**Arka Chakraborty,
Pradeep Kumar Mandal and
Namasivayam Gautham***

CAS in Crystallography and Biophysics,
University of Madras, Guindy Campus,
Chennai, Tamil Nadu 600 025, India

Correspondence e-mail:
n_gautham@hotmail.com

Received 27 May 2012
Accepted 7 August 2012

PDB Reference: d(CCGGGACCGG)₄, 4ez2

Structure of d(CCGGGACCGG)₄ as a four-way junction at 1.6 Å resolution: new insights into solvent interactions

The crystal structure of the decamer sequence d(CCGGGACCGG)₄ has previously been reported at 2.16 Å resolution as a four-way junction. Here, the structure of this sequence is reported at the significantly higher resolution of 1.6 Å, which is the highest resolution reported for a four-way junction. This allowed the unambiguous identification of an extensive hydration network with distinct patterns and solvent-mediated interactions that shed new light on the role of water in the formation and stabilization of junction structures.

1. Introduction

A large number of four-way junction structures with decamer sequences have been deposited in the Nucleic Acid Data Bank (<http://ndbserver.rutgers.edu/>; Berman *et al.*, 1992). The first such structure was reported by Ortiz-Lombardía *et al.* (1999), in which the sequence d(CCGGGACCGG)₄ with a GA mismatch formed a so-called 'stacked-X' four-way junction. Soon afterwards, Eichman *et al.* (2000) reported the structure of d(CCGGTACCGG)₄ with the same conformation. Junction structures are observed to be highly hydrated and while comparison of the hydration patterns in the above two sequences (Eichman *et al.*, 2002) provided some insights into the role of water-mediated interactions in junction structures, the conclusions were constrained by the low resolution of the structures, particularly for the GA-mismatch sequence. Here, we report the structure of d(CCGGGACCGG)₄, again as a four-way junction, but at the significantly higher resolution of 1.6 Å. This is the highest resolution reported to date for a junction structure. Thus, we are able to describe a number of solvent-mediated interactions and general hydration patterns that were not detected in the previous report of the structure (Ortiz-Lombardía *et al.*, 1999).

2. Materials and methods

2.1. Crystallization, X-ray diffraction data collection and data processing

PAGE-purified oligonucleotides and other chemicals were purchased from Sigma–Aldrich Chemicals, Bangalore, India. Crystals were grown by the hanging-drop vapour-diffusion technique at 293 K from a drop consisting of 1 µl 1 mM DNA, 1 µl 50 mM sodium cacodylate trihydrate buffer pH 7.0, 1 µl 10 mM MnCl₂, 1 µl 1 mM spermine. [The first report of this structure (Ortiz-Lombardía *et al.*, 1999) used magnesium instead of manganese as the cation. The initial objective of our study was to investigate the nature of the interaction of the transition metals with the junction structure. However, as reported below, we could not identify the Mn ion in the structure.] The hanging drop was equilibrated against a reservoir consisting of 50% 2-methyl-2,4-pentanediol (MPD). Thin diamond-shaped plate crystals of dimensions 0.1 × 0.05 × 0.001 mm were obtained after 5 d. For data collection at 100 K, the crystals were flash-cooled in liquid nitrogen. The mother liquor was sufficient for cryoprotection. X-ray diffraction data (Table 1) were collected on the BM14 synchrotron beamline at the ESRF, France at a wavelength of 1.00 Å. The diffraction images were recorded using a MAR CCD detector (MAR Research). Data processing was performed using *iMOSFLM* (Battye

et al., 2011) from the CCP4 suite (Winn *et al.*, 2011). The crystal belonged to the monoclinic space group *C2*, with unit-cell parameters $a = 64.01$, $b = 24.35$, $c = 38.18$ Å, $\beta = 112.53^\circ$.

