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

Comparison between the orthorhombic and tetragonal forms of the heptamer sequence d[GCG(xT)GCG]/d(CGCACGC)

Koen Robeyns,[a] Piet Herdewijn[b] and Luc Van Meervelt[a]*

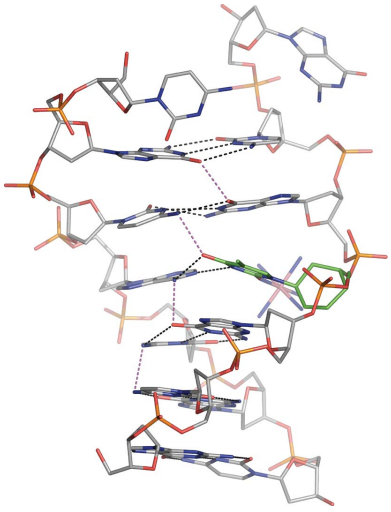
[a]Department of Chemistry, Biomolecular Architecture and BioMacS, Katholieke Universiteit Leuven, Belgium, and [b]Laboratory of Medicinal Chemistry, Rega Institute for Medical Research and BioMacS, Katholieke Universiteit Leuven, Belgium

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

Received 25 May 2010
Accepted 6 August 2010

PDB Reference: d[GCG(xT)GCG]/d(CGCACGC), 3lln.

NDB Reference: d[GCG(xT)GCG]/d(CGCACGC), NA0435.



Cyclohexene nucleic acid (CeNA) building blocks can be introduced into natural DNA sequences without a large conformational influence because of the ability of the six-membered sugar ring to mimic both the C2′-*endo* and C3′-*endo* conformations of the naturally occurring ribofuranose sugar ring. The non-self-complementary DNA sequence d[GCG(xT)GCG]/d(CGCACGC) with one incorporated CeNA (xT) moiety crystallizes in two forms: orthorhombic and tetragonal. The tetragonal form, which diffracts to 3 Å resolution, is a kinetically stable polymorph of the orthorhombic form [Robeyns *et al.* (2010), *Artificial DNA*, **1**, 1–7], which diffracts to 1.17 Å resolution and is the thermodynamically stable form of the CeNA-incorporated duplex. Here, the two structures are compared, with special emphasis on the differences in crystal packing and the irreversible conversion of the kinetic form into the high-resolution diffracting thermodynamic form.

1. Introduction

Cyclohexene nucleic acids (CeNAs) are known to mimic both the C2′-*endo* and C3′-*endo* conformations of natural occurring DNA. They do so by adopting one of two extreme conformations: the $^{2}H_{3}$ (similar to C2′-*endo*) and $^{3}H_{2}$ (similar to C3′-*endo*) half-chair conformations. These conformations (Fig. 1) are obtained by replacing the natural 2′-deoxyribofuranose ring with a cyclohexene ring. CeNAs have been explored in the context of the synthetic and enzymatic production of xeno-nucleic acids as an alternative to DNA and RNA biosynthesis (Herdewijn & Marlière, 2009).

Recently, the structure of the orthorhombic form of the non-self-complementary DNA sequence d[GCG(xT)GCG]/d(CGCACGC), where xT is a cyclohexene nucleotide, was determined to 1.17 Å resolution (NDB code BD0108; Robeyns *et al.*, 2010). This high-resolution structure crystallized in space group $P2_12_12_1$ and will be referred to as the orthorhombic form. Crystals of a tetragonal form were harvested from the same crystallization conditions and diffracted to 3.0 Å resolution. Here, we describe the overall structure of the tetragonal polymorph and its conversion to the orthorhombic form.

2. Materials and methods

2.1. Oligonucleotide synthesis and crystallization conditions

2.1.1. Oligonucleotide synthesis. The CeNA nucleoside (Wang & Herdewijn, 1999) with a thymine base, as well as the protected

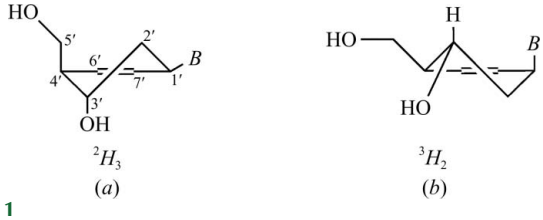


Figure 1
Schematic representation of the two half-chair conformations of CeNA, (*a*) $^{2}H_{3}$ and (*b*) $^{3}H_{2}$, which mimic the C2′-*endo* and C3′-*endo* conformations of the deoxyribofuranose ring, respectively. The CeNA numbering scheme is indicated in (*a*).

