Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of d(TGCGCA)$_2$ at 293 K: comparison of the effects of sequence and temperature

S. Thiyagarajan, P. Satheesh Kumar, S. S. Rajan and N. Gautham*

Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai 600025, India

Correspondence e-mail: crystal@giasmd01.vsnl.net.in

Received 24 December 2001
Accepted 17 May 2002

NDB Reference:
d(TGCGCA)$_2$, ZD0007

The crystal structure of a hexameric DNA fragment with the sequence d(TGCGCA)$_2$ has been solved and refined at 293 K at a resolution of 1.64 Å. The molecule adopts a left-handed Z-type helical conformation which is common in alternating pyrimidine–purine sequences. The presence of A·T base pairs at the two terminals does not perturb the structure to any great degree. However, several sequence-specific microstructural changes are noticeable. The structure of the identical sequence determined at 120 K involving somewhat different crystallization conditions has been reported previously [Harper *et al.* (1998), *Acta Cryst.* D**54**, 1273–1284]. A comparison of the present structure with that at low temperature and with that of d(CGCGCG)$_2$ shows that the effect of the change in sequence is greater than the combined effect of changes in temperature and environment.

1. Introduction

The alternating copolymer d(CG)$_n$ adopts a left-handed helical structure, named Z-DNA, when exposed to high salt or ethanol concentration (Pohl *et al.*, 1972). Crystal structures of several oligonucleotides (Wang *et al.*, 1979; Crawford *et al.*, 1980; Gessner *et al.*, 1985) with alternating pyrimidine–purine sequences showed that the conformation of the Z-type helix is maintained more rigidly than in the case of the other two major helical types, *viz*. A-type and B-type DNA (Ho & Mooers, 1997). Nevertheless, sequence-specific microstructural effects have also been observed in Z-type DNA (Coll *et al.*, 1988; Sadasivan & Gautham, 1995). As part of a systematic programme to study such effects, we have undertaken the determination of the structures of hexameric DNA fragments with sequences based upon the 'canonical' Z-DNA sequence d(CGCGCG)$_2$ (Wang *et al.*, 1981). We have previously reported the crystal structures of d(CACGCG)·d(CGCGTG) and d(CGCACG)·d(CGTGCG) (Sadasivan & Gautham, 1995). As a continuation of these studies, we have undertaken the solution of the structures of hexameric sequences with two A·T base pairs instead of a single one. In this paper, we report the structure of d(TGCGCA)$_2$ at 293 K (henceforth referred to as HAT61). The structure was solved and refined to a resolution of 1.64 Å. A report of the structure of the identical sequence (HAT61/120) at 120 K (Harper *et al.*, 1998) became available while we were in the final stages of the present work. The crystallization solvents involved in HAT61 and HAT61/120 were also somewhat different. This facilitated a comparison of the effect of temperature and environment on the one hand and sequence on the other on the microstructure of Z-DNA.

2. Experimental

The synthetic self-complementary oligonucleotide was purchased from M/s Microsynth, Switzerland. A PAGE gel photograph supplied by the manufacturer suggested that the sample could be used for crystallization without further purification. Crystals were grown at room temperature (293 K) by the hanging-drop vapour-diffusion method. The crystallization drop contained 10 m*M* DNA at pH 7.0 together with 150 m*M* cobalt hexammine chloride and 25 m*M* magnesium chloride. The drop was equilibrated against 35% methyl pentanediol in the reservoir. A single crystal of dimensions 0.15 × 0.1 × 0.1 mm was used for data collection on a MAR Research imaging-plate system at the National Area Detector Facility, Indian Institute of Science, Bangalore, India. The structure was solved by the molecular-replacement method using *AMoRe* (Navaza, 1994) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The starting model was based on the coordinates of Z-DNA (Wang *et al.*, 1981) with A·T replacing the terminal C·G base pairs. The structure was first refined as a rigid body using *CNS* (Brünger *et al.*, 1998) with data in the

resolution range 8–2 Å, after applying a $1\sigma(F_o)$ cutoff to the input reflections. After several cycles of refinement, the R factor converged to 40%. Keeping aside 10% of the data in the resolution range 19–1.6 Å for cross-validation, the remaining reflections were used for positional refinement by Powell minimization. The bulk-solvent correction was applied. The structure was then subjected to simulated annealing using the slow-cool protocol (Brünger *et al.*, 1987, 1989, 1990). An electron-density map was calculated using $(2F_o - F_c)$ Fourier coefficients. The structure fitted well into the map after some minor changes. 30 water molecules could be identified in the map and were added to the structure. It was once again subjected to a few cycles of positional refinement followed by identification of some further water molecules. These steps were repeated until a total of 58 water molecules had been added. Two strong densities that appeared alone even at the 4σ level of the electron-density map were identified as magnesium ions. Both the magnesium ions had four hydration sites as revealed from the electron-density map. The final R factor is 21.2% ($R_{\rm free}$ = 26.8%). There are no gross unexplained features in the map. The temperature factors of all the atoms of the structure, the water molecules and the two Mg ions are found to fall within acceptable ranges.