2.2. Structure determination and refinement

The structure was solved using molecular replacement. As a lower resolution crystal structure of the oligomer was already available (PDB entry 467d; Ortiz-Lombardía *et al.*, 1999), this was used as a search model. The model consisted of a noncrossing strand and a crossing strand (the crystallographic twofold symmetry axis generates the full junction by a twofold rotational symmetry operation). Molecular replacement was performed using the program *AMoRe* (Navaza, 1994) from the CCP4 suite (Winn *et al.*, 2011) and a solution with a correlation coefficient of 88.3 and an *R* factor of 25.4% was obtained. The structure was initially refined using *REFMAC5* (Murshudov *et al.*, 2011) with maximum-likelihood targets and the *REFMAC5* dictionary (Vagin *et al.*, 2004) and was finally refined using *phenix.refine* from the *PHENIX* suite (Adams *et al.*, 2010). Water molecules were added only if such additions were warranted by a $2F_o - F_c$ difference map. After initial rounds of refinement, a positive peak was observed at 9.25σ in the $F_o - F_c$ difference map on the crystallographic twofold symmetry axis (the so-called 'sodium cavity'; Hays, Watson *et al.*, 2003). The peak was initially modelled as a manganese (Mn^{2+}) ion with an occupancy of 0.50 and was subjected to coordinate and real-space refinement in *phenix.refine*. A blob of negative density was observed around the manganese ion in the $F_o - F_c$ map (Fig. 1*a*). Modelling the peak as a water molecule with subsequent coordinate and real-space refinement showed a clear extra positive density at 2.0σ in the $F_o - F_c$ map (Fig. 1*b*). Therefore, in a final trial the peak was modelled as a sodium (Na^+) ion with 0.50 occupancy and was subjected to the same refinement protocol. No extra density was observed at the position in the resulting $F_o - F_c$ difference map even at 1.0σ (Fig. 1*c*). The interatomic distances from sodium to the phosphate O atoms of A_6 of the two crossing strands were cross-verified using the junction structure reported by Hays, Vargason *et al.* (2003) (PDB entry 1p54). Additionally, the coordination number of sodium (four) and its coordination distance to oxygen were verified using a number of protein structures (PDB entries 1svy, 2occ, 1cm5, 1c24 and 1dpo; Puius *et al.*, 2000; Yoshikawa *et al.*, 1998; Becker *et al.*, 1999; Lowther *et al.*, 1999; Earnest *et al.*, 1991) and the sodium–oxygen phosphate distances were found to be in the range of values for these structures (maximum sodium–oxygen

Table 1

Summary of data-processing and refinement statistics.

Values in parentheses are for the outermost shell.

Diffraction data	
Resolution (Å)	22.52–1.60 (1.69–1.60)
Space group	<i>C2</i>
Unit-cell parameters (Å, °)	$a = 64.01$, $b = 24.35$, $c = 38.18$, $\beta = 112.53$
R_{merge} (%)	5.4 (26.8)
Mean $I/\sigma(I)$	11.7 (3.6)
Completeness (%)	99.3 (99.0)
Multiplicity	3.8 (3.8)
No. of observations	27940 (4002)
No. of unique reflections	7323 (1047)
Refinement	
No. of DNA atoms	408
No. of solvent atoms	91
<i>R</i> factor (%)	21.6
R_{free} (%)	24.9
R.m.s.d. bond lengths (Å)	0.005
R.m.s.d. bond angles (°)	1.26
Average <i>B</i> factor (Å ²)	22.6
PDB code	4ez2

coordination distance of 3.29 Å). The sodium–oxygen distances in the present structure were also lower than the maximum acceptable value of 3.25 Å (<http://tanna.bch.ed.ac.uk/mphos/mphos3w.htm>; Harding, 2004). In view of these indications, the peak was finally fixed as a sodium ion. However, owing to the fact that the distinction between sodium and a well ordered water is ambiguous in general, the presence of the sodium ion is indicative and cannot be conclusively verified. TLS refinement was carried out in *phenix.refine* with TLS parameters generated in *phenix.refine* using the two strands of the asymmetric unit as two separate TLS groups. The final refinement statistics are given in Table 1.

