Structure of a Triple Helical DNA with a Triplex−Duplex Junction‡

Sangkee Rhee, Zong-jin Han, Keliang Liu, H. Todd Miles, and David R. Davies*

Laboratory of Molecular Biology, NIDDK, Building 5, Room 338, National Institutes of Health, Bethesda, Maryland 20892

Received August 4, 1999; Revised Manuscript Received October 4, 1999

ABSTRACT: Extended purine sequences on a DNA strand can lead to the formation of triplex DNA in which the third strand runs parallel to the purine strand. Triplex DNA structures have been proposed to play a role in gene expression and recombination and also have potential application as antisense inhibitors of gene expression. Triplex structures have been studied in solution by NMR, but have hitherto resisted attempts at crystallization. Here, we report a novel design of DNA sequences, which allows the first crystallographic study of DNA segment containing triplexes and its junction with a duplex. In the 1.8 Å resolution structure, the sugar−phosphate backbone of the third strand is parallel to the purine-rich strand. The bases of the third strand associate with the Watson and Crick duplex via Hoogsteen-type interactions, resulting in three consecutive C^+·GC, BU·ABU (BU = 5-bromouracil), and C^+·GC triplets. The overall conformation of the DNA triplex has some similarity to the B-form, but is distinct from both A- and B-forms. There are large changes in the phosphate backbone torsion angles (particularly γ) of the purine strand, probably due to the electrostatic interactions between the phosphate groups and the protonated cytosine. These changes narrow the minor groove width of the purine−Hoogsteen strands and may represent sequence-specific structural variations of the DNA triplex.

The formation of DNA triple helices has the potential to block transcription, thus controlling gene expression (*1*, *2*). The identification of triplex DNA-binding proteins provides evidence for a role of triplexes in cellular processes (*3*−*5*). Formation of nucleic acid triple helices was first observed in 1957 in the binding of poly U to poly(A)poly(U) by pairing of pyrimidines to the N7 and 6-substituent positions of purines in double helices and was later extensively studied in other polynucleotides (*6*−*8*). To date, solution structures of intramolecular DNA triplexes have been obtained from NMR studies (*9*−*11*), and an earlier report of fiber X-ray diffraction has appeared (*12*). Unusual DNA triplet interactions were observed in crystal contacts (*13*). Therefore, there has been no previous report of a high-resolution three-dimensional X-ray crystallographic study of a DNA triple helix.

The formation of isolated triple base interactions of the kind proposed for triplex DNA has been observed in many RNA structures (*14*). Also, a triplex complex has been observed for two PNA (protein nucleic acid) pyrimidine strands bound to a purine strand with normal deoxyribose-phosphate backbone (*15*). In the PNA strands, the bases were linked by a peptidic backbone, and the complex formed a triplex structure with unique helical dimensions termed P-form DNA.

Previous efforts in this laboratory to crystallize a triplex DNA construct yielded only a fiber-type X-ray diffraction pattern from apparently well-formed single crystals (*16*). Oligomeric triple helices also frequently showed smearing by gel electrophoresis, indicating equilibria among multiple species and possibly causing difficulties in crystallization. Therefore, electrophoresis was used to screen candidate structures, and only those showing sharp bands on gels were investigated further. Since numerous DNA duplexes were known to diffract well, the construct selected (Figure 1) was designed with duplex portions in the expectations that these would favor crystal-packing interactions that would improve ordering in the crystal. To establish that the triple helix is intact under the conditions of crystallization, we measured temperature profiles of absorbance as a function of pH at the same concentrations of Na, Mg, spermidine, and MPD as those used for crystallization. The characteristic temperature profiles and pH dependence show that the triplex remains intact under the conditions of crystallization.

MATERIALS AND METHODS

Crystallization. Oligonucleotides were synthesized using a solid-phase cyanoethylphosphoramidite method with 5-bromouracil used as an analogue for thymine and purified on gel-electrophoresis. The crystals were grown at room temperature and were flash frozen for data collection.[1] In summary, the best crystals were grown with additive solution containing 0.04 M sodium cacodylate (pH 5.04), 0.2 M NaCl, 0.1 M $MgCl_2$, 0.01 M spermine 4 HCl, and 12% MPD and reservoir solution containing 47% MPD and 0.17 M NaCl. Modified 12-mer oligonucleotide with phosphorothiolate linkage between T5 and C6 gave the best diffracting crystals for no apparent reason.

