Acta Cryst. (1998). D**54**, 577–583

Structure of the DNA Decamer d(GGCAATTGCG) Contains both Major- and Minor-Groove Binding G·(G·C) Base Triplets

Christine M. Nunn* and Stephen Neidle

The CRC Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, England.
E-mail: c.nunn@icr.ac.uk

(*Received 22 August 1997; accepted 6 November 1997*)

Abstract

The crystal structure of the decamer d(GGCAATTGCG) has been determined at 2.4 Å resolution. The central eight bases of each DNA single-strand base pair with a self-complementary strand to form an octamer B-DNA duplex. These duplexes lie end-to-end within the unit cell. The terminal 5′-G and G-3′ bases of each decamer strand are unpaired, and interact with the neighbouring duplexes *via* interactions within both the major and minor groove. This results in base triplets of the type G—(G·C) and G*(G·C), with the third guanine base binding to a Watson–Crick G·C base pair from the major groove and the minor groove, respectively. The triplet interaction of the type G—(G·C) involves Hoogsteen hydrogen-bonding interactions between the two guanine bases. The minor- and major-groove base triplet interactions which exist within this structure act to stabilize the $d(GCAATTGC)_2$ B-DNA octamer duplex.

1. Introduction

Base triplets in crystal structures of DNA can involve third-base binding to either the minor- or major-groove side of a Watson–Crick base pair. The resultant triplets, denoted in this paper as $X—(Y·Z)$ and $X*(Y·Z)$ for third strand X binding to the major groove and minor groove, respectively, of a Watson–Crick base pair $(Y·Z)$, are stabilized by hydrogen-bonding interactions from the third base to the Watson–Crick base pair. These types of interactions have been observed in a number of crystal structures.

Within A-DNA octamer structures crystal packing results in close contact between the minor grooves of symmetry-related duplexes to give interduplex interactions and the formation of base triplets. These interactions include G*(G·C) (Cervi *et al.*, 1992; Shakked *et al.*, 1989), C*(G·C) (Tippin & Sundaralingam, 1996) and G*(A·T) (Courseille *et al.*, 1990), with one hydrogen-bonding interaction from the third base to an A-DNA base pair. Other examples of third-base binding to the minor groove side of a Watson–Crick base pair include the triplet A*(G·C) in the structure of the decamer sequence d(AGGCATGCCT) (Nunn & Neidle, 1996) where one hydrogen-bonding interaction exists between adenine atom N7 and guanine atom N2. This contrasts with an A*(G·C) triplet observed in the structure of the A-DNA octamer d(GGTATACC) (Shakked *et al.*, 1981) in which the adenine base binds *via* atom N3. A triplet of the type G*(G·C) was reported in the structure of

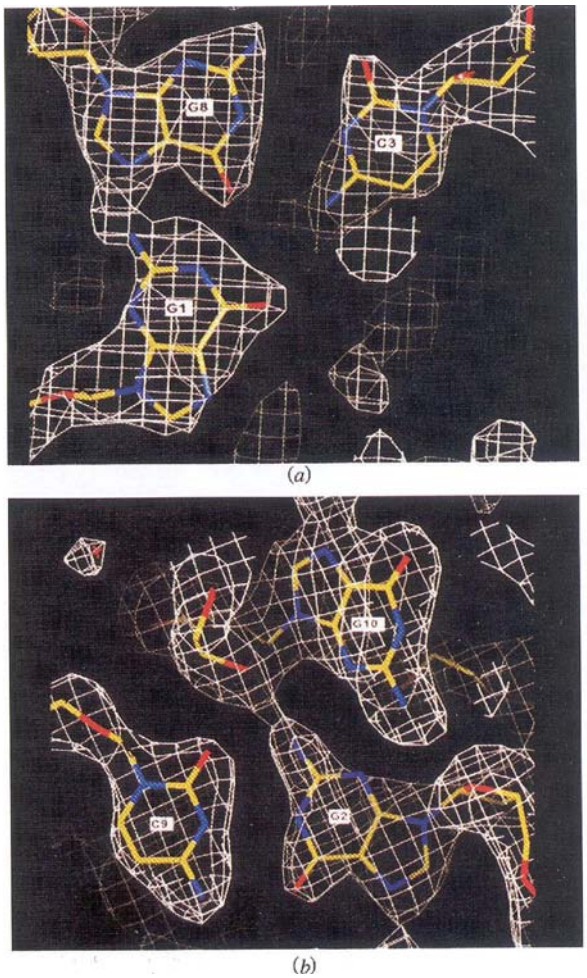

Fig. 1. $1\sigma 2F_o - F_c$ electron-density maps calculated with the final model showing the base triplets (*a*) G1—(G8·C3) and (*b*) G10*(G2·C9).