Table 1
Data-collection statistics for the modified heptamer sequence d[GCG(xT)GCG]/d(CGCACGC), where (xT) is a cyclohexene residue.

Space group	$P4_12_12$
Resolution range (Å)	18.55–3.00
No. of measured/unique reflections	7757/681
Completeness (%)	99.6
R_{merge} (%)	10.5
Mean $I/\sigma(I)$	10.6
Multiplicity	11.4
Mosaicity (°)	1.00

Table 2
Refinement statistics for the modified heptamer sequence d[GCG(xT)GCG]/d(CGCACGC), where (xT) is a cyclohexene residue.

Resolution range (Å)	18.55–3.00
No. of reflections	678
No. of atoms	298
Final R value (all data) (%)	22.81
Final R_{free} value (%)	n.a.
R.m.s. deviation from restraint target value	
Bond lengths (Å)	0.012
Angles (°)	1.925
Distances from restraint planes (Å)	0.010
Mean B value (Å^2)	71

phosphoramidite nucleosides that were used in the oligonucleotide synthesis, were synthesized by the Laboratory of Medicinal Chemistry at the Rega Institute, Leuven (Gu *et al.*, 2004). The preparation of the non-self-complementary nucleic acid duplex is described in Robeyns *et al.* (2010).

2.1.2. Crystallization conditions. Crystallization conditions for the heptamer were screened using a 24-matrix screen developed for oligonucleotides (Berger *et al.*, 1996). The crystallization conditions and the crystal morphology of the heptamer duplex were identical for the tetragonal and orthorhombic forms and have been described previously (Robeyns *et al.*, 2010). The orthorhombic form of the heptamer sequence was mostly harvested from the crystallization droplets; the occurrence of the tetragonal form was somewhat uncontrollable, possibly because the tetragonal form converts to the orthorhombic form over time. The orthorhombic form is believed to be the thermodynamically stable form, as the tetragonal form was only detected in fresh crystallization droplets.

2.2. Data collection and structure determination

2.2.1. Data collection. Data were collected in $\Delta\varphi$ increments of 2° over a total of 180° on beamline X12 (λ = 0.9537 Å) at the EMBL synchrotron facility, Hamburg. A crystal with dimensions of 0.15 × 0.15 × 0.44 mm was harvested on-site and flash-cooled to 100 K. The diffraction data were processed with *MOSFLM* and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994) truncated to 3.0 Å. The crystal belonged to the tetragonal space group $P4_12_12$, with unit-cell parameters $a = b = 25.48$, $c = 81.15$ Å. Compared with the previously determined 1.17 Å resolution orthorhombic form (unit-cell parameters a = 25.720, b = 33.570, c = 81.200 Å), this is a volume reduction of about 25%. The data-collection statistics are summarized in Table 1.

2.2.2. Structure determination. The molecular-replacement program *Phaser* (McCoy *et al.*, 2005) was used to solve the crystal structure of the tetragonal form of the heptamer sequence. The heptamer sequence in the orthorhombic form consists of six base pairs, with the terminal base pair flipped away. The helical region of one such duplex was used as a search model in the molecular-replacement procedure. The flipped-away bases were fitted into the $2|F_o| - |F_c|$ electron-density maps after subsequent cycles of refinement with *REFMAC* (Murshudov *et al.*, 1999). The standard dictionary files were used, with an additional dictionary entry for the CeNA moiety. Target values for the new cyclohexene thymine residue were obtained from the CSD database (Allen, 2002) and have been validated in previously refined crystal structures containing CeNA residues (Robeyns *et al.*, 2008*a*,*b*, 2010).

The likelihood-based refinement using *REFMAC* resulted in an R value of 22.67% for 298 atoms and 679 unique reflections (99.56% complete data set). Water molecules were located using *Coot* (Emsley & Cowtan, 2004) and were monitored during refinement. One cobalt hexamine residue was also located in the difference map and was refined at 50% occupancy; this is in contrast to the orthorhombic form, in which seven cobalt hexamine moieties were located. Some refinement statistics are given in Table 2.

No R_{free} was calculated as a 5–10% data set would not constitute a statistically robust cross-validation set (Kleywegt & Brünger, 1996).

