Solution Structure of an Intramolecular Pyrimidine−Purine−Pyrimidine Triplex Containing an RNA Third Strand

Charlotte H. Gotfredsen,[†] Peter Schultze,[‡] and Juli Feigon*[,‡]

Contribution from the Department of Chemistry, Odense University, DK-5230 Odense M, Denmark, and Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, California 90095-1569

Received September 12, 1997

Abstract: We have used two-dimensional NMR spectroscopy to study an intramolecular pyrimidine−purine−pyrimidine triplex composed of a Watson−Crick paired DNA duplex and an RNA third strand connected by two hexakis(ethylene glycol) linkers. The solution structure of the triplex was calculated using restrained molecular dynamics and relaxation matrix refinement. The overall helical structure of the Watson−Crick duplex part of the triplex more closely resembles B-DNA than A-DNA. Thus, binding of a third-strand RNA in the major groove does not cause a structural transition toward A-DNA. Both the spectroscopic data and the calculated structures show a narrowing of the minor groove relative to an all-DNA triplex of the same sequence. The sugar puckers of the DNA residues are mostly C2′-endo while the RNA third strand residues show only partial C3′-endo character.

Introduction

Oligodeoxyribonucleotide, oligoribonucleotide, or hybrid oligonucleotide triplexes can be formed when a single oligonucleotide strand binds in the major groove of a homopurine−homopyrimidine duplex.[1−8] The third strand will bind either parallel or antiparallel to the purine strand in the Watson−Crick duplex, depending on the nucleotide sequence. In general, if the third strand is a homopurine strand or a combination of purines and thymidines, it will bind in the antiparallel, purine motif, via reverse Hoogsteen base pairing, G with G·C and A or T with A·T.[9−12] If the third strand is a homopyrimidine strand, it will bind in a parallel, pyrimidine motif, via normal Hoogsteen base pairing, T with A·T and C^+ with G·C.[3,13−15] Mismatched triplets can also be accommodated but cause a destabilization of the triplex that is dependent on the neighbor sequence.[16−26]

The possibility of using the high sequence selectivity of triplex formation for targeting specific genes makes them a potential therapeutic tool and an object of many interesting studies.[27−30] Formation of hybrid triplexes composed of both DNA and RNA strands have much interest due to their possible role in antisense technology, in which synthetic oligonucleotides are used to block translation of mRNA. The relative stability of hybrid DNA−RNA triplexes, as for pure DNA or RNA triplexes, depends strongly on the sequence, length, and solution conditions.[19,31−35]

[†] Odense University.
[‡] University of California, Los Angeles.
* Corresponding author. Telephone: 310-206-6922. Fax: 310-825-0982. E-mail: feigon@mbi.ucla.edu.

(1) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023−2024.
(2) Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlrab, F. *FASEB J.* **1988**, *2*, 2939−2949.
(3) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645−650.
(4) Sun, J.-S.; Hélène, C. *Curr. Opin. Struct. Biol.* **1993**, *3*, 345−356.
(5) Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666−690.
(6) Radhakrishnan, I.; Patel, D. J. *Biochemistry* **1994**, *33*, 11405−11416.
(7) Soyfer, V. N.; Potaman, V. N. *Triple-Helical Nucleic Acids*; Springer-Verlag: New York, 1996; pp 360.
(8) Wang, E.; Feigon, J. *Structures of Nucleic Acid Triplexes* in *Oxford Handbook of Nucleic Acid Structures*; Neidle, Ed.; 1998, in press.
(9) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360−1363.
(10) Broitman, S. L.; Im, D. D.; Fresco, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5120−5124.
(11) Radhakrishnan, I.; de los Santos, C.; Patel, D. J. *J Mol. Biol.* **1991**, *221*, 1403−1418.
(12) Radhakrishnan, I.; Patel, D. J. *Structure* **1993**, *1*, 135−152.
(13) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282−7289.
(14) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859−7870.
(15) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637−640.
(16) Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weis, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3840−3844.
(17) Belotserkovskii, B. P.; Veselkov, A. G.; Filippov, S. A.; Dobrynin, V. N.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1990**, *18*, 6621−6624.
(18) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967−971.
(19) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannangeli, C.; Garestier, T.; Hélène, C. *Biochemistry* **1992**, *31*, 9269−9278.
(20) Macaya, R. F.; Gilbert, D. E.; Malek, S.; Sinsheimer, J.; Feigon, J. *Science* **1991**, *254*, 270−274.
(21) Wang, E.; Malek, S.; Feigon, J. *Biochemistry* **1992**, *31*, 4838−4846.
(22) Ji, J.; Hogan, M. E.; Gao, X. *Structure* **1996**, *4*, 425−435.
(23) Dittrich, K.; Gu, J.; Tinder, R.; Hogan, M.; Gao, X. *Biochemistry* **1994**, *33*, 4111−20.
(24) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, *2*, 17−32.
(25) Radhakrishnan, I.; Patel, D. J. *J. Mol. Biol.* **1994**, *241*, 600−619.
(26) Roberts, R. W.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9397−9401.
(27) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1993**, *49*, 6123−6194.
(28) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456−459.
(29) Hélène, C.; Toulmé, J. J. *Biochim. Biophys. Acta* **1990**, *1049*, 99−125.
(30) Maher, L. J. *Bioessays* **1992**, *14*, 807−815.
(31) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463−1466.
(32) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3806−10.
(33) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995−11003.
(34) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957−6965.

S0002-7863(97)03221-6 CCC: $15.00 © 1998 American Chemical Society
Published on Web 04/23/1998

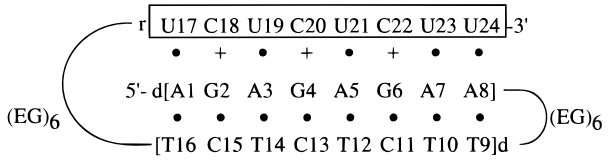

Figure 1. Folding scheme of the RDD-EG triplex with the residue numbering scheme used. The boxed third-strand bases are ribonucleotides. The Hoogsteen base pairs are indicated with a + for the protonated cytosines and filled circles for the others. The Watson−Crick base pairs are indicated by filled circles. $(EG)_6$ is a hexaethylene glycol loop.

Only six of the eight possible DNA−RNA hybrid pyrimidine−purine−pyrimidine triplexes form; the ones with a DNA third strand and an RNA purine strand are not stable.[31,32,35] Studies of stability and structure of chimera nucleic acid triplexes have been conducted mainly by use of UV absorption spectrophotometry,[31,35−37] vibrational spectroscopy,[36,38,39] gel electrophoresis,[31] and quantitative affinity cleavage titration.[32] These have given information on triplex formation, sugar conformations, thermodynamic stability, and their pH dependence. Although not totally in agreement on relative stabilities, all studies indicate that an RNA third strand bound to a DNA duplex is one of the most stable chimera triplexes.