ISSN 0907-4449

d(CCCGGCCGGG) with the third-strand guanine base hydrogen bonding to both bases of the (G·C) base pair (Ramakrishnan & Sundaralingham, 1993).

Major groove triplets are biologically important due to their presence within DNA triple helices. DNA triplexes are triple-stranded DNA formed by either two purine-rich and one pyrimidine-rich strand or one purine-rich and two pyrimidine-rich strands. The two like-strands in both cases, are oriented antiparallel with respect to one another (Frank-Kamenetskii & Mirkin, 1995; Sun *et al.*, 1996). Crystals of DNA triplexes have been reported in the literature (Lui *et al.*, 1994), however they exhibit poor quality and fibre diffraction. DNA triplex structures have been studied to date using NMR and fibre diffraction techniques (Radhakrishnan & Patel, 1994; Lui *et al.*, 1996), however at the present time there is no high-resolution three-dimensional structure of an extended DNA triplex.

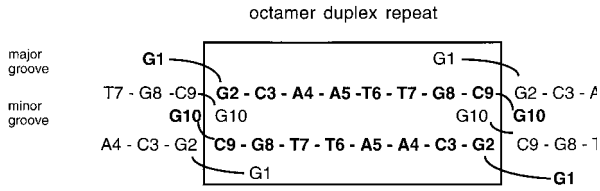

Fig. 2. Schematic view of the overall structure together with the numbering scheme.

Triplet interactions involving third-base binding from the major groove of a Watson–Crick base pair have been observed in only four existing DNA crystal structures. The structures of d(CGCAATTGCG) (Spink *et al.*, 1995), d(GCGAATTCG) (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996*a*), d(GGCCAATTGG) (Vlieghe *et al.*, 1996*b*) and d(CGCAATTGCG)-netropsin (Nunn *et al.*, 1997) contain triplets of the type C—(G·C), G—(G·C), G—(G·C) and C^+—(G·C), respectively. In these structures the terminal 5′ bases are unpaired and interact with the neighbouring pseudocontinuous duplex within the unit cell to form a short stretch of triple-stranded DNA. In the case of the G—(G·C) and C^+—(G·C) triplets, the third-base hydrogen bonds to the