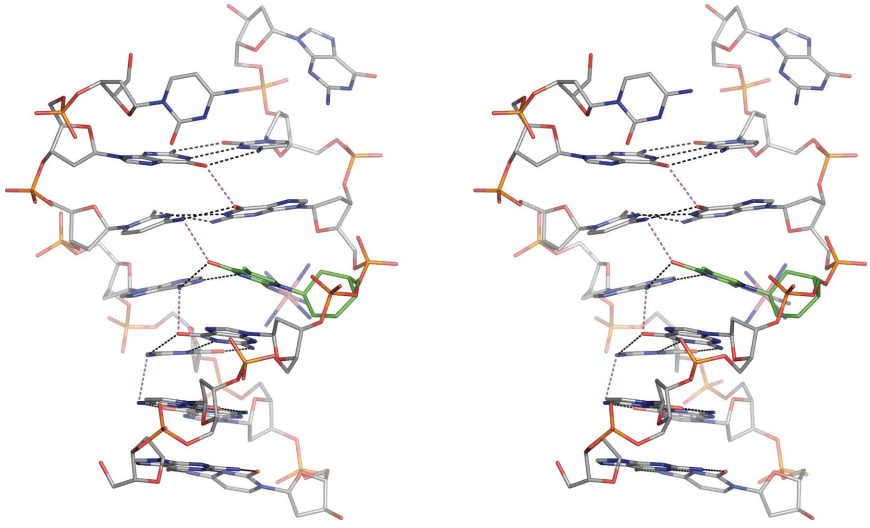


Figure 2
Stereo representation of the heptamer sequence d[GCG(xT)GCG]/d(CGCACGC) with the CeNA residue (xT) shown in green. Hydrogen bonds are indicated by dashed lines, with standard Watson–Crick hydrogen bonds in black; other inter-strand interactions are shown in magenta. The two central inter-strand interactions can be regarded as bifurcated hydrogen bonds. The two other interactions are a carbonyl–carbonyl and an amine–amine interaction.

All molecular figures were created using the program *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Overall structure and comparison of tetragonal and orthorhombic forms

The non-self-complementary heptamer sequence d[GCG(xT)GCG]/d(CGCACGC) forms a right-handed helix belonging to the B-type family. As in the orthorhombic form, six bases are engaged in Watson–Crick base pairing, with the terminal G·C base pair flipped out (Fig. 2). The cyclohexene moiety refines to the ^{2}E envelope conformation, which can be considered as a distorted $^{2}H_{3}$ half-chair conformation and is the expected C2′-*endo*-like conformation for CeNA monomers when incorporated into B-type helices. In the orthorhombic form, there was a discrimination in the adopted CeNA puckering between the two distinct helices in the asymmetric unit, in which one duplex incorporated a CeNA monomer with a $^{3}H_{2}$ half-chair conformation and the other showed an ^{2}E envelope conformation similar to that observed here. Although the 3 Å resolution is not sufficient to directly observe the CeNA sugar puckering in the electron-density maps, the refined ^{2}E envelope conformation is likely to be correct. This conformation has previously been reported in a crystallographic study in which a single CeNA residue was incorporated into the Dickerson sequence CGCGA(xA)TTCGCG (Robeyns *et al.*, 2008*a*). Moreover, the $^{3}H_{2}$ half-chair conformation observed in the orthorhombic form causes the phosphate groups to be closer to one another, with an inter-phosphate distance around the CeNA modification of about 5.8 Å (which is more likely for A-type DNA). This behaviour is not observed for the CeNA puckering in the tetragonal form; the inter-phosphate distance of 7.2 Å is further evidence for the assignment of the CeNA sugar puckering as an ^{2}E envelope conformation.

The r.m.s. deviations between the tetragonal form and the two distinct duplexes of the orthorhombic form are 0.97 and 1.02 Å, respectively. The largest deviations are found for the phosphate groups linking base pairs xT(4)·A(14) and G(5)·C(13).

As in the orthorhombic form, the helical regions of neighbouring helices stack onto each other, with the first base pair (of the next duplex in the helical column) sandwiched between the flipped-away bases. The flipped-away guanine base interacts with this first base pair to form a minor-groove G*(G·C) triplet (Nunn & Neidle, 1998). The flipped-away cytosine base also interacts with the guanine from this base pair and is further stabilized by interactions with a phosphate group in a neighbouring helical column.

