X-ray analyses of d(GCGAXAGC) containing G and T at X: the base-intercalated duplex is still stable even in point mutants at the fifth residue

Jiro Kondo, Shun-ichi Umeda, Kazuhiro Fujita, Tomoko Sunami and Akio Takénaka*

Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan. E-mail: atakenak@bio.titech.ac.jp

DNA fragments containing the sequence d(GCGAAAGC) prefer to adopt a base-intercalated (zipper-like) duplex in crystalline state. To investigate effects of point mutation at the 5th residue on the structure, two crystal structures of d(GCGAGAGC) and d(GCGATAGC) have been determined by X-ray crystallography. In the respective crystals, the two octamers related by a crystallographic two-fold symmetry are aligned in an anti-parallel fashion and associated to each other to form a duplex, suggesting that the base-intercalated duplex is stable even when the 5th residue is mutated with other bases. The sheared G_3:A_6 pair formation makes the two phosphate backbones closer and facilitates formation of the A-X*-X-A* base-intercalated motif. The three duplexes are assembled around the three-fold axis, and their 3rd and 4th residues are bound to the hexamine cobalt chloride. The central 5th residues are bound to another cation.

Keywords: base-intercalated duplex; A-X*-X-A* base-intercalated motif; sheared G:A pair; DNA; crystal structure.

1. Introduction

For storage of genetic information, DNA is highly sophisticated to be chemically stable and to form a duplex containing redundant information through the complementary base-base interactions. On the other hand, RNA exists in general as a single strand but it is folded to form a complicated three-dimensional structure so that it serves as a functional molecule similar to proteins. The typical examples are found in ribosomal RNAs (Yusupov *et al*., 2001), hammerhead ribozymes (Pley *et al*., 1994), group I intron ribozymes (Cate *et al*., 1996) and so on, in which stems composed of anti-parallel two strands are folded into a globular form by tertiary interactions involving loops and bulges.

In living organisms, it is difficult to find a functional DNA similar to RNA, because the life was established and evolved on the basis of the complementary system with DNA. Recent *in vitro* selection technique, however, makes it possible to create a functional DNA, which catalyze RNA-cleaving reaction (Breaker & Joyce, 1994). As a material for this purpose, DNA may be more useful due to its chemical stability. Such a single strand DNA would adopt a specific structure folded by intra-strand and inter-strand interactions. In order to establish the structural basis for designing such molecules, it is necessary to reveal the structural features of single stranded DNA.

DNA fragments containing the sequence d(GCGAAAGC) were found with their extraordinary properties and postulated to adopt a mini-hairpin structure from CD and NMR studies (Hirao *et al*., 1989). In crystalline states, however, it has been found that the octamer with the sequence d(GCGAAAGC) form a base-intercalated duplex (Sunami *et al*., 2001; Sunami *et al*., 2003), similar to a zipper-like duplex reported by Shepard *et al*., 1998. As a basic structure to support a specific function, it is necessary to examine the structural stability of this duplex. X-Ray analyses of d(GCGAXAGC) containing guanine and thymine residues at X have been performed.[#]

2. Materials and methods

2.1. Synthesis and crystallization

Two kinds of DNA oligomers with the sequence d(GCGAXAGC) containing G and T at X were synthesized on a DNA synthesizer (hereafter abbreviated as 8hmt-G_5 and 8hmt-T_5, respectively). To apply the MAD method for phase determination, 2'-deoxy-5-bromocytidine was introduced at the 2nd residue. The octamers were purified by HPLC, and converted to Na^+ salt on cation-exchange column. Furthermore, excess Na^+ cation was removed by dialysis in milliQ water. Crystallization conditions were surveyed at 4 °C and 20 °C by the hanging-drop vapor diffusion method, each droplet (2 µl initial volume) being equilibrated to 1000 µl of 30% (v/v) 2-methyl-2,4-pentanediol. Crystals of 8hmt-G_5 were obtained at 4 °C for four days from a droplet prepared by merging 1 µl of 0.5 mM DNA solution and 1 µl of 50 mM sodium succinate buffer solution (pH 7.0) containing 1 mM hexamine cobalt chloride, 10 mM magnesium chloride and 20% (v/v) 2-methyl-2,4-pentanediol. Those of 8hmt-T_5 were at 4 °C for two weeks from a droplet prepared by merging 1 µl of 1.5 mM DNA solution and 1 µl of 50 mM sodium cacodylate buffer solution (pH 7.0) containing 2 mM hexamine cobalt chloride, 100 mM sodium chloride and 10% (v/v) 2-methyl-2,4-pentanediol.

