Four highly pseudosymmetric and/or twinned structures of $d(\text{CGCGCG})_2$ extend the repertoire of crystal structures of Z-DNA

Zhipu Luo,^a Zbigniew Dauter^{a*} and Mirosław Gilski^{b,c}

Received 21 August 2017

Accepted 16 October 2017

Edited by Z. S. Derewenda, University of Virginia, USA

This paper is dedicated to Professor Zofia Kosturkiewicz on her 90th birthday.

Keywords: Z-DNA crystal structure; Z-DNA hexamer duplex; pseudomerohedral twinning; hectohedral twinning.

PDB references: $d(\text{CGCGCG})_2$, Z3A, 6aqt; Z3B, 6aqv; Z4A, 6aqw; Z4B, 6aqx

Supporting information: this article has supporting information at journals.iucr.org/d

^aSynchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA, ^bDepartment of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland, and ^cCenter for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland. *Correspondence e-mail: dauter@anl.gov

DNA oligomer duplexes containing alternating cytosines and guanines in their sequences tend to form left-handed helices of the Z-DNA type, with the sugar and phosphate backbone in a zigzag conformation and a helical repeat of two successive nucleotides. Z-DNA duplexes usually crystallize as hexagonally arranged parallel helical tubes, with various relative orientations and translation of neighboring duplexes. Four novel high-resolution crystal structures of $d(\text{CGCGCG})_2$ duplexes are described here. They are characterized by a high degree of pseudosymmetry and/or twinning, with three or four independent duplexes differently oriented in a monoclinic $P2_1$ lattice of hexagonal metric. The various twinning criteria give somewhat conflicting indications in these complicated cases of crystal pathology. The details of molecular packing in these crystal structures are compared with other known crystal forms of Z-DNA.

1. Introduction

Since the first crystal structures of left-handed DNA oligonucleotides were determined (Wang *et al.*, 1979), many crystal structures of Z-DNA have been deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000) and have been thoroughly analyzed (see, for example, Ho & Mooers, 1997; Mandal *et al.*, 2008). The majority of these structures are hexamer duplexes, with alternating sequences of purines and pyrimidines, usually cytosines and guanines, sometimes with bases that are modified by various substitutions. In all Z-DNA structures pyrimidines have the *anti* conformation of the glycosidic bond between the base and sugar and purines have the *syn* conformation, which causes the DNA phosphate-sugar backbone to adopt a left-handed zigzag conformation, instead of the smooth thread characteristic of the A-DNA and B-DNA forms. As a consequence, the repeating unit along the double helix in the Z-DNA consists of two pairs of nucleotides, instead of a single one as in right-handed DNA. The double helix in Z-DNA is narrower and has a longer repeat per base of about 3.7 Å, compared with 2.6 and 3.4 Å in A-DNA and B-DNA, respectively. The twist of the Z-DNA double helix is 60° per two pairs of nucleotides, so that two hexamer duplexes (12 nucleotides each) stacked in an end-to-end fashion, and related by a twofold screw axis, correspond to a full turn of the helix, with a length of about 44 Å.

Four new types of $d(\text{CGCGCG})_2$ crystals were obtained in the course of our investigations of Z-DNA crystallized under various conditions. They are all characterized by a high degree of twinning, and are referred to in the following as Z3A, Z3B, Z4A and Z4B. Their high-resolution structures are described

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the last resolution shell.

Structure	Z3A	Z3B	Z4A	Z4B	Z4A_ortho
PDB code	6aqt	6aqv	6aqw	6aqx	
Data collection					
No. of passes	2	2	1	1	
Beamline	SER-CAT 22-ID	SER-CAT 22-ID	SER-CAT 22-ID	SER-CAT 22-ID	
Detector	Rayonix MX300HS	Rayonix MX300HS	Rayonix MX300HS	Rayonix MX300HS	
Wavelength (Å)	0.9756	1.0000	0.9756	1.0000	
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Unit-cell parameters					
a (Å)	30.92	30.84	35.43	35.20	17.72
b (Å)	42.89	42.57	44.00	42.22	30.68
c (Å)	30.95	30.84	35.43	35.21	44.00
β (°)	119.98	120.05	119.83	119.83	
Resolution (Å)	27–1.05 (1.09–1.05)	27–1.30 (1.35–1.30)	31–1.30 (1.35–1.30)	31–1.55 (1.61–1.55)	31–1.30 (1.35–1.30)
Total reflections	193622 (8617)	120071 (10180)	84388 (8148)	48432 (4763)	
Unique reflections	32555 (3031)	17047 (1685)	22010 (2242)	13145 (1319)	
Completeness (%)	99.0 (93.3)	99.2 (97.9)	99.1 (98.8)	99.8 (100.0)	
Multiplicity	5.9 (2.8)	7.0 (6.4)	3.7 (3.6)	3.7 (3.6)	
Mosaicity range (°)	0.4–1.0	1.0–2.5	0.5–1.2	0.7–1.5	
R_{merge}	0.069 (0.586)	0.063 (0.646)	0.075 (0.405)	0.069 (0.493)	
R_{meas}	0.075 (0.730)	0.068 (0.703)	0.089 (0.476)	0.081 (0.579)	
$\langle I/\sigma(I) \rangle$	21.7 (1.6)	23.5 (2.5)	13.8 (2.2)	16.0 (2.1)	
$CC_{1/2}^\dagger$	0.74	0.82	0.75	0.99	
Wilson B factor (Å ²)	12.6	16.7	16.7	17.4	
Refinement					
Program	REFMAC5	SHELXL	REFMAC5	SHELXL	SHELXL
No. of nucleosides	12 × 3	12 × 3	12 × 3	12 × 3	12 × 1
Total No. of reflections	32538	32538	17033	16998	22007
R factor	0.1953	0.1443	0.1318	0.1420	0.1689
No. of free reflections‡	1408	1408	873	873	1020
R_{free}	0.2310	0.1921	0.1474	0.1981	0.1842
No. of reflections > $4\sigma(F_o)$		25155		14556	14954
R factor > $4\sigma(F_o)$		0.1267		0.1230	0.1289
R.m.s.d., bonds (Å)	0.010	0.012	0.010	0.010	0.008
R.m.s.d., angles (°)	1.70	1.57	1.59	0.95	0.97
Average B factor (Å ²)					
Overall	21.9	27.6	28.8	28.2	26.4
DNA	20.3	26.0	28.1	27.9	24.8
Solvent	33.3	36.9	35.2	30.7	36.3

[†] Correlation between intensities from random half-sets in the highest resolution bin as defined by Karplus & Diederichs (2012). [‡] The HKLF 5 reflection format of *SHELX* does not allow use of the R_{free} flag.

and compared with other crystal forms of Z-DNA duplexes. Two of the forms, with four duplexes in the asymmetric unit, are almost isomorphous with the recently published structure of the chimeric DNA–RNA hexamer duplex (PDB entry 5ebi; Gilski *et al.*, 2016), but differ in the details of molecular packing.

