex DNA structures. *Annu. Rev. Biochem.* 64:65–95.

Gallego, J., S. H. Chou, and B. R. Reid. 1997. Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *J. Mol Biol.* 273:840–856.

Gilbert, D. E., and J. Feigon. 1999. Multistranded DNA structures. *Curr. Opin. Struct. Biol.* 9:305–314.

Grigoriev, M., D. Praseuth, P. Robin, A. Hemar, T. Saison-Behmoaras, A. Dautry-Varsat, N. T. Thuang, C. Helene, and A. Harel-Bellan. 1992. A triple helix-forming oligonucleotide-intercalator conjugate acts as a transcriptional repressor via inhibition of NF-κB binding to interleukin-2 receptor α-regulatory sequence. *J. Biol. Chem.* 267:3389–3395.

Hare, D. R., and B. R. Reid. 1986. Three-dimensional structure of a DNA hairpin in solution: two-dimensional NMR studies and distance geometry calculations on d(CGCGTTTTCGCG). *Biochemistry.* 25: 5341–5350.

Helene, C. 1991. The anti-gene strategy: control of gene expression by triplex-forming-oligonucleotides. *Anticancer Drug Des.* 6:569–584.

Koshlap, K. M., P. Schultze, H. Brunar, P. B. Dervan, and J. Feigon. 1997. Solution structure of an intramolecular DNA triplex containing and N^7-glycosylated guanine which mimics a protonated cytosine. *Biochemistry.* 36:2659–2668.

Macaya, R., E. Wang, P. Schultze, V. Sklenar, and J. Feigon. 1992. Proton nuclear magnetic resonance assignments and structural characterization of an intramolecular DNA triplex. *J. Mol. Biol.* 225:755–773.

Maher, L. J., 3rd, B. Wold, and P. B. Dervan. 1989. Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation. *Science.* 245:725–730.

Maher, L. J., 3rd. 1992. DNA triple-helix formation: an approach to artificial gene repressors? *BioEssays.* 14:807–815.

Mariappan, S. V., A. E. Garcia, and G. Gupta. 1996. Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy. *Nucleic Acids Res.* 24: 775–783.

Mauffrett, O., A. Amir-Aslani, R. G. Maroun, M. Monnot, E. Lescot, and S. Fermandjian. 1998. Comparative structural analysis by [^{1}H,^{31}P]-NMR and restrained molecular dynamics of two DNA hairpins from a strong DNA topoisomerase II cleavage site. *J. Mol. Biol.* 283:643–655.

Mirkin, S. M., V. I. Lyamichev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, and M. D. Frank-Kamenetskii. 1987. DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature.* 330:495–497.

Mirkin, S. M., and M. D. Frank-Kamenetskii. 1994. H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.* 23:541–576.

Musso, M., L. D. Nelson, and M. W. VanDyke. 1998. Characterization of purine-motif triplex DNA-binding proteins in HeLa extracts. *Biochemistry.* 37:3086–3095.

Praseuth, D., A. L. Guieysse, and C. Helene. 1999. Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta.* 1489:181–206.

Radhakrishnan, I., C. de los Santos, and D. J. Patel. 1991. Nuclear magnetic resonance structural studies of intramolecular purine·purine·pyrimidine DNA triplexes in solution. *J. Mol. Biol.* 221:1403–1418.

Radhakrishnan, I., and D. J. Patel. 1994a. Solution structure and hydration patterns of a pyrimidine·purine·pyrimidine DNA triplex containing a novel T·CG base-triple. *J. Mol. Biol.* 241:600–619.

Radhakrishnan, I., and D. J. Patel. 1994b. Solution structure of a pyrimidine·purine·pyrimidine DNA triplex containing T·AT, C+·GC and G·TA triples. *Structure.* 2:17–32.

Rajagopal, P., and J. Feigon. 1989a. Triple-strand formation in the homopurine:homopyrimidine DNA oligonucleotides $d(G\text{-}A)_4$ and $d(T\text{-}C)_4$. *Nature.* 339:637–640.

Rajagopal, P., and J. Feigon. 1989b. NMR Studies of triple-strand formation from the homopurine-homopyrimidine deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$. *Biochemistry.* 28:7859–7870.

Sklenar, V., and J. Feigon. 1990. Formation of a stable triplex from a single DNA strand. *Nature.* 345:836–838.

Tarkoy, M., A. K. Phipps, P. Schultze, and J. Feigon. 1998. Solution structure of an intramolecular DNA triplex linked by hexakis(ethyleneglycol) units: d(AGAGAGAA-(EG)6-TTCTCTCT-(EG)6-TCTCTCTT). *Biochemistry.* 37:5810–5819.

Van Dongen, M. J., M. M. Mooren, E. F. Willems, G. A. van der Marel, J. H. van Bloom, S. S. Wijmenga, and C. W. Hilbers. 1997. Structural features of the DNA hairpin d(ATCCTA-GTTA-TAGGAT): formation of a G-A basepair in the loop. *Nucleic Acids Res.* 25:1537–1547.

Vigneswaran, N., C. A. Mayfield, B. Rodu, R. James, H. G. Kim, and D. M. Miller. 1996. Influence of GC and AT specific DNA minor groove binding drugs on intermolecular triplex formation in the human c-Ki-*ras* promoter. *Biochemistry.* 35:1106–1114.

Volkmann, S., J. Jendis, A. Frauendorf, and K. Moelling. 1995. Inhibition of HIV-1 reverse transcription by triple-helix forming oligonucleotides with viral RNA. *Nucleic Acids Res.* 23:1204–1212.

Wang, E., K. M. Koshlap, P. Gillespie, P. B. Dervan, and J. Feigon. 1996. Solution structure of a pyrimidine-purine-pyrimidine triplex containing the sequence-specific intercalating non-natural base D_3. *J. Mol. Biol.* 257:1052–1069.

Wells, R. D., D. A. Collier, J. C. Hanvey, M. Shimizu, and F. Wohlrab. 1988. The chemistry and biology of unusual DNA structures adopted by oligopurine-oligopyrimidine sequences. *FASEB J.* 2:2939–2949.

Wijmenga, S. S., M. M. Mooren, and C. W. Hilbers. 1993. NMR of Macromolecules: A Practical Approach. G. C. Roberts, editor. Oxford University Press, New York, NY.

Xu, Z., D. S. Pilch, A. R. Srinivasan, W. K. Olson, N. E. Geacintov, and K. J. Breslauer. 1997. Modulation of nucleic acid structure by ligand binding: induction of a DNA-RNA-DNA hybrid triplex by DAPI intercalation. *Bioorg. Med. Chem.* 5:1137–1147.

Yuhasz, S. C., N. Kanhouwa, T. Shida, and L. S. Kan. 1987. DNA synthesis: a look at sequence dependent conformational anomalies. *Biotechniques.* 5:54–59.