2.2. Data collection

X-Ray data of 8hmt-G_5 and 8hmt-T_5 crystals were collected at 100 K using synchrotron radiations at BL18B (λ = 1.00 Å) of Photon Factory in Tsukuba and at BL44XU (λ = 0.90 Å) of SPring-8 in Harima, respectively. Diffraction patterns were processed with the program *DPS/MOSFLM* (Leslie, 1992; Steller *et al*., 1997; Rossmann & van Beek, 1999; Powell, 1999) at resolution 2.0 and 1.7 Å, respectively. Intensity data were then put on a relative scale and merged into the independent reciprocal space using the program *SCALA* and *TRUNCATE* of *CCP4* suite (Collaborative Computational Project, Number 4, 1994). Statistics of the data collections and crystal data are summarized in Table 1. The two crystals 8hmt-G_5 and 8hmt-T_5 are isomorphous, the space group being the same $P6_322$ with similar unit-cell dimensions ($a = b$ = 37.7 and c = 65.5 Å for 8hmt-G_5, and $a = b$ = 37.0 and c = 65.3 Å for 8hmt-T_5). R_{merge} of 8hmt-G_5 is 7.5% for the 2142 unique reflections and R_{merge} of 8hmt-T_5 is 10.1% for the 2919 unique reflections.

2.3. Structure determination and refinement

The crystal data of 8hmt-G_5 and 8hmt-T_5 suggested that they are isomorphous to the hexagonal crystal of d(GCGAAAGC) (8hmt-A_5 hereafter) (Sunami *et al*., 2003). Initial phases of the observed data were derived uniquely by the molecular replacement method with the program *AMoRe* (Navaza, 1994) using that atomic coordinates of 8hmt-A_5. Therefore, the MAD method was not applied, although the oligonucleotides contain a bromo-derivative at the 2nd position. The molecular structures were constructed and modified on a graphic

[#] To describe the detailed structures, the following numbering expressions are employed for nucleotide residues and atoms hereafter, X_n: the n-th residue X and Xn: the n-th atom X according to the IUPAC-IUPAB nomenclature.

workstation by inspecting $|F_o|$-$|F_c|$ omit maps with the program *QUANTA* (Molecular simulation Inc.). The atomic parameters were refined with the program *CNS* (Brünger *et al*., 1998) through a combination of rigid-body, simulated-annealing, crystallographic conjugate gradient minimization refinements and B-factor refinements, followed by interpretation of an omit map at every nucleotide residues. In the initial stages of refinement, all sugar puckers were assumed to be C2'-*endo*, but released later. No restraints were applied to base-base interactions. In the 8hmt-G_5 crystal, a hexamine cobalt cation and a chloride ion were found on the crystallographic three-fold axis, and one magnesium ion and seventy water molecules were included in the final refinement. In addition to these solvents, another hexamine cobalt cation was found at the cross of the crystallographic three-fold and two-fold axes in the 8hmt-T_5 crystal. The final *R*-factor of 8hmt-G_5 was 22.0% for 20.0 - 2.0 Å resolution data (R_{free} = 27.8% for 10% of the observed data) and that of 8hmt-T_5 was 22.4% for 65.3 - 1.7 Å resolution data (R_{free} = 23.9% for 10% of observed data). Statistics of the structure refinement are summarized in Table 1. All local helical parameters including torsion angles and pseudorotation phase angles of ribose rings were calculated by using the program *NUPARAM* (Bansal *et al*., 1995). Figure 1 shows local electron density maps drawn with the program *O* (Jones *et al*., 1991). Figures 2 and 3 were depicted with the program *RASMOL* (Sayle & Milner-White, 1995).