Two NMR studies on hybrid triplexes have been reported[36,40] which show that RNA−DNA−DNA (RDD) triplexes, where the RNA is the third strand, form the expected hydrogen-bonded base triplets. Here we present the first high-resolution NMR structure of an RDD intramolecular triplex (RDD-EG). RDD-EG is composed of two octadeoxyribonucleotides and one octaribonucleotide linked by two hexakis(ethylene glycol) moieties, designed to form an intramolecular DNA duplex with an RNA third strand binding in its major groove. The sequence and numbering scheme of RDD-EG are shown in Figure 1. The ethylene glycol linkers were used to decrease the spectral complexity as well as to eliminate possible structural effects on the triplex caused by the presence of nucleotide loops. In addition, substitution of ethylene glycol loops renders resistance to nuclease degradation, a possible advantage for in vivo applications.[41] The refined three-dimensional structures show that binding of a third-strand RNA in the major groove has little effect on the structure of the DNA duplex part. The structure shows that the minor groove is narrow compared to a normal B-DNA helix and to an all-DNA triplex of the same sequence (DDD-EG).[42] The sugar puckers for the RNA third strand are not pure C3′-endo. The RDD-EG triplex structure is compared to the structure of DDD-EG.

Materials and Methods

Sample Preparation. dAGAGAGAA-$(EG)_6$-dTTCTCTCT-$(EG)_6$-rUCUCUCUU (RDD-EG) was chemically synthesized on an ABI 392 DNA/RNA synthesizer using commercially available deoxyribo- and ribonucleotide (ABI) and the hexakis(ethylene glycol) ("Spacer 18", Glen Research) phosphoramidites. Coupling times for the ethylene glycol linker and the RNA residues were increased from 25 s to 10 min. The oligonucleotide was deprotected in a 3:1 aqueous ammonia/ethanol solution at 55 °C for 12 h and subsequently evaporated to dryness. The 2′-TBDMS protecting group was removed with 1 M tetrabutylammonium fluoride/tetrahydrofuran (Lancaster) in excess for 27 h at room temperature, and the reaction was quenched with water. The crude product was concentrated and then desalted on a Sephadex G25-column eluted with water. The eluted products were subsequently precipitated with ethanol and purified by denaturing polyacrylamide gel electrophoresis. The nucleic acid was visualized by UV shadowing, and the desired bands were cut out, electroeluted, and ethanol precipitated. The precipitate was dissolved in H_2O and purified on a DEAE sephacel column, concentrated by ethanol precipitation, and desalted on a G15 column. The final conditions of the NMR samples were ∼2 mM RDD-EG and 100 mM NaCl (pH 5.7) in 450 μL of 99.999% D_2O or 90% H_2O/10% D_2O.

NMR Spectroscopy. The NMR experiments were performed on Bruker DRX500 or AMX500 spectrometers and processed with XWINNMR. One-dimensional spectra in H_2O were obtained for a range of temperatures from 1 to 45 °C using a $1\bar{1}$ spin−echo pulse sequence[43] to suppress the water resonance. Each spectrum was acquired with 4096 complex points, 128 scans, and a spectral width of 11 364 Hz and apodized with a 60° phase-shifted squared sine bell. Nuclear Overhauser effect (NOESY) spectra[44] were obtained in a States-TPPI mode.[45] For spectra in D_2O, the residual H_2O peak was suppressed by irradiation during the recycle delay and, in some cases, by additional use of gradients. Suppression of H_2O in water samples was achieved by using a $1\bar{1}$ spin−echo read pulse.[43] The carrier was centered on the water resonance, and the excitation maximum for the water suppression was set to 13 ppm. Further elimination of the water signal was achieved by use of gradients.

For assignment purposes, the following spectra were acquired: NOESY spectra in D_2O at temperatures of 10, 25, and 30 °C with mixing times from 100 to 400 ms, NOESY spectra in H_2O at temperatures of 1 and 10 °C with a mixing time of 150 ms, TOCSY[46] spectra at 30 °C and several mixing times ranging from 40 to 140 ms, homonuclear TOCSY−NOESY 3D spectra[47] at 25 °C with NOESY mixing time of 250 ms and TOCSY mixing times of 80−140 ms, and a heteronuclear $^1H-^{13}C$ natural abundance HSQC spectrum.[48] A $^1H-^{31}P$ correlation spectrum was also acquired.[49]

T_1 relaxation rates were obtained from inversion−recovery experiments[50] acquired at 30 °C. Forty variable delays (τ) were used between the inversion pulse and the observation pulse. The acquisition parameters used were a spectral width of 5000 Hz, 64 scans, and a recycle delay of 15 s. The 1D data series was transferred to give a 2D acquisition file, and T_1 values were fitted ($I[t] = I[0][1 - 2A \exp[-t/T_1]]$) using the T_1/T_2 relaxation package from BRUKER.

To obtain data on proton coupling constants, a DQF−COSY spectrum[51] was acquired at 25 °C. The acquisition parameters used were a spectral width of 5000 Hz in both t_2 and t_1, 2048 points in t_2, 524 points in t_1, 40 scans per t_1 block, and a recycle delay of 1.6 s. Apodization in t_1 and t_2 was a 60° phase-shifted squared sine bell. The spectrum was zero-filled to 4096 points in both dimensions, giving a digital resolution of 1.22 Hz/point for cross-peak pattern fitting.

(35) Wang, S.; Kool, E. T. *Nucleic Acids Res.* **1995**, *23*, 1157−1164.
(36) Liquier, J.; Taillandier, E.; Klinck, R.; Guittet, E.; Gouyette, C.; Huynh-Dinh, T. *Nucleic Acids Res.* **1995**, *23*, 1722−1728.
(37) Escude, C.; Sun, J. S.; Rougée, M.; Garestier, T.; Hélène, C. *C. R. Acad. Sci. Paris, Ser. III* **1992**, *315*, 521−525.
(38) Liquier, J.; Coffinier, P.; Firon, M.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1991**, *9*, 437−445.
(39) Dagneaux, C.; Liquier, J.; Taillandier, E. *Biochemistry* **1995**, *34*, 16618−16623.
(40) van Dongen, M. J.; Heus, H. A.; Wymenga, S. S.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *Biochemistry* **1996**, *35*, 1733−1739.
(41) Rumney, S.; Kool, E. T. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 1617−1619.
(42) Tarköy, M.; Phipps, A. K.; Schultze, P.; Feigon, J. *Biochemistry* **1998**, in press.
(43) Sklenář, V.; Bax, A. *J. Magn. Reson.* **1987**, *74*, 469−479.
(44) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6.
(45) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−399.
(46) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355−360.
(47) Oschkinat, H.; Cieslar, C.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1989**, *81*, 212−216.
(48) Santoro, J.; King, G. C. *J. Magn. Reson.* **1992**, *97*, 202−207.
(49) Sklenář, V.; Miyashiro, H.; Zon, G.; Miles, H. T.; Bax, A. *FEBS Lett.* **1986**, *208*, 94−98.
(50) Vold, R. L.; Waugh, J. S.; Klein, M. P.; Phelps, D. E. *J. Chem. Phys.* **1968**, *48*, 3831−3832.
(51) Piantini, U.; Sørensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1982**, *104*, 6800−6801.