2. Materials and methods

2.1. Crystallization and diffraction data collection

The salt-free hexamer d(CGCGCG)₂ was purchased from Eurofins MWG Operon (Huntsville, USA) and was used without further purification. The oligomer was dissolved in 1 mM Tris–HCl pH 7.5 at a concentration of 2 mM, kept at 90°C for 1 min, annealed at 60°C for 10 min and then slowly cooled to room temperature. Crystals were grown using the sitting-drop vapor-diffusion method at room temperature by mixing 1 µl of the DNA oligomer solution and 1 µl of a precipitant solution prepared as follows: crystal Z3A, 200 mM MgCl₂, 50 mM Tris–HCl pH 7.0, 10% MPD; crystal Z3B,

100 mM KCl, 10 mM MgCl₂, 50 mM Tris–HCl pH 8.5, 30% PEG 400; crystal Z4A, 200 mM CaCl₂, 50 mM Tris–HCl pH 7.0, 10% MPD; crystal Z4B, 80 mM KCl, 20 mM MgCl₂, 40 mM sodium cacodylate pH 6.0, 12 mM spermine, 45% MPD. The Z3A and Z4A crystallization drops were equilibrated against a well solution consisting of 50% MPD, and the Z3B and Z4B drops were kept above their own precipitant solutions. The crystals were dipped into a cryoprotecting solution consisting of 50 mM Tris–HCl pH 7.0, 40% MPD for a few seconds before data collection and were then vitrified in liquid nitrogen.

High-resolution diffraction data sets were collected on the SER-CAT beamline 22-ID at the Advanced Photon Source, Argonne National Laboratory, USA using a Rayonix MX300HS fast CCD detector. Data sets Z3A and Z3B were recorded in two passes, with different effective exposures, in order to adequately measure the strongest (low resolution) and the weakest (high resolution) reflection intensities. The data were processed using *HKL-2000* (Otwinowski & Minor, 1997); the data-collection statistics are presented in Table 1. In

Table 2
Twin fractions obtained from refinement by *REFMAC5* and *SHELXL* (in parentheses) and R_{merge} on F_{obs} and F_{calc} for individual twinning operations, various twinning criteria, native Patterson peaks and statistics of normalized structure factors for selected Z-DNA crystals, calculated with *phenix.xtriage* from the *PHENIX* software package (Adams *et al.*, 2010; Zwart *et al.*, 2005).

	Untwinned	Twinned	Z3A	Z3B	Z4A	Z4B	5ebi	3e9w	3p4j
Space group			$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Z'			3	3	4	4	4	1	1
R_{merge} for individual symmetry operators on $I_{\text{obs}}/I_{\text{calc}}$									
$-h, k, -l^\dagger$	Twofold $\parallel$ 010		0.069/—	0.063/—	0.075/—	0.0069/—			
$l, -k, h$	Twofold $\parallel$ 101		0.046/0.355	0.048/0.316	0.251/0.644	0.220/0.487			
$-h - l, -k, l$	Twofold $\parallel$ 100		0.049/0.323	0.048/0.332	0.248/0.253	0.198/0.459			
$h, -k, -h - l$	Twofold $\parallel$ 001		0.051/0.346	0.037/0.343	0.136/0.659	0.031/0.307			
$l, k, -h - l$ and $-h - l, k, h$	Threefold $\parallel$ 010		0.040/0.337	0.039/0.367	0.251/0.689	0.211/0.513			
R_{merge} in 622			0.093	0.083	0.414	0.255	0.269	0.065	
Refined twinning fractions from <i>REFMAC5</i> (<i>SHELXL</i>)									
h, k, l			0.17 (0.14)	0.17 (0.17)	0.27 (0.27)	0.36 (0.38)	0.28		
$l, -k, h$			0.20 (0.16)	0.20 (0.20)	0.02 (0.02)	0.07 (0.06)	0.10		
$-h - l, k, h$			0.12 (0.13)	0.14 (0.14)	0.10 (0.08)	0.06 (0.05)	0.11		
$-h - l, -k, l$			0.16 (0.19)	0.16 (0.17)	0.13 (0.10)	0.38 (0.38)	0.28		
$l, k, -h - l$			0.15 (0.17)	0.12 (0.13)	0.34 (0.32)	0.09 (0.07)	0.11		
$h, -k, -h - l$			0.20 (0.21)	0.21 (0.19)	0.14 (0.21)	0.06 (0.06)	0.12		
Intensity statistics									
$\langle I^2 \rangle / \langle I \rangle^2$	2.000	1.500	2.626	2.647	2.845	4.778	2.732	2.018	2.138
$\langle F^2 \rangle / \langle F \rangle^2$	0.785	0.885	0.752	0.758	0.693	0.510	0.717	0.765	0.771
$\langle E^2 - 1 \rangle$	0.736	0.541	0.816	0.804	0.891	1.244	0.837	0.732	0.758
$\langle L \rangle$	0.500	0.375	0.453	0.450	0.482	0.493	0.482	0.521	0.501
$\langle L^2 \rangle$	0.333	0.200	0.281	0.278	0.312	0.325	0.315	0.372	0.337
$\langle H \rangle$	0.500	0.000	0.060	0.059	0.294	0.033	0.244	0.032	
$\langle H^2 \rangle$	0.333	0.000	0.006	0.006	0.141	0.002	0.092	0.002	
$\langle E^2 \rangle$									
All			0.99	0.99	0.99	0.98	1.00		
$h - l = 3n$ ($\neq 3n$)			1.70 (0.64)	1.47 (0.75)					
$l = 2n$ ($\neq 2n$)					1.14 (0.85)	1.51 (0.45)	1.50 (0.49)		
$h + l = 2n$ ($\neq 2n$)					1.55 (0.45)	1.77 (0.19)	1.20 (0.79)		

Native Patterson peak heights (as a fraction of the origin peak).

	Domains 1, 2			Domains 3, 4			Domains 5, 6		
	Z4A	5ebi	Z4B	Z4A	5ebi	Z4B	Z4A	5ebi	Z4B
Color in Fig. 4(b)	Green			Red			Blue		
Twin operations									
1	h, k, l			$l, k, -h - l$			$-h - l, k, h$		
2	$-h - l, -k, l$			$h, -k, -h - l$			$l, -k, h$		
Rotation versus domain 1 (axis) ($^\circ$)	0			120 (b)			240 (b)		
	180 (c^*)			180 (a^*)			180 ($a^* + c^*$)		
Position of Patterson peak	(0, 0, 1/2)			(1/2, 0, 0)			(1/2, 0, 1/2)		
Twin fractions									
1	0.27	0.28	0.36	0.34	0.11	0.09	0.10	0.11	0.06
2	0.13	0.28	0.38	0.14	0.12	0.06	0.02	0.10	0.07
Total twin fraction	0.40	0.56	0.74	0.48	0.23	0.15	0.12	0.12	0.13
Patterson peak height versus origin	0.41	0.52	0.87	0.47	0.20	0.59	0.16	0.18	0.57

$\dagger$ The values of R_{obs} for the $(-h, k, -l)$ operation are from the original merging of data during data processing. There are, therefore, no R_{calc} values for this operation.

all four cases the lattice metric corresponded to the hexagonal system but, taking into account the fact that these structures could not have trigonal or hexagonal symmetry (see below), all four data sets were scaled and merged in space group $P2_1$. For the subsequent representation of the Z4A structure in the orthorhombic space group $P2_12_12_1$ and the Z4B structure in a smaller $P2_1$ unit cell (see §3.3), several *ad hoc* programs written in Fortran, as well as dedicated scripts, were used to appropriately handle the reflections with different parities of indices.

