Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of d(Gm^{5}CGm^{5}CGCGC) and d(GCGCGm^{5}CGm^{5}C): effects of methylation on alternating DNA octamers

Ke Shi, Baocheng Pan, Douglas Tippin and Muttaiya Sundaralingam*

Departments of Chemistry and Biochemistry, The Ohio State University, 200 Johnston Laboratory, 176 West 19th Avenue, Columbus, OH 43210, USA

Correspondence e-mail: sundaral@chemistry.ohio-state.edu

Received 17 June 2003
Accepted 2 October 2003

PDB References:
d(Gm5CGm5CGCGC), 1r3z;
d(GCGCGm5CGm5C), 1r41.

The crystal structures of two alternating octamers d(Gm^{5}CGm^{5}CGCGC) and d(GCGCGm^{5}CGm^{5}C) have been determined at 1.7 and 1.9 Å resolution with R_{work}/R_{free} of 16.9/19.6% and 17.9/19.8%, respectively. The two octamers crystallized in the same tetragonal space group $P4_32_12$, with unit-cell parameters $a = b = 43.0$, $c = 24.9$ Å and $a = b = 43.1$, $c = 24.7$ Å, respectively. Both structures adopt the A-type double helix and the abutting crystal packing. Despite the different positions of the methylated cytosine in the sequence, the two octamers display essentially the same conformation. The close similarity of the two structures indicates that methylation at different positions in a DNA sequence does not affect the helical conformation of the alternating DNA octamer. Investigation of other related crystal structures implies that methylation of cytosine stabilizes the A-type conformation for alternating DNA sequences beginning with a 5′-purine but is beneficial to the Z-type conformation for alternating DNA sequences beginning with a 5′-pyrimidine.

1. Introduction

Methylation of cytosine in DNA is involved in many biological processes such as control of cellular differentiation, gene regulation, gene expression and X-chromosome inactivation in eukaryotic organisms (Antequera & Bird, 1993; Singer-Sam & Riggs, 1993; Sasaki *et al.*, 1993). The methylation may exert its effects through alterations in local chromatin structure and subsequent modulation of protein binding. It also plays an important role in the crystallization of DNA molecules. Experimental results have shown that methylation can stabilize the A-type (Mooers *et al.*, 1995; Frederick *et al.*, 1987) and Z-type conformations (Fujii *et al.*, 1982; Wang *et al.*, 1984; Schroth *et al.*, 1993). Methylation of cytosine can convert the Z-type helix d(GCGCGCGCGC) (Ban *et al.*, 1996) to the A-type conformation (Tippin *et al.*, 1997) and induce stable intermediates in the B-type to A-type transition (Vargason *et al.*, 2000) and in the helical transition from B-type to A-type (Vargason *et al.*, 2001).

Alternating DNA oligomers containing only guanines and cytosines display quite different potentials to form helical structures. Crystallographic studies have shown that $(CG)_n$ oligomers (n = 2, 3, 4) adopt the Z-type helix (Drew *et al.*, 1980; Wang *et al.*, 1979; Fujii *et al.*, 1985). The preference for the Z-DNA structure is attributed to the greater stacking stability of the n CpG steps compared with the $(n - 1)$ GpC steps (Jain *et al.*, 1987). In contrast, alternating sequences beginning with a 5′-purine, $d(GC)_n$, may adopt A-type and Z-type conformations. Circular dichroism (CD) studies have demonstrated that $d(GC)_n$ sequences have a left-handed

Table 1
Crystal data and refinement parameters of the A-DNA octamer.

Values in parentheses are for the highest resolution shell, 1.78–1.70 Å for $d(G^mCG^mCGCGC)_2$ and 1.99–1.90 Å for $d(GCGCG^mCG^mC)_2$.