While the structure in the orthorhombic form is extensively stabilized by interactions with seven cobalt hexamine complexes, only one such complex is present in the tetragonal form. This cobalt hexamine stabilizes the CeNA thymine by interacting with the O2 base atom.

The xT·A base pair and the preceding and following G·C base pairs are characterized by a large negative propeller twist (−11°, −24° and −18° for the consecutive base pairs as calculated using 3*DNA*; Lu & Olson, 2003). This causes some base atoms on the major-groove side to be positioned in between the base pairs, making two inter-strand bifurcated hydrogen bonds (the two central dashed magenta bonds in Fig. 2) or favourable amine–carbonyl overlaps [O4(T4)···N4(C13) and O6(G3)···N6(A14)] between opposite strands. This type of behaviour has previously been described for consecutive A·T base pairs (Nelson *et al.*, 1987), where the large propeller twist also causes an increase in base-pair overlap that stabilizes the duplex. A similar increase in overlap is observed in the tetragonal form compared with the orthorhombic form, in which cobalt hexamine interactions stabilize the CeNA sugar conformation and crystal packing. The sum of the base-pair overlap (as calculated by 3*DNA*) for the two steps around the central xT·A base pair is 16 Å^{2} for the tetragonal form and 10.5 and 13.1 Å^{2} for the orthorhombic form. Two other short

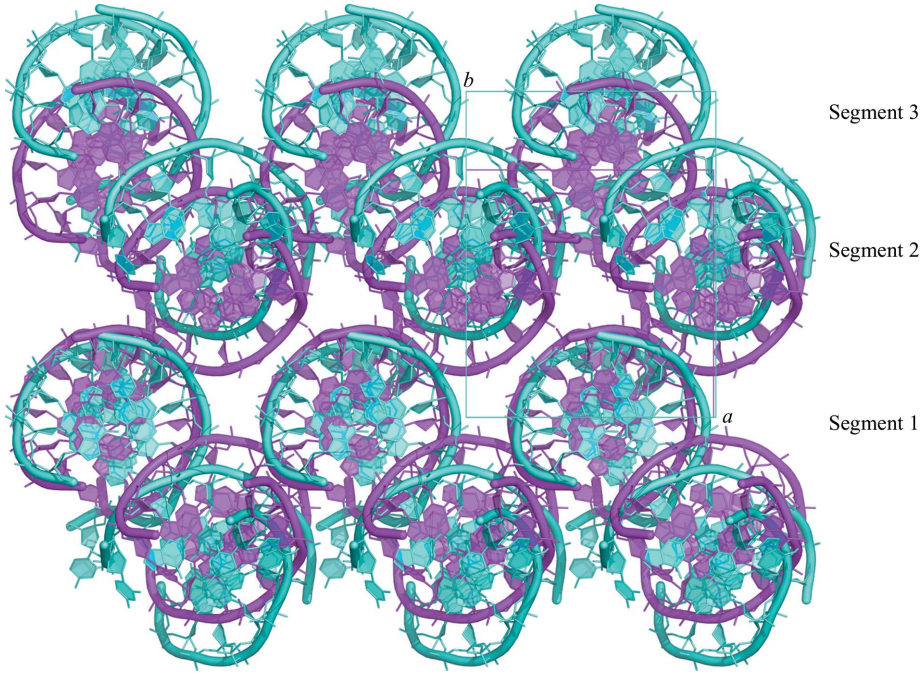


Figure 3
Superposition of the crystal packing of the tetragonal form (magenta) and the orthorhombic form (cyan). The helical columns in segment 1 are aligned. Conversion into the orthorhombic form moves the helical columns (magenta) in segment 3 by 8 Å in the direction of the *b* axis to the orthorhombic form (cyan). The larger cavity is then occupied by the helical columns in segment 2.

interactions are also observed, O6(G5)···O(G12) and N4(C2)···N4(C15), which are most likely to be destabilizing (Fig. 2).

