Monica Purushothaman,^a Anna Varghese,^a Pradeep Kumar Mandal^b and Namasivayam Gautham^{a*}^aCAS in Crystallography and Biophysics, University of Madras, Guindy, Chennai, Tamil Nadu 600 025, India, and ^bInstitut Européen de Chimie et Biologie (IECB), 2 Rue Robert Escarpit, 33607 Pessac CEDEX, FranceCorrespondence e-mail:
n_gautham@hotmail.comReceived 16 March 2014
Accepted 12 May 2014**PDB reference:** d(CCCCGGTACCGGGG)₂,
4okl

Structure of d(CCCCGGTACCGGGG)₂ at 1.65 Å resolution

The crystal structure of the tetradecanucleotide d(CCCCGGTACCGGGG)₂ has previously been reported as an A-type double helix at a resolution of 2.5 Å in space group *P*₄₁. Here, the structure of this sequence was determined at a significantly higher resolution of 1.65 Å in space group *P*₄₁₂₂. The differences in crystal packing between the former and latter are described. The crystallographic asymmetric unit consists of one tetradecanucleotide duplex that spans more than one full turn of the A-helix. This structure allowed the unambiguous identification of solvent interactions.

1. Introduction

We report here the structure of the tetradecanucleotide d(CCCCGGTACCGGGG)₂ at a resolution of 1.65 Å in the tetragonal space group *P*₄₁₂₂. The DNA sequence crystallizes as an A-type double helix. The structure of this sequence has been previously reported at a resolution of 2.5 Å (Mandal *et al.*, 2012). Apart from the higher resolution, which allows identification of the solvent interactions and supports a discussion on the economy of hydration, the present crystal differs in another way from that reported previously. There is a difference in the space group, as a probable consequence of the slightly different inter-helical interactions. This could be due to the addition of ZnCl₂ to the crystallization setup, unlike MnCl₂ in the previous report, although in the present case the ion is not visible in the electron-density map. The sequence had been designed to form a Holliday junction with ACC as the trinucleotide core, which is postulated to stabilize the junction (Eichman *et al.*, 2002). In the event, the sequence crystallized as an A-DNA helix.

2. Materials and methods

2.1. Crystallization, data collection and data processing

The PAGE-purified oligonucleotide sequence d(CCCCGGTACCGGGG) and other chemicals were purchased from Sigma-Aldrich Chemicals Pvt. Ltd (Bangalore, India). The self-complementary strands were annealed by raising the temperature of the solution in a water bath to 343 K for an hour and bringing it to room temperature gradually at a rate of about 10 K per 30 min. The crystallization drop consisted of 1 µl 1 mM DNA, 1 µl 50 mM sodium cacodylate buffer pH 7.0, 0.5 µl 10 mM ZnCl₂, 1 µl 10 mM spermine; the reservoir comprised 33% (v/v) MPD. Tetragonal crystals appeared in 48 h with dimensions of 0.1 × 0.1 × 0.15 mm. Crystals were flash-cooled in liquid nitrogen at 100 K with the mother liquor as the cryoprotectant.

Diffraction data were collected on the BM14 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at a wavelength of 0.95372 Å using a MAR CCD 225 detector. 180 images were collected at 1° oscillation angle with an exposure time of 7 s. The data were processed using *iMosflm* (Battye *et al.*, 2011) in the *CCP4* suite (Winn *et al.*, 2011). The crystals belonged to the tetragonal space group *P*₄₁₂₂. The unit-cell parameters are *a* = *b* = 41.27, *c* = 87.67 Å. Data-collection and processing statistics are provided in Table 1.

Table 1

Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	30.05–1.65 (1.74–1.65)
Completeness (%)	99.8
σ cutoff	$F > 1.35\sigma(F)$
No. of reflections, working set	9654 (1188)
No. of reflections, test set	964 (131)
Final R_{cryst}	0.227 (0.223)
Final R_{free}	0.243 (0.265)
No. of non-H atoms	
Nucleic acid	568
Water	44
Total	612
R.m.s. deviations	
Bonds (Å)	0.003
Angles (°)	0.846
Average B factors (Å ²)	
Nucleic acid	43.2
Water	46.1