	$d(G^mCG^mCGCGC)_2$	$d(GCGCG^mCG^mC)_2$
Space group	$P4_32_12$	$P4_32_12$
Unit-cell parameters (Å)	$a = b = 43.0$, $c = 24.9$	$a = b = 43.1$, $c = 24.7$
Asymmetric unit	1 strand	1 strand
Volume per base pair (Å^3)	1435	1437
Resolution range (Å)	10.0–1.7	10.0–1.9
No. reflections [$F > 2\sigma(F)$]	2540	1972
No. reflections (all data)	2711	2192
Completeness (%) [$F > 2\sigma(F)$]	90.0 (80.5)	86.5 (84.8)
R_{sym} (%)	8.1 (20.0)	8.7 (30.9)
R_{work}/R_{free} (%) [$F > 2\sigma$(F)]	16.9/19.6 (25.0/29.1)	17.9/19.8 (34.3/33.2)
R_{work}/R_{free} (%) (all data)	19.3/21.7 (26.2/31.3)	20.4/21.9 (35.1/37.6)
Final model		
Nucleic acid atoms	163	163
Water O atoms	25	19
Deviations from ideal geometry		
Bond lengths (Å)	0.020	0.048
Bond angles (°)	2.5	5.3

Z-type helix for $n > 7$, but a right-handed B-type helix for n = 3–7 (Quadrifoglio *et al.*, 1984). However, methylated $d(GC)_3$ adopts the A-type helix (Mooers *et al.*, 1995), while $d(GC)_5$ adopts the Z-type helix in the crystalline state (Ban *et al.*, 1996). These results are different from those of CD studies and may indicate differing environments in the crystalline state and solution. On the other hand, unmethylated $d(GC)_3$ could not be crystallized (Mooers *et al.*, 1995) and it is not yet known which helical form the hexamer adopts in the crystalline state. In this paper, we intended to determine the helical form of the sequence $d(GC)_4$ in the crystalline state, the role of methylation in DNA crystals and the influence of methylated cytosine at different positions in a sequence. The octamer d(GCGCGCGC) and its methylated sequences $d(Gm^5CGm^5CGCGC)$ (called 5GmC) and d(GCGC-Gm^5CGm^5C) (called 3GmC) were synthesized and crystallized. The unmethylated octamer diffracted very poorly and only yielded a few diffraction spots. The 5GmC and 3GmC duplexes diffracted to 1.7 and 1.9 Å resolution, respectively. Here, we describe the crystal structures of the two methylated duplexes and discuss the effects of methylation in alternating DNA oligomers.

2. Materials and methods

2.1. Synthesis, crystallization and data collection

The DNA octamers 5GmC and 3GmC were synthesized on our in-house ABI 391 synthesizer and were cleaved from the solid support using ammonia. The solutions were then incubated at 328 K overnight. They were precipitated with ethanol in the presence of 2.5 *M* ammonium acetate at 248 K and were purified on an ion-exchange column using FPLC. The samples were lyophilized and then crystallized by the hanging-drop vapor-diffusion method at room temperature. The two octamers crystallized under the same conditions. Crystals were grown using the following conditions: 1 m*M* DNA (single-stranded concentration), 25 m*M* sodium cacodylate buffer pH 6.0, 1 m*M* magnesium chloride, 0.5 m*M* spermine tetrachloride. The droplets were equilibrated against 500 µl 40% MPD in the reservoir. Crystals appeared in one week and grew to dimensions of 0.3 × 0.3 × 0.2 mm in a further week. The crystals were mounted in a thin-walled glass capillary and sealed with wax. The intensity data were collected at room temperature on our R-AXIS IIC imaging plate with a graphite-monochromated Cu $K\alpha$ X-ray beam at 50 kV and 100 mA. The crystal-to-detector distance was 70 mm and the angle 2θ was set to 0°. 15 oscillation frames with 4° φ-scan width yielded 2540 independent reflections [$F \geq 2.0\sigma(F)$] to 1.7 Å resolution for the 5GmC duplex and 1972 independent reflections [$F \geq 2.0\sigma(F)$] to 1.9 Å resolution for the 3GmC duplex. The data were processed using version 2.1 of the software supplied by Molecular Structure Corporation and the crystal data are given in Table 1. The completeness of the data in different resolution ranges is given in the supplementary material.[1]