Data Collection and Structure Determination. The crystal structure of brominated oligonucleotides was determined by

‡ The coordinates of the structure have been deposited in the Protein Data Bank (PDB) under the name 1D3R.

* To whom correspondence should be addressed. Phone: (301) 496-4295. Fax: (301) 496-0201. E-mail: David.davies@nih.gov.

[1] Han, Rhee, Liu, Miles, and Davies (1999) *Acta Crystallogr*. (in press).

10.1021/bi991811m
Published on Web 12/03/1999

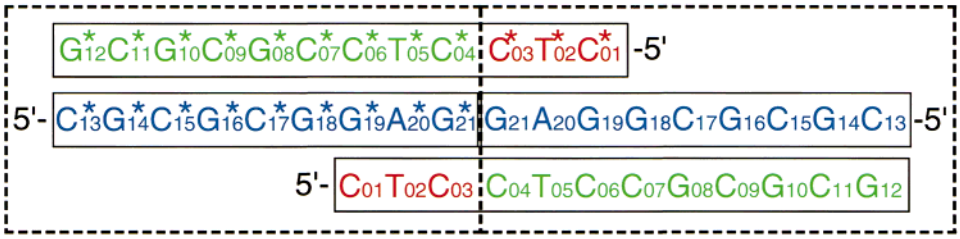

FIGURE 1: Sequences and numbering of oligonucleotides in an asymmetric unit. The 12-mer numbers from 1 to 12 and the 9-mer from 13 to 21 with the second half of duplex being indicated by an asterisk mark. The bases are colored as follows: the Hoogsteen bases in red, the purine-rich (Crick) strand in blue, and the pyrimidine-rich (Watson) strand in green. Two independent triplex and duplex DNA, which are used for analysis, are enclosed with the dashed boxes. The two halves are related by a 2-fold symmetry and the corresponding atoms are superimposed with rms deviations of 0.5 Å. In these crystals, 5-bromouracil was used as an analogue of thymine.

Table 1: Data Collection and Refinement Statistics

	wavelength (Å)	resolution (Å)	no. of reflections (redundancy)	completeness[a] (%)	R_{sym}[a] (%)	MAD analysis (20−2.0 Å) $R_{cullis}[R_{cullis(ano)}]$[b]	phasing power[c]	FOM[d]
Br λ_1	0.920 04	20.0−1.8	11 408 (11.2)	99.2 (99.8)	4.3 (40.4)	(0.39)		0.78
Br λ_2	0.919 39	20.0−1.8	11 077 (10.2)	99.2 (99.9)	3.8 (40.6)	0.75 (0.40)	1.04	
Br λ_3	0.914 98	20.0−1.8	11 005 (9.4)	98.6 (99.8)	3.9 (36.7)	0.60 (0.40)	1.73	
House	1.5418	40.0−2.0	8 106 (6.4)	98.0 (98.0)	3.0 (13.7)			

Refinement Statistics

resolution (Å)	reflections ($\|F\| > 4\sigma$)	R_{factor} (R_{free})[e]	no. of atoms DNA	no. of atoms water	average *B*-factors (Å^2) DNA	average *B*-factors (Å^2) water	rms deviations bond length (Å)	rms deviations bond angles (deg)
20.0−1.8	19 206	0.210 (0.275)	823	157	28.0	37.3	0.012	2.16

[a] Reflections with ($I/\sigma I \geq -3.0$) were included in data processing and values in parentheses are for the shell 1.80−1.86 Å of MAD data and 2.05−2.00 Å of house data. $R_{sym} = \sum|I - \langle I\rangle|/\sum I$. [b] $R_{cullis} = \sum||F_{PH} \pm F_P| - F_H|/\sum|F_{PH} \pm F_P|$ where F_P and F_{PH} are the native and derivative observed structure amplitudes, respectively, and F_H is the calculated heavy-atom structure amplitude. $R_{cullis(ano)} = \sum|D_{PHobs} - D_{PHcal}|/\sum D_{PHobs}$, where D_{PHobs} and D_{PHcal} are the observed and calculated anomalous differences for F_{PH}. [c] Phasing power = rms ($\langle F_H\rangle/E$), where E is the residual lack of closure error. [d] Mean figure of merit (FOM) = $\sum|F_{best}|/F$. [e] $R = \sum|F_C - F_O|/\sum F_O$ R_{free} is the same as R, but for 10% of the data that was not used for the refinement. With the zero σ cutoff ($|F| > 0\sigma$, 22 407 reflections), R_{factor} is of 0.224, and R_{free} of 0.283.

