Nucleic Acids Research, 1992, Vol. 20, No. 24 6637–6647

Crystal and molecular structure of the alternating dodecamer d(GCGTACGTACGC) in the *A*-DNA form: comparison with the isomorphous non-alternating dodecamer d(CCGTACGTACGG)

Craig Bingman[1,3,+], Sanjeev Jain[1,§], Gerald Zon[2] and Muttaiya Sundaralingam[1,3,*]
[1]Crystallography Laboratory, Department of Biochemistry, University of Wisconsin, Madison, WI 53706, [2]Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404 and [3]Laboratory of Biological Macromolecular Structure, Department of Chemistry and Ohio State Biotechnology Center, The Ohio State University, 1060 Carmack Road, Columbus, OH 43210-1002, USA

Received August 28, 1992; Revised and Accepted November 18, 1992

ABSTRACT

The crystal structure of the alternating dodecamer d(GCGTACGTACGC) (5′-GC) has been determined to a resolution of 2.55Å using oscillation film data. The crystals belong to space group $P6_1$ 22, a = b = 46.2Å, c = 71.5Å with one strand in the asymmetric unit, and are isomorphous with a previously described non-alternating dodecamer, d(CCGTACGTACGG) (5′-CC). Refinement by X-PLOR/NUCLSQ gave a final R factor of 14.2% for 1089 observations. The molecule adopts the *A*-DNA form. The interchange of the terminal base pairs in the two dodecamers results in differences in the intermolecular contacts and may account for the differences in the bending. This dodecamer shows an axial deflection of 30°, in the direction of the major groove compared to 20° in 5′-CC and may be a consequence of additional contacts generated in 5′-GC by the interchange of end base pairs. The high helical axis deflection appreciably influences the local helical parameters. The molecule exhibits relatively high inclination angles, and has a narrow major groove. The helical parameters when described relative to the dyad-related hexamer halves of the molecule give more reasonable values. The crystal packing, local helical parameters, torsion angles, and hydration are described and also compared with the non-alternating 5′-CC dodecamer.

INTRODUCTION

The self-complementary dodecamer, d(GCGTACGTACGC), (hereafter referred to as 5′-GC) with a 5′-purine start was investigated to determine if it will crystallize either as *Z*-DNA or in a right-handed structure. The inner 10 base sequence d(CGTACGTACG) was found to crystallize in the expected *Z*-DNA form, typical of all known alternating DNA with a 5′-pyrimidine start (Brennan & Sundaralingam, 1985, Brennan *et al*, 1986). If the dodecamer were to crystallize as Z-DNA, it would represent a full helical turn of this conformation, with 12 base-pairs. If the dodecamer crystallized in a right-handed form, it would provide additional confirmation that alternating sequences with a 5′-purine start prefer to be right-handed rather than left-handed *Z*-DNA. Our earlier crystallization studies had shown that many 5′-purine start sequences crystallized as *A*-DNA duplexes, such as the octamers d(GTGTACAC) (Jain *et al*, 1987, Jain *et al*, 1989) and d(GTGCGCAC) (Bingman *et al*, 1992b). These observations are consistent with those of Quadrifoglio *et al*.(1984). The present alternating 5′-GC dodecamer differs from the non-alternating dodecamer d(CCGTACGTACGG) (hereafter referred to as 5′-CC) (Bingman *et al*, 1992a) only in the interchange of the two end base-pairs. In this the structure of 5′-GC dodecamer is presented and compared with the 5′-CC dodecamer.

EXPERIMENTAL

Synthesis and crystallization

This compound was synthesized using phosphoramidite chemistry on an Applied Biosystems synthesizer. The material was removed from the solid support, purified with the 5′-trityl group attached, detritylated, ethanol precipitated, exchanged with D_2O and subjected to NMR spectroscopy to confirm that all the triethylamine had been removed by the ethanol precipitation, and dried (Zon & Thompson, 1986). The sample was dissolved in sodium cacodylate buffer, pH = 6.5, annealed by heating to 100°C and cooled slowly first to room temperature, and finally to cold room temperature (4°C). Since the Z-DNA decamer d(CGTA-

* To whom correspondence should be addressed

Present addresses: [+]Department of Biochemistry & Molecular Biophysics, Columbia University, New York, NY 10032 and [§]Department of Biochemistry, St Louis University, St Louis, MO 63104, USA

CGTACG) had been crystallized in the presence of cobaltic hexammine and spermine with 1-propanol as the precipitant, it was decided to use similar initial crystallization conditions. Preliminary titrations with cobaltic hexammine and spermine were performed in a cold room to determine the approximate concentrations of these modifiers required to effect crystallization. The crystals were grown by vapor diffusion of a droplet containing 1mM DNA duplex, 1mM spermine, 3mM cobaltic hexammine in an aqueous sodium cacodylate buffered (pH=6.5) solution against a reservoir composed of 30% 1-propanol in water.

