Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Exploration of triple-helical fragments: crystallization and preliminary X-ray diffraction of d(TGGCCTTAAGG)

Kristof Van Hecke, Koen Uytterhoeven and Luc Van Meervelt*

Biomolecular Architecture, Department of Chemistry, Katholieke Universiteit Leuven, B-3001 Leuven (Heverlee), Belgium

Correspondence e-mail: luc.vanmeervelt@chem.kuleuven.be

Received 3 October 2006
Accepted 15 December 2006

The nonamer d(GCGAATTCG) and decamer d(GGCCAATTGG), containing one and two overhanging guanines, respectively, form G·GC triplets in their crystal packing. In order to introduce a third subsequent T·AT triplet, the decamer was further extended by one overhanging thymine residue. Two different crystal morphologies of the sequence d(TGGCCTTAAGG) were obtained by hanging-drop vapour diffusion and diffracted to 2.5 and 2.3 Å resolution, respectively. However, both crystals belong to the orthorhombic space group $P2_12_12_1$, with similar unit-cell parameters. Therefore, the two data sets could be merged to a resolution of 2.4 Å with unit-cell parameters $a = 26.97$, $b = 41.12$, $c = 52.72$ Å.

1. Introduction

Nucleic acid triplexes can be considered to be the result of sequence-specific association between a single-stranded oligonucleotide and a double helix. Triple-helix formation is believed to play a role in numerous biochemical processes, *e.g.* regulation of transcription and replication, genetic recombination of homologous sequences, chromosome folding *etc*.

One of the most important applications of triple-stranded complexes is their potential role in regulating gene expression *in vivo*. The antigene strategy exploits the possibility of binding exogenous compounds to DNA inside genes or regulatory regions in order to hamper mRNA synthesis. This might be an efficient technique, as the antigene strategy is directed against the primary process of gene expression.

However, in order to overcome technical problems in the formation of a stable triplex, *e.g.* the relatively low stability of the triplex under physiological conditions and restriction of the target sequence, the development of novel and chemically modified triplex-forming oligonucleotides (TFOs) is required. For example, oligonucleotide analogues with the phosphodiester linkage replaced by a phosphoramidate linkage show a high binding affinity towards dsDNA (Gryaznov, 1999). Bridged nucleic acids (BNAs) with a methylene bridge between the O2′ and C4′ atoms (2′,4′-BNA) effectively inhibit the dsDNA–transcription factor interaction (Obika *et al.*, 2001). Moreover, the recognition of the TA and CG base pairs by a nucleobase to form a triplex structure can be further enhanced. A nucleoside analogue with a phenylimidazole derivative can recognize the shape of the PyPu base pair in the major groove (Griffin *et al.*, 1992). An analogue for recognizing TA has been developed (Guianvarc'h *et al.*, 2001) and a novel thymine nucleobase analogue which lacks a carbonyl O atom at the 4-position effectively recognizes a CG base pair (Prévot-Halter & Leumann, 1999).

As little structural knowledge is available on triplex DNA, we have introduced a novel method of obtaining structural details of DNA triplets by the use of overhanging bases (Van Meervelt *et al.*, 1995). The ability of overhanging bases to form triplexes opens up possibilities for obtaining ordered crystals of triple-helical fragments by extending the length of the overhanging strands and applying crystal-engineering techniques.

Detailed models of a parallel G·GC triplet and of a parallel and antiparallel $(G{\cdot}GC)_2$ triple-helical fragment have been obtained using the carefully chosen nonamer d(GCGAATTCG) and decamer d(GGCCAATTGG) (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996*b*), leading to an interaction in the major groove between the unpaired guanine residues and the GC Watson–Crick base pairs. The nonamer d(GCGAATTCG) crystallized in the B-DNA conformation with unpaired guanine bases at its ends. Two crystallographic independent parallel Hoogsteen G·GC base triplets are formed by interaction of the guanine bases with the terminal CG base pairs of neighbouring double helices (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996*a*). The decamer d(GGCCAATTGG) forms an octamer B-DNA helix with two overhanging G bases, which are able to form both parallel Hoogsteen and antiparallel reverse-Hoogsteen $(G{\cdot}GC)_2$ triple-helical fragments (Vlieghe *et al.*, 1996*b*; Vlieghe, Turkenburg *et al.*, 1999).