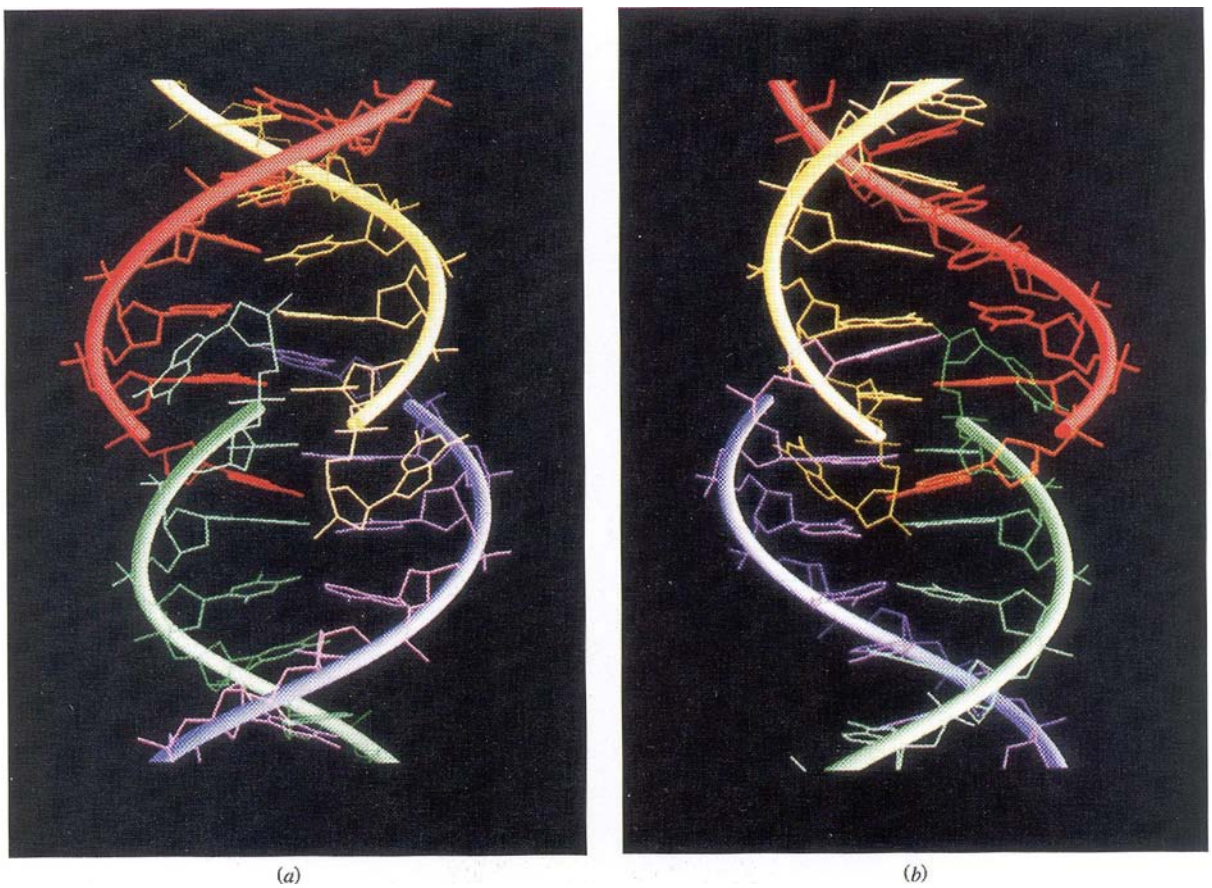

Fig. 3. The interaction between duplexes within the unit cell viewed towards (*a*) the minor groove and (*b*) the major groove. The terminal guanosines lie in the minor groove of neighbouring duplexes whilst the guanine base at the 5′ end of each strand extends into the major groove of a neighbouring duplex. Each symmetry-equivalent strand is shown in a different colour.

(G·C) Watson–Crick base pair of the neighbouring duplex *via* Hoogsteen hydrogen-bonding interactions.

Two crystal structures of the native DNA decamer sequence d(CGCAATTGCG) have been determined in this laboratory, in addition to the structure of the drug-bound complex d(CGCAATTGCG)–netropsin (Nunn *et al.*, 1997). The two native structures have different conformations by virtue of the different crystallizing conditions used. One structure exists as an octamer duplex formed between two self-complementary strands, with the terminal 5′-C and G-3′ bases of each single strand unpaired and interacting with the terminal base pairs of a neighbouring duplex within the unit cell (Spink *et al.*, 1995). This DNA conformation is also observed in the structure of d(CGCAATTGCG)–netropsin (Nunn *et al.*, 1997). The second structure of d(CGCAATTGCG) is a decamer duplex with no unpaired bases. In this structure the decamer duplexes stack end-to-end within the unit cell to form a crossed-helix packing arrangement (Wood *et al.*, 1997).

In this paper we present the crystal structure of the decamer d(GGCAATTGCG). This sequence differs from that found for the structure of d(CGCAATTGCG) in having a 5′-G base replacing the 5′-C. It was anticipated that this non-self-complementary sequence would exhibit a crystal-packing arrangement similar to that found for d(CGCAATTGCG) (Spink *et al.*, 1995) and d(CGCAATTGCG)-netropsin (Nunn *et al.*, 1997) and that the change of sequence from 5′-C to 5′-G should produce a structure with the 5′-G base binding within the major groove of an octamer duplex to form a triplet of the type G—(G·C).