2.2. Structure solution and refinement

The two octamer duplexes are isomorphous to d(GTGTACAC) (Jain *et al.*, 1989) and therefore its coordinates were used in molecular replacement for the 5GmC duplex and the 3GmC duplex. 10% of the data were randomly selected for R_{free} calculation in order to monitor the refinement process. The refinements were performed with *CNS* (Brünger *et al.*, 1998). Initial rigid-body refinement of the 5GmC duplex gave an R_{work} of 0.223 using 1489 reflections in the resolution range 10–2.5 Å. Positional refinement using all 2540 reflections to 1.7 Å gave an R_{work} of 0.323. A difference map was calculated and the bases were changed to the correct sequence with the program *CHAIN* (Sack & Quiocho, 1992). Further positional and individual *B*-factor refinement with the correct sequence lowered R_{work} to 0.280. Next, simulated annealing was performed in order to remove conformational bias of the starting model by heating the system to 3000 K and cooling slowly to 300 K in steps of 0.5 fs time interval, yielding an R_{work} of 0.235. Water molecules were added according to $|F_o - F_c| > 3\sigma$ and $|2F_o - F_c| > \sigma$ electron-density maps. Refinement with 25 water molecules gave final R_{work} and R_{free} values of 0.169 and 0.196, respcetively. The structure of the 3GmC duplex was refined similarly and the final R_{work} and R_{free} values were 0.179 and 0.198, respectively. The R_{work} and R_{free} in different resolution ranges are listed in the supplementary material.[1] The refinement statistics are summarized in Table 1.

3. Results and discussion

3.1. Overall conformation

Both the 5GmC and 3GmC duplexes crystallized in the tetragonal space group $P4_32_12$ with one strand in the asym-

[1] Supplementary data have been deposited in the IUCr electronic archive (Reference: AD5004). Details for accessing these data are described at the back of the journal.

Table 2
Helical parameters of present structures.

	d(Gm^{5}CGm^{5}CGCGC)			d(GCGCGm^{5}CGm^{5}C)		
	Twist (°)	Rise (Å)	Roll (°)	Twist (°)	Rise (Å)	Roll (°)
G1–C2†	34	2.6	6	35	3.2	−2
C2–G3	30	3.3	6	31	3.2	16
G3–C4	37	3.2	4	38	3.1	6
C4–G5	22	3.2	0	17	3.5	−1

† The values for the symmetry-related half of the duplex are the same and therefore are not listed.

metric unit. The strand and its symmetry-related strands form a right-handed A-type duplex with all-Watson–Crick G·C base pairs (Fig. 1). The DNA duplexes display a wide and flat minor groove, with an average width of 9.5 Å. There is only one measurement of the shortest P—P distance across the major groove. This distance is 15.1 Å, corresponding to a 9.3 Å groove width, which is much wider than the usual 4.6 Å for the A-DNA helix. Careful study of A-DNA duplex structures indicates that this may not be an accurate measurement of the major-groove width. Investigation of A-DNA decamers and dodecamers indicates that the major-groove widths are usually determined by the P—P distance that is seven base pairs apart, namely (i + 7). In other words, the major-groove width can usually only be measured accurately for nanomer or longer duplexes.

The torsion angles and helical parameters were calculated using *CURVES* (Lavery & Sklenar, 1989) and are shown in Table 2. The helical twist for the 5′-purine-–pyrimidine-3′ step is low and that for the 5′-pyrimidine–purine-3′ step is high, which is the opposite to that observed in B-DNA alternating sequences (Chen *et al.*, 1994, 1997; Mitra *et al.*, 1999). In other words, the helical twist is not only related to base-step identity but also to the helical form. Even though the sequences alternate between guanine and cytosine, only the helical twist displays an obvious pattern of 'high–low' alternation. This is quite different from the alternating B-DNA duplex (Chen *et al.*, 1994, 1997; Mitra *et al.*, 1999). All the nucleotides adopt the usual values of the torsion angles for an A-type duplex except for G5 in both duplexes and m^{5}C8 in the 3GmC duplex. G5 adopts a *trans*/*trans* conformation for the torsion angles O5′—P (α) and C4′—C5′(γ) instead of the usual g^-/g^+ conformation, which results in low twist angles for base step C4–G5 (Table 2). m^{5}C8 in the 3GmC duplex adopts the g^+/g^- conformation for O5′—P (α) and C4′—C5′ (γ), the opposite of the usual g^-/g^+ conformation. The average helical twist and rise are 32° and 3.1 Å for the 5GmC duplex and 32° and 3.2 Å for thet 3GmC duplex, respectively, corresponding to 11.3 base pairs per turn. All the sugar puckers belong to the C3′-*endo* conformation except for the 3′-terminal nucleotides, which adopt the C2′-*endo* conformation in the 5GmC duplex and the C4′-*endo* conformation in the 3GmC duplex.