This crystal-engineering technique, which mimics triple-helical fragments in the crystal lattice of d(GGCCAATTGG), can at the same time be used to improve the resolution of the diffraction data obtained. We have previously reported the 1.9 Å resolution structure determination of the minor-groove binder DAPI (4′,6-diamidino-2-phenylindole) with d(GGCCAATTGG), revealing a novel off-centred binding with a hydrogen bond between the drug and a CG base pair (Vlieghe, Sponer *et al.*, 1999). Structure determinations of the same decamer with distamycin at 2.38 and 1.85 Å revealed two 1:1 binding modes for distamycin in the minor groove (Uytterhoeven *et al.*, 2002). The same crystal-engineering technique could be used to improve the resolution of the 1:1 d(GGCCAATTGG)–netropsin complex to 1.75 Å (Van Hecke *et al.*, 2005).

The decamer is now further extended by one overhanging thymine residue. These thymines can interact with AT base pairs, forming T·AT triplets, and hence extend the triple-helical fragment to three triplets. Owing to geometric restrictions of the T·AT triplet, the central AATT sequence had to be inverted to TTAA in comparison with the nonamer and decamer (Fig. 1).

Structure determination by molecular replacement is currently in progress using the decamer d(GGCCAATTGG) (Vlieghe, Turkenburg *et al.*, 1999; PDB code 431d) as a model.

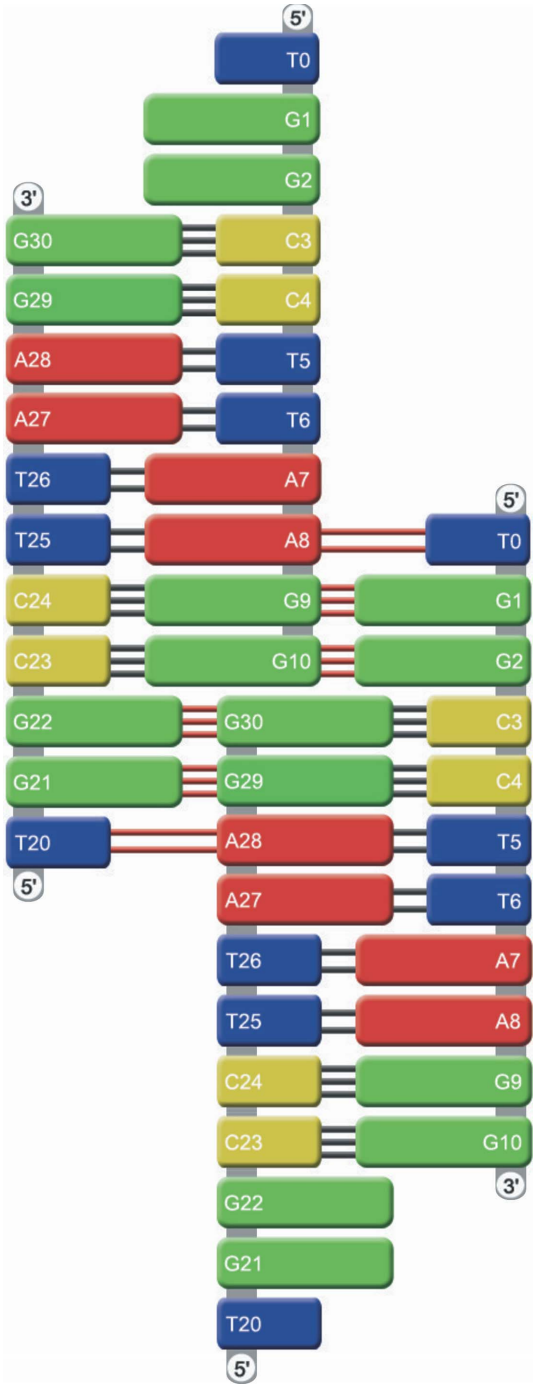
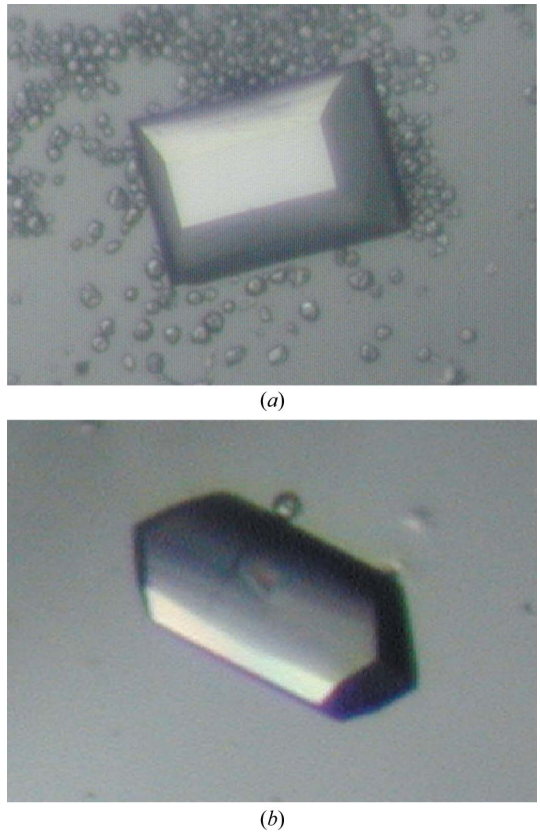


