Solution Structure of an Intramolecular DNA Triplex Linked by Hexakis(ethylene glycol) Units: d(AGAGAGAA-$(EG)_6$-TTCTCTCT-$(EG)_6$-TCTCTCTT)[†,‡]

Markus Tarköy,[§] A. Kathryn Phipps,[§] Peter Schultze, and Juli Feigon*

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, California 90095-1569

Received November 17, 1997; Revised Manuscript Received January 30, 1998

ABSTRACT: A DNA molecule was designed and synthesized with three octanucleotide stretches linked by two hexakis(ethylene glycol) chains to form an intramolecular triplex in solution. The structural data obtained from a series of NMR NOESY spectra yielded interproton distances, and COSY experiments provided dihedral angle information for analysis of deoxyribose ring pucker. Using distance geometry followed by simulated annealing with restrained molecular dynamics and relaxation matrix refinement, a well-refined ensemble of conformations was calculated. Although some NOE cross-peaks involving protons of the hexakis(ethylene glycol) linker could be identified, most could not be assigned and the conformations of the linkers were not determined. The deoxyribose conformations are predominantly of the S type, except for the protonated cytosine residues in the third strand which show hybrid N and S character. Overall, the duplex part of the molecule resembles a B-DNA double helix with the third strand bound in its major groove by Hoogsteen hydrogen bonds. This structure provides a basis for comparison with triplexes containing noncanonical or nonnatural nucleotides.

Triple-stranded DNA was discovered nearly four decades ago by Felsenfeld et al. (*1*), who observed that poly(A) and poly(U) not only can form duplex structures but also are able to bind a second poly(U) strand to form a triplex. In recent years, there has been a resurgence of interest in these unusual structures because of their potential as therapeutics targeted to control regions in the genome (*2*, *3*). The inhibition of mRNA translation (antisense strategy, duplex formation) or of the transcription of a DNA segment (antigene strategy, triplex formation) resulting in the suppression of "undesired" gene fragments is an attractive therapeutic strategy because the mechanism for the interaction of the oligonucleotide with the target sequence is well understood and targeting is highly sequence specific. However, problems with this approach include the lack of resistance of these compounds to enzymatic degradation, length requirements for sufficient thermodynamic stability and specificity of the formed adducts, and their transport across the cell membrane (*2*). To overcome these limitations, oligonucleotides with modified backbones, bases, or sugar moieties which can form either stable duplex or triplex structures have been investigated as well (*2*−*5*). It has been demonstrated that the binding of sequence specific proteins (i.e. restriction endonucleases) can be blocked by triple strands (*4*, *6*). In addition to their potential therapeutic applications, triplexes may also have a natural role in vivo in transcriptional regulation and in replication (*7*, *8*). Purine-rich mirror repeats of DNA in topologically constrained supercoiled plasmids can form H-DNA (*9*), which is the proposed in vivo structure composed of both triplex and looped out single strand.

Two general classes of triplexes have now been characterized: the parallel motif, composed of two pyrimidines and one purine strand for forming canonical T·AT and C^+·GC triplets with the third pyrimidine strand in an orientation parallel to that of the central purine strand, and the antiparallel motif, in which the third strand consists primarily of purines and is antiparallel to the central purine strand (*10*, *11*). Although the general characteristics of parallel motif triplexes, i.e. binding of the third strand in the major groove via Hoogsteen base pairs, are now well established, the details of the triple-helix conformation are less well elucidated. From early fiber diffraction studies of DNA triplexes, it was concluded that they formed A-form helices (*12*). However, NMR (*10*, *13*) and IR (*14*) studies on intermolecular triplexes have shown that the sugar conformations do not generally conform to the N type. After it was shown that intramolecular triplexes could be formed from a single DNA strand of the appropriate sequence with DNA nucleotides as linkers in the loops (*15*), it became possible to study triplex structure in solution in more detail. Analysis of coupling constants for sugars in pyrimidine-purine-pyrimidine intramolecular triplexes (parallel motif) indicated that the sugar puckers for all but some of the cytosines in the Hoogsteen paired strand are predominately of the S type, as in B-DNA (*13*). A number of NMR-based structures of parallel motif DNA triplexes have appeared in the literature but contain other triplets or base pairs besides the canonical T·AT and C^+·GC triplets (*17*−*21*) or have not been refined in high detail (*16*). No X-ray crystal structures of parallel motif DNA triplexes have been published to date.

[†] This work was supported by NIH Grant GM 37254 to J.F.

[‡] Coordinates for these structures have been deposited in the Brookhaven Protein Data Base (entry 1D3X).

* To whom correspondence should be addressed.

[§] These two authors contributed equally to this work.

S0006-2960(97)02810-9 CCC: $15.00
Published on Web 03/24/1998

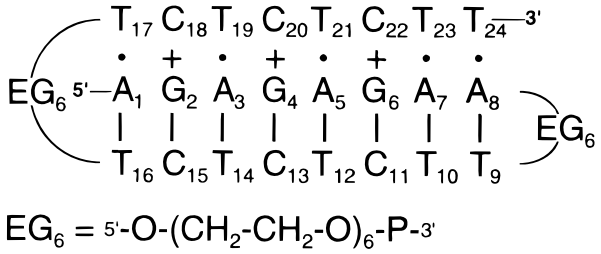

FIGURE 1: Sequence and base pairing scheme of DDD-EG. Lines between strands symbolize Watson−Crick base pairing, and bullets denote Hoogsteen-type pairing. Cytidines 18, 20, and 22 are protonated. $(EG)_6$ stands for the ethylene glycol linker [5′-$(OCH_2$-$CH_2)_6$O-P3′].

We have used two-dimensional (2D) NMR spectroscopy to gain insight into the structural characteristics of the intramolecular pyrimidine-purine-pyrimidine DNA triplex d(AGAGAGAA-$(EG)_6$-TTCTCTCT-$(EG)_6$-TCTCTCTT) (DDD-EG), where the loops between the Watson−Crick paired duplex and between the two pyrimidine strands are each replaced by one hexakis(ethylene glycol) unit (Figure 1). It has been shown that the loop configuration where the 5′-end of the third (Hoogsteen paired) strand is linked to the duplex oligonucleotide yields an overall stabilization of about 2 K in the T_m value and about 0.4 kcal mol^{-1} in free energy compared to the case where the third strand is linked via its 3′-end (*22*). This observation is consistent with the fact that in H-DNA a 5′-type bridging loop is preferred over a 3′-type bridging loop as well (*23*). An intramolecular triplex structure containing a triplex−duplex junction and linked by two hexakis(ethylene glycol) units but set up in a 3′-type bridging manner has been reported (*17*). The use of a synthetic loop decreases the size of the investigated oligonucleotide by at least eight nucleotides which reduces spectral overlap and facilitates the interpretation of NMR data (*24*). The flexible ethylene glycol linkers also ensure that the helical conformation is minimally perturbed by constraints from the loops. Additionally, such a substitution increases the resistance of triplex systems to enzymatic degradation.

Here we present the refined three-dimensional solution structure of the intramolecular DNA triplex formed by d(AGAGAGAA-$(EG)_6$-TTCTCTCT-$(EG)_6$-TCTCTCTT) (DDD-EG). The structure was calculated from a set of distance geometry starting structures and refined by restrained molecular dynamics. The 10 best structures were further refined with relaxation matrix calculations. The structure of this triplex, which is composed solely of canonical T·AT and C^+·GC triplets, provides a point of reference for studies on similar triplexes containing modified nucleotides in the third strand (*25*, *26*).