The previously reported unit cell for this sequence (Mandal *et al.*, 2012) had unit-cell parameters $a = b = 29.34$, $c = 87.69$ Å in space group $P4_1$. This is related to the present unit cell by a rotation of 45° about the c axis and extension of the a and b axes by a factor of $2^{1/2}$. In other words, the ab face diagonals of the previously reported cell constitute the a and b axes of the present cell. To check whether we had the correct cell, appropriate transformations were made and the data were reprocessed with both cell dimensions using *XDS* (Kabsch, 2010). The R_{merge} values for the smaller cell and the larger cell are 19 and 4.8%, respectively. These values indicate that the crystal belongs to the tetragonal system with the larger cell dimensions. The unequivocal systematic absences of the reflections $h00$, $h = 2n + 1$; $0k0$, $k = 2n + 1$ and $00l$, $l \neq 4n$ indicated either space group $P4_12_12$ or $P4_32_12$. The structure could be solved only in $P4_12_12$. Table 1 gives the data-processing statistics using *iMosflm* (Battye *et al.*, 2011). We discuss below the changes in the packing interactions that could have led to the change in the unit cell.

2.2. Structure determination and refinement

The Matthews coefficient (Matthews, 1968) was calculated as $2.22 \text{ Å}^3 \text{ Da}^{-1}$ indicating the presence of a duplex in the asymmetric unit with a solvent content of 63%. A structure-factor check using *SFCHECK* (Vaguine *et al.*, 1999) revealed that the crystal was not twinned. The overall Wilson B factor was 29.8 Å^2 .

The structure was solved by molecular replacement using *AMoRe* (Navaza, 1994) in the *CCP4* suite (Winn *et al.*, 2011). The previously reported crystal structure of the present sequence (PDB entry 3v9d; Mandal *et al.*, 2012) was used as the search model. A solution with a correlation coefficient of 89.4 and an R factor of 26.9% was obtained. The structure was refined using *REFMAC5* (Murshudov *et al.*, 2011) with maximum-likelihood targets and *phenix.refine* from the *PHENIX* suite (Adams *et al.*, 2010). Although the tetradecanucleotide was crystallized in the presence of Zn^{2+} , no prominent peaks were found in the $2F_o - F_c$ map, indicating the absence of Zn^{2+} in the structure. 45 water molecules were added across different stages of refinement as indicated by the electron density in the appropriate $F_o - F_c$ map. The difference Fourier map was contoured at the 3σ level to place the water molecules. Giving full occupancy to the water molecules resulted in negative electron density. Occupancy refinement was carried out to ensure that there was no negative density at those positions in subsequent maps.

The refinement statistics are provided in Table 1. The model and electron-density map were analysed using *Coot* (Emsley & Cowtan,

2004). *X3DNA* (Lu & Olson, 2003) was used to carry out structural analysis and geometrical calculations. Visual representations were generated using *PyMOL* (DeLano, 2002). The PDB code is 4okl.

3. Results and discussion

3.1. Sequence as A-DNA

Although it was designed to form a Holliday junction with ACC as its core trinucleotide, the sequence forms an A-type double helix. Despite the presence of sequence symmetry, there is no structural symmetry observed between the two heptameric halves. The crystallographic asymmetric unit consists of a duplex with about 1.3 turns of the A-type helix.

Based on the differences in the hydration of the DNA surface, Basham *et al.* (1995) have attempted to predict a triplet code for the A-DNA duplex formation. This technique involves the calculation of the difference in solvent free energy of the A and B forms of DNA. Their results show that the triplet GTA favours the formation of A-DNA. This sequence forms the central triplet of the present tetradecamer. Also, studies (Langridge, 1969; Arnott & Hukins, 1972) have shown that the presence of oligo(dC)-oligo(dG) sequences, such as those that occur here, favour the formation of A-type DNA.

3.2. Overall structure

A right-handed A-type DNA duplex occupies the crystallographic asymmetric unit (Fig. 1). Unsurprisingly, the structure is almost identical to the previously reported structure for the present sequence, PDB entry 3v9d (Mandal *et al.*, 2012), with an r.m.s.d. of 0.53 Å when the two are superposed. The details of the structures are also the same and no significant variations are found in the helical parameters. 3v9d has Mn^{2+} bound at the major-groove side of the $\text{C}_2\text{-G}_{13}$ base pair. In the present structure, no Zn^{2+} ions are observed in the electron density, although the conformation at the metal-binding site is almost exactly the same.

A-type DNA is characterized by a wide and shallow minor groove and a narrow and deep major groove. The average minor-groove width is 9.7 Å, less than the value for the fibre model A-DNA

Figure 1

Wall-eyed stereoview of d(CCCCGGTACCGGGG)₂ with the electron-density map contoured at 1.0σ .