3.2. Crystal packing

The DNA duplexes are clustered about the 4_3 screw axes in the $P4_32_12$ unit cell, forming elliptically shaped solvent channels (10 × 20 Å) along the 2_1 screw axes (Fig. 2). Both the 5GmC and the 3GmC duplexes adopt the abutting crystal packing, with the terminal base pairs abutting into the middle of the minor groove of the adjacent duplexes. In 5GmC duplex, there is one hydrogen bond between N2(G5) and O4′(G1#) (3.0 Å) and a water bridge connecting N3(5) and O5′(G1#) in the lattice interface (where # indicates the adjacent duplex). In addition to the hydrogen-bonding interactions, there is a C—H···O interaction between C1′(G1) and O2(C4#) (3.2 Å). Similar results have been observed in the 3GmC duplex. This packing mode resembles those observed in other A-DNA octmers (Wahl & Sundaralingam, 1997).

3.3. Hydration

Of the 25 water molecules in the 5GmC duplex, 11 hydrate phosphate groups. In contrast, only three of the 19 water molecules in 3GmC duplex hydrate phosphate groups. Some

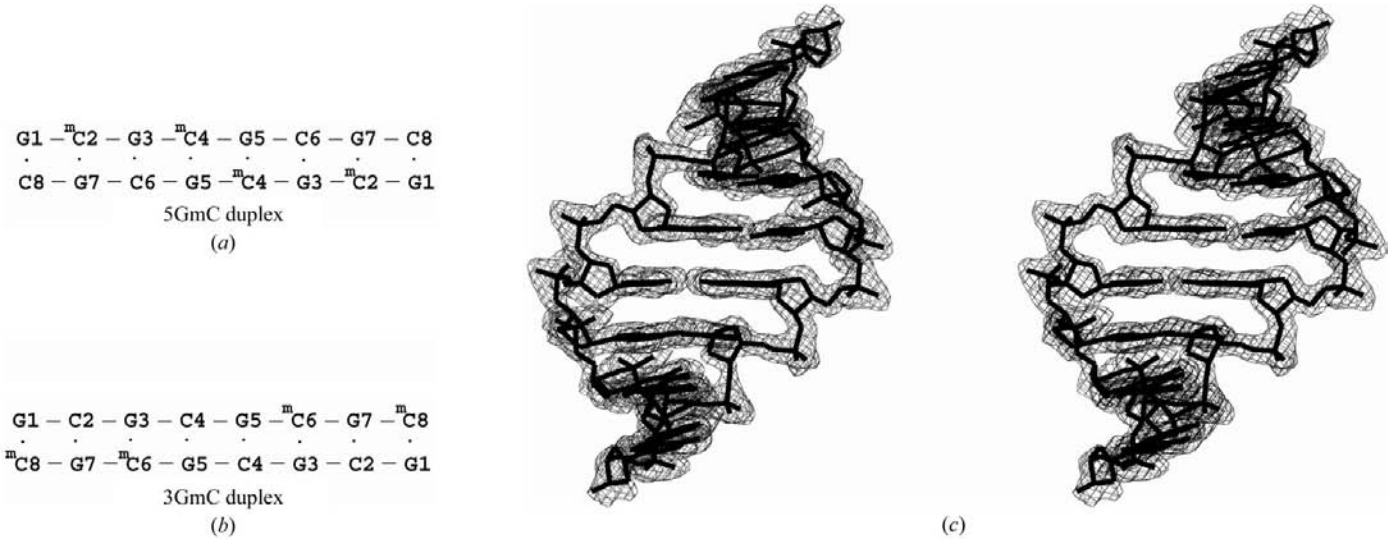

Figure 1
(*a*) and (*b*) Numbering schemes of the two present structures. (*c*) A stereoview of the 5GmC duplex superposed with the $|2F_o - F_c|$ map at 1σ contours.