Table 1 Crystal data, statistics of data collection, and statistics of structure refinement.

Crystal sample	8hmt-G_5	8hmt-T_5
Crystal data		
Space group	$P6_322$	$P6_322$
Unit cell (Å)	$a = b = 37.7$, $c = 65.5$	$a = b = 37.0$, $c = 65.3$
Asymmetric unit[a]	1	1
Data collection		
Resolution (Å)	29 - 2.0	65 - 1.7
Observed reflections	104504	223225
Unique reflections	2142	2919
Completeness (%)	100.0	89.7
in the outer shell	100.0	83.5
R_{merge}[b] (%)	7.5	10.1
in the outer shell	32.3	21.5
Structure determination		
Resolution range (Å)	20 - 2.0	65 - 1.7
Used reflections	2030 ($F_o > 3\sigma$)	2707 ($F_o > 3\sigma$)
R-factor[c] (%)	22.0	22.4
R_{free}[d]	27.8	23.9
Number of DNA atoms	166	164
Number of water molecules	70	91
Number of hexamine cobalt cations	1/3	1/2
Number of chloride ions	1/3	1/3
Number of magnesium ions	1	0
Number of sodium ions	0	1
R.m.s. deviation		
Bond length (Å)	0.004	0.005
Bond angles (°)	0.8	0.9
Improper angles (°)	0.8	0.9

[a]A single strand d(G^{Br}CGAGAGC) or d(G^{Br}CGATAGC).
[b]$R_{merge} = 100 \times \Sigma_{hklj}|I_{hklj} - \langle I_{hklj}\rangle| / \Sigma_{hklj}\langle I_{hklj}\rangle$.
[c]*R*-factor = $100 \times \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.
[d]Calculated using a random set containing 10% of observations that were not included throughout refinement (Brünger, 1992).

Table 2 Local helical parameters and intra base-pair parameters. Values for one end of the duplex are shown due to the crystallographic two-fold symmetry.

8hmt-G_5	Inclin[a]	Tip	Twist	D_z[b]	Prop[a]	Buck[a]	Open[a]	C1...C1[b]
G_1:C_8					5	4	20	10.8
	-2	-2	24	3.5				
$^{Br}C_2$:G_7					6	-18	15	10.5
	42	-41	9	2.7				
G_3:A_6					16	-30	7	8.4
8hmt-T_5								
G_1:C_8					8	6	20	10.7
	1	2	23	3.5				
$^{Br}C_2$:G_7					10	-13	16	10.6
	39	-47	8	2.4				
G_3:A_6					19	-26	7	8.4

[a]Inclin, inclination angle (°); Prop, propeller twist angle (°); Buck, buckle angle (°); Open, opening angle (°).
[b]Distance in Å.

Table 3 Pseudorotation phase angle (°) and sugar conformation.

Sequence	Phase[a]	Conformation	Sequence	Phase[a]	Conformation
8hmt-G_5	(°)	Pucker	**8hmt-T_5**	(°)	Pucker
G_1	-168	C3'-*exo*	G_1	-158	C3'-*exo*
$^{Br}C_2$	134	C1'-*exo*	$^{Br}C_2$	127	C1'-*exo*
G_3	151	C2'-*endo*	G_3	153	C2'-*endo*
A_4	29	C3'-*endo*	A_4	31	C3'-*endo*
G_5	152	C2'-*endo*	T_5	128	C1'-*exo*
A_6	176	C2'-*endo*	A_6	-157	C3'-*exo*
G_7	94	O4'-*endo*	G_7	119	C1'-*exo*
C_8	44	C4'-*exo*	C_8	53	C4'-*exo*

[a]Pseudorotation phase angle of ribose ring.