Graphical analyses of the model and the electron-density maps were carried out using *Coot* (Emsley & Cowtan, 2004). *PyMOL* was used to prepare the figures (DeLano, 2002). The coordinates and structure factors have been deposited in the PDB (Berman *et al.*, 2000) as entry 4ez2.

3. Results and discussion

3.1. Overall structure

Fig. 2 shows the topology and numbering scheme. As reported by Ortiz-Lombardía *et al.* (1999), the sequence crystallizes as a tightly

Figure 1

Results of fixing manganese, water and sodium in the 'sodium cavity'. The $2F_o - F_c$ maps are contoured at 1.0σ . (*a*) Fixing manganese (Mn^{2+}) ion (magenta) shows negative density in the $F_o - F_c$ map (grey; contoured at 3.0σ). The $2F_o - F_c$ map (contoured at 1.0σ) is shown in blue. (*b*) Fixing a water molecule (green) shows extra positive density (red) in the $F_o - F_c$ map at 2.0σ . The $2F_o - F_c$ map is shown in grey. (*c*) Fixing a sodium (blue) results in no extra density. The $F_o - F_c$ map at the position (contoured at 2.0σ) is featureless. The $2F_o - F_c$ map is shown in grey.

packed right-handed antiparallel stacked-X four-way junction. The full junction structure consists of four short helices or arms formed by base-pairing interactions between crossing and noncrossing strands. Two long arms, each six base pairs in length, are formed *via* hydrogen-bonded base pairs between bases C₁ to A₆ of a crossing strand and bases G₁₀ to G₅ of a noncrossing strand. Two short arms of four base pairs each in length are formed by the pairing of bases C₇ to G₁₀ of the crossing strand with bases G₄ to C₁ of an alternative noncrossing strand. The DNA sequence contains a G–A mismatch at the exchange point of the strands. However, this does not lead to any major distortions, except for the presence of a three-centred hydrogen-bonded interaction with a high propeller twist (as also reported by Ortiz-Lombardía *et al.*, 1999). The arms stack to form two semi-continuous pseudo-helices with breaks in the phosphate backbones at the exchange points. The root-mean-square deviation (r.m.s.d.) in atomic positions obtained after least-squares superposition of the present structure with the previously reported structure of the same sequence (Ortiz-Lombardía *et al.*, 1999) is 0.25 Å.

Figure 2

Sequence topology of the d(CCGGGACCGG)₄ four-way junction. The strands are numbered from 1 to 10 in the 5'-to-3' direction and are colour-coded as follows: strand A, red; strand B, yellow; strand C, purple; strand D, pink. This colour convention is followed in all subsequent figures. The black boxes show the G–A mismatch.

All of the interactions and structural features that were previously reported by Ortiz-Lombardía *et al.* (1999) were also observed in the present structure. In addition, we observed a number of new solvent-mediated interactions that were not reported previously which may have implications in the formation and stabilization of the junction structure.

3.2. Core interactions

The crystal structures of four-way junctions point to a common trinucleotide A₆–C₇–C₈ sequence as a 'core' sequence that is required for formation of the junction. The following interactions are associated with this core sequence (Eichman *et al.*, 2000; Thorpe *et al.*, 2002, 2003): (i) a direct hydrogen bond between the N4 nitrogen of the C₈ base and the C₇ nucleotide phosphate oxygen (C₈–C₇ interaction), (ii) a solvent-mediated interaction between the complementary G₃ nucleotide and the A₆ nucleotide phosphate oxygen (G₃–A₆ interaction) and (iii) a direct hydrogen bond from the N4 nitrogen of the C₇ base to the A₆ phosphate oxygen (C₇–A₆ interaction). In the previously reported structure of the present sequence (Ortiz-Lombardía *et al.*, 1999) the first two core interactions were observed. The third was absent, as also reported for many other structures (Hays, Watson *et al.*, 2003), implying that this interaction is not crucial for junction formation. In the present structure all three interactions were observed (Fig. 3). However, the third interaction was not a direct interaction between C₇ and A₆ but was mediated by a water molecule. It is possible that this interaction helps to stabilize the junction despite the GA mismatch.