Data collection

The preliminary precession and Weissenberg photographs indicated either space group $P6_122$ or $P6_522$. The majority of the data were collected by oscillation photography (Arndt & Wonacott, 1977) using an Elliot GX6 rotating anode operating at 40 kV, 40 mA. There were two Kodak DEF-5 films per cassette, each cassette was exposed over an angular range of two degrees at 7500 seconds per degree, covering a total angular range of 32 degrees. The films were processed and dried in Madison, then scanned and digitized at the University of Chicago. The intensity data were extracted from the digitized film data with DENZO (Otwinowski *et al*, 1988) Approximately half the reflections were partially recorded. Two film packs were rejected due to high R-factors relative to the rest of the packs, with individual R-factors of 0.08 and 0.11. This left 2008 observations of 1115 unique reflections with $I>2\sigma(I)$, and an R-merge on intensity of 0.05. The oscillation films showed in addition to the sharp Bragg diffraction maxima, diffuse scattering, the most prominent was along c* at a spacing of 3.4Å. This suggested that the crystal contained disordered B-DNA conformation, as in the 5'-CC (Bingman et al., 1992a). Thirty one high-intensity reflections that were overloaded on the second film were replaced from diffractometer measurements on another crystal. The diffractometer data set contained 128 reflections with $I \geq 1.5\sigma(I)$ in the resolution range 40−5Å and were scaled to the oscillation data set.

Structure solution

Several attempts were made to solve this structure by molecular replacement. On the basis of the appearance of the Patterson function at the w=1/6 section, which (for these space groups) in principle contains a c-axis projection of the electron density.

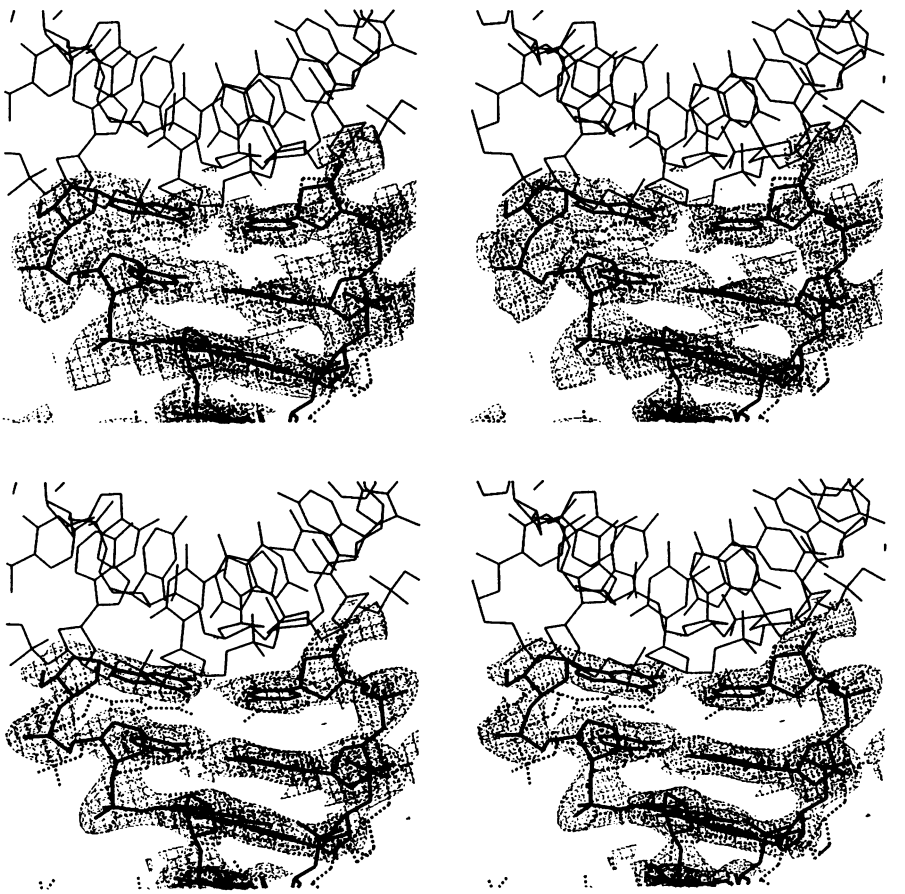

Figure 1. Top: Starting electron density for d(GCGTACGTACGC). The starting model is shown in heavy dashed lines, the final refined model in heavy solid lines, and the symmetry related molecule is in lighter solid lines. The 5'-end is on the left and the 3'-end is on the right. The electron density is rendered as fine dashed lines. The phases for this $3F_o-2F_c$ electron density map were obtained from the fiber diffraction model placed at the correct position and orientation in the unit cell. The map is contoured at approximately 1σ. The density for the terminal base pairs is weak and discontinuous, although it is clearly shifted from the modeled position toward the position given by the final model, even though no parts of the model have been omitted in determining the phases. **Bottom:** Electron density after X-PLOR refinement. The various elements in this figure are rendered as in figure 1a. The electron density is much clearer after simulated annealing refinement.

