Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of a d(gcGAACgc) duplex containing two consecutive bulged A residues in both strands suggests a molecular switch

Jiro Kondo,‡ Tomoko Sunami and Akio Takénaka*

Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan

‡ Present address: Institut de Biologie Moléculaire et Cellulaire du CNRS, Université Louis Pasteur, Strasbourg, France.

Correspondence e-mail: atakenak@bio.titech.ac.jp

Received 17 February 2007
Accepted 17 March 2007

PDB Reference:
d(gcGAACgc), 2got, r2gotsf.

In previous studies, it was reported that DNA fragments with the sequence d(gcG*XY*Agc) (where *X* = A or G and *Y* = A, T or G) form a stable base-intercalated duplex (Bi-duplex) in which the central *X* and *Y* residues are not involved in any base-pair interactions but are alternately stacked on each other between the two strands. To investigate the structural stability of the Bi-duplex, the crystal structure of d(gcGAACgc) with a point mutation at the sixth residue of the sequence, d(gcGAAAgc), has been determined. The two strands are associated in an antiparallel fashion to form two types of bulge-containing duplexes (Bc-duplexes), I and II, both of which are quite different from the Bi-duplex of the parent sequence. In both Bc-duplexes, three Watson–Crick G·C base pairs constitute the stem regions at the two ends. The A_4 residues are bulged in to form a pair with the corresponding A_4 residue of the opposite strand in either duplex. The $A_4{\cdot}A_4$* pair formation is correlated to the orientations of the adjacent A_5 residues. A remarkable difference between the two Bc-duplexes is seen at the A_5 residue. In Bc-duplex I, it is flipped out and comes back to interact with the G_3 residue. In Bc-duplex II, the A_5 residue extends outwards to interact with the G_7 residue of the neighbouring Bc-duplex I. These results indicate that *trans* sugar-edge/Hoogsteen (sheared-type) $G_3{\cdot}A_6$* base pairs are essential in the formation of a Bi-duplex of d(gcG*XY*Agc). On the other hand, the alternative conformations of the internal loops containing two consecutive bulged A residues suggest molecular switching.

1. Introduction

Recent human-genome projects have revealed that coding sequences comprise only 1.2% of the euchromatin (International Human Genome Sequencing Consortium, 2004), while large proportions of the remaining parts are occupied by many different types of repetitive sequences, such as transposon-derived sequences, simple sequence repeats (SSRs) and segmental duplications, which are dispersed throughout the human genome (International Human Genome Sequencing Consortium, 2001). It might be expected that such repetitive DNA sequences exist in single-stranded states during the dynamic processes of replication, transcription, recombination and cell development and that they could also potentially form more complicated three-dimensional structures in order to carry out specific biological functions. For example, the G-rich DNA repeats found in some SSRs and telomeres have been considered to be capable of forming quadruplexes with G-quartets, while the complementary C-rich repeats can also form quadruplexes with intercalated C–C pairs (Patel *et al.*,

1999). These examples would be analogous to the complex folding of single-stranded RNA into a functional structure as seen in hammerhead ribozymes (Pley *et al.*, 1994; Scott *et al.*, 1995; Doudna, 1995), group I intron ribozymes (Cate *et al.*, 1996; Golden *et al.*, 2005; Stahley & Strobel, 2005), ribosomal RNAs (Harms *et al.*, 2001; Ban *et al.*, 2000; Wimberly *et al.*, 2000; Carter *et al.*, 2000; Yusupov *et al.*, 2001; Schuwirth *et al.*, 2005) and so on. Therefore, structural studies of DNAs with repeated sequences and/or non-Watson–Crick pairing will help us to understand their possible biological significance in the single-stranded state.

We have previously reported that the DNA sequence d(gcG*XY*Agc) (where *X* = A or G and *Y* = A, T or G) forms a base-intercalated duplex (hereafter termed a Bi-duplex, see Fig. 1; Kondo, Umeda *et al.*, 2004; Kondo, Adachi *et al.*, 2004; Kondo, Tanashaya *et al.*, 2006; Sunami *et al.*, 2004*a*,*b*) and have proposed that this structure may be involved in certain biological processes[1]. In this structure, two Watson–Crick G·C base pairs are formed at either end of the duplex. The subsequent G_3 residue forms a *trans* sugar-edge/Hoogsteen base pair[2] (a well known sheared-base pair) with the A_6 residue of the opposite strand, involving the two hydrogen bonds N3(G)···H—N6(A) and N2—H(G)···N7(A). The central X_4 and Y_5 residues, however, are not involved in any base pairs. Instead, they form an intercalated stacked X_4–Y_5*–Y_5–X_4* base column. In the case where *X* = A and *Y* = G, d(gcGAGAgc), four Bi-duplexes come together to form an octaplex at low potassium concentrations through double G-quartet formation. The octaplex splits into its two halves at a slightly higher potassium concentration (Kondo, Adachi *et al.*, 2004). These remarkable structural features suggest a folding mechanism involving a double Greek-key motif for groups of eight tandem repeats of d(ccGAGGGGAgg), such as the variable number of tandem repeats (VNTR) found next to the human pseudoautosomal telomere (Inglehearn & Cooke, 1990). Such a folding mechanism could induce slippage of the repeats during DNA replication. Another example of octaplex formation is also found in the case where *X* = *Y* = A, d(gcGAAAgc) (Sunami *et al.*, 2004*b*; Sato *et al.*, 2006). Its central four nucleotides, d(GAAA), are well known as an SSR in the human (Williams *et al.*, 1997), canine (Shibuya *et al.*, 1996), *Meloidogyne artiellia* (De Luca *et al.*, 2002) and *Oryza sativa* (McCouch *et al.*, 2002) genomes and they exhibit length polymorphism. It has been suggested by modelling analysis that this repetitive sequence can be folded into a cluster like an octaplex, which may be responsible for the slippage mechanism of the repeat (Sato *et al.*, 2006). In addition to these octaplex structures, it was found that Bi-duplexes with *X* = A and *Y* = A, T or G and with *X* = *Y* = G are associated to form a hexa-assembly in the presence of hexamine-cobalt cations or in a mixture of calcium and sodium cations (Kondo, Umeda *et al.*, 2004; Kondo, Tanashaya *et al.*, 2006; Sunami *et al.*, 2004*a*). These studies suggest that the G·A base pairs adjacent to the *X* and *Y* residues facilitate the formation of the Bi-duplexes. Similar sequences are dispersed throughout the eukaryotic genome, especially in repetitive sequences such as $d(GTACGGGACCGA)_n$ in the *Drosophila* centromere (Chou & Chin, 2001), d(gGAGGAc) in the human VNTRs located in the D1S80 (MCT118) locus (Kasai *et al.*, 1990) and d(tgGAGGGAca) in the human 3′-flanking mini-satellite pMS51 (Armour *et al.*, 1989).

In the present study, the crystal structure of a DNA octamer d(gcGAACgc) has been determined in order to investigate changes in the structure of the Bi-duplex when the A residue at the sixth position is replaced by other nucleotides (*e.g.* an A6C mutation). It has been found that such a mutation replaces the sheared G_3·A_6* base pairs in the Bi-duplex with canonical G_3·C_6* base pairs in the two stem regions. The central bases are released from the stacked column. These changes allow them to form two types of new bulge-containing duplexes (termed Bc-duplexes): one containing an intra-duplex hand-in-pocket motif and the other containing an inter-duplex hand-in-pocket motif. From the structural features of Bc-duplexes and their similarity to certain RNA motifs, it is expected that the symmetrical internal loop with the two consecutive bulged A residues found in this study could have a potential as a molecular switch.

