Crystal and Molecular Structure of d(GTGCGCAC): Investigation of the Effects of Base Sequence on the Conformation of Octamer Duplexes†

Craig Bingman,‡,§,⊥ X. Li,‡ Gerald Zon,‖ and Muttaiya Sundaralingam*,‡,§

Crystallography Laboratory, Department of Biochemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, Laboratory of Biological Macromolecular Structure, Department of Chemistry and Ohio State Biotechnology Center, The Ohio State University, Rightmire Hall, 1060 Carmack Road, Columbus, Ohio 43210-1002, and Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404

Received June 10, 1992; Revised Manuscript Received October 13, 1992

ABSTRACT: The structure of the self-complementary deoxyoctanucleotide d(GTGCGCAC), which crystallized as an A-type helix in the space group $P4_32_12$, with one strand in the crystallographic asymmetric unit has been determined and refined to a final R-value of 0.154 using 1.64-Å diffraction data collected on an area detector. In contrast to the closely related sequence $d(GTGTACAC)_{tet}$, there was no evidence for an ordered spermine molecule in the major groove of this octamer. Ordered water is found associated with almost all the exposed hydrogen bonding groups of the octamer. A pentagonal ring of water molecules is hydrogen bonded to O6 and N7 of G3 and the N4 and O6 of the C4-G13 base pair. A detailed comparison of the local helical parameters of d(GTGCGCAC) and $d(GTGTACAC)_{tet}$ is presented. The base sequence change at the center of the octamers affects several of the local helical parameters, via both intra- and interduplex interactions within the crystal.

Structural studies on d(GTGCGCAC) were initiated for several reasons. This sequence is closely related to d(GTGTACAC), which crystallized in two space groups, in $P4_32_12$ (Jain et al., 1987), in which a bound spermine molecule was found in the major groove of the octamer and cross-strand hydrogen bonds were observed between adjacent purines (Jain et al., 1989), and in the hexagonal space group $P6_122$, where no spermine could be identified in the major groove, and in which the deoxyribose residue at the seventh position occurred in the C2′-endo conformation, the first such occurrence observed within an A-DNA oligomer (Jain & Sundaralingam, 1989). Accordingly, the sequence d(GTGCGCAC) was synthesized, in which the inner TA base-pairs were replaced with CG base-pairs while the purine–pyrimidine alternation was maintained in the molecule. We were interested to see if this oligomer would exhibit similar polymorphism, and if it crystallized in the tetragonal space group, whether a spermine molecule would still be found in the major groove. Also, this oligonucleotide served as an additional test of the observation that in solution and in crystals small oligomers with an alternating pyr–pur sequence initiating with a 5′-purine were resistant to conversion to left-handed forms (Quadrifoglio, 1984; Jain et al., 1987).

We have only obtained tetragonal crystals of d(GTGCGCAC). Like crystals of d(GTGTACAC), these crystals also diffract to high resolution. To obtain a model with accurate positional and thermal parameters as well as accurate positions for the ordered solvent surrounding the DNA within the crystal lattice, high-resolution data were collected on an area detector.

MATERIALS AND METHODS

Synthesis and Crystallization. The oligonucleotide was synthesized by the solid support phosphoramidite method and purified by reverse-phase chromatography (Zon & Thompson, 1986). The purified oligonucleotide was crystallized by the vapor diffusion technique using conditions identical to those used to crystallize the tetragonal form of d(GTGTACAC) (Jain, 1990). A droplet consisting of 2 mM octamer duplex as the sodium salt, 3 mM $MgCl_2$, 2 mM spermine, and 10 mM sodium cacodylate, pH 6.5, was equilibrated against a reservoir solution of 30‰ 2-methyl-2,4-pentanediol (MPD). Tetragonal crystals formed over a period of a few weeks, with space group $P4_32_12$ and cell constants $a = b = 42.22$ Å, $c = 25.07$ Å, very similar to tetragonal d(GTGTACAC) ($a = b = 42.43$ Å, $c = 24.75$ Å).

Data Collection. Data on d(GTGCGCAC) were collected at the Argonne National Laboratory on a Siemens/Nicolet multiwire area detector using a crystal cooled to −10 °C with an FTS cooling device. The crystal to detector distance was 10 cm, and the data were collected at a swing angle of 30°, 100 s per frame, using CuKα radiation from an Elliot GX20 rotating anode operating at 40 kV, 65 mA. The data were collected in three scans: first a 130° ϕ scan was performed followed by two ω scans covering 45° and 60° angular range at χ 90° and 45°. The frames were reduced using Xengen 1.3 (Howard et al., 1987) to yield a total of 11 665 reflections out of 12 578 predicted. Each unique reflection was observed on average 4.2 times, giving an R_{symm} of 3.5% on intensity and 1.9% on F. A summary of the statistics of the data set is given in Table I.

Refinement. Since d(GTGCGCAC) crystallized isomorphously with the tetragonal d(GTGTACAC), the latter was used as a starting model for refinement. All the water molecules were removed from the atomic model of d(GTGTACAC) and thermal parameters were reset to a uniform 25 Å^2. The correlation coefficient and R-factor for the data between 5 and 1.64 Å from the present structure were 0.914 and 0.302, respectively. The TA to CG sequence changes in the inner two base-pairs were made with FRODO on an Evans and Sutherland PS340 (Jones et al., 1985). The resultant model of d(GTGCGCAC) was then refined with NUCLSQ (Hendrickson & Konnert, 1981; Westhof et al., 1985). Since

† This work was supported by NIH Grant GM-17378.
* Author to whom correspondence should be addressed.
‡ University of Wisconsin—Madison.
§ The Ohio State University.
‖ Applied Biosystems.
⊥ Present address: Department of Biochemistry, Columbia University, 650 W. 168th St., New York, NY 10032.

0006-2960/92/0431-12803$03.00/0

Table I: Data Collection Statistics

resolution (Å)	no. of refl (% coverage)	$\langle I \rangle^a$	$\langle I/\sigma(I) \rangle$	R-sym
99–2.99	544 (98.4)	3878	154	0.023
2.99–2.37	509 (100.0)	666	54	0.051
2.37–2.07	500 (100.0)	253	25	0.085
2.07–1.88	474 (98.2)	146	15	0.114
1.88–1.75	466 (94.9)	90	8	0.165
1.75–1.64	274 (56.4)	43	4	0.245
total	2767 (91.6)	975	49	0.035

[a] I = intensity.

NUCLSQ does not allow positional restraints between symmetry-related molecules, in particular between the Watson–Crick base-pairs of the two strands related by a crystallographic dyad, it was necessary to refine the duplex model in $P4_3$, rather than the crystallographic asymmetric unit of one strand in $P4_32_12$. Therefore, the data were expanded to the Laue symmetry $4/m$, and the octamer duplex was treated as the asymmetric unit for the refinement. For electron density map calculations, the positions and thermal parameters of the two strands were averaged, and one strand was used for phasing in $P4_32_12$.

The "mutated" model was initially refined against the data in the 8 to 4Å shell. Approximately 300–400 reflections were added after each round of refinement, until all the reflections between 5 and 2 Å were included. At this point, $2F_o - F_c$ and $F_o - F_c$ omit maps were calculated, omitting one residue per map. The model was adjusted to fit the electron density with FRODO. Difference density maps were used to identify solvent molecules; when peaks appeared at 3σ or above and were 2–3.5 Å from potential hydrogen bonding DNA atoms, they were included in the atomic model as ordered water. The acceptance criteria for solvent were relaxed during later stages to include peaks appearing at 2σ and peaks within 2–3.5 Å of any atom in the model, either DNA or preexisting solvents. After the first map fitting at 2-Å resolution, thermal parameter refinement was initiated, and the data were gradually increased to include all 4483 reflections between 8- and 1.64-Å resolution with $F > 2\sigma(F)$ ($P4_3$). After several rounds of refinement, refitting, and inclusion of a total of 81 ordered waters per duplex, the R-factor dropped to the current value of 0.154. No bound spermine was identified during the course of refinement.

