Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An A-DNA structure with two independent duplexes in the asymmetric unit

G. Savitha and M. A. Viswamitra*

Department of Physics, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: mav@physics.iisc.ernet.in

Received 25 July 1998
Accepted 22 February 1999

NDB Reference: d(G_4CGC_4), ad0002.

The crystal and molecular structure of the self-complementary A-DNA decamer sequence d(G_4CGC_4) was solved at 1.9 Å resolution. The decamer crystallizes in space group $P2_1$ with two independent duplexes in the asymmetric unit. Duplex 1 has interactions which are distributed symmetrically about its length compared with duplex 2. The two end base pairs of duplex 1 have a similar NH···O hydrogen-bond pattern involving GGC segments of duplex 2 and a symmetry-related neighbour, while the end base pairs of duplex 2 interact with the GCC and GGG segments of its symmetry-related neighbours through NH···O and NH···N hydrogen bonds and a water-mediated hydrogen bond between the carboxyl groups of C40 and C8. In addition to the C4′–C5′ torsion angle γ assuming the *trans* conformation in certain steps, this angle also adopts the *gauche*$^-$ conformation at C37 as opposed to the preferred *gauche*$^+$ conformation, with a concomitant change in phosphodiester P—O5′ (α) in the opposite sense. This facilitates stacking between adjacent bases. The study suggests that the structural alterations in the two molecules in the asymmetric unit originate from an inherent propensity of the d(G_4CGC_4) base sequence for varied intermolecular interactions and malleability.

1. Introduction

Since the first crystal structure study of the deoxy-oligonucleotide dp A–T–A–T (Viswamitra *et al.*, 1978), X-ray crystal structure analyses have established the plasticity of DNA molecules in crystals, whether it be a consequence of the base sequence of the DNA or the intermolecular packing forces in the crystal. Local base sequences are recognized by proteins and drugs to the same extent as by other copies of the base sequence, and there can be different modes in which this takes place. The crystal structure of the d(G_4CGC_4) decamer reported here is an A-DNA and is unique, crystallizing in space group $P2_1$, which is uncommon for A-DNA decamers, and having two molecules in the asymmetric unit, the two having different modes of interaction and conformational diversity. This allows comparison of the same duplex in different packing environments within the same crystal.

2. Materials and methods

2.1. Synthesis and crystallization

The sample was synthesized using solid-support phosphoramidite chemistry on an Applied Biosystems 380B synthesizer, purified by reverse-phase chromatography using a PepRPC HR10/10 column and desalted using a Fast Desalting column on the FPLC. Single crystals of d(G_4CGC_4) were

grown at 291 K in sitting drops by the vapour-diffusion method. The drop was set up by sequential addition of 10 µl of 2 m*M* DNA, 20 m*M* sodium cacodylate buffer (pH 6.5), 30 m*M* $MgCl_2$, 2 m*M* spermine tetrachloride and 5% MPD. The drop was equilibrated against a reservoir solution containing 50% MPD. Rectangular crystals were observed within three weeks. When the concentration of the reservoir solution was reduced to 5% MPD, the crystals dissolved in 3–4 d, and upon increasing the concentration to 50% MPD larger crystals could be grown. This process of dissolution and re-crystallization was reversible and repeatable any number of times. The crystal used for data collection was approximately 0.25 × 0.32 × 0.4 mm in size.

2.2. Data collection

Intensity data were collected on a MAR Research imaging-plate system supported by a Rigaku rotating-anode X-ray source (Cu $K\alpha$; λ = 1.5418 Å) with a power rating of 40 kV, 58 mA. The crystal was mounted in a glass capillary with a droplet of mother liquor. The crystal-to-film distance was maintained at 120 mm with an exposure time of 320 s frame^{-1} and an oscillation angle of $\delta\varphi = 1^\circ$. The crystals were found to be stable in the X-ray beam for over 4 d. The diffraction data were processed and autoindexed using the procedure of Kabsch (1993) as implemented in the *XDS* package. The crystal is monoclinic, space group $P2_1$, with unit-cell parameters $a = 45.504$, $b = 46.084$, $c = 28.465$ Å and $\beta = 100.74^\circ$. The processed and reduced data have a cumulative completeness of 92.4% at 1.8 Å resolution, with 41278 measurements of 10033 unique reflections, an overall multiplicity of 4.11 and a merging R factor of 6.16%. Though the crystals diffract to 1.8 Å, the reflections around that shell were mostly weak and therefore data to only 1.9 Å were used for diffraction analysis, with 3956 reflections >2σ. Assuming an average density of $\rho = 1.5$ g cm^{-3} and a solvent content of 50%, which is often the case with most oligonucleotide crystals, there are four single-stranded d(G_4CGC_4) molecules in the asymmetric unit, or two d(G_4CGC_4) duplexes.

Surprisingly, $P2_1$ is an uncommon space group for oligonucleotide single crystals, in contrast to protein crystals where $P2_1$ is the most prevalent space group. Additionally, there are very few structures with two oligonucleotide duplexes in the asymmetric unit. In fact, this is the first structure observed with two A-type DNA duplexes in the asymmetric unit.

2.3. Structure determination

A self-rotation map calculated using the small *POLARRFN* (Kabsch, unpublished work) routine in the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) shows persistent non-origin peaks in the *xz* plane, as indicated by the values of the direction cosines in Table 1. The two non-origin peaks are disposed at 90° to each other and are staggered by approximately 13° from the crystallographic *c* axis and are close to the *a* axis. The 222 symmetry in the $\kappa = 180^\circ$ section shows that these two duplexes could be related by a non-crystallographic dyad perpendicular to the crystallographic dyad along *b*.

Table 1
Results of the self-rotation search at $\kappa = 180^\circ$.

