Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

DNA variability in five crystal structures of d(CGCAATTGCG)

Núria Valls,[a] Glenford Wright,[a] Roberto A. Steiner,[b] Garib N. Murshudov[b] and Juan A. Subirana[a]*

[a]Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Avinguda Diagonal 647, E-08028 Barcelona, Spain, and [b]York Structural Biology Laboratory, Department of Chemistry, University of York, York Y010 5YW, England

Correspondence e-mail: juan.a.subirana@upc.es

Received 3 November 2003
Accepted 5 February 2004

PDB Reference: d(CGCAATTGCG), 1s23, r1s23sf.

NDB Reference: BD0066.

The deoxyoligonucleotide d(CGCAATTGCG) has previously been crystallized in four different space groups. The crystals diffract to moderate resolution (2.3–2.9 Å). Here, a fifth crystal form that diffracts to higher resolution (1.6 Å) is presented which was obtained thanks to the use of Co^{2+} and cryogenic temperatures. The availability of five different crystal structures allows a thorough analysis of the conformational variability of this DNA sequence. It is concluded that the central hexamer sequence CAATTG has a practically constant conformation under all conditions, whilst the terminal base pairs at both ends vary considerably as a result of differing interactions in the crystals. The new crystal structure presented here is stabilized by guanine–Co^{2+}–guanine interactions and the formation of $C1^{+}$–G8·C3 triplexes between neighbouring duplexes. As a result of the higher resolution of the crystal structure, a more regular structure was obtained and a clear definition of the spine of hydration was observed which was not visible in the four previous structures.

1. Introduction

The $d(CGCAATTGCG)_2$ duplex has been subjected to several crystallization attempts with different minor-groove-binding drugs (Spink *et al.*, 1995; Nunn *et al.*, 1997; Wood *et al.*, 1997) and peptides (Soler-López *et al.*, 2002). Except in the case of the drug netropsin (Nunn *et al.*, 1997), complexes have not yet been obtained. Crystallization was carried out in the presence of either Mg^{2+} or Zn^{2+} ions. Each crystallization experiment yielded a different space group. This indicates the effect of variable packing forces as a function of the presence of either ions, drugs or peptides. In all cases MPD was used as a precipitant.

Prompted by our finding that *N*-α-(9-acridinoyl)-tetra-arginine intercalates in an AA/TT step of the dodecamer duplex d(CGCGAATTCGCG) (Malinina *et al.*, 2002), we have tried to obtain a complex of d(CGCAATTGCG) with an acridine compound. However, no intercalation was found in the crystals obtained. We carried out several crystallization attempts. The best diffracting crystals (1.6 Å) were obtained using Co^{2+} ions. This metal ion participates in guanine–Co^{2+}–guanine bridges in the crystal.

In this paper, we describe the structure of the $d(CGCAATTGCG)_2$ duplex at 1.6 Å. This is the most accurate structure of this decamer achieved to date. In addition, we compare this structure with the other four available structures. This is a unique case in which the same oligonucleotide has been crystallized in five different environments. We will try to

ascertain the type of interactions that stabilize each crystal structure. We can also determine the variability of the oligonucleotide when found in different environments.

2. Materials and methods

2.1. Crystallization

The deoxydecanucleotide d(CGCAATTGCG) was supplied by Oswell DNA service (Southampton, UK) as the ammonium salt. The intercalating drug acridine-(Arg-Gly-Arg) was synthesized by solid-phase methods. After the Boc-peptide was assembled on a *p*-methylbenzhydrylamine resin, the N-terminus was deprotected and 9-acridinecarboxylic acid was coupled by means of PyBOP and diisopropylethylamine (5, 5 and 10 equivalents, respectively) in 1:1 DMSO/DMF for 1 h. Acidolysis with HF/anisole (9:1, 273 K, 1 h) provided the target compound in a highly homogeneous form (>95% by HPLC) and with a correct MALDI–TOF mass spectrum.

Crystals were grown by vapour diffusion in a hanging drop containing 0.5 m*M* DNA duplex, 1 m*M* acridine-(Arg-Gly-Arg) drug adduct, 20 m*M* sodium cacodylate buffer pH 6.5, 4 m*M* $CaCl_2$, 1 m*M* $CoCl_2$, 0.5 m*M* spermine and 10% MPD equilibrated against a 25% MPD reservoir. Colourless crystals grew at 293 K in approximately three weeks to typical dimensions of 0.2 × 0.06 × 0.06 mm.

