FEBS Letters 416 (1997) 86–89 FEBS 19278

A model for the [C^+-G•C]$_n$ triple helix derived from observation of the C^+-G•C base triplet in a crystal structure

Christine M. Nunn, John O. Trent, Stephen Neidle*

CRC Biomolecular Structure Unit, The Institute of Cancer Research, Cotswold Road, Sutton, Surrey SM2 5NG, UK

Received 13 May 1997; revised version received 27 August 1997

Abstract A molecular modelling study on the [C^+-G•C]$_n$ triple helix is reported. We have observed the C^+-G•C base triplet in the crystal structure of an oligonucleotide-drug complex, between the minor-groove drug netropsin and the decanucleotide d(CGCAATTGCG)$_2$. The complex was crystallised at pH 7.0, but the crystal structure, at a resolution of 2.4 Å, shows that a terminal cytosine has become protonated and participates in a parallel C^+-G•C base triplet. The structure of this triplet and its associated sugar-phosphate backbones have been energy-refined and then used to generate a triple helix. This has characteristics of the B-type family of DNA structures for two strands, with the third, the C^+ strand, having backbone conformations closer to the A family.

Key words: DNA triple helix; Molecular modelling; Base triplet

1. Introduction

It is now well established that purine stretches of duplex DNA can interact with a parallel pyrimidine third strand to form triple-helical DNA structures [1–3]. Formation of triple helices in a sequence-specific manner is currently being exploited as a means of selectively inhibiting the transcription of therapeutically-relevant genes [4–6] such as the *c-erbB-2* [7] and *bcl-2* oncogenes [8] Stabilisation of triple helices is achieved by Hoogsteen hydrogen bonding between third-strand pyrimidines and duplex purines, to form T-A•T and C^+-G•C base triplets (Fig. 1a,b). Two Hoogsteen hydrogen bonds can only be formed in the latter instance at a pH of ca. 5, since the N3 atom of cytosine has a pK_a of 4.3. Accordingly, formation of triplex DNA with a mixed guanine/adenine purine target strand is pH-dependent and of low affinity under physiological conditions. Two approaches to stabilising cytosine-containing parallel DNA triplexes have been described. In one, cytosine base modifications or mimetics have been devised (see for example [9–11]) which are fully protonated at pH 7.0. Few of these have as yet found widespread use. The second approach utilises DNA-binding ligands which selectively interact with triplex rather than duplex DNA [12–14], and so stabilise the triplex form, although most effective stabilisation probably occurs at T-A•T sequences [15].

Structural information on nucleic acid triple helices at the atomic level is limited to low-resolution fibre diffraction analyses (see, for example, [16,24]) although a number of molecular modelling studies have examined the various plausible arrangements, with both A- and B-type helices having been described [17–19]. NMR studies, especially on intramolecular triplexes, have elucidated a number of features, such as sugar puckers and qualitative indications that the global conformation of a parallel triple helix has features of both A- and B-form duplex DNA [3]. The recent crystal structure [20] of a very stable DNA-PNA triplex, with a peptide-like backbone for the two pyrimidine strands instead of deoxyribose-phosphodiester ones, has C^+-G•C triplets in a novel 'P-form' helical structure. Several recent oligonucleotide crystal structures with non-base-paired ends have revealed intermolecular triplex base arrangements [21,22] of the G-G•C type, as found in anti-parallel triplexes, but the classical parallel C^+-G•C arrangement has not hitherto been observed to date in any DNA structure that is not constrained by the particular characteristics of a peptide nucleic acid.

We have recently determined the crystal structure of a decanucleotide complexed with the minor groove drug netropsin [23], which also shows unpaired bases at both ends of the duplex. These bases are involved in classic parallel C^+-G•C triplets, and their precise arrangement has been used as the starting-point for the construction of a molecular model of the [C^+-G•C]$_n$ triple helix.