A sodium ion was fixed at the centre of the junction on the crystallographic twofold axis. While some previous reports (Eichman *et al.*, 2000) have suggested a role for the ion in mediating a close approach of the phosphate O atoms on the crossing strands, the absence of a metal ion at this position in a number of other junction structures (Hays, Watson *et al.*, 2003) suggests that this ion-mediated interaction is not crucial for the overall geometry of the junction.

3.3. Hydration and solvent-mediated interactions

Comparison of the helical parameters of the two pseudo-continuous helices of the present structure with those of the B-DNA fibre model (Table 2) revealed no significant differences. The B-DNA

Figure 3

Stereoview showing the core interactions of the present junction structure. Water molecules are shown as green spheres. The inset shows the junction structure, with the location of the core interactions demarcated by a box.

Table 2

Average helical parameters calculated using *w3DNA* for the quasi-continuous helices in the four-way junction and comparison with the fibre B-DNA model (Chandrasekaran & Arnott, 1989).

Minor-groove and major-groove widths were measured as the 'refined' phosphate-phosphate distances based on the method proposed by El Hassan & Calladine (1998) and incorporated in the *w3DNA* web server (Lu & Olson, 2003; Zheng *et al.*, 2009).

	d(CCGGGACCGG) ₂	B-DNA fibre model
Helical rise (Å)	3.3	3.3
Helical twist (°)	36.7	35.6
Roll (°)	2.3	1.8
Tilt (°)	−0.5	0.0
Propeller twist (°)	−13.9	−14.6
X-displacement (Å)	1.8	0.6
Slide (Å)	1.5	0.5
Major-groove width (Å)	11.0	11.4
Minor-groove width (Å)	7.3	5.9

structure is characterized by a hydrogen-bonded network of water molecules in the minor groove of the DNA which consists of specific interactions between the solvent and the DNA and is referred to as the 'spine of hydration' (Drew & Dickerson, 1981). It has been reported (Eichman *et al.*, 2002) that in the two junction structures d(CCGGTACCGG)₄ and d(CCGGGACCGG)₄ the spine of hydration does not extend over the full length of the pseudo-continuous helices and that the water molecules are clustered at the ends of the arms while the central residues are not hydrated. In the present structure of d(CCGGGACCGG)₄ we observe a continuous spine of hydration in the minor groove, similar to that in B-DNA (Fig. 4). Specifically, we find the central residues, particularly base pairs 4, 5 and 6, to be adequately hydrated.

Figure 4

Stereoview showing the spine of hydration in the present structure. The water molecules are shown as green spheres. The $2F_o - F_c$ map (cyan, contoured at 1.0σ) is also shown for the water molecules. The sodium ion and its density are shown in blue.

Figure 5

Stereoview of the junction structure showing the locations of the six solvent-mediated interactions. The numbering follows the order of description in the text. Symmetry-related interactions are not shown.

As reported by Eichman *et al.* (2002), a number of first-shell solvent interactions were found to be conserved between d(CCGGTACCGG)₄ and d(CCGGGACCGG)₄. In the present structure we observed 15 more previously undetected water molecules which also appear in the structure of d(CCGGTACCGG)₄. Of these, nine are located in the minor grooves of the pseudo-helices. In conformity with the observations made by Eichman and coworkers, the majority of the conserved water molecules occupy the cavity formed at the interface of the two pseudo-helices, where they may play important roles in the stability of the junction structure *via* interactions with both the crossing and the noncrossing strands. In particular, the following interactions (Fig. 5) appear to play such roles.