Distance Restraints. Distance restraints for the nonexchangeable protons used for the distance geometry (DG) structure calculation were obtained from a NOESY experiment with $\tau_m = 250$ ms in D_2O at 30 °C. For the relaxation matrix structure calculation, the nonexchangeable NOE restraints were obtained from a set of NOESY spectra in D_2O acquired with mixing times of 40, 80, 150, and 250 ms at 30 °C. The acquisition parameters used were a spectral width of 5000 Hz in both t_2 and t_1, 2048 points in t_2, 300 points in t_1, 32 scans per t_1 block, and a recycle delay of 4.1 s. All spectra were apodized with a 75° phase-shifted squared sine bell. The NOE restraints involving the exchangeable resonances were obtained from H_2O $1\bar{1}$ echo NOESY spectra acquired at 1 and 10 °C with a mixing time of 150 ms. The acquisition parameters used were a spectral width of 11 364 Hz in both t_2 and t_1, 2048 points in t_2, 290 points in t_1, 64−96 scans per t_1 block, and a recycle delay of 1.9 s.

Torsion Angle Restraints. To determine the coupling constants in the deoxyriboses, the H1′−H2′ and H1′−H2″ cross-peaks in a DQF-COSY spectrum were fitted using the FORTRAN program CHEOPS (Schultze and Feigon, unpublished program). Acceptable correlation coefficients ranging from 91 to 97% were obtained for 12 of the 16 DNA residues. The dihedral angles ν_1 and ν_2 were then determined from these coupling constants by use of the program PSEUROT,[52] which fits them to a two-state model for the sugar pucker. The pseudorotation amplitudes were left constant at 38° for the fit, so that two pucker values and their relative population were obtained. The values of ν_1 and ν_2 associated with the majority conformation were used for restraints in the X-PLOR calculations. The RNA sugar dihedral angles were left unrestrained because of the lack of data for this type of analysis, due to the absence of H2″ and unassignable H3′ resonances.

Structure Calculations. Distance restraints between nonexchangeable protons used for the initial structure calculation were obtained from the 250-ms mixing time NOESY experiment. Cross-peaks were picked and integrated using the AURELIA software package.[53] The obtained NOESY peak lists were used in a semiautomatic assignment procedure. An expanded peak list was generated containing all possible assignments closer than 0.01 ppm in each dimension from the chemical shift list. As an additional decision criterion, the expanded list also contained the corresponding H−H distances from a preliminary model for each possible cross-peak assignment. The expanded list was inspected, corrected by hand, and reimported into AURELIA for obtaining distances and peak integrals in an X-PLOR readable format. For the D_2O spectra, the distances were calibrated using the strongest resolved H5−H6 cross-peak set to 2.4 Å. All lower limits were set to 2.0 Å, and the upper bounds for most of the peaks were set to the calibrated distance plus a margin of 0.5 Å, except for the methyl groups where 1.5 Å was added. In the case of overlapping cross-peaks, an upper bound of 5 Å was used. Peaks that consistently gave rise to violations in initial rounds of calculation were evaluated for incorrect assignment or overlap. In the case of inaccurate peak integrals caused by noise or overlap, the upper bound was increased. For the exchangeable protons, a uniform upper bound of 5 Å was 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 a typical calculation, 20 coordinate sets were obtained by metric matrix distance geometry embedding of all atoms in the initial step of X-PLOR (Version 3.1)[54] calculations. In the next step, the initial structures were subjected to the distance geometry-simulated annealing protocol (DGSA),[55] consisting of template fitting to improve local geometry followed by simulated annealing. Dihedral angle restraints involving the four ligands of each chiral center were introduced to enforce correct configurations. To increase the number of converging structures, the weight and the bond angle potential was increased 8-fold over the standard X-PLOR force field. As described previously,[56] the ceiling of the energy term for the distance restraints was set to 100, i.e., 1/10 of the value in the published[54] protocol. 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. The geometry of the base triplets was restrained using standard hydrogen bond distances. 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 20 initial structures were carried through to the third refinement step. At this point the coordinate sets were sorted according to overall energy terms. The first 10 of these coordinate sets were subjected to relaxation matrix refinement consisting of a short simulated annealing protocol with inclusion of the NOE potential. The nonexchangeable restraints for resolved cross-peaks were converted to cross-peak volumes for each of the four spectra, while the hydrogen bond restraints and the restraints obtained from the exchangeable resonance and overlapping nonexchangeable resonance cross-peaks were left as distances.

Several test calculations were also done to assess the effect of the planarity and hydrogen bond restraints on the final structures. First, a family of structures was generated using only the hydrogen bond and planarity constraints as input for the structure calculations. This gave models in which the three strands are properly base paired, but the triplex is almost completely unwound and has a greatly increased rise giving an elongated double-ladder-like appearance. Second, test calculations were run without the planarity restraints. This results in somewhat more poorly defined structures. However, these structures also agree slightly less well with the experimentally observed peak integrals as measured by their average $R^{1/6}$ factors (0.1176 without planarity vs 0.1158 with planarity for each ensemble of 10 structures). We note that the planarity restraints do not restrict the base triplets to be perfectly planar and that the triplets in the refined structures calculated with planarity restraints have buckle ranges from −11° to 12° and propeller twist ranges from −28° to 17°. These ranges correspond very closely to values observed in oligonucleotide crystal structures. Finally, a test calculation was done without using hydrogen bond restraints for the base pairs. This results in very poorly defined structures where not all base pairs are formed. However, this clearly is not in agreement with the additional non-NOE experimental data on the line widths and chemical shifts of the observed imino and amino proton signals (with the exception of the 5′-terminal triplet).

The helical parameters were calculated from the duplex portion of the triplex structures using the program CURVES.[57]

UV Melting Spectroscopy. UV melting studies were carried out on a Varian Cary1E spectrophotometer with a temperature probe. The heating rate was 0.5 °C/min. For determination of the T_m, the first derivative was plotted and the melting temperature read at the maximum. The samples were 1 mL with 0.35−0.5 A_{260} of DNA/RNA, and the buffer conditions were 50 mM NaOAc and 0.1 M NaCl with or without 5 mM $MgCl_2$ at pH 5.2.

Results

Thermal Stability of RDD-EG. The thermal stability of RDD-EG as measured by the change in UV absorbance at 260 nm as a function of the temperature was compared with that of an all-DNA triplex of the same sequence (DDD-EG), with dT and dC in the third strand in place of rU and rC[42] (Figure 2). Under the conditions used, 100 mM NaCl and 5 mM $MgCl_2$ (pH 5.2), both triplexes melt in a single cooperative transition. The RNA third strand increases the melting temperature by only 2.5 °C over the all-DNA triplex, from 68.0 to 70.5 °C. In the absence of Mg^{2+}, the melting temperature of RDD-EG is only slightly lower than with Mg^{2+} (69.0 °C). Since Mg^{2+} is known

(52) de Leeuw, F. A. A. M.; Altona, C. J. *J. Comput. Chem.* **1983**, *4*, 428−437.
(53) Neidig, K. P.; Geyer, M.; Gorler, A.; Antz, C.; Saffrich, R.; Beneicke, W.; Kalbitzer, H. R. *J. Biomol. NMR* **1995**, *6*, 255−270.
(54) Brünger, A. T. *X-PLOR (Version 3.1) Manual*; Yale University Press: New Haven, CT, and London, 1992; pp 291.
(55) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317−324.
(56) Koshlap, K. M.; Schultze, P.; Brunar, H.; Dervan, P. B.; Feigon, J. *Biochemistry* **1997**, *36*, 2659−2668.
(57) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.

