Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystal structure of rGd(CGCGCG): a Z-DNA hexamer duplex with a 5′-(rG) overhang

Baocheng Pan and Muttaiya Sundaralingam*

Departments of Chemistry and Biochemistry, 200 Johnston Laboratory, 176 West 19th Avenue, Columbus, Ohio 43210-1002, USA

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

Received 22 July 2002
Accepted 9 December 2002

PDB Reference:
rGd(CGCGCG), 1m6r, r1m6rsf.

The crystal structure of the heptamer rGd(CGCGCG) has been determined at 1.54 Å resolution with R_{work} and R_{free} of 0.191 and 0.235, respectively. The crystal belongs to space group $P2_12_12_1$, with unit-cell parameters a = 17.96, b = 31.47, c = 44.73 Å and two independent strands in the asymmetric unit. The chimera forms a Z-DNA hexamer duplex d(CGCGCG)$_2$, with the 5′-overhang rG invisible in the density. The replacement of rG in the 5′ terminus of d(GCGCGCG) changes the reverse Hoogsteen G·G base pairing of the 5′-overhang dG (Pan *et al.*, 1997). However, the replacement does not change the helix from Z-form to A-form or the water-structure motifs in the Z-DNA double-helical structure.

1. Introduction

DNA molecules are polymorphous and their conformations depend on the sequence and also the environment, such as hydration and interaction with ions. Alternating oligonucleotides beginning with a 5′-purine, such as d(GCGCGC), have a tendency to crystallize in the A-form (Jain *et al.*, 1987; Mooers *et al.*, 1995), while those beginning with a 5′-pyrimidine, such as d(CGCGCG), crystallize in the Z-form (Wang *et al.*, 1979). The helical forms of DNA can change with a change in environment. It has been shown that B-form helices can be converted to the A-form upon dehydration (Franklin & Gosling, 1953) and to the Z-form in solution upon addition of high salt concentrations (Pohl & Jovin, 1972). DNA can also change its helical forms in the crystalline state. DNA can be changed from B-form to A-form when one or more residues are substituted by ribonucleotides (Ban *et al.*, 1994*a*,*b*; Wahl & Sundaralingam, 2000). However, Z-form DNA displays quite different behavior to B-form DNA. Even though Z-form DNA can be destabilized by AT tracts (Wang *et al.*, 1984), the replacement with r(CG) in the center of the hexamer d(CG**CG**CG) does not convert the DNA to A-form (Teng *et al.*, 1989). This may result from the inherent stability of Z-form DNA, which is characterized by alternating glycosyl conformations (*anti* and *syn*) and sugar puckers (C2′-*endo* and C3′-*endo*).

Overhang ribonucleotides have been shown to stabilize the double-helical conformation of oligomers (Martin *et al.*, 1971; Romaniuk *et al.*, 1978; Petersheim & Turner, 1983; Freier *et al.*, 1985; Sugimoto *et al.*, 1987). They have also been suggested to be important in determining the stability of codon–anticodon associations (Grosjean *et al.*, 1976; Ayer & Yarus, 1986). The stability contributed by overhang residues arises mainly from their base stacking in the helical junction (Burkard *et al.*, 1999). Recently, we have determined the crystal structure of d(GCGCGCG), which adopts the Z-form hexamer with

5′-overhang dG residues that are internally stacked in the helix and form a reverse Hoogsteen G·G base pairing (Pan *et al.*, 1997). This result indicates that the Z-form helical conformation dominates over the A-form in the crystalline state for this sequence. In this paper, we designed the heptamer gCGCGCG (where the lower case letter represents RNA, while the upper case letters represent DNA) to study the effects of g upon the conversion of helical forms. The sequence can adopt an A-form duplex $(gCGCGC)_2$ with a 3′-overhang G (Fig. 1*a*) or a Z-form duplex $(CGCGCG)_2$ with a 5′-overhang g (Fig. 1*b*). As observed in the previous crystal structures (Ban *et al.*, 1994*a*,*b*; Wahl & Sundaralingam, 2000), substitution of g in the 5′-terminus of the heptamer (GCGCGCG) may increase the tendency of the sequence to form the A-form helix.