2. Materials and methods

Crystals of a d(CGCAATTGCG)$_2$-netropsin complex were grown by vapour diffusion. X-ray diffraction data were collected to a resolution of 2.4 Å. The final R factor for all data with $F > 2\sigma(F)$ is 20.1% in the resolution range 8–2.4 Å, with 85 water molecules. Full details of the data collection, structure determination, refinement and netropsin binding, are given elsewhere [23]. Final atomic coordinates and structure factors have been deposited with the Nucleic Acid Database as entry GDJ046.

The three nucleotides forming the C^+-G•C triplet were used as the basis for molecular modelling of a triple helix. This was built in the following manner: values for helical rise and twist were taken from the recent fibre diffraction study of a general sequence DNA triplex [24]. A triplet exactly duplicating the crystallographically-observed one was placed on a helical z axis, perpendicular to the G•C base pair plane, with a rise of 3.2 Å and a helical twist of 30°. The backbone geometries of the three dinucleotides were optimised by using a molecular mechanics/dynamics/mechanics protocol with the Macromodel v4.5 package [25]. This employed a continuum water solvent model, 500 steps of steepest descent followed by 1000 steps of conjugate gradient minimisation, then 5ps equilibration and a 40ps production phase with time-averaging of 100 structures. The time-averaged structure was then minimised by the same procedure. The base pairs were kept rigid throughout the procedure. The resulting nucleotides were used as the starting point for the generation of a full helix of (C^+-G•C)$_n$, again using the published [24] helical parameters. The (C^+-G•C)$_n$ model derived from the fibre diffraction coordinates was generated by the literature method [24].

---

*Corresponding author. Fax: (44) (181) 643 1675.
E-mail steve@iris5.icr.ac.uk

0014-5793/97/$17.00 © 1997 Federation of European Biochemical Societies. All rights reserved.
PII S0014-5793(97)01130-7

3. Results

The crystal structure of the netropsin-d(CGCAATTGCG)$_2$ complex shows that three of the four 3′ and 5′ terminal bases are not Watson-Crick hydrogen bonded within the duplex formed by the central eight base pairs. Instead they are oriented away from the octamer helix, so that they are in positions to interact with the adjacent stacked helices in the crystal lattice (Fig. 2). (The fourth, C11, is not observed in the structure, presumably since it is highly mobile or disordered.) Cytosine C1 interacts with the G8•C13 base pair, which is itself one base pair away from the junction between two octamer helices. This cytosine is in the duplex major groove, and it is in an appropriate position for two Hoogsteen hydrogen bonds to be formed to guanine G8, in an identical manner to those postulated to occur for the C$^+$-G•C triplet of bases in parallel DNA triplexes (Fig. 1b). The N3C11...N7G8 and N4C11...G8 distances are both 2.7 Å, confirming that two hydrogen bonds have been formed (Fig. 3). This can only be rationalised by postulating that the N3 atom of the cytosine is protonated. Alternative tautomers for the guanine with protonation of the N7 position are highly unlikely, especially since the guanine is also involved in a standard G•C base pair. The Watson-Crick hydrogen bonds between G8 and C13 have lengths in a normal range of values, of between 2.7 and 2.9 Å. The backbones attached to C1 and G8 are in a parallel orientation to each other, as in a true parallel DNA triple helix. The Watson-Crick bases in the triplet are closely coplanar to each other, with a propeller twist of −6° and a buckle of 0° (values calculated with the CURVES program [26]). By contrast the third-strand cytosine is significantly distorted out of this plane, with a tilt of 22° with respect to the guanine. All three deoxyribose sugars in the triplet have pseudorotation phase angles close to 180° and C2′-*endo* puckers. This triplet of bases and the associated sugar-phosphate backbones thus provide an appropriate experimental basis for the building of molecular models for the C$^+$-G•C triple helix.