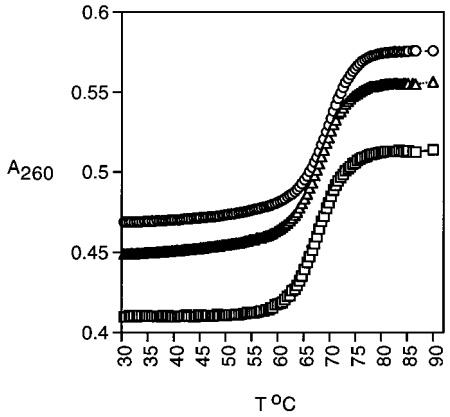

Figure 2. UV melting study of RDD-EG without (triangles) and with (circles) 5 mM $MgCl_2$ and DDD-EG (squares) with 5 mM $MgCl_2$ from 30 to 90 °C.

to catalyze RNA strand cleavage, the NMR sample was made up without magnesium to avoid sample degradation.

NOESY Spectrum of Exchangeable Resonances of RDD-EG. Analysis of the H_2O NOESY spectra confirms that RDD-EG forms an intramolecular triplex. The imino region of a NOESY spectrum of RDD-EG is shown in Figure 3a. Sequential imino−imino connectivities for both the Watson−Crick and Hoogsteen hydrogen-bonded base pairs, characteristic of a triplex structure, are seen for all but the terminal base triplet next to the Hoogsteen loop, A1, T16, and U17. The imino to amino/aromatic region from the same spectrum is shown in Figure 3b. The downfield shifted rC^+ amino protons show the same characteristic pattern seen for DNA triplexes, where each pair of rC^+ amino protons has the possibility of showing the following connectivities to five imino resonances: (i) its own imino proton, (ii) the imino proton in the 5′-neighbor Hoogsteen base pair, (iii) the imino proton in the 3′-neighbor Hoogsteen base pair, (iv) the imino proton on G in its own triplet (spin diffusion), and (v) the imino proton of the Watson−Crick base pair in the 3′ direction (with respect to the purine strand).[42,58] For $rC18^+$, only weak signals between its own amino and imino protons are seen. Each Watson−Crick A·T imino proton resonance shows a strong NOE to its own triplet AH2, while each Hoogsteen A·U and G·C^+ imino shows a strong NOE to its own triplet AH8 or GH8, as reported earlier for intramolecular DNA triplexes.[59] The imino, amino, H2, and some H6/H8 resonances were assigned from these spectra following previously described protocols for other intramolecular DNA triplexes[14,20,59,60] and are given in Table 1.

NOEs were observed between Hoogsteen base-paired imino resonances and H2′/H2″ resonances from the 5′-neighboring purine on the homopurine strand and between rC^+ amino and A amino protons on both the 5′- and 3′-neighboring residue (data not shown). These unusual NOE cross-peak patterns have been observed in DNA triplexes[25,59] and were also reported for an RNA−DNA triplex formed by a hairpin DNA duplex and a single-strand RNA.[40]

Assignments of the Nonexchangeable Resonances. Figure 4 shows a natural abundance $^1H-^{13}C$ HSQC spectrum of RDD-EG. Most of the RNA and DNA base and sugar and the ethylene glycol resonances are separated into well-resolved

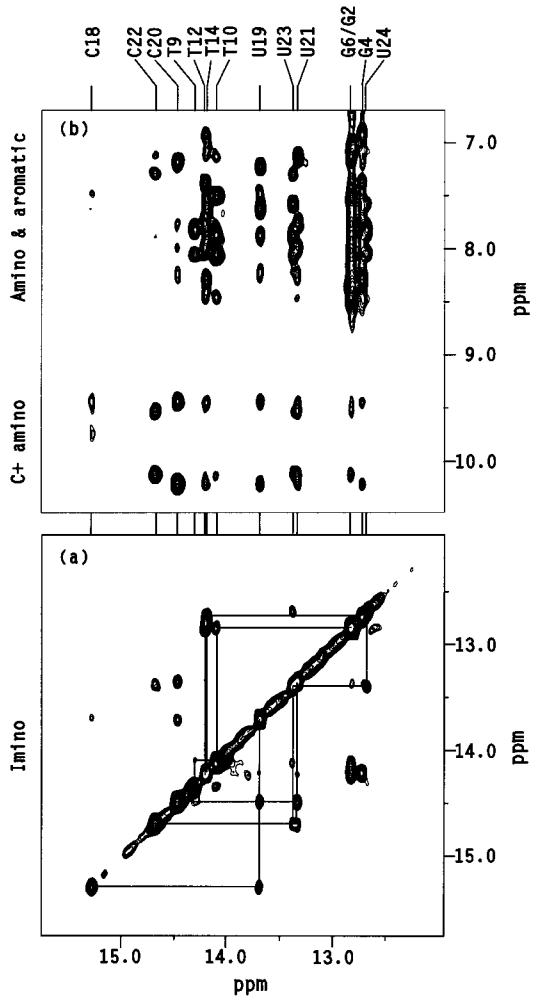

Figure 3. Portions of a NOESY spectrum of RDD-EG in 90% H_2O/10% D_2O at 10 °C with a mixing time of 150 ms. (a) Imino−imino region. The Watson−Crick imino sequential connectivities are indicated above the diagonal, and the Hoogsteen imino sequential connectivities are indicated below the diagonal. (b) The imino−amino, aromatic region. The NOESY spectrum was acquired with a spectral width of 11 364 Hz in both dimensions, 2048 points in t_2, 290 points in t_1, 64 scans per t_1 block, and a recycle delay of 2.1 s. The spectrum was apodized with Gaussian multiplication (LB = −18, GB = 0.1, and LB = −25, GB = 0.15 in t_2 and t_1, respectively). The final data matrix was 2K × 1K.

regions in the spectrum. This allows clear differentiation between DNA and RNA resonances as well as between CH5 and UH5 resonances on the basis of their different ^{13}C chemical shifts.[61] Thus, this spectrum provided a useful starting point for the assignment process. The sequential assignments of the nonexchangeable resonances for the RDD-EG triplex were for the most part done as previously described for intramolecular DNA triplexes using NOESY, DQF-COSY, and TOCSY spectra.[59] The H1′, H5 to aromatic region of a 250-ms NOESY spectrum of the RDD-EG triplex in D_2O at 30 °C is shown in Figure 5. The H1′ aromatic sequential assignment pathway can be followed along each strand, from dG2−dA8, dT9−dT16, and rU17−rU24. The assignment of A1 is not unambiguous here but can be identified elsewhere in the spectra. Sequential connectivities can also be traced for the DNA strands in the aromatic−H2′,2″ and aromatic−H3′ regions (data not shown). We note that the aromatic H6 resonances of the third strand

(58) Feigon, J.; Koshlap, K. M.; Smith, F. W. ^1H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods in Enzymology*; Academic Press: San Diego, 1995; Vol. 261, pp 225−255.