2. Materials and methods

2.1. Preparation

The oligonucleotide d(gcGAACgc) (referred to hereafter as A6C) and its bromine derivative, which contains a 2′-deoxy-5-bromocytidine at the second position (hereafter termed

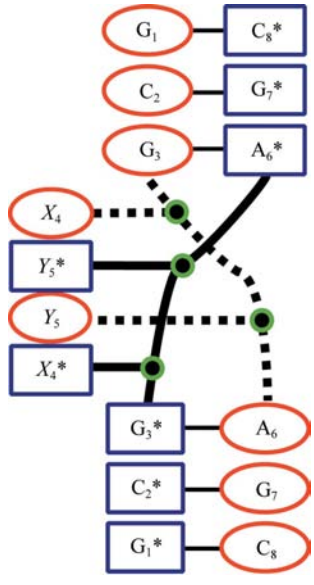

Figure 1
Schematic diagram of the Bi-duplex of d(gcG*XY*Agc) (where *X* = A or G and *Y* = A, T or G; Kondo, Umeda *et al.*, 2004; Kondo, Adachi *et al.*, 2004; Kondo, Tanashaya *et al.*, 2006; Sunami *et al.*, 2004*a*,*b*).

[1] Lower case characters in the sequence indicate residues that can form a Watson–Crick base pair when the fragments are aligned in an antiparallel fashion for self-assembly or folding.

[2] Geometric nomenclature and classification of nucleic acid base pairs is according to the Leontis–Westhof classification (Leontis & Westhof, 2001).

Table 1
Crystal data and data-collection and structure-determination statistics.

Values in parentheses are for the outer shell (0.1 Å width).

MAD phasing			
Crystal data			
Space group	$P6_122$		
Unit-cell parameters (Å)	$a = b = 26.8$, $c = 226.3$		
Z†	2		
Data collection			
Beamline	BL18B, Photon Factory		
Wavelength (Å)	1.00 (remote)	0.91955 (peak)	0.92010 (edge)
Resolution (Å)	45–2.6	38–2.8	25–2.7
Observed reflections	63062	61244	52625
Unique reflections	1729	1615	1670
Completeness (%)	92.8 (64.9)	99.8 (100.0)	99.5 (99.7)
$R_{\rm merge}$‡ (%)	5.4 (26.6)	6.5 (32.3)	6.3 (34.2)
$R_{\rm anom}$§ (%)	1.3 (7.0)	4.1 (7.5)	2.2 (11.4)
Redundancy	15.6 (12.7)	16.3 (16.9)	14.4 (10.0)
$I/\sigma(I)$	8.2 (2.5)	7.4 (2.3)	7.1 (2.1)
Theoretical f'/f''	−2.219/0.585	−7.084/3.819	−9.941/3.823
Refined f'/f''	−2.139/0.341	−5.569/2.888	−7.043/2.257
Structure determination			
Refinement			
Reflections used	1729		
Completeness (%)	92.8		
R factor¶ (%)	25.3		
$R_{\rm free}$†† (%)	26.9		
No. of DNA atoms	326		
No. of waters	5		
R.m.s. deviation			
Bond length (Å)	0.02		
Bond angles (°)	2.6		
Improper angles (°)	1.7		

† Number of ssDNA in the asymmetric unit. ‡ $R_{\rm merge} = 100 \times \sum_{hkl}\sum_{j}|I_{hklj} - \langle I_{hklj}\rangle|/\sum_{hkl}\sum_{j}\langle I_{hklj}\rangle$. § $R_{\rm anom} = 100 \times \sum_{hkl}|I_{hklj}(+) - I_{hklj}(-)|/\sum_{hkl}[I_{hklj}(+) + I_{hklj}(-)]$. ¶ R factor $= 100 \times \sum||F_{\rm o}| - |F_{\rm c}||/\sum|F_{\rm o}|$, where $|F_{\rm o}|$ and $|F_{\rm c}|$ are the observed and calculated structure-factor amplitudes, respectively. †† Calculated using a random set containing 10% of observations that were not included throughout refinement (Brünger, 1992).

A6C-Br), were synthesized using a DNA synthesizer. These octamers were purified by HPLC and gel filtration. Crystals of A6C-Br suitable for X-ray diffraction were grown at 277 K by the hanging-drop vapour-diffusion method from solutions containing 1 µl 1.5 m*M* DNA solution and 1 µl reservoir solution containing 40 m*M* sodium cacodylate pH 7.0, 12 m*M* spermine tetrahydrochloride, 80 m*M* sodium chloride and 10%(*v*/*v*) 2-methyl-2,4-pentanediol. Needle-shaped crystals of A6C-Br (0.1 × 0.1 × 2.0 mm in size) grown in 4 d were mounted in nylon cryoloops (Hampton Research) with the crystallization solution containing 35%(*v*/*v*) MPD as a cryoprotectant and stored in liquid nitrogen prior to X-ray experiments.

2.2. Data collection

The multiwavelength anomalous diffraction (MAD) method was applied for phase determination. To measure the anomalous scattering effects of Br atoms, three data sets were collected at 100 K using synchrotron radiation with wavelengths of 0.91955 Å for the peak, 0.92010 Å for the edge and 1.00 Å for the remote at the BL18B beamline of PF (Tsukuba, Japan). X-ray diffraction patterns were recorded on a CCD detector (Quantum 4R; Area Detector Systems Corporation) positioned at a distance of 265 mm from the crystal, using 1° oscillations and 40 s exposure per frame over a total range of 180°, and were processed at resolutions of 2.6 Å for the remote data, 2.8 Å for the peak data and 2.7 Å for the edge data with the program *DPS*/*MOSFLM* (Leslie, 1992; Steller *et al.*, 1997; Rossmann & van Beek, 1999; Powell, 1999). The intensities were then scaled and merged using *SCALA* from the *CCP*4 suite of crystallographic programs (Collaborative Computational Project, Number 4, 1994). The crystal belongs to space group $P6_122$, with unit-cell parameters $a = b = 26.8$, $c = 226.3$ Å. The data-collection statistics and the crystal data are summarized in Table 1.

2.3. Structure determination and refinement

The program *SOLVE* (Terwilliger & Berendzen, 1999) was used for MAD phasing. Two Br atoms were found from the anomalous difference Patterson map, from which the initial phases were estimated with a figure of merit of 0.55. The statistics for phasing are summarized in Table 1. After density modification with solvent flattening (solvent content 43.7%) using the program *CNS* (Brünger *et al.*, 1998), two types of Bc-duplexes, I and II, were easily traced around the different crystallographic twofold axes, with one of the two strands of each duplex being in the asymmetric unit. Only the electron densities for the A_4 residues were broadened, suggesting a conformational disordering between the two antiparallel strands in respective Bc-duplexes. From the shape of the density, the two A_4 bases were assumed to form a Hoogsteen/Hoogsteen pair in combination with *syn* and *anti* conformations around the glycosidic bonds in the Bc-duplex I, as seen in Fig. 2(*a*). In the case of the Bc-duplex II, the two A_4 bases were assumed to form a Watson–Crick/Hoogsteen pair in combination with *syn* and *syn* conformations, as seen in Fig. 2(*b*).