Figure 1
Schematic representation of the possible triple-helix formation of the undecamer d(TGGCCTTAAGG). Guanine, adenine, thymine and cytosine bases are shown in green, red, blue and yellow, respectively. The sugar-phosphate backbone is shown as a grey ribbon. Watson–Crick hydrogen bonds are shown in black and (reverse) Hoogsteen bonds in red.



(*a*)

(*b*)

Figure 2
Typical crystals of the undecamer d(TGGCCTTAAGG) obtained by sitting-drop vapour diffusion. (*a*) Bar-shaped block of 0.25 × 0.15 × 0.15 mm and (*b*) hexagonal block of 0.3 × 0.15 × 0.1 mm.

Table 1
Data-collection statistics for the undecamer d(TGGCCTTAAGG).

The values in parentheses are for the outer resolution shell.

	Bar-shaped crystal	Hexagonal-shaped crystal	Merged data
No. of reflections (full and summed partials)	17818 (2822)	6741 (986)	23187 (3533)
No. of unique reflections	2312 (350)	2203 (307)	2489 (349)
Resolution range (Å)	26.91–2.30	32.36–2.50	27.05–2.40
Outer resolution shell (Å)	2.42–2.30	2.64–2.50	2.53–2.40
Completeness (%)	84.8 (88.3)	98.5 (98.1)	99.1 (99.6)
$R_{\rm merge}$ (%)	5.6 (57.5)	5.6 (58.3)	9.0 (61.9)
Mean $I/\sigma(I)$	31.0 (9.0)	14.9 (1.8)	27.9 (9.1)
Multiplicity	7.7 (8.1)	3.1 (3.2)	9.3 (10.1)

2. Methods and results

2.1. Crystallization

The DNA undecamer d(TGGCCTTAAGG) was purchased from Oswel DNA service (University of Southampton, England).

Two different crystal forms suitable for X-ray diffraction were obtained using the sitting-drop vapour-diffusion method at 289 K. Bar-shaped blocks (Fig. 2*a*) grew after approximately four months, using 35 µl droplets of an optimized condition containing 7.5 m*M* sodium cacodylate buffer pH 6.0, 52.5 m*M* $MgCl_2$, 1.5 m*M* spermine, 10%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) and 0.8 m*M* ssDNA equilibrated against 20 µl 50%(*v*/*v*) MPD stock solution. Notably, after a few weeks the sharp crystal planes started to smoothen and the crystals finally became droplets. More stable crystals with a hexagonal-shaped block morphology (Fig. 2*b*) were obtained after approximately 12 months, using 35 µl droplets of a condition containing 21.1 m*M* sodium cacodylate buffer pH 6.0, 47.4 m*M* $MgCl_2$, 21.1%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) and 0.4 m*M* ssDNA equilibrated against 20 µl 50%(*v*/*v*) MPD stock solution. Remarkably high concentrations of MPD were needed for successful crystallization of the undecamer d(TGGCCTTAAGG).