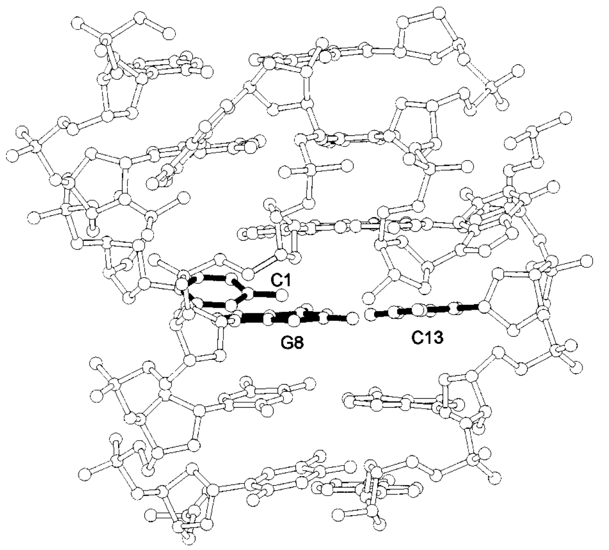

Fig. 2. A view of the junction between two d(CGCAATTGCG)$_2$ duplexes in the crystal structure of the netropsin complex [23]. The three bases forming the C$^+$-G•C triplet are shown with their bonds shaded.

The molecular model constructed here for a [C$^+$-G•C]$_n$ triple helix, based on the observed C$^+$-G•C triplet and energy-refined backbone conformations, is shown in Fig. 4, where it is compared with that from fibre diffraction studies. Both models are of overall B-family appearance in terms of the bases of the Watson-Crick pair being perpendicular to the helix axis and all the sugar groups on all three strands having puckers in the C2′-*endo* range (Table 1). Since the modelling has preserved the tilt of the cytosine third base in each triplet, the third strand of the triplex has some A-like character with the cytosines inclined to the helix axis (Fig. 4b). Both crystallographically derived and fibre diffraction models have the major groove significantly widened compared to canonical B-DNA, in order to fully accommodate the third strand. There are small but significant differences between the two models in terms of all three groove widths, as calculated by CURVES (Fig. 5). The minor groove in the fibre model [24] is wider, 7.2 Å compared to 6.2 Å. The two grooves formed

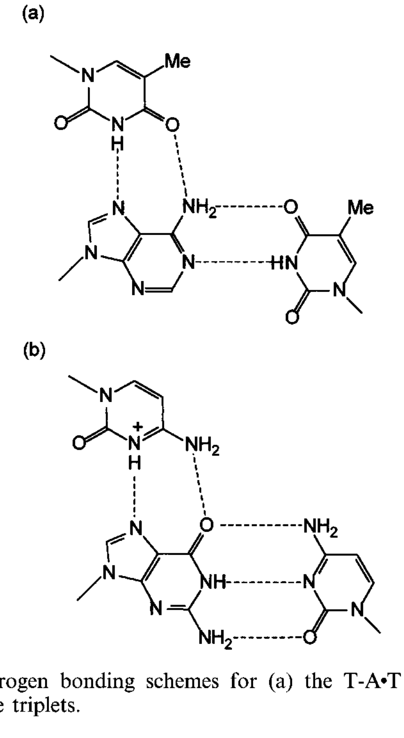

Fig. 1. Hydrogen bonding schemes for (a) the T-A•T and (b) the C$^+$-G•C base triplets.

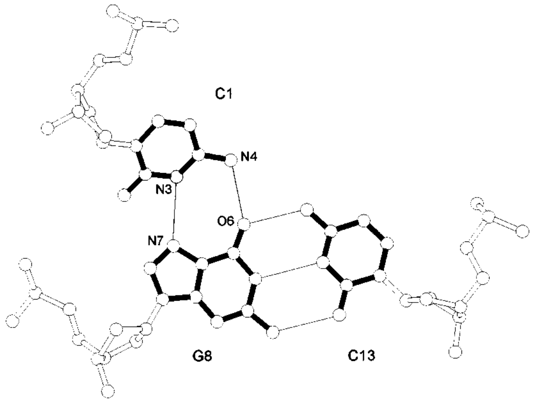

Fig. 3. The structure of the C$^+$-G•C triplet as observed in the crystal structure, viewed onto the plane of the bases. Hydrogen bonds are shown as continuous thin lines.