MATERIALS AND METHODS

Sample Preparation. The oligonucleotide d(AGAGAGAA-$(EG)_6$-TTCTCTCT-$(EG)_6$-TCTCTCTT) was synthesized on a ABI 381A DNA synthesizer using the standard phosphoramidite chemistry protocol (*27*), except that the coupling time for the hexakis(ethylene glycol) linker was extended from 25 s to 10 min. The deoxyribonucleotide phosphoramidites and the hexakis(ethylene glycol) linkers were purchased from Pharmacia and Glen Research, respectively. After deblocking with concentrated aqueous ammonia, the crude sequence was purified by column chromatography on Sephadex G50 as previously described (*28*). The final NMR sample was 2 mM in strand with 100 mM NaCl (pH 5.2) and 5 mM $MgCl_2$ in 200 μL (Shigemi NMR tube). This one sample was used for both H_2O and D_2O experiments, where the solvent was exchanged between 90% H_2O/10% D_2O (H_2O experiments) and 99.996% D_2O (D_2O experiments) by drying the sample under a stream of N_2(g) in the NMR tube as needed.

NMR Spectroscopy. The NMR spectra were acquired at 500 MHz on GE GN500 and Bruker AMX500 spectrometers. The measuring temperature for spectra in D_2O was 298 K, and for spectra taken in H_2O, it was 274 K. NOESY spectra in D_2O were acquired with mixing times of 40, 80, 100, 140, 200, and 250 ms. A NOESY spectrum in H_2O was acquired with a mixing time of 150 ms. Suppression of the residual water peak in D_2O spectra was accomplished by presaturation during the recycle delay and in H_2O spectra by a $1\bar{1}$ spin echo pulse sequence. All NOESY experiments (*29*, *30*) were carried out in the pure absorption mode using States−TPPI phase cycling (*31*). A HOHAHA spectrum (*32*, *33*) was obtained with a 100 ms mixing time using the MLEV17 cycle (*34*). Coupling constants were obtained from a P.COSY spectrum (*35*, *36*) with a 90° flip angle mixing pulse. A natural abundance $^1H-^{13}C$ HSQC spectrum (*37*) was also acquired for this sample to aid in the assignments. The NMR data were processed with FELIX 2.30 (Biosym) or XWIN-NMR (Bruker Instruments, Inc.).

UV Melting Studies. UV melting studies were carried out on a Varian Cary1E spectrophotometer equipped with a temperature probe. The heating rate was 0.5 °C/min. The T_m was determined from the maximum of the first numerical derivative. The DNA samples used, synthesized and purified as described above, were d(AGAGAGAA-$\underline{(EG)_6}$-TTCTCTCT-$\underline{(EG)_6}$-TCTCTCTT), d(AGAGAGAA-$\underline{GCAA}$-TTCTCTCT-$\underline{CTTTG}$-TCTCTCTT), and d(AGAGAGAA-$\underline{CCCC}$-TTCT-CTCT-$\underline{TATA}$-TCTCTCTT), where the underlined sequences indicate the loops. The sample volumes were 1 mL with $0.45-0.65A_{260}$ of DNA, and buffer conditions were 50 mM NaOAc, 0.1 M NaCl, and 5 mM $MgCl_2$ (pH 5.20).

Determination of Deoxyribose Coupling Constants and Pseudorotation Angles. The FORTRAN program CHEOPS (coupling constants from high-resolution NMR experiments by optimized parameter simulation) (P. Schultze and J. Feigon, unpublished program) (*13*) was used to simulate the H1′−H2′/H2″ region of a P.COSY spectrum by automatic optimization of the fit between the experimental and simulated spectra. The *J* coupling constants for H1′−H2′, H1′−H2″, H2′−H3′, and H2″−H3′ thus obtained were fitted with a two-state model for ribose conformation using the program PSEUROT (*38*, *39*). The relative population and the pseudorotation phase angles for both conformational states were kept variable during the optimization, whereas both amplitudes were held constant at 38° to avoid overfitting the four input parameters. The values of ν_1 and ν_2 obtained for the majority conformation were used with bounds of ±3° for dihedral angle restraints in the X-PLOR calculations.

Structure Calculations and Refinement. Nonexchangeable proton−proton distances obtained from the 200 ms mixing time NOESY experiment were used as initial restraints for the structure calculation. Cross-peaks were picked and integrated using the AURELIA software package (*40*). The resulting peak lists were assigned using a semiautomatic

procedure requiring the picked peak list, the chemical shifts, and a model of the molecule. An expanded peak list containing all possible assignments within 0.01 ppm in each dimension was generated from the chemical shift list and then compared to that of a model triplex to check for reasonable distances for each of the proposed assignments. In cases of multiple ambiguous assignments, the one corresponding to the smallest H−H distance in the model was automatically selected and marked without deleting the remaining ones. This list was inspected by hand to correct any erroneous assignments and then reimported into AURELIA for peak integration and output of distance and volume files in X-PLOR format. The distances in the D_2O spectra were calibrated using the H2′−H2″ cross-peak at 1.9 Å. For resolved cross-peaks, all lower limits were set to 1.8 Å, and the upper bounds were set to the calibrated distance plus a margin of 1.0 Å, except for the methyl groups which had a margin of 2.5 Å and H5′−H5″ which had a margin of 2.0 Å. In the case of overlapping cross-peaks, an additional 0.5 Å was added to the existing upper bounds. Peaks that consistently gave rise to violations in initial rounds of calculation were re-examined; misassigned peaks were removed, while correctly assigned peaks were checked for the possibilities of errors due to spin diffusion or overlap. For the exchangeable protons, a lower bound of 2 Å and a uniform upper bound of 5 Å were used, to account for inaccuracies in peak integrals due to exchange with water and the $\sin^3$ excitation profile of the $1\bar{1}$ spin echo read pulse.

In the initial step of X-PLOR Version 3.1 (*41*) calculations, 30 coordinate sets were obtained by metric matrix distance geometry embedding of all atoms. Next, the initial structures were subjected to the distance geometry−simulated annealing (DGSA) procedure (*42*), which consists of template fitting for improving local geometry and simulated annealing. To increase the number of converging structures to approximately 90% (in this case, 27 out of 30), the weight of the bond angle potential was increased 8-fold over the standard X-PLOR force field, and dihedral angle restraints involving the four ligands of each chiral center were introduced to enforce the correct chiral configurations. As described previously (*20*), the ceiling of the energy term for the distance restraints was set to 100, i.e. $^1/_{10}$ of the value in the published protocol (*41*). In the third step, those modified parameters were returned to their standard values for further refinement by molecular dynamics and energy minimization based on distance restraints. Standard hydrogen bond distances were used to restrain the geometry of the base triplets. In addition, planarity restraints with a low weight factor of 3 kcal/Å^2 were applied to each triplet in all refinement steps. This favors overall planarity of the triplets if no specific distance restraints cause out-of-plane tilting of particular bases.

All 30 initial structures were carried through to the third refinement step. These structures were then sorted according to overall energy terms. The first 10 coordinate sets were subjected to relaxation matrix refinement, which consisted of a short simulated annealing protocol under inclusion of the NOE potential. The nonexchangeable restraints were substituted by the cross-peak volumes for each of four NOESY spectra acquired with mixing times of 40, 80, 140, and 200 ms, while the exchangeable and hydrogen bond restraints were left as distances. Only cross-peaks that gave good integrals at all four mixing times were included at this step; i.e. overlapping cross-peaks were discarded.