The following numbering scheme is adopted in the analysis of the conformation.

5′ T1 p G2 p C3 p G4 p C5 p A6 3′

3′ A12 p C11 p G10 p C9 p G8 p T7 5′

The conformational calculations were carried out using the program *FREEHELIX*98 (Dickerson, 1998).

3. Results and discussion

The overall structure of the self-complementary duplex falls in the Z-DNA family (Wang *et al.*, 1979). The structure is very similar to that of d(CGCGCG)$_2$ (Wang *et al.*, 1979; Wang & Teng, 1988). The root-mean-square deviation (r.m.s.d.) in the positions of the common atoms after a least-squares superposition on d(CGCGCG)$_2$ is 0.93 Å. When superposed on its low-temperature counterpart HAT61/120, the r.m.s.d. is 0.87 Å. These numbers indicate little difference in the overall structure arising from either changes in sequence or temperature. However, as may be seen in Fig. 1, a closer analysis reveals that many structural parameters differ significantly between the present structure, d(CGCGCG)$_2$ and HAT61/120. Fig. 1 also shows that despite these differences, the structure of HAT61 is closer to that of HAT61/120 than to d(CGCGCG)$_2$. The structure of the latter sequence is, in turn, closer to that of its low-temperature counterpart (Bancroft *et al.*, 1994) than to that of HAT61 or HAT61/120. These results are discussed below.