The present model possesses good stereochemistry, as exemplified by the rms deviation from ideal sugar base bond distances of 0.011 Å, and 1.4° in bond angles. A complete summary of the restrained parameters and their deviations from ideality is provided in Table II. There are 2.78 observations per refined parameter in the atomic model. Electron density for two especially important residues (Figure 1) shows holes in the density of the six membered rings, characteristic of refined models at this resolution. These residues were chosen because they depart from the usual A-DNA conformation. These departures will be discussed later in this paper. The electron density for all other nucleotides is as good as or better than these examples. The coordinates for the DNA and ordered solvent and structure factor data have been deposited in the Brookhaven Data Bank.

Table II: Refinement Summary for d(GTGCGCAC)

number of reflections	4614	($P4_3$)
R_{merge}	0.035	
final R value	0.17	
final correlation coefficient	0.97	

restraint	no.	RMS dev	σ	no. dev $> 2\sigma$
distance restraints				5
bonds	304	0.005 Å	0.025 Å	
angles	448	0.017 Å	0.050 Å	
PO_4bonds	56	0.023 Å	0.050 Å	
PO_4angles, h-bonds	344	0.036 Å	0.075 Å	
planarity restraints	168	0.015 Å	0.030 Å	0
chiral center restraints	48	0.053 Å^3	0.100 Å	0
nonbonded contacts				
single torsion	0	NA	0.090 Å	0
multiple torsion	57	0.142 Å	0.090 Å	28
thermal parameter restraints				
bonds	304	2.7 Å^2	5.0 Å^2	0
angles	448	3.8 Å^2	7.5 Å^2	0
PO_4bonds	56	4.4 Å^2	5.0 Å^2	0
PO_4bonds, h-bonds	344	5.0 Å^2	7.5 Å^2	0
noncrystallographic symmetry				
DNA atoms				
position	161	0.007 Å	0.200 Å	0
thermal param	161	0.14 Å^2	5.0 Å^2	0
solvent atoms				
position	41	0.032 Å	0.200 Å	0
thermal param	41	0.45 Å^2	5.0 Å^2	0

RESULTS AND DISCUSSION

Crystal Packing. The octamer belongs to the class of A-DNA octamers crystallizing in the tetragonal space group $P4_32_12$. The packing motif of this class of A-DNA crystals has been described in detail elsewhere (Wang et al., 1982; Conner et al., 1982; Jain & Sundaralingam, 1989), and only the pertinent features will be presented here. As with all A-DNA crystals, the most prominent packing mode involves the abutment of the terminal base-pairs of one molecule against the deoxyribose groups on the minor grooves of symmetry-related molecules (Shakked et al., 1981; Wang et al., 1982). This is the only type of direct intermolecular interaction in these crystals.

A stereoview down the c direction of the crystal is shown in Figure 2. The DNA oligomers are clustered along the 4_3 axis. There are large, oval-shaped solvent channels measuring approximately 10 by 20 Å running along the vertical 2_1 axes, parallel to the 4_3 axis. In contrast to the channels found in some other A-DNA structures (Doucet et al., 1989; Bingman et al., 1992), the channels in these octamer crystals are too narrow to accept a disordered B-DNA oligomer, and no diffuse meridional fiberlike diffraction pattern was observed. The electron density maps become essentially featureless beyond one or two layers of water molecules from the DNA, indicating that the solvent in this region is quite mobile.

Distortions in the DNA Duplex. It has been previously noted that octamers crystallizing in the tetragonal space group $P4_32_12$ are bent by approximately 15° (Jain et al., 1989). A 15° discrepancy was also found for d(GTGCGCAC) between the helical axes computed for the two half-molecules. The overall and half-molecule axis (Figure 3) pass within 0.33 Å of each other, and the two half-molecule axes nearly intersect, passing within 0.01 Å of each other. Inspection of the axes and marker atoms on the graphics confirmed that the two half-molecules are rotated away from the major groove side of the molecule.

The observed axial deflection in the tetragonal octamers gives rise to an extremely wide major groove, P2–P10 8.51 Å. It should be noted that this P–P distance in one step short of the closest approach of cross-strand phosphates in fiber A-DNA (Chandrasekaran et al., 1989) and in the A-DNA dodecamer (Bingman et al., 1992). The width of the minor groove is fairly constant in the octamer (P4–P16 9.73 Å, P5–