The atomic parameters of the crystal structure were refined using the remote data through a series of rigid-body approximations, simulated-annealing runs and *B*-factor balancing using the program *CNS* with crystallographic conjugate-gradient minimization techniques, followed by interpretations of OMIT maps at every nucleotide residue. In the final refinement, the two asymmetrical strands with the two alternative conformers of the A_4 residues, along with the phosphate backbone of the A_5 residues, were refined using *REFMAC*5 (Murshudov *et al.*, 1997). Statistics of structure refinement are summarized in Table 1. The final R and $R_{\rm free}$ values are relatively high because of the conformational disorder at the central A_4 residues, as seen in their final OMIT maps (see Figs. 2*a* and 2*b*). The other parts are well fitted as shown in Figs. 2(*c*) and 2(*d*). All local helical parameters including torsion angles and pseudo-rotation phase angles of ribose rings were calculated using the program *3DNA* (Lu & Olson, 2003). Fig. 2 was drawn with the *PyMOL* molecular-graphics system (http://www.pymol.org). Figs. 3(*a*), 3(*b*), 4, 5, 6 and 7 were drawn using the program *RASMOL* (Sayle & Milner-White, 1995).

3. Results

3.1. Overall structures

The crystal of A6C-Br is composed of two types of antiparallel Bc-duplexes, I and II, shown in Fig. 3. Each duplex is composed of $G_1{\cdot}C_8$*, $C_2{\cdot}G_7$*, $G_3{\cdot}C_6$*, $A_4{\cdot}A_4$*, $C_6{\cdot}G_3$*, $G_7{\cdot}C_2$* and $C_8{\cdot}G_1$* pairs, along with the bulged A_5 and A_5*. The final OMIT maps for $A_4{\cdot}A_4$* pairs are shown in Figs. 2(*a*) and 2(*b*). The two Bc-duplexes are well fitted overall. Only the A_4 residues in both duplexes exhibit a broadened density between the two antiparallel strands, owing to their disorder, even in the final OMIT maps. In Bc-duplex I, the two A_4 bases form a Hoogsteen/Hoogsteen pair in combination with *syn* and *anti* conformations around the glycosidic bonds (see Fig. 2*a*). The electron density of the *syn* conformer is poorer at the piperidine ring. This might be ascribed to the mobility of the adenine base, as the *anti* A_4 base is completely sandwiched by the base pairs above and below, whereas the piperidine ring of the *syn* A_4 base is exposed. The short C1′···C1′ distance only allows the antiparallel *anti–syn* combination or *vice versa*. In Bc-duplex II, however, the wider space and longer C1′···C1′ distance allow the antiparallel *syn–syn* combination, so that the two A_4 bases form a Watson–Crick/Hoogsteen or a Hoogsteen/Watson–Crick pair (see Fig. 2*b*). The disordered residues have root-mean-square deviations (r.m.s.d.s) of 1.1 Å in Bc-duplex I and 0.6 Å in Bc-duplex II. If the two Bc-duplexes are superimposed, their r.m.s. coordinate difference is 4.7 Å. A remarkable difference occurs at the flipped-out A_5 residues. In Bc-duplex I, the two A_5 residues are flipped out but fold back to interact with the G_3 residues of the same strands. In Bc-duplex II, however, the flipped-out A_5 residues do not fold back and are fully extended, interacting with the G_7 residue of the neighbouring Bc-duplex I, as shown in Fig. 3(*c*). As a result, Bc-duplex I accepts two A_5 residues from two different Bc-duplexes II, so that the Bc-duplexes are linked together alternately in the crystal.

Table 2
Sugar puckers of A6C-Br.

The sugar pucker in the alternative conformer is given in parentheses.

Sequence	Duplex I	Duplex II
G_1	C3′-*exo*	C2′-*endo*
$^{Br}C_2$	C1′-*exo*	C2′-*endo*
G_3	C2′-*endo*	C2′-*endo*
A_4	C2′-*endo* (C2′-*endo*)	C1′-*exo* (C4′-*exo*)
A_5	C2′-*endo*	C4′-*endo*
C_6	C2′-*endo*	C2′-*endo*
G_7	C3′-*exo*	C2′-*endo*
C_8	C4′-*exo*	O4′-*endo*

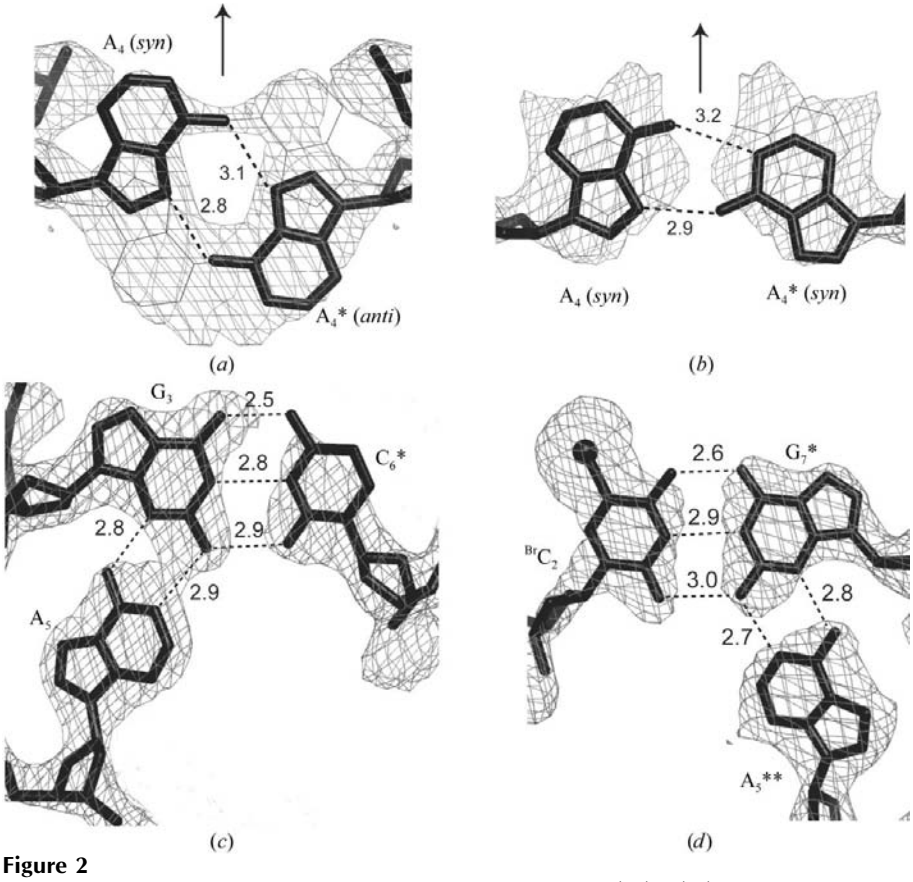

Figure 2
Local OMIT maps around the A·A* base pairs (*a*, *b*) and local $2|F_o| - |F_c|$ maps around C·G·A triplets (*c*, *d*). All maps were calculated using the final phases and contoured at the 1.0σ level. Two types of $A_4{\cdot}A_4$* base pairs are assigned to the two Bc-duplexes I (*a*) and II (*b*). In both Bc-duplexes the pair is disordered around the crystallographic twofold axis, shown by arrows. Two alternative conformers are drawn in black sticks and thin grey lines, respectively. Two triplets, $C_6{*}{\cdot}G_3{\cdot}A_5$ and $^{Br}C_2{\cdot}G_7{*}{\cdot}A_5{**}$, are found in Bc-duplex I (*c*, *d*). Broken lines indicate possible hydrogen bonds with distances (Å). * indicates the counter-strand. The A_5** residue is from Bc-duplex II.