3. Results and discussion

3.1. Overall structures of 8hmt-G_5 and 8hmt-T_5

In the crystals of 8hmt-G_5 and 8hmt-T_5, the two octamers related by a crystallographic two-fold symmetry are aligned in an anti-parallel fashion and associated to each other to form a duplex, as shown in Figure 2 (*a* and *b*). At both ends of the duplexes, two Watson-Crick G:C base pairs, G_1:C_8* (C_8:G_1*) and C_2:G_7* (G_7:C_2*), are formed (see Figure 1 (*a* and *b*)). The subsequent G_3 residue forms a sheared pair with the A_6 residue of the counter strand, G_3:A_6* and A_6:G_3* through the two hydrogen bonds, N2H...N7 and N3…HN6 (see Figure 1 (*c*)). It is interesting to note that the remaining two residues, A_4 and X_5, are not involved in any base-pair formations. Their base moieties are respectively stacked on those of the other strand so that A_4 is intercalated between G_3 and X_5*, and X_5 is intercalated between X_5* and A_4*. These four bases, A_4, X_5*, X_5 and A_4*, expose their Watson-Crick-type hydrogen bonding sites[†] into the solvent region to interact with the surrounding molecules. As a result, there are three different stacked columns in a duplex. One is a long column of G_1-C_2-G_3-A_4-X_5*-X_5-A_4*-G_3*-C_2*-G_1* and two short columns of A_6-G_7-C_8 and A_6*-G_7*-C_8*. These structural features of 8hmt-G_5 and 8hmt-T_5 are quite similar to those of the base-intercalated duplex of 8hmt-A_5 (Sunami *et al*., 2003), as shown in Figure 2 (*c*). The root-mean-square deviations between 8hmt-G_5 and 8hmt-A_5 and between 8hmt-T_5 and 8hmt-A_5 are 0.5 Å and 0.4 Å, respectively, when the non-hydrogen atoms except the 5th residues are superimposed. It shows that bromination at the 2nd residues have no significant effects on the tertiary structures, because 8hmt-A_5 has no chemical modification.

[†] "Watson-Crick sites" means the donor and the acceptor sites for hydrogen bonds to form the Watson-Crick base pairs.

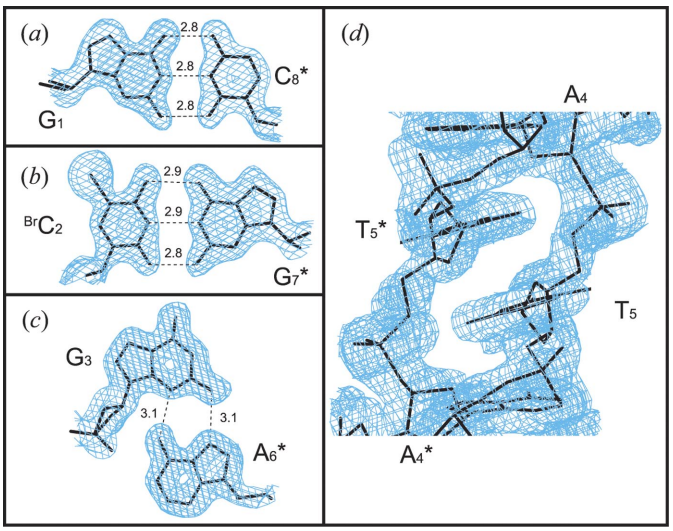

Figure 1 Local $2|F_o|-|F_c|$ maps for Watson-Crick base pairs (*a* and *b*), a sheared G:A pair (*c*) and the central two strands (*d*) of 8hmt-T_5. Density is contoured at 1σ level. Broken lines indicate possible hydrogen bonds with the distances.