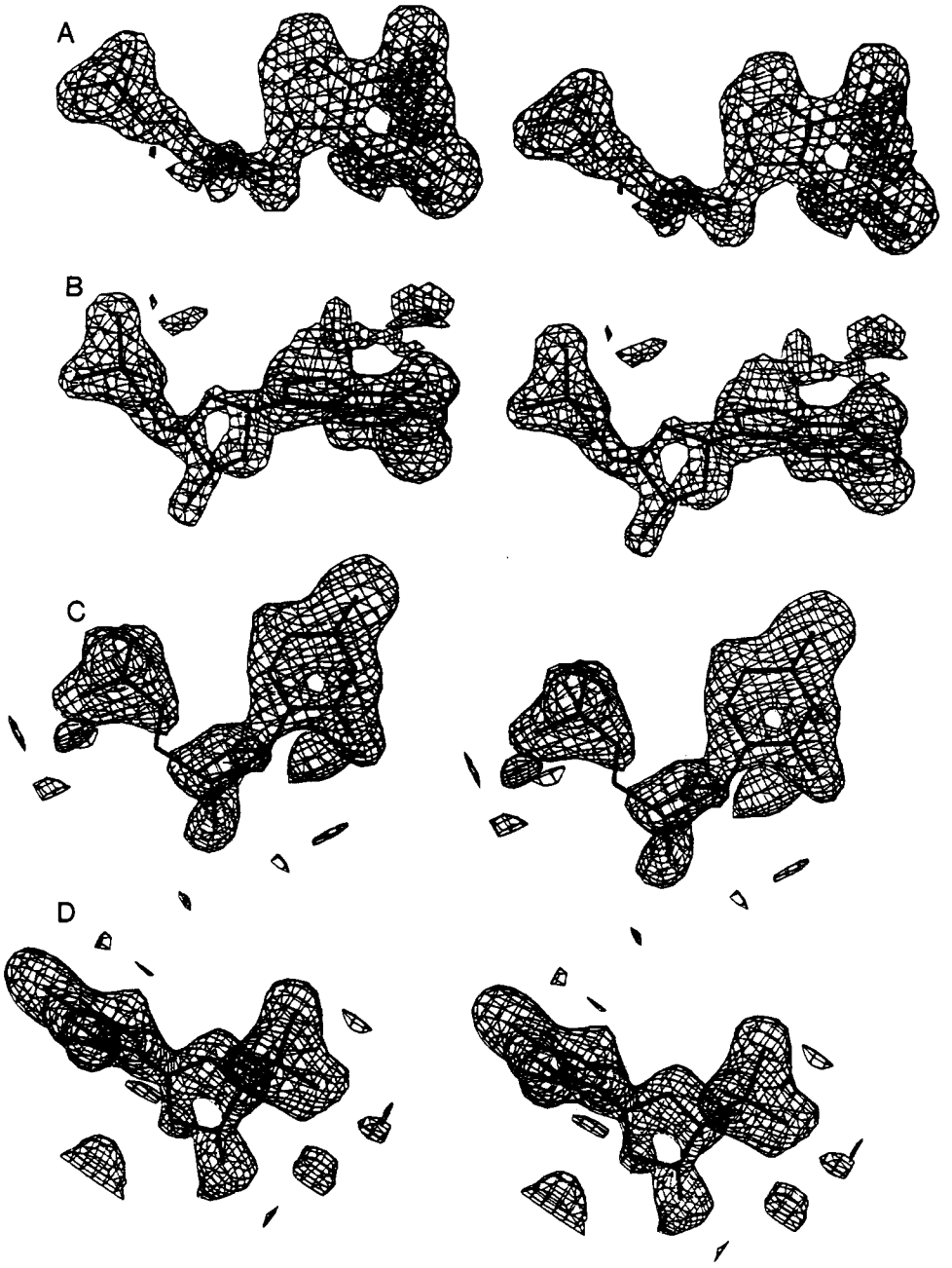

FIGURE 1: Electron Density Maps of d(GTGCGCAC). (A, B) "Omit" $F_o - F_c$ map of residue 5, showing the extended "all trans" backbone conformation. This map is contoured at 2σ. The backbone clearly falls within the extended tube of density at this position. Also noteworthy are the "holes" in the density of the six-membered ring of the purine and the five-membered deoxyribose ring. (C, D) "Omit" $F_o - F_c$ map of residue 8, also contoured at 2σ. The deoxyribose ring adopts the C2′-endo conformation, which is unusual for A-DNA. Again, the pyrimidine shows a hole in the center of the six-membered ring, although this density does seem somewhat more smeared than other pyrimidine electron density in this structure. There are typically three "handles" on the electron density of the deoxyribose ring that are especially important in fitting the electron density maps: the base density, the phosphate density, and the density for the O3′ oxygen. The relative positions for these three groups extending from the deoxyribose ring are quite different for C2′-endo and C3′-endo sugars. Even though the electron density is slightly washed out at the C3′ and C2′ positions, it is clear that the C2′-endo conformation fits both density and matches the surrounding "handles" well and probably represents the conformation of the majority of sugars at this position, averaged over time and space in the crystal. Evidence for a smaller population of C3′-endo sugars at this position may be taken from the higher than average thermal parameters for the deoxyribose group and attached base.

P15 9.70 Å, P6–P14 9.79 Å, P7–P13 9.70 Å, and P8–P12 9.73 Å).

When the descriptive helical parameters were calculated for the overall and half-molecules, the standard deviation in the helical rise per residue for the overall axis was 0.38 Å/bp, while the half-molecule axis gave a larger standard deviation of 0.45 Å/bp. This is in contrast to the case of d(CCG-TACGTACGG) (Bingman et al., 1992), where the standard deviation in helical rise decreased by a factor of 2 when calculated with respect to the half-molecule axes. This, along with inspection of the plots for both parameter sets, indicated that the half-molecule analysis was not especially useful in this case. Consequently, only the helical parameters referred to an overall helical axis are presented here.

Base Stacking and Backbone Torsion Angles. Since the sequence of the octamer is composed entirely of alternating purine–pyrimidine (R–Y and Y–R) steps, there is primarily intrastrand stacking at the R-Y steps and cross-strand purine–purine stacking at the Y–R steps. Stereo figures of the individual dinucleotide steps are provided in Figure 4. The

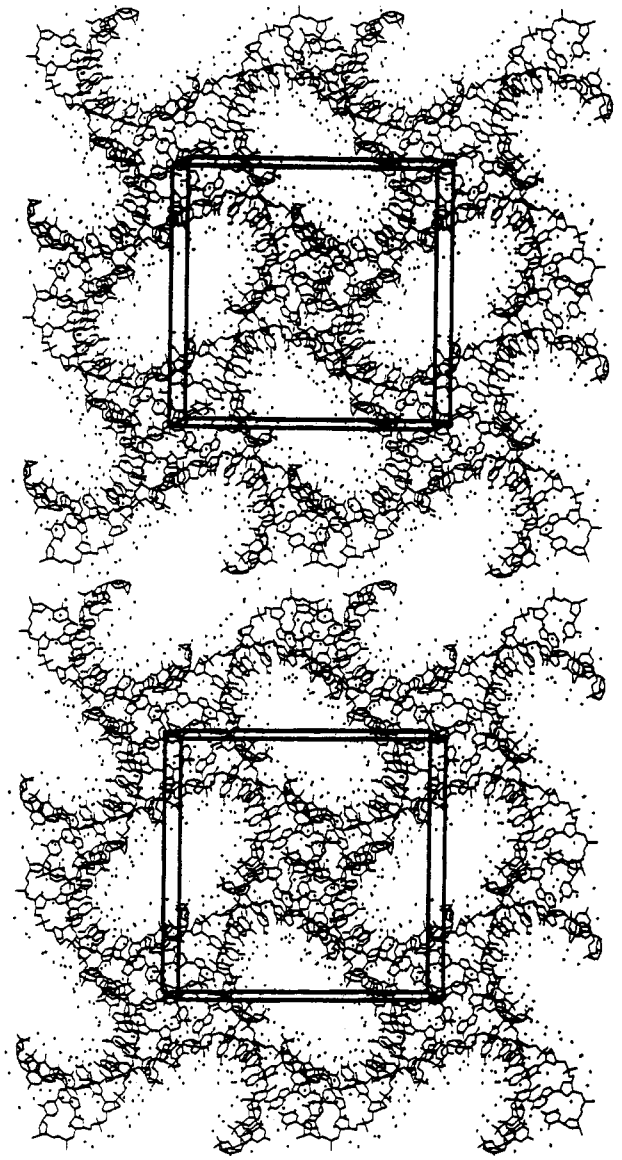

FIGURE 2: Crystal packing of d(GTGCGCAC). The view is down the *c* axis of the crystal. The DNA is rendered in skeletal form, and solvent molecules are shown as crosses. The DNA duplexes are clustered around the vertical 4_3 axes. The elliptical solvent channels in the crystal run along the vertical 2_1 axes and are approximately 10 by 20 Å. The only direct interaction between neighboring DNA duplexes in the crystal involves the abutment of the terminal base-pairs of one molecule against the sugar phosphate backbone of symmetry related molecules.

