Structure and Recognition of Sheared Tandem G·A Base Pairs Associated with Human Centromere DNA Sequence at Atomic Resolution†

Yi-Gui Gao,‡ Howard Robinson,‡ Ruslan Sanishvili,§ Andrzej Joachimiak,§,|| and Andrew H.-J. Wang*,‡

Department of Cell & Structural Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, Structure Biology Center, Bioscience Division, Argonne National Laboratory, Argonne, Illinois 60439, and Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208

Received June 25, 1999; Revised Manuscript Received October 13, 1999

ABSTRACT: G·A mismatched base pairs are frequently found in nucleic acids. Human centromere DNA sequences contain unusual repeating motifs, e.g., $(GAATG)_n$·$(CATTC)_n$ found in the human chromosome. The purine-rich strand of this repeating pentamer sequence forms duplex and hairpin structures with unusual stability. The high stability of these structures is contributed by the "sheared" G·A base pairs which present a novel recognition surface for ligands and proteins. We have solved the crystal structure, by the multiple-wavelength anomalous diffraction (MAD) method of d(CCGAATGAGG) in which the centromere core sequence motif GAATG is embedded. Three crystal forms were refined to near-atomic resolution. The structures reveal the detailed conformation of tandem G·A base pairs whose unique hydrogen-bonding surface has interesting interactions with bases, hydrated magnesium ions, cobalt(III)hexaammine, spermine, and water molecules. The results are relevant in understanding the structure associated with human centromere sequence in particular and G·A base pairs in nucleic acids (including RNA, like ribozyme) in general.

A number of DNA sequences have been suggested to form stable duplex and hairpin structures incorporating unusual base pairs (*1*). Of particular interest are those associated with repetitive sequences including the centromere $(GAATG)_n$ sequence (*2*−*5*), the triplet $(GGA)_n$ sequence associated with Friedreich's ataxia (*6*), the $(GTACGGGACCGA)_n$ dodeca repeat in the centromeric Drosophila dodeca satellite DNA (*7*). Extensive biochemical and biophysical studies on these sequences suggested that the so-called "sheared" G·A base pairs (Figure 1A) are involved in these structures and the high stability of these structures is contributed by the tandem "sheared" G·A base pairs which present a novel recognition surface for ligands and proteins.

In fact, G−A mismatched base pairs (single or tandem) are also frequently found in RNA (*8*), and they play important structural [e.g., in ribosome (*9*, *10*)] and functional roles [in hammerhead ribozymes (*11*−*13*)]. Tandem G·A base pairs in RNA are very stable (*14*). Several RNA structures containing sheared tandem G·A base pairs have been analyzed, including duplexes solved by NMR (*15*, *16*) and a duplex (*17*) and two ribozymes (*11*, *12*) solved by X-ray diffraction.

It is interesting to note that when the GpA step is embedded in the sequence of 5′-(pyrimidine)-GA-(purine), the tandem G−A mismatches adopt a sheared conformation (*1*, *18*−*20*), but when it is embedded in the sequence of 5′-(purine)-GA-(pyrimidine), the tandem G−A mismatches adopt the nonsheared conformation (*21*). The structural basis for this switch of base pair scheme is not yet clear.

While several NMR structures of DNA containing tandem sheared G·A sequences have been analyzed (*1*, *4*, *18*−*20*), surprisingly, no X-ray structure has been solved so far. In this paper, we present three high-resolution crystal structures of d(CCGAATGAGG) (which incorporates the centromere core sequence motif GAATG), determined by the MAD method. The structures are compared with other NMR DNA structures and with RNA structures derived from NMR and X-ray diffraction analysis. In addition, the observation of various molecules interacting with the novel surface of the "sheared" G·A base pairs provides clues for their recognition and interaction by small ligands and proteins.

MATERIALS AND METHODS

The DNA molecules, including the brominated derivatives, were synthesized at the DNA synthesis facility at University of Illinois at Champaign-Urbana and purified by gel-filtration column chromatography. Three crystal forms were obtained using the vapor diffusion method (*22*), and their crystallographic statistics are listed in Table 1. The long $C222_1$ form of d(CCGAA[br^{5}U]GACC) was crystallized from 1.3 mM DNA duplex, 2 mM Tris/HCl buffer (pH 7.5), 5 mM $MgCl_2$, 5 mM $Co(NH_3)_6^{3+}$, and 2.6% 2-methyl-2,4-pentanediol (2-MPD) solution, equilibrated with 30 mL of 15% 2-MPD. The short $C222_1$ form of CCGAA[br^{5}U]GACC) was crystallized from the same condition without $Co(NH_3)_6^{3+}$, except no reservoir solution was used in the initial stage of the

† This work was supported by NIH (GM41612) and NSF (MCB98-08298) to A.H.-J.W. and by the U.S. Department of Energy, Office of Health and Environmental Research, under contract W-31-109-Eng-38 to A.J.

* To whom correspondence should be addressed. Fax: (217) 244-3181. E-mail: ahjwang@uiuc.edu.

‡ University of Illinois at Urbana−Champaign.

§ Argonne National Laboratory.

|| Northwestern University.