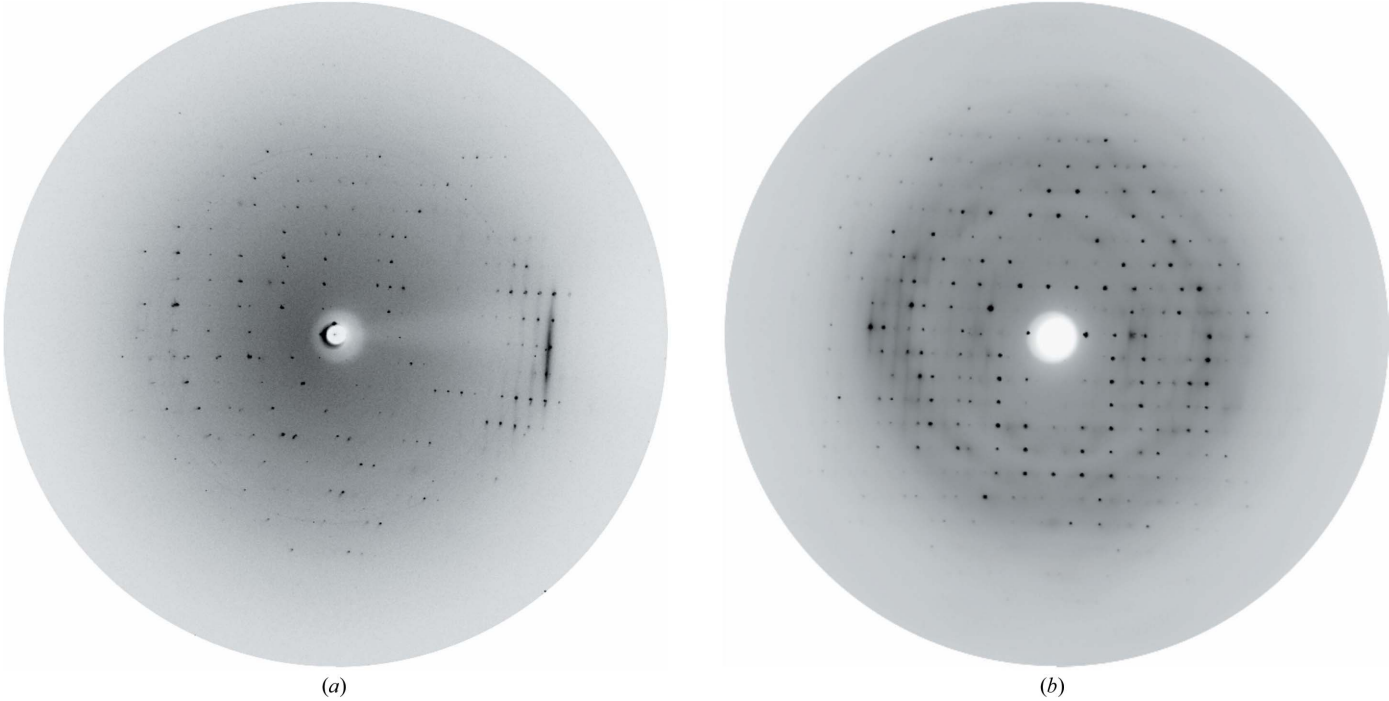
2.2. Data collection and processing

Both crystal types (bar-shaped and hexagonal-shaped) were used for data collection at the EMBL Outstation of the DESY synchrotron in Hamburg. A bar-shaped single crystal was used to collect a 84.8% complete data set at beamline BW7b with λ = 0.8428 Å, a φ range of 90°, a φ increment of 1.5° and a crystal-to-detector distance of 200 mm. A total of 2312 unique reflections were observed in the resolution range 26.91–2.30 Å ($R_{\rm merge}$ = 0.056). Although the diffraction pattern showed diffraction to 2.00 Å (Fig. 3*a*), careful analysis showed that the data in the outermost shells were of much lower quality; a 2.30 Å resolution cutoff was therefore chosen. Overloaded reflections were collected separately and included in the data set.