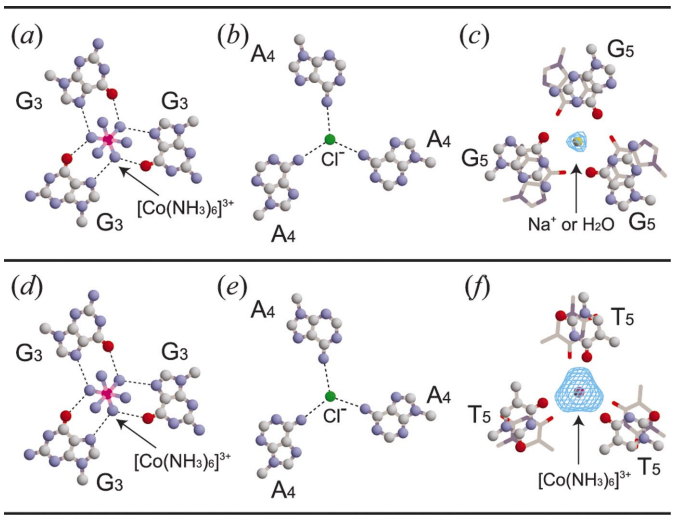

Figure 3 Assemblies of the three base-intercalated duplexes around the crystallographic three-fold axis. The G_3, A_4 and X_5 residues of 8hmt-G_5 are shown in (*a*), (*b*) and (*c*), respectively, and those of 8hmt-T_5 are shown in (*d*), (*e*) and (*f*). $2|F_o|-|F_c|$ maps in (*c*) and (*f*) are contoured at 1σ level to show the central ions. Broken lines indicate possible hydrogen bonds.