(11.28 Å; Arnott & Hukins, 1972). The average major-groove width is 6.47 Å, which is higher than in the case of 3v9d (5.97 Å) as well as the fibre value of 4.19 Å.

3.3. Crystal packing

It has been observed in A-DNA octamers that sequences with a central pyrimidine–purine step mostly crystallize in one of the space groups in the tetragonal system (Tippin & Sundaralingam, 1996). The present tetradecanucleotide has a central TpA step, and has crystallized in the tetragonal space group $P4_12_12$. There are eight duplex helices in the unit cell (Fig. 2a). The interactions of the duplex with its symmetry-related neighbours are the same as those seen previously in 3v9d (Mandal *et al.*, 2012; Fig. 2b), with one significant exception: that between the reference duplex and the duplex related by $x, y, -z$ symmetry. The nearest O1P–O1P distance between these two duplexes is 4.2 Å in the present structure, with a water molecule positioned to mediate an interaction (Fig. 3a). In the $P4_1$ structure, a

similar arrangement of helices is found but they are related by 4_1 and have neither direct interactions nor water-mediated interactions (Fig. 3b).

The change in space group could also follow from the absence of any specific ion-binding site in the present structure, as compared with a clearly defined Mn^{2+} -ion position in the $P4_1$ structure. The presence of the ion on one half of the duplex and not on the other breaks the twofold symmetry within the duplex. Thus, the pair of helices related by a crystallographic twofold axis in the $P4_12_12$ structure is related by translation in the $P4_1$ structure.

A common packing pattern in crystals of A-DNA helices is that the terminal base pairs of one helix stack on the nearly flat minor groove of a neighbouring molecule (Rabinovich *et al.*, 1988; Shakked *et al.*, 1990; Ramakrishnan & Sundaralingam, 1993). Such an interaction is observed in the present structure. In the other (octamer) structures reported in the tetragonal space group (Eisenstein & Shakked, 1995), the stacking occurs at the central base pairs. In the present structure, the interaction occurs at the C and G residues two base steps away from the central TpA step.

Figure 2

(a) Wall-eyed stereoview of the unit cell of the present structure. The reference helix is in blue and the helix related by $y, x, -z$ is in red. There is direct interaction between these two helices. (b) Wall-eyed stereoview of the unit cell of 3v9d (Mandal *et al.*, 2012). The reference helix is in blue and the helix related by the fourfold screw axis is in red. There is no interaction between the two helices. This difference gives rise to the different space groups.

Figure 3

(a) Water-mediated interaction between the reference helix (red) and its twofold symmetry-related molecule in the present structure. (b) The distance between the reference helix (red) and its 4_1 symmetry-related molecule in 3v9d shows the absence of direct or water-mediated interactions. The metal ion Mn^{2+} (shown in purple) breaks the possibility of the twofold symmetry in this structure.

Table 2

Comparison of phosphate hydration in A-DNA octamers (Eisenstein & Shakked, 1995) and in d(CCCCGGTACCGGGG)₂.

Oligomer	Resolution (Å)	Phosphate hydration	
		Total atoms [†]	No. of water atoms/ No. of hydrated O atoms
GGGCGCCC	1.75	28	37/24
GGCCGGCC	2.1	28	35/22
CCCCGGGG	2.5	28	29/18
CCCCGGTACCGGGG	1.65	52	11/10

[†] Only O1P and O2P phosphate O atoms are taken into consideration.

3.4. Hydration

45 water molecules were added at various stages of refinement as indicated by the appropriate difference Fourier map. The major groove and the minor groove are hydrated by 26 and 13 water molecules, respectively. Out of a total of 42 phosphate O atoms in the duplex, ten are hydrated by 11 water molecules. A significant decrease in the hydration around phosphate groups was observed when compared with the octamers (Table 2). The average number of water molecules per base pair in d(CCCCGGTACCGGGG)₂ is 2.42.