Number	ψ	φ	κ	Direction cosines			Peak
1	90.0	90.0	180.0	0.0000	1.0000	0.0000	100.0
2	13.0	0.0	180.0	0.2250	0.0000	0.9744	58.0
3	77.0	180.0	180.0	−0.9744	0.0000	0.2250	58.0
4	103.0	0.0	180.0	0.9744	0.0000	−0.2250	58.0
5	167.0	180.0	180.0	−0.2250	0.0000	−0.9744	58.0
6	29.0	180.0	180.0	−0.4847	0.0000	0.8747	40.2
7	119.0	180.0	180.0	−0.8747	0.0000	−0.4847	40.2
8	61.0	0.0	180.0	0.8747	0.0000	0.4847	40.2
9	151.0	0.0	180.0	0.4847	0.0000	−0.8747	40.2

The molecular-replacement method gave the most promising solutions when the structure of d(CCGGC)r(G)d(CCGG) (Ban *et al.*, 1994) was used as a starting model. Cross-rotation and translation searches followed by rigid-body refinement for the first 50 peaks in the cross-rotation search were carried out using the automated package *AMoRe* (Navaza, 1994). The best R factor was 39.6%, with a correlation factor of 58.5%. When the sequence of d(CCGGC)r(G)d(CCGG) was mutated (using the graphics package *InsightII*) to d(G_4CGC_4), the R factor dropped further to 38.9%, with a correlation factor of 59.8%. The disposition of the two d(G_4CGC_4) duplexes in the cell was found to be in agreement with the direction cosines predicted by the self-rotation search.

Refinement was carried out using the package *X-PLOR* Version 3.851 (Brünger, 1992*b*). First, the rigid-body solution from *AMoRe* was subjected to torsion-angle refinement (Rice & Brünger, 1994) using the $R_{\rm free}$ option (Brünger, 1992*a*). $R_{\rm free}$ dropped from 40 to 29% while the R factor fell to 23%, in the 8–2.4 Å resolution range for $F > 2\sigma(F)$, 3418 reflections with no NCS (non-crystallographic symmetry).

The two molecules were found to be related by 182.2° along with a translational component. The resolution range and step-size limits used for the self-rotation search had made it appear that the two molecules were related by an exact dyad symmetry. Any attempt to use NCS restraints (Weis *et al.*, 1990; Braig *et al.*, 1995), which provide the advantage of doubling the reflection-to-parameter ratio, worsened $R_{\rm free}$. Therefore, the two molecules were refined independently. The $R_{\rm free}$ option was not used after this stage. The resolution was increased to 1.9 Å in steps of 0.1 Å and the structure was subjected to simulated-annealing molecular-dynamics (Brünger *et al.*, 1987, 1990) and group *B*-factor refinement, resulting in an R factor of 20.2%. Map inspection and water fitting were carried out using the graphics package *TOM* Version 2.9. After several rounds of positional and group *B*-factor refinement and water fitting, 116 water molecules could be picked with a final R factor of 17.4%. The r.m.s. deviations from the ideal values for bond lengths, bond angles, torsion angles and improper deviations are 0.005 Å, 1.56, 32.8 and 0.95°, respectively (dna-rna.param file available in *X-PLOR* Version 3.851, Parkinson *et al.*, 1996). The mean group temperature factors were close to 20 Å^2 for the bases,

Table 2
Average helical parameters.

Helical parameters	Duplex 1	Duplex 2	Model†	A-DNA fibre‡
Residues/turn	11.25	11.41	11.50	11.00
Helical twist (°)	31.3	31.6	31.8	32.7
Rise (Å)	2.77	2.64	2.51	2.56
Slide (Å)	−1.68	−1.71	−1.79	−1.50
Inclination (°)	11.8	14.6	16.8	20.1
x displacement (Å)	−4.37	−4.87	−5.18	−4.49
Roll (°)	4.0	4.5	6.3	10.9
Propeller (Å)	−5.3	−4.7	−7.4	11.7
Average backbone torsions (°)				
α	279.7	253.4	286.6	276.1
β	184.6	183.5	170.0	207.9
γ	60.3	89.6	65.8	45.5
δ	98.3	98.6	78.2	84.3
ε	196.0	196.9	204.5	179.5
ζ	283.4	280.6	287.0	311.0
χ	205.8	206.7	198.3	205.7

† Ban *et al.* (1994). ‡ Arnott & Selsing (1974).

and the largest values were found for phosphate groups (50 Å^2), especially those present at the ends of the helix. Duplex 1 and duplex 2 have an r.m.s.d. of 1.056 Å in atomic positions on least-squares superposition.

Analysis of the refinement results shows that a large number of the solvent peaks have temperature factors as low as those of the nucleic acid atoms, especially the solvent molecules which link anionic phosphate O atoms to anionic phosphate O atoms or to other exposed nucleic acid atoms such as O4′ of the sugar, N4 or O2 of cytosine, N2 or O6 of guanine or terminal O3′ or O5′ atoms. Water molecules also link different asymmetric units.

3. Results

3.1. Analysis of the structure

The overall helical parameters of the two molecules conform to the ideal A-DNA geometry, as seen in Table 2. The variations are described in the context of the packing. Residues are numbered 1–20 and 21–40 for the two duplexes in the asymmetric unit, respectively, and the helical parameters were calculated using *NEWHEL*93 (distributed by R. E. Dickerson) based on the global helix axis.