3.2. Stem regions of Bc-duplexes I and II

At one end of Bc-duplexes I and II, three contiguous G·C base pairs are formed, $G_1{\cdot}C_8$*, $C_2{\cdot}G_7$* and $G_3{\cdot}C_6$*, and similarly at the opposite end. It seems that bromination at the second cytidine residue has no effect on Watson–Crick base-pair formation. These residues adopt C2′-*endo* puckers to maintain the canonical B-form conformation, except for the terminal C_8 residues, which have the C4′-*exo* (close to C3′-*endo*) conformation in the Bc-duplex I and the O4′-*endo* (between C2′-*endo* and C3′-*endo*) conformation in the Bc-duplex II (see Table 2). This difference may be ascribed to the surrounding solvent structure and crystal packing.

3.3. $A_4{\cdot}A_4$ base pair in Bc-duplex I

In Bc-duplex I, the Hoogsteen/Hoogsteen $A_4{\cdot}A_4$* pair, with hydrogen bonds N6—H(A_4)···N7(A_4*) and N7(A_4)···H—N6(A_4*), fits well to the

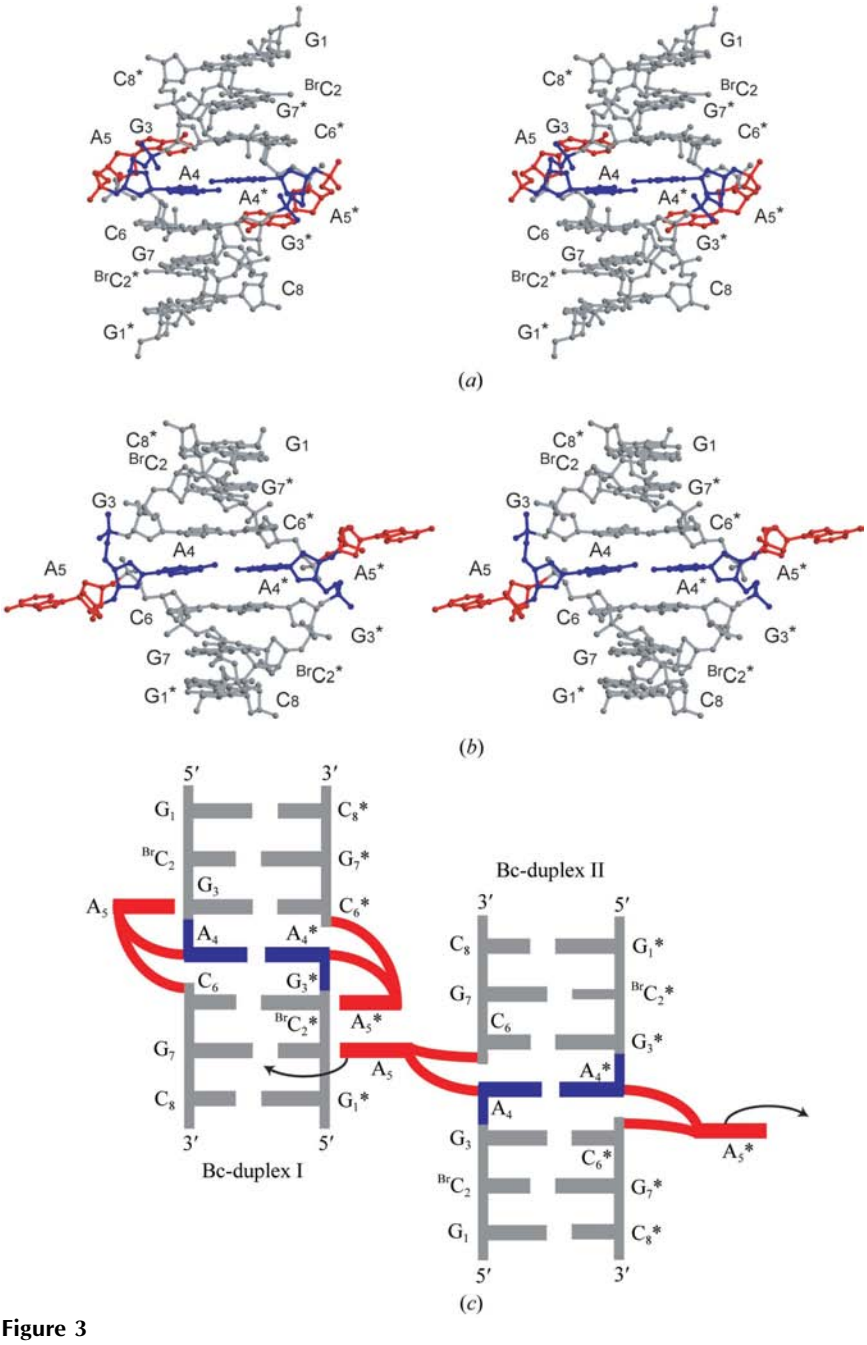

Figure 3
Stereo diagrams of Bc-duplex I (*a*), Bc-duplex II (*b*) and their schematic interaction modes (*c*). The stem regions, the $A_4{\cdot}A_4$* base pairs and the bulged-out A_5 residues are coloured grey, blue and red, respectively. Arrows indicate inter-duplex interactions.

OMIT map (see Fig. 2*a*). The geometry appears symmetric around the normal to the base planes, but the crystallographic twofold axis is in fact parallel to the base planes. The distance between the C1′ atoms (11.0 Å) is slightly longer than that (10.7 Å) of standard Watson–Crick pairs (see Table 3). This pairing is similar to that found in a parallel DNA duplex (Sunami *et al.*, 2002), in which the two adenine bases are, however, positioned *trans* to each other. In the present structure, the pair formation between the two Hoogsteen edges occurs between the two antiparallel strands, so that one of the two A_4 residues must have a *syn* conformation around the glycosidic bond, adopting the C2′-*endo* sugar pucker (see Table 2).

3.4. $A_4{\cdot}A_4$ base pair in Bc-duplex II

In Bc-duplex II, the $A_4{\cdot}A_4$* pair, with hydrogen bonds N1(A_4)···H—N6(A_4*) and N6—H(A_4)···N7(A_4*) between the Watson–Crick edge and the Hoogsteen edge of the A_4 residues, fits well to the OMIT map (see Fig. 2*b*). The distance between the C1′ atoms (12.3 Å) is longer than that of the *trans* Hoogsteen/Hoogsteen A·A base pair in Bc-duplex I (see Table 3). This type of pairing has not yet been observed in any X-ray structures of DNA duplexes[3]. To form a *trans* Watson–Crick/Hoogsteen A·A pair in an antiparallel duplex, the two A_4 residues must have a *syn* conformation and adopt C1′-*exo* (close to C2′-*endo*) and C4′-*exo* (close to C3′-*endo*) sugar puckers (see Table 2).