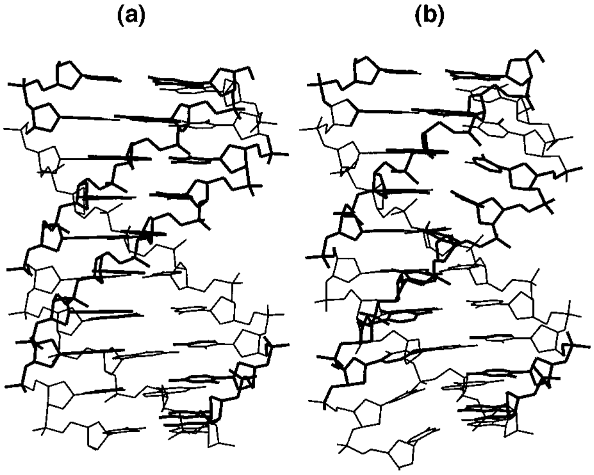

Fig. 4. Views of (a) fibre diffraction [24] and (b) the crystallographically derived models for the [C^+-G•C]$_n$ triple helix. The helix axis is vertical. The pyrimidine third strand is highlighted in bold.

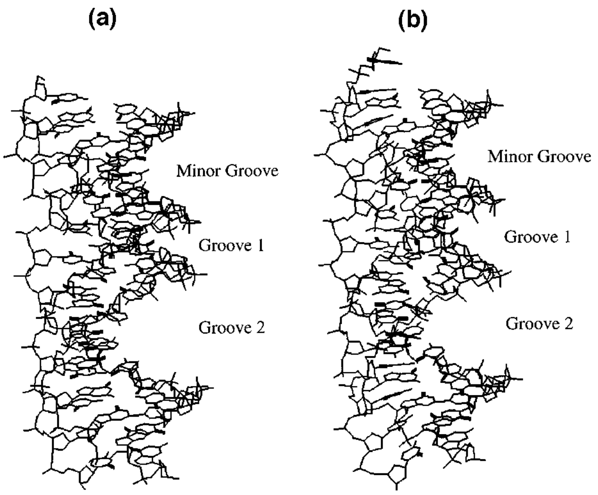

Fig. 5. As in Fig. 4a,b, but with the helix axis tilted in order to view along the grooves. Grooves 1 and 2 together constitute the major groove of the [G•C]$_n$ duplex component of the triple helix.

from the major groove of the parent [G•C]$_n$ duplex, are 1.0 and 7.9 Å wide in the fibre model compared to 3.0 and 6.5 Å in the crystallographically-derived one. These differences are a direct consequence, in the latter, of the tilt of the C^+ bases in the third strand out of the plane of the G•C base pairs. Adjacent N3 cytosine atoms are some 4.9Å apart, comparing favourably with the separation of 4.2 Å found in an NMR triplex structure [27].

The backbone torsion angles in the two models (Table 1) show considerable variations between the three strands. It is notable that only one strand of the crystallographic model (strand 1; C) has all backbone angles close to those in the fibre model. This reflects the conformational plurality of triple helices, as well as the slight non-equivalence of the third C^+ strand due to its tilt. The backbone angles of this third strand are closer to those for A-DNA than B, with angles β, δ and ε having typical A-like values.