10.1021/bi9914614 CCC: $18.00
Published on Web 11/24/1999

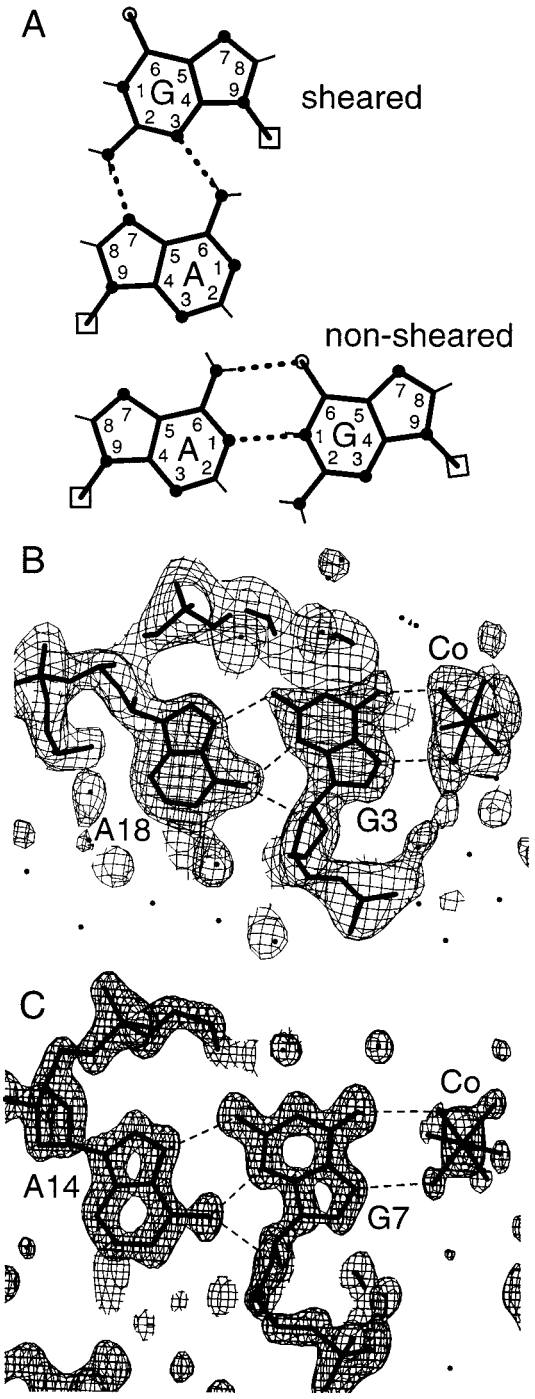

FIGURE 1: (A) A schematic diagram showing the "sheared" and "nonsheared" G−A base pairs. Other types of G−A base pair also exist (see, for example, ref *49*). (B) The Fourier electron density map (contoured at 1σ level) of the G3−A18 base pair of the long $C222_1$ form after density modification of the MAD phases at 2.5 Å resolution was applied. (C) The refined ($2F_o - F_c$) Fourier electron density map (contoured at 1σ level) of the G7−A14 base pair of the C2 form at 1.2 Å resolution.

crystallization process for 1 week and then 30 mL of 40% 2-MPD reservoir was used. The *C*2 form of d(CCGAATGACC) was obtained using the same condition as that of the long $C222_1$ form. Diffraction data were collected either at ambient temperature or at 123 K on a Rigaku R-Axis IIc image plate area detector system. The data have been processed using the software (d*trek, v.4.7) from Molecular Structure Co (The Woodlands, Texas).

Experimental phases from the d(CCGAA[br^{5}U]GAGG) crystals of the long $C222_1$ form were obtained by the MAD method using the approach similar to that described recently (*23*). The exact energy of the absorption edge of brominated DNA was determined by recording the fluorescence spectra from a crystal. These data were analyzed using program CHOOCH (*24*), which produces a plot of both f' and f''. The data at three wavelengths (i.e., at inflection point, at peak and at low energy remote from the peak) were measured at Structural Biology Center undulator beamline 19ID at the Advanced Photon Source using fundamental undulator harmonics and 3 × 3 mosaic CCD detector (*25*). The experiment was carried out at 110 K from a single crystal. The crystal was not aligned along any particular crystal axis, and at each wavelength and complete data set was collected in one continuous scan.

Crystallographic data integration and reduction was done with the program package HKL2000 (*26*). Two bromine sites were found using the Patterson heavy-atom search method, and the phases were calculated semiautomatically as implemented in the crystallographic suite CNS (*27*). The figure-of-merit is 0.79 and 0.89 before and after density modification, respectively. The resulting electron density maps were of high quality, revealing not only the oligonucleotide chain but also the metal ions and many water molecules bound to DNA (Figure 1B). Some of the essential statistics of the MAD phasing are given in Table 1. The structures of two other crystal forms were solved by the molecular replacement method using structures modified from the brominated DNA as starting search models.

For the short $C222_1$ form, the molecular replacement search result revealed that the decamer structure cannot be a full duplex because two end base pairs from adjacent symmetry-related duplexes were superimposed on top of each other. This is consistent with the significantly shorter (by ~4 Å) *c*-axis. Subsequently, the search was performed using only the nonamer duplex. After the initial refinement, the missing terminal nucleotides were located from the difference Fourier maps and they were found to be in extended conformations as discussed in more detail later.

All three structures have been refined, first by the simulated annealing procedure incorporated in X-PLOR (*28*), and then by SHELX97 (*29*). The DNA force field parameters of Parkinson et al. (*30*) with modifications to allow sugar pucker to vary were used. Water molecules were located by the procedure incorporated in SHELX97. We were careful in the criteria of the inclusion of water molecules. Only first shell waters and well-defined higher shell waters were included. For the C2 form, anisotropic temperature factors were applied for all atoms of DNA, $Co(NH_3)_6^{3+}$ ion, and hydrated magnesium ions, and isotropic temperature factors were applied for nonliganded water molecules. No hydrogen atom was included. For the two orthorhombic forms, all atoms were treated with isotropic temperature factors and the HOPE procedure (*31*) was applied. For all three structures, the procedure of SWAT in SHELX97 by Moews and Kretsinger (*32*) has been applied to model diffuse solvent. The atomic coordinates of the three crystal structures have been deposited at Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (accession numbers 1D8X, 1D9R, and 1DCR).

Table 1: Crystallographic and Refinement Data of Three Crystal Forms of CCGAATGAGG

	$C222_1$ (long)[a]					
	0.9611 Å long_remote	0.9192 Å peak	0.9198 Å inflection point	0.9611 Å long_remote	C2	$C222_1$ (short)[a]
crystallographic data						
a (Å)	22.71	22.80	22.80	22.81	62.78	20.95
b (Å)	67.22	67.11	67.13	67.12	35.69	61.86
c (Å)	72.50	72.28	72.31	72.33	22.47	68.49
β(deg)					104.38	
resolution (Å)	1.50	2.00	2.00	2.00	1.20	1.60
no. of observed reflections		60 339	60 243	60 179		
$\langle I/\sigma(I)\rangle$		13.8	12.4	14.7		
phasing power		3.34	3.47	2.73		
no. of unique reflections [$>1.0\sigma(F)$]	8923	7242	7246	7238	12 550	5749
R_{merge} (%)	5.5	3.9	4.1	3.4	4.50	7.50
completeness (%)	99.9	99.7	99.7	99.3	82.0	85.1
temp of data collection (K)	110	110	110	110	123	293
refinement data						
no. of reflections [$>2.0\sigma(I)$]	8120				11 236	5409
R-factor/*R*-free (5% data)	0.231/0.274				0.182/0.238	0.192/0.261
rmsd bond distance (Å)	0.010				0.009	0.009
rmsd bond angle (deg)	1.98				2.16	2.03
no. of DNA atoms	412				412	412
no. of waters	133				128	77
no. of cations	0.5 $Co(NH_3)_6^{3+}$				1 $Co(NH_3)_6^{3+}$ 2 $Mg(H_2O)_5^{2+}$	3 $Mg(H_2O)_5^{2+}$ 0.5 Spermine

[a] DNA is CCGAA[br^{5}U]GAGG.