The major and minor grooves are well hydrated and, together with the hydration shells of the sugar-phosphate backbone, they possess extensive water networks. Wahl & Sundaralingam (1997) observed that in the major groove of the octamers, the N7 atom of the purines

or the N4 atoms of the pyrimidines are linked *via* two consecutive water molecules to the phosphate O1P. They also observed that the O4' sugar O atoms were frequently seen to be linked *via* one water molecule to O2 of the preceding pyrimidine base or N3 of the preceding purine base. In the present tetradecanucleotide, the following interactions are found: the N7 atoms of the G₅, G₆ and G₁₁ bases interact with the neighbouring phosphate O atom (Fig. 4*a*); the N4 atoms of the C₉ and C₁₀ bases also make such an interaction (Fig. 4*b*); the N3 atom of A₈ (Fig. 4*c*) and the O2 atoms of C₃ and C₄ (Fig. 4*d*) interact with the O4' atoms of their successive ribose residues *via* a water molecule.

Various water-mediated interactions between the duplexes generated by the fourfold screw axis are classified in the octamers (Eisenstein & Shakked, 1995). Of these, the interaction between a terminal base and the phosphate group of a neighbouring helix is observed in the present structure: the N7 of G₁₄ is linked to the phosphate O atom of the neighbouring duplex (Fig. 5*a*). In addition, the interaction between two phosphate O atoms belonging to symmetry-related neighbours, mentioned by Eisenstein & Shakked (1995), is also observed in the present structure (Fig. 5*b*). Furthermore, there are other water-mediated interactions between symmetry-related molecules that were not seen in the octamers: the N3 or N7 atom of the purines is linked to one of the sugar O atoms. An example is given in Fig. 5(*c*). Some of the interactions observed in the present structure are also present in 3v9d: the water-mediated interactions between the N3 atom of guanine bases and the C5' sugar

Figure 4

(*a*), (*b*) The water interactions in the major groove, *i.e.* the N7 atom of the purines or the N4 atoms of the pyrimidines interact with phosphate O1P. (*c*), (*d*) O4' sugar O atoms interact with O2 of the preceding pyrimidine base or N3 of the preceding purine base.

Figure 5
Water-mediated interactions between the molecules related by 4_1 and twofold axes: (a) terminal base-pair-backbone interactions, (b) backbone-backbone interactions, (c) N3 of G₁₁ interacts with C5' of C₂. (The reference helix is in pink.)

atoms of the neighbouring helices, and interactions between the phosphate O atoms of neighbouring molecules.

While in general the hydration pattern and the interactions that arise as a result of the water molecules are similar to those found in A-DNA of different lengths, there are variations in the interactions which are specific to the present structure, as is the absence of certain interactions. This could be an effect of the sequence or of the length. In particular, the central TpA base pair causes some differences in the terminal base-pair-minor-groove interactions. Some of these interactions may serve as an analogy to the interactions in DNA-DNA or protein-DNA complexes.

We would like to thank the Department of Biotechnology (DBT), Government of India for beamline time at BM14, ESRF, France, where the X-ray diffraction data for the present structure were collected. We also gratefully acknowledge the financial assistance of

the University Grants Commission, Government of India under the CAS program.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Arnott, S. & Hukins, D. W. (1972). *Biochem. Biophys. Res. Commun.* **47**, 1504–1509.
- Basham, B., Schroth, G. P. & Ho, P. S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6464–6468.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Eichman, B. F., Ortiz-Lombarda, M., Aymami, J., Coll, M. & Ho, P. S. (2002). *J. Mol. Biol.* **320**, 1037–1051.
- Eisenstein, M. & Shakked, Z. (1995). *J. Mol. Biol.* **248**, 662–678.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Langridge, R. (1969). *J. Cell. Physiol.* **74**, Suppl. 1, 1–20.

- Lu, X.-J. & Olson, W. K. (2003). *Nucleic Acids Res.* **31**, 5108–5121.
- Mandal, P. K., Venkadesh, S. & Gautham, N. (2012). *Acta Cryst.* **F68**, 393–399.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- 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.
- Rabinovich, D., Haran, T., Eisenstein, M. & Shakked, Z. (1988). *J. Mol. Biol.* **200**, 151–161.
- Ramakrishnan, B. & Sundaralingam, M. (1993). *J. Biomol. Struct. Dynam.* **11**, 11–26.
- Shakked, Z., Guershtein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. (1990). *Structure & Methods*, edited by R. H. Sarma, & M. H. Sarma, Vol. 3, pp. 55–72. New York: Adenine Press.
- Tippin, D. B. & Sundaralingam, M. (1996). *Acta Cryst.* **D52**, 997–1003.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Wahl, M. C. & Sundaralingam, M. (1997). *Biopolymers*, **44**, 45–63.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.