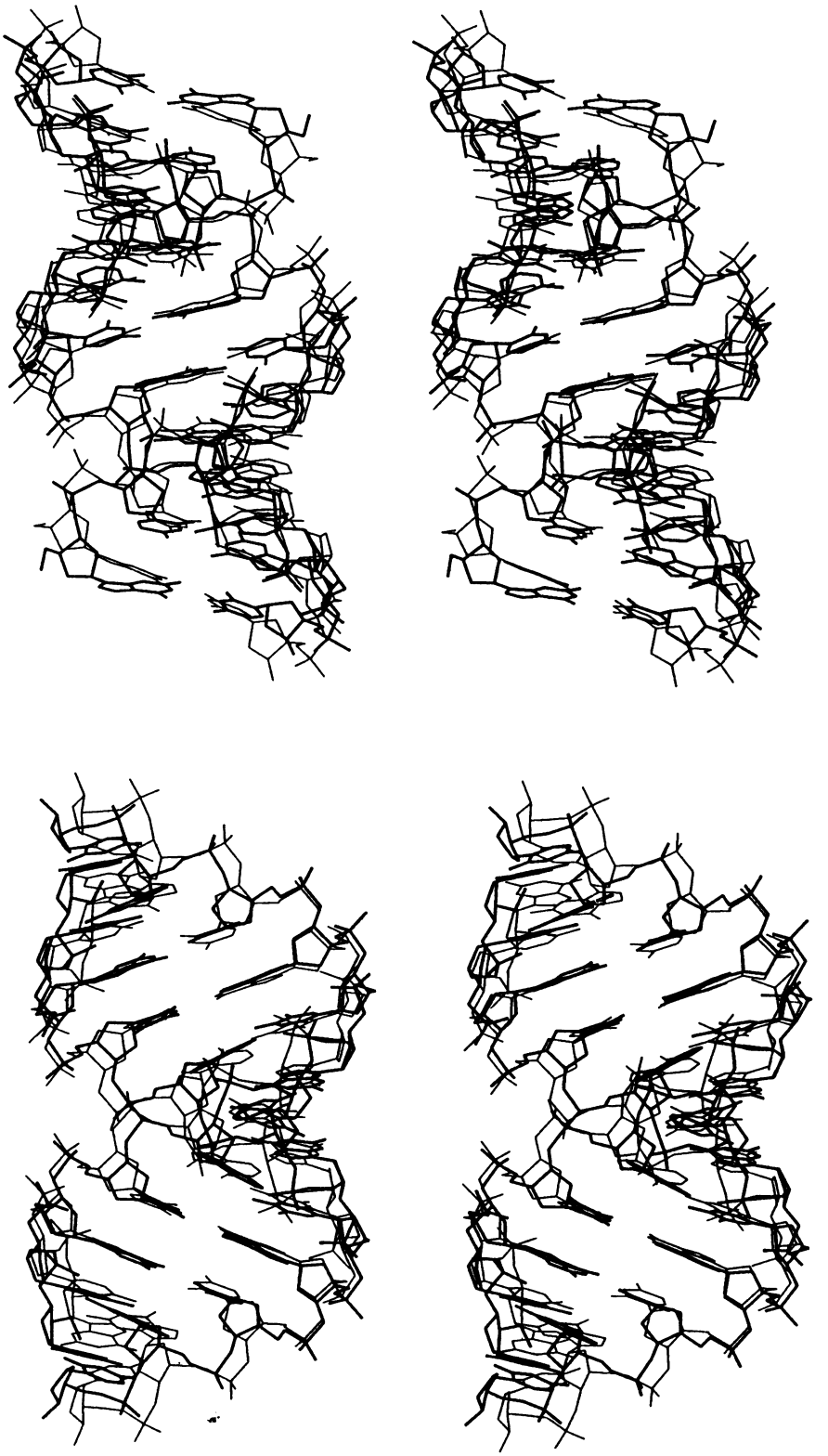

Figure 2. Superposition of the starting and refined models of d(GCGTACGTACGC). The refined model is rendered in heavy lines, the starting fiber diffraction model is in light lines. Major positional shifts can be seen at the molecular termini, where there is both a translational shift of more than a bond length. The upper and lower halves of the molecule seem to be rotated towards the major groove side. Both of these changes seem to be related to the need to form a close but non-interpenetrating vdW contact with symmetry related molecules. The deflection in the helical axis is quite apparent. The overall helical axes of the starting and refined molecules is vertical in both views. **Top:** The view is into the major groove. The 5′-end is on top right and the 3-end on top left. **Bottom:** The view is with the molecular dyad in the plane of the page. Overall, the major groove of the molecule is quite narrow compared and the bases are highly inclined compared to A-DNA octamer crystals.

Duplex models of *A*-, *B*- and *Z*-DNA were initially restricted to the class of twofold running in the a-direction in both space groups $P6_122$ and $P6_522$, using the 104 reflections between 10–5.5Å resolution. However, no acceptable models were found in this restricted search.

Concurrently, structural study of the closely related isomorphous dodecamer d(CCGTACGTACGG) (Bingman *et al*, 1992a) was also underway. The cell constants of the two crystals were identical, a=46.2, c=71.5Å. The structure of d(CCGTACGTACGG) was determined to be *A*-DNA with a real-space structure factor calculation search along the alternative twofold axes at z=1/12, sometimes called the diagonal twofold, that had been previously neglected in the study of d(GCGTACGTACGC). A fiber diffraction model (Arnott *et al*, 1972) of the sequence d(GCGTACGTACGC) was then placed in the crystal lattice at the orientation discovered for d(CCGTACGTACGG), and a fine-grid search (±1Å and ±5°) was performed. The highest correlation coefficient 0.48 and lowest R-factor 0.45 in this section was found at the same orientation and translated 0.2Å from the solution for d(CCGTACGTACGG). At this point, a six-parameter minimization with RVMIN (Sekharudu and Jain, unpublished) was conducted which increased the correlation coefficient to 0.60 and reduced the R-value to 0.38 for the 104 reflections used in the searches. In retrospect, only three potential solutions encountered in other searches had comparable correlation coefficients but none of these could be refined. The solution could have been obtained *de novo* by searches along both twofold axes. The c-axis projection of the correct solution also showed remarkable similarity to the w=1/6 section of the Patterson map.