RESULTS AND DISCUSSION

Structural Features. The decamer d(CCGAATGAGG) has been crystallized in three different crystal forms (Table 1), depending on the subtle variations of crystallization conditions. Two orthorhombic $C222_1$ forms were produced using the 5-bromodeoxyuridine (replacing the T6 residue) derivative. The long $C222_1$ form of the decamer duplex was used to solve the structure by the MAD method (Figure 1B). The short $C222_1$ form was subsequently solved by the molecular replacement (MR) method. As described above in the Materials and Methods, the latter structure contains an unexpected feature of a nine base-pair duplex plus two terminal swung-out G and C bases, resulting in an open end of the helix, with those bases involved in ion binding and crystal contacts. This is similar to that found in some DNA structures (*33*, *34*). The C2 form, which was also solved by the MR method, consists of a complete decamer duplex. All three structures have been well refined as evident from the refinement statistics (Table 1) and the quality of the electron density map (Figure 1C). The structures of the C2 form and the short $C222_1$ form are shown in Figure 2.

Although the incorporation of two tandem GpA steps does not cause the duplex conformation to depart radically from B-DNA, it causes the two pentanucleotide segments of the decamer duplex to kink toward the major groove at the central A5pT6 step (Figure 3). Notably, the base stacking at the GpA step is unusual, involving extensive *inter*strand base stacking. The consequence is that while the base stacking is continuous with a gentle left-handed twist, the base stacking along a strand is actually interrupted at each GpA step (indicated by a heavy bracket in Figure 2, right panels). This characteristic stacking pattern was also noted in the homo-base-paired parallel-stranded Π-DNA structure (*35*).

There are six independent copies of the CGAA/TGAG tetranucleotide motifs obtained from the three crystal forms, thus providing an excellent insight into the detailed conformation of the sheared tandem G−A mismatched base pair (Figure 4, panels A and B). There are several outstanding features. First, the consistency of the conformations is high, with the rmsd from those six copies being 0.62 Å (Figure 4B). Second, the sheared G−A base pair has large propeller twist angle (avg 26.2°) and buckle angle (avg 25°). Interestingly, the buckles associated with the sheared G−A base pairs in the 5′-CGAT sequence are not the same. The G−A base pair close to the C−G side has a high buckle value of 29°, whereas the one close to the A−T base pair has a low value of 19°. Third, the sheared G−A base pair conformation produces highly uneven helical twist angles (Ω) in the duplex, resulting in an irregular right-handed double helix. The Ω angle values of the GpA steps, because of the sheared base pair conformation, range 91.7−97.5°, with an average value of 93.1°. Fourth, three strong hydrogen bonds are formed between the G and A nucleotides. The detailed helical parameters are listed in Tables 1S, 2S, and 3S (Supporting Information) for the *C*2, $C222_1$ (long form), and $C222_1$ (short form) structures, respectively.

The unique structural features work together in concert to provide the unusual stability for the tandem G−A base pairs (*36*−*38*). The sheared conformation allows the N6/N7 side of the adenine base to pair with the N2/N3 side of the guanine base such that two hydrogen bonds are formed between the G and A bases: AN7−GN2 (avg 2.95 Å) and AN6−GN3 (avg 3.00 Å). The high propeller twist angle and buckle further allow the adenine base to move toward the deoxyribose of the guanine nucleotide so that the A−N6 amino group can form an additional hydrogen bond to the O4′ of the deoxyribose of G (avg 3.01 Å). Thus, the three hydrogen bonds of the sheared G·A base pair, plus the extensive *inter*strand base stacking, provide the molecular basis for the observed stability.

The backbone torsion angles adopt B-DNA-like values with a few exceptions (Tables 1S, 2S, and 3S). The average