3.2. Packing

Each d(G_4CGC_4) decamer molecule in the asymmetric unit has strong direct interactions with four neighbouring molecules, as seen in Fig. 1(*a*). Three base pairs are spanned by the end base pair of the neighbouring molecules, as shown schematically in Fig. 2. These interactions involve hydrogen bonds and van der Waals contacts between base pairs, O3′ of the end sugar and O4′. Fig. 2 also shows that the base pairs involved in interactions with the neighbours are different in duplex 1 and duplex 2. The two end base pairs of duplex 1 have similar NH$\cdots$O hydrogen-bonding interactions which both involve the GGC segment (Figs. 3*a* and 3*c*), one belonging to the other molecule in the asymmetric unit and the second to its

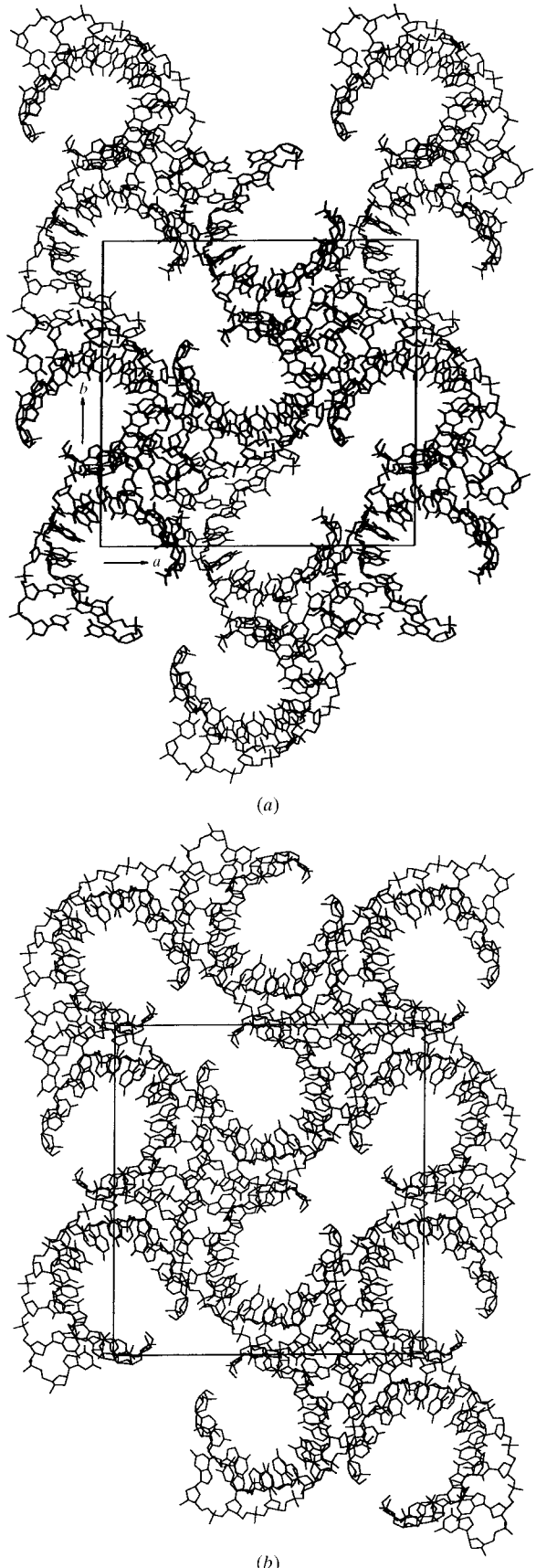

Figure 1
Packing diagrams looking down the *c* axis for (*a*) d(G_4CGC_4) in the $P2_1$ cell and (*b*) d(CCGGC)r(G)d(CCGG) (Ban *et al.*, 1994) in the $P2_12_12_1$ cell.

symmetry-related neighbour. On the other hand, the end base pairs of duplex 2 interact with the GCC and GGG segments of its symmetry-related neighbours, involving NH···O and NH···N hydrogen-bonding interactions (Figs. 3*b* and 3*d*) as well as a water-mediated interaction between the carboxyl groups of C40 and C8 at the G21·C40 end base pair. Consequently, duplex 1 has interactions distributed symmetrically about its length compared with duplex 2.

3.3. Conformational parameters

The backbone structures of both duplexes vary considerably (Table 3). The torsion angles of some residues in one strand differ considerably from their equivalents in the other strand. The average value of the angle δ, which is diagnostic of the A-form, corresponds to a sugar pucker in the C3′-*endo* region. The backbone torsion angles β, δ, ε, ζ and P have values which are clustered, while α and γ show strongly correlated jumps in their values. At residues G6, G13, G22, G33, G31, G36 and C38, the intermolecular contacts seem to shift the values of α and γ from their preferred *gauche*$^-$ and *gauche*$^+$ conformations, respectively, to the *trans* conformation. This feature has been observed in some other A-DNA structures (Haran *et al.*, 1987; Rabinovich *et al.*, 1988; Jain & Sundaralingam, 1989; Frederick *et al.*, 1989; Ban *et al.*, 1994; Ramakrishnan & Sundaralingam, 1993*a*,*b*). At C37, α shifts to *gauche*$^+$ with a concomitant shift in γ to *gauche*$^-$ while maintaining the stacking of the bases (the value marked * in Table 3). This crankshaft motion has been predicted from theoretical calculations (Yathindra & Sundaralingam, 1976; Malathi & Yathindra, 1985). Fig. 4 is a superposition of the backbones of the two duplexes. The GG steps show the usual stacking of adjacent guanine residues on the same strand with very little stacking of the cytosine residues, except at the G23·C38::G24·C37 stack of duplex 2, where G24 has poor overlap owing to an interaction involving hydrogen bonding of N3 of the G24·C37 base pair to an amino group of the end

Table 3
Backbone torsion angles (°) of the two molecules in the duplex.

The sharp fluctuations seen in the values of α and γ are shown in bold.