2.2. Structure determination and refinement

All four structures were solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007), with the Z-DNA hexamer

duplex from the ultrahigh-resolution crystal structure PDB entry 3p4j (Brzezinski *et al.*, 2011) serving as a search model. Using MR_AUTO mode, three copies of the hexamer duplex were found in crystals Z3A and Z3B, whereas four duplexes were located in the Z4A and Z4B crystals.

All four structures were refined according to the same protocol and the results are presented in Table 1. The initial models were first refined with *REFMAC5* (Murshudov *et al.*, 2011) in space group $P2_1$, initially in isotropic mode and then anisotropically, with the TWIN option to automatically refine the six identified component twinning fractions. The H atoms were introduced as ‘riding’ on their parent atoms. About 1000 reflections were used for the calculation of R_{free} , but the validity of this criterion is doubtful in cases with such a high

level of twinning. Manual corrections to atomic models and the selection of solvent water atoms were executed with *Coot* (Emsley *et al.*, 2010).

Subsequently, all models were also refined with *SHELXL* (Sheldrick, 2015). Since the twinning operators do not form a cyclic set of symmetry elements, the HKLF 5 option was used after appropriate preparation of the input reflection files. In addition, the Z4A structure was interpreted and refined in the orthorhombic space group $P2_12_12_1$, with the input reflection file prepared as explained below in §3.3. R_{free} was not used in the *SHELXL* refinements because it is incompatible with the HKLF 5 command.

3. Results and discussion

3.1. Twinning analysis of diffraction data

All four crystals investigated here have unit cells that are very close to the typical hexagonal metrics, with a rhombic base of two equal edges and a third dimension of about 44 Å. Since almost all previously known structures of $d(\text{CGCGCG})_2$ are arranged as a pseudohexagonal set of helical $d(\text{CGCGCG})_2$ tubes with pairs of duplexes of ~44 Å in length related by a 2_1 screw axis, it was assumed that they are also arranged in an analogous way in these structures.

The possibility of crystallizing the tube-like Z-DNA oligomers in monoclinic, pseudo-hexagonal unit cells with several tubes in each cell raises the question of which topological realizations of such packing are possible. Inspection of Fig. 1 shows that the appropriate principle is the same as the triangulation rule applied to icosahedral viruses (Caspar & Klug, 1962). If the coordinates of the cell edge in the simplest hexagonal lattice are h and k , the number of crystallographically independent helices in the larger lattice (*i.e.* the triangulation number) is $T = h^2 + hk + k^2$ (Fig. 1). The structure with PDB code 3e9w corresponds to the combina-

Figure 1

Various triangulation possibilities for pseudohexagonal packing of Z-DNA helical 'tubes' represented in different colors. If the a (or b) unit-cell edge has coordinates h , k in the hexagonal pattern of circles, different combinations of h and k lead to crystal unit cells with $T = h^2 + hk + k^2$ symmetrically independent helices.

tion $h = 1$, $k = 0$ and $T = 1$, and the structures Z3A and Z3B to $h = 1$, $k = 1$, $T = 3$, whereas the structures Z4A and Z4B and that with PDB code 5ebi correspond to $h = 2$, $k = 0$, $T = 4$.

The only possible packing of three helices in the structures Z3A and Z3B is when they are located as follows: the first one at $x = 0$, $y = 0$, the second one at $x = 1/3$, $y = 2/3$ and the third one at $x = 2/3$, $y = 1/3$. Similarly, the Z4A and Z4B structures must have four helices at the following positions: the first one at $x = 0$, $y = 0$, the second one at $x = 1/2$, $y = 0$, the third one at $x = 0$, $y = 1/2$ and the fourth one at $x = 1/2$, $y = 1/2$. None of the trigonal space groups support such arrangements, and the only hexagonal space groups that are theoretically acceptable for Z4A or Z4B are $P6_3$ or $P6_322$, each possessing the 2_1 axis included in the 6_3 screw. However, the R_{merge} values for these two data sets in hexagonal symmetry were high (Table 2) and it was not possible to solve these structures by simply assuming hexagonal symmetry. Similarly, diffraction patterns for all four crystals could be indexed in six orientations of $C2$ symmetry, but the diffraction data did not merge satisfactorily in any version of this space group. A postmortem analysis of refined structures confirmed that they did not agree with the $C2$ symmetry.

Both of the programs, *REFMAC5* and *SHELXL*, used to refine the structures led to similar values for the estimated twinning fractions of all four crystals, differing by at most three percentage points. Such differences are not surprising, taking into account the high amount of pseudosymmetry in these structures and the different refinement algorithms that are used by the programs: the maximum-likelihood target in *REFMAC5* and full-matrix least squares in *SHELXL*.

It was therefore concluded that all four of these crystals belonged to space group $P2_1$ and were pseudomerohedrally twinned. All four structures were successfully solved and refined in this symmetry as sixfold (hectohedrally) twinned. However, analysis of the various twinning criteria for the diffraction data is somewhat unclear and confusing. Table 2 summarizes the various twinning criteria for the four crystal structures solved here and, in addition, for the highly twinned $P2_1$ structure 5ebi (Gilski *et al.*, 2016) and, for comparison, two nontwinned structures, PDB entries 3e9w (Venkadesh *et al.*, 2009) in $P2_1$ and 3p4j (Brzezinski *et al.*, 2011) in $P2_12_12_1$.

The orthorhombic structure 3p4j is the only one in Table 2 that cannot be merohedrally twinned, and indeed all twinning criteria are very close to the values expected for an untwinned crystal. The structure 3e9w shows intensity statistics that are not very different from the ideal values for untwinned crystals, but some other criteria differ significantly from the expected values. The R_{merge} value in point group 622 for these data is 0.065 and the similarity of intensities of potentially sixfold axis-related reflections is evident from the very low average $\langle |H| \rangle$ and $\langle H^2 \rangle$ values. This structure was, however, successfully refined without invoking twinning (Venkadesh *et al.*, 2009).

Diffraction data for the five monoclinic structures of Z-DNA with multiple molecules in the asymmetric unit show unusual and confusing results from the various twinning tests (Table 2). The average ratios of the moments of intensities and

amplitudes, as well as statistics of the normalized structure factors, have values that exceed the limits for untwinned crystals, especially for the structure Z4B. This syndrome is characteristic of the presence of translational noncrystallographic symmetry (tNCS), where two classes of reflections have either distinctly low or high intensities, as evident in the values of the average squares of normalized structure factors $\langle E^2 \rangle$ for the appropriate parities of indices. This is corroborated by the presence of significant non-origin peaks in the native Patterson map. In the structure Z4B these peaks are very high, in keeping with the especially unusual intensity statistics. Indeed, inspection of the molecular packing of these structures (Figs. 2*d–h*) confirms the presence of pairs of parallel duplexes related by the observed Patterson vectors (see §3.5).