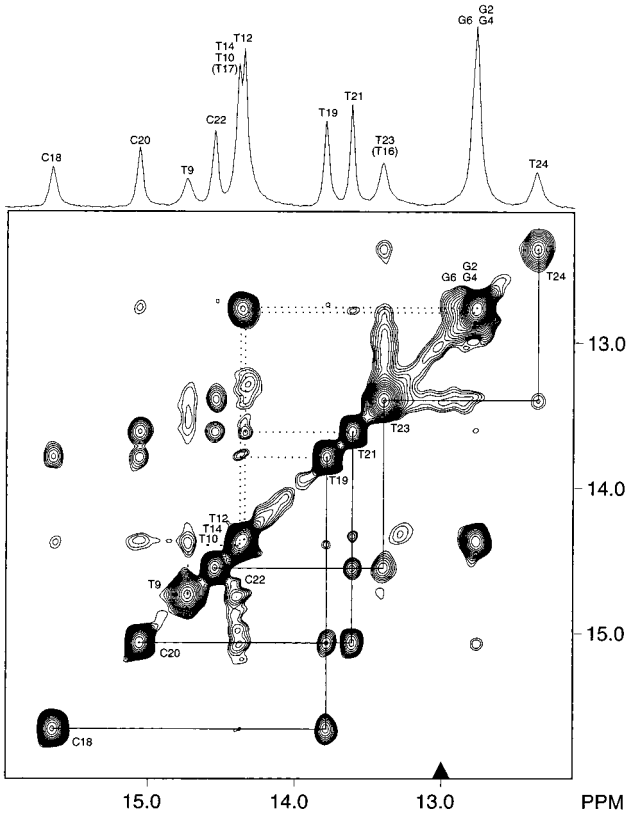

FIGURE 2: Portion of a NOESY in 90% H_2O/10% D_2O at 1 °C and a mixing time of 150 ms containing the exchangeable imino proton resonances. Assignments are indicated above the spectrum. The lines below the diagonal of the spectrum show the sequential connectivities in the Hoogsteen paired strand. The dashed lines denote connectivities between T9 and T10 and interstrand NOEs between T14 and T19, T12 and T21, T14 and G4, and T12 and G6. A total of 2048 and 200 complex points were collected in t_2 and t_1, respectively, with 96 scans per t_1 block. The sweep width was 11 364 Hz, and the excitation maximum for the $1\bar{1}$ echo delay was at 13 ppm (▲). The processed matrix consisted of 2048 × 2048 complex data points. The FIDs in t_2 and t_1 were apodized by a Gauss function with a Lorentzian broadening of −21 and −25 and a Gaussian broadening of 0.15 and 0.18, respectively.

The helical parameters were calculated for the Watson−Crick duplex portion of the triplex structures using the program CURVES (version 5.1) (*43*). These values given are averaged over the 10 relaxation matrix refined structures and have been calculated using nonlinear and linear helical axes for comparison.

RESULTS

Exchangeable Proton Resonances and Assignments. The NOESY spectrum of DDD-EG in a H_2O at 1 °C shows characteristics similar to those of other previously described intramolecular triplexes (*15*). An expanded region of the exchangeable imino proton region is shown in Figure 2. Sequential imino−imino connectivities can be traced through the Watson−Crick paired strands from T9 to G2 and through the Hoogsteen paired strands from T17 to T24. Resonances for nucleotides T16 and T17 which are part of the first triplet are overlapped. The protonated cytosine residues are easily identified from their strong imino−amino cross-peaks and the characteristic downfield shift of their amino resonances

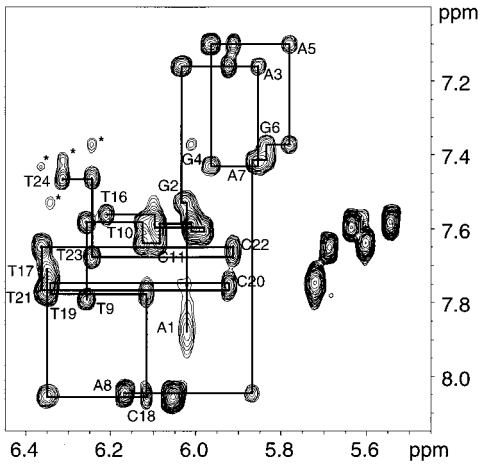

FIGURE 3: Portion of a NOESY spectrum in D_2O at 25 °C with a mixing time of 200 ms showing the aromatic−H1′/H5 cross-peaks. Solid lines indicated sequential base−H1′ connectivities from A1 to A8, T9 to T16, and T17 to T24. The strong cross-peaks between 5.8 and 5.5 ppm are the H6−H5 NOEs from the cytosines. Peaks denoted with an asterisk are interstrand NOEs between purine H8 and H1′ of the neighboring residue (3′-direction) in the Hoogsteen strand. The sweep width was 5000 Hz in both dimensions; 2048 and 512 complex points were collected in t_2 and t_1, respectively, with 48 scans per t_1 increment. The final data matrix was 2048 × 2048 complex points; t_2 and t_1 were apodized with a Lorentzian broadening of −15 and Gaussian broadening of 0.15 and 90° qsine, respectively, and the baseline was flattened with a first-order polynomial.

in comparison to those of the other amino groups. Very weak interstrand imino cross-peaks are also observed between T12−T21 and T14−T19 which help to confirm the assignments. Some interstrand cross-peaks from the Hoogsteen strand to the amino protons of C15 and C19 are also visible (not shown).

Sequential Assignment of Nonexchangeable Protons. Assignments of the nonexchangeable proton resonances in DDD-EG were obtained by tracing the sequential base−H1′ connectivities for each of the three strands, as previously described (*16*). The identification of the CH5−CH6 and TMe−TH6 cross-peaks (obtained from a HOHAHA spectrum, not shown) provides a good starting point for the assignment. The assignment process was aided by the information obtained from a natural abundance $^1H-^{13}C$ HSQC spectrum (not shown). The distinct carbon chemical shifts allowed for unambiguous identification of AH2 resonances and for obtaining the upper and lower ^1H chemical shift values for deoxyribose regions that overlap, e.g. H4′ and H5′−H5″. Figure 3 shows the expanded aromatic−H1′/H5 region of a NOESY spectrum in D_2O at 25 °C. Sequential connectivities can be traced along each strand, from A1 to A8, T9 to T16, and T17 to T24. Some interstrand cross-peaks are observed between H1′ protons of the Hoogsteen strand and the purine H8 protons in the 5′-neighboring triplet, e.g. T24−A7. Two relatively strong cross-peaks in this category are C20H1′−A3H8 and C22H1′−A5H8. These interstrand peaks are particularly important in the structure determination because they constitute proof of the actual formation of the intramolecular triplex topology as designed. There are also interstrand contacts between the Watson−Crick duplex strands, between the AH2 protons of the purine strand and H1′ protons of the pyrimidine strand in the 5′-neighboring triplet, e.g. C13−A5.

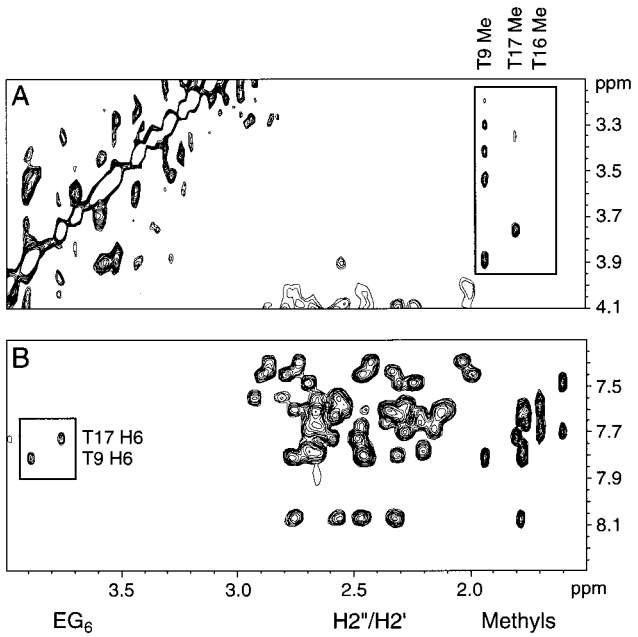

FIGURE 4: Portions of a NOESY spectrum showing cross-peaks between base protons and ethylene protons of the glycol linker. In panel A, cross-peaks between the EG and T9Me, T16Me, and T17Me groups are boxed. In panel B, cross-peaks between the 3′-terminal glycol protons and the 5′-terminal T9H6 and T17H6 are boxed. The spectrum is the same as Figure 3.