(59) Macaya, R.; Wang, E.; Schultze, P.; Sklenář, V.; Feigon, J. *J. Mol. Biol.* **1992**, *225*, 755−773.

(60) Sklenář, V.; Feigon, J. *Nature* **1990**, *345*, 836−838.

(61) Varani, G.; Tinoco, I. *Q. Rev. Biophys.* **1991**, *24*, 479−532.

Table 1. Chemical Shifts for the Exchangeable and Nonexchangeable Resonances at 10 and 30 °C, Respectively

base	H6, H8	H2, H5, Me	H1′	H2′	H2″	H3′	H4′	imino	amino
A1	7.74	7.74	6.01	2.71	2.71	5.03	4.23		
G2	7.51		6.02	2.58	2.97	5.10	4.37	12.81	
A3	7.28	7.57	6.04	2.35	2.84	4.81	4.56		7.64, 7.89
G4	7.17		5.93	2.48	2.86	4.99	4.41	12.71	
A5	7.18	7.42	5.99	2.37	2.78	4.88	4.41		8.00, 7.77
G6	7.28		5.82	2.44	2.79	4.95	4.49	12.81	
A7	7.61	7.55	5.99	2.41	2.8	5.00	4.34		7.89, 8.07
A8	8.04	8.08	6.25	2.48	2.62	4.92	4.37		7.83
T9	7.75	1.91	6.31	2.46	2.76	4.94		14.42	
T10	7.61	1.80	6.19	2.41	2.74	4.95	4.30	14.10	
C11	7.65	5.69	6.01	2.28	2.66	4.73	4.31		7.12, 8.46
T12	7.65	1.74	6.06	2.43	2.68	4.92	4.27	14.22	
C13	7.59	5.61	5.94	2.21	2.59	4.72	4.21		6.93, 8.29
T14	7.66	1.74	6.00	2.28	2.59	4.92	4.20	14.22	
C15	7.68	5.69	6.12	2.19	2.56	4.76	4.22		
T16	7.56	1.68	6.22	2.25	2.54	4.87	4.16		
T17	7.96	5.87	6.06	4.58		4.84			
C18	8.10	6.19	5.69	4.45		4.66		15.29	9.48, 9.99
T19	8.26	5.67	5.95	4.56				13.72	
C20	8.24	5.97	5.68	4.46				14.43	9.44, 10.19
T21	8.35	5.68	5.81	4.49				13.36	
C22	8.25	6.02	5.72	4.31				14.62	9.56, 10.11
T23	8.23	5.64	5.86	4.24				13.38	
T24	7.93	5.63	5.98	4.13		4.26		12.73	

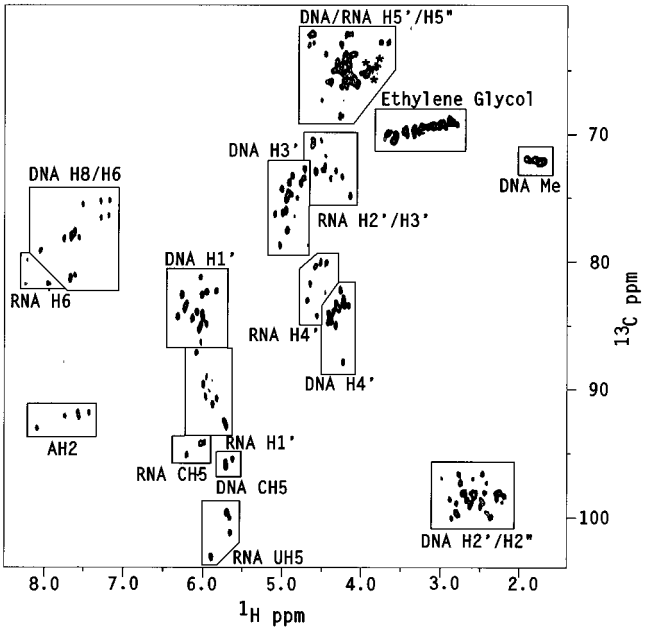

Figure 4. $^{1}H-^{13}C$ HSQC spectrum at 25 °C of RDD-EG. Ethylene glycol residues that are shifted into the H5′/H5″ area are indicated with (*). The HSQC spectrum was acquired with a spectral width of 5000 Hz in t_2 and 7546 Hz in t_1, 2048 points in t_2, 297 points in t_1, 160 scans per t_1 block, and a recycle delay of 1.7 s. The spectrum was apodized with a Gaussian multiplication (LB = −15, GB = 0.1) in t_2 and a 90° phase-shifted square sine bell in t_1.

are all downfield-shifted relative to the third-strand aromatic residues for the DNA analogue.[42]

A further indication of triplex formation is the presence of interstrand cross-peaks between the purine H8 and the H1′ resonances on the 3′-neighboring Hoogsteen base-paired pyrimidine. These are all seen except for A7H8−U24H1′ (Figure 5). Interstrand connectivities from AH2 to H1′ resonances are also present; for example, A5H2 connects to A5H1′, A6H1′, T12H1′, and C13H1′ (Figure 5). The intensities of the H1′ aromatic intraresidue cross-peaks indicate that all residues are in the anti conformation.

We were unable to obtain specific assignments for the ethylene glycol linker protons. However, in both the D_2O and H_2O NOESY spectra, a large number of NOEs are observed between the last base triplet before the Hoogsteen loop (U17•-A1•T16) and the ethylene glycol linker. Some NOEs are also observed between the ethylene glycol linker forming the Watson−Crick loop and the last base pair (A8•T9) in the duplex part. This differs from the results reported on an intramolecular DNA triplex with nine base pairs and seven triplets with the same linkers used in this study, where no NOE cross-peaks were observed to the linker.[62]

A $^{31}P-^{1}H$ heteroCOSY spectrum was also obtained on RDD-EG (not shown). All of the phosphorus resonances appear in a narrow range of about 2 ppm. We were unable to resolve sequential assignments using $^{31}P-^{1}H$ correlation spectra. The fact that none of the phosphates is shifted significantly out of the normal chemical shift range does not give any evidence that hydrogen bonds are formed between the third-strand 2′-OH and the P−O on the purine strand, as proposed for an all-RNA A-form triple helix.[63]

Sugar Conformation. Sugar conformations were determined in part from the cross-peaks observed in a DQF-COSY spectrum of RDD-EG. For deoxyribose, H1′−H2′ and H1′−H2" coupling constants were determined as described in the Materials and Methods. This analysis showed that all of the deoxyriboses are predominantly in the S-type sugar conformation, in the range normally observed for B-DNA and also found for DNA triplexes.[64] The conformation of the riboses was qualitatively determined from the H1′−H2′ cross-peak intensities. For N-type ribose puckers, the H1′−H2′ γ-coupling is <2 Hz and no cross-peak is expected to be observed. For the RNA strand of RDD-EG, all of the rU sugars had observable H1′−H2′ cross-peaks in the DQF−COSY spectrum. This indicates that $J_{H1',H2'}$ > 2 Hz and that they are therefore not pure N-type conformation. The protonated rC's in the third strand had smaller

(62) Bornet, O.; Lancelot, G. *J. Biomol. Struct. Dyn.* **1995**, *12*, 803−814.