short water bridges connecting the adjacent phosphate groups have been observed. However, continuous water bridges along the backbone were not observed in either duplex. The hydration of the bases is listed in Table 3. The major groove is much more heavily hydrated than the minor groove because of the abutting crystal packing in the minor groove. Owing to the right-handed A-DNA conformation, water molecules interacting with O6 atoms of guanines in the adjacent base steps are close enough to form water bridges between these guanines. Comparison between the hydration in the major groove of methylated cytosine and the corresponding unmethylated cytosine indicates that methylation of cytosine does not affect the hydration in the major groove of the present octamer duplex. This result may imply that an H atom and a methyl group at the C5 position of cytosine have a similar behavior in interaction with water molecules. The hydration features observed here are quite similar to those analyzed in A-DNA octamers (Eisenstein & Shakked, 1995).

Table 3
Hydration of the major groove and minor groove of the present structures.

d(Gm^{5}CGm^{5}CGCGC)$_2$				d(GCGCGm^{5}CGm^{5}C)$_2$			
Residue	Atom	H_2O	Distance (Å)	Residue	Atom	H_2O	Distance (Å)
Major groove							
G1	O6	W106	2.7	G1	O6	W113	2.7
	N7	W107	2.8		N7	W114	2.8
m^{5}C2				C2			
G3	O6	W106	2.6	G3	O6	W104	2.6
	N7	W109	2.8		N7	W103	2.6
m^{5}C4	N4	W125	3.1	C4	N4	W110	2.8
G5	O6	W105	2.4	G5	O6	W106	2.6
					N7	W105	2.7
C6	N4	W110	3.4	m^{5}C6	N4	W125	3.3
G7	O6	W110	2.6	G7	O6	W111	2.5
	N7	W118	2.7		N7	W102	2.7
C8	N4	W111	2.9	m^{5}C8	N4	W112	2.9
Minor groove							
m^{5}C2	O2	W115	2.7				
G5	N3	W102	2.9	G5	N3	W101	2.9
G7	N2	W101	3.3	G7	N2	W101	3.2
	N3	W101	2.8		N3	W101	2.8
				m^{5}C8	O2	W117	3.3

Table 4
Crystal structures with alternating DNA sequences containing 5-methylated cytosine.

NDB code	Sequence	Helical form	Reference
ZDFB06	d(m^{5}CG TA m^{5}CG)	Z	Wang *et al.* (1984)
ZDFB24	d(m^{5}CG UA m^{5}CG)	Z	Zhou & Ho (1990)
ZDFB03	d(m^{5}CG m^{5}CG m^{5}CG)	Z	Fujii *et al.* (1982)
ADFB72	d(Gm^{5}CGCG C)	A	Mooers *et al.* (1995)
ADFB62	d(Gm^{5}CG m^{5}CG C)	A	Mooers *et al.* (1995)
	d(Gm^{5}CG m^{5}CG CGC)	A	Present study
	d(GCGCGm^{5}CG m^{5}C)	A	Present study
ADJB88	d(Gm^{5}CG m^{5}CG CGCGC)	A	Tippin *et al.* (1997)

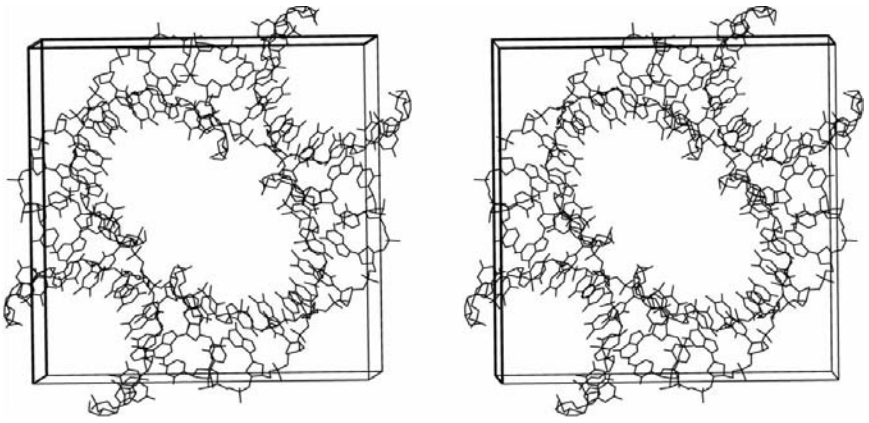

Figure 2
A stereoview of the crystal packing of the 5GmC duplex in the unit cell.