Code	α	β	γ	δ	P	ε	ζ	χ
G1	—	—	58.5	100.4	17.6	194.9	280.7	194.1
G2	290.8	182.5	50.9	100.9	22.9	196.1	294.5	200.5
G3	291.5	186.1	46.0	100.3	32.4	189.9	276.2	205.1
G4	291.0	173.8	65.2	101.9	29.7	197.2	277.6	197.1
C5	293.1	196.6	34.5	99.2	15.7	196.1	292.1	220.6
G6	**142.3**	181.8	**182.9**	98.5	16.1	211.0	281.2	196.2
C7	297.7	176.0	45.7	98.3	20.1	198.8	289.6	213.4
C8	304.5	188.8	37.6	102.2	27.1	186.7	286.8	218.9
C9	299.0	191.2	37.8	102.1	31.2	180.2	283.9	212.5
C10	296.7	189.0	45.0	87.0	30.8	—	—	211.2
G11	—	—	70.0	101.6	18.8	187.2	286.8	201.3
G12	301.4	186.7	47.1	97.5	23.4	192.2	282.5	201.4
G13	**147.9**	190.4	**182.6**	98.3	16.2	216.6	277.3	178.5
G14	299.2	182.9	42.2	99.7	23.6	197.3	284.0	202.6
C15	300.5	190.2	33.1	100.9	20.8	187.7	284.8	203.7
G16	280.4	193.8	49.1	97.8	27.2	203.4	279.8	210.6
C17	300.5	170.8	49.6	99.5	26.7	196.6	280.1	213.9
C18	286.7	168.1	67.1	98.3	30.1	205.3	280.3	200.0
C19	300.9	191.3	25.4	100.7	27.6	191.5	283.3	213.3
C20	310.1	182.9	36.0	80.3	28.4	—	—	222.1
G21	—	—	62.3	98.4	19.3	200.5	283.0	190.1
G22	**165.3**	185.4	**164.3**	95.5	14.5	209.0	288.4	187.8
G23	301.2	183.7	47.5	99.3	25.6	183.0	280.6	200.9
G24	279.3	200.4	51.7	100.3	15.8	214.3	288.3	204.9
C25	311.8	182.8	20.9	101.5	30.2	180.7	279.9	230.3
G26	288.3	189.3	48.5	99.6	28.5	202.0	284.6	209.7
C27	311.6	175.1	29.6	103.0	30.2	183.8	280.7	229.0
C28	299.6	185.1	46.6	103.4	25.5	186.6	286.1	213.2
C29	294.0	196.7	36.5	102.5	23.2	189.7	290.1	208.2
C30	281.7	194.5	50.4	89.4	33.5	—	—	226.2
G31	—	—	**178.2**	101.0	17.0	193.2	287.8	193.9
G32	304.4	185.8	40.9	97.8	27.7	186.0	277.9	209.6
G33	**145.2**	198.9	**181.7**	97.8	19.7	219.3	273.9	176.1
G34	302.8	171.3	49.1	101.5	28.7	196.0	279.2	197.6
C35	296.3	191.9	33.3	99.7	19.1	200.5	286.4	206.4
G36	**151.2**	182.8	**177.6**	99.0	21.6	211.8	243.1	199.8
C37	**25.9***	135.3	**342.9**	101.1	22.6	179.1	286.9	231.6
C38	**162.6**	189.7	**164.1**	95.9	14.2	217.6	274.4	199.4
C39	320.5	170.9	38.4	100.8	29.7	191.5	280.1	205.8
C40	318.8	183.1	28.5	85.0	29.5	—	—	214.4

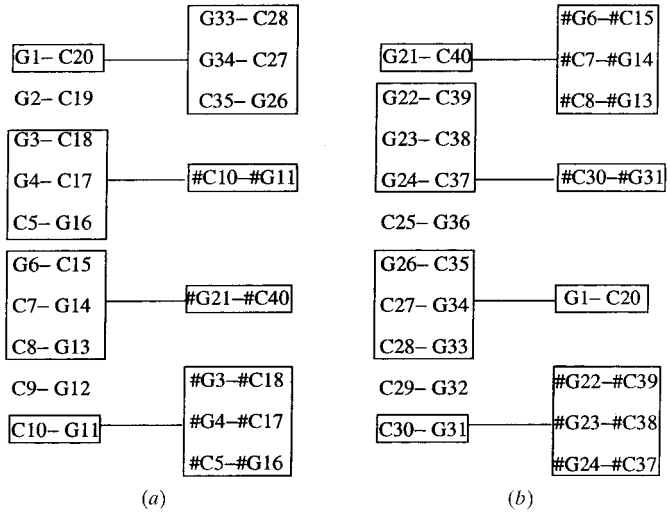

Figure 2
Schematic sketch showing the interaction of duplex 1 (*a*) and duplex 2 (*b*). The symmetry-related neighbours are represented by #. See text for details.

base pair of the symmetry-related molecule. This results in a large deviation in the torsion angle from normal in that region (Fig. 3*d*). The GC steps show a high degree of overlap between purine and pyrimidine bases of the same strand, while the CG steps show cross-strand interaction of the guanine base. Although the guanine base shifts into the minor groove at the CC steps, it still has substantial stacking across strands, with the cytosines showing little or no overlap.

The base-pair parameters, as seen in Fig. 5, show considerable local variation. In A-DNA, since the bases are inclined and placed away from the helix axis, the rise parameter measured at the C1′ atoms varies in a complex manner as a result of the local changes in orientation of the bases: twist, roll and buckle. Duplex 2 exhibits greater inclination of the base pairs than duplex 1, with the end base pairs least inclined and with a positive tip at the 5′-G·C end base pair and a negative tip at the C·G-3′ end base pair. The changes in the propeller, buckle, x displacement and y displacement parameters, though resembling the other A-form crystals, are not easy to interpret, except that the C30·G31 base pair shows a

high propeller twist value, presumably the result of the hydrogen-bonding interaction of the N3 of guanine with the amino group of a neighbouring molecule in the crystal.