Sequential connectivities can also be traced in the aromatic−H3′ and aromatic−H2′/H2" regions of the NOESY spectrum (data not shown), and confirmed the assignments. The assignments could be further confirmed by analysis of the aromatic−aromatic region (not shown). In this triplex, NOEs are resolved between all consecutive nucleotides in the purine strand except for A1 and G2 and G6 and A7. Some unusual, very weak sequential cross-peaks between H1′ resonances are observed between G2 and A3, G4 and A5, A7 and A8, T17 and C18, and T19 and C20, even in an 80 ms mixing time 2D NOESY experiment. Sequence specific assignments were obtained for almost all the nonexchangeable and exchangeable proton resonances except the H5′−H5″ (Table 1).

Ethylene Glycol Linker Resonances. Most of the protons from the hexakis(ethylene glycol) linkers resonate in a resolved region from 2.6 to 3.9 ppm, between the H5′/H5″ and H2′/H2″ regions (not shown). The glycol methylene groups next to the ends of the nucleotide strands, however, are shifted upfield compared to the others and resonate slightly downfield of the A1 H5′/H5″ peaks, which are themselves slightly upfield of the other H5′/H5″ resonances due to the presence of the 5′-terminal hydroxyl group. In contrast to previous reports on ethylene glycol-linked triplexes (*17*, *24*), NOE contacts from the protons on the DNA to the methylene groups of the ethylene glycol chain are observed for this triplex. The alkyl chains of the glycols attached to the 5′-ends of the nucleotide strands have NOE cross-peaks to the H6 and methyl resonances for the terminal residues T9 and T17 and can be used in confirming those assignments (Figure 4). Weak cross-peaks from the glycols attached to the 3′-ends of the nucleotide strands, A8 and T16, can also be seen in the H1′ region of the DNA (not shown). There are also cross-peaks between the various glycol proton resonances. However, many of these NOE cross-peaks are

Table 1: Proton Chemical Shifts of DDD-EG (Parts per Million) at 25 °C

residue	H6, H8	H2, H5, Me	H1′	H2′	H2″	H3′	H4′	imino[a]	amino[a]
A1	7.87	na	6.02	2.63	2.79	4.97	4.17		na
G2	7.53		6.03	2.55	2.90	5.06	4.30	12.75	na
A3	7.16	7.52	5.85	1.97	2.77	4.79	4.59		7.44, 7.85
G4	7.43		5.96	2.44	2.88	4.91	4.32	12.75	na
A5	7.10	7.37	5.78	2.01	2.71	4.73	4.53		7.46, 7.90
G6	7.37		5.83	2.40	2.85	4.83	4.35	12.77	na
A7	7.41	7.58	5.87	2.32	2.72	4.89	4.29		7.72, 7.94
A8	8.04	8.02	6.16	2.45	2.54	4.87	4.34		na
T9	7.79	1.91	6.26	2.42	2.73	4.90	4.27	14.72	
T10	7.58	1.74	6.12	2.34	2.67	4.91	4.22	14.39	
C11	7.64	5.60	6.08	2.19	2.64	4.76	4.26		7.39, 8.58[b]
T12	7.58	1.73	6.01	2.31	2.63	4.90	4.22	14.33	
C13	7.58	5.54	6.01	2.11	2.56	4.74	4.17		7.18, 8.36[b]
T14	7.61	1.76	5.98	2.28	2.56	4.88	4.17	14.38	
C15	7.59	5.63	6.10	2.08	2.51	4.73	4.15		7.21, 8.35[b]
T16	7.56	1.68	6.21	2.25	2.52	4.83	4.14	na	
T17	7.70	1.78	6.35	2.46	2.57	4.93	4.37	na	
C18	8.05	6.05	6.11	2.30	2.76	4.62	4.37	15.64	9.32, 10.17[b]
T19	7.77	1.76	6.34	2.45	2.68	4.94	4.37	13.77	
C20	7.74	5.72	5.92	2.18	2.65	4.63	4.37	15.05	9.13, 9.99[b]
T21	7.76	1.76	6.36	2.44	2.62	4.94	4.37	13.60	
C22	7.65	5.69	5.91	2.13	2.61	4.68	4.41	14.53	9.86, 10.11[b]
T23	7.67	1.67	6.24	2.26	2.67	4.91	4.27	13.39	
T24	7.46	1.58	6.31	2.22	2.29	4.51	4.07	12.35	

[a] Chemical shifts at 1 °C. [b] Lower field resonance is the hydrogen-bonded amino proton H4(1). [c] na, not assigned.

unresolved, and we were unable to specifically assign the glycol resonances.

Deoxyribose Conformation. Figure 5 shows the H1′−H2′/H2″ cross-peak region of the experimental and the simulated P.COSY spectrum of the DNA triplex. Acceptable correlation coefficients ranging from 90 to 98% were obtained for 18 of the 19 DNA residues (of 24 total) that were sufficiently resolved to be accurately simulated. On the basis of the values obtained from the PSEUROT program, all 18 of the sugars were predominantly in the S-type conformation (Supporting Information Table 1), and dihedral restraints for ν_1 and ν_2 were used accordingly in the structure calculations. The values obtained for the sugar conformations are in good agreement with those obtained on an intramolecular triplex of the same sequence but with nucleotide loops (*13*).

Thermal Stability. The stability of the triplex structure discussed here was compared to those of other triplexes differing only in the loop region, by monitoring the change of absorbance as a function of temperature. Figure 6 shows the melting curves and numerical first derivatives of three triplexes which differ only in their loop sequences. The optical melting curves of DDD-EG with ethylene glycol linkers are compared with those of two triplexes containing Watson−Crick CCCC and Hoogsteen TATA loops (*44*) and Watson−Crick GCAA and Hoogsteen CTTTG loops. The GCAA Watson−Crick linker was chosen on the basis of evidence that it forms an especially stable DNA tetraloop (*45*) and the CTTTG Hoogsteen linker on the basis of loop stability studies of Kool and co-workers (*22*). The T_m values derived for the DDD-EG, CCCC/TATA, and GCAA/CTTTG triplexes were 68.7, 68.2, and 72.8 °C, respectively. This result is in agreement with findings on a series of triplexes with $(EG)_{3-8}$ loops between two pyrimidine domains targeting a purine strand (*46*).