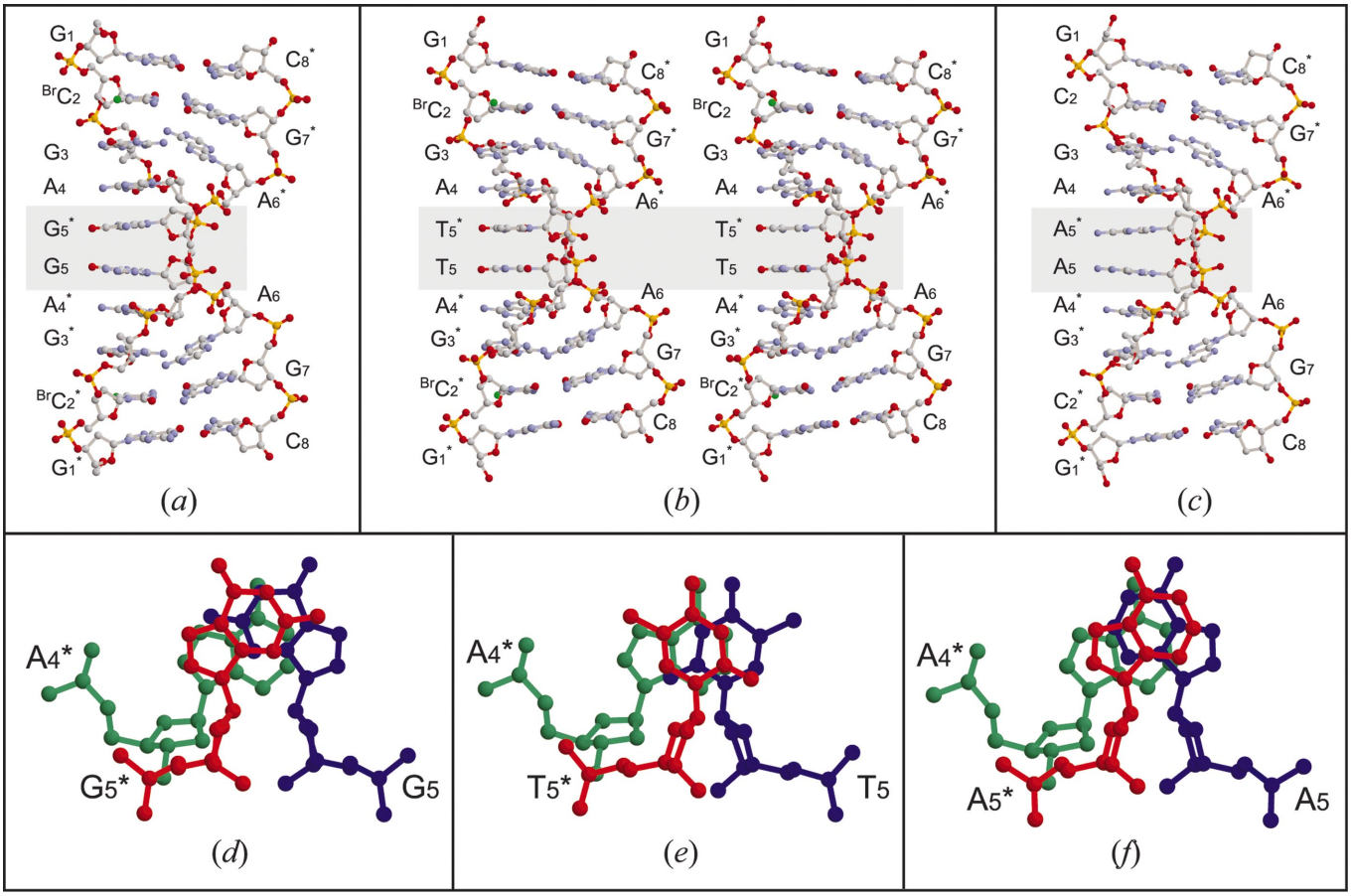

Figure 2 Overall structures of the base-intercalated duplexes, 8hmt-G_5 (*a*), 8hmt-T_5 (*b*; stereo-pair diagram), 8hmt-A_5 (*c*) and the intercalated A_4*-X_5-X_5* stackings viewed down the *c*-axes, in 8hmt-G_5 (*d*), in 8hmt-T_5 (*e*) and in 8hmt-A_5 (*f*). The superscript * indicates the counter strand. (Atomic coordinates of 8hmt-A_5 from Sunami *et al*., 2003).

The helical parameters and the base-pair geometry are given in Table 2, and the pseudorotation phase angles in Table 3. In the stem regions of 8hmt-G_5 and 8hmt-T_5, the G_1, C_2 and G_7 residues adopt a C2'-*endo* pucker to maintain the canonical *B*-form conformation. The C_8 residues, however, adopt a C3'-*endo* conformation. This difference may be due to the surrounding solvent structure. At the 3rd residue, the C1'-C1' distance is shorter (8.4 Å) than that (10.7 Å) of the standard Watson-Crick pairs, and the buckle angle larger to form the sheared pair between the minor-groove site of the G_3 residue and the major-groove site of the A_6 residue. However, the two residues adopt a conformation close the C2'-*endo* pucker. To support the A_4-X_5*-X_5-A_4* intercalated stacking, the G:A pairs serve as scaffolds at both ends of the duplex. The A_4 and X_5 residues adopt a C3'-*endo* and a C2'-*endo* (C1'-*exo*) puckers, respectively. The former pucker is required to make an open space for intercalation of the X_5* bases between the A_4 and X_5 residues. Another notable feature is the torsion angle of the P-O5' and C5'-C4' bonds of the X_5 residue. These angles also support the backbone extension.

3.2. Effect of ions on assembly of three base-intercalated duplexes

In both crystals 8hmt-G_5 and 8hmt-T_5, the three base-intercalated duplexes are assembled around the crystallographic three-fold axis. Three G_3 residues are bound to a hexamine cobalt cation, the coordinated ammonia molecules being hydrogen-bonded to the O6 and N7 atoms in the major groove sites, as shown in Figure 3. Three N6 atoms of the A_4 residues are bound to a chloride ion. At the 5th residues, the solvent effect is different depending on the type of bases. In the 8hmt-G_5 crystal, the six G_5 residues locating around the cross of the crystallographic three-fold and the two-fold axes make an ellipsoidal isolated space, in which a sodium cation or a water molecule is disordered at the two sites on the three-fold axis (see Figure 3 (*c*)). In the 8hmt-T_5 crystal, however, a cobalt atom occupies the cross-center of the corresponding space surrounded by the six T_5 residues, the coordinated six ammonia molecules being disordered (see Figure 3 (*f*)). In addition, the hydrated magnesium cation in 8hmt-G_5 and the hydrated sodium cation in 8hmt-T_5 are bound in the major groove site of the G_7 residues.