4. Discussion

The observation of the C^+-G•C triplet in a drug-oligonucleotide crystal structure, is unexpected since the crystals were obtained at pH 7.0. Protonated cytosines have been implicated in the C•C base pairs found in the crystal structures of intercalated four-stranded cytosine telomere structures [28,29], as well as in polymers of cytidylic acid. In all these instances, the existence of hemi-protonated cytosines at pH 7 has been ascribed to the effects of base stacking, which presumably is able to raise the pK_a of N3 by several units. By contrast, C1 in the present structure is not well stacked with other bases in the crystal lattice. We have previously reported [30] on the crystal structure of the d(CGCAATTGCG)$_2$ sequence in the absence of a bound drug. This structure also has terminal bases unravelled from the central duplex, but with the non-duplex cytosines hydrogen-bonded via the major groove to both bases in a Watson-Crick G•C base pair, in a manner not requiring N3 protonation, and thus quite distinct from that observed here. This suggests that the drug binding, which is the major difference between the two structures, may contribute to the change in the N3 pK_a value, even though the netropsin molecule is no closer than ca. 12 Å to the C1 base. It is not possible to provide a definitive explanation for the observation of N3 protonation; and the experimental data [31–35] on minor groove drugs binding to triplexes is not entirely relevant, since such studies have focussed on the binding of these drugs to [T-A•T]$_n$-type triple helices themselves. Our results do suggest that a minor groove drug if targeted to an adjacent duplex region of appropriate sequence, could improve the stabilisation of C^+-G•C triplets at physiological pH.

The model derived for the triplex is similar to several other B-type py-pu•py triple helices proposed on the basis of molecular modelling and fibre diffraction studies [17,19,24,36–38]. A number of features such as C2′-*endo* sugar puckers and the perpendicular orientation of the Watson-Crick bases to the helix axis are in accord with a B-type helix. The third cytosine strand has some features in common with the

Table 1
Backbone torsion angles and sugar puckers for the nucleotide repeating units in the fibre model, where all three strands are conformationally equivalent, and the present model, where the three strands are non-equivalent

	Fibre [24]	Strand 1; C	Strand 2; G	Strand 3; C^+
α (P–O5′)	−37	−62	−66	−65
β (O5′–C5′)	171	−168	−134	153
γ (C5′–C4′)	35	47	27	92
δ (C4′–C3′)	129	114	140	59
ε (C3′–O3′)	−177	178	153	−129
ζ (O3′–P)	−109	−89	−83	−75
P, in °	139	171	175	182
Pucker	C1′-*exo*	C2′-*endo*	C2′-*endo*	C3′-*exo*

P represents the pseudorotation phase angle for the deoxyribose sugar rings.

A-DNA family, with in particular a significant tilt of bases from the helix axis (Fig. 4) and an A-like backbone conformation. This type of behaviour has been observed in molecular dynamics simulations [19] of triplex DNA. The fact that these aspects of our model are directly derived from single-crystal structural data, provides further credence to the view that DNA triple helices are not as structurally homogeneous as duplexes. Thus they cannot be simply described in classic A- or B-type terms, and they can even show features of both, as illustrated in the present model.

Acknowledgements: This work was supported by Programme Grant SP1384 from the Cancer Research Campaign.