3.2. Crystal packing and conversion to the orthorhombic form

The tetragonal and orthorhombic crystal forms can easily be interconverted into each other. Basically, the tetragonal and orthorhombic forms differ only in the length of the *b* axis. The stacking into the helical columns is similar, as are the flipped-away bases. It is quite striking that the volume per base pair (V_{bp}) is only 1098 Å^3, taking into account only the six base pairs with Watson–Crick interactions, with a Matthews coefficient of 1.56 Å^3 Da^{-1}, while the V_{bp} range for B-type DNA is typically 1175–1462 Å^3 (Heinemann, 1991). Relaxation of the tight crystal packing in the tetragonal form moves the helical columns 8 Å from each other along the *b* axis and converts it into the lower symmetry orthorhombic form (V_{bp} is 1460.6 Å^3, with a Matthews coefficient of 2.06 Å^3 Da^{-1}). In Fig. 3 the orthorhombic form is superposed onto the tetragonal form. The helical spacing along the *a* axis is unchanged, as is the stacking in the *c* direction. When converting the tetragonal form into the orthorhombic form, the fourfold symmetry is lost and the helical columns in segment 3 (as indicated in Fig. 3) are translated 8 Å away from segment 1. This allows the helical columns in segment 2 to occupy the enlarged space between segments 1 and 3.

The conversion into the orthorhombic form is probably triggered by the uptake of water molecules, resulting in an increase in volume and unit-cell expansion along the *b* axis. This water-mediated transformation results in solvent-accessible channels and pockets which are large enough to transport cobalt hexamine throughout the crystals. A probe with a 2.8 Å radius was used to contour the internal surface of the crystal structure using the Python utility *HOLLOW* (Bosco & Gruswitz, 2008). Visualization shows solvent channels of at least 5.6 Å in diameter throughout the crystal (cobalt hexamine has a Co—N distance of 1.97 Å and a cation volume of approximately 60 Å^3, which is comparable to a sphere of 2.4 Å radius; Kucharski *et al.*, 2000). The cobalt hexamine residues stabilize the packing and possibly make the conversion irreversible. The conversion from the kinetic tetragonal form to the thermodynamic orthorhombic form is probably driven by an enthalpy gain at the level of the crystal (lower space-group symmetry), at the level of the duplexes (displacement of 8 Å, resulting in a less tight packing) and at the level of the individual chains (the occurrence of alternative backbone conformations). In combination with the stabilizing effect of the cobalt hexamine interactions and a further optimization of the duplex structure, this results in a base-pair rotation to a position more perpendicular to the helical axis and eliminates the short inter-strand contacts (Fig. 2).

Dehydration of the orthorhombic form would not necessarily result in the tetragonal form, since the cobalt hexamine atoms are thought to reside within the crystal packing even after the expected volume decrease caused by the dehydration. Structural investigation of the dehydrated orthorhombic form is ongoing and might provide further insight into the conversion mechanism.

KR thanks the K. U. Leuven Research Fund for a postdoctoral position. We thank the staff of beamline X12 at EMBL Hamburg for help with the synchrotron experiments. BioMacS, the K. U. Leuven Interfacultary Centre for Biomacromolecular Structure, is supported by the Impulse Project of K. U. Leuven.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Berger, I., Kang, C., Sinha, N., Wolters, M. & Rich, A. (1996). *Acta Cryst.* D**52**, 465–468.
Bosco, K. H. & Gruswitz, F. (2008). *BMC Struct. Biol.* **8**, 49.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2002). *The PyMOL Molecular Viewer.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gu, P., Griebel, C., Van Aerschot, A., Rozenski, J., Busson, R., Gais, H. J. & Herdewijn, P. (2004). *Tetrahedron*, **60**, 2111–2123.
Heinemann, U. (1991). *J. Biomol. Struct. Dyn.* **8**, 801–811.
Herdewijn, P. & Marlière, P. (2009). *Chem. Biodivers.* **6**, 791–808.
Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
Kucharski, L. M., Lubbe, W. J. & Maguire, M. E. (2000). *J. Biol Chem.* **272**, 16767–16773.
Lu, X.-J. & Olson, W. K. (2003). *Nucleic Acids Res.* **31**, 5108–5121.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). *Nature (London)*, **330**, 221–226.
Nunn, C. M. & Neidle, S. (1998). *Acta Cryst.* D**54**, 577–583.
Robeyns, K., Herdewijn, P. & Van Meervelt, L. (2008*a*). *Nucleic Acids Res.* **36**, 1407–1414.
Robeyns, K., Herdewijn, P. & Van Meervelt, L. (2008*b*). *J. Am. Chem. Soc.* **130**, 1979–1984.
Robeyns, K., Herdewijn, P. & Van Meervelt, L. (2010). *Artificial DNA*, **1**, 1–7.
Wang, J. & Herdewijn, P. (1999). *J. Org. Chem.* **64**, 7820–7827.