2. Materials and methods

2.1. *Crystallization and data collection*

The DNA decamer d(GGCAATTGCG) was purchased from the Oswel DNA Service (University of Southampton, UK). Crystals were grown by vapour diffusion at 277 K from sitting drops containing 8 µl 3.0 m*M* DNA, 2 µl 350 m*M* $MgCl_2$ and 2 µl 30% 2-methyl-2,4-pentanediol, equilibrated against a 1 ml reservoir of 30% 2-methyl-2,4-pentanediol. The DNA and MPD solutions were prepared using 30 m*M* sodium cacodylate buffer at pH 7.0. Crystals of the decamer were grown within 6 d.

Data collection was carried out at 289 K on a crystal of approximate dimensions 0.5 × 0.2 × 0.2 mm. A Rigaku R-axis IIC image-plate system mounted on a Rigaku RU-200 rotating-anode X-ray generator was used, operating at 100 mA, 50 kV with an MSC double-focusing mirror system. A crystal-to-detector distance of 10 cm produced data with a maximum resolution of 2.0 Å at the detector edge. 100 frames, each with an oscillation angle of 1.5° and an exposure time of 25 min were collected.

Data processing was carried out using the program *DENZO* (Otwinowski & Minor, 1996) and the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The crystal has cell parameters of a = 27.92, b = 39.21, c = 54.04 Å, $\alpha = \beta = \gamma = 90^\circ$, and space group $I2_12_12_1$ ($I222$, allowed by the systematic absences, was excluded by packing considerations, see later). $R_{\rm merge}(I)$ is 8.0% for all 1944 reflections to 2.0 Å. Data between 2 and 2.4 Å were excluded from the structure solution and refinement as $R_{\rm merge}(I)$ values for data shells in this region were greater than 30%. 1177 reflections were used in structure solution and refinement, to the resolution limit of 2.4 Å, representing a 95.5% complete data set.

2.2. *Structure solution and refinement*

The crystallographic asymmetric unit contains a decamer single strand. Inspection of the diffracted intensities suggested a B-DNA structure with the base-pair stacking lying perpendicular to the c axis of the unit cell. The c-axis length of 54.04 Å corresponds to a 16 base-pair repeat along this axis.

The unit cell for this structure are similar to that reported for the DNA decamer d(CGCAATTGCG) (Spink *et al.*, 1995). The space group is $I2_12_12_1$. Space group $I222$ is prohibited as the packing within this space group requires parallel rather than antiparallel strand orientation within the DNA duplexes. The structure of d(CGCAATTGCG) (Spink *et al.*, 1995) has a decamer single strand within the crystallographic asymmetric unit and one base change from the present structure. The nine common bases GCAATTGCG comprise the eight bases GCAATTGC, which are involved in Watson–Crick base-pairing interactions with a complementary strand and the terminal G-3′ base which together with the associated furanose sugar lies within the minor groove of a neighbouring duplex. These nine like-bases common between the two structures were used as an initial refinement model.

Crystallographic refinement was carried out using the program *X-PLOR* (Brünger *et al.*, 1987). Initial rigid-group refinement was carried out for the DNA, firstly as one rigid body and subsequently as nine nucleotides. Positional refinement of the nine like-bases and temperature-factor refinement was carried out prior to the addition of the terminal 5′ guanine base. This terminal base although being poorly defined in the $F_o - F_c$ electron-density maps lies within the major groove of the neighbouring duplex within the unit cell, and within hydrogen-bonding distance of a duplex G·C base pair. The average value for atomic temperature factors for the 5′-G base is 68.9 $Å^2$ as compared with values between 41.0 and 24.2 $Å^2$ for the other nine bases. The final refined position for both the 5′-G (G1) and the G-3′ (G10) bases are shown in Fig. 1 together with the $2F_o - F_c$ map generated at the end of the structure

refinement. After inclusion of the 5′-G base into the refinement, solvent molecules were located and added to the model. At the conclusion of the refinement 68 water molecules had been added, together with one magnesium ion. The final R value is 19.5% for all data with $F > 2\sigma(F)$ in the resolution range 8–2.4 Å. The root-mean-square deviations from idealized values for bond lengths and angles within the DNA are 0.02 Å and 2.8°, respectively. Final atomic coordinates and experimental details have been deposited with the Nucleic Acid Database.† All graphics work was carried out using the program *O* (Jones *et al.*, 1991).