3.1. Temperature effects

Table 1 gives the group B factors of HAT61 and compares them with those of

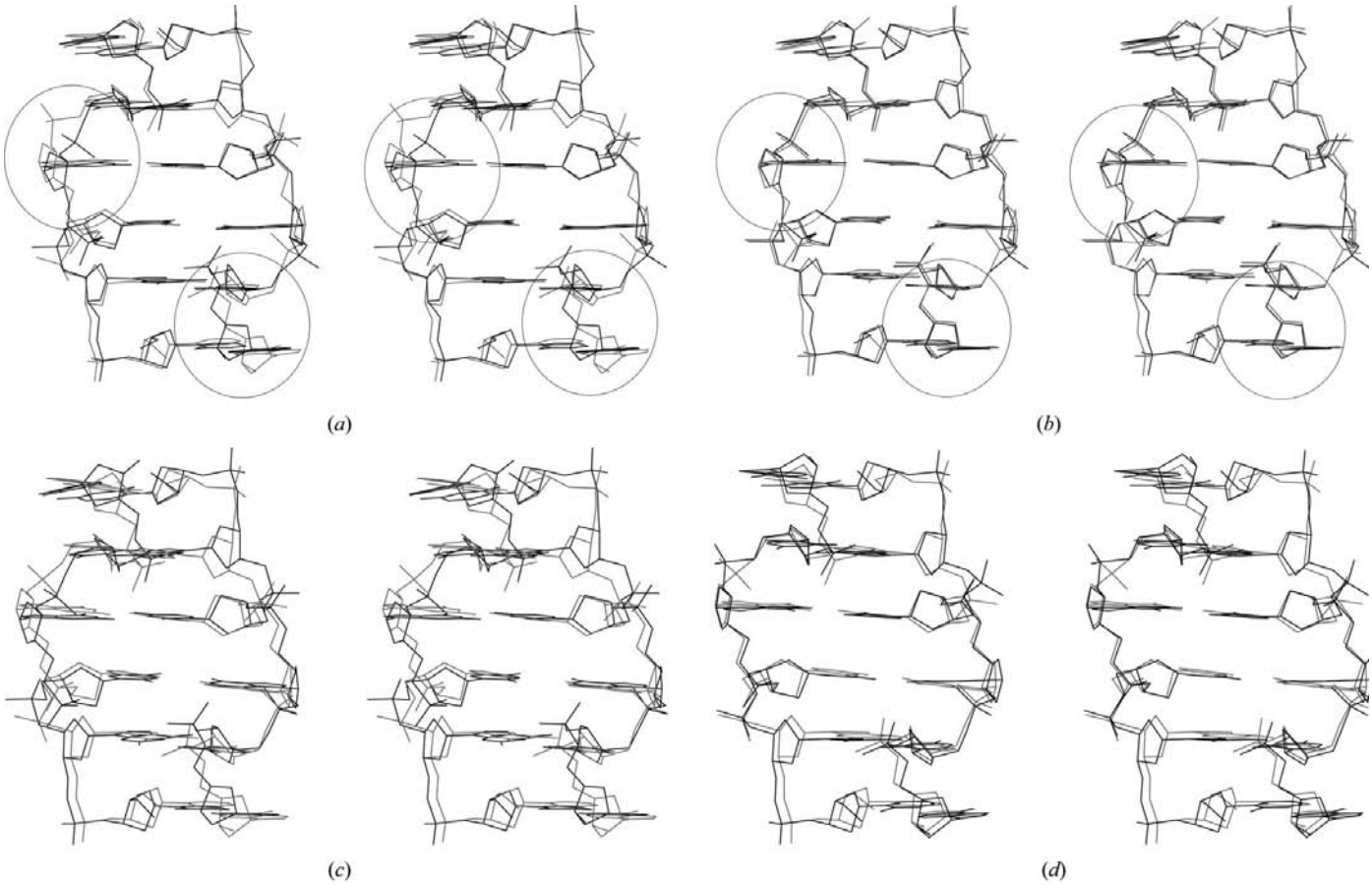

Figure 1
(*a*) Least-squares superposition of HAT61 (thick lines) on d(CGCGCG)$_2$ at room temperature. (*b*) Least-squares superposition of HAT61 (thick lines) on HAT61/120. (*c*) Least-squares superposition of HAT61/120 (thick lines) on d(CGCGCG)$_2$ at 163 K. (*d*) Least-squares superposition of d(CGCGCG)$_2$ at 163 K (thick lines) on d(CGCGCG)$_2$ at room temperature. Regions highlighted by circles indicate that the HAT61 structure is closer to HAT61/120 than to d(CGCGCG)$_2$ in (*a*) and (*b*). Clearly, the differences are greater between structures of different sequences than between structures of the same sequence at different temperatures.

Table 1
The group *B* factors (Å^2) for HAT61.

Values for HAT61/120 are given in parentheses for comparison.

	B factors (Å^2)		
Residues (base pair)	Base	Ribose	Phosphate
T1·A12	15.3/18.1 (9.7/9.3)	22.5/23.1 (16.0/9.3)	23.9 (9.1)
G2·C11	14.3/12.9 (7.4/6.9)	18.4/19.8 (10.1/8.3)	23.2/30.9 (23.1/9.5)
C3·G10	11.5/14.4 (3.8/4.6)	14.3/19.7 (5.4/5.9)	19.6/16.3 (10.8/5.7)
G4·C9	13.3/10.0 (4.1/3.6)	19.1/14 (5.2/4.2)	17.8/17.1 (8.5/5.1)
C5·G8	10.8/10.4 (4.1/3.6)	14.8/15.9 (4.7/5.2)	24.3/20.5 (5.4/12.4)
A6·T7	13.8/11.7 (6.3/6.6)	15.8/14.0 (8.4/7.1)	16.4 (6.5)

Table 2
Base-step parameters (calculated using *FREEHELIX*98; Dickerson, 1998) for HAT61, HAT61/120 and d(CGCGCG)$_2$.

Sequence	Base step	Tilt (°)	Roll (°)	Slide (Å)	Twist (°)	Rise (Å)
HAT61	T1·A12:G2·C11	0.73	3.68	3.90	−16.20	3.72
HAT61/120	T1·A12:G2·C11	−4.33	−2.52	5.42	−13.85	3.53
d(CGCGCG)$_2$	C1·G12:G2·C11	0.87	5.09	5.23	−7.35	3.81
HAT61	G2·C11:C3·G10	0.26	0.83	−0.47	−51.45	3.38
HAT61/120	G2·C11:C3·G10	−0.43	2.13	−0.43	−53.43	3.37
d(CGCGCG)$_2$	G2·C11:C3·G10	0.01	3.54	−0.82	−50.40	3.83
HAT61	C3·G10:G4·C9	2.19	1.43	5.28	−10.84	3.61
HAT61/120	C3·G10:G4·C9	2.39	1.82	5.28	−9.97	3.73
d(CGCGCG)$_2$	C3·G10:G4·C9	−0.69	4.08	5.35	−7.73	3.65
HAT61	G4·C9:C5·G8	−0.71	4.06	−1.07	−46.15	3.95
HAT61/120	G4·C9:C5·G8	−0.40	5.28	−1.21	−46.44	3.96
d(CGCGCG)$_2$	G4·C9:C5·G8	0.54	1.73	−0.65	−51.83	3.61
HAT61	C5·G8:A6·T7	−8.16	−0.95	3.58	−10.38	3.72
HAT61/120	C5·G8:A6·T7	−2.59	8.71	5.03	−9.19	3.55
d(CGCGCG)$_2$	C5·G8:G6·C7	−0.67	2.28	5.37	−10.70	3.84