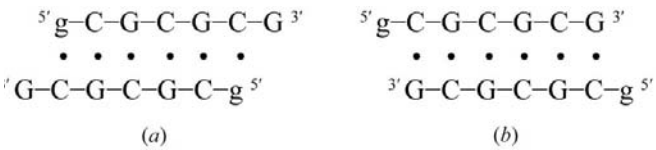

Figure 1
(*a*) A-form duplex $(gCGCGC)_2$, (*b*) Z-form duplex $(CGCGCG)_2$.

2. Materials and methods

2.1. Synthesis, crystallization and data collection

The DNA–RNA chimera heptamer gCGCGCG was synthesized by the phosphoramidite method using an in-house Applied Biosystems 391 DNA synthesizer (Foster City, CA, USA). The coupling of the bases to the existing oligomer was checked by visual monitoring of the orange color during the elution of the trityl group. To ensure good coupling of ribonucleotide to the oligomer, the coupling time was 10 min instead of the 15 s used for deoxyribonucleotides. The deep orange color in the elution of the trityl group removed from the 5′-g ensured the incorporation of the 5′-terminal g into the sequence. The sample was purified by ion-exchange FPLC (Wahl, Ramakrishnan *et al.*, 1996). Crystallization was carried out by the hanging-drop vapor-diffusion method at room temperature. Suitable crystals were obtained with 1 m*M* chimera (single-stranded concentration) in the presence of 40 m*M* sodium cacodylate buffer pH 6.5, 20 m*M* $MgCl_2$ and 1 m*M* spermine tetrachloride equilibrated against a reservoir of 0.2 ml of 10% 2-propanol. A crystal of dimensions 0.2 × 0.3 × 0.3 mm was mounted in a thin-walled glass capillary with some mother liquor at one end. The intensity data were collected at room temperature using our R-AXIS IIc imaging plate and graphite-monochromated Cu $K\alpha$ X-ray beam and were processed using Version 2.1 of the software from Molecular Structure Corporation (Rigaku Co.). The crystal-to-detector distance was 55 mm and 26 frames of data were collected with 2° framewidth and 20 min exposure time, giving 12 118 reflections $[F > \sigma(F)]$ with 3576 independent reflections to 1.54 Å resolution. The data set has an $R_{\rm merge}$ of 4.9% (19.5%) and 87.1% (79.9%) comple-