A useful criterion for the characterization of twinned crystals is the behavior of the R_{merge} values calculated for pairs of

reflections related by individual twinning operations on the originally measured I_{obs} values and the $I_{\text{calc}} (= F_{\text{calc}}^2)$ values obtained after refining the structure, known as the RvR analysis (Lebedev *et al.*, 2006). The results of this analysis are included in Table 2 and illustrated in Fig. 3. As is evident after comparison of the scatter plot in Fig. 3 with Fig. 1 of Lebedev *et al.* (2006), the RvR behavior confirms the partial twinning of all four investigated crystal structures, especially with the influence of the rotational pseudosymmetry (RPS).

For pseudohexagonal $P2_1$ structures there are four independent R_{merge} values, three based on three twofold axes perpendicular to the crystallographic 2_1 direction and one on a threefold axis parallel to this direction. The reflection pairs resulting from the remaining three possible twofolds are equivalent to the pairs resulting from the previous three twofolds (perpendicular and in the same plane) by the action

Figure 2

Schematic representations of packing Z-DNA hexamer duplexes in various pseudohexagonal crystal structures with the orientations and vertical heights of their local, noncrystallographic twofold axes marked in blue in the following structures: (a) 3e9w, (b) 3p4j, (c) 1i0t, (d) Z3A, (e) Z3B, (f) Z4A, (g) Z4B, (h) Sebi.

Figure 2 (continued)

of the dyad (along [010]) present in the reciprocal lattice. Similarly, the reflection pairs related by the sixfold rotations give the same R_{merge} values as those related by the threefold rotations.

For the structures Z3A and Z3B, the R_{obs} values for all operators are similar, in the range 0.037–0.051, with the R_{calc} values in the range 0.32–0.37, falling in the region of ‘twin with RPS’ in Fig. 1 of Lebedev *et al.* (2006). The low values of R_{obs} in these two structures result from similar twinning fractions of all six twin domains, also resulting in the low overall R_{merge} values in 622 symmetry (0.093 and 0.083; Table 2). The refined model corresponds to only one, principal twin domain without the ‘horizontal’ twofold axes, and consequently the R_{calc} values based on calculated structure factors are much larger. However, the presence of some approximate RPS causes these values to be lower than the value of ~ 0.8 usually obtained for pairs of symmetry-independent intensities.

The RvR behavior for Z4A and Z4B is different. For most symmetry operators the RvR points fall into the ‘partial twin’ category of Lebedev *et al.* (2006), with an R_{calc} larger than 0.45 (up to 0.7). However, for Z4A the $R_{\text{obs}}/R_{\text{calc}}$ point at 0.248/0.253 for the $(h + l, -k, -l)$ symmetry has a lower R_{calc} value as a result of this dyad being one of the twofold axes of the true orthorhombic 222 point group of this crystal. For Z4B the

$R_{\text{obs}}/R_{\text{calc}}$ point at 0.031/0.307 corresponds to $(l, -k, h)$ symmetry and has an unusually low R_{obs} value because this dyad connects two major twin domains with comparable volume fractions and two pairs of minor domains also with comparable twin fractions. This ‘artificially’ introduces the twofold correspondence of intensities for reflections related

Figure 3

RvR analysis of the four Z-DNA structures, specified by a different color for each crystal. The R_{merge} values are obtained from the observed (R_{obs}) and calculated (R_{calc}) reflection intensities based only on reflection pairs related by the individual twinning operations, three twofold axes marked by dots and one threefold axis marked by crosses. See the text for a detailed explanation.

Table 3
Representative pseudohexagonally packed crystal structures of Z-DNA deposited in the PDB.

PDB code	Space group	Resolution (Å)	Unit-cell parameters				Reference
			<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	
3p4j	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁ A	0.55	17.88	31.42	43.90		Brzezinski <i>et al.</i> (2011)
1i0t	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁ B	0.60	18.32	30.68	42.49		Tereshko <i>et al.</i> (2001)
3uln	<i>P</i> ₃ ₂	2.72	17.82		43.01		Mandal, Chandrasekaran <i>et al.</i> (2012)
1dn8†	<i>P</i> ₆ ₅	1.50	17.93		43.41		Brennant <i>et al.</i> , 1986)
1d53†	<i>P</i> ₆ ₅	1.50	31.00		43.70		Kumar <i>et al.</i> (1992)
4dy8	<i>P</i> ₆ ₅	1.76	35.21		44.45		Mandal, Venkadesh <i>et al.</i> (2012)
279d†	<i>P</i> ₆ ₅ 22	1.90	18.10		43.10		Ban <i>et al.</i> (1996)
3e9w	<i>P</i> ₂ ₁	2.05	17.85	43.44	17.85	119.87	Venkadesh <i>et al.</i> (2009)
4ocb†	<i>I</i> 2‡	0.75	31.22	19.55	44.46	102.30	Luo <i>et al.</i> (2014)
6aqt (Z3A)	<i>P</i> ₂ ₁	1.05	30.92	42.89	30.95	119.98	This work
6aqv (Z3B)	<i>P</i> ₂ ₁	1.30	30.85	42.57	30.84	120.05	This work
6aqw (Z4A)	<i>P</i> ₂ ₁	1.30	35.43	44.00	35.43	119.83	This work
Z4A_ortho	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	1.55	17.72	30.68	44.00		This work
6aqx (Z4B)	<i>P</i> ₂ ₁	1.30	35.20	42.22	35.21	119.83	This work
5ebi	<i>P</i> ₂ ₁	1.09	35.72	44.10	35.72	119.98	Gilski <i>et al.</i> (2016)

† The 1dn8 and 4ocb structures are dodecamer duplexes, 279d is a decamer duplex and 1d53 is a octamer duplex; all others are hexamer duplexes. ‡ The cell is here presented in *I*2 symmetry to underline its relation to other structures. In Luo *et al.* (2014) the crystal is described in the standard C2 space group.

by this dyad, despite the lack of such symmetry in the real Z4B structure.

3.2. Variations in the packing of Z-DNA duplexes

In almost all crystal structures the hexamer duplexes of Z-DNA are packed in a pseudo-hexagonal arrangement. The duplexes follow each other along the 2₁ screw axes, forming infinite helical tubes with equally spaced CG base pairs, in which every sixth phosphate group is missing in the backbone between each pair of consecutive duplexes. These tubes are arranged in a parallel fashion, and in cross-section correspond to the dense hexagonal packing of circles. Preserving this general arrangement, various crystal forms of Z-DNA differ in the mutual rotational orientation and relative translation of the neighboring tubes, resulting in different unit-cell and space-group symmetries of the crystals. Only a few Z-DNA structures reported to date display a much more complex packing mode with helical tubes running along two different directions and with consecutive hexameric duplexes tilted with respect to the column axis; for example, the structures of the d(pCGCGCG)₂ hexamers with phosphorylated 5' ends (PDB entries 390d and 391d; Malinina *et al.*, 1998) and the structure of a racemic mixture of D and L Z-DNA duplexes (PDB entry 5jzq; Drozdal *et al.*, 2016).