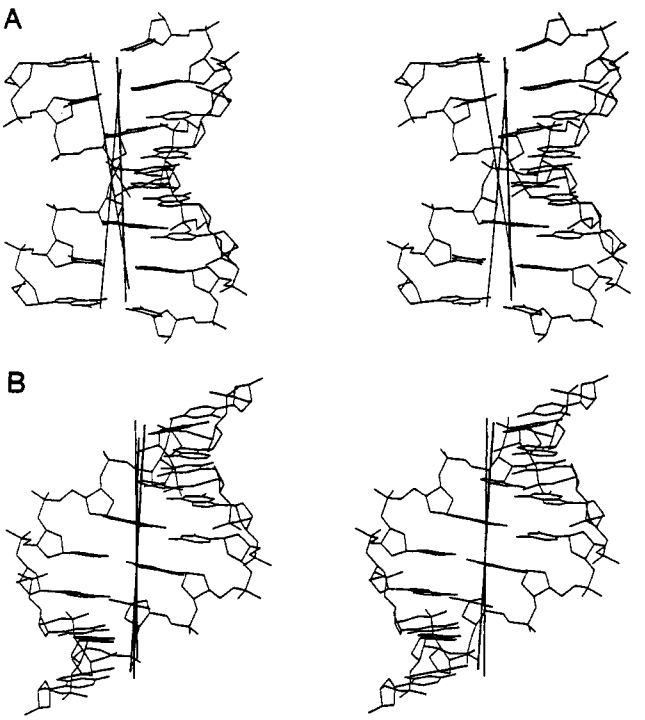

FIGURE 3: Global and local helical axes of d(GTGCGCAC). (A) The view is with the molecular dyad in the plane of the page. The global, upper, and lower half molecule axes are shown. The angle between upper and lower half axes is approximately 15°. (B) View into the minor groove. Most of the bend is directed in the plane defined by the molecular dyad and the overall helical axis.

stacking pattern of the central step is of special significance. The backbone torsion angles for the octamer are given in Table III and show that the α,γ torsion angles of G5 deviate from the typical A-DNA range of α gauche$^-$, γ gauche$^+$ to α trans, γ trans. Several authors have inferred that this altered backbone conformation, which increases the intrastrand P–P separation to nearly 7 Å, enhances the degree of interstrand purine–purine stacking at this step, although it has been observed that the degree of cross-strand purine overlap in the hexagonal crystal form of d(GTGTACAC), which did not have the so-called "all-trans" backbone conformation, was comparable to the degree of overlap seen in the tetragonal form of d(GTGTACAC) (Jain & Sundaralingam, 1989; Jain et al., 1990). In either case, this unusual backbone conformation is found at the fifth position of all tetragonal octamers and seems to be a prerequisite for A-DNA duplexes existing in this crystal lattice. At the other R–Y step in this structure, T2–G3, there is also considerable purine–purine overlap. The overlap does seem somewhat less than at the central step, where the extended backbone conformation has slid the cross-strand purines closer together and altered their rotation slightly to provide more overlap on the major groove side of the bases. It also seems that there is a slight increase in intrastrand stacking at the center of d(GTGCGCAC). The extended backbone conformation has caused a smaller relative translation toward the minor groove of the purine from the adjacent pyrimidine, which leads to better intrastrand stacking. The other notable feature of the backbone torsion angles in this tetragonal octamer is the change in δ of the eighth residue, which is related to the change in sugar pucker from C3′-endo, characteristic of A-DNA, to C2′-endo. Unusual sugar puckers have previously been noted at the ends of A-DNA structures [e.g., Kennard et al. (1985) and Heinemann et al. (1987)].

COMPARISON WITH d(GTGTACAC)

Conserved Parameters. The descriptive helical parameters of d(GTGCGCAC) and d(GTGTACAC) are compared in Figure 5 and Table IV. The intrastrand phosphorus–phosphorus distances vary between 5.7 and 6.9 Å. The all-trans backbone conformation at G5 dramatically increases the P–P separation between the fifth and sixth phosphate groups, to slightly more than 6.8 Å in both cases. There seems to be some sequence-related variation in this parameter, since in both structures, the P–P separation of the R–Y steps is high (average 6.7 Å) and low (average 5.9 Å) in the Y–R steps. The all-trans backbone conformation at the fifth residue marks the extreme upper range in this alternation, but the trend is also apparent in the other R–Y steps in the octamer as well, even though they have backbone torsion angles within the preferred range for A-DNA. The G3–G4 step is particularly noteworthy. The P–P separations are fairly well-determined,

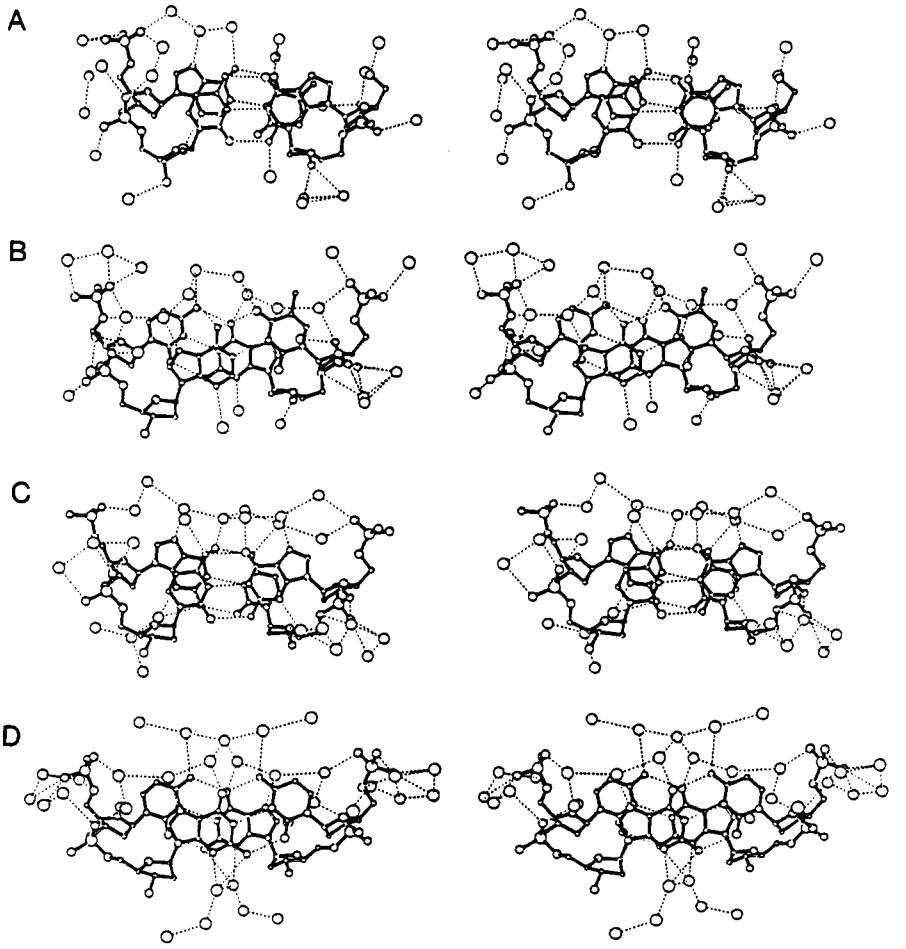

FIGURE 4: Dinucleotide stacking diagrams and dinucleotide hydration. Stacking: In all cases, the "top" base-pair (containing the lower numbered residue in strand 1) is rendered in filled bonds, and the bottom base-pair is in open bonds. The base-pair steps are oriented so that the normal to the least-squares plane of the upper base-pair projects vertically out of the plane of the page. (A) Upper base-pair G1·C16, lower base-pair T2·A15. Most of the stacking interactions of this step are between the upper and lower bases of the same strand. (B) Upper base-pair T2·A15, lower base-pair G3·C14. The dominant stacking interactions at this Y–R step are between cross-strand purines. (C) Upper base-pair G3·C14, lower base-pair C4·G13. (D) Upper base-pair C4·G13, lower base-pair G5·C12. Note in particular the "all-trans" extended backbone conformation of G5 and G13. This figure should be compared to the stacking pattern of the other unique R–Y step in this structure, shown in (B). Dinucleotide hydration: Water molecules are represented as spheres. Waters within 3.5 Å of each other and the DNA are joined by dashed lines. Note the long string of ordered solvent on the minor groove side of the third step and the pentagonal network on the major groove side of this base-pair.