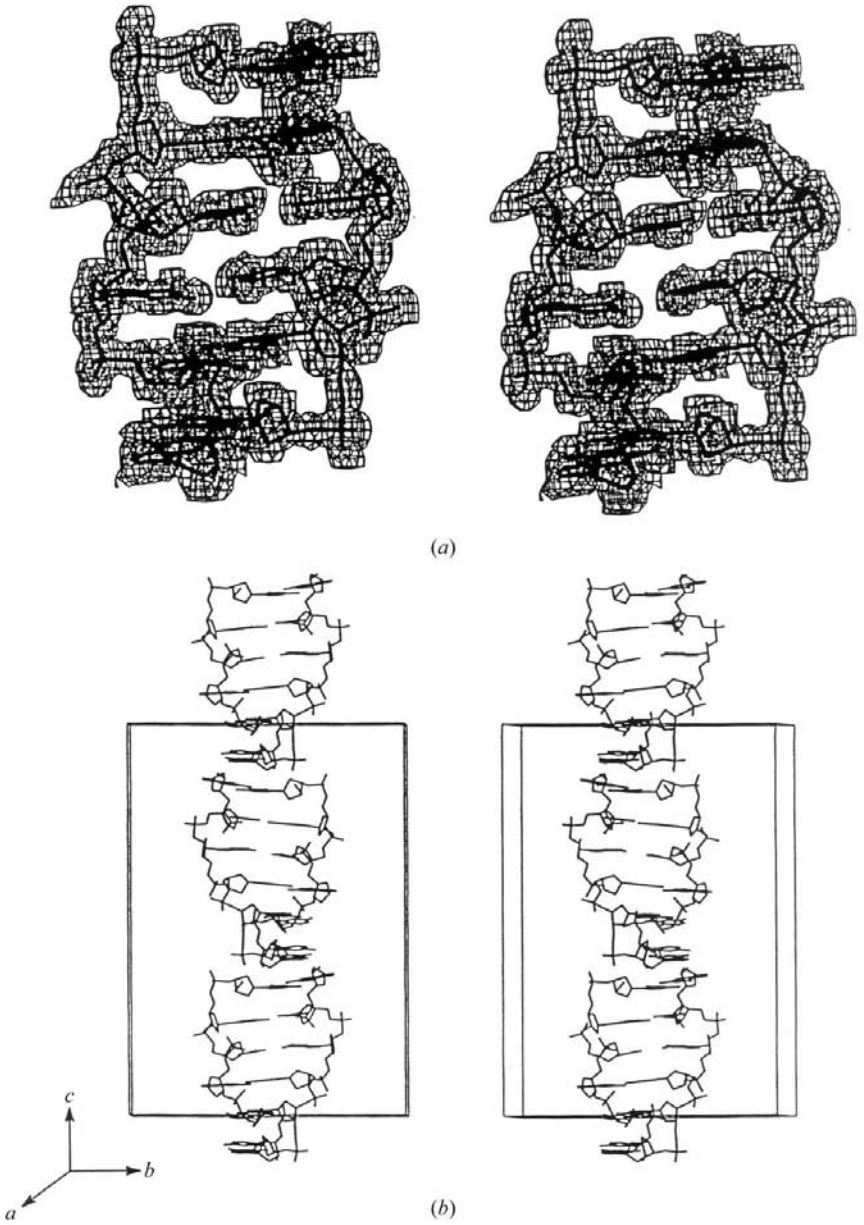

Figure 2
(*a*) Stereoview of the final duplex structure of gCGCGCG superposed on the $|2F_o - F_c|$ density map at the 1σ level. (*b*) Stereoview of the unit-cell packing of three duplexes along the *c* axis. Note the absence of the electron density for g.

teness, where the values in parentheses correspond to the highest resolution shell (1.65–1.54 Å). The crystal data and refinement parameters are summarized in Table 1.

2.2. Structure refinement

Because the crystal has the same space group as and similar unit-cell parameters to the Z-DNA hexamer (CGCGCG) (a = 17.87, b = 31.55, c = 44.58 Å; Wang *et al.*, 1979), the coordinates of the hexamer Z-DNA were used as the starting model for structural refinement. Refinement was performed using the *X-PLOR* program (Brünger, 1992) with 3536 reflections [$F > 2.0\sigma(F)$] in the resolution range 8.0–1.54 Å. 10% of the reflections (352 reflections) were randomly selected for the $R_{\rm free}$ calculation. A rigid-body refinement using 1726 reflections in the resolution range 8.0–3.0 Å gave an $R_{\rm work}$ and $R_{\rm free}$ of 0.315 and 0.341, respectively. Positional refinements performed with a gradual increase of resolution to 1.54 Å gave an $R_{\rm work}$ and $R_{\rm free}$ of 0.280 and 0.317, respectively. B-factor refinement lowered $R_{\rm work}$ and $R_{\rm free}$ to 0.269 and 0.281, respectively. The model was then annealed by heating the system to 3000 K and slowly cooling to room temperature with 0.5 fs sampling intervals, lowering $R_{\rm work}$ and $R_{\rm free}$ to 0.238 and 0.263, respectively. 60 water molecules were located in consecutive steps with $2F_o - F_c$ and $F_o - F_c$ maps. Further refinement with these solvent molecules gave the final $R_{\rm work}$ and $R_{\rm free}$ of 0.191 and 0.235, respcetively. The 5′-g in either strand could not be located in the density maps, which may be owing to disorder of these overhang residues. The final model contains 240 nucleic acid atoms and 60 water molecules. The crystallographic refinement parameters are listed in Table 1. The atomic coordinates and structure factors have been deposited with the Nucleic Acid Database (NDB code ZD0008) and the PDB (Berman *et al.*, 1992).