Structure Calculations. A total of 573 NOE restraints were used to calculate the structure, and 36 torsion angle restraints and 70 hydrogen bond restraints. Although the

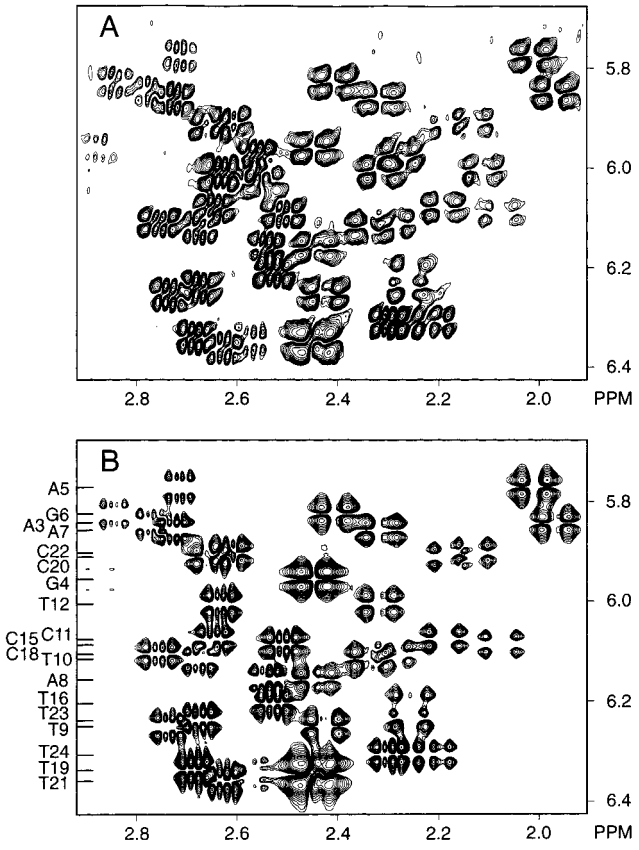

FIGURE 5: (A) Portion of a P.COSY spectrum in D_2O at 25 °C showing the H1′−H2′/H2″ cross-peaks. (B) Same region of a simulated spectrum calculated using CHEOPS (see Materials and Methods). Assignments of the simulated resonances are indicated on the left border of the spectrum. The P.COSY spectrum was acquired with a flip angle mixing pulse of 90, spectral width of 5000 Hz in both dimensions, 1024 and 420 complex points collected in t_2 and t_1, respectively, and eight scans per t_1 increment. Both dimensions were apodized by a 65° phase-shifted qsine bell function with a skew factor of 1.0 and zero-filled to 4096 complex points.

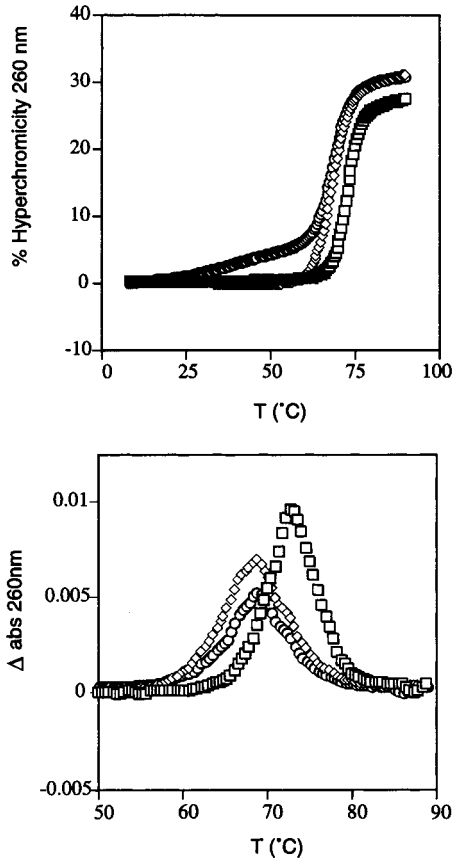

FIGURE 6: Optical denaturation curves (A_{260}) and first derivatives of different triplex structures differing only in the loop region. Buffer conditions were 50 mM NaOAc, 0.1 M NaCl, and 5 mM $MgCl_2$ (pH 5.20). The triplex sequence is 5′-AGAGAGAA-loop$_{WC}$-TTCTCTCT-loop$_{HS}$-TCTCTCTT-3′. The loop between the purine strand and the pyrimidine strand is denoted as Watson−Crick (WC) and the one between pyrimidine and pyrimidine as Hoogsteen (HS) loop: loop$_{WC}$ GCAA and loop$_{HS}$ CTTTG (□), loop$_{WC}$ $(EG)_6$ and loop$_{HS}$ $(EG)_6$ (◇), and loop$_{WC}$ CCCC and loop$_{HS}$ TATA (○) (this sample was provided by R. Macaya).

triplex with $(EG)_6$ loops eliminated some of the overlap in the spectra, there were still many overlapping cross-peaks,

Table 2: Input and Results of Structure Refinement

NOE restraints	573
intranucleotide	344
sequential	183
nonsequential internucleotide	46
involving exchangeable protons	46
torsion angle restraints	36
hydrogen bond restraints	70
relaxation matrix refinement	
number of peak integrals used at each mixing time	223
average $R^{1/6}$ factor before	0.1133 ± 0.0017
average $R^{1/6}$ factor after	0.0578 ± 0.0016
refinement statistics (10 final lowest-energy structures)	
NOE violations >0.5 Å	0
dihedral angle violations >5°	0
average pair wise rmsd (all heavy atoms) (Å)	1.25 ± 0.25
average rmsd from ideal covalent geometry	
bond lengths (Å)	0.021
bond angles (deg)	5.53

and these peaks had to be accounted for in the structure calculations. To make sure that the use of overlapped cross-peaks in the initial calculations was not distorting the structure, test calculations were run without using any of the distances derived from overlapped cross-peaks. These runs still gave the same overall structure as the runs using all the cross-peaks. The top 10 refined structures, using the full data set, were then used for the relaxation refinement. This led to significant improvement of the agreement between calculated and observed peak integrals, as shown by the decrease of the $R^{1/6}$ factor by a factor of 2 (Table 2). However, a comparison of the structures before and after relaxation matrix refinement shows only small local changes (Figure 7). Thus, it appears that the calibration method and conservative error bounds used to obtain distance restraints already yield a reasonably good approximation of the final structures obtained by taking into account spin diffusion during relaxation matrix refinement. Furthermore, the small local changes which are observed after the relaxation matrix

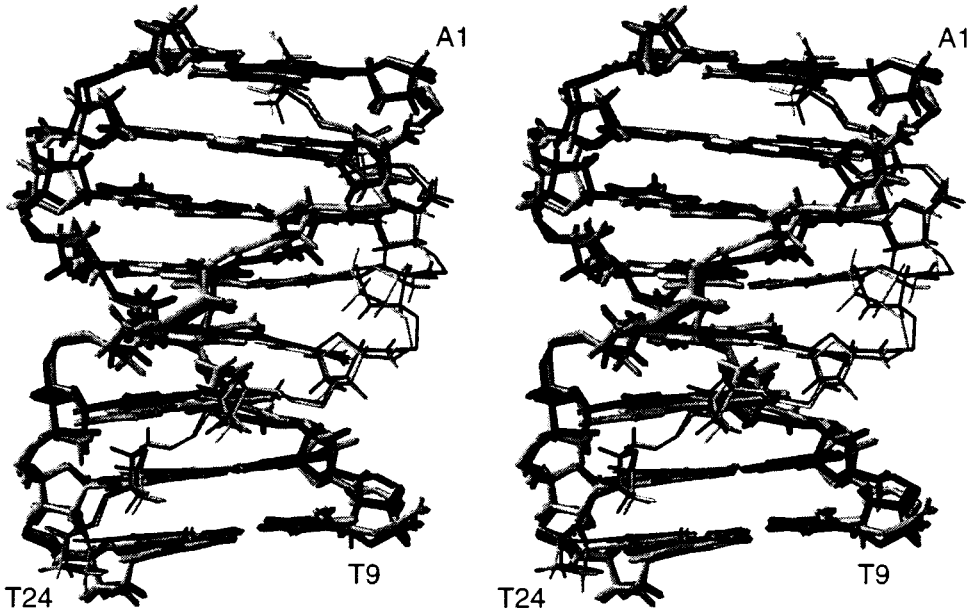

FIGURE 7: Stereoview of a superposition of the mean structures calculated for DDD-EG before (black) and after (gray) relaxation matrix refinement. The Hoogsteen paired strand is shown with thinner lines. All atoms except O1P and O2P are shown.