References

[1] Felsenfeld, G., Davies, D.R. and Rich, A. (1957) J. Am. Chem. Soc. 79, 2023–2024.
[2] Soyfer, V.N. and Potaman, V.N. (1996) Triple-Helical Nucleic Acids, Springer-Verlag, New York.
[3] Radhakrishnan, I. and Patel, D.J. (1994) Biochemistry 33, 11405–11416.
[4] Thuong, N.T. and Hélène, C. (1993) Angew. Chem. Int. Ed. Eng. 32, 666–690.
[5] Moser, H. and Dervan, P.B. (1987) Science 238, 645–660.
[6] Duval-Valentin, G., Thuong, N.T. and Hélène, C. (1992) Proc. Natl. Acad. Sci. USA 89, 504–508.
[7] Porumb, H., Gousset, H., Letellier, R., Salle, V., Briane, D., Vassy, J., Amor-Gueret, M., Israël, L. and Taillandier, E. (1996) Cancer Res. 56, 515–522.
[8] Olivas, W.M. and Maher, L.J. (1996) Nucleic Acids Res. 24, 1758–1764.
[9] Xiang, G.B., Soussou, W. and McLaughlin, L.W. (1994) J. Am. Chem. Soc. 116, 11155–11156.
[10] Huang, C.Y., Bi, G. and Miller, P.S. (1996) Nucleic Acids Res. 24, 2606–2613.
[11] Bates, P.J., Laughton, C.A., Jenkins, T.C., Capaldi, D.C., Roselt, P.D., Reese, C.B. and Neidle, S. (1996) Nucleic Acids Res. 24, 4176–4184.
[12] Pilch, D.S., Waring, M.J., Sun, J.-S., Rougée, M., Nguyen, C.-H., Bisagni, E., Garestier, T. and Hélène, C. (1993) J. Mol. Biol. 232, 926–946.
[13] Fox, K.R., Polucci, P., Jenkins, T.C. and Neidle, S. (1995) Proc. Natl. Acad. Sci. USA 92, 7887–7891.
[14] Lee, J.S., Latimer, L.J.P. and Hampel, K.J. (1993) Biochemistry 32, 5591–5597.
[15] Cassidy, S.A., Strekowski, L., Wilson, W.D. and Fox, K.R. (1994) Biochemistry 33, 15338–15347.
[16] Arnott, S. and Selsing, E. (1974) J. Mol. Biol. 88, 509–521.
[17] Howard, F.B., Miles, H.T., Liu, K., Frazier, J., Raghunathan, G. and Sasisekharan, V. (1992) Biochemistry 31, 10671–10677.
[18] Laughton, C.A. and Neidle, S. (1992) J. Mol. Biol. 223, 519–529.
[19] Weerasinghe, S., Smith, P.E., Cheng, Y.-K. and Pettitt, B.M. (1995) J. Am. Chem. Soc. 117, 2147–2159.
[20] Betts, L., Josey, J.A., Veal, J.M. and Jordan, S.R. (1995) Science 270, 1838–1841.
[21] van Meervelt, L., Vlieghe, D., Dautant, A., Gallois, B., Précigoux, G. and Kennard, O. (1995) Nature 374, 742–744.
[22] Vlieghe, D., van Meervelt, L., Dautant, A., Gallois, B., Précigoux, G. and Kennard, O. (1996) Science 273, 1702–1705.
[23] Nunn, C.M., Garman, E. and Neidle, S. (1997) Biochemistry 36, 4792–4799.
[24] Liu, K., Sasisekharan, V., Miles, H.T. and Raghunathan, G. (1996) Biopolymers 39, 573–589.
[25] Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T. and Still, W.C. (1990) J. Comp. Chem. 11, 440–467.
[26] Lavery, R. and Sklenar, H. (1988) J. Biomol. Struct. Dyn. 6, 63–91.
[27] Radhakrishnan, I. and Patel, D.J. (1992) J. Mol. Biol. 241, 600–619.
[28] Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. and Rich, A. (1994) Proc. Natl. Acad. Sci. USA 91, 11636–11640.
[29] Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. and Rich, A. (1995) Proc. Natl. Acad. Sci. USA 92, 3874–3878.
[30] Spink, N., Nunn, C.M., Vojtechovsky, J., Berman, H.M. and Neidle, S. (1995) Proc. Natl. Acad. Sci. USA 92, 10767–10771.
[31] Pilch, D.S. and Breslauer, K.J. (1994) Proc. Natl. Acad. Sci. USA 91, 9332–9336.
[32] Vigneswaran, N., Mayfield, C.A., Rodu, B., James, R., Kim, H.-G. and Miller, D.M. (1996) Biochemistry 35, 1106–1114.
[33] Durand, M. and Maurizot, J.C. (1996) Biochemistry 35, 9133–9139.
[34] Pilch, D., Kirolos, M.A. and Breslauer, K.J. (1996) Biochemistry 35, 16107–16124.
[35] Kim, H.-K., Kim, J.-M., Kim, S.K., Rodger, A. and Nordén, B. (1996) Biochemistry 35, 1187–1194.
[36] Raghunathan, G., Miles, H.T. and Sasisekharan, V. (1993) Biochemistry 32, 455–462.
[37] Raghunathan, G., Miles, H.T. and Sasisekharan, V. (1995) Biopolymers 36, 333–343.
[38] Cheng, Y.-K. and Pettitt, B.M. (1992) J. Am. Chem. Soc. 114, 4465–4474.