Table 1
Crystal data and refinement parameters of gCGCGCG.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	17.96
b	31.47
c	44.73
Contents of asymmetric unit	1 duplex
Volume per base pair (Å^3)	903
Resolution range (Å)	8.0–1.54
No. of reflections used [$F > 2.0\sigma(F)$]	3536
Final $R_{\rm work}/R_{\rm free}$	0.191/0.235
Final model	
Nucleic acid atoms	240
Water molecules	60
Average thermal parameters (Å^2)	
Nucleic acid atoms	11.3
Water molecules	28.6
R.m.s.d.s from ideal geometry	
Bond lengths (Å)	0.007
Bond angles (°)	71.3
Dihedral angles (°)	34.4
'Improper' angles (°)	1.2

3. Results and discussion

3.1. Overall structure

The heptamer gCGCGCG crystallized as a Z-DNA duplex $(CGCGCG)_2$ of residues 2–7/9–14 with the 5′-terminal residues g1 and g8 invisible in the electron-density map (Fig. 2*a*). The molecules are packed in the head-to-tail fashion to form a pseudo-continuous helix along the c axis (Fig. 2*b*). The two independent strands in the duplex show an r.m.s. deviation of 0.82 Å for their superposition and are related by an approximate twofold axis: 179.1° rotation and 0.03 Å translation. The base pairs in the helical junction stack in the 5′, 3′/5′, 3′ arrangement, with a twist angle of 52° and a rise of 3.2 Å. Superposition with the duplex portions of the all-DNA structure G**CGCGCG** gives an r.m.s. deviation of 0.83 Å, showing that their duplexes are similar even though their 5′-overhang residues are in different conformations (Fig. 3). The great difference between the two duplexes lies in the phosphate groups of pyrimidines C4, C6 and C11, which have an r.m.s. deviation of more than 2.5 Å. The residue C atoms adopt the C2′-*endo* sugar pucker and *anti* glycosidic conformation as in the heptamer GCGCGCG, while the residues G3, G5, G10, G12 show the O4′-*endo* sugar pucker (Table 2) instead of the C3′-*endo* in the heptamer GCGCGCG (Pan *et al.*, 1997).

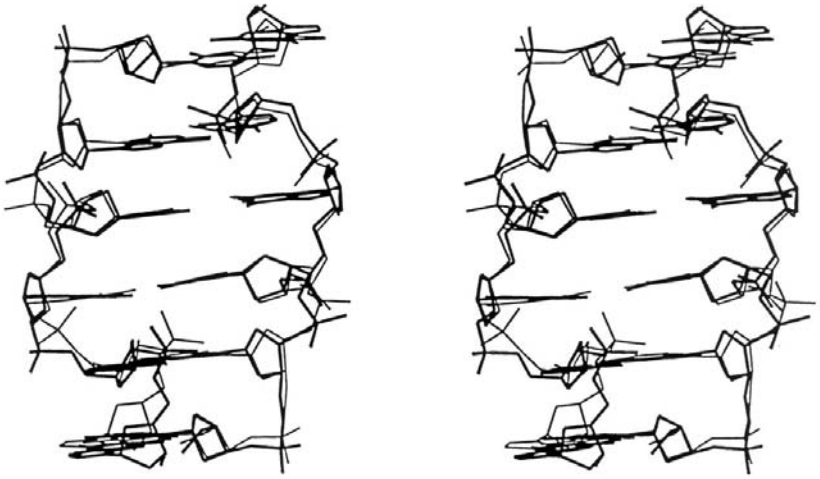

Figure 3
Stereoview of the superposition of the hexamer duplex portion (in bold) in (G**CGCGCG**) (thin lines) with the present hexamer duplex g**CGCGCG** (thick lines).

3.2. The effects influencing the Z-form duplex

It has been shown that the Z-form duplex d(GCGC**GC**GCGC) (Ban *et al.*, 1996) can switch to the A-form duplex when the GC bases in bold in the sequence are substituted by AT bases (Ban & Sundaralingam, 1996) or when the C bases are methylated (Tippin *et al.*, 1997). On the other hand, a 2′-hydroxyl group is beneficial to the formation of the A-form duplex (Ban *et al.*, 1994*a*,*b*; Wahl & Sundaralingam, 2000). However, replacement by r(CG) in the central

position of the hexamer Z-form duplex d(CG**CG**CG) does not convert the duplex to the A-form (Teng *et al.*, 1989). The present sequence provides two alternative assemblies: an A-DNA hexamer with a 3′-overhang G (Fig. 1*a*) or a Z-DNA hexamer with a 5′-overhang g (Fig. 1*b*). The crystal structures have shown that replacement of deoxyribonucleoside with ribonucleosides in B-DNA can convert the helical form to A-type (Ban *et al.*, 1994*a*,*b*; Wahl & Sundaralingam, 2000).