Table 1. Restrained parameters for d(GCGTACGTACGC)

Restraint	#	RMS Dev.	Σ	# Dev. > 2σ
Distance restraints				
bonds	456	0.006 Å	0.025 Å	8
angles	672	0.020 Å	0.050 Å	
PO_4bonds	88	0.0024 Å	0.050 Å	
PO_4angles, h-bonds	524	0.045 Å	0.075 Å	
Planarity restraints	252	0.015 Å	0.0100 Å	0
Chiral center restraints	72	0.048 Å^3	0.0100 Å	0
Non-bonded contacts				
single torsion	26	0.146 Å	0.090 Å	10
multiple torsion	24	0.111 Å	0.090 Å	8
Thermal parameter restraints				
bonds	456	2.4 Å^2	3.0 Å^2	0
angles	672	3.3 Å^2	4.5 Å^2	0
PO_4bonds	88	3.1 Å^2	3.0 Å^2	0
PO_4bonds, h-bonds	524	3.9 Å^2	4.5 Å^2	0
Non-Crystallographic Symmetry				
DNA atoms				
position	243			
0.018 Å	0.200 Å	0		
thermal param.	243	0.33 Å^2	5.0 Å^2	0
solvent atoms				
position	33	0.055 Å	0.200 Å	0
thermal param.	33	0.63 Å^2	5.0 Å^2	0

Refinement

The initial refinement of this rigid-body minimized fiber-diffraction model was performed with X-PLOR (Brünger, 1988). The starting R-value was 0.514 for the data between 5 and 2.55Å. Figure 1 (top) shows the electron density map calculated with phases from the properly oriented fiber model. This map shows more noise than the comparable map of d(CCGTACGTACGG).

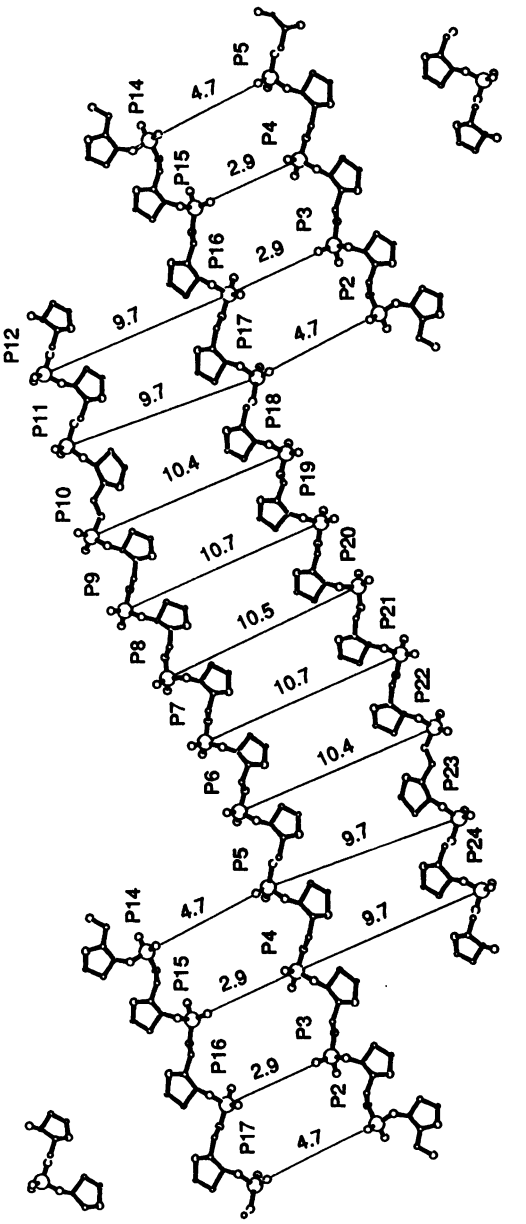

Figure 3. Cylindrical projection of backbone atoms. This cylindrical projection was calculated at 10Å and covers an angular range of approximately 540°. The view is from outside the cylinder of projection. The groove dimensions are indicated next to solid lines connecting the limiting phosphate groups. As with d(CCGTACGTACGG), there is an approximately 1Å variation in the width of the minor groove, increasing from 9.7Å at the termini of the duplex to 10.7Å near the center. There are two independent measurements of the major groove dimension, 2.9 and 4.7Å. By contrast, the major groove dimension of fiber diffraction A-DNA is 2.4Å.

The model was subjected to energy minimization with a soft van der Waals' (vdW) repulsive potential to relieve short contacts. This initial minimization with amplitude restraints reduced the R-value to 0.323. The model was then subjected to the simulated annealing procedure, at a nominal temperature of 2000K. This reduced the R-factor to 0.245 over the 5−2.55Å data. The model was then cooled to 300K and additional energy minimization with X-ray terms was performed, lowering the R-factor to 0.198. Figure 1 (bottom) shows the electron density for the X-PLOR refined model.