Fig. 6 shows that duplex 2 has a large positive roll parameter of nearly 13° and a large negative cup value at the GC step, compensated by a low tilt value. The adjacent CG step, which also has a positive roll value of 13°, also has a large negative tilt value. Duplex 2 shows fluctuations in the cup parameter all along the entire length of the helix. The large positive roll value opens the angle between the base pairs G24·C37 and C25·G36 towards the minor groove, while compressing the major groove, in addition to the opening of the angle between base pairs C25·G36 and G26·C35 toward strand II (31–40 residue strand). The large negative tilt value is accompanied by severe distortion in the backbone torsions of G36, C37 and C38 residues, implying bending of the helix in that region.

It is apparent from Fig. 7 that the local twist parameter has a clear inverse correlation with the rise parameter of each base pair. The central CG step in both the molecules is characterized by a low helical twist value and a relatively low slide value, indicating considerable intra-strand stacking. The inverse correlation found between the twist and rise parameters indicates that the base pairs adjust their positions so that a large local twist value is compensated by a small rise

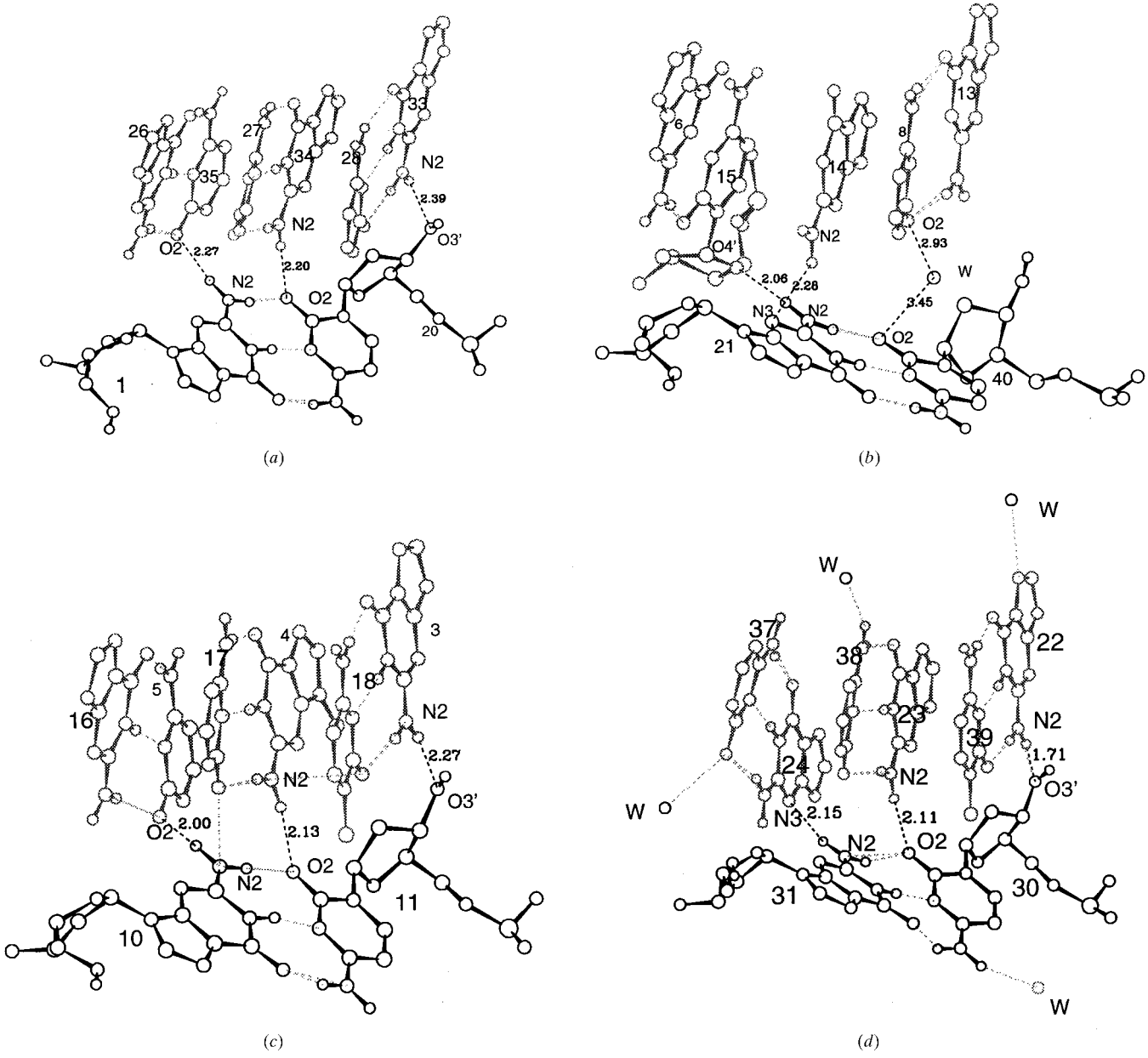

Figure 3
Interactions involving terminal base pairs. (*a*) G1·C20 of duplex 1, (*b*) G21·C40 of duplex 2, (*c*) C10·G11 of duplex 1 and (*d*) C30·G31 of duplex 2. Water molecules are indicated by W. Hydrogen-bond distances between H atoms located geometrically and acceptor O atoms are given. The distances involving water molecules are between O atoms.

value. As there are backbone constraints, the changes are reflected in the variation of the stacking between base pairs so that the helical structure is conserved.