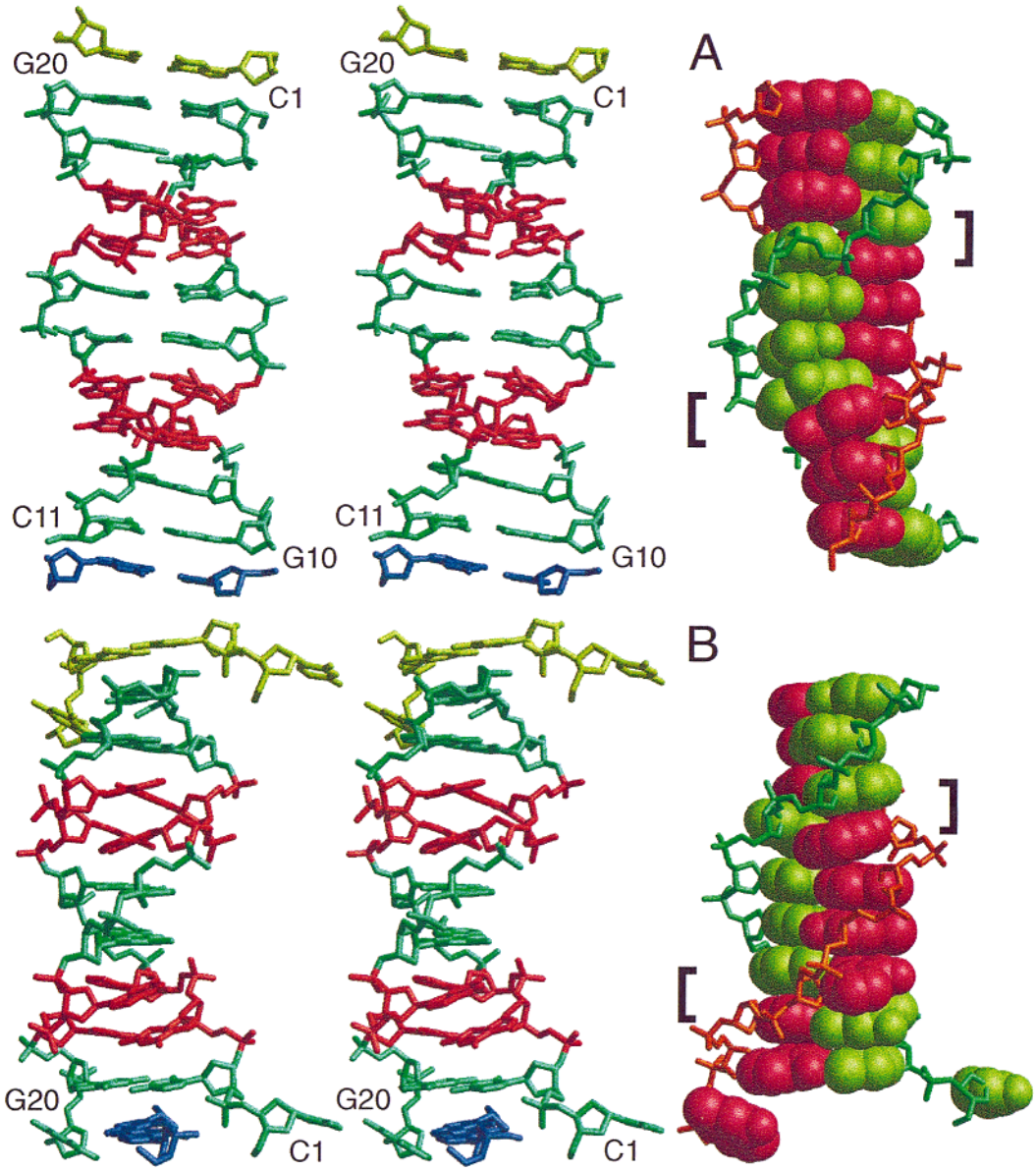

FIGURE 2: (A) The structure of the d(CCGAATGAGG) duplex of the C2 form crystal. The two tandem G−A base pairs are colored red in the stereoscopic skeletal drawings (left panel). Duplexes related by the C-centered operation are stacked end-over-end to form a virtual long DNA polymer. The terminal base pairs of the neighboring duplexes are shown. The base pair stacking switches strands at the GpA step (marked by the heavy brackets), clearly visible in the left panel. (B) The structure of the short $C222_1$ form crystal. The terminal G20−C1 base pair is unraveled with the C1 and G20 residues hang in the minor groove and major groove, respectively, of the adjacent duplex.

values (excluding outliers) are α(290°), β(170°), γ(55°), δ(130°), ϵ(207°), ζ(275°), and χ(260°). The most apparent deviations are the ζ values of the G residues (avg 162°, a trans conformation) of the GpA steps. Most of the sugars have the S-type pucker (C2′-endo) with an averaged pseudorotation angle of 175°. However, all A nucleotides (A4, A5, A14, and A15) adopt N-type (C3′-endo) sugar puckers with an averaged pseudorotation angle of 7.5°. Moreover, some other small variations exist among the three crystal forms.

The tandem G−A base pairs, related by a local 2-fold symmetry, have extensive *inter*strand G-over-G and A-over-A stacking interactions (Figure 4A). The two adenines, as well as the two guanines, from the opposite strands are nearly parallel to each other. The two outer base pairs in the (CGAA)·(TGAG) fragment have normal propeller twist angles in the range of 6−9°. Therefore, the high propeller twist of the tandem G·A base pairs does not propagate much beyond the (CGAA)·(TGAG) fragment. The averaged base pair inclination angle is 7.7°. The overall helix length of the present decamer duplex remains very similar to that of B-DNA. As noted before, the inclusion of a tandem G·A step into B-DNA can cause a significant kink in the helical axis toward the major groove at the interface of the CGAA motif and the adjoining B-DNA helix. Such a distortion plus the local unusual sheared G−A conformation may be important for biological activity. The sheared G−A base pair conformation leaves the O6−N1−N2 edge of guanine base in the major groove completely accessible for interactions with other molecules. It is possible that those unique surface features associated with the sheared G−A base pair conformation may be sites for specific recognition by proteins or nucleic acids. For example, DNase I does not recognize the DNA conformation with a sheared G−A base pair, whereas

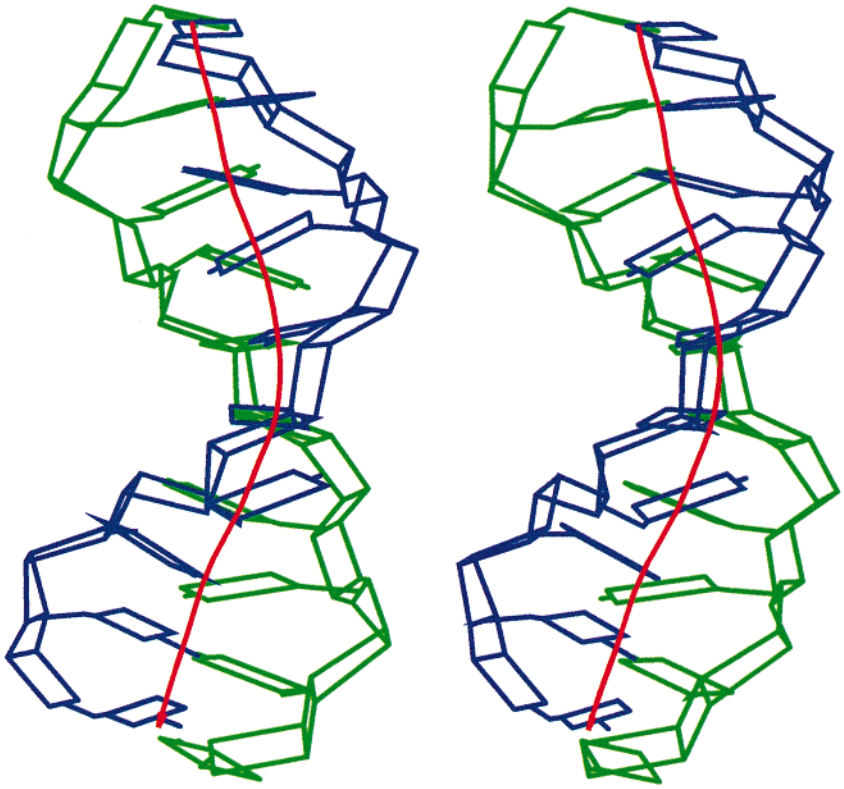

FIGURE 3: Ribbon diagram of the structure of the d(CCGAATGAGG) duplex in the C2 form crystal depicted by the program CURVES v5.3 (50). The helix axis is kinked by 38° at the central A6pT6 step toward the major groove. The kink angles in the long and short $C222_1$ forms are 28° and 31°, respectively.

it can act on the DNA backbone with the G−A base pair in the nonsheared $G_{anti}-A_{anti}$ conformation (i.e., with GN1−AN3 hydrogen bond) (*39*).