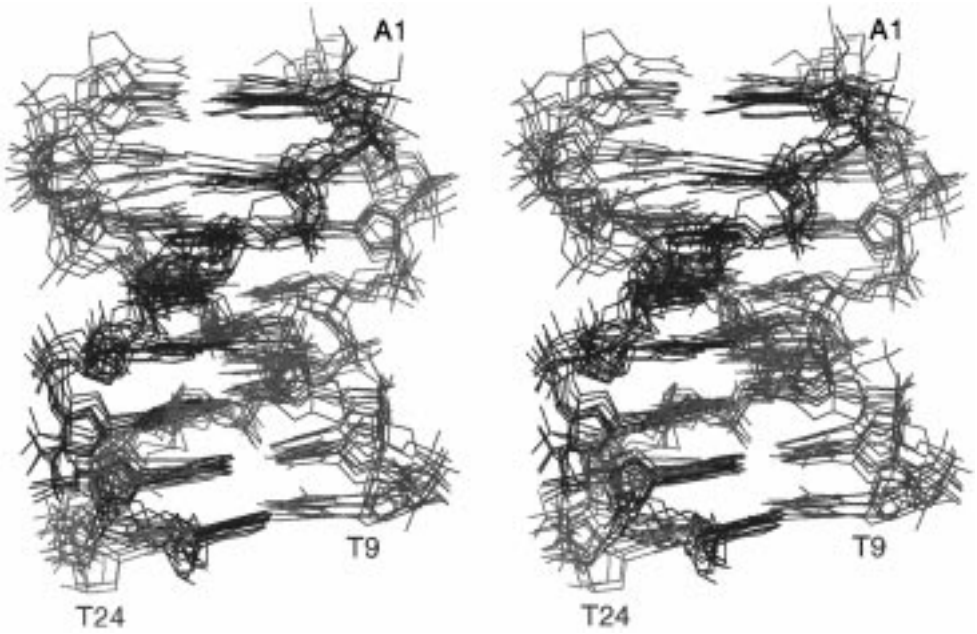

FIGURE 8: Stereoview of a superposition of the eight lowest-energy structures of DDD-EG. The view is into the major groove. The Watson−Crick paired pyrimidine strand (T9−T16) is green, the purine strand (A1−A8) blue, and the Hoogsteen paired pyrimidine strand (T17−T24) cyan. Only the DNA nucleotides are shown, since the ethylene glycol linkers were not included in the structure calculation.

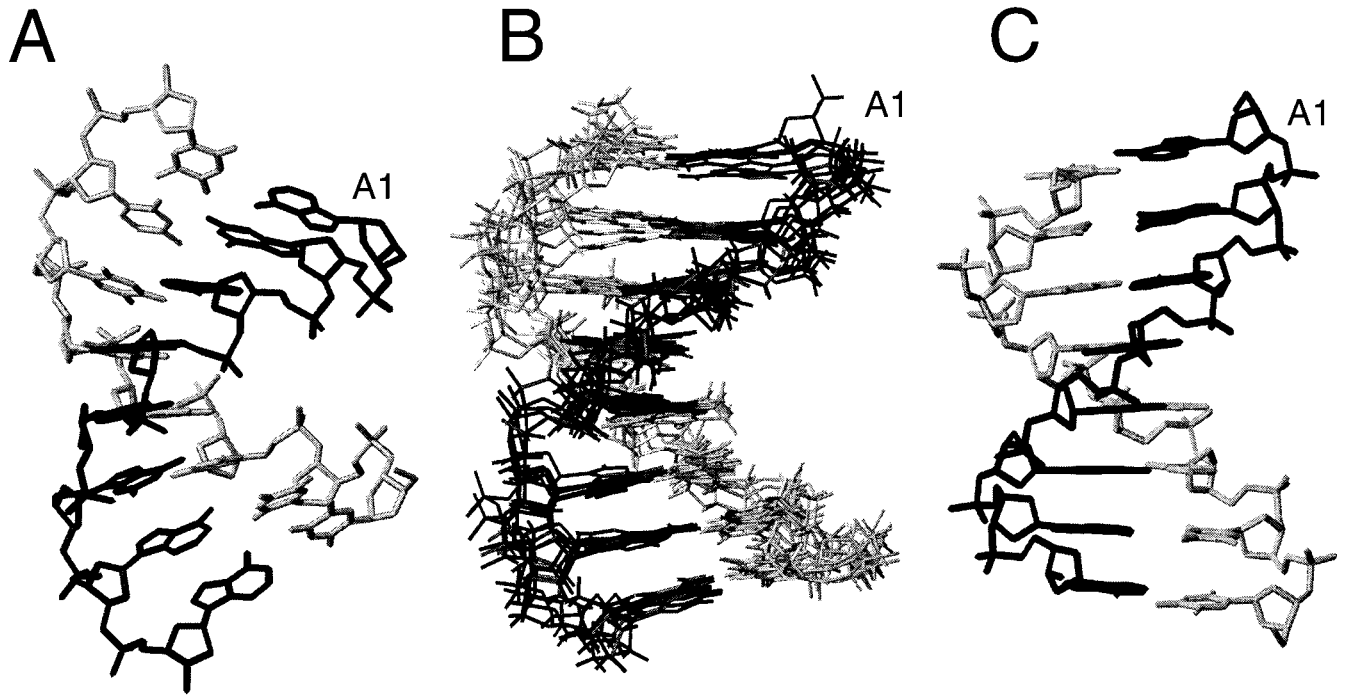

FIGURE 9: Comparison of the structures of (A) A-DNA, (B) Watson−Crick paired strands of DDD-EG, and (C) B-DNA. The purine strand is black and the pyrimidine strand gray. The view is with minor groove in front at the top and the major groove in front at the bottom. The standard A-DNA and B-DNA helices were constructed from fiber diffraction-derived coordinates (*12*) using Insight II software (Biosym Technologies). The Watson−Crick paired strands of the five lowest-energy structures of the DNA triplex are shown.

refinement may in fact have arisen partially from distance errors due to cross-peak overlap in the initial calculations.

DISCUSSION

Triplex Formation and Stability. In this work, we have studied the structure of an intramolecular DNA triplex with ethylene glycol linkers. Triplex formation was initially assayed both from analysis of the UV melting profile and from analysis of the imino and amino proton resonances observed in NMR spectra of the sample in H_2O. Resonances and NOE cross-peaks in the NMR spectra of the molecule are characteristic of formation of an intramolecular triplex. Imino resonances are observed for each of the base pairs in both the Watson−Crick and Hoogsteen paired strands, and sequential imino−imino connectivities can be traced. The UV melting study at pH 5.2 shows that the DDD-EG triplex melts in a single transition under the conditions studied. As can be seen in Figure 6, the stabilities as measured by T_m for the triplex with the ethylene glycol linkers ($T_m = 68.7$) and the CCCC/TATA loops ($T_m = 68.2$) are nearly the same.