multiwavelength anomalous diffraction (MAD)[2] method. MAD data were collected on a MAR imaging plate at beam line X9B at the Brookhaven National Synchrotron Light Source. The diffraction data were collected at 95 K with an inverse beam method in 1.0° oscillation frames. Wavelengths λ_1, λ_2, and λ_3 near or at the K absorption edge of bromine were chosen so that the dispersive differences were maximized between λ_1 and λ_3 and that the anomalous differences were maximized at λ_2. The collected data at three wavelengths were independently integrated and scaled using the program package HKL (*17*). The crystal belongs to a space group $P4_2$ with unit cell parameters $a = b = 53.8$, $c = 43.1$ Å and each asymmetric unit contains two molecules of duplex DNA. The data set collected at the wavelength λ_1 were considered as the native (Table 1). A program suite CCP4 package (*18*) was used for phasing. Four Br sites were initially located in various difference Patterson maps and further verified by SHELXS-97 (*19*). The initial positions of four Br sites were refined using MLPHARE with data of 20.0−2.0 Å and the experimental phases were further extended by solvent flattening and histogram matching implemented in DM. The map at this stage clearly shows every nucleic acid atom except for highly disordered C13*. The model was built using O (*20*) with help of strong anomalous peaks (>10σ) for bromine atom in the Br-uracil bases and was then refined to a resolution of 1.8 Å using CNS (*21*) with the Parkinson et al. parameters (*22*) and with conformational restraints of the B-form DNA. Additional data were collected using an in-house X-ray generator (CuKα; $\lambda = 1.5418$ Å) with a R-axis IV imaging plate.

[2] MAD, multiwavelength anomalous diffraction.

RESULTS AND DISCUSSION

Overall Features. The crystal structure formed by the 1:1 complex of the 12-mer and 9-mer was determined using a multiwavelength anomalous dispersion (MAD) method (Table 1, Figure 2). The Hoogsteen strand is positioned within the major groove of the B DNA structure. This results in the formation of two additional minor grooves, between the Hoogsteen strand and the Crick and Watson strands.

In the crystal, the asymmetric unit contains one copy of the construct shown in Figure 1. The two halves of the structure are related by a noncrystallographic 2-fold symmetry. Due to the palindromic orientations of these four strands, there are two independent triple and double helical regions within the asymmetric unit. The three bases, C1, BU2, and C3, which form the triplex region, are at the 5′ end of a dodecanucleotide. The next three bases, C4, BU5, and C6, form Watson and Crick bonds with G21, A20, and G19, of the purine-rich strand, which is also bonded in a triplex arrangement to C3*, BU2*, and C1*. The remaining six nucleotides of this strand form a regular Watson and Crick duplex. The two purine-rich strands in blue are antiparallel in this structure and the 3′ guanosine ends (G21 and G21*) are positioned by the triplex interactions. The