HAT61/120. As expected, the *B* factors are relatively higher for HAT61 than for its low-temperature counterpart. However, the variation in the group *B* factors follows roughly the same pattern in both structures, indicating that sequence-specific effects remain even at different temperatures. At both temperatures the terminal bases and sugars have higher group *B* factors than the bases in the interior. Table 1 also indicates an interesting asymmetry in the values at the two ends in both structures, which is not in consonance with the symmetry in the sequence. The base and sugar atoms of the terminal base pair T1·A12 have larger *B* values than the A6·T7 base pair at the other end. The highest *B* factor (30.9 Å^2) in the molecule is observed at the phosphate group of C11. In HAT61/120, this phosphate is disordered (Harper *et al.*, 1998). It is likely that at the low temperature at which HAT61/120 was studied, these atoms froze in the two positions. In the present structure, these atoms display their flexibility in the form of higher temperature factor.

3.2. Base-step and base-pair parameters

Table 2 gives the local helical and base-centred parameters corresponding to the five base steps in HAT61, as well as HAT61/120 and d(CGCGCG)$_2$. The average helical parameters for the hexamer are similar to those seen in other Z-type structures. Only the differences are discussed below. The twist is −16.2° at the terminal T1G2:C11A12 step. This value is higher than in d(CGCGCG)$_2$ (−7.7°; Wang *et al.*, 1981), but closer to that in HAT61/120 (−13.9°; Harper *et al.*, 1998). This may be a consequence of the presence of the terminal A·T base pair. Compared with HAT61/120, the present duplex contracts slightly at C5A6:T7G8 but expands slightly at all other base steps. The value of rise at the virtual step between the symmetry-related duplexes is only 3.14 Å in HAT61 compared with 3.70 Å in HAT61/120. The net effect of these changes is to increase the length of the unit-cell *c* axis by 0.11 Å in the present structure. In the base-pair parameters there are small variations that appear to fall into a pattern of sequence dependence. For example, the A·T base pairs at the two termini and the C5·G8 base pair are slightly more buckled than the others. This pattern is also seen in the low-temperature structure, but not in d(CGCGCG)$_2$. Similarly, the T1·A12 base pair has a low value for the inclination in both HAT61 and HAT61/120, but is higher in d(CGCGCG)$_2$. For some of the parameters such as *x* displacement, however, the values in HAT61 are different from those in

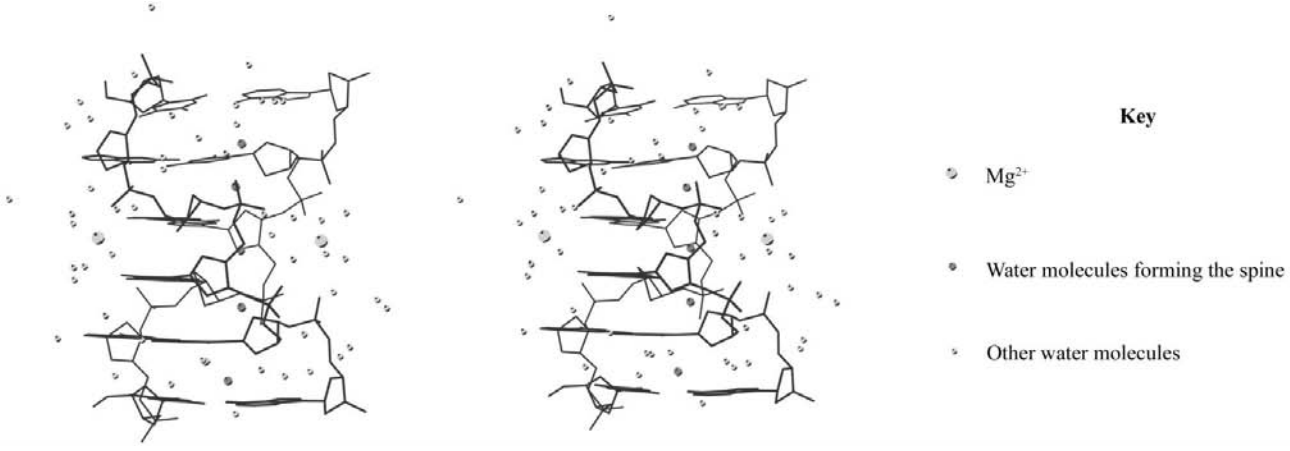

Figure 2
Solvent molecules and ions in HAT61, plotted using *BOBSCRIPT* (Esnouf, 1997)

both HAT61/120 and d(CGCGCG)$_2$.