3.5. Bulged-out A_5 residue of Bc-duplex I

In the Bc-duplex I, the two A_5 residues are bulged out and then folded back toward their own phosphate backbones, entering pockets in the minor groove of the duplex. In each pocket, the adenine base interacts with the $G_3{\cdot}C_6$ pair to form a triplet. Two hydrogen bonds, N1(A_5)···H—N2(G_3) and N6—H(A_5)···N3(G_3) (see Fig. 2*c*) are formed with the G_3 residues of the same strand (see Fig. 4). The A_5 residue adopts a C2′-*endo* sugar pucker (see Table 2). This type of A·G·C triplet is the first such example in DNA crystal structures. This type of interaction is designated here as the 'intra-duplex hand-in-pocket' motif.

3.6. Bulged-out A_5 residue of Bc-duplex II

The two bulged-out A_5 residues of Bc-duplex II protrude from the double helix, like extended arms, on either side of the duplex (see Fig. 5). The A_5 residue adopts a C4′-*endo* sugar pucker (close to C2′-*endo*) (see Table 2). The adenine moiety is in contact with the G_7 residues in the pocket of the minor grooves of a neighbouring Bc-duplex I through the two hydrogen bonds N1(A_5)···H—N2(G_7) and N6—H(A_5)···N3(G_7) (see Fig. 2*d*). This triplet formation is of the same type as that found in Bc-duplex I (see Fig. 2*c*). The only difference is that it occurs between the two Bc-duplexes. This

[3] This type of A·A pairing has been seen in the crystal structures of small organic molecules (Takimoto-Kamimura *et al.*, 1986).

type of interaction is designated here as the 'inter-duplex hand-in-pocket' motif.

3.7. Minor-groove pockets

Bc-duplex I accepts four bulged-out A_5 residues: two involved in intra-duplex interactions and the other two in inter-duplex interactions at the two sites in the minor groove. A mixed pair of A_5 residues, one internal and one external, are stacked on each other and inserted into each site, contacting the $G_3{\cdot}C_6$* and $G_7{\cdot}C_2$* base pairs, as well as the corresponding pairs in the symmetry-related site, as shown in Fig. 6. In the Bc-duplex II, however, the corresponding minor grooves at the $G_3{\cdot}C_6$* ($C_6{\cdot}G_3$*) and $G_7{\cdot}C_2$* ($C_2{\cdot}G_7$*) base pairs are open and free from interactions with other adenosine residues.

Table 3
Local base-pair parameters of A6C-Br.

The corresponding values for the remaining half of the duplexes are omitted owing to crystallographic symmetry.

Base pair	Inclination angle (°)	Tip (°)	Twist (°)	Rise (Å)	Propeller twist angle (°)	Buckle angle (°)	Opening angle (°)	C1′···C1′ (Å)
Duplex I								
$G_1{\cdot}C_8$*					−4	2	0	10.7
	−4	−2	33	3.5				
$^{Br}C_2{\cdot}G_7$*					−9	−5	−5	10.8
	−1	−6	40	3.0				
$G_3{\cdot}C_6$*					7	14	−5	10.5
	n.d	n.d	n.d	n.d				
$A_4{\cdot}A_4$*					−6	−3	−177	11.0
Duplex II								
$G_1{\cdot}C_8$*					−6	−1	−3	10.7
	−4	0	34	3.3				
$^{Br}C_2{\cdot}G_7$*					−12	1	−2	10.9
	17	5	33	3.3				
$G_3{\cdot}C_6$*					−14	3	0	10.8
	n.d	n.d	n.d	n.d				
$A_4{\cdot}A_4$*					−9	−9	−127	12.3
A-form	15	0	33	2.8	−12	0	1	10.7
B-form	2	0	37	3.3	−11	1	1	10.7

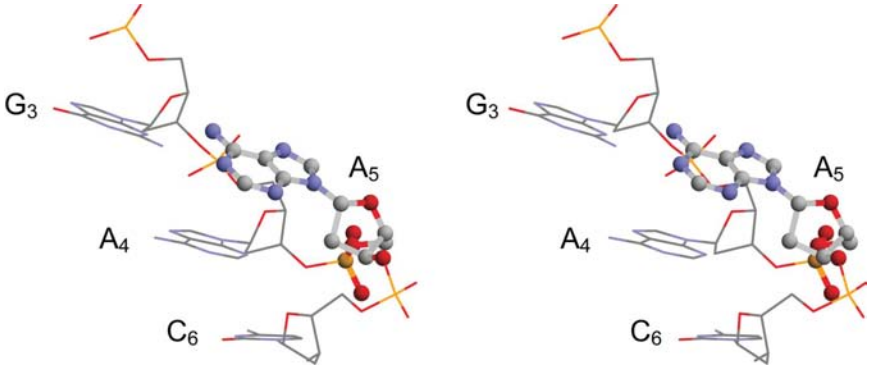

Figure 4
A stereoview of the bulged-out A_5 residue of Bc-duplex I, which interacts with the G_3 residue of the same strand.

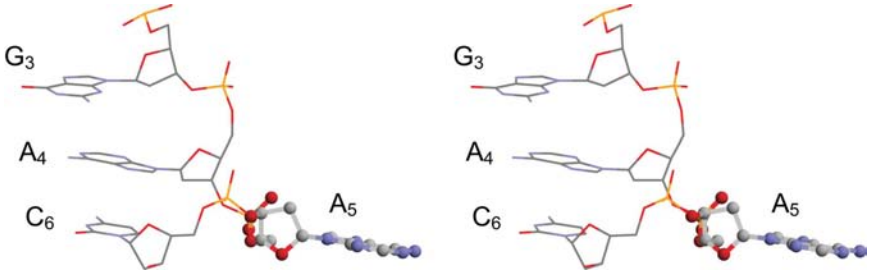

Figure 5
A stereoview of the bulged-out A_5 residue of Bc-duplex II, which extends outside to interact with the G_7 residue of the neighbouring Bc-duplex I.

3.8. Backbone conformation

There are several notable conformational features in Bc-duplex I, especially involving torsion angles around the P—O5′ bonds (α) of the A_5 residues. Small α angles (166–175°) allow the bulged-out A_5 residues to fold back to form the intra-duplex hand-in-pocket motifs. On the other hand, the C_6 residues of the Bc-duplex II, which flank the bulged-out A_5 residues, have small α angles (61°) and large C5′–C4′ (γ) angles (185°), allowing the protrusion of the A_5 nucleotide.