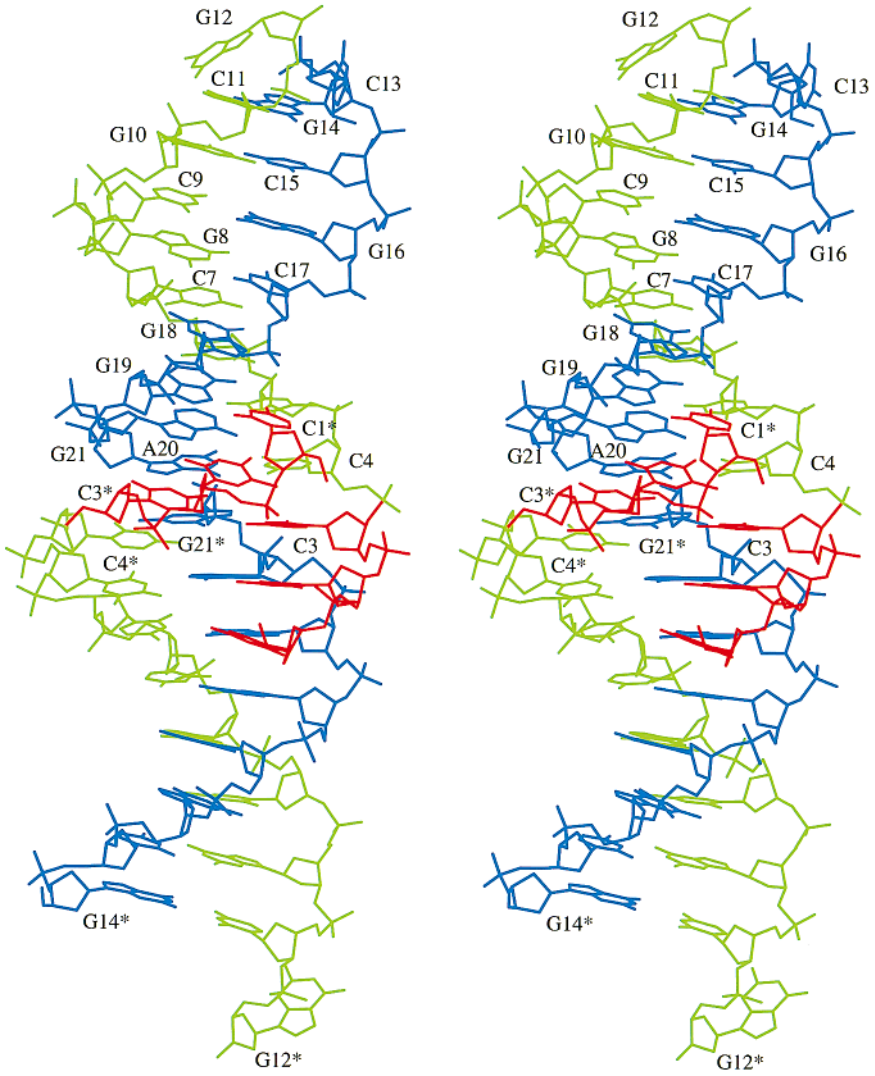

FIGURE 2: Stereodiagram of the duplex and triplex in the asymmetric unit. Color codes are identical as were shown in Figure 1. There is a narrow groove of the purine-rich (blue)— Hoogsteen (red) strands at about C3* (see text more details). Program MOLSCRIPT (*29*) was used for generating this figure.

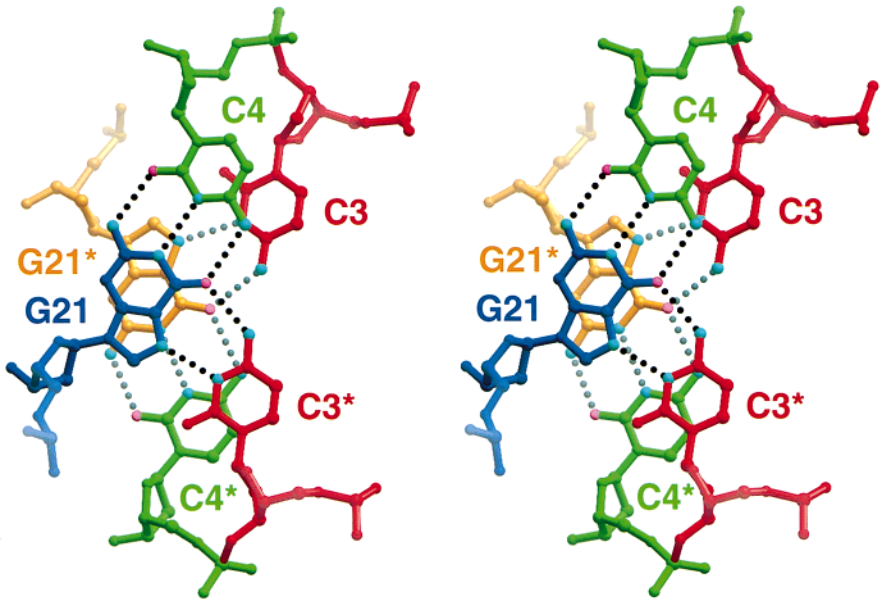

FIGURE 3: A close view of the area in the G21 and G21* along the DNA helical axis. Due to the 2-fold symmetry, there are switches between the Hoogsteen (red) and pyrimidine-rich (green) strand in the 12-mer strand. For clarity, G21* is indicated with yellow. Program RIBBONS (*30*) was used for generating this figure.