Comparison with Other G−A Motifs. As mentioned above, G−A mismatched base pairs are frequently found in nucleic acids. Until this work, no DNA crystal structure containing tandem sheared G·A base pairs has been solved. Only one DNA crystal structure (at 2.1 Å resolution) is available that contains two nontandem sheared G·A base pairs embedded in an unusual zipper-like duplex (*40*). The conformation of the individual G·A base pairs in that structure is similar to that found in our structure.

Several NMR structures of DNA containing tandem G·A sequences have been analyzed (*1*, *4*, *18*−*20*). Interestingly, the structure of GCGAATGAGC, obtained by the NMR distance geometry method (*4*), is remarkably similar to our crystal structure. The rmsd of the central eight base pairs between the structure of the C2 form and the NMR structure is 1.33 Å. The major differences between the two structures are (1) the G·A pair base in the NMR structure is less sheared, causing the A−N6 to G−O4′ distance to be out of the hydrogen bonding distance range; (2) some torsional angles associated with the phosphate group are different, which is not surprising due to the difficulty in the determination of those angles from proton NMR data.

When the conformation of RNA tandem base pairs, exemplified by that found in the hammerhead ribozyme (Figure 4C), is compared with the DNA conformation (Figure 4A), several differences are found. The sheared G·A base pair in RNA is less propeller-twisted. Moreover, the *intra*-base-pair hydrogen-bonding scheme in RNA is more variable. In Figure 4C, A8 is base paired to G12 with three hydrogen bonds including one from AN6 to GO2′. G12 has an additional hydrogen bond between its N2 to the O2′ of G9, and the latter is in turn hydrogen bonded to the phosphate of A8. Interestingly, the lower G9−A13 base pair has three hydrogen bonds such as those found in DNA. It has been suggested that such variability in the hydrogen-bonding patterns in RNA is sequence-context dependent (*16*).

Interactions of Ligands with Tandem G−A Motif. The availability of three independent crystal forms of the d(CCGAATGAGG) decamer at atomic or near-atomic resolution provides an excellent opportunity to inspect the interactions of water molecules, ions, and other molecules with the tandem G−A base pairs. In all three crystal forms, one end of the DNA duplex is stacked on the other end of a 2_1-axis symmetry-related duplex, forming an infinite pseudo DNA duplex (along the crystallographic *a*-axis in the C2 form and along the crystallographic *b*-axis in the two $C222_1$ forms). The crystal-packing interactions are shown in Figures 1S, 2S, and 3S (Supporting Information) for the *C*2, $C222_1$ (long form), and $C222_1$ (short form) structures, respectively. The adjacent columns of the duplexes in both $C222_1$ forms have left-handed crossover angles (∼32°) (Figures 2S and 3S) similar to that found in some other DNA crystals (*41*).

Hydration. As noted above, the sheared G−A base pair conformation leaves the O6−N1−N2 edge of guanine base completely accessible for hydrogen-bonding interactions. It is not surprising that all three sites of this edge are fully hydrated (sites 1, 2, and 3 in Figure 5), unless it is involved in other interactions with nonwater ligands. In fact, this is true for all base pairs of the decamer in all three crystal forms. Site 1's water molecules bridge the G−N2 with the phosphate *pro*-R_p oxygen of the base-paired adenine nucle-