2.2. X-ray data collection

Crystals were flash-cooled at 120 K and diffraction data were collected on a MAR CCD detector using synchrotron radiation at ESRF beamline BM14. Two data sets were collected at resolution cutoffs of 2.25 and 1.6 Å, respectively. The low-resolution pass was needed to accurately measure the high-intensity reflections, which resulted in detector saturation in the high-resolution pass. The data were integrated and scaled with the *HKL* suite (Otwinowski & Minor, 1997).

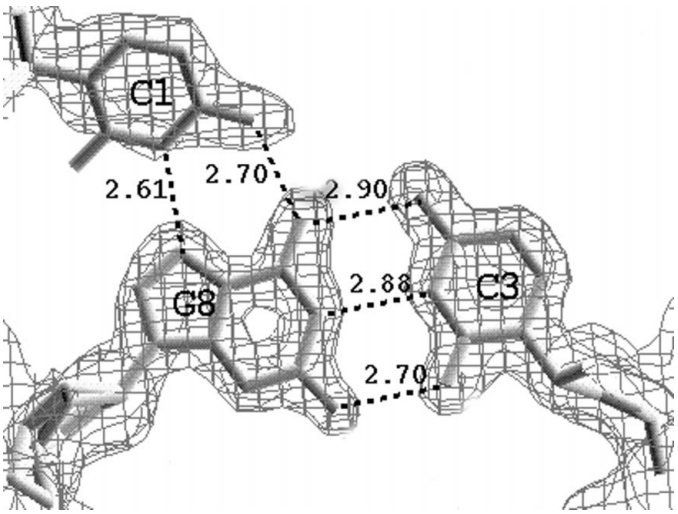

Figure 1
Electron-density map at the 1σ level of the flipped-out base C1 which forms a triplet with base pair C3·G8 in a neighbouring duplex. Hydrogen bonds are shown as dashed lines with distances in Å. Note that the triplet is not planar, as is apparent from Fig. 3.

Table 1
Crystallographic and refinement statistics.

Values in parentheses are for the last resolution shell.

Space group	$I2_12_12_1$
Unit-cell parameters	a = 26.22, b = 44.36, c = 52.34
Asymmetric unit contents	One DNA strand, 0.5 Co^{2+} ions, 46 H_2O
Temperature (K)	120
Resolution range (Å)	25–1.6 (1.66–1.6)
Completeness (%)	98.4 (89.1)
Unique reflections	4172
Overall redundancy†	17.7
R_{sym}‡	0.077 (0.60)
R_{work}§/R_{free}¶	0.21/0.281
R_{work} (all reflections)	0.214

† Total reflections registered divided by the number of unique reflections. ‡ $R_{sym}(I) = \sum_{hkl}\sum_j |I_j(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_j I_j(hkl)$ calculated for the whole data set. § $\sum_{hkl} |F_o(hkl) - kF_c(hkl)| / \sum_{hkl} F_o(hkl)$. ¶ R factor of reflections used for cross-validation in the refinement.

Crystallographic data and refinement statistics are given in Table 1.

2.3. Structure determination and refinement

The structure of d(CGCAATTGCG) was determined by the molecular-replacement method using the program *EPMR* v.2.5 (Kissinger *et al.*, 1999) starting from the low-resolution model (NDB code UDJ031) published by Spink *et al.* (1995). Data in the resolution range 15–3 Å were used to determine the molecular-replacement solution. Bases in the decamer were numbered 1–10 starting with the cytosine at the 5′ end. The terminal bases C1 and G10 were not used in the model since they were not part of the duplex. The initial model was subjected to rigid-body refinement using the program *CNS* v.1.1 (Brünger *et al.*, 1998) with data in the resolution range 25–2.0 Å. At this stage, 10% of the unique reflections were set apart to calculate a free R factor (Brünger, 1992) as an independent cross-validation indicator of the progress of refinement. After several cycles of refinement, electron density clearly showed the base G10 and an ion that was modelled as Co^{2+}. Extra density for the C1 base was seen next to G2. However, it was not possible to model its exact position at this stage. Refinement was continued using the program *REFMAC*5 (Murshudov *et al.*, 1997) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) using the same resolution limits. After several cycles of maximum-likelihood isotropic restrained refinement, the C1 base appeared clearly with a flipped-out geometry in $mF_o - DF_c$ and $2mF_o - DF_c$ electron-density maps. At this stage of refinement, water molecules were added to the model at positions corresponding to peaks in the difference Fourier higher than the 3σ level if they satisfied hydrogen-bonding criteria. Additional water molecules could be located in the electron-density maps by including the high-resolution data (to 1.6 Å). This also allowed a more reliable inclusion in the refinement of H atoms at their riding positions. We also found two positions for the phosphate of C9 and the sugar of G8. Note that data in the last shells contained reflections with a significant intensity (>3σ) which contributed to the improvement of the electron-density map. Manual rebuilding was