The relative freedom in the mutual rotational and translational disposition of neighboring duplexes is a consequence of the relatively weak and nonspecific interactions between them. The outer surface of each duplex, formed by the DNA backbone, consisting of deoxyriboses and phosphate groups, contains polar O atoms that may serve as acceptors but not as donors of hydrogen bonds, except for the two free hydroxyl groups at the 3' and 5' ends of each hexamer. The potentially suitable N2 amino group of Gua is inside the large minor groove, and the N4 amino group of Cyt, located at the outer side of the duplex, is too distant from the adjacent duplexes to participate in direct hydrogen bonds. The interactions

between adjacent duplexes are therefore only executed indirectly through the mediating water molecules or ions present in the crystallization medium, such as various metals, ammonium ions or polyamines.

It is also characteristic of the d(CGCGCG)₂ duplex that it has an internal twofold axis of symmetry perpendicular to the helix at the middle of its length. In the crystals, such twofold axes of symmetry are present in each helical tube at the center of each duplex and in each gap between them. These dyads may form either the crystallographic symmetry elements of the crystal space group, or act only within one helical tube as a noncrystallographic symmetry element. In fact, in all hitherto known crystal forms of d(CGCGCG)₂ they are always of noncrystallographic character.

To form a periodic crystal, the arrangement of the tubes has to exhibit the three-dimensional translational symmetry, but the identical structural motifs (asymmetric units) may consist of more than one duplex. Because a single unit of two Watson–Crick pairs (CG)₂ in Z-DNA is followed along the left-handed helix by another such unit rotated by exactly 60°, three such units forming the hexamer duplex correspond to a total rotation of 180°. The translational repeat along the helical tubes is therefore always equal to the length of two duplexes related by a 2₁ screw axis, corresponding to 12 stacked base pairs (about 44 Å), and indeed almost all crystal forms of d(CGCGCG)₂ have such a cell dimension along this direction. The repeat of duplexes in directions perpendicular to the helices depends on their mutual disposition, but the crystal unit-cell parameters in this plane form a base with an area close to a multiple of a single rhombus of about 280 Å².

A number of Z-DNA crystal structures are interpreted in space groups *P*₃₂, *P*₆₅ or *P*₆₅22 but with duplexes that are disordered in their exact positioning along the helix axis and a cell edge of about 44 Å (Table 3). A linear arrangement of duplexes along the hexagonal or trigonal screw axes is not compatible with the symmetry of the tubes of duplexes, characterized by a lack of phosphate groups along the

'pseudo-infinite' helix between two successive duplexes. This requirement is only fulfilled if the consecutive hexamer duplexes are related by a 2_1 axis. Apart from these, a small number of Z-DNA hexamer duplex structures crystallize in different, unrelated cells and symmetries.

The simplest crystal structure of $d(\text{CGCGCG})_2$, built from helical tubes that are parallel to each other and with the unit cell containing two duplexes related by a 2_1 axis of symmetry, is PDB entry 3e9w (Venkadesh *et al.*, 2009). This crystal form is monoclinic, space group $P2_1$, with cell dimensions corresponding to the hexagonal lattice metric, but appears not to be twinned (Table 3, Fig. 2a).

Most $d(\text{CGCGCG})_2$ structures found in the PDB were crystallized in the orthorhombic space group $P2_12_12_1$, with one duplex in the asymmetric unit. The ratio of cell parameters b/a is about $31/18 \approx 3^{1/2}$, corresponding to a pseudohexagonal arrangement of the helices. Two versions of the orthorhombic packing have almost identical cell dimensions, but differ in the orientation and translation of the duplexes with respect to the

unit-cell edges and crystallographic twofold axes (Table 3, Figs. 2b and 2c).

The four crystal structures of $d(\text{CGCGCG})_2$ described here satisfactorily refine in the monoclinic space group $P2_1$, with three or four duplexes in the asymmetric parts of the unit cells characterized by a hexagonal lattice metric (Table 3). All four crystals are highly pseudomerohedrally twinned, since the content of the monoclinic unit cell can be accommodated in the hexagonal lattice in several (up to six) different orientations of the potential twin domains. Schematic representations of the molecular packing of these four structures are presented in Figs. 2(d)–2(g). The structure of the Z-DNA–RNA chimera (PDB entry 5ebi; Gilski *et al.*, 2016; Fig. 2h) is also hectohedrally twinned in an analogous cell and symmetry as Z4A and Z4B, but with the duplexes oriented differently. The structure Z4A can be interpreted as threefold twinned in the orthorhombic space group $P2_12_12_1$ (see below).

3.3. Orthorhombic representation of Z4A

The pseudohexagonal structures of Z-DNA hexamer duplexes could theoretically be interpreted in the orthorhombic space group $C222_1$ in three orientations differing by 120° . This was attempted by the authors of PDB entry 5ebi (Gilski *et al.*, 2016), but it was eventually concluded that the orientation of duplexes in this structure does not support the

Figure 4
(a) Packing of Z4A duplexes as in Fig. 2(f), but marked with the orthorhombic unit cell and symmetry operations of space group $P2_12_12_1$ in green and purple colors. (b) Packing of 5ebi duplexes as in Fig. 2(h), but marked with the pseudo-orthorhombic unit cell and symmetry operations of space group $P2_1$ in green and purple colors. (c) The reciprocal lattice of the Z4A crystal with three reciprocal orthorhombic unit cells of the three twinned domains represented in different colors. Reflections belonging simultaneously to all three domains have contributions from all colors. This figure also illustrates the diffraction pattern of the 5ebi crystal, with each color representing pairs of superposed reflections resulting from two twin domains related by one of the three twofold axes in the plane of the figure.

presence of crystallographic twofold axes perpendicular to the axes of the pseudo-helices. Indeed, in the structure 5ebi the directions of the local dyads in the individual duplexes are rotated from the cell edge by -21° and $+39^\circ$ (Fig. 2*h*), which is not in agreement with any potential direction of the theoretically possible orthorhombic twofold axes.

However, inspection of Fig. 2(*f*) shows that the local dyads in the structure Z4A are rotated from the direction of the monoclinic cell edge *c* by $+25^\circ$ in two of the individual duplexes and by -25° in the other two. Moreover, the centers of these two classes of inversely rotated duplexes lie at equal distances below and above the middle of the long unit-cell parameter, at the level of 0.5 ± 0.13 . This arrangement agrees with space group $P2_12_12_1$, as illustrated in Fig. 4(*a*). Since this structure can be presented as hectohedrally twinned monoclinic in a pseudo-hexagonal lattice, the new interpretation can therefore be expressed as a tritohedrally twinned (also called ‘drilling’) orthorhombic structure.