3. Results and discussion

3.1. *Overall structure*

The structure of the decamer d(GGCAATTGCG) is shown schematically in Fig. 2 and comprises a Watson–Crick base-paired B-DNA octamer duplex with the two self-complementary strands GCAATTGC related *via* a twofold symmetry axis through the centre of the duplex. The terminal G-3′ base of each single strand lies within the minor groove of a symmetry-related duplex within the unit cell. This base hydrogen bonds across the minor groove–minor groove interface, with the terminal guanine base of a symmetry-related duplex. The terminal 5′-G base of each strand lies in the major groove of a symmetry-related duplex with hydrogen-bonding contacts to the guanine base of the penultimate G·C base pair of the octamer duplex.

3.2. *DNA structure*

Within the unit cell, each of the octamer duplexes lie end-to-end with a symmetry-related octamer duplex, to form pseudocontinuous DNA helices lying collinear with the direction of the crystallographic c axis. At the intersection between duplexes there is a rise of 3.36 Å between the terminal C9·G2 and G2·C9 base pairs and a pseudo-helical twist angle of 65°. At the junction of the two octamer duplexes, the 3′ end of each strand contacts the minor groove of a symmetry-related duplex, while the 5′ end contacts the major groove. These duplex junction interactions are illustrated in Fig. 3.

The backbone torsion angles are close to standard values for canonical B-DNA with average values of 298 (α), 186 (β), 51 (γ), 142 (δ), 171 (ε), 261 (ζ) and 267° (χ). The sugar puckers are of C2′-*endo* type for all except those associated with bases G1 and G7, for which the pseudorotation angles are slightly greater than 180° and the sugar puckers are C3′-*exo*. The intrastrand $P\cdots P$ separations along each strand vary between 6.1 and 7.3 Å. The central AT-rich region of the $d(GCAATTGC)_2$ duplex has large propeller twist values and a narrow minor groove. Fig. 4 shows the minor-groove width for d(GGCAATTGCG) and the similar structure d(CGCAATTGCG) (Spink *et al.*, 1995) together with the values for propeller twist. d(GGCAATTGCG) displays high propeller twist values for all A·T base pairs and a more narrow minor groove than d(CGCAATTGCG). The high propeller twist for the A·T base pairs are associated with bifurcated hydrogen bonds; in addition to the Watson–Crick