Table III: Comparison[a] of Backbone Torsion Angles (°), χ, P, and τ_m in d(GTGCGCAC) and d(GTGTACAC)$_{tet}$

sequence	α	β	γ	δ	ϵ	ζ	χ	P	τ_m
G1			57	82	211	290	192	7	39
G1			45	82	209	295	196	19	41
T2	267	168	53	84	201	286	204	12	37
T2	283	172	57	79	202	288	201	19	42
G3	296	172	53	86	178	302	202	23	36
G3	273	172	82	77	189	301	192	19	41
C4	290	177	52	78	192	284	207	21	39
T4	293	188	43	87	192	279	210	19	35
G5	152	198	**179**	85	215	286	186	4	36
A5	141	198	**186**	84	214	290	191	−5	42
C6	299	171	45	81	201	285	202	14	40
C6	310	159	46	79	203	285	204	20	44
A7	303	169	57	80	200	295	204	21	38
A7	295	177	58	89	202	288	202	20	33
C8	298	182	54	124			232	**138**	23
C8	293	175	57	105			219	**59**	18

[a] Notations and definitions are as recommended by EMBO Workshop (1989).

by virtue of the electron-dense phosphorus atom. The present octamer provides an example of a structure determined at near atomic resolution where a remarkably long P–P distance of 6.4 Å is not associated with an all-trans backbone conformation.

The helical twist shows considerable variation in the tetragonal octamers. It has been noted that the central step of these structures is underwound by approximately 10° with respect to the flanking dinucleotide steps (Wang et al., 1982; Takusagawa, 1990), consequently increasing the cross-strand

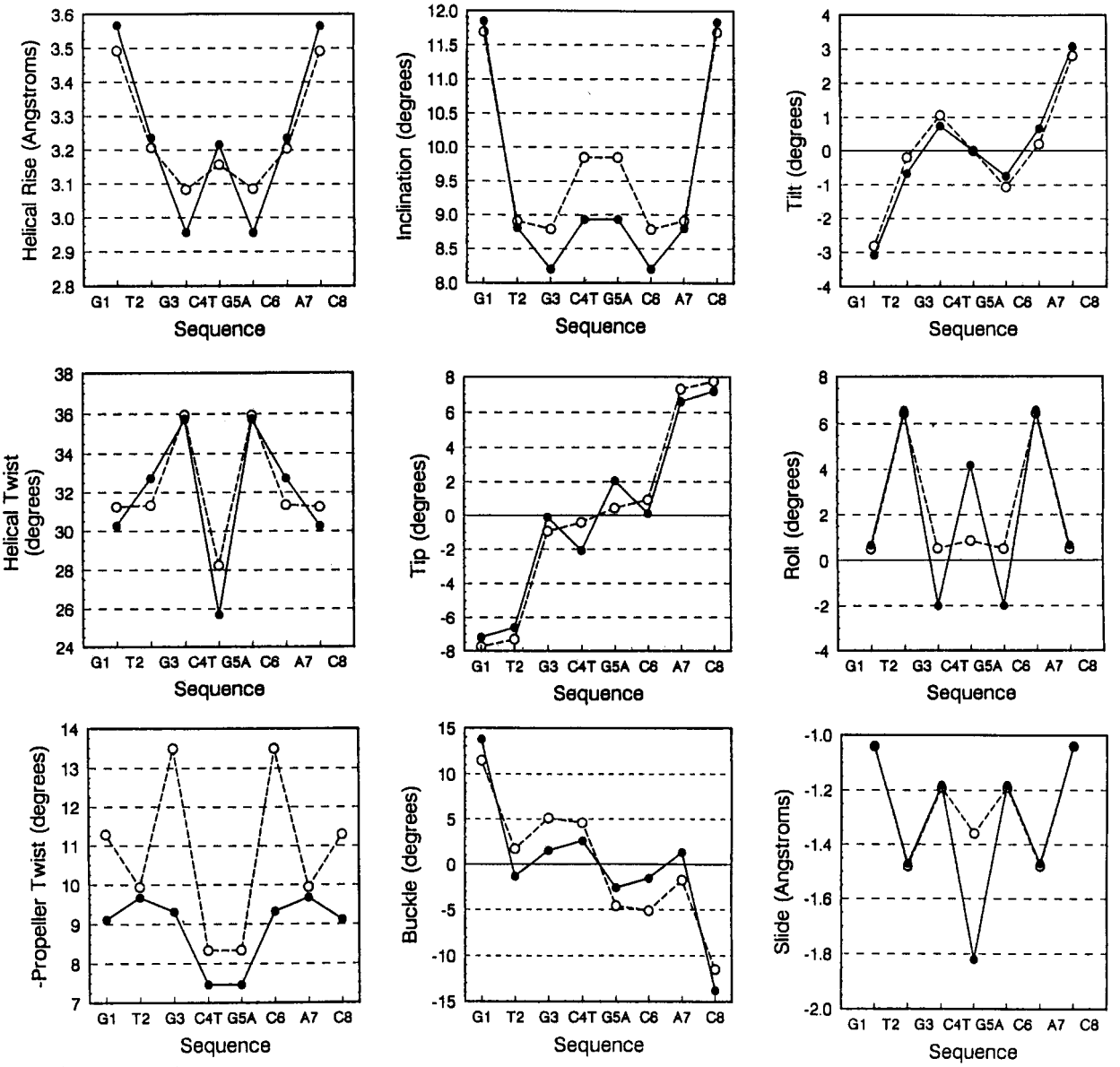

FIGURE 5: Comparison of descriptive helical parameters of d(GTGCGCAC) and d(GTGTACAC)$_{tet}$. d(GTGCGCAC) is represented by solid lines and filled circles; d(GTGTACAC)$_{tet}$ is represented by dashed lines and open circles.

purine–purine stacking at these steps. However, in the hexagonal structure of d(GTGTACAC), it was seen that the same central dinucleotide step exhibited virtually the same degree of cross-strand stacking overlap with the normal backbone conformation at A5 (Jain et al., 1989.) The helical twist of fiber diffraction A-DNA is approximately 32.7° per base-pair. In the octamer structure, the third and fifth dinucleotide steps are overwound by over 3° with respect to fiber diffraction A-DNA, the central step is underwound by approximately 6°, the second step has a helical twist similar to fiber diffraction A-DNA, and the terminal steps are slightly underwound. Conformational aberrations at the ends of DNA oligomers are quite common and are usually rationalized as the result of the termination of base stacking within the duplex and intermolecular interactions with symmetry-related molecules in the crystal. Hence the unusual backbone geometry at the central steps of the tetragonal octamers disrupts the helical twist and the crystal packing interactions at the ends reduce the unperturbed length of the octamer to less than two tetramers.