carried out with the programs *TURBO-FRODO* (Roussel *et al.*, 1998) and *QUANTA* (Accelrys Inc.). The refinement converged at $R = 21.0\%$ and $R_{free} = 28.1\%$ for the resolution range 25–1.6 Å. Further refinement details are given in Table 1. No residual electron density was found that could be attributed to the drug. An example of the quality of the final electron-density map in the C1 region is shown in Fig. 1. This figure was prepared with the program *SETOR* (Evans, 1993). Figs. 2–5 were drawn with the program *Cerius*2 (Accelrys Inc.).

3. Results

3.1. Description of the present structure

The oligonucleotide d(CGCAATTGCG) has been crystallized in space group $I2_12_12_1$ with one strand in the asymmetric unit. Continuous columns of duplexes in the B-DNA conformation with antiparallel Watson–Crick base pairs are formed for eight bases of the decamer (the packing of the crystal is shown in Fig. 2). The terminal bases C1 and G10 are flipped out. Columns of duplexes are linked through a Co^{2+} ion that interacts with two flipped-out guanines from neighbouring molecules with the N7 of G10 at a typical distance of 2.2 Å from the metal ion. The interaction of the Co^{2+} ion has previously been described in other oligonucleotide structures in the presence of either Co^{2+} (Yang *et al.*, 2000) or Ni^{2+} (Abrescia *et al.*, 1999). Thus, the present structure adds a new crystallization environment for this oligonucleotide. Our work also shows that Co^{2+} ions may be useful to obtain crystals that diffract to high resolution, since they form intermolecular bridges between symmetry-related molecules.

The terminal base G10 enters the minor groove of the next duplex in a column. It is hydrogen bonded to G2. Since its description by Spink *et al.* (1995), this type of interaction has been frequently found in different structures as reviewed by Subirana & Abrescia (2000).

The terminal cytosine (C1) is also flipped out. This observation is clearly supported by electron-density maps (Fig. 1). However, the average *B* factor for this base suggests a greater disorder compared with the rest of the structure. C1 is protonated at N3 and interacts with G8 by hydrogen bonding. The interactions of the terminal bases are shown in detail in Fig. 3.

The oligonucleotide crystal contains an extensive network of water molecules, some of which form bridges between neighbouring oligonucleotide columns. All phosphate groups are hydrogen bonded to water molecules. Most bases also show direct hydrogen bonds to water molecules. An extensive network covers the major-groove side of the central AATT region. A well defined water spine is found at the centre of the minor groove, as shown in Fig. 4. It is virtually identical to that described in other high-resolution structures of oligonucleotides that also contain a central AATT sequence (Soler-López *et al.*, 1999).