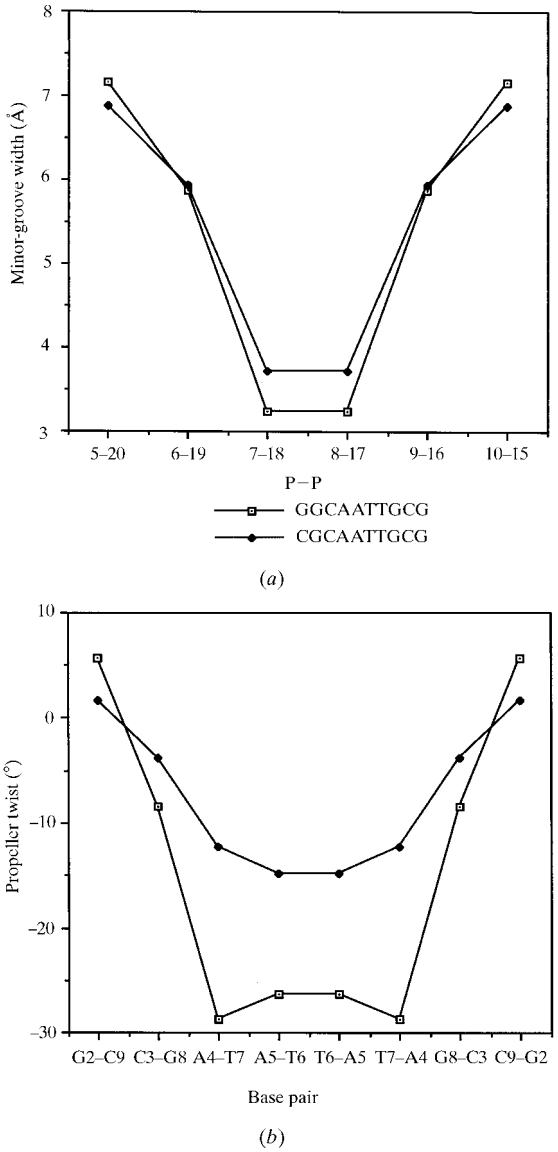

Fig. 4. (*a*) Minor-groove width (intrastrand P–P distance minus 5.8 Å) and (*b*) the propeller twist angles for the structures d(GGCAATTGCG) and d(CGCAATTGCG) (Spink *et al.*, 1995) calculated with *NEWHEL*92.

† Atomic coordinates and structure factors have been deposited with the Nucleic Acid Database (Reference: UDJ061). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: HA0165).

hydrogen bonds between the adenine and thymine bases additional hydrogen bonds are present between atom N4 of base C3 and atom O4 of base T7 (3.02 Å) and atom N6 of base A4 and atom O4 of base T6 (2.65 Å).

3.3. *The guanine-3′ base forms a minor-groove G*(G·C) triplet*

The 3′-terminal guanosine nucleotides lie within the minor groove of a neighbouring duplex within the unit cell. Hydrogen-bonding stabilizes this interaction *via* N2···N3 contacts between guanine bases across the minor-groove interface (Fig. 5). The dihedral angle between the two guanine bases is 46°. Base G10 makes hydrogen-bonding interactions with base G2 *via* two interactions of the type N2···N3. In addition, atom N2 of G10 hydrogen bonds to the O4′ deoxyribose sugar ring atom of base G2, while atom N2 of base G2 similarly binds to the O4′ deoxyribose sugar ring atom of base G10.

These guanine–guanine base interactions are similar to those observed in all the dodecamer structures of the type d(CGX_8GC) where minor-groove interlocking occurs for the terminal two C·G bases at each end of the duplex. This type of interaction has also been observed in the structure of the native decamer sequence d(CGCAATTGCG) (Spink *et al.*, 1995) and for d(CGCAATTGCG) bound to the antibiotic netropsin (Nunn *et al.*, 1997). In all of these structures the entire guanosine lies within the minor groove of the duplex formed between two self-complementary strands of the type GCAATTGC, with hydrogen-bonding interactions between the terminal guanine base and the **G**CAATTGC base of each DNA strand. It is interesting to compare the position of the terminal O3′ atom of the guanosine within the minor groove of these three structures, as it is the separation between these symmetry-related O3′ atoms which defines the length of the ligand-accessible minor groove. In the case of d(CGCAATTGCG) (Spink *et al.*, 1995) the two 10O3′ atoms lie approximately coplanar with the C3·G8 base pair. There is a hydrogen-bonding interaction from 10O3′ to atom N2 of base G8 and five base-pair steps lie between the two symmetry-equivalent 10O3′ atoms along the minor groove. The recently reported crystal structure of netropsin binding to this sequence has lost the twofold symmetry axis through the centre of the structure and there is an asymmetry for the position of 10O3′. The position of the two O3′ atoms has moved and there is no longer a hydrogen-bonding interaction with atom N2 of the guanine base in the penultimate base pair of the octamer duplex. Five base-pair steps lie between the two terminal O3′ atoms in the structure of d(CGCAATTGCG)-netropsin. For d(GGCAATTGCG) the two O3′ atoms are symmetry-equivalent and lie midway between base pairs C3·G8 and A4·T7 with no hydrogen-bonding interactions to the DNA. This results in a shortened ligand-accessible minor groove and four base-pair steps lie between these symmetry-equivalent O3′ atoms.