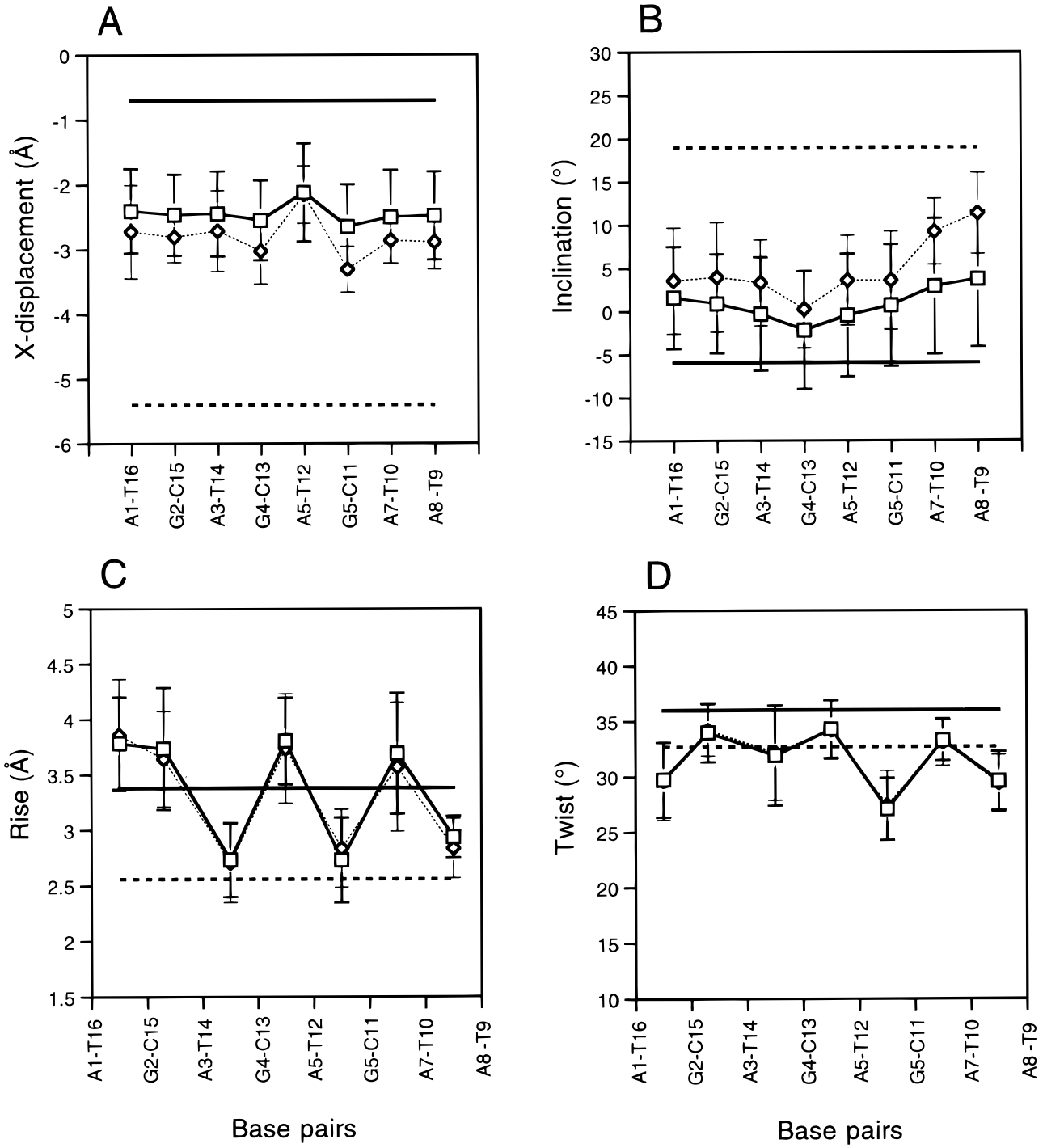

FIGURE 10: Helical parameters for DDD-EG: (A) X displacement, (B) inclination, (C) rise, and (D) twist. Error bars indicate the standard deviation of the measurement from the 10 lowest-energy structures. Values using both nonlinear (squares and solid line) and linear (diamonds and dotted line) axes are plotted. The values from fiber diffraction models for standard A-DNA and B-DNA (*12*) are indicated by dashed and solid lines, respectively.

However, the latter triplex shows a broad sloping transition before the sharp duplex melting, indicating that it may dissociate in two steps. The DDD-EG triplex and the GCAA/CTTTG triplex only display one sharp melting transition. The stable GCAA/CTTTG nucleotide loops result in an increase of about 4 °C in the T_m compared to those of the other two triplexes.

Comparison to Spectra of Triplexes with Nucleotide Loops. Triplexes with the same triplet core AGAGAGAA, but with the nucleotide loops (1) CCCC/TTT (*16*), (2) CCCC/TATA (*44*), and (3) GCAA/GTTTC (unpublished results), have chemical shifts for the triplet resonances that are similar to those of DDD-EG. The advantage of using the ethylene glycol loops is that it not only simplifies the spectra by eliminating the seven to nine nucleotides of the loops but also results in better chemical shift dispersion of the nucleotides to which the glycol moieties are attached (*24*). This is because the glycols create an environment at the end of each nucleotide "strand" (residues 8, 9, 16, and 17) similar to that for terminal residues 1 and 24 which have only one nucleotide neighbor. The triplexes with nucleotide loops have only two terminal residues, and therefore, the ends of

the triplex core can sometimes be more difficult to identify, especially if the sequential connectivity is broken somewhere in the loop. Cross-peaks from these nucleotide residues are also more obscured by overlap, which made it more difficult to calculate an accurate structure.

Triplex Structure. A superposition of the lowest-energy structures of the DNA triplex is presented in Figure 8. Structure statistics for the 10 relaxation matrix refined structures are summarized in Table 2. The average pairwise rmsd for all heavy atoms for the 10 structures is 1.25 ± 0.25 Å. The single-stranded oligodeoxynucleotide folds into a well-defined, right-handed triple-helical structure. The Watson−Crick paired strands of the triplex form a duplex that is most similar to an underwound B-form DNA helix.

Visually, the Watson−Crick strands of the triplex fit well to the overall structure of ideal B-form DNA (Figure 9). It is not surprising that there is not a perfect fit between these two molecules, since the B-form is based on a continuous DNA fiber while the first two strands of the triplex not only are finite but also are bound to loops which might have an effect at the terminal nucleotides. The third strand of the triplex can be thought of as a ligand which binds in the major groove of the duplex DNA (target molecule). To accommodate the third strand, one would expect that the duplex strands will be displaced somewhat from the ideal B-DNA structure. A slightly widened minor groove width (close to the width of A-form DNA) compared to that of duplex B-DNA is observed with this triplex (average twist value of 31°), which is even less than the A-DNA average of 32.7°. This is due to the unwinding of the duplex which accommodates the third strand in the major groove and minimizes the electrostatic repulsion of the negative charge introduced from the third strand's phosphate backbone.

To analyze the structure and obtain a more quantitative comparison to A-DNA and B-DNA helices, helical parameters for the Watson−Crick duplex part of the triplex were calculated. Plots of the helix parameters as a function of nucleotide position are shown in Figure 10. Since somewhat different results are obtained if one calculates the values using a nonlinear axis or a linear axis using CURVES, values obtained both ways are presented for comparison (*25*), although they are very similar for DDD-EG. The X displacement of the Watson−Crick base pairs, one of the most characteristic parameters for the distinction between A- and B-DNA, varies from about −2.1 to −2.6 Å with an average of −2.4 ± 0.7 Å (−2.81 ± 0.5 Å for linear). The X displacement is about halfway between that of standard A-DNA and B-DNA duplexes. No axial hole is observed for DDD-EG, which clearly distinguishes this structure from a pure A-DNA conformation. The rise per residue ranges from 2.7 to 3.8 Å, with an average of about 3.35 ± 0.40 Å (3.31 ± 0.43 Å for linear). The helical twist varies from 27 to 34°, with an average of 31.4 ± 2.9° (31.4 ± 2.9° for linear), which is lower than that for B-DNA due to unwinding of the duplex which accommodates the third strand. The inclination varies from about −2.1 to 3.8° with an average closer to that of B-DNA than that of A-DNA.