The variation in the helical parameters shows that the sequence d(G_4CGC_4) has the potential to adopt a diverse range of conformations. The differences observed may in part be a consequence of the crystal environment, but it is to be noted that the two molecules could have arranged themselves perfectly with a twofold NCS in the $P2_12_12_1$ space group, where the helical parameters for the two units would have been exactly identical. Thus, it is likely that the sequence d(G_4CGC_4) has an intrinsic tendency to adopt the range of conformations observed.

3.4. Hydration

The geometries of the nucleic acid helices are such that water molecules are able to bridge hydration sites of the same residue, of adjacent residues on the same strand or of distant residues on the two strands. Intra-base water bridges are observed between N7 and O6 of guanines (*e.g.* G4, G6, G12, G23, G26, G34). Water bridges between O1P anionic O atoms and N7 of guanine (*e.g.* G26) or N4 of cytosine (*e.g.* C5) are sometimes composed of two water molecules in the d(G_4CGC_4) structure, which is a characteristic feature in the A-form of DNA. In general, inter- and intra-strand two-water bridges are a common feature in the major groove. Both in the A and B forms, no dependence on sequence is yet clearly apparent (Westhof *et al.*, 1985; Westhof, 1987). In the minor groove, water molecules bind to N3 of purines, O2 of pyrimidines and O4′ of sugar rings (*e.g.* G6, G9, C10, C20, C29, G33, C40). The water-mediated interaction seen at the G21·C40 base pair indicates that such water molecules can be used for recognition at specific sites (Otwinowski *et al.*, 1988; Rodgers & Harrison, 1993; Shimon & Harrison, 1993; Shakked *et al.*, 1994).

About half of the 116 water molecules located interact with bases, the remainder interacting with the phosphodiester backbone, and none of them appear in the same relative position in the two duplexes. A significant number of water molecules are located in the major groove, while the minor groove is mostly involved in interaction with neighbouring oligomers. Some of the phosphates, amino groups and the end base pair O3′ have more than one hydrogen-bond contact. The anionic phosphate O atoms are the most hydrated. They interact with 45 waters, followed by the carboxyl O atom O2 of cytosine, which interacts with 29 waters. Five water molecules were located which were associated with the sugar-ring O atom O4′, which is frequently involved in water binding through a water bridge with the O2 of the preceding pyrimidine (*e.g.* C9) or the N3 of a preceding purine (*e.g.* G23). The end O3′ O atom is relatively more hydrated, forming water bridges between neighbours (*e.g.* G11 and C30). The molecules are packed in the unit cell such that a network of channels is formed connecting the major grooves of adjacent molecules, in which the solvent molecules and counterions may be located.

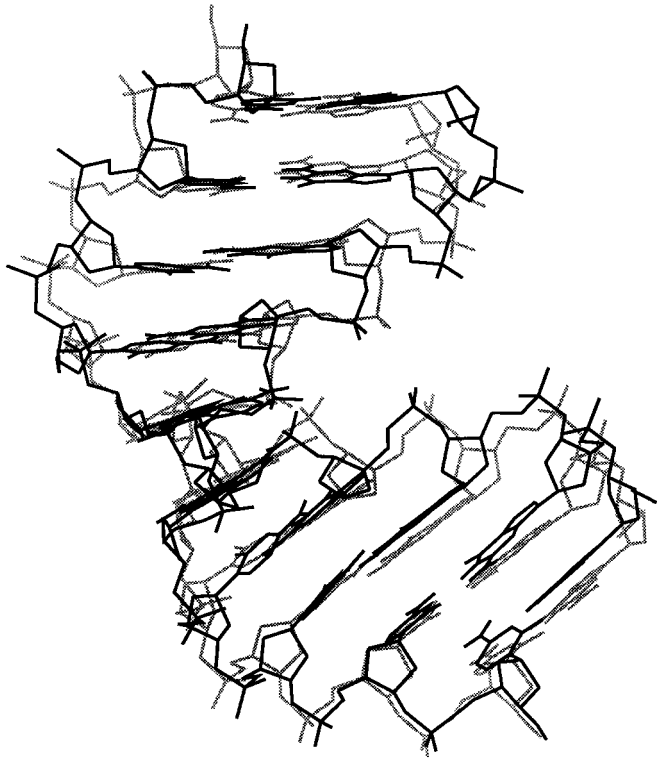

Figure 4
Superposition of the two duplexes in the asymmetric unit.

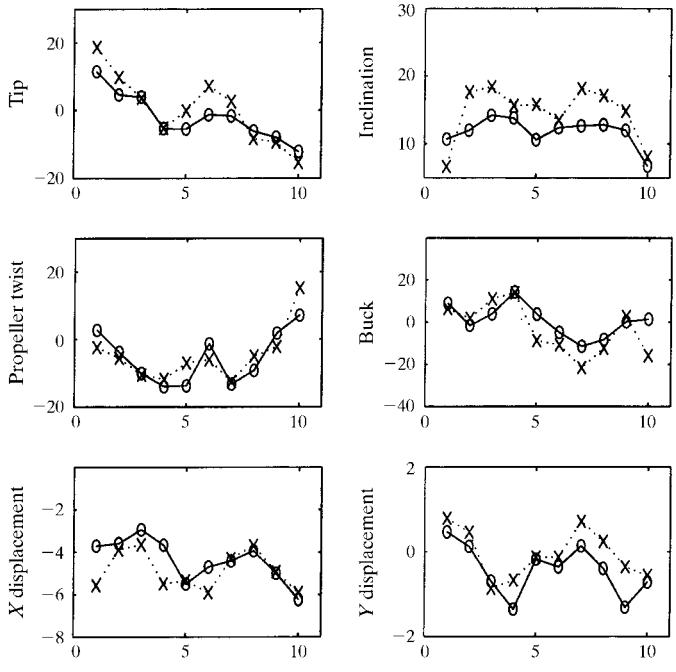

Figure 5
Base-pair parameters of the two molecules in the asymmetric unit. For the sake of clarity, both duplexes are numbered 1–10 in the standard 5′-to-3′ direction. In this and the following figures, duplex 2 is represented by a broken line.