3.4. *The 5′-guanine base forms a major-groove G—(G·C) triplet*

At the 5′ end of the DNA strand an unpaired guanine base lies in the major groove of the neighbouring octamer duplex within the unit cell to form a G—(G·C) base triplet. The position of this base is well defined by the omit difference density map, and lies within hydrogen-bonding distance of the penultimate base pair C3·G8 of the octamer duplex. The hydrogen-bonding interactions between the three bases are of the type known to be present within triple-stranded or triplex DNA (Thuong & Hélène, 1993) and are shown in Fig. 6. The third-strand guanosine lies parallel with respect to

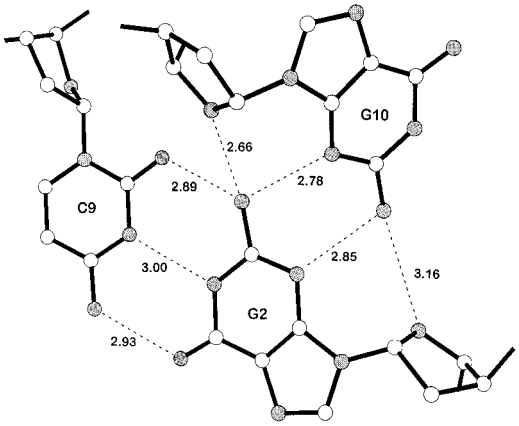

Fig. 5. The G*(G·C) triplet interaction, together with hydrogen-bonding distances in Å.

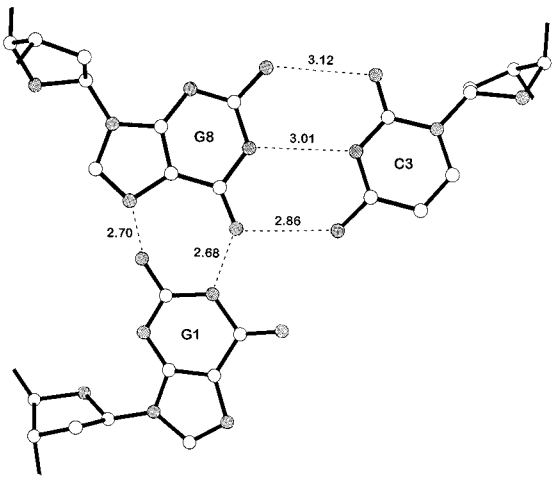

Fig. 6. The G—(G·C) triplet interaction, together with hydrogen-bonding distances in Å.

the purine strand of the Watson–Crick duplex with Hoogsteen base pairing. The dihedral angle between base G1 and base G8 is 10°.

Recent crystal structure determinations have been carried out for the sequences d(GCGAATTCG) and d(GGCCAATTGG). Both of these structures contain major-groove base triplets of the type observed in this structure (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996*a*,*b*) with the terminal 5′-G and 5′-GG bases binding to the next-neighbour pseudocontinuous duplex with a parallel orientation of the 5′-G bases with respect to the guanine base of the d(CGAATTCG)$_2$ duplex. In these two structures there is a hydrogen-bonding interaction with the terminal base pair of the octamer duplex *via* Hoogsteen hydrogen-bonding interactions which involves an additional hydrogen-bonding interaction between atom O6 of base G1 and atom N4 of base C3. An interaction of this type is not seen in this structure (O6—N4 = 3.72 Å).

Within the structure d(GGCAATTGCG) the 5′-G base chooses not to hydrogen bond with the terminal base pair, but with the penultimate base pair of the octamer duplex. A similar situation was observed in the crystal structure of d(CGCAATTGCG)–netropsin (Nunn *et al.*, 1997). Binding to the terminal base pair would require the third-strand guanine base to hydrogen bond with the cytosine base of the terminal C·G base pair. This is unfavourable due both to a lack of potential hydrogen-bonding interactions and the close proximity of two amine groups.

3.5. *Crystal packing*

X-ray crystal structural studies of DNA octamer sequences have produced many octamer structures of A-DNA type. Coordinates for 37 A-DNA structures are currently available from the Nucleic Acid Database (Berman *et al.*, 1992). Only within the last year has the structure of a B-DNA octamer structure been reported in the literature (Tereshko *et al.*, 1996). This structure, for the sequence d(CGCTAGCG), differs from the other octamer DNA sequences in having terminal bases of the type d(CGX_4CG). As has previously been observed for dodecamer sequences of the type d(CGX_8CG) (Dickerson *et al.*, 1987) the crystal packing for d(CGX_4CG) involves the interlocking of the two terminal base pairs at each end of the octamer duplex via interactions between guanine bases across the minor groove. Interlocking of this type involves hydrogen-bonding interactions between atoms N2 and N3 of the two guanine bases at both the terminal and the penultimate base positions at each end of the octamer.