4. Discussion

4.1. Effect of the A6C change on the Bi-duplex formation

The present study shows that the sequence d(gcGAACgc) forms two unique types of Bc-duplexes, I and II, that differ dramatically from the Bi-duplex found in the parent d(gcGAAAgc) sequence (Sunami *et al.*, 2004*a*,*b*). It is thus clear that a single mutation at the sixth residue induces a large change in the duplex formation, which is ascribed to the difference in geometry between Watson–Crick and sheared base pairs. Fig. 7 shows a superimposition of the stem regions of Bc-duplex I and the Bi-duplex (Sunami *et al.*, 2004*a*). In the latter duplex, there is a short distance between the C1′ atoms of the $G_3{\cdot}A_6$* pair (8.4 Å) compared with that of a Watson–Crick duplex (10.7 Å). This shortening brings the two strands closer, so that the subsequent residues have no space to form base pairs between the strands. The present study definitely supports our previous proposal that the scaffold of the Bi-duplex is the two G·A sheared pairs followed by Watson–Crick pairing at both ends of the duplex. Here, again, we conclude that without these pairings the central *X* and *Y* residues cannot adopt the base-intercalated *X*–*Y**–*Y*–*X** base–base stacking essential to

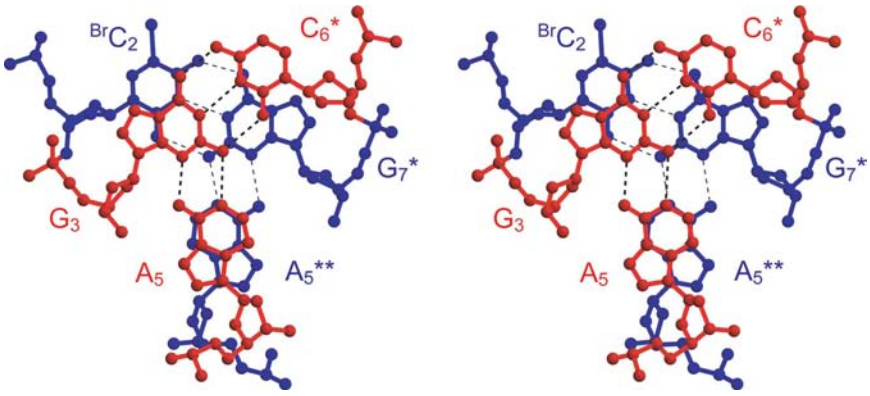

Figure 6
A stereoview showing the A_5–A_5** stacking between the intra-duplex (red) and the inter-duplex (blue) hand-in-pocket motifs in the minor groove of Bc-duplex I. The single asterisk indicates the opposite strand in Bc-duplex I. The A_5** residue is from Bc-duplex II.

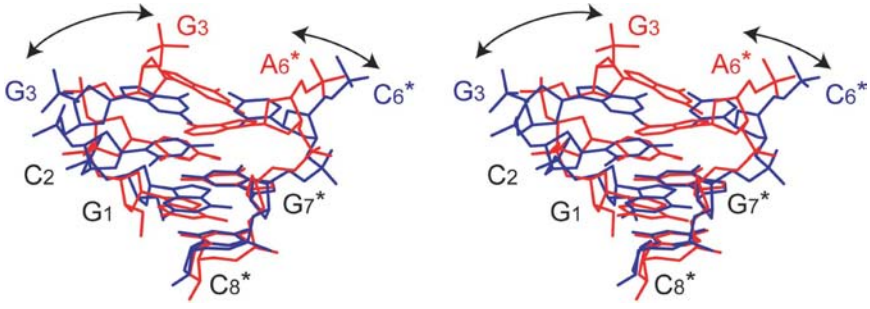

Figure 7
Superimposition of the stem regions of Bc-duplex I (blue) and the Bi-duplex (red) (Sunami *et al.*, 2004*a*). Arrows indicate the conformational change of the phosphate backbones between the Watson–Crick-type pairing and the sheared-type pairing.

multiplex formation. Interest will next be focused on the length of the central column.

4.2. A·A pair and A·G·C triplet formation

Two types of A·A pair formations have been found in the present study (see Fig. 8*a*). One occurs between the Hoogsteen edges. This pairing is similar to that found in a parallel DNA duplex (Sunami *et al.*, 2002). To form this pairing in an antiparallel DNA duplex, the two adenine bases must adopt *syn* and *anti* conformations or *vice versa* around the glycosidic bonds. This contrasts with the parallel duplex, in which the two adenines adopt *anti* conformations (see Fig. 8*b*). As another example, a similar A(*anti*)·A(*syn*) mismatch is found in the antiparallel DNA duplex complexed with the *Escherichia coli* DNA-mismatch repair enzyme (MutS; Natrajan *et al.*, 2003).

The other A·A base pair of the Bc-duplex II is formed between the Watson–Crick edge and the Hoogsteen edge, both of which adopt a *syn* conformation. This is the first example of such a pairing in a DNA duplex, although it has been reported in several antiparallel RNA duplexes (Baeyens *et al.*, 1996), including the 50S ribosomal RNA from *Haloarcula marismortui* (Ban *et al.*, 2000) and the 30S ribosomal RNA from *Thermus thermophilus* (Wimberly *et al.*, 2000; Carter *et al.*, 2000), in which the adenine bases are in the *anti* conformation.

The bulged-out A residues from the two Bc-duplexes form an A·G·C ternary triplet in the Bc-duplex I (see Fig. 9*a*). This triplet formation differs from that found in d(AGGCATGCCT) (see Fig. 9*b*; Nunn & Neidle, 1996). In the present triplet, the Watson–Crick edge of the adenine residue interacts with the sugar edge of the guanine residue through two hydrogen bonds, N1···H—N2 and N6—H···N3. This A(*anti*)·G(*anti*) interaction is also a new interaction that has never previously been found in structural studies of DNA molecules. Recently, a similar interaction was found in functional RNA molecules: the ribosomal frameshifting viral pseudoknot (Su *et al.*, 1999) and the malachite green aptamer (Baugh *et al.*, 2000). Moreover, although they are rare in DNA crystal structures (Sunami *et al.*, 2004*b*; Joshua-Tor *et al.*, 1992), bulged-in and bulged-out bases have been found in X-ray structures of functional RNAs, including the prokaryotic and eukaryotic cytoplasmic ribosomal decoding A site (Wimberly *et al.*, 2000; Carter *et al.*, 2000; Vicens & Westhof, 2001, 2002, 2003; François *et al.*, 2004, 2005; Shandrick *et al.*, 2004; Zhao *et al.*, 2005; Han *et al.*, 2005; Kondo, François, Russell *et al.*, 2006; Kondo, Urzhumtsev *et al.*, 2006; Kondo, François, Urzhumtsev *et al.*, 2006), the dimerization-initiation site of genomic HIV-1 RNA (Ennifar *et al.*, 2001; Ennifar & Dumas, 2006), helix II of *Xenopus laevis* 5S rRNA (Xiong & Sundaralingam, 2000) and group II self-splicing introns (Zhang & Doudna, 2002). Therefore, it is expected that the present interaction motifs may be involved in various biological processes and may be useful for designing new functional DNA molecules.