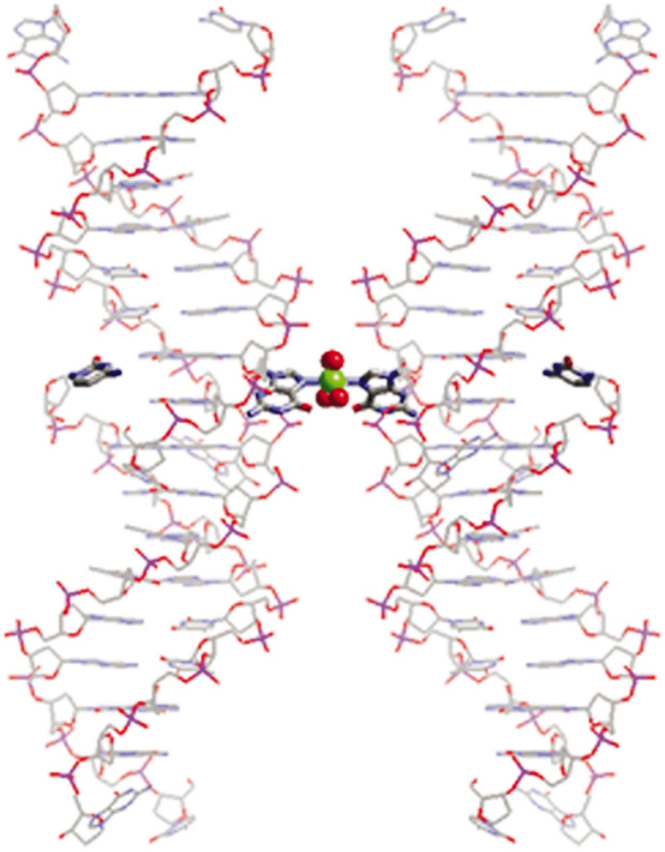

Figure 2
Two neighbouring columns in the crystal structure. A cobalt ion (green) interacts with the N7 atom of two guanines. The bonds of the flipped-out residues in the central region are shown with a larger diameter. Hydration waters associated with the Co^{2+} ion are represented as red spheres.

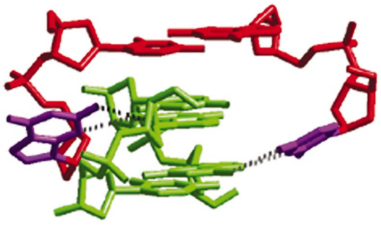

Figure 3
A detailed view of the interaction between the ends of two neighbouring duplexes (red and green) in a column. The flipped-out residues (C1 and G10) are shown in purple. Hydrogen bonds between G10 and G2 and between C1 and G8 are indicated as dashed lines.

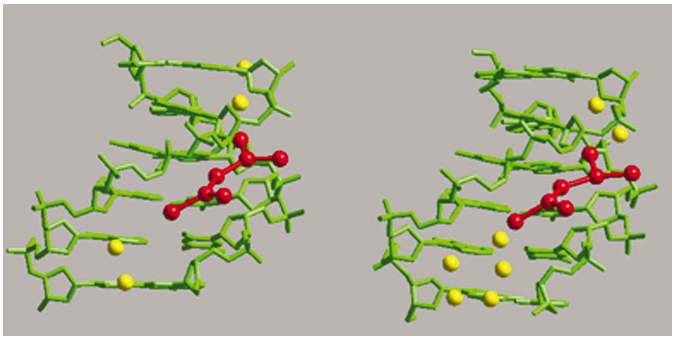

Figure 4
Comparison of the hydration spine in our structure (left) with the spine present in d(GCGAATTCG) (right) (Soler-López *et al.*, 1999). Red spheres indicate waters in practically identical positions, with hydrogen bonds between them shown as virtual bonds. Hydration waters away from the central region (yellow) have different positions.

Table 2
Structural features of d(CGCAATTGCG) in different crystal forms.

Structure	NDB code	Space group	Contents of AU	Resolution (Å)	Ions	C1	G10	$\langle D_z \rangle$† (Å)	Minor-groove width‡ (Å)	Volume per bp (Å^3)	Reference
1	UDJ031	$I2_12_12_1$	One strand	2.5	Mg^{2+}	Flipped out	Minor groove	3.27 (0.07)	3.8	1432	Spink *et al.* (1995)
2	GDJ046	$P2_12_12_1$	Duplex	2.4	Mg^{2+}, netropsin	C1$^+$–G8·C13	Minor groove	3.36 (0.14)	4.1	1337	Nunn *et al.* (1997)
3	BDJ069	$C2$	Duplex	2.3	Mg^{+2}	Duplex	Duplex	3.37 (0.35)	5.0	1237	Wood *et al.* (1997)
4	UD0012	$C222_1$	One strand	2.9	Zn^{+2}, peptide	Flipped out	Minor groove	3.34 (0.10)	3.1	1638	Soler-López *et al.* (2002)
5	BD0066	$I2_12_12_1$	One strand	1.6	Co^{+2}	C1$^+$–G8·C13	Minor groove	3.25 (0.07)	3.2	1510	This work

† $\langle D_z \rangle$ is the average value of rise. Its standard deviation is given in parentheses. ‡ The minor-groove width at the centre of the octamer was calculated with the 3*DNA* program (Lu *et al.*, 2003). The van der Waals radius (5.8 Å) of the phosphate group was substracted.