In addition to having helical parameters close to those of a B-form DNA, the restrained sugars of the triplex also have predominantly S-type puckers (near C2′-endo). The final ensemble of structures has pseudorotation phase angles ranging from 108 to 189° as computed by CURVES. Four of the unrestrained sugars also fall within this range, while C13 and C20 had pseudorotation angles <108°. Although the unrestrained sugars are in a wide range of sugar conformations, additional NMR data such as the lack of downfield-shifted C1′ resonances (*47*) suggest that these sugars are more likely in the same conformational range as the rest of the residues. The sugar conformations agree with results from previous NMR and FTIR studies (*14*, *48*) that show that triplexes have been found to have predominantly S-type sugars with some N-like character observed in triplexes containing C^+·GC base triplets. The three protonated cytosine residues show a relatively low ratio of S/N conformation according to the PSEUROT analysis. This is consistent with the general observation that protonation of bases shifts the equilibrium toward the N-type (C3′-endo) conformation due to an increase in the anomeric effect (*49*−*51*). Therefore, we conclude that the conformation of the third strand is essentially B-form, and that the higher %N conformation of the protonated C's is due to the anomeric effect rather than to a structural effect of triplex formation.

The high-resolution structure of an unmodified pyrimidine motif triplex reported here provides a basis for comparison of helical parameters to those of other triplexes containing modified or ribonucleotides in the third strand (*25*, *26*) or noncanonical triplets (*20*, *21*). A detailed comparison of this DNA triplex to one with the identical sequence, except that the thymine methyl groups in the Hoogsteen paired strand are replaced with propynes, is presented in the following paper (*25*).

NOTE ADDED IN PROOF

While this paper was under review, a similar intramolecular DNA triplex structure was published (*52*).

SUPPORTING INFORMATION AVAILABLE

Coupling constraints and correlation coefficients for DDD-EG obtained using CHEOPS and sugar conformations obtained using PSEUROT (1 page). Ordering information is given on any current masthead page.

REFERENCES

1. Felsenfeld, G., Davies, D. R., and Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023−2024.
2. Uhlmann, E., and Peyman, A. (1990) *Chem. Rev. 90*, 543−584.
3. Soyfer, V. N., and Potaman, V. N. (1996) *Triple-Helical Nucleic Acids*, Springer-Verlag, New York.
4. Thuong, N. T., and Hélène, C. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 666−690.
5. Neidle, S. (1997) *Anti-Cancer Drug Des. 12*, 433−442.
6. Maher, L. J. d., Wold, B., and Dervan, P. B. (1989) *Science 245*, 725−730.
7. Ussery, D. W., and Sinden, R. R. (1993) *Biochemistry 32*, 6206−6213.
8. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) *Annu. Rev. Biochem. 64*, 65−95.
9. Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987) *Nature 330*, 495−497.
10. Wang, E., and Feigon, J. (1998) *Structures of Nucleic Acid Triplexes*, Oxford University Press, Oxford (in press).
11. Radhakrishnan, I., and Patel, D. J. (1994) *Biochemistry 33*, 11405−11416.
12. Arnott, S., and Selsing, E. (1974) *J. Mol. Biol. 88*, 509−521.

13. Macaya, R. F., Schultze, P., and Feigon, J. (1992) *J. Am. Chem. Soc. 114*, 781−783.
14. Liquier, J., Coffinier, P., Firon, M., and Taillandier, E. (1991) *J. Biomol. Struct. Dyn. 9*, 437−445.
15. Sklenář, V., and Feigon, J. (1990) *Nature 345*, 836−838.
16. Macaya, R. F., Wang, E., Schultze, P., Sklenář, V., and Feigon, J. (1992) *J. Mol. Biol. 225*, 755−773.
17. Bornet, O., and Lancelot, G. (1995) *J. Biomol. Struct. Dyn. 12*, 803−814.
18. Radhakrishnan, I., and Patel, D. J. (1994) *J. Mol. Biol. 241*, 600−619.
19. Radhakrishnan, I., and Patel, D. J. (1994) *Structure 2*, 17−32.
20. Koshlap, K. M., Schultze, P., Brunar, H., Dervan, P. B., and Feigon, J. (1997) *Biochemistry 36*, 2659−2668.
21. Wang, E., Koshlap, K. M., Gillespie, P., Dervan, P. B., and Feigon, J. (1996) *J. Mol. Biol. 257*, 1052−1069.
22. Wang, S., Booher, M. A., and Kool, E. T. (1994) *Biochemistry 33*, 4639−4644.
23. Broitman, S. L. (1995) *Prog. Biophys. Mol. Biol. 63*, 119−129.
24. Dittrich, K., Gu, J., Tinder, R., Hogan, M., and Gao, X. (1994) *Biochemistry 33*, 4111−4120.
25. Phipps, A. K., Tarköy, M., Schultze, P., and Feigon, J. (1998) *Biochemistry 37*, 5820−5830.
26. Gotfredsen, C. H., Schultze, P., and Feigon, J. (1998) *J. Am. Chem. Soc.* (in press).
27. Beaucage, S. L., and Caruthers, M. H. (1981) *Tetrahedron Lett. 22*, 1859−1862.
28. Feigon, J., Sklenář, V., Wang, E., Gilbert, D. E., Macaya, R. F., and Schultze, P. (1992) in *Methods in Enzymology* (Lilley, D. M. J., and Dahlberg, J. E., Eds.) pp 235−253, Academic Press, San Diego, CA.
29. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
30. Macura, S., and Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
31. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
32. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
33. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820−2821.
34. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
35. Mueller, L. (1987) *J. Magn. Reson. 72*, 191−196.
36. Marion, D., and Bax, A. (1988) *J. Magn. Reson. 80*, 528−533.
37. Santoro, J., and King, G. (1992) *J. Magn. Reson. 97*, 202−207.
38. de Leeuw, F. A. A. M., and Altona, C. J. (1983) *J. Comput. Chem. 4*, 428−437.
39. van Wijk, J., Huckriede, B. D., Ippel, J. H., and Altona, C. J. (1992) *Methods Enzymol. 211*, 286−306.
40. Neidig, K. P., Geyer, M., Gorler, A., Antz, C., Saffrich, R., Beneicke, W., and Kalbitzer, H. R. (1995) *J. Biomol. NMR 6*, 255−270.
41. Brünger, A. T. (1992) *X-PLOR (Version 3.1) Manual*, Yale University Press, New Haven, CT.
42. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
43. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
44. Macaya, R. F., Gilbert, D. E., Malek, S., Sinsheimer, J., and Feigon, J. (1991) *Science 254*, 270−274.
45. Antao, V. P., and Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 819−824.
46. Rumney, S., and Kool, E. T. (1995) *J. Am. Chem. Soc. 117*, 5635−5646.
47. Lankhorst, P. P., Erkelens, C., Haasnoot, C. A., and Altona, C. (1983) *Nucleic Acids Res. 11*, 7215−7230.
48. Akhebat, A., Dagneaux, C., Liquier, J., and Taillandier, E. (1992) *J. Biomol. Struct. Dyn. 10*, 577−588.
49. Luyten, I., Thibaudeau, C., Sandstrom, A., and Chattopadhyaya, J. (1997) *Tetrahedron 53*, 6433−6464.
50. Thibaudeau, C., Plavec, J., and Chattopadhyaya, J. (1996) *J. Org. Chem. 61*, 266−286.
51. Plavec, J., Tong, W. M., and Chattopadhyaya, J. (1993) *J. Am. Chem. Soc. 115*, 9734−9746.
52. Bartley, J. P., Brown, T., and Lane, A. N. (1997) *Biochemistry 36*, 14502−14511.

BI9728102