Interpretation of such twinning in the reciprocal lattice is somewhat complicated and is presented in Fig. 4(*c*). Since the volume of the orthorhombic unit cell is half that of the monoclinic cell, the corresponding reciprocal cell is twice as large, and each of the three orthorhombic reciprocal lattices related by a threefold axis (the twinning operator) contains only half of all the observed reflections corresponding to the pseudo-hexagonal lattice. In Fig. 4(*c*) reflections belonging to each of the three orthorhombic lattices are presented in three different colors: green, red and blue. It is evident that a quarter of the reflections belong only to the green twin domain, another quarter only to the red domain and another quarter only to the blue domain, while the last quarter of all observable reflections contain contributions from all three domains at once.

The relation between indices of reflections in the monoclinic lattice (*h*, *k*, *l*) and in the three orthorhombic lattices (with H^g , K^g and L^g for ‘green’ and analogously for ‘red’ and ‘blue’ reflections, where the monoclinic b^* axis becomes the orthorhombic c^* axis), are

$$\begin{aligned} H^g &= h/2 & K^g &= h/2 + l & L^g &= k, \\ H^r &= l/2 & K^r &= -h - l/2 & L^r &= k, \\ H^b &= -h/2 - l/2 & K^b &= h/2 - l/2 & L^b &= k. \end{aligned}$$

The conditions for ‘green’ reflections in the monoclinic lattice are $h = 2n$ and $l = 2n - 1$, those for ‘red’ reflections are $h = 2n - 1$ and $l = 2n$, and those for ‘blue’ reflections are $h + l = 2n$. The ‘mixed’ reflections have monoclinic indices $h = 2n$ and $l = 2n$, but in orthorhombic lattices they have $H + K = 2n$.

Table 4
Intensities of reflections for the Z4A crystal corresponding to the axial directions in the three orthorhombic twin domains.

Their indices in the monoclinic and orthorhombic lattices are also given in the first rows. See the text for an explanation of the index symbols.

Monoclinic	$2h, 0, -h$	$0, 0, h$	$h, 0, -2h$	$h, 0, 0$	$h, 0, h$	$h, 0, -h$	$0, k, 0$
Orthorhombic	$H^g, 0, 0$	$0, K^g, 0$	$H^r, 0, 0$	$0, K^r, 0$	$H^b, 0, 0$	$0, K^b, 0$	$0, L, 0$
1	16	12	8	2	3	5	—
2	21148	29128	22333	5308	3761	3155	121
3	6	6	9	11	4	10	1
4	3383	8089	4179	13769	16818	12732	2781
5	43	44	39	4	9	8	−1
6	6903	4616	6456	26725	23624	23322	997
7	7	7	14	1	5	6	5
8	2147	2350	1590	16311	6529	12297	6537
9	9	10	13	16	13	13	1
10		55	46	1001	1036	1407	862
11		−8	6	15	17	13	—
12		504	333	9507	10470	8960	—
13		3	3	8	17	11	—
14				12091	14100	12632	292
15				8	7	6	1
16				431	464	363	145
17				1	1	−3	3
18				198	255	357	138
19				1	4	7	3
20				12	19	179	1087
21				8	9	−1	2
22				4	−1	70	34
23				−4	−1	−1	3
24							10294

The presence of reflections originating from three individual twin domains (of purely one color) provides an opportunity to estimate the twin fractions by comparison of the intensities of reflections acquiring the same indices after reindexing to three orthorhombic lattices. Scaling of re-indexed sets of reflections leads to the following twin fractions: $\alpha^g:\alpha^r:\alpha^b = 0.35:0.10:0.55$. Merging of these three data sets gives an R_{merge} value of 0.163.

Interpretation of the Z4A crystal in space group $P2_12_12_1$ is strongly supported by very clear systematic absences among the axial reflections in the three orientations of the orthorhombic cell (Table 4), clearly identifying the presence of three perpendicular twofold screw axes in each of the orthorhombic crystal twin domains.

A crystal structure with such a special case of twinning, in which a subgroup of reflections originates from individual twin domains and another subgroup contains intensity contributions from all domains, can only be refined by the program *SHELXL* with the ‘HKLF 5’ option, with individual reflections in the input file separately marked as belonging (and appropriately indexed) to appropriate orthorhombic lattices. The following extract from the *SHELXL* reflection file—contains the first three symmetry-equivalent reflections from three different individual domains (green, red and blue with $M = 1, 2$ and 3 , respectively), and in the next three consecutive lines the same (‘multicolor’) reflection indexed in three different orthorhombic domains and marked as two domains (blue, $M = -3$; red, $M = -2$) sharing their intensity with the principal domain (green, $M = 1$).

H	K	L	I	$\sigma(I)$	M
0	-3	1	17452.1	939.9	1
0	3	1	4541.6	188.1	2
0	3	1	12142.1	91.4	3
-1	1	1	1356.6	87.4	-3
1	1	1	1356.6	87.4	-2
1	-2	1	1356.6	87.4	1

After appropriate reorientation of the duplex, the tritohedrally twinned $P2_12_12_1$ model of Z4A was refined with *SHELXL* to an *R* factor slightly lower than that obtained for this structure expressed as a hectohedrally twinned $P2_1$ case (Table 1). However, the orthorhombic model is simpler, as it contains only one crystallographically independent duplex in the asymmetric unit of the $P2_12_12_1$ cell.

3.4. Pseudosymmetry of the Z4B structure

Inspection of Figs. 2(g) and 2(h) reveals that the Z4B and PDB entry 5ebi structures contain two pairs of duplexes related by a half-cell translation along the *c* edge of the monoclinic unit cell. We translated two duplexes in each of these structures by exactly *c*/2, overlapped them on the remaining two molecules and calculated the r.m.s. deviations between pairs of corresponding atoms (the disordered phosphate groups were omitted). For the 5ebi structure this value was 0.03 Å and for Z4B it was 2.07 Å. Taking into account that the uncertainties of bond lengths, obtained from inversion of the least-squares full-matrix in the *SHELXL* refinement of the Z4B model, are in the range 0.02–0.03 Å, it may be concluded that in the Z4B unit cell the Z-DNA duplexes are packed in a pseudosymmetric fashion, with a high degree of translational NCS, but this structure cannot be faithfully expressed in a smaller unit cell.

However, in the 5ebi structure the analogous *c*/2 translation is much more exact within the expected accuracy of atomic positions. It can be therefore expressed in simpler terms, with two independent duplexes in the twice smaller $P2_1$ cell with all cell angles right, as presented in Fig. 4(b). The reciprocal lattice of the 5ebi crystal can be interpreted analogously to that of the Z4A crystal in Fig. 4(c), but with each color representing a pair of reflections superposed by the action of a different twofold rotation on a pair of twin domains in this hectohedrally twinned crystal.