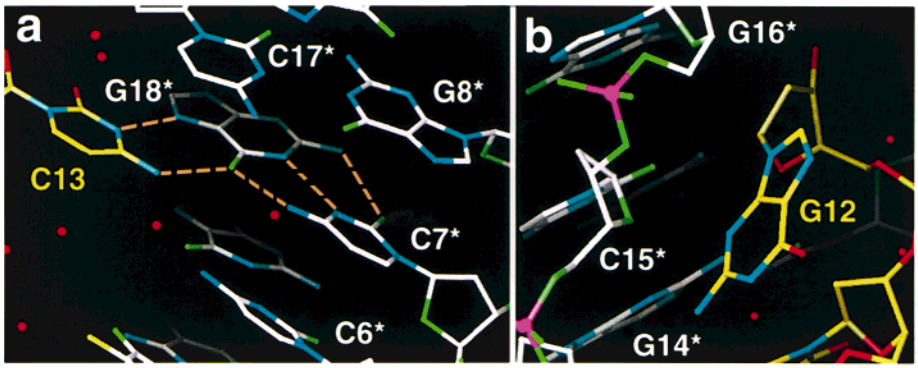

FIGURE 4: Crystal-packing interactions of (a) C13 and (b) G12. A symmetry-related molecule is indicated in white.

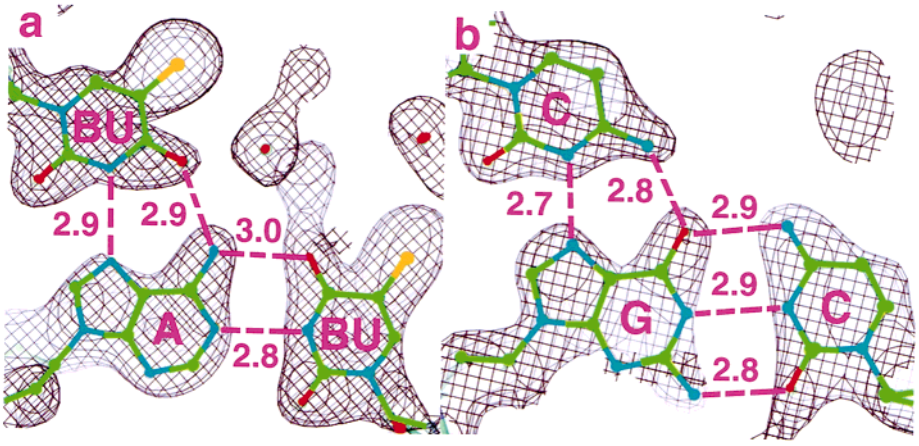

FIGURE 5: Initial electron density map and interatomic distances (Å) for possible hydrogen bonds in triplet bases (a) for BU·ABU and (b) for C^+·GC. Carbon is colored green, nitrogen blue, and oxygen red, and bromine orange. The map is contoured at 1σ.

Table 2: Helical Parameters and Average Torsion Angles of DNA

	Helical Parameters[a]					
	twist (deg)	rise (Å)	inclination (deg)	X-displacement (Å)	opening (deg)	sugar pucker
duplex	34.5	3.2	12.3	−1.2	−0.5	C2′−endo
triplex[b]	32.5	3.3	5.6	−2.0	2.7	C2′−endo
A-DNA[c]	33	3.0	13.0	−4.5		C3′−endo
B-DNA[c]	36	3.4	2.4	0.8		C2′−endo

	Torsion Angles (deg)						
molecule	α	β	γ	δ	ϵ	ζ	χ
duplex	−42	165	27	151	−151	−139	−89
triplex[b]	−17	−163	−52 (−33)	155	−163	−97	−104
Hoogsteen[c]	−40	−170	37	144	−166	−98	−118
A-DNA[d]	−50	172	41	79	−146	−78	−154
B-DNA[d]	−61	180	57	122	173	−91	−119

[a] Helical parameters and torsion angles are calculated with the program CURVES (*21*). [b] Values for triplex are from the Watson−Crick pair in the triplexes excluding those involved in crystal packing. For the γ angle, an averaged value is in a parenthesis that does not include the values of A20 and A20*. [c] Values for Hoogsteen are from C1, T2, C3, C1*, T2*, and C3*. [d] The values are given for the standard A- and B-DNA (*13*).

unique 2-foldsymmetry around G21 and G21* makes the transition between the Hoogsteen and Watson strands possible without significantly altering the conformation of the sugar−phosphate backbones (Figure 3).