Group thermal parameters were refined in the final stages of X-PLOR, which had a small effect on the R-value of the model, bringing it to 0.194. Since there are no restraints relating the thermal parameters of adjacent groups, the thermal parameters of the model were reset to 15, and the model was subjected to NUCLSQ least squares refinement (Hendrickson & Konnert, 1981, Westhof *et al*, 1985). Initially, all the data between 5 and 2.55Å were used, with an R-factor of 0.204. The first round of 18 cycles of NUCLSQ least squares refinement of both thermal and positional parameters brought the R-value to 0.203. The geometry as defined in the library of constrained distances was substantially improved, with no bond distance and bond angle distance constraints deviating more than 2σ from target values, no violations greater than 2σ of planarity constraints, chiral center constraints and local symmetry between the two strands of the dodecamer. There were 6 close non-bonded contacts. No further

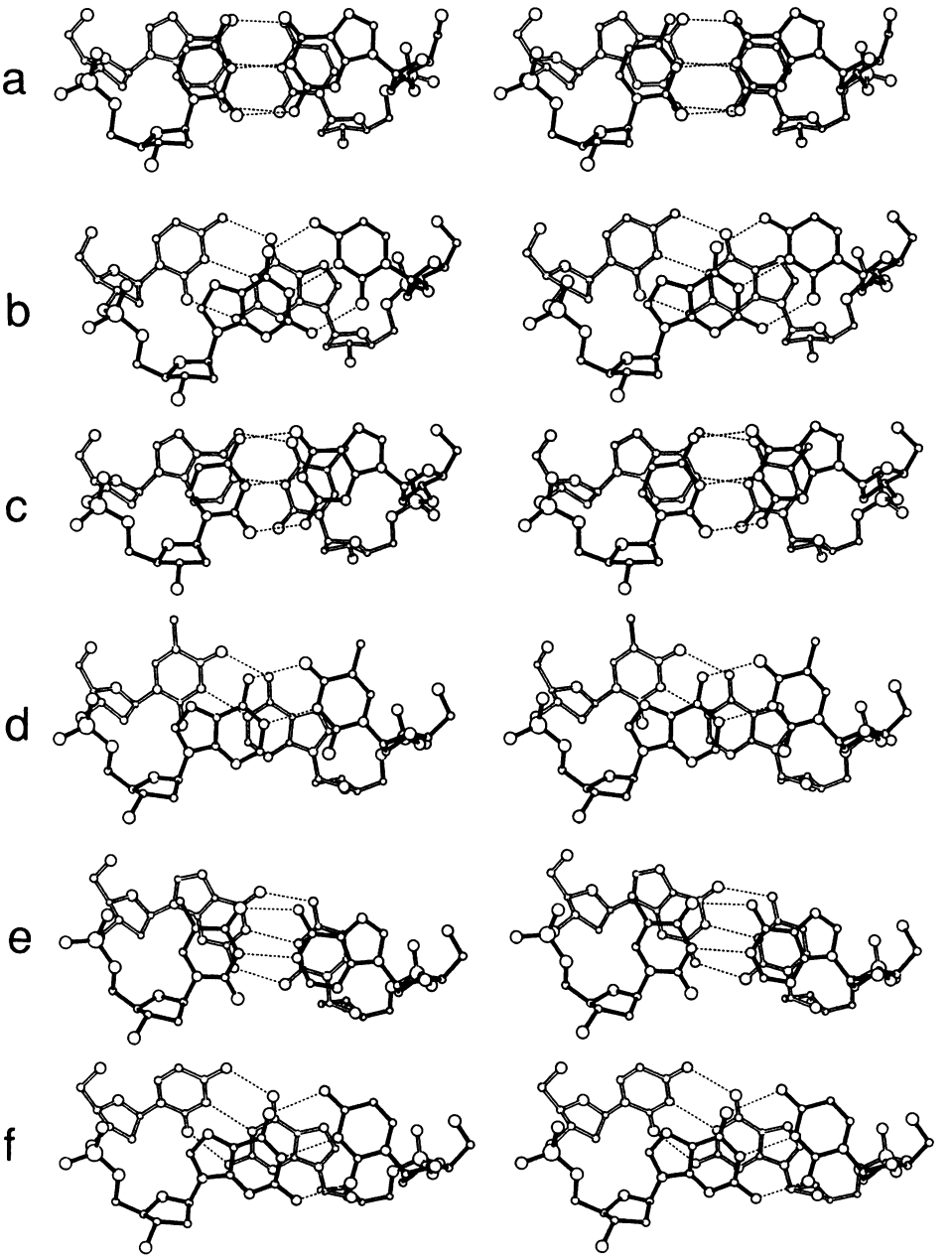

Figure 4. Dinucleotide stacking diagrams. The view is down the upper base normal. The top base pair is drawn with filled bonds, the bottom base pair with open bonds. A: G1-C2. B: C2-G3. C: G3-T4. D: T4-A5. E: A5'-C6. F: C6-G7.