3.2. Comparison with other structures

3.2.1. General features. Various studies of the oligonucleotide that we have crystallized are summarized in Table 2. The central octamers of the five structures are compared in Fig. 5. All of them are rather similar, particularly in the central region. The main differences are found in the phosphodiester backbone of both ends, as shown in Fig. 5. It should be noted that only structure 3 has all bases in a duplex conformation; all other structures have the terminal C1 and G10 bases flipped out and interacting with neighbouring duplexes.

Structures 1–3 in Table 2 were collected at room temperature and are expected to have higher disorder than structures 4 and 5, which were collected at 120 K. However, structure 4 also shows high disorder owing to the presence of a mobile peptide moiety which contains four arginines. The presence of the peptide also gives a higher volume per base pair, as shown in Table 2.

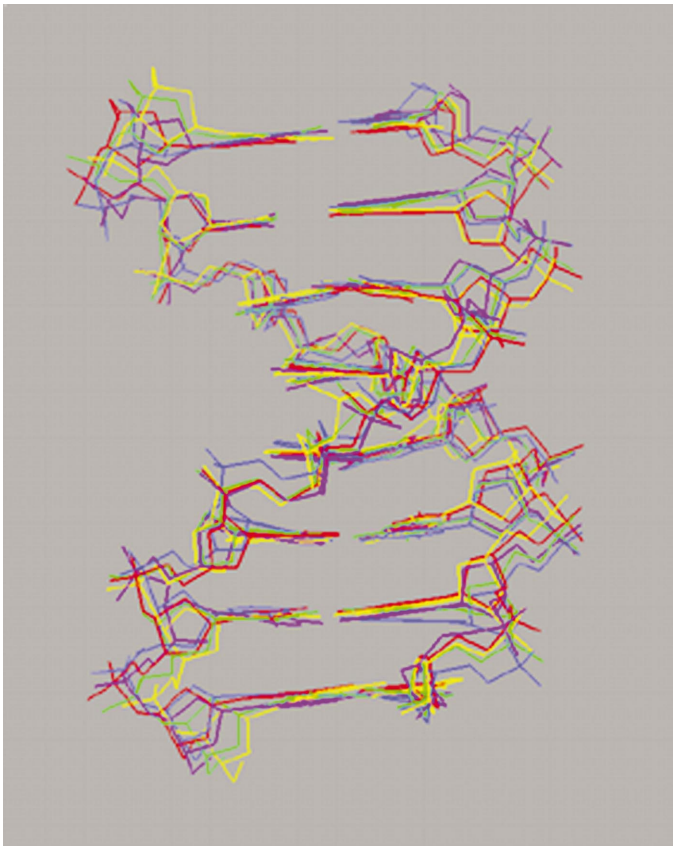

Figure 5
Superposition of all structures. Structure 1 (green) has been taken as reference. Colours correspond to the following structures: 2, blue; 3, purple; 4, yellow; 5, red.

3.2.2. Packing. All structures reported in Table 2 display similar packing, except for structure 3, which crystallized with a very different symmetry in space group *C*2. The duplexes also packed in layers of parallel columns, but the helices in neighbouring layers have their axes crossed at an angle of 42° (Wood *et al.*, 1997). Structures 1 and 5 were crystallized under very different conditions. However, the structures are almost identical. They both have the base G10 flipped out. In structure 5 the base G10 interacts with a Co^{2+} ion, whereas in structure 1 water molecules connect the different columns in a way similar to that shown in Fig. 2. Base C1 also presents a different position. In our structure the interaction C1–G8 through two hydrogen bonds is clearly seen, while in structure 1 the base C1 was very disordered and tentatively modelled as interacting with the base pair G2·C9 of the next duplex. The higher resolution of structure 5 may help to define a more accurate position for the base C1. The netropsin structure 2 has C1 in the same position as shown in Fig. 1, whereas in structure 4 it shows high disorder owing to the fact that it is bound to the disordered peptide chain.

In all cases except structure 3, the terminal guanines G10 enter the minor groove of the next duplex in the oligonucleotide columns as shown in Fig. 2 and in more detail in Fig. 3.

The oligonucleotide columns in structures 1 and 5 have the same relative positions, as shown in Fig. 2, whereas in structures 2 and 4 they are also parallel but displaced; there is no guanine–guanine interaction between neighbouring columns.

Ions play a very important role in crystallization. Crystals 1, 2 and 3 contain Mg^{2+} but no ion was present in the final refined structures. Zn^{2+} and Co^{2+} are clearly seen in structures 4 and 5, respectively. In crystal 4, two kinds of Zn^{2+} ions were found: one interacts with the oxygen of phosphates and the other interacts with N7 of a single guanine; it does not form a bridge between columns.