Crystal Packing. The terminal bases of the duplex regions, C13−G12, do not form Watson−Crick base pairs but swing out (Figure 2) and engage in crystal-packing contacts (Figure 4). C13 makes Hoogsteen hydrogen bonds to G18* of the symmetry-related molecule and forms a C^+·GC triplet. The other terminal base G12 is stacked against the sugar ring of another symmetry-related molecule. These interactions help to establish crystal packing and are consistent with the hypothesis that duplex regions would favor crystal packing interactions.

Triplet Base Pairs. Figure 5 shows a 2.0 Å initial map from MAD phasing, which is overlaid on the final model and interatomic distances of possible hydrogen bonds in the triplet base pairs. In particular, Hoogsteen pairs between C and G require protonation of the N3 atom of cytosine. The distance (2.7 Å) between N3 of C^+ and N7 of G clearly indicates the presence of a hydrogen bond under our crystallization conditions (pH of 5.2). Beside these hydrogen

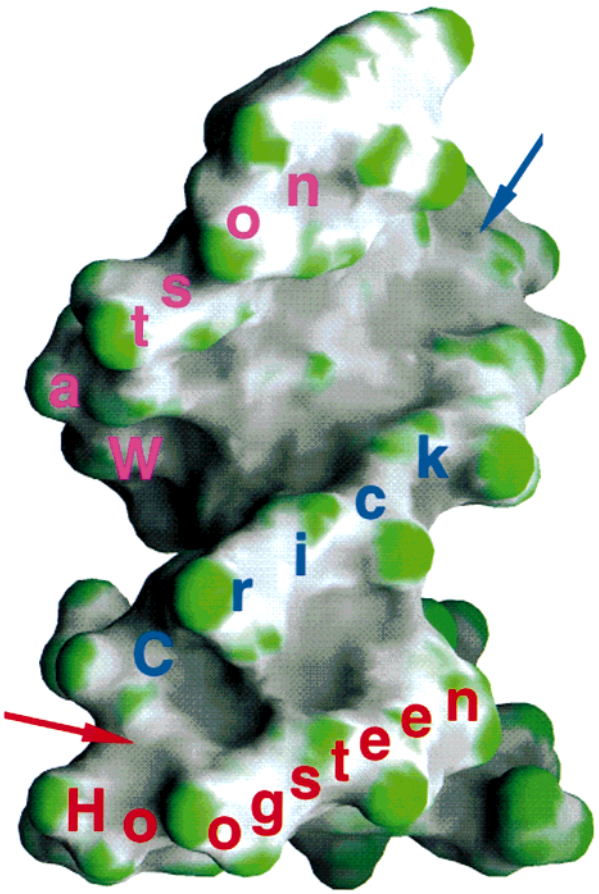

FIGURE 6: Molecular surface of the triplex and duplex regions. The orientation is identical to one in Figure 2. The minor groove is indicated with a blue arrow for the Watson−Crick strands and a red arrow for the Crick-Hoogsteen strands. The most convex, concave, and planar regions of the surface are coded green, gray, and white, respectively. Figure is prepared by using the program GRASP (*31*).

bonds, the 2-keto groups of the Hoogsteen pyrimidines are stacked over the aromatic rings of the 3′-flanking adjacent bases as found in other unusual DNA structures (*14*). In addition, some of these keto oxygen atoms make water-mediated hydrogen bonds with phosphate groups of the parallel purine strand or sugar oxygens on the same Hoogsteen strand. These interactions contribute to the triplex stability.

Analysis of Triplex DNA. In this structure each segment of triplex DNA is quite short, consisting of three base triplets. Nevertheless, the structure does provide overall dimensions for the helical regions which lie between the range of canonical dimensions for A- and B-DNA (Table 2). Most of the nucleotides clearly adopt an S-type (C2′-endo) sugar conformation as found in B-DNA, except for the two 5′-terminal Hoogsteen bases C1 and C1* in which the electron density map indicates an equilibrium between the C2′-endo and the C3′-endo conformation.