(63) Holland, J. A.; Hoffman, D. W. *Nucleic Acids Res.* **1996**, *24*, 2841−2848.

(64) Macaya, R. F.; Schultze, P.; Feigon, J. *J. Am. Chem. Soc.* **1992**, *114*, 781−783.

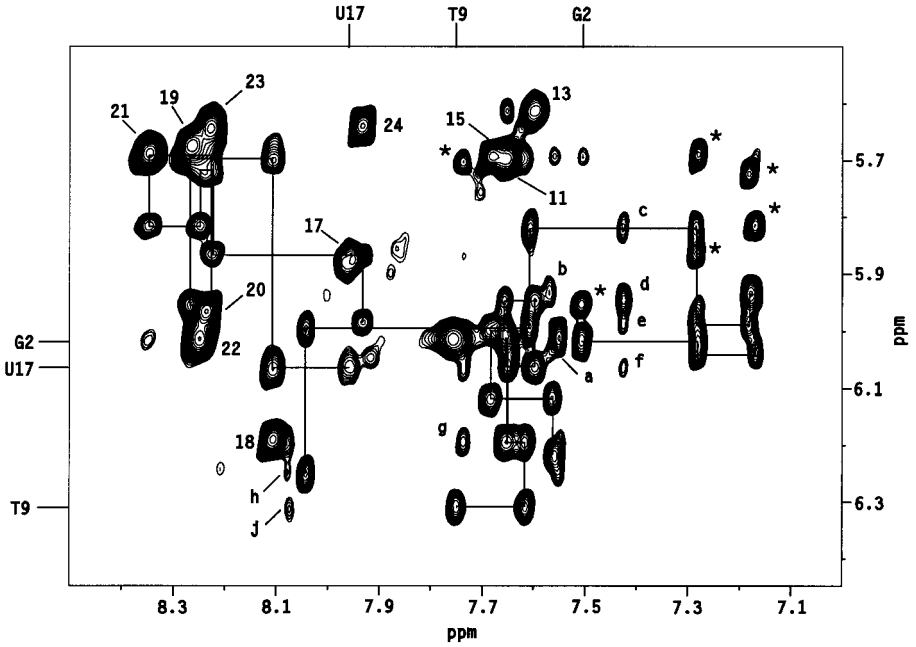

Figure 5. H1′,H5 aromatic region of the 2D NOESY spectrum of RDD-EG in D_2O at 30 °C with mixing time of 250 ms. The sequential assignments along the three strands are indicated with lines. CH5−CH6 and UH5−UH6 cross-peaks are labeled by residue number. Interstrand cross-peaks between the purine H8 and the H1′ on the 5′-neighboring Hoogsteen base-paired pyrimidine are indicated with (*). Cross-peaks between AH2 and H1′ protons are labeled (a) A1H2-G2H1′/A1H2-A1H1′, (b) A3H2-G4H1′, (c) A5H2-G6H1′, (d) A5H2−C13H1′, (e) A5H2-A5H1′, (f) A5H2-T12H1′, (g) A7H2-T10H1′, (h) A8H2-A8H1′, and (j) A8H2-T9H1′. The NOESY spectrum was acquired with a spectral width of 5000 Hz in both dimensions, 2048 points in t_2, 300 points in t_1, 48 scans per t_1 block, and a recycle delay of 4.1 s. The spectrum was apodized with a 75° phase-shifted squared sine bell in both t_1 and t_2. The t_1 dimension was zero-filled to 1K points.

Table 2. Relaxation Times of Some Resolved Protons

assignment		T_1 (s)
A5	H2	3.90
A8	H2	3.13
U21	H6	2.14 (RNA)
T10	H6	1.63
G2	H8	2.04
A5	H8	2.22
G2	H3′	2.37
G6	H1′	2.12
T10	Me	1.06

H1′−H2′ coupling constants indicative of mostly N-type conformation. This is also observed in general for third-strand protonated rC's in DNA triplexes.[59,64] Overall, RDD-EG does not appear to have sugar conformations significantly different from those of DDD-EG and other DNA triplexes.

***T_1* Analysis.** In studies on duplex RNA−DNA chimeras, it has been found by Reid and co-workers[65,66] that there is a significant difference in the relaxation rates of DNA and RNA proton resonances. This can significantly affect observed NOE cross-peak intensities if the relaxation delay used in the NOESY experiment is too short. We therefore obtained an inversion−recovery experiment on RDD-EG. From a qualitative examination of the spectra, it can be concluded that the RNA protons do not have a significantly longer T_1 than DNA protons of the same kind. Calculated T_1's for the few resolved proton resonances are given in Table 2. As expected, the AH2 have the longest T_1 relaxation times, e.g. A5H2 has a T_1 of 3.9 s. There is no significant difference between the relaxation time of U21H6 and the DNA base H6 and H8 resonances. On the basis of these results, we chose a 4.1-s relaxation delay in the mixing time series NOESY experiments.

Table 3. Input and Results of Structure Refinement

no. of NOE-derived distance restraints	
intranucleotide ($\Delta n = 0$)	
DNA (res 1−8, 9−16)	204
RNA (res 17−24)	38
sequential ($\Delta n = 1$)	
DNA−DNA	115
RNA−RNA	46
internucleotide ($\Delta n > 1$)	
DNA−DNA	29
DNA−RNA	79
RNA−RNA	0
total	511
no. of hydrogen bond distance restraints	70
no. of torsion angle restraints	24
no. of integrated peak intensities for relaxation matrix refinement	4 × 174
refinement results for 10 lowest energy structures	
total no. of NOE violations > 0.5 Å	4
total no. of dihedral angle violations > 5°	0
relaxation matrix refinement	
average *r* factor before	0.1158 ± 0.0014
average *r* factor after	0.0540 ± 0.0031
average pairwise RMSD (all residues) (Å)	1.12 ± 0.21
average RMSD from ideal covalent geometry	
bond length (Å)	0.017
bond angle (deg)	5.2

Relaxation Matrix Refinement of RDD-EG. A total of 581 distance restraints were used in the DG structure calculation prior to the relaxation matrix refinement. Of these, 70 are hydrogen bond restraints, 407 are restraints between nonexchangeable protons, and 104 involve exchangeable protons (Table 3). For the relaxation matrix refinement, peak integrals in the four NOESY spectra at different mixing times were determined for all 407 NOE cross-peaks involving only nonexchangeable protons. Analysis of these data showed that

(65) Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R. *Biochemistry* **1993**, *32*, 4207−4215.
(66) Wang, A. C.; Kim, S. G.; Flynn, P. F.; Chou, S.-H.; Orban, J.; Reid, B. R. *Biochemistry* **1992**, *31*, 3940−3946.