3.3. Solvent and cation interactions

58 water molecules and two magnesium hexahydrate ions were located in the structure (Fig. 2). No cobalt hexammine could be located, although the crystallizing solution contained 150 m*M* cobalt hexammine chloride. The structure of HAT61/120 was solved from crystals grown from solutions without any magnesium, but containing cobalt hexammine and arginamide chloride, and showed two cobalt hexammine ions instead of the magnesium ions in the present structure. Magnesium *A* in HAT61 is shifted 0.53 Å away from the helix compared with the corresponding cobalt ion in HAT61/120. The position of the other ion *B* is substantially different in HAT61 and HAT61/120. It is shifted 1.68 Å closer to the helix in the present structure. Both magnesium ions make water-mediated contacts with the atoms of the DNA helix and with other water molecules. Several attempts were made during the course of the present work to identify the ions as cobalt, but each time this was performed, the refinement proceeded badly, with a rise in the values of the *R* factor (from 21.6 to 22.0%) and $R_{\rm free}$ (from 26.7 to 27.5%). Also, the cobalt ions showed higher thermal factors (54 and 36 Å^2, respectively) than the corresponding magnesium ions (20 and 13 Å^2, respectively). Clearly, in the present structure hydrated magnesium has displaced cobalt hexammine.

The solvent structure of Z-DNA is mainly characterized by a spine of hydration (Wang *et al.*, 1981) and this is completely present in HAT61 (Fig. 2). On the other hand, in HAT61/120 this spine is disrupted in the middle. The rest of the water structure in HAT61 is not significantly different from that in HAT61/120 and in other Z-DNA structures.

To summarize the results, the comparison of the present room-temperature structure with the low-temperature structure of the same sequence shows that higher temperature and the presence of Mg^{2+} in the crystallization solution of HAT61 has an impact on some of the details of the conformation. This is particularly true of the phosphate structure at C11. However, comparison with d(CGCGCG)$_2$ reveals that differences in the structure are specified more by the sequence than by the temperature and by the presence of different metal ions, since the structures of HAT61 and HAT61/120 are closer to each other than to the structure of d(CGCGCG)$_2$[1].

This work was supported by funds to the department under the DST-FIST and the UGC-SAP programmes of the Government of India. ST and PS thank CSIR, India for SRF. The National Area Detector Facility at MBU, IISc, Bangalore, India is acknowledged.

References

Bancroft, D., Williams, L. D., Rich, A. & Egli, M. (1994). *Biochemistry*, **33**, 1073–1086.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: VJ0050). Services for accessing these data are described at the back of the journal.

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.
Brünger, A. T., Karplus, M. & Petsko, G. A. (1989). *Acta Cryst.* A**45**, 50–61.
Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Coll, M., Fita, I., Lloveras, J., Subirana, J. A., Bardella, F., Huynh-Dinh, T. & Igolen, J. (1988). *Nucleic Acids Res.* **16**, 8695–8705.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Crawford, J. L., Kolpak, F. J., Wang, A. H.-J., Quigley, G. J., Van Boom, J. H., Van Der Marel, G. A. & Rich, A. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 4016–4020.
Dickerson, R. E. (1998). *Nucleic Acids Res.* **26**, 1906–1926.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Gessner, R. V., Quigley, G. J., Wang, A. H.-J., Van Der Marel, G. A.,Van Boom, J. H. & Rich, A. (1985). *Biochemistry*, **24**, 237–240.
Harper, N. A., Brannigan, J. A., Buck, M., Lewis, R. J., Moore, M. H. & Schneider, B. (1998). *Acta Cryst.* D**54**, 1273–1284.
Ho, P. S. & Mooers, B. H. M. (1997). *Biopolymers Nucleic Acid Sci.* **44**, 65–90.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Pohl, F. M., Jovin, T. M., Baehr, W. & Holbrook, J. J. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 3805–3809.
Sadasivan, C. & Gautham, N. (1995). *J. Mol. Biol.* **248**, 918–930.
Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., Van der Marel, G. & Rich, A. (1979*a*). *Nature (London)*, **282**, 680–686.
Wang, A. H.-J., Quigley, G. J., Kolpak, van Boom, J. H., Van der Marel, G. & Rich, A. (1981) *Science*, **211**, 171–176.
Wang, A. H.-J. & Teng, M. K. (1988). *J. Cryst. Growth*, **90**, 295–310.