The structure of d(GGCAATTGCG) and those of the sequences d(CGCAATTGCG) (Spink *et al.*, 1995) an d(CGCAATTGCG)–netropsin (Nunn *et al.*, 1997), all exist as B-DNA duplexes with the central eight bases of each decamer strand forming a B-DNA octamer duplex with a complementary strand. Hydrogen-bonding interactions occur between atoms N2 and N3 of the terminal G-3′ base and the guanine base of the terminal base pair of the neighbouring octamer duplex within the unit cell. These hydrogen-bonding interactions between guanine bases produce triplets of the type G*(G·C), which are presumably highly stabilizing in much the same way as found for the two terminal C·G base pairs in octamer and dodecamer B-DNA structures of the type (CGX_nGC) where n = 4 or 8 (Dickerson *et al.*, 1987). These interactions allow DNA helices to pack in columns within the unit cell and act to stabilize B-DNA octamer duplexes.

This work was supported by the Cancer Research Campaign (grant No. SP1384).

References

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasen, A. R. & Schneider, B. (1992). *Biophys. J.* **63**, 751–759.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Cervi, A. R., D'Estaintot, B. L. & Hunter, W. N. (1992). *Acta Cryst.* B**48**, 714–719.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Courseille, C., Dautant, A., Hospital, M., D'Estaintot, B. L., Precigoux, G., Molko, D. & Teoule, T. (1990). *Acta Cryst.* A**46**, FC9–FC12.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L. & Pjura, P. E. (1987). *J. Biomol. Struct. Dyn.* **5**, 557–579.

Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995). *Annu. Rev. Biochem.* **64**, 65–94.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Lui, K., Miles, H. T., Parris, K. D. & Sasisekharan, V. (1994). *Nature Struct. Biol.* **1**, 11–12.

Lui, K., Sasisekharan, V., Miles, H. T. & Raghunathan, G. (1996). *Biopolymers*, **39**, 573–589.

Nunn, C. M., Garman, E. & Neidle, S. (1997). *Biochemistry*, **36**, 4792–4799.

Nunn, C. M. & Neidle, S. (1996). *J. Mol. Biol.* **256**, 340–351.

Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.

Radhakrishnan, I. & Patel, D. J. (1994). *Biochemistry*, **33**, 11405–11416.

Ramakrishnan, B. & Sundaralingham, M. (1993). *J. Mol. Biol.* **231**, 431–444.

Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. (1989). *Nature (London)*, **342**, 456–460.

Shakked, Z., Rabinivich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A. & Viswamitra, M. A. (1981). *Proc. R. Soc. London Ser. B*, **213**, 479–487.

Spink, N., Nunn, C. M., Vojtechovsky, J., Berman, H. M. & Neidle, S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10767–10771.

Sun, J., Garestier, T. & Hélène, C. (1996). *Curr. Opin. Struct. Biol.* **6**, 327–333.

Tereshko, V., Urpi, L., Malinina, L., Huynh-Dinh, T. & Subirana, J. A. (1996). *Biochemistry*, **35**, 11589–11595.

Thuong, N. T. & Hélène, C. (1993). *Angew. Chem. Int. Ed. Engl.* **32**, 666–690.

Tippin, D. B. & Sundaralingam, M. (1996). *Acta Cryst.* D**52**, 997–1003.

Van Meervelt, L., Vlieghe, D., Dautant, A., Gallois, B., Precigoux, G. & Kennard, O. (1995). *Nature (London)*, **374**, 742–744.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B., Précigoux, G. & Kennard, O. (1996*a*). *Acta Cryst.* D**52**, 766–775.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B., Précigoux, G. & Kennard, O. (1996*b*). *Science*, **273**, 1702–1705.

Wood, A. A., Nunn, C. M., Trent, J. O. & Neidle, S. (1997). *J. Mol. Biol.* **269**, 827–841.