The triplex−duplex junctions along each pyrimidine-rich strand of the complex (i.e., between C6 and C7) do not show any abrupt changes in helical parameters. However, the lower twist (32.5°) and higher opening (2.7°) angles (*23*) of the Watson−Crick pairs within the triplexes indicate that the transition from duplex to triplex results in a partial unwinding of the DNA. In addition, the Watson−Crick pairs within the triplex become more perpendicular to the helix axis (i.e., lower inclination angles defined as in ref *23*) and are displaced toward the minor groove (i.e., −2 Å in X-displacement). These general features are similar to the conclusions from NMR studies using intramolecular triplexes despite some differences in the helical parameters and sugar conformations (*9*−*11*). In the duplex regions (C7−C11), no big deviations from canonical B-DNA are observed. These differences in conformation between the triplex and duplex are relatively small, and consequently, the passage of each strand from triplex to duplex is achieved quite smoothly.

Effects of C^+ on the DNA Conformation. In the vicinity of the C^+ base, there is an additional narrowing of the groove between the purine-rich (Crick, blue) strand and the Hoogsteen strand (red). This narrowing is accompanied by a change in the torsional angle γ (the exocylic angle about the C5′−C4′ axis) at A20 and A20* (Table 2). In contrast to other nucleotides, A20 and A20* have γ angles of −104° and −179°, respectively. These alterations of γ in A20 (and A20*) are associated with changes in the groove width of the Crick−Hoogsteen strands, leading to a locally narrow groove (distance between the nearest phosphate groups is about 6.0−6.1 Å; Figure 6). These changes probably result from the electrostatic interactions between the positive charge on a protonated cytosine in the Hoogsteen strand and the negative charge on the phosphate group in the purine-rich strand, in agreement with the conclusions of Asensio et al. (*24*). The interatomic distance between the protonated cytosine bases (C1, C3, C1*, and C3*) in the Hoogsteen strand and the nearest phosphate group of the Crick strand is within a range of 3.9−5.0 Å, while the uncharged bromouracil bases in Hoogsteen strand are about 6.5 Å from the corresponding phosphate group.

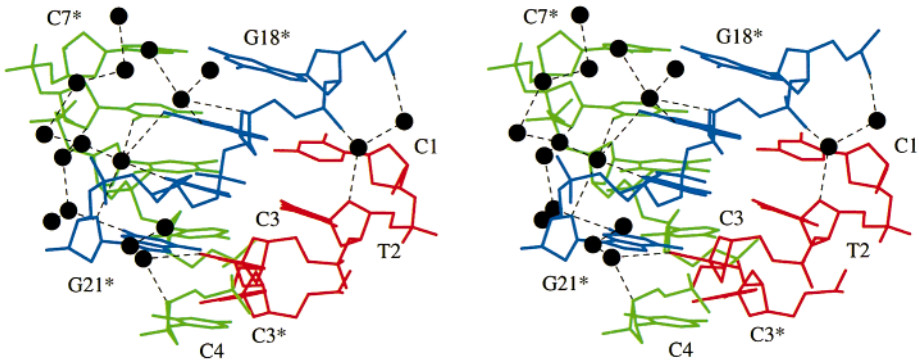

FIGURE 7: Stereoview of solvent molecules in the triplex region. Color codes for the strands are same as before, and solvent molecules are indicated with a filled circle. Possible hydrogen bonds are indicated with the dashed lines.

Water and Metal-Binding Sites. The current model contains a total of 157 solvent molecules in the asymmetric unit. Most are assigned as water molecules and interact with DNA via direct and water-mediated indirect interactions. Figure 7 shows water molecules near the triplex region. As seen in other DNA structures, there is a spine of water in the minor groove (*25*−*27*) of the Watson−Crick duplex, but not in the restricted groove of the Crick−Hoogsteen duplex, in particular, near the phosphate groups of A20 and A20*. The geometry of the coordination and anomalous peak from data using CuKα radiation suggest that, in addition to the bromine sites, there are at least three more heavy-atom binding sites. Two of these sites are near the phosphate oxygens of G12 and have an octahedral coordination to water molecules with average distance of 2.1 Å. On the basis of the coordination geometry and the interatomic distances, these two sites are considered to be occupied by Mg, which is included in our crystallization buffer. A third site with a strong anomalous peak ($>10\sigma$) is observed in the triplex region. It has a square planar coordination with two water molecules and N4 atoms of pyrimidine C3 and C3* and an average interatomic distance of about 3.1 Å, and we have assigned this to a chloride ion. It is noteworthy that this site involves interactions with the two positively charged cytosines. Although divalent cations were required to form a three stranded nucleic acid (*6*), it is not apparent from this crystal structure that metal ions have any specific structural roles in stabilizing the triple-helical DNA as was shown in the G-tetraplexes (*28*).