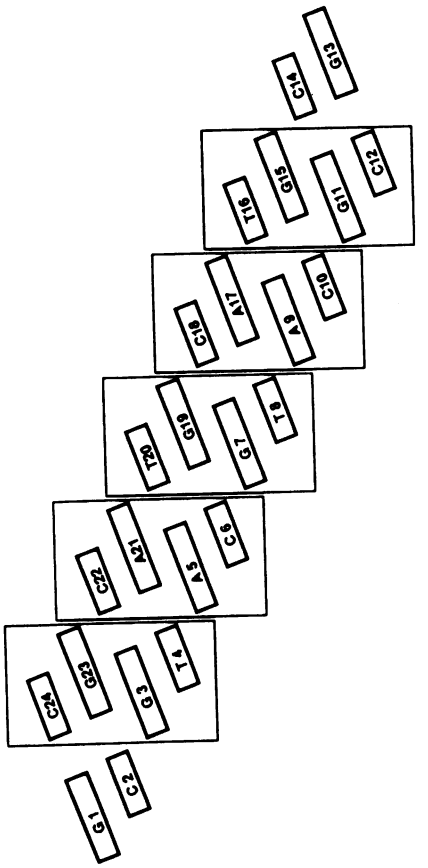

Figure 5. Schematic view of the base stacking. Bases are drawn approximately to scale, and view is similar to a cylindrical projection at r=7Å, as viewed from within the cylinder. The vertical stacking domains are shown as boxes. Watson-Crick hydrogen bonding interactions hold the adjacent stacks together. Although this stacking pattern is apparent even in fiber diffraction models, it may be important in conferring conformational flexibility to this sequence.

Table 2. Comparison of backbone torsion angles(°), χ, P and τ_m d(GCGTACGTACGC) and d(CCGTACGTACGG)

Sequence	α	β	γ	δ	ϵ	ζ	χ	ϱ	τ_m
G1	–	–	133	90	257	256	180	-4	33
C1	–	–	139	84	222	278	190	16	36
C2	313	132	63	76	218	276	184	19	43
C2	288	161	77	74	221	271	187	28	46
G3	301	171	70	86	204	291	197	27	36
G3	301	162	77	84	212	291	193	26	36
T4	293	187	49	78	192	284	206	24	45
T4	282	185	61	78	195	283	205	24	45
A5	309	175	42	86	201	287	212	7	42
A5	275	181	79	79	217	273	201	14	42
C6	269	181	73	79	215	278	201	12	46
C6	331	168	30	84	204	286	212	13	42
G7	291	173	66	79	225	268	194	28	40
T8	315	164	41	83	194	295	204	17	39
T8	316	168	35	78	206	281	205	18	42
A9	301	187	44	87	189	303	210	9	34
A9	284	164	73	78	218	282	196	17	43
C10	244	179	101	81	210	277	185	25	39
C10	289	176	65	82	209	281	192	10	43
G11	300	177	48	80	199	291	203	16	41
G11	304	176	47	81	191	292	197	23	41
C12	275	182	69	80	–	–	209	14	40
G12	285	191	63	79	–	–	205	11	46

Table 3a. Comparison of overall and fragment helical axes, d(GCGTACGTACGC)

Fragment	Angle	Distance to global helical axis	Helix Rotation	Base Pairs per turn	Helical Rise per residue
1–12 (entire molecule)	***	***	31.9(4.5)°	11.27	2.62(58)Å
1–6	14.7°	0.48Å	31.7(3.4)°	11.37	3.26(25)Å
2–6	13.6°	0.38Å	32.5(3.1)°	11.08	3.15(23)Å
3–6	13.8°	0.59Å	32.3(2.4)°	11.16	3.21(27)Å
7–10	13.7°	0.61Å	32.3(2.4)°	11.16	3.20(27)Å
7–11	13.6°	0.38Å	32.5(3.1)°	11.08	3.15(23)Å
7–12	14.7°	0.48Å	31.7(3.4)°	11.36	3.27(26)Å

*** not defined

Table 3b. Pair-wise comparison of fragment axes: angle between fragment axes and distance between fragment axes

Fragment	1–6	2–6	3–6
7–12	29.4° 0.80Å		
7–11		27.2° 0.94Å	
7–10			27.5° 1.18 Å

reduction in R-factor while maintaining reasonable geometry was possible, therefore, a series of $3F_o$-$2F_c$ and F_o-F_c 'omit' electron density maps were calculated. For each map, two adjacent residues were omitted from the phasing model, and the entire molecule was spanned by this series of omit maps. The model was refitted into these maps on an Evans & Sutherland PS340 with FRODO. At this stage, solvent molecules were picked when roughly spherical peaks appeared at 2σU in the F_o-F_c map within 2 to 3.5Å of potential DNA hydrogen bonding partners. Eleven ordered solvents per strand were identified in the first round of map fitting. In subsequent rounds of refinement, all the reflections between 8–2.55Å were used (1089 of 1146 independent reflection in $P6_122$). Two more rounds of map fitting and solvent identification followed by additional least squares refinement brought the final R-value to 0.142, with 33 ordered solvent per DNA strand. The stereochemistry of the final model is quite good, as exemplified by RMS deviations of 0.005Å in bond lengths and 1.2° in bond angles of the deoxyribose groups and bases from library values. The final refinement statistics are detailed in Table 1. The coordinates and the structure factors have been deposited with the Brookhaven Protein Data Bank.