3.4. Comparison of the two structures

The two present structures have the same sequence with methylated cytosines at different positions of the sequence. These two structures possess the same space group and display the same crystal packing and inter-duplex interactions. Comparison between these two structures may reveal an effect upon DNA conformation caused by the different positions of the methylated cytosine in the sequence. Superposition of the two structures gives an r.m.s.d. of 0.2 Å, with some differences in the phosphate and sugar backbone of the 3′-terminal cytosine (Fig. 3). These results imply that methylation in different positions does not affect the DNA structure or crystal packing in the present octamer A-form helix.

3.5. Factors affecting the structures of alternating sequences

Even though the two methylated sequences of d(GCGCGCGC) adopt the A-form helix, we do not yet know the helical form of the unmethylated sequence. The present results suggest two possibilities: (i) d(GCGCGCGC) adopts the Z-form helix and methylation converts the helix from the Z-form to the A-form, as in the case of d(GCGCGCGCGC) (Ban *et al.*, 1996; Tippin *et al.*, 1997), or (ii) d(GCGCGCGC) adopts the A-form helix and methylation only improves the resolution as observed in d(GCGCGC) (Mooers *et al.*, 1995). Further research on the unmethylated sequence d(GCGCGCGC) may solve this ambiguity. No matter which helical form d(GCGCGCGC) adopts, the present study indicates that a methylated alternating sequence beginning with 5′-purine adopts the A-form regardless of the positions of the 5-methylated cytosine in the sequence. This result is consistent with the effects of methylation on the hexamer d(GCGCGC) (Mooers *et al.*, 1995).

DNA molecules are flexible and dynamic. Their conformations are dependent on molecular subtlety. Methylation of cytosine is usually believed to stabilize a Z-form helix (Behe & Felsenfeld, 1981; Fujii *et al.*, 1982; Schroth *et al.*, 1993). It was unexpected that methylation of cytosine would convert the Z-DNA duplex d(GCGCGCGCGC) (Ban *et al.*, 1996) to the A-form (Tippin *et al.*, 1997). Also, methylation of cytosine resulted in a helical transition from the B-form to the A-form (Vargason *et al.*, 2000, 2001). These

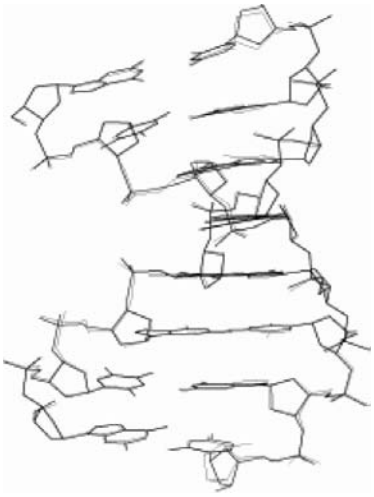

Figure 3
A view of the superposition of all atoms of the two present structures. The 5GmC duplex is shown in thick lines and the 3GmC duplex in thin lines.

results indicate that the role of methylation is not as simple as we had previously expected. Further studies are needed to investigate its role in helical structures. We noticed the fact that methylation of cytosine has a tremendous effect in promoting crystallization in some DNA sequences. For instance, d(CGATCG) could not be crystallized by itself but the methylated sequence readily crystallized in the Z-form helix (Wang *et al.*, 1984). Similarly, only methylated sequences of d(GCGCGC) could be crystallized in the A-form helix and a doubly methylated sequence yielded higher resolution than the singly methylated sequence (Mooers *et al.*, 1995). Table 4 lists all the published reports of alternating DNA crystal structures containing 5-methylated cytosine. It is clear from the table that methylation of cytosine in an alternating sequence beginning with a 5′-pyrimidine stabilizes the Z-form helix, while methylation of cytosine in an alternating sequence beginning with a 5′-purine stabilizes the A-form.