The helical rise per residue in the octamer is also generally similar to that seen in d(GTGTACAC). The rise at the ends of the molecules is quite high, over 3.5 Å/base-pair, larger than the rise observed even for B-DNA. It is not apparent why this is so. It may be a consequence of the crystal packing interactions between the terminal base-pairs and the backbone of the symmetry-related molecules, or the C2′-endo sugar of the 3′-terminal residue. The corresponding sugar in d(GTGTACAC) falls in the unusual O4′-endo conformation, intermediate between C2′-endo and C3′-endo (Altona & Sundaralingam, 1972). The electron density for the 3′-terminal residue is shown in Figure 1. The density for the cytosine base is fairly well-defined, and although the density for the deoxyribose residue displays a central hole, it is somewhat washed out around C2′ and C3′ atoms. The position of these atoms changes between C2′- and C3′-endo puckering modes, so it may be that there is some dynamic or static disorder in the conformation of the terminal sugar. The position of the O3′ atom relative to the furanose ring also changes considerably between C2′- and C3′-endo puckering modes. There is clear density for the O3′ atom, and it is on this basis that the C2′-endo pucker is assigned. Also, it is noteworthy that this sugar pucker changed during NUCLSQ refinement and was not changed during map fitting. Presumably this occurred since

Table IV: Comparison between d(GTGCGCAC) and $d(GTGTACAC)_{tet}$

(a) Descriptive Helical Parameters[a] Involving Adjacent Base Pairs

step	twist (°)	rise (Å)	roll (°)	tilt (°)	cup (°)	slide (Å)
G1|T2	30.3	3.57	0.7	−3.1	−15.1	−1.0
G1|T2	31.2	3.49	0.5	−2.9	−9.8	−1.0
T2|G3	32.7	3.24	6.6	−0.7	2.8	−1.5
T2|G3	31.3	3.21	6.5	−0.2	3.4	−1.5
G3|C4	35.7	2.96	−2.0	0.7	1.1	−1.2
G3|T4	35.9	3.09	0.6	1.0	−0.5	−1.2
C4|G5	25.7	3.22	4.2	0.0	−5.2	−1.8
T4|A5	28.2	3.16	0.9	0.0	−9.2	−1.4
G5|C6	35.7	2.96	−2.0	−0.7	1.1	−1.2
A5|C6	35.9	3.09	0.6	−1.0	−0.5	−1.2
C6|A7	32.7	3.24	6.6	0.7	2.8	−1.5
C6|A7	31.3	3.21	6.5	0.2	3.4	−1.5
A7|C8	30.3	3.57	0.7	3.1	−15.1	−1.0
A7|C8	31.2	3.49	0.5	2.9	−9.8	−1.0

(b) One-Base Pair Descriptive Helical Parameters

sequence	tip (°)	inclination (°)	prop. twist (°)	buckle (°)
G1 (G1)	−7.2 (−7.8)	11.9 (11.7)	−9.1 (−11.3)	13.8 (11.5)
T2 (T2)	−6.6 (−7.4)	8.9 (8.9)	−9.7 (−9.9)	−1.3 (1.7)
G3 (G3)	−0.1 (−1.0)	8.2 (8.7)	−9.3 (13.5)	1.5 (5.1)
C4 (T4)	−2.1 (−0.4)	8.9 (9.8)	−7.5 (−8.3)	2.6 (4.6)
G5 (A5)	2.1 (0.4)	8.9 (9.8)	−7.5 (−8.3)	−2.6 (−4.6)
C6 (C6)	0.1 (1.0)	8.2 (8.7)	−9.3 (−13.5)	−1.5 (−5.1)
A7 (A7)	6.6 (7.4)	8.8 (8.9)	−9.7 (−9.9)	1.3 (−1.7)
C8 (C8)	7.2 (7.8)	11.9 (11.7)	−9.1 (−11.3)	−13.8 (−11.5)

[a] Definitions of base pair parameters are as recommended by EMBO Workshop (1989).

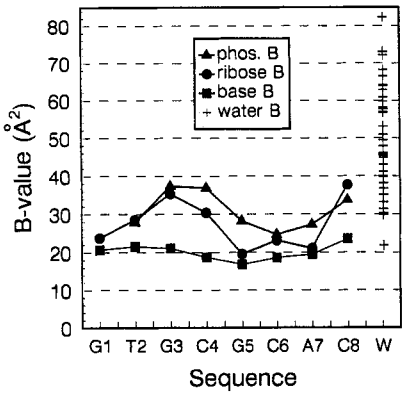

FIGURE 6: Group thermal parameters (Å²) for the residues. Symbols: phosphate groups, triangles; deoxyribose groups, circles; bases, squares. The *B* values for the ordered water molecules are shown as residue W.

P and τ_m were not explicitly restrained during the refinement. The plot of group thermal parameter for the residue (Figure 6) also shows that the thermal parameter of the 3′-terminal deoxyribose group is relatively high, comapred with the 5′-terminal residue, consistent with at least some disorder in its conformation. The thermal parameter of the attached base is also the highest in the structure. This is consistent with the observation that the glycosyl torsion angle and consequentially the orientation of the base relative to the deoxyribose ring and sugar pucker are linked, with C2′-endo sugars primarily occurring in the high χ range, and C3′-endo sugars having lower χ values. If the sugar is in dynamic or static disorder, then the orientation of the base should also be somewhat disordered. Residues 3 and 4 show high values similar to the 3′-end. The bases exhibit a more uniform value of around 20 Å². The *B* values for the ordered water molecules are shown under residue W, and they all exhibit *B* values in the range 20–80 Å².

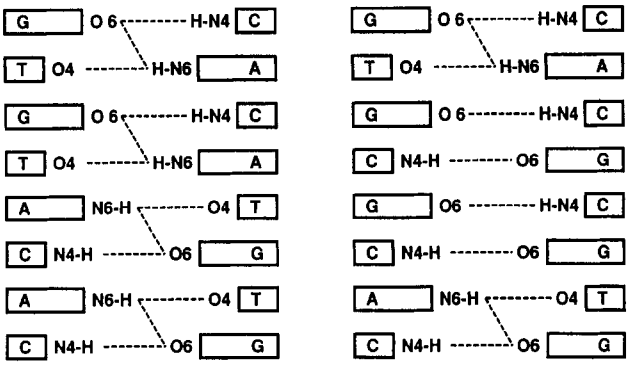

FIGURE 7: Schematic illustration of cross-strand hydrogen bonds in d(GTGCGCAC) (right) and $d(GTGTACAC)_{tet}$ (left). Bases are represented as rectangles, with hydrogen bonds rendered as dashed lines. The sequence change from T4,A5 to C4,G5 disrupts the central two purine–purine cross-strand hydrogen bonds.

The overall value for helical rise in the tetragonal octamers is much larger than expected from fiber diffraction studies. This group of molecules represents an important part of several correlations observed between helical rise per residue and several other descriptive parameters, such as displacement from the helical axis and base-pair inclination (Heinemann et al., 1987; Jain et al., 1987).

Affected Base-Pair Parameters. The structure of d(GTGTACAC) was noteworthy not only for the observation of a bound spermine in the major groove but also for the observation of interstrand hydrogen bonds between the cross-stand adenine N6 and guanine O6 at the 5′-R–Y steps in the octamer. The base change in d(GTGCGCAC) replaced the hydrogen bond donor N6 at position 5 (and the symmetry-related position 13) with an O6 atom, which eliminated the possibility of this cross-strand hydrogen bonding interaction. This is illustrated schamatically in Figure 7, which suggests that the sequence change at position 5 will potentially have a direct effect on the base conformation of residues 3, 4, 5, and 6 in that strand. The distance between G3 O6 and A13 N6 is 3.26 Å in d(GTGTACAC) (sum of group vdW radii = 3.2 Å), the corresponding distance between G3 O6 and G13 O6 increases to 3.54 Å in d(GTGCGCAC) (sum of group vdW radii = 2.8 Å), consistent with this interaction being disrupted in the latter structure. Evidence for a similar cross-strand purine–purine hydrogen bond at the ends of the molecule between G1 O6 and A15 N6 was somewhat weaker, this distance being 3.53 Å in the structure of d(GTGTACAC) and 3.49 Å for d(GTGCFCAC).