RESULTS AND DISCUSSION

The refinement of the 5'-GC structure was carried out with 1089 independent reflections, compared to 1664 reflections for the 5'-CC (Bingman et al., 1992a). Some of the differences between the two structures may arise from the difference in their precision.

Overall conformation

The overall conformation of d(GCGTACGTACGC) (5'-GC) is presented in figure 2. The molecule is very similar to d(CCGTACGTACGG) (5'-CC). The helix axes of both molecules are

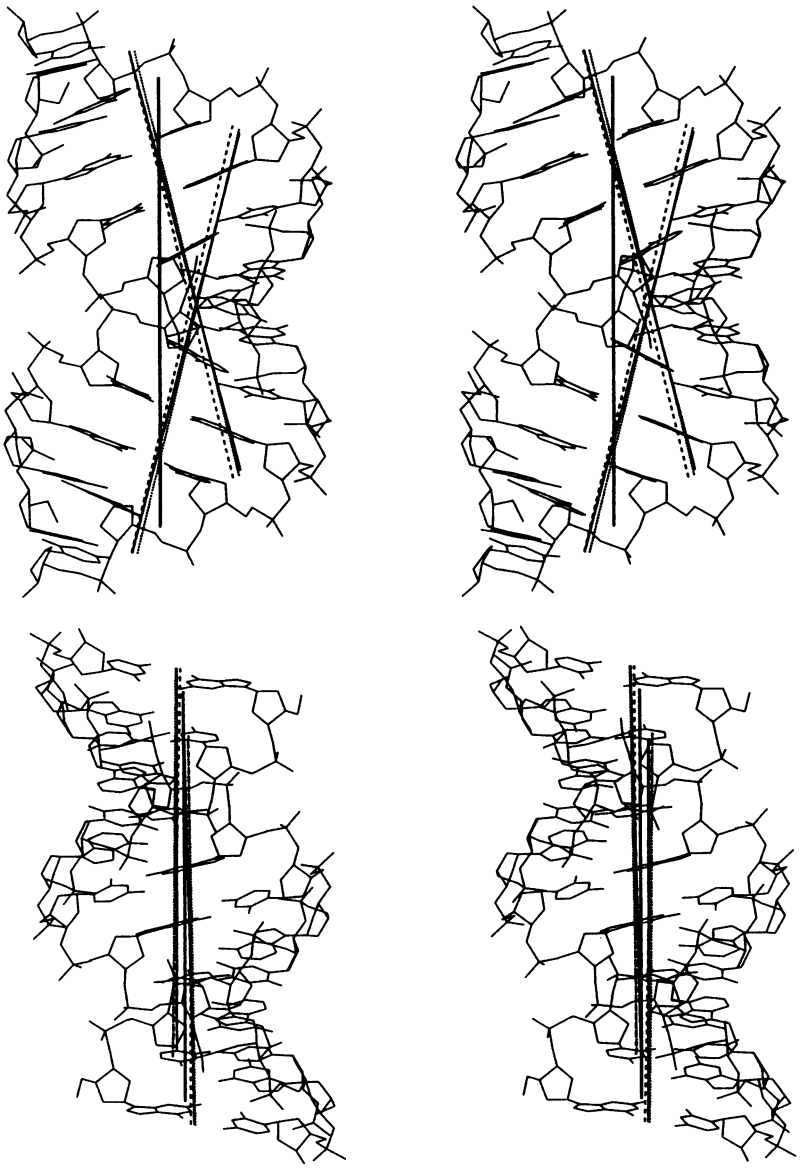

Figure 6. Plot of overall and fragment helical axes. **Top:** The view is with the overall helical axis vertical, and the molecular dyad horizontal. The overall helical axis is drawn with a solid line, as are the two longer fragment axes. The shorter fragment axes are drawn with progressively finer dashed patterns. All the fragment axes agree well with each other. For a quantitative treatment of these axes, see table 3. **Bottom:** The view is with the molecular dyad projecting vertically from the page. The 5'-end is on top right and the 3'-end on top left. There is a translational discrepancy in the fragment axes, which is described in tables 3.

deflected, and the major groove is narrow compared to the A-DNA octamers. A cylindrical projection of the backbone atoms and the phosphate-phosphate separations delimiting the major and minor grooves are shown in figure 3. Like 5'-CC, the width of the minor groove is more variable than most other A-DNA structures, 9.7Å at its narrowest and 10.7Å at its widest point. The base reversal at the ends of 5'-CC has converted the molecule to a purely alternating sequence. Like 5'-CC, the molecule shows clearly separated vertical stacks of bases when viewed down the helix axis (figure 4, 5) and the stacking patterns are also quite similar. The backbone conformation of both dodecamers lies squarely in the range of values expected from fiber diffraction studies (Chandrasekaran *et al*, 1989) and previous single crystal studies of A-DNA. The backbone angles and the deoxyribose parameters are presented in tabular form in table 2. The all-trans backbone conformation does not appear at any location in either of the independently refined dodecamer structures.