3.5. Analysis of the native Patterson peaks

All analyzed Z-DNA structures show distinct non-origin peaks in the native Patterson maps, calculated here with *XPREP* (Sheldrick, 2015). The two structures Z3A and Z3B show significant peaks at positions (1/3, 0, 2/3). Their heights are 0.35 and 0.24, respectively, of the origin peak. They correspond to the translational relation between two of the three symmetrically independent duplexes that are parallel

and positioned at 1/3 and 2/3 of the long diagonal of the unit cell *ac* base (Figs. 2d and 2e). The heights of the Patterson peaks are close to the expected 1/3 of the origin, since in all six twin domains vectors between this 'internal' pair of duplexes contribute to the same Patterson peaks.

Analysis of the Patterson peaks in structures with four independent molecules is more revealing. All three structures, Z4A, Z4B and PDB entry 5ebi, if presented in a pseudo-hexagonal $P2_1$ cell, are characterized by the presence of either crystallographic (as seems to be the case in Z4A) or non-crystallographic translation of one pair of duplexes from the remaining pair by one half of the *c* cell edge. This, of course, manifests itself in the presence of characteristic non-origin peaks in the native Patterson map in the position (*u*, *v*, *w*) = (0, 0, 1/2). Since the Patterson map calculated with data from a twinned crystal corresponds to the superposition of Patterson functions originating from individual twin domains, it contains peaks corresponding to all different orientations of the contributing domains, but with their heights weighted according to their twin fractions. Indeed, inspection of Table 2 shows a very good correlation between the heights of the Patterson peaks at three positions, (0, 0, 1/2), (1/2, 0, 0), (1/2, 0, 1/2), and the total twin fractions of domains contributing to each of these peaks.

The situation with the Z4B structure is somewhat more complicated, and here there is not the clear correlation observed for the two previous structures. This is most probably a result of the presence of additional pseudosymmetric relations between Z-DNA duplexes in this structure. Inspection of Fig. 2(g) shows that the two pairs of highly parallel duplexes are positioned in the cell at almost the same *z* level and their orientation differs by 10°. The additional Patterson vectors, rotated by 120 and 240°, contribute to the total heights of the observed peaks, increasing these peaks that in principle originate from very small twin domains.

3.6. *L*-tests of Padilla and Yeates

The *L*-test, introduced by Padilla & Yeates (2003) and usually presented as an *L*-plot, is based on comparison of the intensities of pairs of reflections that are close in reciprocal space but not related by any twin operation. It is especially useful and robust in cases characterized by tNCS, where regular subsets of reflections have different average intensities. Since it relates reflections in pairs, the *L*-test works well for hemihedral twins, where the twinning operations are the twofold axes, but may not provide reliable results in cases of a higher degree of twinning. In our cases all reflections indexed in a pseudo-hexagonal monoclinic cell can be classified into four groups of different 'colors', as explained in the previous sections and illustrated in Fig. 4(c). All reflections for each structure may potentially be affected by six twinning operations, but the single-color reflections may only be affected by one of the three twofold axes perpendicular to the principal pseudo-hexagonal axis.

The *L*-plots based on various sets and subsets of data for all of the investigated structures are collected in the Supporting

Table 5

Conformation of the phosphate groups in the backbone of the investigated Z-DNA hexamers. ZI/ZII means that both conformers are present at half occupancy.

The structure 3p4j is included for comparison.

	C1/G2	G2/C3	C3/G4	G4/C5	C5/G6	C7/G8	G8/C9	C9/G10	G10/C11	C11/G12
Z3A-AB	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z3A-CD	ZI	ZI	ZI	ZII	ZI	ZI	ZI/ZII	ZI	ZI	ZI
Z3A-EF	ZI	ZI	ZI	ZII	ZI	ZI	ZI/ZII	ZI	ZI	ZI
Z3B-AB	ZI	ZI	ZI	ZI	ZI	ZI	ZII	ZI	ZI	ZI
Z3B-CD	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z3B-EF	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z4A-AB	ZI	ZII	ZI	ZI/ZII	ZI	ZI	ZI/ZII	ZI	ZI	ZI
Z4A-CD	ZI	ZII	ZI	ZI/ZII	ZI	ZI	ZI/ZII	ZI	ZI	ZI
Z4A-EF	ZI	ZI/ZII	ZI	ZI	ZI	ZI	ZI/ZII	ZI	ZI/ZII	ZI
Z4A-GH	ZI	ZI/ZII	ZI	ZI	ZI	ZI	ZI/ZII	ZI	ZI/ZII	ZI
Z4A_orth-AB	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z4B-AB	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z4B-CD	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z4B-EF	ZI	ZII	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI
Z4B-GH	ZI	ZI	ZI	ZII	ZI	ZI	ZII	ZI	ZI	ZI
3p4j-AB	ZI	ZI	ZI	ZII	ZI	ZI	ZI	ZI	ZI	ZI

Information. Inspection of these plots leads to somewhat confusing conclusions, which are often in disagreement with previous findings.

None of the *L*-plots for Z3A and Z3B show any significant indication of twinning, and the plots are close to the straight line characteristic of nontwinned crystals. However, the plot based on F_{calc} values, calculated from the coordinates refined by *REFMAC*, displays a significant departure from linearity opposite to those characteristic of twinning.

The *L*-plots for the Z4A data indicate different degrees of twinning. Among the three 'colored' subsets, the green reflections suggest almost perfect twinning, the red data suggest partial twinning and the blue data do not show any significant amount of twinning. Here, the discrepancy with the previous interpretation is most striking. If the Z4A structure is orthorhombic, as is suggested by the arrangement of duplexes in the cell and previous analyses, the twofold axes in the *ac* plane are crystallographic, and thus the monocolored reflections originate from nontwinned domains. The *L*-plots based on all or multicolor reflections show partial twinning characteristics.

The *L*-plots for the Z4B structure are similar, but here the reflections of two colors, green and red, suggest perfect twinning and the blue reflections show only a very low degree of twinning. All data give an *L*-plot showing partial twinning but, surprisingly, the multicolor reflections do not show any amount of twinning. This supports the expectation that *L*-plots are not suitable as twinning criteria for multiple twinned crystals. The *L*-plots based on calculated intensities for the Z4A and Z4B structures are not as abnormal as for the Z3A and Z3B structures, although they still deviate from ideality.

Various *L*-plots for the 5ebi data show different amounts of partial twinning but, surprisingly, the multicolor reflections suggest the highest, almost perfect twinning.

3.7. Details of the structures

Although all crystals were obtained from solutions containing either Ca^{2+} or Mg^{2+} ions, it was not possible to

convincingly identify any ions in these structures, and all solvent molecules are represented as water O atoms. There was also no electron density in the structure of Z4B that could be interpreted as a molecule of spermine, in spite of its presence in the crystallization medium.

A large number of solvent water sites were modeled in all four structures, but their arrangement in the minor grooves of Z-DNA and in the inter-helical space is not as regular as in the other, atomic resolution crystal structures of Z-DNA. All water sites are treated as fully occupied but, similarly to even the highest resolution structure PDB entry 3p4j, modeling of the solvent region is not fully comprehensive. There are still some peaks in the final difference map that could be tentatively filled with water molecules, but would not be acceptable by the objective validation criteria. The high degree of twinning may exacerbate the difficulties in modeling the solvent sites in these four crystal structures.