The exchange of a G for an A at position 5 is also expected to cause a substantial change in the minor groove side of the double helix, since guanine bears the additional N2 group in the minor groove. This addition not only changes the steric environment of the minor groove but also adds a potential hydrogen bond donor in this area, and the hydrogen bonding can be satisfied by contact with a either neighboring DNA molecule or solvent. In the crystal structure of d(GTGCGCAC), the N2 amino group of guanine 5 lies within 3.1 Å of the O4′ atoms of a symmetry-related G9 residue. In the structure of d(GTGTACAC), only the deoxyribose group of A5 is involved in an intermolecular interaction. The new intermolecular interaction in the present structure probably causes some of the changes in the helical parameters between the two structures.

A comparison of the propeller twist values of the two structures (Figure 5) clearly shows the importance of the cross-strand purine hydrogen bond in d(GTGTACAC). In $d(GTGTACAC)_{tet}$, G3 adopts the highest propeller twist value in

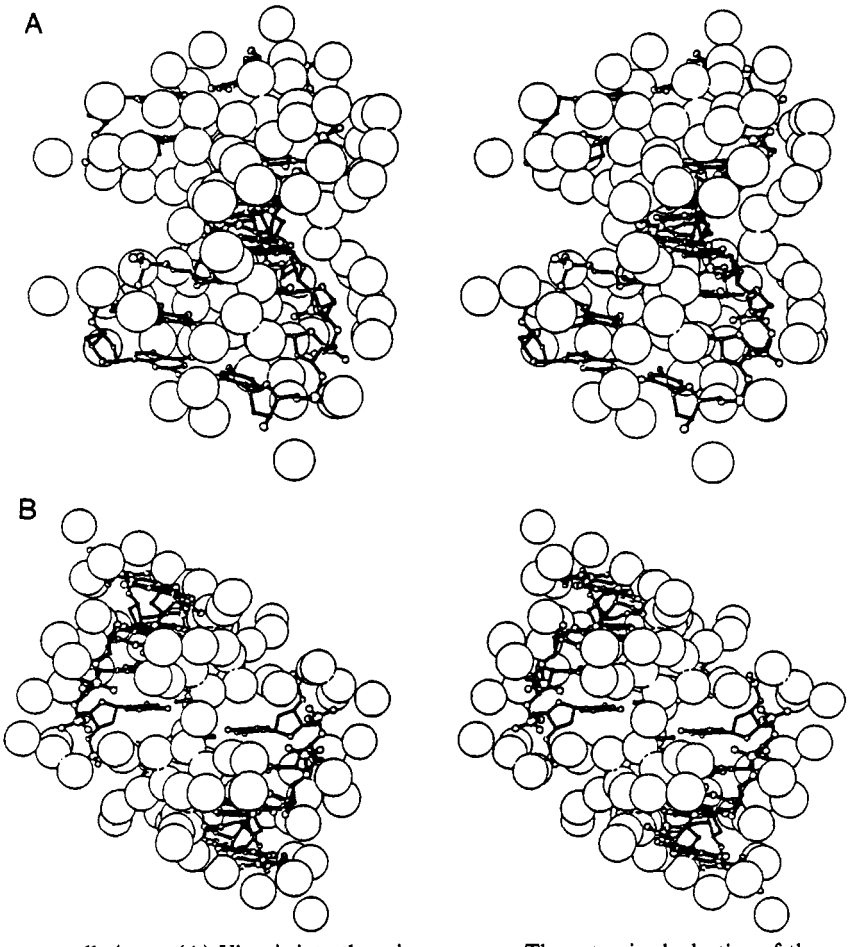

FIGURE 8: Ordered water, overall views. (A) View is into the minor groove. The extensive hydration of the central base-pair is apparent. The contact of symmetry-related molecules leaves two nonhydrated patches on the backbone. (B) View is with molecular dyad in the plane of the page.

the entire structure, over 13°. In d(GTGCGCAC), the propeller twist of the third base-pair is dramatically reduced, with a somewhat smaller impact on the propeller twist of the fourth base-pair, the purines of which also participate in this interaction in d(GTGTACAC). There is also a 2° decrease in the propeller twist of the terminal base-pair. This small change may also be an indirect consequence of hydrogen bond connecting the N2 group of G5 with the O4′ of G9.

What seemed to be a simple exchange of a CG for an TA base-pair at the center of the duplex is seen to have directly influenced the conformation of three of the four independent base-pairs in the octamer duplex at least within the crystal lattice. The propeller twist of the second base-pair is essentially constant in the two structures, consistent with the lack of direct interactions with the central base-pairs in the crystal.

The roll angles for the two molecules are quite similar, in the two crystal structures, and the primary changes in roll are at the third and central dinucleotide steps. The alternating roll angles is caused by clashes occurring between cross-strand purines, alternately in the major groove and minor groove. The roll angles of d(GTGCGCAC) are actually in better agreement with the Dickerson sum function (1983) than the roll angles of d(GTGTACAC), although given the number of changes in the packing environment of the molecule caused by this base change, it is not clear how this result should be interpreted. The roll angles of the two outer dinucleotide steps are essentially constant in both structures. The inclination angles of the inner four base-pairs are also somewhat different. The degree of similarity in the parameters for the two outer base-pairs, especially those of the second base-pair, is quite remarkable.

The remaining descriptive parameters, buckle, cup, tilt, and tip, show similar trends. The deviations in these parameters are primarily concentrated at the inner two base-pairs. Of the remaining translational parameters, the base-pair slide at the central dinucleotide deviates dramatically at only the central step of the structure.

Hydration of d(GTGCGCAC). During the course of refinement and map fitting, 81 ordered solvent peaks per duplex were identified, one of which was on the crystallographic 2-fold axis. The overall pattern of the ordered solvent surrounding the DNA is indicated in Figure 8. The major groove is essentially filled with a monolayer of ordered solvents. The polar atoms in the backbone are also highly hydrated, especially around the anionic oxygen atoms of the phosphate groups. The minor groove side of the molecule is clear of solvent in places where there are crystal packing interactions; otherwise, it is well-solvated, especially around the central dinucleotide step. Figure 4 shows the ordered solvent surrounding each unique dinucleotide step in the octamer. Water–water and water–polar DNA atoms distances less than 3.4 Å are indicated by dashed lines to represent possible hydrogen bonds. As with all A-DNA structures, the major groove hydrogen bond donors are nearly systematically hydrated. Since the characteristic packing motif of A-DNA involves the apposition of the terminal base-pairs of one molecule with the minor groove of another, it might be expected that the minor groove would be poorly solvated in these structures. This is not the case.

Table V: DNA Base–Solvent and DNA Backbone–Solvent Interactions of d(GTGCGCAC)

(a) DNA Base–Solvent Interactions

	major grv site 1[a]			major grv site 2			minor grv site 1			minor grv site 2		
res	atom	solv	*d* (Å)	atom	solv	*d* (Å)	atom	solv	*d* (Å)	atom	solv	*d* (Å)
G1	O6	21W07	2.7	N7	21W06	2.8	N3			N2	18W18	2.8
T2	O4	18W02	3.0				O2	18W06	2.9			
G3	O6	13W02	2.6	N7	17W01	2.7	N3	18W18	2.6	N2		
C4	N4						O2					
G5	O6	18W13	2.4	N7	18W15	2.6	N3	17W09	2.8	N2	17W09	3.4
C6	N4	18W01	2.8				O2					
A7	N6	18W08	3.2	N7	17W02	2.7	N3	17W04	3.0			
C8	N4	18W07	2.9				O2	17W04	3.3			

(b) DNA Backbone–Solvent Interactions

	O4′	*d* (Å)	O3′ (*n* − 1)	*d* (Å)	O1P	*d* (Å)	O2P	*d* (Å)	O5′	*d* (Å)
G1			NA		NA		NA		18W04	2.6
									18W10	2.7
T2					18W03	3.3	18W05	2.9	18W03	3.4
					18W04	2.6				
G3			19W02	3.4	20W02	3.3	18W03	3.2		
			20W02	3.4						
C4	18W18	2.9	20W03	3.5	17W08	3.0	20W05	3.3	20W05	3.3
			20W05	2.5			20W07	3.0		
							23W06	3.4		
G5			19W02	3.0	19W02	2.6	20W01	2.9		
					20W02	2.3				
C6	17W05	3.2			18W09	3.0	18W17	2.9	18W09	3.3
					19W04	2.6				
A7	20W06	2.8			19W06	3.3	17W03	2.9		
					20W06	2.6	18W09	2.7		
C8	17W04	3.3			19W06	2.6	18W07	2.7		
					19W07	2.6				
*9	NA		24W03	3.3	NA		NA		NA	

[a] Site 1 is for the pyrimidine portion of the purine ring and site 2 is for the imidazole portion.