4. Discussion

All crystalline A-DNA duplexes studied so far interact through their minor grooves. Most often, the terminal base pair of one duplex stacks against the shallow minor groove of a neighbouring molecule related to it by a twofold, fourfold or sixfold screw axis depending on the specific space group. In the present case, the sequence d(G_4CGC_4), which crystallizes in space group $P2_1$, packs in a similar fashion, with the two molecules related by a near dyad symmetry. The similarity of packing of d(G_4CGC_4) in $P2_1$ and d(CCGGC)r(G)d(CCGG) in $P2_12_12_1$ (Ban *et al.*, 1994) is shown in Figs. 1(*a*) and 1(*b*), respectively.

There have been detailed comparisons of the polymorphous crystal forms of the DNA sequence d(GCG_3C_3GC) and they have also been compared with an isomorphous crystal structure with a differing base sequence, d($C_3G_2C_2G_3$) (Ramakrishnan & Sundaralingam, 1993*a*,*b*). The similar conformations observed for different base sequences in the same crystal form and different conformations for the same base sequence in different crystal forms indicate that crystal-packing interactions dominate over base-sequence effects. It appears that packing requirements might be the determining factor for the loss of symmetry of the molecule upon crystallization. In the present structure, the d(G_4CGC_4) decamer in space group $P2_1$ does not utilize the chemical symmetry present, even though an almost identical packing of decamers could be achieved in space group $P2_12_12_1$ with one strand in the asymmetric unit. The lack of symmetry originates from intrinsic features in the d(G_4CGC_4) sequence, predisposing the structure to such intermolecular interactions as to give optimal packing in the crystal or counterion binding. Thus, the sequence, the oligomer length and the crystallization solvent may all have a strong influence on the DNA form and structure adopted by oligonucleotides upon crystallization.

A striking feature of d(G_4CGC_4) is that the phosphodiester chain torsion angles, the glycosidic angle and the average twist angle, in spite of large local variations, have average values which are almost identical to those found in fibres (Table 2). It appears that intramolecular forces stabilize the overall DNA conformation, so that when there is a strong local variation in structure it is compensated elsewhere in the molecule by variations with the opposite sign. In this context, it is interesting to note that those molecules which have most extreme twist angles present an alternation of large to small values, as found in d(GGCCGGCC) (Wang *et al.*, 1982), d(CCCCGGGG) (Haran *et al.*, 1987) and d(CTCTAGAG) (Hunter *et al.*, 1989). Such structural compensations appear to be a natural consequence of the tendency to maintain optimal stacking and helical backbone interactions and demonstrate the inherent conformational flexibility of DNA to adapt to the environment and assume a variety of perturbed conformations. Moreover, it appears that though the base sequence is initially responsible for seeding the three-dimensional arrangement, there is a fair amount of competition from the packing forces. This is reflected in the significant deviations seen in the local conformation of the two molecules in the asymmetric unit. This behaviour is probably a consequence of small differences in energy between conformations in the sequence.

It is interesting to compare the present structure with the related structures d($C_3G_2C_2G_3$) (Ramakrishnan & Sundaralingam, 1993*b*) and d(CCGGC)r(G)d(CCGG) (Ban *et al.*, 1994), which was used as the starting model. Both these structures are in the orthorhombic space group $P2_12_12_1$ and

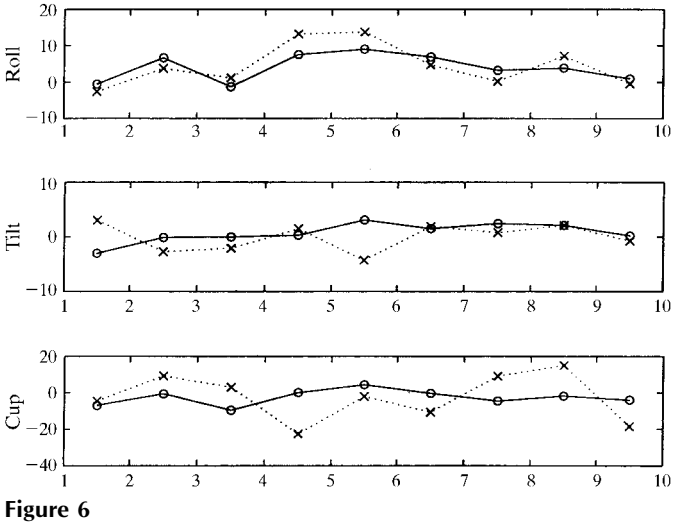

Figure 6
The roll, tilt and cup base-step parameters in the two molecules in the asymmetric unit.