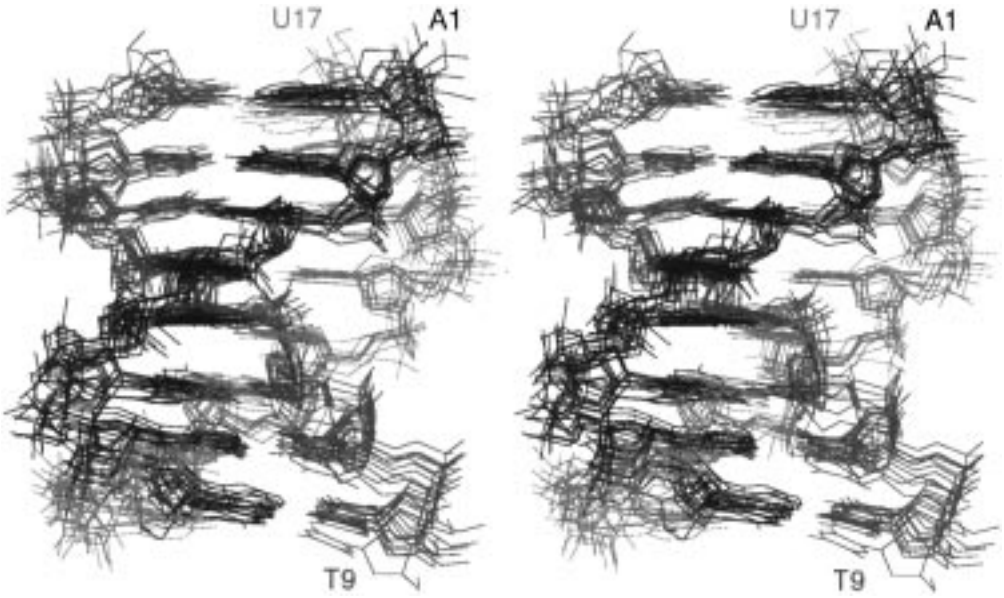

Figure 6. Stereoview of a superposition of the 10 lowest energy structures of RDD-EG. View is into the major groove. The Watson−Crick purine strand is blue, the Watson−Crick pyrimidine strand is green, and the Hoogsteen pyrimidine strand is magenta. The 5′ nucleotides for each "strand" are labeled.

233 of these peak integrals for the series had to be discarded. In those cases, the integration algorithm had failed at one or more mixing times because of peak overlap or spectral artifacts. To assess the quality of relaxation matrix refinement with significantly fewer peak integrals than distance restraints, two different methods of refinement were compared. In method A, only the 174 remaining volume restraints were used. In method B, the 233 distance restraints from the previous refinement step, corresponding to the deleted volume data, were used in conjunction with the 174 volume restraints. Both methods reached almost exactly the same minimum relaxation matrix term, as shown by the average *R* factors of 0.0527 (A) and 0.0540 (B). The changes which occur during relaxation matrix refinement are minor. A direct comparison of each of the 10 structures before and after relaxation matrix refinement gives an average rmsd of 1.19 ± 0.05 Å for method A and 1.06 ± 0.06 Å for method B of the two sets of 10 pairs. The additional distances lead to greater precision of the ensemble of structures obtained from method B, with an average pairwise rmsd of 1.12 Å, compared to 1.25 Å with method A. For this reason, the set of structures resulting from method B was chosen for final analysis.

A stereoview of the 10 best superimposed structures is shown in Figure 6. Refinement statistics are given in Table 3. The average pairwise rmsd's were calculated for the structure without the loops, which were not included in the structure calculation. The refined structure shows a well-defined triple helix for the central six triplets, with the conformation of the end triplets being somewhat less precisely defined.

Discussion

Triplex Formation and Stability. Here we have studied the structural effects of introducing a third-strand RNA in an intramolecular triplex. The NMR spectra show that the oligonucleotide folds into the designed intramolecular triplex with the RNA strand located in the major groove. From the optical melting studies, the intramolecular chimera triplex is found to be slightly more stable than its DNA analogue, DDD-EG. The fact that the increase in T_m is only 2.5 °C for the RDD chimera compared to the DDD triplex was surprising in view of other studies in which it was found that RDD triplexes are significantly more stable than DDD triplexes.[31,32] This relatively small increase in stability may be due to the short length of the triplex and/or the possibility that the Hoogsteen loop of six ethylene glycol units [$(EG)_6$] might be too short, effectively destabilizing the terminal triplet. No imino−imino cross-peaks were observed for this U17·A1·T16 triplet. That $(EG)_6$ may not be the optimal linker in either the Watson−Crick loop or the Hoogsteen loop was reported by Rumney and Kool.[67] They find that, although the $(EG)_6$ linker is the one most commonly used (and commercially available), $(EG)_7$ for the Watson−Crick loop and $(EG)_8$ or $(EG)_9$ for the Hoogsteen loop seem to be more favorable.

The effect of the divalent ion Mg^{2+} on the RDD-EG triplex stability is also smaller than expected ($\Delta T_m = 1.5$ °C) (Figure 2), since divalent ions are known to stabilize triplex formation.[19,58] However, there are no detailed studies of the effect of Mg^{2+} on the stability of triplexes with RNA third strands.

RDD Triplex Structure Is Similar to That of Other DDD Triplexes. The refined family of structures in Figure 6 show a well-defined triplex with the RNA strand Hoogsteen paired to the purine strand in the major groove, as expected. As discussed below, the helical parameters of the RDD triplex are very similar to those of previously determined DDD triplexes. However, one interesting feature is the narrow minor groove. Figure 7 shows a comparison of the structures of RDD-EG and DDD-EG, to illustrate the difference in the width of the minor groove. The minor groove of RDD-EG is also narrower than that of a standard B-DNA duplex.

To help analyze the structure, helical parameters for the Watson−Crick duplex part of the triplex were calculated (Figure 8). The *X* displacement of the Watson−Crick base pairs varies from −2.3 to −2.9 Å with an average value of −2.5 Å. These values are almost identical to those found for DDD-EG.[42] The *X* displacement is closer to that of standard B-DNA (−0.7 Å) than A-DNA (−5.4 Å). The average value for the rise is 3.2 ± 0.5 Å. It is not apparent why there is an especially large rise between nucleotides 5 and 6. It is possible that this may

(67) Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 5635−5646.

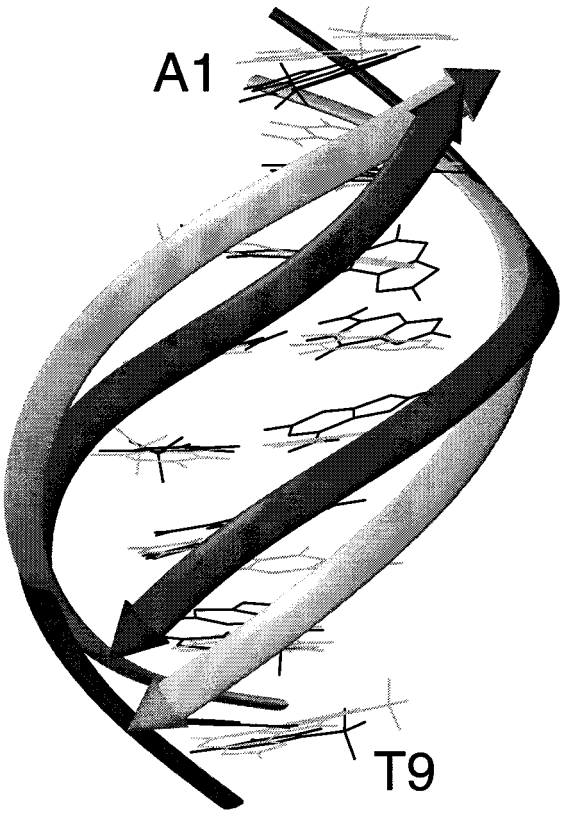

Figure 7. View into the minor groove of the lowest energy structures of RDD-EG and DDD-EG.[42] The backbone ribbons of RDD-EG (dark gray) and DDD-EG (light gray) are shown to highlight the difference in minor groove width.

be due to insufficient restraints due to overlap for these residues. The helical twist has an average value of 31 ± 7°.