Table 2
Backbone torsion angles for gCGCGCG (°).

	α†	β	γ	δ	ε	ζ	χ	P
Strand 1								
C2	—	—	53	146	265	77	211	161 (C2′-*endo*)
G3	54	177	187	112	236	288	72	81 (O4′-*endo*)
C4	208	231	58	148	260	71	212	156 (C2′-*endo*)
G5	65	178	189	119	196	43	67	105 (O4′-*endo*)
C6	169	164	48	143	266	77	217	165 (C2′-*endo*)
G7	78	175	180	147	—	—	82	169 (C2′-*endo*)
Strand 2								
C9	—	—	50	146	267	78	217	163 (C2′-*endo*)
G10	56	183	181	107	241	285	78	82 (O4′-*endo*)
C11	216	224	55	147	265	78	204	163 (C2′-*endo*)
G12	61	178	184	111	244	292	72	95 (O4′-*endo*)
C13	213	226	54	142	263	67	207	159 (C2′-*endo*)
G14	81	180	185	156	—	—	79	168 (C2′-*endo*)

† The backbone torsion angles as defined by IUPAC–IUB (1983) are O3′-P-*α*-O5′-*β*-C5′-*γ*-C4′-*δ*-C3′-*ε*-O3′-*ζ*-*P*-O5′. *P* is the pseudorotation phase angle.

These results indicate that ribonucleotides are beneficial to the formation of the A-form helix. The present structure shows that the sequence still adopts the Z-form conformation. Together with the previous result in the replacement of ribonucleotides in Z-form DNA (Teng *et al.*, 1989), it seems that in converting helical form ribonucleotides may have a different effect upon Z-form DNA compared with B-form DNA.

3.3. Overhang residues

The less constrained overhang residues are more flexible in adopting different conformations than those that stay inside the duplex. Observed unusual base-pairing schemes of overhang residues include (G·C)*G base triplets (van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996), the reverse Hoogsteen G·G base pair (Pan *et al.*, 1997; Mooers *et al.*, 1997), the reverse wobble G·T and the reverse Watson–Crick G·C base pairs (Mooers *et al.*, 1997) in DNA and the *trans* U·U base pairs in RNA (Wahl, Rao *et al.*, 1996). Unlike the 5′ terminal nucleotides in GCGCGCG, in which one G swings out of the helix and the other stays in the helix to participate in the reverse Hoogsteen base pairing (Pan *et al.*, 1997), both 5′ terminal ribonucleotides in the present structure are missing and cannot be located in the electron density. Lack of hydrogen bonding and base-stacking interactions may be the reason for the disordered overhang residues. Similar phenomena have been observed for the blunt-end terminal residues in the DNA–RNA hybrid duplex crystal structure (Xiong & Sundaralingam, 1998) and the overhang residues in the RNA duplex structure (Shi *et al.*, 1999). Previous studies have shown that replacement of the deoxyribonucleotides by ribonucleotides may result in a switch from B-form to A-form duplex with the base-pairing schemes remaining intact (Ban *et al.*, 1994*a*,*b*). The present structure provides evidence that such replacement in the terminal nucleotide in a Z-form helix cannot change the helical form to A-form.