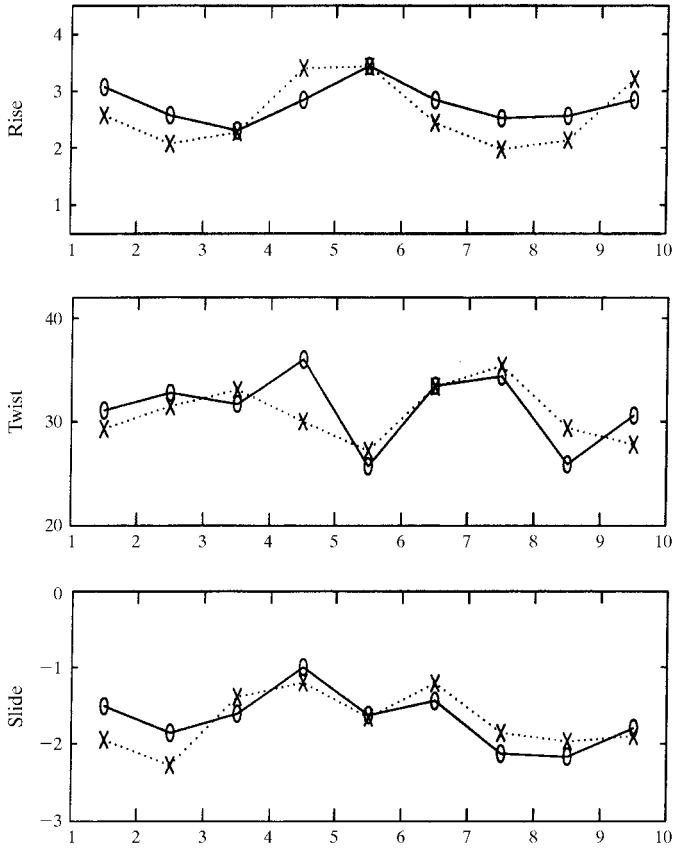

Figure 7
The rise, twist and slide base-step parameters in the two molecules.

have almost identical unit-cell dimensions to those of the $d(G_4CGC_4)$ crystal. They interact with their respective neighbours by forming a novel base-paired triple G4*G10·C11, hydrogen bonding through their minor-groove sites N2 and N3 to the minor-groove atoms N2 and O2 of both bases of the G10·C11 base pair of the symmetry-related molecule. Also, both these structures have different intermolecular contacts with their respective neighbours when compared with the $d(G_4CGC_4)$ structure, primarily owing to the fact that the end base pairs involved are different. The end base pair is G·C in the structures compared, while in $d(G_4CGC_4)$ it is a C·G pair. The observed intermolecular base-pairing interactions, in this case, are a consequence of the available donor and acceptor atoms or, in other words, of the base sequence. Hence, it follows that the deviations in the helical geometry experienced upon interaction would also be different, which can curtail the disposition of the molecules with respect to each other and hence affect the packing.

The authors are grateful to the Department of Biotechnology, India for financial aid. MAV thanks the Indian National Science Academy for financial support. GS thanks University Grants Commission for the scholarship provided. The X-ray data were collected at the National Area Detector facility at MBU, Indian Institute of Science, Bangalore. This study forms part of work carried out under Jawaharlal Nehru Centre for Advanced Scientific Research and the Centre for Human Genetics, Bangalore, India.

References

Arnott, S. & Selsing, E. (1974). *J. Mol. Biol.* **88**, 551–552.
Ban, C., Ramakrishnan, B. & Sundaralingam, M. (1994). *J. Mol. Biol.* **236**, 275–285.
Braig, K., Adams, P. D. & Brünger, A. T. (1995). *Nature Struct. Biol.* **2**, 1083–1094.
Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–474.
Brünger, A. T. (1992*b*). *X-PLOR Manual, Version* 3.1. Yale University, USA.
Brünger, A. T., Krukowski, A. & Erickson, J. (1990). *Acta Cryst.* A**46**, 585–593.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Frederick, C. A., Quigley, G. J., Teng, M.-K., Collect, M., van der Marel, G. A., van Boom, J. H., Rich, A. & Wang, A. H.-J. (1989). *Eur. J. Biochem.* **181**, 295–307.
Haran, T. E., Shakked, Z., Wang, A. H.-J. & Rich, A. (1987). *J. Biomol. Struct. Dynam.* **5**, 199–217.
Hunter, W. N., Langlois D'Estaintot, B. & Kennard, O. (1989). *Biochemistry*, **28**, 2444–24551.
Jain, S. & Sundaralingam, M. (1989). *J. Biol. Chem.* **264**, 12780–12784.
Kabsch, W. (1993). *J. Appl. Cryst.* **25**, 795–800.
Malathi, R. & Yathindra, N. (1985). *J. Biomol. Struct. Dynam.* **3**, 127–134.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Nature (London)*, **335**, 321–329.
Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). *Acta Cryst.* D**52**, 57–64.
Rabinovich, D., Haran, T., Eisenstein, M. & Shakked, Z. (1988). *J. Mol. Biol.* **200**, 151–161.
Ramakrishnan, B. & Sundaralingam, M. (1993*a*). *Biochemistry*, **32**, 11458–11468.
Ramakrishnan, B. & Sundaralingam, M. (1993*b*). *J. Mol. Biol.* **231**, 431–444.
Rice, L. M. & Brünger, A. T. (1994). *Proteins Struct. Funct. Genet.* **19**, 277–290.
Rodgers, D. W. & Harrison, S. C. (1993). *Structure*, **1**, 227–240.
Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). *Nature (London)*, **368**, 469–473.
Shimon, L. J. W. & Harrison, S. C. (1993). *J. Mol. Biol.* **219**, 321–334.
Viswamitra, M. A., Kennard, O., Jones, P. G., Sheldrick, G. M., Salisbury, S., Falvello, L. & Shakked, Z. (1978). *Nature (London)*, **273**, 687–688.
Wang, A. H.-J, Fujii, S., Van Boom, J. H. & Rich, A. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 3968–3972.
Weis, W. I., Brünger, A. T., Skehel, J. J. & Wiley, D. C. (1990). *J. Mol. Biol.* **212**, 737–761.
Westhof, E. (1987). *Int. J. Biol. Macromol.* **9**, 186–192.
Westhof, E., Prange, T., Chevrier, B. & Moras, D. (1985). *Biochimie*, **67**, 811–817.
Yathindra, N. & Sundaralingam, M. (1976). *Nucleic Acids Res.* **3**, 729–747.