Interaction I. The interaction between the O1P oxygen of C₇ of strand A and the O1P oxygen of C₈ of strand C mediated by two water molecules. The oxygen interacting with O1P of C₈ of strand C is involved in a second hydrogen bond to the N3 nitrogen of A₆ of strand A (Fig. 6a). This interaction may play a role in allowing the pseudo-helix to retain the B-DNA conformation across the crossover region.

Interaction II. The hydrogen-bond chain formed by two water molecules linking the O1P oxygen of C₈ of strand A to the N3 nitrogen of A₆ of strand C. The water molecule interacting with O1P of C₈ of strand A forms another hydrogen bond to the O3' oxygen of the deoxy-ribose sugar of C₇ of strand A (Fig. 6b). This set of interactions, like the previous one, is also located at the point where one crossing strand leaves the pseudo-helix and another enters. Such interactions could have implications in the transition of junction structures from the X-stacked conformation observed for junctions as DNA-only constructs to the open form observed for junction

Figure 6
The six solvent-mediated interactions observed in the present structure. (a) Interaction I. (b) Interaction II. (c) Interaction III. (d) Interaction IV. (e) Interaction V. (f) Interaction VI. Water molecules are shown as green spheres.

sequences in complex with proteins such as RuvA (Roe *et al.*, 1998). As proposed by Ortiz-Lombardía *et al.* (1999), a plausible mechanism for such a transition would be a rotation of the stacked arms in opposite directions accompanied by unwinding of the continuous helix at the stacking point. This may require displacement of the water molecules by the binding of RuvA. In the absence of such a protein, the solvent interactions probably impose the stacked-X structure on the molecule.

Interaction III. The interaction between the O2P oxygen of C₈ of one crossing strand A and the O2P oxygen of C₈ of the other crossing strand C mediated by hydrogen-bonding interaction from two water molecules (Fig. 6c). The water molecule interacting with O2P of C₈ of strand A is also involved in a second hydrogen-bonding interaction with O1P of C₇ of strand C.

Interaction IV. The hydrogen-bonding chain formed by two water molecules linking the O4' oxygen of G₁₀ of strand A and the N2 nitrogen of G₃ of strand D. The water molecule interacting with the O4' oxygen of G₁₀ of strand A also forms a hydrogen bond to the N3 nitrogen of G₉ of strand A (Fig. 6d). This interaction is located in the minor groove of the pseudo-helix formed by strands A, C and D. It is possible that during junction formation, subsequent to the crossing over, such interactions serve to hold the terminal part of the incoming strand in close proximity to the noncrossing strand, aiding the formation of base pairing between the two, which then stabilizes the B-DNA conformation of the pseudo-helices.

Interaction V. The interaction involving the O2P oxygen of C₇ of strand A and the N4 nitrogen of C₂ of strand D which are linked *via* a chain of three hydrogen bonds by two water molecules (Fig. 6e). This set of interactions may play a role similar to the second core interaction in holding the phosphate group of C₇ in place and providing a hinge point for strand crossover to take place.

Interaction VI. The interaction involving the O1P oxygen of C₈ of strand C and the N3 nitrogen of A₆ of strand B which are linked

through a chain of three hydrogen bonds formed between them by two water molecules (Fig. 6f). The water molecules lie in the minor groove of the pseudo-helix formed by strands A, B and C and the interaction appears to hold the phosphate of C₈ in proximity to the noncrossing B strand, thus contributing to the stability of the junction structure.