3.2.3. Minor-groove structure. In general (Table 2), all five decamers show the narrow minor groove typically found in AT-rich sequences, but because of the presence of netropsin,

Table 3
Twist (°) of individual base steps in d(CGCAATTGCG).

Only the central octamer in the duplex conformation is analyzed. Expected values and their standard deviation are taken from Subirana & Faria (1997). Values which deviate by more than two standard deviations are shown in bold. Note that in structures 2 and 3 the asymmetric unit is the whole duplex instead of a single strand. Therefore, there are two values for each step, except in the central ApT.

Structure	GpC	CpA/TpG	ApA/TpT	ApT	⟨Ω⟩†
1	37.4	38.1	30.0	**41.5**	36.1 (4.4)
2	34.0/38.2	32.3/32.3	37.1/37.2	34.5	35.1 (2.4)
3	**24.9**/37.9	35.8/38.7	38.2/39.6	30.4	35.1 (5.4)
4	39.0	31.7	**43.3**	29.6	36.8 (5.7)
5	**31.8**	33.2	36.2	34.9	33.9 (1.8)
Expected	37.5 (2.7)	38.5 (8.6)	35.3 (3.2)	32.0 (2.8)	

† Average twist value, with standard deviation in parentheses.

structure 2 has a wider minor groove. In the case of structure 3, the interaction with neighbouring duplexes by interpenetration of their grooves also results in a wider minor groove.

3.2.4. Structural details. As shown in Fig. 5, all five structures are rather similar; the r.m.s. deviations between the central octamers are rather small. They range between 0.71 Å (for structures 1 and 3) and 1.22 Å (for structures 2 and 4). The conformational parameters of all structures were compared using the data available in the NDB (Berman *et al.*, 1992). Those for the present structure were calculated with 3*DNA* v.1.5 (Lu *et al.*, 2003). Only small differences were found between the structures, with the exception of the structure 3, in which several phosphates adopted the BII conformation. The central A·T base pairs all show a large negative propeller twist (around −14°) as expected for this sequence. As a result, the distance between the C2 atoms of adenine and the O2 atom of thymine in neighbouring base pairs (AA/TT steps) of our structure is shortened to 3.27 Å, which may be considered to be an additional weak hydrogen bond (Madhumalar & Bansal, 2003). However, the angle C2—H2···O2 is only 114°, which is not optimal for a hydrogen bond.

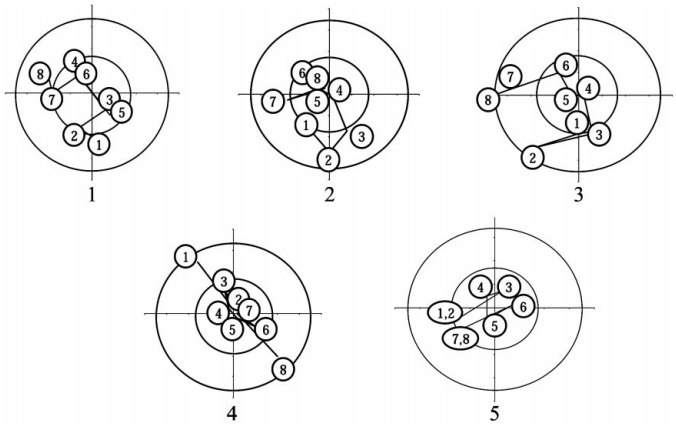

Figure 6
Plots of base-pair normal vectors for the central octamer in all five structures. Inner and outer circles represent angles of 5 and 10° with respect to the centre of the duplex (between base pairs 4 and 5). Deviations are larger at either end of the duplexes.

As an example of the small differences found, the twist values are compared in Table 3 for all five structures. Most of them fall within the expected average range, with some exceptions that are indicated in bold in the table. These differences probably arise from the variable interactions of the oligonucleotides in the crystal, although some of them, in particular the unexpected changes in the central AATT region, might arise from inaccuracy of the coordinates related to the low resolutions of some of the structures. It is also worth noting that the twist values for the CA/TG step are rather constant. This is surprising as this is a very variable step (Subirana & Faria, 1997).

An interesting feature of our structure 5 is that the variations in twist (Table 3) and in the conformational angles of all bases are rather small (results not shown), which we attribute to the higher accuracy of the coordinates related to higher resolution of our structure. It has a regular duplex structure with small local variations independent of the local sequence.