We gratefully thank the NIH (grant GM-17378) and an Ohio Regents Eminent Scholar Endowment for supporting this work. We also acknowledge partial support for the purchase of an R-AXIS IIc imaging plate from the Ohio Regents Investment Fund and Ohio Regents Eminent Scholar Award to MS.

References

Antequera, F. & Bird, A. (1993). *DNA Methylation: Molecular Biology and Biological Significance*, edited by J. P. Jost & H. P. Saluz, pp. 169–185. Boston: Birkhauser Verlag.

Ban, C., Ramakrishnan, R. & Sundaralingam, M. (1996). *Biophys. J.* **71**, 1215–1221.

Behe, M. & Felsenfeld, G. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 1619–1623.

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.

Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994). *Nature Struct. Biol.* **1**, 169–175.

Chen, X., Ramakrishnan, B. & Sundaralingam, M. (1997). *J. Mol. Biol.* **267**, 1157–1170.

Drew, H., Takano, T., Itakura, K. & Dickerson, R. E. (1980). *Nature (London)*, **286**, 567–573.

Eisenstein, M. & Shakked, Z. (1995). *J. Mol. Biol.* **248**, 662–678.

Frederick, C. A., Saal, D., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. (1987). *Biopolymers*, **26**, S145–S160.

Fujii, S., Wang, A. H.-J., Quigley, G. J., Westerink, H., van der Marel, G. A., van Boom, J. H. & Rich, A. (1985). *Biopolymers*, **24**, 243–250.

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H. & Rich, A. (1982). *Nucleic Acids Res.* **10**, 7879–7892.

Jain, S. C., Zon, G. & Sundaralingam, M. (1987). *J. Mol. Biol.* **197**, 141–145.

Jain, S. C., Zon, G. & Sundaralingam, M. (1989). *Biochemistry*, **28**, 2360–2364.

Lavery, R. & Sklenar, H. (1989) *J. Biomol. Struct. Dyn.* **4**, 655–667.

Mitra, S. N., Wahl, M. C. & Sundaralingam, M. (1999). *Acta Cryst.* D**55**, 602–609.

Mooers, B. H., Schroth, G. P., Baxter, W. W. & Ho, P. S. (1995). *J. Mol. Biol.* **249**, 772–784.

Quadrifoglio, F., Manzini, G. & Yathindra, N. (1984). *J. Mol. Biol.* **175**, 419–423.

Sack, J. & Quiocho, F. A. (1992). *CHAIN: Crystallographic Modeling Program.* Baylor College of Medicine, Houston, TX, USA.

Sasaki, H., Allen, N. D. & Surani, M. A. (1993). *DNA Methylation: Molecular Biology and Biological Significance*, edited by J. P. Jost & H. P. Saluz, pp. 469–486. Boston: Birkhauser Verlag.

Schroth, G. P., Kagawa, T. F. & Ho, P. S. (1993). *Biochemistry*, **32**, 13381–13392.

Singer-Sam, J. & Riggs, A. D. (1993). *DNA Methylation: Molecular Biology and Biological Significance*, edited by J. P. Jost & H. P. Saluz, pp. 358–384. Boston: Birkhauser Verlag.

Tippin, D. B., Ramakrishnan, B. & Sundaralingam, M. (1997). *J. Mol. Biol.* **270**, 247–258.

Vargason, J. M., Eichman, B. F. & Ho, P. S. (2000). *Nature Struct. Biol.* **7**, 758–761.

Vargason, J. M., Henderson, K. & Ho, P. S. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 7265–7270.

Wahl, M. C. & Sundaralingam, M. (1997). *Biopolymers*, **44**, 45–63.

Wang, A. H.-J., Hakoshima, T., van der Marel, G. A., van Boom, J. H. & Rich, A. (1984). *Cell*, **37**, 321–331.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A. & Rich, A. (1979). *Nature (London)*, **282**, 680–686.

Zhou, G. & Ho, P. S. (1990). *Biochemistry*, **29**, 7229–7236.