4. Conclusion

The base-intercalated duplex was reported for 8hmt-A_5 with the sequence d(GCGAAAGC). The present work confirms that the base-intercalated duplex is stable even when the 5th residue is mutated with other bases. These results apparently indicate that the sequence of d(GCGAXAGC) is specific to form a base-intercalated duplex as a strong intrinsic property. The sheared G:A base pair formations at the two positions make the two phosphate backbones closer and facilitate the A-X*-X-A* intercalated base-base stacking. In addition, several ions are found to support formation of these characteristic structures. Therefore, it is concluded that the duplex formation with a base-intercalated motif is preferred under a high-salt condition, and that the structure itself is very stable due to the sheared G:A pairs.

The most interesting question is whether the present structures is involved in some biological function. As the major parts of living systems are established by using the Watson-Crick pairing between complementary strands, it may be difficult to find such unique sequence in the stable double helix structure. However, it is expected that when DNA is unwound to perform a function in a single stranded state, the base-intercalated motif will occur as a local structure to have a function. Single-strand DNA from DNA viruses or phages could adopt many types of structures (like RNA) with specific functions. For regulation of the telomere length in eukaryote, single stranded region of the telomeric DNA is assumed to form quadruplex with G-quartet (Kang *et al*, 1992) and intercalated cytosine motif (*i*-motif) (Chen *et al*., 1994). Moreover, the base-intercalated motif will be useful for designing various functional DNA, which are being developed for several purposes, such as anti-genes (Sioud, 2001) and RNA-cleaving DNA enzymes (Breaker & Joyce, 1994).

We thank N. Sakabe, M. Suzuki, N. Igarashi and A. Nakagawa for synchrotron radiation facilities and help during data collection. This work was supported in part by Grants-in-Aid for Scientific Research (No.12480177 and 14035217) from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the Structural Biology Sakabe Project.

Data bank accession code The atomic coordinates have been deposited in the Protein Data Bank (PDB) with the ID codes 1UHX for d(GCGAGAGC) and 1UHY for d(GCGATAGC).

Reference

Bansal, M., Battacharyya, D. & Ravi, B. (1995). *Comput. Appl. Biosci*. **11**, 281-287
Breaker, R. R. & Joyce, G. F. (1994). *Chem. Biol*. **1**, 223-229.
Brünger, A. T. (1992). *Nature*, **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.* **D54**, 905-921.
Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). *Science*, **273**, 1678-1685.
Chen, L., Cai, L., Zhang, X. & Rich, A. (1994). *Biochemistry*, **33**, 13540-13546.
Collaborative Computational Project, Number 4. (1994). *Acta Cryst.* **D50**, 760-763.
Hirao, I., Nishimura, Y., Naraoka, T., Watanabe, K., Arata, Y. & Miura, K. (1989). *Nucleic Acids Res.* **17**, 2223-2231.
Jones, T. A., Zou, J. Y., Cowan, S.W. & Kjeldgaard, M. (1991). *Acta Cryst*. **A47**, 110-119.
Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992). *Nature*, **356**, 126-131.
Leslie, A. G. W. (1992). *Molecular data processing. Crystallography Computing* **5**, *From chemistry to Biology*, edited by Moras, D., Podjarny, A. D. & Thierry, J. C. pp. 50-61. Oxford University Press, Oxford, UK.
Navaza, J. (1994). *Acta Cryst.* **A50**, 157-163.
Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). *Nature*, **372**, 68-74.
Powell, H. R. (1999). *Acta Cryst.* **D55**, 1690-1695.
Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst*. **D55**, 1631-1640.
Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci*. **20**, 374.
Shepard, W., Cruse, W. B., Fourme, R., de la Fortelle, E. & Prangé, T. (1998) *Structure*, **6**, 849-861
Sioud, M. (2001). *Curr. Mol. Med*. **1**, 575-588.
Steller, I., Bolotovsky, R. & Rossmann, M. G. (1997). *J. Appl. Cryst.* **30**, 1036-1040.
Sunami, T., Kondo, J., Chatake, T., Hirao, I., Watanabe, K., Miura, K. & Takénaka, A. (2001). *Nucleic Acids Res. Suppl.* **1**, 191-192.
Sunami, T., Kondo, J., Hirao, I., Watanabe, K., Miura, K. & Takénaka, A. (2003) *Acta Cryst*. **D59**, in press.
Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). *Science*, **292**, 883-896.