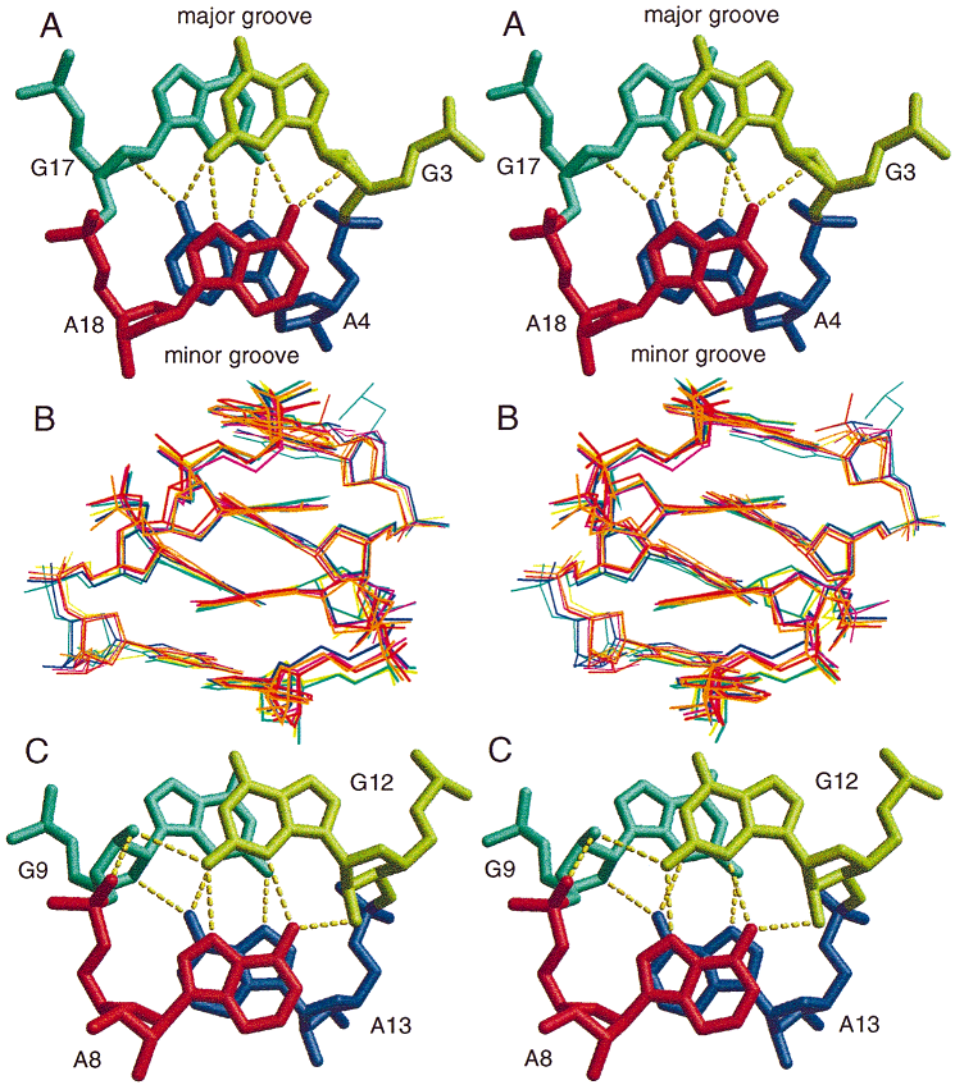

FIGURE 4: (A) The detailed conformation of the sheared tandem G−A base pairs in DNA using the G3pA4:G17pA18 step of the C2 structure as an example. There is a large propeller twist between the G and A bases within a G−A base pair. Each A base forms three hydrogen bonds (AN7-GN2, AN6-GN3, and AN6-GO4′) with the paired G nucleoside. The tandem G−A step is further stabilized by the *inter*strand G-over-G and A-over-A stacking interactions. (B) Superposition of six independent 5′-(Py)GA(Pu) tetranucleotide steps from three crystal structures, with a mean rmsd. of 0.62 Å. The large propeller twist of the G−A base pair is evident. (C) The detailed conformation of the sheared tandem G−A base pairs in RNA using the G12pA13:G9pA8 step of the hammerhead ribozyme structure (PDB accession no. 1hmh) as an example.

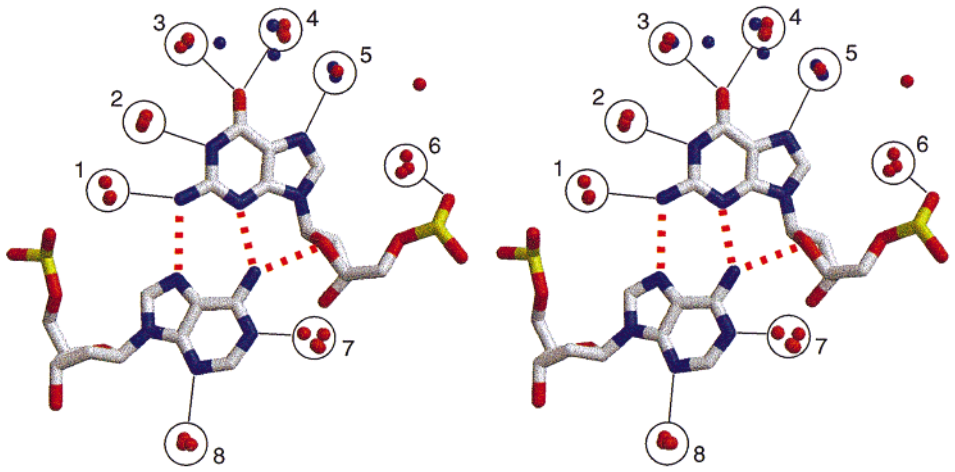

FIGURE 5: The first shell water structure surrounding the sheared tandem G−A base pairs in DNA. The water molecules (red spheres) around the four independent G−A base pairs of the C2 structure are superimposed. Eight conserved water binding sites are identified. At sites 3, 4, 5, and 6, water molecules are replaced by the NH_3 molecules (blue spheres) from the bound $Co(NH_3)_6^{3+}$ ion.