Most of the interactions between symmetry related molecules consist of vdW interactions between the terminal base-pairs and the aliphatic surface of the deoxyribose residues. Burial of potential hydrogen bond donors and acceptors in the minor groove would presumably exact a large energetic penalty. Consequentially, most of these hydrogen bonding sites are satisfied by solvent molecules.

The inner shell hydration patterns with the backbone and base polar atoms of d(GTGCGCAC) are summarized in Table V. In general, the major groove hydrogen bonding sites in both molecules are systematically hydrated. At the N4 position of C4 in d(GTGCGCAC), no ordered solvent has been identified. This in itself does not mean that the site is blocked from access to solvent; instead, it may simply imply that whatever water is associated with that site in the crystal was not sufficiently ordered to give rise to a strong peak in the difference electron density maps.

Of the 81 water molecules in this duplex structire, 32 are within 0.6 Å of the water molecules found in the tetragonal d(GTGTACAC). Most of the common solvent molecules are found at the terminal base-pairs, in the conserved hydration of the minor groove, and along the sugar phosphate backbone. The major groove hydration in the central part of the molecule changes substantially, due to the bound spermine in the tetragonal d(GTGTACAC).

A puckered pentagonal ring of water molecules is hydrogen bonded in the major groove of step 3, G3·C14 and C4·G13. The major groove base sites that are involved in the hydrogen bonding are O6 and N3 of G3 and N4 and O6 of C4 and G13, respectively. Most of the water molecules of the pentagonal ring are engaged, in turn, in hydrogen bonding with additional water molecules in the groove. Thus, the pentagonal water ring hangs in the floor of the groove like a spider web. This motif of pentagonal water rings fused together has also been observed in the major groove in another A-DNA (Kennard, 1986). Another common hydration motif observed in A-DNA is waters bridging adjacent phosphate groups in the same strand (Saenger et al., 1986; Thota et al., 1993). About half of the phosphate groups in this structure are bridged by single water molecules, while the other half are independently hydrated. Three cases of bridging waters between adjacent phosphates have been found. In two of these cases, waters span the gap between O1P of residue *n* and O2P of residue *n* + 1 (18W03 and 18W09, Table V). In a third case, a single water bridges O1P to O1P of the next residue (19W06, Table V). In a fourth case, a single water bridges the O5′ oxygen of residue 2 to the O2P oxygen of residue 3 (18W03, Table V). In $d(GTGTACAC)_{tet}$, four of the six pairs of adjacent phosphate groups were found to be linked by single water molecules.

There is a very distinctive hydration pattern of the purine bases in this structure. The two major groove hydrogen bonding sites of the purine bases (N6 and N7 of A, O6 and N7 of G) are filled separately by water molecules. These waters are in turn within hydrogen bonding distance of each other, and this chain of solvents is extended toward the phosphate group of the same residue in the backbone. One or two additional water molecules form a bridge to one of the anionic phosphate oxygens. Such a string of 3–4 waters is found for 3 of the four independent purine bases in this structure, except G1. This is not surprising, since the 5′-terminal residue does not have a 5′-phosphate group. In the closely related d(GTGTACAC), no such purine to phosphate-bridged networks were found (Jain, 1990). However, three of the four purines in d(GTGTACAC) have a pair of water molecules bound to each other and to the two major groove hydrogen bonding sites. Binding of waters to G3 is blocked by the presence of the bound spermine molecule. There is a

conserved patch of solvent molecules in the minor groove of both d(GTGCGCAC) and d(GTGTACAC). As shown in Figure 5, there is a long string of water molecules extending from the minor groove of the central dinucleotide.

ACKNOWLEDGMENT

We thank Dr. Edwin Westbrook and Mary Westbrook of Argonne National Laboratories for their helpful assistance in data collection and processing. We thank the National Institutes of Health for Grant GM-17378.

REFERENCES

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–8212.

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 48*, 1392–1398.

Bingman, C. A., Zon, G., & Sundaralingam, M. (1992) *J. Mol. Biol. 227*, 738–756.

Chandrasekaran, R., Wang, M., He, R.-G., Pugianer, L. C., Byler, M. A., Millane, R. P., & Arnott, S. (1989) *J. Biomol. Struct. Dyn. 6*, 1189–1202.

Conner, B. N., Takano, T., Tanaka, S., Ikatura, K., & Dickerson, R. E. (1982) *Nature 295*, 294–299.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419–441.

Doucet, J., Benoit, J. P., Cruse, W. B., Prange, T., & Kennard, O. (1989) *Nature 337*, 190–192.

EMBO Workshop (1989) *EMBO J. 8*, 1–4.

Heinemann, U., Lauble, H., Frank, R., & Blöcker, H. (1987) *Nucleic Acids Res. 15*, 9531–9550.

Hendrickson, W. A., & Konnert, J. H. (1981) in *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., Ed.) Vol. 1, p 43, Pergamon, Oxford.

Howard, A. J., Gilliand, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 12*, 225–238.

Jain, S. (1990) Ph.D. Thesis, University of Wisconsin—Madison.

Jain, S., & Sundaralingam, M. (1989) *J. Biol. Chem. 264*, 12780–12784.

Jain, S., & Sundaralingam, M. (1991) *Biochemistry 30*, 3567–3576.

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

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

Jones, T. A. (1985) in *Methods in Enzymology* (Wycoff, H. W., Hirs, C. H. W., & Timasheff, S. N., Eds.) Vol. 115, pp 157–171, Academic Press, New York.

Kennard, O., Cruse, W. B. T., Nachmann, J. Prange, T., Shakked, Z., & Rabinovich, D. (1986) *J. Biomol. Struct. Stereodyn. 3*, 623–647.

McCall, M., Brown, T., & Kennard, O. (1985) *J. Mol. Biol. 183*, 385–396.

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

Saenger, W., Hunter, W. N., & Kennard, O. (1986) *Nature 322*, 661–664.

Shakked, Z., Rabinovich, D., Cruse, W. B. T., Egert, E., Kennard, O., Sala, G., Salisbury, S. A., & Viswamitra, M. A. (1981) *Proc. R. Soc. London B213*, 479–487.

Takusagawa, F. (1990) *J. Biomol. Struct. Dyn. 7*, 795–809.

Thota, N., Li, X.-H., Bingman, C., & Sundaralingam, M. (1993) *Acta. Crystallogr. Sect. D* (in press).

Wang, A. J.-H., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968–3972.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.

Zon, G., & Thompson, J. A. (1986) *BioChromatography 1*, 22–32.