4.3. Biological significance of the consecutive bulged A residues in a DNA duplex

In the present study, a new type of duplex containing two consecutive bulged A residues has been found, with the two residues bulged-in and bulged-out, respectively. The bulged-in A_4·A_4* pairs are either Hoogsteen/Hoogsteen or Watson–Crick/Hoogsteen (see Fig. 8*a*) depending on the directions of the bulged-out A_5 residues. One is folded back to interact with the minor groove of the same duplex, while the other is extended to interact with the minor groove of the neighbouring duplex. This suggests that the bulged-in A·A pairs are flexible in the Bc-duplex, consistent with their slight disordering, so that the central bulged-in A·A pairs are responsible for the movement of the bulged-out residues. The sequence d(gcGAACgc) itself does not have any obvious biological

meaning, but a pair of bulged-in and bulged-out A residues could be useful as a molecular switch. Such a pair is exemplified with the ribosomal decoding A site. It has been revealed by X-ray analyses that the prokaryotic A site as well as the eukaryotic cytoplasmic A site have two conformational states, the 'off' and the 'on' states (Carter *et al.*, 2000; Kondo, Urzhumtsev *et al.*, 2006). In the 'off' state, two adenine residues (A1492 and A1493) or one of the two are bulged in. On the other hand, the 'on' state has two bulged-out adenine residues. When the cognate tRNA is delivered to the A site, the A site changes its conformation from the 'off' state to the 'on' state. The two bulged-out adenine residues can recognize the first two Watson–Crick base pairs in the cognate codon–anticodon helix, which leads to the high fidelity of tRNA-selection step. A similar situation is also proposed for the catalytic action of U2-snRNA (Berglund *et al.*, 2001). In the present study, the two hand-in-pocket motifs of the bulged-out residues are the first examples to be observed. It is expected that these interaction motifs could be involved in various biological processes in which DNA is in a single-stranded state and has a complicated tertiary structure, like RNA. In addition, these motifs could be applicable for the design of new functional DNA molecules. Recently, several RNA molecules were designed, including dimeric nanoparticles (Jaeger & Leontis, 2000; Jaeger *et al.*, 2001; Liu *et al.*, 2004), micrometre-long RNA filaments (Jaeger & Leontis, 2000; Nasalean *et al.*, 2006) and a new class of self-folding RNA molecules similar to domain P4-P6 of the natural *Tetrahymena* group I intron ribozyme (Ikawa *et al.*, 2002). These RNA molecules were designed by RNA architectonics based on the accumulated structural database of natural RNA (Westhof *et al.*, 1996; Jaeger & Chworos, 2006). Since DNA molecules are chemically stable, they could be very useful for several purposes. However, our knowledge of structural motifs of DNA is still limited. More extensive studies of DNA in single-stranded states, including non-Watson–Crick pairings and/or repeated sequences are required.

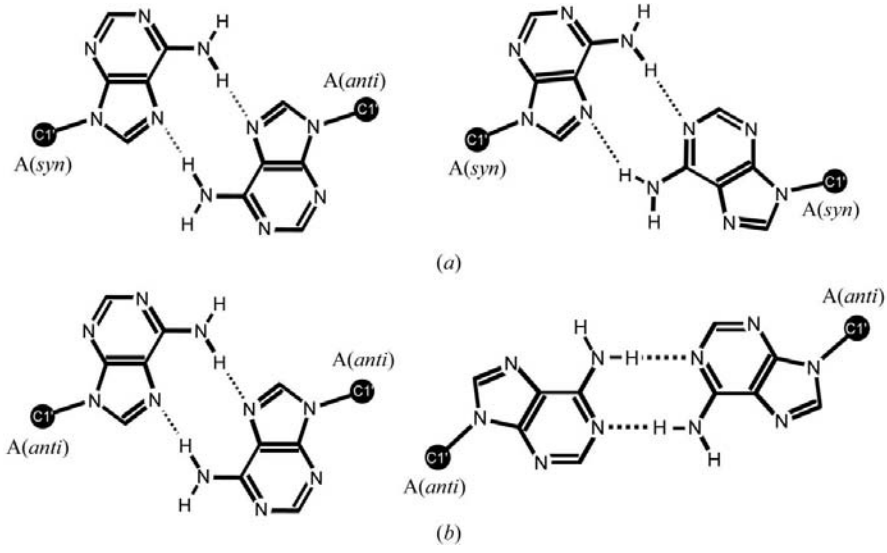

Figure 8
A·A pairing modes of antiparallel DNA duplexes found in the present structure (*a*) and those of the parallel DNA duplex (*b*) (Sunami *et al.*, 2002).

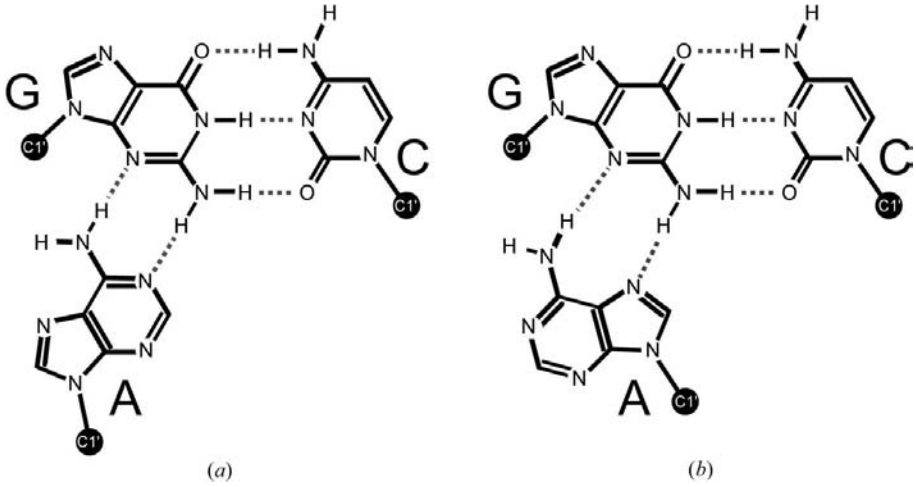

Figure 9
The type of A·G·C triplet formation found in the present structure (*a*) and that reported previously in X-ray analysis of DNA fragments (*b*) (Nunn & Neidle, 1996).

We thank M. Suzuki, N. Igarashi and A. Nakagawa for facilities and help during data collection and Thomas Simonson for proofreading the original manuscript. This work was supported in part by Grants-in-Aid for Scientific Research (Nos. 12480177 and 14035217) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Armour, J. A., Wong, Z., Wilson, V., Royle, N. J. & Jeffreys, A. J. (1989). *Nucleic Acids Res.* **17**, 4925–4935.
Baeyens, K. J., De Bondt, H. L., Pardi, A. & Holbrock, S. R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12851–12855.
Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–920.
Baugh, C., Grate, D. & Wilson, C. (2000). *J. Mol. Biol.* **301**, 117–128.
Berglund, J. A., Rosbash, M. & Schultz, S. C. (2001). *RNA*, **7**, 682–691.
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.

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 340–348.

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.

Chou, S.-H. & Chin, K. H. (2001). *J. Mol. Biol.* **314**, 139–152.

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

De Luca, F., Reyes, A., Veronico, P., Di Vito, M., Lamberti, F. & De Giorgi, C. (2002). *Gene*, **293**, 191–198.

Doudna, J. A. (1995). *Structure*, **3**, 747–750.

Ennifar, E. & Dumas, P. (2006). *J. Mol. Biol.* **356**, 771–782.

Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. & Dumas, P. (2001). *Nature Struct. Biol.* **8**, 1064–1068.

François, B., Russell, R. J., Murray, J. B., Aboul-ela, F., Masquida, B., Vicens, Q. & Westhof, E. (2005). *Nucleic Acids Res.* **33**, 5677–5690.