In Fig. 6 we also present a bending plot of the central octamer. It can be noted that the central region is rather straight; bending is only detected in the terminal base pairs, which should be attributable to their variable interactions in the crystal structure.

4. Discussion

One of the main conclusions of the comparison we have presented is that the crystallization conditions do not have a great influence on the overall conformation of the oligonucleotide except at the ends, which show variable interactions in the crystal. On the other hand, it is not clear why the crystallization conditions have such a strong influence on the crystal packing. For example, structures 1 and 3 were crystallized under very similar conditions. A different drug was present, but it was not incorporated in the crystal lattice. In this respect our structure was also surprising, as under similar crystallization conditions using Ni^{2+} instead of Co^{2+} a related decamer crystallized in space group $P4_12_12$ (Abrescia *et al.*, 1999). Flipped-out guanines formed similar ion bridges between neighbouring columns, but instead of being parallel as shown in Fig. 2, they were perpendicular when Ni^{2+} was used.

The high resolution achieved for crystal 5 allows us to very clearly visualize a spine of hydration (Fig. 4). This was not seen in all previous low-resolution structures of the decamer. Interestingly, related structures solved at similar resolution instead display a well defined spine of hydration as in Vlieghe *et al.* (1996).

It seems plausible that cryogenic temperatures may be required to stabilize the water molecules. In general, it appears that our structure, which has the highest resolution, is less variable than the others.

We are grateful to Drs G. Parkinson and L. Malinina for valuable help in the molecular-replacement stage. We thank Drs L. Campos and C. Cáceres for help and discussion. NV

acknowledges a fellowship from the Generalitat de Catalunya. This work has been supported by grants BI2002-00317 from the Ministerio de Ciencia y Tecnología and 2001 SGR 00250 from the Generalitat de Catalunya. Support from the European Project HPRN-CT-2000-00009 is also acknowledged. We thank the BM14 Spanish beamline staff of ESRF (Grenoble) for assistance with data collection. RAS and GNM thank the Wellcome Trust for support.

References

Abrescia, N. G. A., Malinina, L., Fernandez, L. G., Huynh-Dinh, T., Neidle, S. & Subirana, J. A. (1999). *Nucleic Acids Res.* **27**, 1593–1599.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). *Biophys. J.* **63**, 751–759.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998) *Acta Cryst.* D**54**, 905–921.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Evans, S. V. (1993). *J. Mol. Graph.* **11**, 134–138.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, B. (1999). *Acta Cryst.* D**55**, 484–491.

Lu, X.-J., Shakked, Z. & Olson, W. K. (2003). *Nucleic Acids Res.* **31**, 5108–5121.

Madhumalar, A. & Bansal, M. (2003). *Biophys. J.* **85**, 1805–1816.

Malinina, L., Soler-López, M., Aymamí, J. & Subirana, J. A. (2002). *Biochemistry*, **41**, 9341–9348.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Nunn, C. M., Garman, E. & Neidle, S. (1997). *Biochemistry*, **36**, 4792–4799.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Roussel, A., Inisan, A. G., Knoops-Mouthuy, E. & Cambillau, E. (1998). *TURBO-FRODO* version OpenGL.1. University of Marseille.

Soler-López, M., Malinina, L., Liu, J., Huynh-Dinh, T. & Subirana, J. A. (1999). *J. Biol. Chem.* **274**, 23683–23686.

Soler-López, M., Malinina, L., Tereshko, V., Zarytova, V. & Subirana, J. A. (2002). *J. Biol. Inorg. Chem.* **7**, 533–538.

Spink, N., Nunn, C. M., Vojtechovsky, J., Berman, M. & Neidle, S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10767–10771.

Subirana, J. A. & Abrescia, N. G. A. (2000). *Biophys. Chem.* **86**, 179–189.

Subirana, J. A. & Faria, T. (1997). *Biophys. J.* **73**, 333–338.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B., Précigoux, G. & Kennard, O. (1996). *Acta Cryst.* D**52**, 766–775.

Wood, A. A., Nunn, C. M., Trent, J. O. & Neidle, S. (1997). *J. Mol. Biol.* **269**, 827–841.

Yang, X.-L., Robinson, H., Gao, Y.-G. & Wang, A. H.-J. (2000). *Biochemistry*, **39**, 10950–10957.