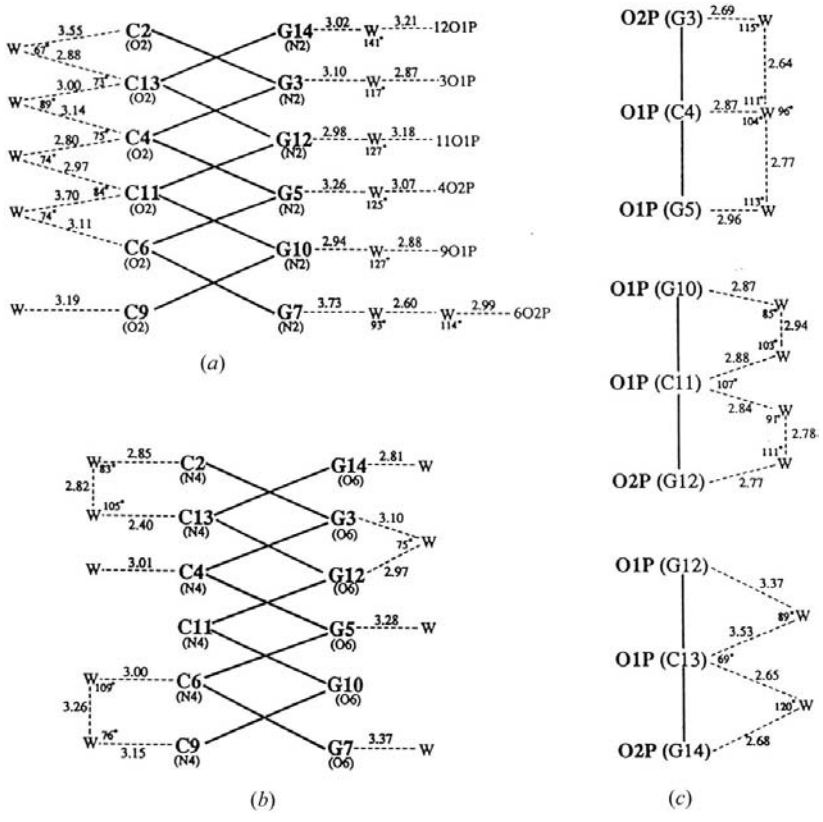

Figure 4
Hydration of the heptamer structure gCGCGCG. (*a*) Water structure in the minor groove. The backbones of the heptamer are shown in dark lines as zigzag chains, while hydrogen bonds are shown as dotted lines. (*b*) Water structure in the major groove. (*c*) Three different hydration patterns for the phosphate groups.

3.4. Hydration of the structure

11 water molecules in the minor groove constitute two different motifs: (i) the water bridge between the O2 keto groups of cytosine from alternating strands and (ii) the water bridge between the N2 amino groups of guanosine and phosphate groups (Fig. 4*a*). The second motif is intact, while the first is disrupted at the end of the duplex. The conservation of the second motif may account for the stability of the *syn* conformation of G in the Z-DNA duplex. There is not much difference in the extent of hydration for O2 atoms of cytosine and N2 atoms of guanine in the minor groove. However, N3 atoms of guanine are not hydrated at all, which may be a consequence of their extreme closeness to the sugar atoms. By comparison, the

following two motifs in the major groove are not as conserved as in the minor groove: (i) the water bridge connecting the N4 amino group of cytosine in alternating strands and (ii) the water bridge linking the O6 keto group of guanine in alternating strands (Fig. 4*b*). Only half of the N7 atoms of G residues in the structure are hydrated. The present results suggest that the extent of hydration for the N4 atom of cytosine in this Z-form helix and the O6 and N7 atoms of guanine are in the order N4 > O6 > N7 in the major groove. Three different hydration patterns have been observed for the phosphate groups for GpCpGp steps, in which two, three and four water molecules are involved in the water bridge (Fig. 4*c*). The present heptamer structure reveals the common water-structure motifs for the Z-DNA hexamer CGCGCG (Gessner *et al.*, 1994) and the heptamer GCGCGCG structure (Pan *et al.*, 1997). Despite the change in the conformation of the terminal residues, the internal water structure is not affected by the replacement of G by g in the 5′-terminal duplex.

We gratefully thank the National Institutes of Health grant GM-17378 and the Board of Regents of Ohio for providing an Ohio Eminent Scholar Endowment in support of this work. We also thank the Board of Regents of Ohio for partially supporting the purchase of the R-AXIS IIc imaging plate through a Consortium Investment Fund.