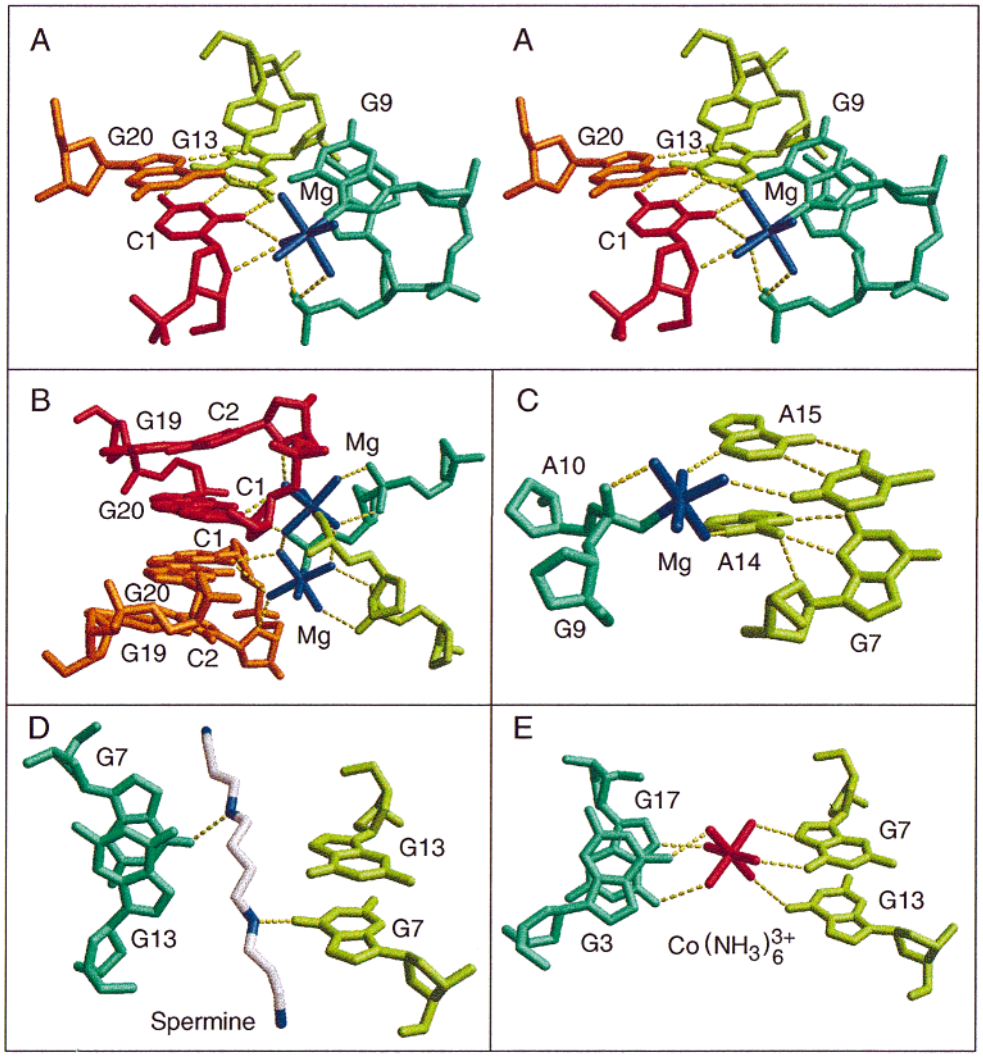

FIGURE 6: Interactions of various ligands/ions with G−A motifs. Nucleotides from different symmetry-related duplexes are colored differently. (A) The tertiary interactions between the terminal C1 and G20 bases and a neighboring helix in the short $C222_1$ form. C1 is base paired with the G13 base (of the G13−A8 base pair) using a Watson−Crick conformation. A pentahydrated Mg^{2+} is coordinated at the N7 site of G8. The hydration sphere of Mg^{2+} ion is extensively hydrogen bonded to DNA. Note that such a coordinated Mg^{2+} ion in RNA would be in a position to interact with the O2′ hydroxyl and the *pro*-R_p oxygen of the 3′-phosphate of a cytosine residue. This arrangement may be relevant in understanding the hammerhead ribozyme cleavage activity. (B) A two Mg^{2+} cluster located among several phosphate groups. (C) A pentahydrated Mg^{2+} ion interacts with a sheared G−A base pair in the minor groove. The Mg^{2+} ion is directly coordinated to the *pro*-S_p oxygen of P10. (D) Interactions of a spermine between two helices using the major groove side of the guanine bases. (E) Interactions of a $Co(NH_3)_6^{3+}$ ion between two helices using the major groove side of the guanine bases.

otide, which may contribute to the stability of the sheared G−A base pair. As discussed below, these three waters may be displaced by a cytosine base so that a Watson−Crick G−C base pair is formed. Sites 4, 5, and 6 are located on the major groove side of the guanine and they are often displaced by metal complexes ($Co(NH_3)_6^{3+}$) or spermine. Site 6 water forms a hydrogen bond with the phosphate, but it does not form hydrogen bonds to the guanine base; instead, it is associated with the second shell waters to form a cage surrounding the hydrophobic G−H8 hydrogen.

On the adenine side, both the N1 (site 7) and N3 (site 8) positions are bound with conserved water molecules. They interact with other first and second shell waters to form a caged network in the minor groove of the duplex.

Tertiary Base Interactions. In the short $C222_1$ form, the terminal G20 and C1 nucleotides are not part of the helix; instead they project away from the duplex (Figure 2B). C1 is paired with G13* from a symmetry-related duplex, forming a G−C Watson−Crick base pair. G20 is in the syn conformation, and it is hydrogen bonded to two cytosine bases (C11* and C12*) (Figure 6A). A penta-hydrated Mg^{2+} ion coordinated to the G9−N7 site was found in a nested cavity in the major groove near the A8pG9 step. The Mg^{2+} ion bound water molecules are hydrogen bonded to the neighboring P8 phosphate *pro*-R_p oxygen. Such a Mg^{2+} ion bound in a cavity near the phosphate may play a catalytic role in ribozyme (*42*). Mg^{2+} ion is essential in the activity of hammerhead ribozyme. While no Mg^{2+} ion has been unequivocally identified in the hammerhead ribozyme structures (*11*, *12*), its heavier analogues Mn^{2+}/Cd^{2+} ion have been used to probe the metal ion binding sites (*43*).

The structure of this penta-hydrated Mg^{2+} ion bound to a base and the adjacent phosphate groups seems to be highly relevant to the magnesium ion mediated ribozyme activity. On the basis of the observed tertiary base pair interactions (seen in the DNA structure of Figure 6A) and other data

describing the hammerhead ribozyme's activity and its structural requirements (*40*), we propose that *the sheared tandem G−A base pairs of the stem II in the hammerhead ribozyme are directly involved in the catalytic cleavage activity* (unpublished results). Specifically, in the hammerhead ribozyme, the C17 residue (the cleavage nucleotide) may base pair in the Watson−Crick conformation with the G12 of stem II so that the O2′ and 3′-phosphate of C17 can be arranged in an in-line conformation, mediated through the hydrated Mg^{2+} ion which is coordinated at the G10.1−N7 site. The proposed model detailing the possible mechanism of the hammerhead ribozyme will be described elsewhere.

Hydrated Magnesium Ions. There are five independent hydrated Mg^{2+} ions located in the three crystal forms (including the one described above), and all of them interact with the phosphate groups, either directly or through the bridging water molecules (Figure 6, panels B and C). They serve to partially neutralize the negative charges of phosphates. The binding of Mg^{2+} ions to the backbone of the decamer helices does not appear to affect the local conformation. When the similar regions from the three crystal forms are compared, they show nearly identical conformation, with or without a hydrated Mg^{2+} ion bound to the phosphate group. This is in contrast to what has been proposed recently in the reexamination of the B-DNA CGCGAATTCGCG dodecamer structure (*44*). Of particular interest is the dinuclear Mg^{2+} cluster, which is located among several phosphate groups (Figure 6B). This type of dinuclear Mg^{2+} center has been found frequently in RNA structures, including the group I ribozyme (*45*) and 5S rRNA (*46*).

Spermine. A spermine molecule is found to bridge two tandem GpA steps from two symmetry-related duplexes (Figure 6D). The spermine makes contacts to four guanines of the G·A base pairs. Thus, it is possible that the lysine side chain of a protein may interact with the surface of the tandem G·A base pair in a similar fashion.

Cobalt(III) Hexaammine Ions. In many DNA crystal structures, $Co(NH_3)_6^{3+}$ ions play important roles in bridging duplexes together to facilitate the crystal packing (*47*). Here is no exception. Two $Co(NH_3)_6^{3+}$ ions, one from the long $C222_1$ form and one from the C2 form, were found in the electron density (Figure 1C). In the long $C222_1$ form, the $Co(NH_3)_6^{3+}$ ion is located on a crystallographic axis and sandwiched between two adjacent duplexes (Figure 6E). One side of the ion has four hydrogen bonds, using three ammonia groups, to the N7/O6 side of two guanines (G3 and G17) in a tandem G−A base pair step. In the C2 form, the $Co(NH_3)_6^{3+}$ ion is bound to the N7/O6 site of guanine G7 and the phosphate of thymine T6. Such interactions involving the guanine N7/O6 sites are consistent with those found in solution (*48*).

CONCLUSION

In this paper, we have analyzed three high-resolution structures of a DNA decamer containing tandem G−A mismatches in the 5′-(pyrimidine)GA(purine) sequence. The sheared G−A base pairs in DNA and RNA share some similar structural features, but some characteristics (e.g., *inter*- and *intra*-base-pair hydrogen bonds, sugar puckers) are different. The observations of tertiary base pair interactions involving the surface of the guanine N1 side (associated with a sheared G−A base pair) may have implications in the biological activity of nucleic acids (e.g., ribozyme).

ACKNOWLEDGMENT

We thank W. Minor (University of Virginia) for making HKL2000 available to us before its official release; A. T. Brünger (Yale University) for providing a prerelease version of CNS.