CONCLUSION

Design of DNA sequences in this study provides a novel approach for crystallographic study of DNA triplex and its junction with a duplex. The overall triplex dimensions differ somewhat from those of B-form DNA, but the general topology remains similar. Presence of C^+·GC triplets causes sequence-specific changes in the groove width of the Crick−Hoogsteen strands by the electrostatic interaction between C^+ and the phosphate group of the Crick strand. These conformational changes restrict the access of water to the minor groove and may be unique structural features for the recognition of triplex DNA by protein molecules.

ACKNOWLEDGMENT

We thank Zbigniew Dauter for help in data collection at beam line X9B at the Brookhaven National Synchrotron Light Source.

REFERENCES

1. Sun, J.-s., Garestier, T., and Helene, C. (1996) *Curr. Opin. Struct. Biol. 6*, 327−333.
2. Chan, P. P., and Glazer, P. M. (1997) *J. Mol. Med. 75*, 267−282.
3. Musso, M., Nelson, L. D., and Van Dyke M. W. (1998) *Biochemistry 37*, 3086−3095.
4. Guieysse, A.-L., Praseuth, D., and Helene, C. (1997) *J. Mol. Biol. 267*, 289−298.
5. Kiyama, R., and Camerini-Otero, R. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10450−10454.
6. Felsenfeld, G., Davies, D. R., and Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023−2024.
7. Howard, F. B., Frazier, J., Lipsett, M. N., and Miles, H. T. (1964) *Biochem. Biophys. Res. Commun. 17*, 93−102.
8. Morgan, A. R., and Wells, R. D. (1968) *J. Mol. Biol. 37*, 63−80.
9. Radhakrishnan, I., and Patel, D. J. (1994) *Biochemistry 33*, 11405−11416.
10. Radhakrishnan, I., and Patel, D. J. (1994) *Structure 2*, 17−32.
11. Tarköy, M., Phipps, A. K., Schultze, P., and Feigon, J. (1998)-*Biochemistry 37*, 5810−5819.
12. Arnott, S., and Selsing, E (1974) *J. Mol. Biol. 88*, 509−521.
13. Vlieghe, D. Meervelt, L. V., Dautant, A., Gallois, B., Precigoux, G., and Kennard, O. (1996) *Science 273*, 1702−1705.
14. Saenger, W. (1984) *Principles of nucleic acid structure*, pp 132−133, Springer-Verlag, New York.
15. Betts, L., Josey, J. A., Veal, J. M., and Jordan, S. R. (1995) *Science 270*, 1838- 1841.
16. Liu, K., Miles, H. T., Parris, K. D., and Sasisekharan, V. (1994) *Nat. Struct. Biol. 1*, 11−12.
17. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
18. Collaborative Computational Project no. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D 50*, 760−763.
19. Sheldrick, G. M. (1990) *Acta Crystallogr., Sect. A 46*, 467−473.
20. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
21. Brünger, A. T., et al. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
22. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T., and Berman, H. M. (1996) *Acta Crystallogr., Sect. D 52*, 57−64.
23. Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn.. 6*, 655−667.
24. Asensio, J. L., Brown, T., and Lane, A. N. (1999) *Structure 7*, 1−11.
25. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535.
26. Soler-Lopez, M., Malinina, L., Liu, J., Huynh-Dinh, T., and Subirana, J. A. (1999) *J. Biol. Chem. 274*, 23683−23686.
27. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc. 121*, 3590−3595.
28. Howard, F. B., and Miles, H. T. (1982) *Biochemistry 21*, 6736−6745.
29. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
30. Carson, M. (1991) *J. Appl. Crystallogr. 24*, 958−961.
31. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.

BI991811M