To sum up, there is a network of solvent-mediated interactions that appears to contribute to the strand-crossover event and its subsequent stabilization of the junction structure. A similar trend of hydration is observed in three other high-resolution junction structures [PDB entries 1p4y (Hays, Watson *et al.*, 2003), 1p54 (Hays, Vargason *et al.*, 2003) and 1zf3 (Hays *et al.*, 2005)]. A majority of the specific interactions discussed above involve atoms belonging to residues from the ACC core sequence, which has been implicated in dictating junction geometry (Hays, Watson *et al.*, 2003). Compromise in the interactions at the core region (which include the three core interactions as well as other solvent-mediated interactions such as those reported here) might trigger the transition of a junction from the stacked-X to the open-X form. Finally, the observation of a large number of conserved water molecules in the crossover region and in the cavities at the interface of the pseudo-helices compared with d(CCGGTACCGG)₄ suggests that, irrespective of the difference in sequence outside the core region, hydration of the core region is a general requirement for junction formation and stability.

The authors would like to thank the Department of Biotechnology (DBT) for beamline time at BM14, ESRF, Grenoble, France, where the X-ray diffraction data for the present structure were collected. PKM thanks CSIR for a Senior Research Fellowship.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D66, 213–221.

- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S. & Volker Wagner, A. F. (1999). *Nature Struct. Biol.* **6**, 969–975.
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). *Biophys. J.* **63**, 751–759.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Chandrasekaran, R. & Arnott, S. (1989). *Landolt–Börnstein Numerical Data and Functional Relationships in Science and Technology*, Vol. VII/Ib, edited by W. Saenger, pp. 31–170. Berlin: Springer-Verlag.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Drew, H. R. & Dickerson, R. E. (1981). *J. Mol. Biol.* **151**, 535–556.
- Earnest, T., Fauman, E., Craik, C. S. & Stroud, R. (1991). *Proteins*, **10**, 171–187.
- Eichman, B. F., Ortiz-Lombardía, M., Aymamí, J., Coll, M. & Ho, P. S. (2002). *J. Mol. Biol.* **320**, 1037–1051.
- Eichman, B. F., Vargason, J. M., Mooers, B. H. & Ho, P. S. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3971–3976.
- El Hassan, M. A. & Calladine, C. R. (1998). *J. Mol. Biol.* **282**, 331–343.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Harding, M. M. (2004). *Acta Cryst.* **D60**, 849–859.
- Hays, F. A., Teegarden, A., Jones, Z. J., Harms, M., Raup, D., Watson, J., Cavaliere, E. & Ho, P. S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 7157–7162.
- Hays, F. A., Vargason, J. M. & Ho, P. S. (2003). *Biochemistry*, **42**, 9586–9597.
- Hays, F. A., Watson, J. & Ho, P. S. (2003). *J. Biol. Chem.* **278**, 49663–49666.
- Lowther, W. T., Zhang, Y., Sampson, P. B., Honek, J. F. & Matthews, B. W. (1999). *Biochemistry*, **38**, 14810–14819.
- Lu, X.-J. & Olson, W. K. (2003). *Nucleic Acids Res.* **31**, 5108–5121.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F. & Coll, M. (1999). *Nature Struct. Biol.* **6**, 913–917.
- Puius, Y. A., Fedorov, E. V., Eichinger, L., Schleicher, M. & Almo, S. C. (2000). *Biochemistry*, **39**, 5322–5331.
- Roe, S. M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I. R. & Pearl, L. H. (1998). *Mol. Cell*, **2**, 361–372.
- Thorpe, J. H., Gale, B. C., Teixeira, S. C. M. & Cardin, C. J. (2003). *J. Mol. Biol.* **327**, 97–109.
- Thorpe, J. H., Teixeira, S. C. M., Gale, B. C. & Cardin, C. J. (2002). *Acta Cryst.* **D58**, 567–569.
- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* **D60**, 2184–2195.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yamashita, E., Inoue, N., Yao, M., Fei, M. J., Libeu, C. P., Mizushima, T., Yamaguchi, H., Tomizaki, T. & Tsukihara, T. (1998). *Science*, **280**, 1723–1729.
- Zheng, G., Lu, X.-J. & Olson, W. K. (2009). *Nucleic Acids Res.* **37**, W240–W246.