References

Ayer, D. & Yarus, M. (1986). *Science*, **231**, 393–395.
Ban, C., Ramakrishnan, B. & Sundaralingam, M. (1994*a*). *J. Mol. Biol.* **236**, 275–285.
Ban, C., Ramakrishnan, B. & Sundaralingam, M. (1994*b*). *Nucleic Acids Res.* **22**, 5466–5476.
Ban, C., Ramakrishnan, B. & Sundaralingam, M. (1996). *Biophys. J.* **71**, 1215–1221.
Ban, C. & Sundaralingam, M. (1996). *Biophys. J.* **71**, 1222–1227.
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). *X-PLOR Manual, Version* 3.0. Yale University, New Haven, CT, USA.
Burkard, M. E., Kierzek, R. & Turner, D. H. (1999). *J. Mol. Biol.* **290**, 967–982.
Franklin, R. E. & Gosling, R. G. (1953). *Acta Cryst.* **6**, 673–677.
Freier, S. M., Alkema, D., Sinclair, A., Neilson, T. & Turner, D. H. (1985). *Biochemistry*, **24**, 4533–4539.
Gessner, R. V., Quigley, G. J. & Egli, M. (1994). *J. Mol. Biol.* **236**, 1154–1168.
Grosjean, H., Soll, D. G. & Crothers, D. M. (1976). *J. Mol. Biol.* **103**, 499–519.
IUPAC–IUB Joint Commission on Biochemistry Nomenclature (1983). *Eur. J. Biochem.* **131**, 9–15.
Jain, S., Zon, G. & Sundaralingam, M. (1987). *J. Mol. Biol.* **197**, 141–145.
Martin, F. H., Uhlenbeck, O. C. & Doty, P. (1971). *J. Mol. Biol.* **57**, 201–215.
Meervelt, L. V. van, Vlieghe, M., Dautant, A., Gallois, B., Precigoux, G. & Kennard, O. (1995). *Nature (London)*, **374**, 742–744.
Mooers, B. H. M., Eichman, B. F. & Ho, P. S. (1997). *J. Mol. Biol.* **269**, 796–810.
Mooers, B. H. M., Schroth, G. P., Baxter, W. W. & Ho, P. S. (1995). *J. Mol. Biol.* **249**, 772–784.
Pan, B., Ban, C., Wahl, M. C. & Sundaralingam, M. (1997). *Biophys. J.* **73**, 1553–1561.
Petersheim, M. & Turner, D. H. (1983). *Biochemistry*, **22**, 256–263.
Pohl, F. M. & Jovin, T. M. (1972). *J. Mol. Biol.* **67**, 375–396.
Romaniuk, P. J., Hughes, D. W., Gregoire, R. J., Neilson, T. & Bell, R. A. (1978). *J. Am. Chem. Soc.* **100**, 3971–3972.
Shi, K., Wahl, M. & Sundaralingam, M. (1999). *Nucleic Acids Res.* **27**, 2196–2201.
Sugimoto, N., Kierzek, R. & Turner, D. H. (1987). *Biochemistry*, **26**, 4554–4558.
Teng, M., Liaw, Y.-C, Van der Marel, G. A., Van Boom, J. H. & Wang, A. H.-J. (1989). *Biochemistry*, **28**, 4923–4928.
Tippin, D. B., Ramakrishnan, B. & Sundaralingam, M. (1997). *J. Mol. Biol.* **270**, 247–258.
Vlieghe, D., van Meervelt, L. V., Dautant, A., Gallois, B., Precigoux, G. & Kennard, O. (1996). *Science*, **273**, 1702–1705.
Wahl, M. C., Ramakrishnan, B., Ban, C., Chen, X. & Sundaralingam, M. (1996). *Acta Cryst.* D**52**, 668–675.
Wahl, M. C., Rao, S. T. & Sundaralingam, M. (1996). *Nature Struct. Biol.* **3**, 24–31.
Wahl, M. C. & Sundaralingam, M. (2000). *Nucleic Acids Res.* **28**, 4356–4363.
Wang, A. H.-J., Hakoshima, T., van der Marel, G., 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. & Rich, A. (1979). *Nature (London)*, **282**, 680–686.
Xiong, Y. & Sundaralingam, M. (1998). *Structure*, **6**, 1493–1501.