François, B., Szychowski, J., Adhikari, S. S., Pachamuthu, K., Swayze, E. E., Griffey, R. H., Migawa, M. T., Westhof, E. & Hanessian, S. (2004). *Angew. Chem. Int. Ed. Engl.* **43**, 6735–6738.

Golden, B. L., Kim, H. & Chase, E. (2005). *Nature Struct. Mol. Biol.* **12**, 82–89.

Han, Q., Zhao, Q., Fish, S., Simonsen, K. B., Vourloumis, D., Froelich, J. M., Wall, D. & Hermann, T. (2005). *Angew. Chem. Int. Ed. Engl.* **44**, 2694–2700.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. & Yonath, A. (2001). *Cell*, **107**, 679–688.

Ikawa, Y., Fukada, K., Watanabe, S., Shiraishi, H. & Inoue, T. (2002). *Structure*, **10**, 527–534.

Inglehearn, C. F. & Cooke, H. J. (1990). *Nucleic Acids Res.* **18**, 471–476.

International Human Genome Sequencing Consortium (2001). *Nature (London)*, **409**, 860–921.

International Human Genome Sequencing Consortium (2004). *Nature (London)*, **431**, 931–945.

Jaeger, L. & Chworos, A. (2006). *Curr. Opin. Struct. Biol.* **16**, 531–543.

Jaeger, L. & Leontis, N. B. (2000). *Angew. Chem. Int. Ed. Engl.* **39**, 2521–2524.

Jaeger, L., Westhof, E. & Leontis, N. B. (2001). *Nucleic Acids Res.* **29**, 455–463.

Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D. & Sussman, J. L. (1992). *J. Mol. Biol.* **225**, 397–431.

Kasai, K., Nakamura, Y. & White, R. (1990). *J. Forensic Sci.* **35**, 1196–1200.

Kondo, J., Adachi, W., Umeda, K., Sunami, T. & Takénaka, A. (2004). *Nucleic Acids Res.* **32**, 2541–2549.

Kondo, J., François, B., Russell, R. J., Murray, J. B. & Westhof, E. (2006). *Biochimie*, **88**, 1027–1031.

Kondo, J., François, B., Urzhumtsev, A. & Westhof, E. (2006). *Angew. Chem. Int. Ed. Engl.* **45**, 3310–3314.

Kondo, J., Tanashaya, C., Juan, E. C. M., Sato, Y., Mitomi, K., Shimizu, S. & Takénaka, A. (2006). *Nucleosides Nucleotides Nucleic Acids*, **25**, 693–704.

Kondo, J., Umeda, U., Fujita, K., Sunami, T. & Takénaka, A. (2004). *J. Synchrotron Rad.* **11**, 117–120.

Kondo, J., Urzhumtsev, A. & Westhof, E. (2006). *Nucleic Acids Res.* **34**, 676–685.

Leontis, N. & Westhof, E. (2001). *RNA*, **7**, 499–512.

Leslie, A. G. W. (1992). *Crystallography Computing 5. From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 50–61. Oxford University Press.

Liu, B., Baudrey, S., Jaegar, L. & Bazan, G. C. (2004). *J. Am. Chem. Soc.* **126**, 4076–4077.

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

McCouch, S. R., Teytelman, L., Xu, Y., Lobos, K. B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y., Zhang, Q., Kono, I., Yano, M., Fjellstrom, R., DeClerck, G., Schneider, D., Cartinhour, S., Ware, D. & Stein, L. (2002). *DNA Res.* **9**, 199–207.

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

Nasalean, L., Baudrey, S., Leontis, N. & Jaeger, L. (2006). *Nucleic Acids Res.* **34**, 1381–1392.

Natrajan, G., Lamers, M. H., Enzlin, J. H., Winterwerp, H. H., Perrakis, A. & Sixma, T. K. (2003). *Nucleic Acids Res.* **31**, 4814–4821.

Nunn, C. M. & Neidle, S. (1996). *J. Mol. Biol.* **256**, 340–351.

Patel, D. J., Bouaziz, A., Kettani, A. & Wang, Y. (1999). *Nucleic Acid Structure*, edited by S. Neidle, pp. 389–453. Oxford University Press.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). *Nature (London)*, **372**, 68–74.

Powell, H. R. (1999). *Acta Cryst.* D**55**, 1690–1695.

Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.

Sato, Y., Mitomi, K., Sunami, T., Kondo, J. & Takénaka, A. (2006). *J. Biochem.* **140**, 759–762.

Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374.

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. & Cate, J. H. (2005). *Science*, **310**, 827–834.

Scott, W. G., Finch, J. T. & Klug, A. (1995). *Cell*, **81**, 991–1002.

Shandrick, S., Zhao, Q., Han, Q., Ayida, B. K., Takahashi, M., Winters, G. C., Simonsen, K. B., Vourloumis, D. & Hermann, T. (2004). *Angew. Chem. Int. Ed. Engl.* **43**, 3177–3182.

Shibuya, H., Collins, B. K., Collier, L. L., Huang, T. H., Nonneman, D. & Johnson, G. S. (1996). *Anim. Genet.* **27**, 59–60.

Stahley, M. R. & Strobel, S. A. (2005). *Science*, **309**, 1587–1590.

Steller, I., Bolotovsky, R. & Rossmann, M. G. (1997). *J. Appl. Cryst.* **30**, 1036–1040.

Su, L., Chen, L., Egli, M., Berger, J. M. & Rich, A. (1999). *Nature Struct. Biol.* **6**, 285–292.

Sunami, T., Kondo, J., Hirao, I., Watanabe, K., Miura, K. & Takénaka, A. (2004*a*). *Acta Cryst.* D**60**, 90–96.

Sunami, T., Kondo, J., Hirao, I., Watanabe, K., Miura, K. & Takénaka, A. (2004*b*). *Acta Cryst.* D**60**, 422–431.

Sunami, T., Kondo, J., Kobuna, T., Hirao, I., Watanabe, K., Miura, K. & Takénaka, A. (2002). *Nucleic Acids Res.* **30**, 5253–5260.

Takimoto-Kamimura, M., Takenaka, A. & Sasada, Y. (1986). *Acta Cryst.* C**42**, 601–603.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Vicens, Q. & Westhof, E. (2001). *Structure*, **9**, 647–658.

Vicens, Q. & Westhof, E. (2002). *Chem. Biol.* **9**, 747–755.

Vicens, Q. & Westhof, E. (2003). *J. Mol. Biol.* **326**, 1175–1188.

Westhof, E., Masquida, B. & Jaeger, L. (1996). *Fold Des.* **1**, 78–88.

Williams, S., Hayes, L., Elsenboss, L., Williams, A., Andre, C., Abramson, R., Thompson, J. F. & Milos, P. M. (1997). *Gene*, **197**, 101–107.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M. Jr, Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 327–339.

Xiong, Y. & Sundaralingam, M. (2000). *RNA*, **6**, 1316–1324.

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). *Science*, **292**, 883–896.

Zhang, L. & Doudna, J. A. (2002). *Science*, **295**, 2084–2088.

Zhao, F., Zhao, Q., Blount, K. F., Han, Q., Tor, Y. & Hermann, T. (2005). *Angew. Chem. Int. Ed. Engl.* **44**, 2–6.