Estimates for the longitudinal T_1 relaxation times were obtained prior to the build-up NOESY experiment to see if the T_1 relaxation time was significantly different for the RNA protons compared to the DNA protons. We find that not to be the case for the RDD-EG triplex (Table 2). This is in contrast to the results obtained by Reid and co-workers,[66] who found that in DNA−RNA hybrid and chimera duplexes the RNA protons have significantly longer T_1 relaxation times than the DNA protons. They suggest that this is not due to internal motion differences between DNA and RNA but probably correlates with the type of helix.[65,66] This would be consistent with our finding no significant difference in T_1, since no pronounced effect on the structure upon binding of the RNA Hoogsteen strand was found, i.e. the RNA strand does not induce a structural transition from a B-DNA-type helix toward an A-DNA-type helix.

Sugar Conformations Are Consistent with a B-Form Helix. One unexpected feature of the RDD-EG triplex is that all of the DNA sugars are in the C2′-endo range and the RNA sugars are not pure C3′-endo. An exact number for the mole fraction of C2′-endo vs C3′-endo was not obtainable for the RNA residues, but the fact that we know that we do have partly C2′-endo sugars for these is a significant result. Other studies of different chimera and hybrid sequences have found the opposite result. Liquier et al.[36] and Dagneaux et al.[39] find, by use of FTIR for evaluation of the sugar puckers in an intramolecular chimera triplex, that upon lowering of the pH the deoxyribose sugars repucker to be mainly N-type sugars. It is possible that the different result obtained here may be due to the difference in pH (pH = 4.8 vs 5.7 here) and ionic strength

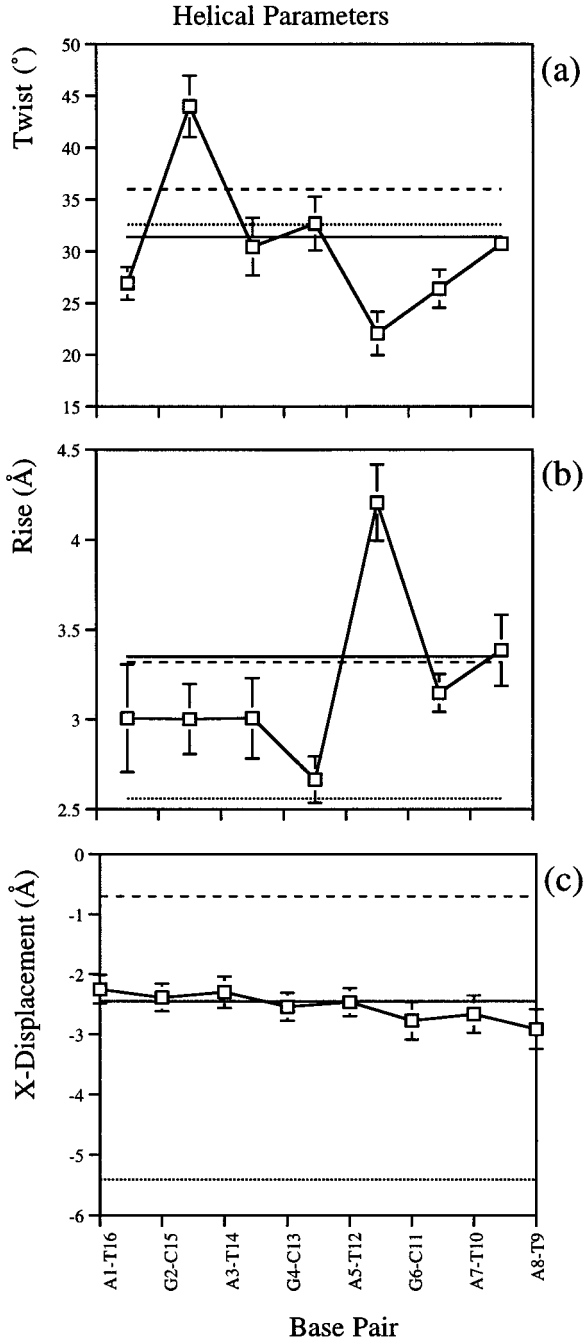

Figure 8. Helical parameters for the RDD-EG: (a) twist, (b) rise, and (c) X-displacement. Error bars indicate the standard deviation of the measurement from the 10 lowest energy structures. The values from fiber diffraction for standard A-DNA and B-DNA[72,73] and the average value for DDD-EG are indicated by the dotted lines, dashed lines, and solid lines, respectively.

(20 mM NaCl[36] vs 100 mM NaCl in our study). For studies of all-RNA (RRR) triplexes only C3′-endo sugar puckers have been reported.[36,63,68,69]

Several NMR studies of intramolecular pyrimidine−purine−pyrimidine triplexes have shown that they form B-type helices with mostly C2′-endo sugar puckers.[6,24,25,42,56,59,62,70,71] The majority of these structures are DNA triplexes containing either

(68) Klinck, R.; Liquier, J.; Taillandier, E.; Gouyette, C.; Huynhdinh, T.; Guittet, E. *Eur. J. Biochem.* **1995**, *233*, 544−553.
(69) Klinck, R.; Guittet, E.; Liquier, J.; Taillandier, E.; Gouyette, C.; Huynh-Dinh, T. *FEBS Lett.* **1994**, *355*, 297−300.
(70) Wang, E.; Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Mol. Biol.* **1996**, *257*, 1052−69.

a mismatch or nonnatural base in the third strand. We have found that both spectral and structural data indicate that, at pH 5.7 and 100 mM NaCl, the RDD triplex studied here forms a B-DNA-type helix similar to the all-DNA triplexes, with all the DNA residues close to the C2′-endo sugar pucker conformation and the RNA residues sugar puckers being between C2′-endo and C3′-endo. Thus, the RNA third strand does not induce a structural change to an A-form helix under the experimental conditions used in this study.

The coordinates for the RDD-EG structures have been deposited in the Brookhaven Protein Data Base (entry 1R3X).

Acknowledgment. This work was supported by NIH Grant GM37254-09 to J.F. and a stipend from The Danish Research Academy to C.H.G. The authors thank Ms. A. K. Phipps for assistance in the relaxation refinement protocol and the design of Figure 7.

JA973221M

(71) Phipps, A. K.; Tarköy, M.; Schultze, P.; Feigon, J. *Biochemistry* **1998**, in press.

(72) Arnott, S.; Hukins, D. W. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504−1509.

(73) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.