SUPPORTING INFORMATION AVAILABLE

Six tables of the torsion angles and helical parameters of the three crystal forms and three figures of crystal packing diagrams of the DNA decamer are included. This material is available free of charge via Internet at http://pubs.acs.org.

REFERENCES

1. Chou, S.-H., Zhu. L., and Reid, B. R. (1997) *J. Mol. Biol. 267*, 1055−1067.
2. Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J., and Moyzis, R. K. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1695−1699.
3. Catasti, P., Gupta, G., Garcia, A. E., Ratliff, R., Hong, L., Yau, P., Moyzis, R. K., and Bradbury, E. M. (1994) *Biochemistry 33*, 3819−3830.
4. Chou, S.-H., Cheng, J.-W., Fedoroff, O., and Reid, B. R. (1994) *J. Mol. Biol. 241*, 467−479.
5. Zhu, L., Chou, S.-H., and Reid, B. R. (1995) *J. Mol. Biol. 254*, 623−637.
6. Suen, I.-S., Rhodes, J. N., Christy, M., McEwen, B., Gray, D. M., and Mitas, M. (1999) *Biochim. Biophys. Acta 1444*, 14−24.
7. Ortiz-Lombardia, M., Cortes, A., Huertas, D., Eritja, R., and Azorin, F. (1998) *J. Mol. Biol. 227*, 757−762.
8. Traub, W., and Sussman, J. L. (1982) *Nucleic Acids Res. 10*, 2701−2708.
9. Gautheret, D., Konings, D., and Gutell, R. R. (1994) *J. Mol. Biol. 242*, 1−8.
10. Rife, J. P., Stallings, S. C., Correll, C. C., Dallas, A., Steitz, T. A., and Moore, P. B. (1999) *Biophys. J. 76*, 65−75.
11. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) *Nature 372*, 68−74.
12. Scott, W. G., Finch, J. T., and Klug, A. (1995) *Cell 81*, 991−1002.
13. Sakamoto, T., Kawai, G., Katahira, M., Kim, M. H., Tanaka, Y., Kurihara, Y., Kohno, T., Watanabe, S., Yokoyama, S., Watanabe, K., and Uesugi, S. (1997) *J. Biochem. 122*, 556−563.
14. Walter, A. E., Wu, M., and Turner, D. H. (1994) *Biochemistry 33*, 11349−11354.
15. SantaLucia, J. J., and Turner, D. (1994) *Biochemisty 32*, 12612−12623.
16. Heus, H. A., Wijmenga, S. S., Hoppe, H., and Hilbers, C. W. (1997) *J. Mol. Biol. 271*, 147−158.
17. Baeyens, K. J., De Bondt, H. L., Pardi, A., and Holbrook, S. R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12851−12855.
18. Li, Y., and Agarwal, S. (1995) *Biochemistry 34*, 10056−10062.
19. Chou, S.-H., Cheng, J.-W., and Reid, B. R. (1992) *J. Mol. Biol. 228*, 138−155.
20. Greene, K. L., Jones, R. L., Li, Y., Robinson, H., Wang, A. H.-J., and Wilson, W. D. (1994) *Biochemistry 33*, 1053−1062.
21. Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., and Dickerson, R. E. (1987) *Science 238*, 498−504.
22. Wang, A. H.-J., and Gao, Y.-G. (1990) *Methods 1*, 91−99.
23. Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I., and Joachimiak, A. (1999) *Acta Crystallogr., Sect. D 55* 1726−1732.
24. Evans, G. CHOOCH Download Page. http://lagrange.mrc-lmb.cam.ac.uk/doc/gwyndaf/chooch.html.

25. Westbrook, E. M., and Naday, I., (1997) *Methods Enzymol. 276*, 44−68.
26. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
27. 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., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
28. Brünger, A. T. (1992) *X-PLOR 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, Connecticut.
29. Sheldrick, G. M. (1997) SHELX-97, crystallographic refinement program. University of Gottingen, Germany.
30. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T., and Berman, H. M. (1996) *Acta Crystallogr. 52*, 57−64.
31. Parkin, S., Moezzi, P., and Hope, H. (1995) *J. Appl. Crystallogr. 28*, 53−56.
32. Moews, P. C., and Kretsinger, R. H. (1975) *J. Mol. Biol. 91*, 201−228.
33. Nunn, C. M., and Neidle, S. (1996) *J. Mol. Biol. 256*, 340−351.
34. Gao, Y., Robinson, H., and Wang, A. H.-J. (1999) *Eur. J. Biochem. 261*, 413−420.
35. Robinson, H., and Wang, A. H.-J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5224−5228.
36. Lane, A., Ebel, S., and Brown, T. (1994) *Eur. J. Biochem. 220*, 717−727.
37. Ebel, S., Brown, T., and Lane, A. (1994) *Eur. J. Biochem. 220*, 703−715.
38. Ke, S.-H., and Wartell, R. M. (1996) *Nucleic Acids Res. 24*, 707−712.
39. Sutton, D. H., Conn, G. L., Brown, T., and Lane, A. N. (1997) *Biochem. J. 321*, 481−486.
40. Shepard, W., Cruse, W. B. T., Fourme, R., de la Fortelle, E., and Prange, T. (1998) *Structure 6***,** 849−861.
41. Timsit, Y., Shatzky-Schwartz, M., and Shakked, Z. (1999) *J. Biomol. Struct. Dyn. 16*, 775−785.
42. Birikh, K., Heaton, P. A., and Eckstein, F. (1997) *Eur. J. Biochem. 245*, 1−16.
43. Feig, A. L., Scott, W. G., and Uhlenbeck, O. C. (1998) *Science 279*, 81−84.
44. Tereshko, V., Minasove, G., and Egli, M. (1999) *J. Am. Chem. Soc. 121*, 3590−3595.
45. Cate, J. H., Hanna, R. L., and Doudna, J. A. (1997) *Nat. Struct. Biol. 4*, 553−558.
46. Correll, C., and Steitz, T. (1997) *Cell 91***,** 705−712.
47. Gao, Y., Robinson, H., van Boom, J. H., and Wang, A. H.-J. (1995) *Biophys. J. 69*, 559−568.
48. Robinson, H., and Wang, A. H.-J. (1996) *Nucleic Acids Res. 24*, 676−682.
49. Pan, B., Mitra, S. N., and Sundaralingam, M. (1999) *Biochemistry 38*, 2826−2831.
50. Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.

BI9914614