A 98.5% complete data was collected from a hexagonal-shaped single crystal at beamline X31 with λ = 1.2000 Å, a φ range of 90°, a φ increment of 1.5° and a crystal-to-detector distance of 200 mm. A total of 2203 unique reflections were observed in the resolution range 32.36–2.50 Å ($R_{\rm merge}$ = 0.056). Although the diffraction pattern showed diffraction to 2.25 Å (Fig. 3*b*), a 2.50 Å resolution cutoff was chosen.

Both crystals belong to the orthorhombic space group $P2_12_12_1$, with similar unit-cell parameters. Therefore, the two data sets could be merged to a resolution of 2.4 Å with unit-cell parameters a = 26.97, b = 41.12, c = 52.72 Å. The Matthews coefficient ($V_{\rm M}$) is 2.2 Å^3 Da^{-1} for one double helix in the asymmetrical unit, resulting in a solvent fraction of 42.8%, with a total unit-cell volume of 58 459.4 Å^3.

The resulting $R_{\rm merge}$ value of the averaged data set is higher than those of the individual data sets, which may be caused by the different



(*a*) (*b*)

Figure 3
1.5° oscillation images of crystals of the undecamer d(TGGCCTTAAGG) taken on a MAR Research image plate (EMBL, Hamburg). (*a*) Image of a bar-shaped crystal taken on beamline BW7b. (*b*) Image of a hexagonal-shaped crystal taken on beamline X31. The resolution edges of the images are 2.00 Å (*a*) and 2.25 Å (*b*), respectively. Base-stacking reflections can be noticed at ∼3 Å in (*a*).

crystallization conditions, crystal dimensions (absorption effects) and data-collection conditions.

For both crystal morphologies, data were collected on a MAR345 imaging-plate detector and a liquid-nitrogen cryostream at 100 K was used to decrease radiation damage to the crystals.

Data were processed with *MOSFLM* v.0.4.5 (Leslie, 1992) and scaled using *SCALA* v.3.2.5 (Evans, 1997). The latter was used as part of the *CCP4* suite (Collaborative Computational Project, Number 4, 1994). Data-collection statistics are given in Table 1.

As the space group and unit cell show isomorphism with the decamer d(GGCCAATTGG) (Vlieghe, Turkenburg *et al.*, 1999), structure determination by molecular replacement is currently in progress using this decamer d(GGCCAATTGG) (PDB code 431d) as a model. However, the unit-cell parameters of the decamer d(GGCCAATTGG) are a = 26.11, b = 36.46, c = 52.56 Å, indicating that the extension with one overhanging thymine residue most influences the structural arrangement in the b direction.

We thank the staff of the EMBL Hamburg Outstation for their support with the synchrotron experiments. Support from the European Community Research Infrastructure Action under the PF6 'Structuring the European Research Area Programme' contract No. RII3/CT/2004/5060008 is gratefully acknowledged.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Griffin, L. C., Kiessling, L. L., Beal, P. A., Gillespie, P. & Dervan, P. B. (1992). *J. Am. Chem. Soc.* **114**, 7976–7982.
Gryaznov, S. M. (1999). *Biochim. Biophys. Acta*, **1489**, 131–140.
Guianvarc'h, D., Benhida, R., Fourrey, J.-L., Maurisse, R. & Sun, J.-S. (2001). *Chem. Commun.* **2001**, 1814–1815.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Obika, S., Uneda, T., Sugimoto, T., Nanbu, D., Minami, T., Doi, T. & Imanishi, T. (2001). *Bioorg. Med. Chem.* **9**, 1001–1011.
Prévot-Halter, I. & Leumann, C. J. (1999). *Bioorg. Med. Chem. Lett.* **9**, 2657–2660.
Uytterhoeven, K., Sponer, J. & Van Meervelt, L. (2002). *Eur. J. Biochem.* **269**, 2868–2877.
Van Hecke, K., Nam, P. C., Nguyen, M. T. & Van Meervelt, L. (2005). *FEBS J.* **272**, 3531–3541.
Van Meervelt, L., Vlieghe, D., Dautant, A., Gallois, B., Précigoux, G. & Kennard, O. (1995). *Nature (London)*, **374**, 742–744.
Vlieghe, D., Sponer, J. & Van Meervelt, L. (1999). *Biochemistry*, **38**, 16443–16451.
Vlieghe, D., Turkenburg, J. P. & Van Meervelt, L. (1999). *Acta Cryst.* D**55**, 1495–1502.
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.