It was observed early on that the phosphodiester groups in the Z-DNA backbone occur in two conformations (Wang *et al.*, 1981; Gessner *et al.*, 1989; Schneider *et al.*, 1997). The more frequent one, designated ZI, has the phosphate group located closer to the inside of the helix, towards the minor groove, whereas in the ZII conformation this group is shifted towards the outside of the helix. The most indicative in this respect is the torsion angle ζ ($\text{C3}'-\text{O3}'-\text{P}-\text{O5}'$), which has a $+sc$ ($+synclinal$, about $+60^\circ$) value for the CG step and a $-sc$ (about -60°) value for the GC step for the ZI conformation, and inverted values for the ZII conformation (Schneider *et al.*, 1992). The ZII conformation most commonly occurs at the GC stage between the fourth and fifth nucleotides in one of the chains in a duplex, but it may also be present at the GC stage between the second and third nucleotides. In some hexamer duplexes the phosphate groups appear to be disordered in both conformations. In the current structures the conformations of phosphate moieties conform in general to these observations, with several molecules having double conformations of their backbone phosphates at both GC stages, as presented in Table 5.

4. Conclusions

The four structures described here expanded the palette of arrangements that Z-DNA hexamer duplexes adopt in their crystals. Owing to the specific shape of Z-DNA duplexes, all observed types of packing of Z-DNA molecules are characterized by a pseudohexagonal disposition of parallel infinite helical 'tubes', differing only in their relative mutual rotational orientation around the helical axis and by respective translation along this axis. Such a pseudosymmetric arrangement of parallel helices leads to the possibility for Z-DNA crystals to exhibit certain crystallographic pathologies, such as disorder or twinning. If the orientation of the individual molecules in neighboring unit cells differs in a random fashion, the electron density resulting from the diffraction experiment corresponds to the average, disordered structure. Twinning occurs if the arrangement of duplexes preserves strict order within macroscopic parts of the whole crystalline specimen, but in different parts (twin domains) the identically arranged molecules are reoriented (according to the twin laws), preserving the whole periodicity of the crystal lattice.

Various Z-DNA crystal structures display three situations. The orthorhombic structures (e.g. PDB entries 3p4j and 1i0t) represent perfect 'healthy' crystals. The structures presented in trigonal and hexagonal space groups (e.g. PDB entries 3uln, 1dn8 and 279d) are disordered. The four structures presented here are monoclinic, and can be successfully interpreted as highly pseudomerohedrally twinned, with multiple domains corresponding to various possible orientations of the monoclinic cell in the lattice with a hexagonal metric. However, the various twinning criteria provide different and conflicting judgments. The two criteria widely recognized as being most robust, the RvR test and *L*-plots, lead to opposite conclusions. This exacerbates difficulties in the unequivocal interpretation of highly complicated cases of twinning.

The hexagonal arrangement of parallel helices in the monoclinic Z-DNA crystals gives rise to the interesting phenomenon of unit-cell triangulation, with a multiple number of crystallographically independent molecules in the cell, in analogy to the known triangulation of protein molecules at the surface of capsids of icosahedral viruses. The structure 3e9w corresponds to the simplest triangulation with $T = 1$, and the current structures represent triangulations with $T = 3$ and $T = 4$.

The four crystal structures of d(CGCGCG)₂ described here add to the range of known structures of Z-DNA oligomers and display a complicated hectohedral twinning phenomena related to the pseudohexagonal packing of DNA helices.

Acknowledgements

We are grateful to the anonymous referee for pointing out the correlation between the Patterson peak heights and twin fractions. This work was supported in part by the Intramural Research Program of the National Cancer Institute, Center for Cancer Research. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science under Contract No. W-31-109-Eng-38.

Funding information

The following funding is acknowledged: National Science Centre, Poland (grant No. 2013/10/M/NZ1/00251).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Ban, C., Ramakrishnan, B. & Sundaralingam, M. (1996). *Biophys. J.* **71**, 1215–1221.
- 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.
- Brennan, R. G., Westhof, E. & Sundaralingam, M. (1986). *J. Biomol. Struct. Dyn.* **3**, 649–665.
- Brzezinski, K., Brzuszkiewicz, A., Dauter, M., Kubicki, M., Jaskolski, M. & Dauter, Z. (2011). *Nucleic Acids Res.* **39**, 6238–6248.
- Caspar, D. L. D. & Klug, A. (1962). *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24.
- Drozdal, P., Gilski, M. & Jaskolski, M. (2016). *Acta Cryst.* **D72**, 1203–1211.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. & Wang, A. H.-J. (1989). *J. Biol. Chem.* **264**, 7921–7935.
- Gilski, M., Drozdal, P., Kierzek, R. & Jaskolski, M. (2016). *Acta Cryst.* **D72**, 211–223.
- Ho, P. S. & Mooers, B. H. M. (1997). *Biopolymers*, **14**, 65–90.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Kumar, V. D., Harrison, R. W., Andrews, L. C. & Weber, I. T. (1992). *Biochemistry*, **31**, 1541–1550.
- Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* **D62**, 83–95.
- Luo, Z., Dauter, M. & Dauter, Z. (2014). *Acta Cryst.* **D70**, 1790–1800.
- Malinina, L., Tereshko, V., Ivanova, E., Subirana, J. A., Zarytova, V. & Nekrasov, Y. (1998). *Biophys. J.* **74**, 2482–2490.
- Mandal, P. K., Chandrasekaran, A. R., Madhanagopal, B. R., Venkadesh, S. & Gautham, N. (2012). *J. Cryst. Growth*, **354**, 20–26.
- Mandal, P. K., Venkadesh, S. & Gautham, N. (2008). *J. Indian Inst. Sci.* **88**, 73–93.
- Mandal, P. K., Venkadesh, S. & Gautham, N. (2012). *Acta Cryst.* **F68**, 1420–1426.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- 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.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* **D59**, 1124–1130.
- Schneider, B., Ginell, S. L., Jones, R., Gaffney, B. & Berman, H. (1992). *Biochemistry*, **31**, 9622–9628.
- Schneider, B., Neidle, S. & Berman, H. M. (1997). *Biopolymers*, **42**, 113–124.
- Sheldrick, G. M. (2015). *Acta Cryst.* **C71**, 3–8.
- Tereshko, V., Wilds, C. J., Minasov, G., Prakash, T. P., Maier, M. A., Howard, A., Wawrzak, Z., Manoharan, M. & Egli, M. (2001). *Nucleic Acids Res.* **29**, 1208–1215.
- Venkadesh, S., Mandal, P. K. & Gautham, N. (2009). *Acta Cryst.* **F65**, 8–13.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979). *Nature (London)*, **282**, 680–686.
- Wang, A. J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H. & Rich, A. (1981). *Science*, **211**, 171–176.
- Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.* **43**, contribution 7.