Helix axis analysis

The present dodecamer displays the same 'sinusoidal' variations in several of the helical parameters as the 5'-CC dodecamer, which also showed mis-matches between the local and global helical axes. The best overall helical axis and derived helical parameters were calculated over the entire duplex using NEWHEL91 (R.E.Dickerson, personal communication.) The local helical axes were calculated using atoms of the half molecules, base pairs 1−6 (and 7−12). Calculations were also done for base pairs 2−6 and 3−6 to check if the aberrations in the helical parameters were due to end effects. The angles between the various calculated helical axes and the closest approach between the axes were calculated with a local program (Bingman, unpublished results.) The results of these analyses are presented in tables 3a,b. The standard deviation of the helical rise per residue is quite high, 0.58Å. When helical axes for the dyad-related half duplexes are calculated, the standard deviation

Table 4a. Two-base pair descriptive helical parameters comparison between d(GCGTACGTACGC) and d(CCGTACGTACGG)

Step	Twist (°)	Rise (Å)	Roll (°)	Tilt (°)	Cup (°)	Slide (Å)
G1|C2	27.0 **27.2**	3.18 **3.49**	**−0.3** **1.0**	**7.3** **0.6**	−6.9	−1.1
C1|C2	28.1 **28.7**	2.73 **2.96**	−4.4 **−3.9**	7.0 **2.1**	−5.3	−1.6
C2|G3	36.3 **35.5**	2.24 **3.05**	−8.1 **−1.4**	5.3 **−0.8**	−7.3	−1.6
C2|G3	33.2 **32.7**	2.39 **2.91**	−5.9 **−2.2**	5.2 **1.0**	−1.9	−1.8
G3|T4	31.5 **29.9**	2.13 **3.03**−12.1	**−3.1**	2.7 **1.3**	1.5	−1.4
G3|T4	34.7 **33.3**	2.26 **2.88**	−9.1 **−2.6**	1.7 **0.2**	−0.2	−1.1
T4|A5	34.1 **32.3**	2.50 **3.32**	−3.2 **4.9**	−1.8 **1.3**	−6.1	−1.2
T4|A5	34.3 **32.9**	2.57 **3.15**	−2.8 **3.2**	−1.0 **0.9**	0.5	−1.4
A5|C6	31.6 **31.0**	2.95 **3.41**	1.6 **6.1**	−5.8 **0.7**	−1.3	−0.8
A5|C6	32.6 **31.7**	2.55 **2.99**	−0.5 **3.1**	−4.5 **−0.1**	−5.2	−1.2
C6|G7	33.1 *31.9*	3.84 *2.79*	13.6 *14.9*	0.0 *7.9*	16.0	−2.1
C6|G7	30.7 *30.3*	3.03 *3.05*	9.1 *10.1*	0.0 *5.0*	12.2	−1.4

Table 4b. One-base pair descriptive helical parameters comparison between d(GCGTACGTACGC) and d(CCGTACGTACGG)

Sequence	Tip (°)	Inclination (°)	Prop. Twist (°)	Buckle (°)
G1	15.1 **0.3**	8.0 **7.8**	−3.9	12.0
C1	18.0 **7.8**	8.4 **9.3**	−4.2	6.2
C2	14.5 **1.4**	15.1 **8.4**	−7.0	5.2
C2	13.3 **4.0**	15.3 **11.4**	−5.0	0.9
G3	6.4 **−0.1**	20.8 **7.6**	−9.1	−2.1
G3	7.4 **1.9**	20.8 **12.4**	−6.7	−1.1
T4	−4.8 **−3.1**	23.5 **9.0**	−8.6	−0.6
T4	−1.2 **−0.7**	22.7 **12.7**	−10.4	−1.3
A5	−7.8 **1.7**	21.5 **10.2**	−11.7	−6.7
A5	−3.8 **2.4**	21.6 **13.5**	−13.5	−0.8
C6	−6.5 **7.7**	15.8 **10.8**	−14.2	−8.0
C6	−4.4 **5.4**	17.1 **13.4**	−13.1	−6.1

Normal: Overall helical axis.
Bold: Half-molecule helical axis.
Italic: Half-molecule, extending over fragment boundary.

in the helical rise is reduced to 0.25Å. Truncation of the helix to shorter double helical fragments give nearly the same standard deviations in the helical rise. For example, deletion of the terminal base pair gives a standard deviation of 0.23Å, and removal of also the penultimate base pair gives a standard deviation of 0.27Å. The number of base pairs per turn increases from 11.27 to 11.37 when the overall axis is abandoned in favor of the half-molecule axes. Deletions of the terminal base pairs reduce this value somewhat, primarily because the terminal base pairs in both dodecamers are somewhat underwound with respect to the rest of the molecule.

The 5′-GC dodecamer shows some differences compared to the 5′-CC dodecamer. First, the overall deflection in the half-molecule helical axes is somewhat larger, 30° for the 5′-GC as opposed to 20° for 5′-CC. The distance between the global helix axis and the local axis is also somewhat larger for this dodecamer, about 0.5Å compared to less than 0.1Å. It is also noteworthy that the axes of the shorter helical fragments seem to agree better with each other in this case than for 5′-CC, the angular scatter with respect to the overall helix axis being only 1.1°. This analysis indicates that this dodecamer seems to be abruptly deflected near the central helical step, whereas in the 5′-CC dodecamer, the deflection locus appears to be somewhat distributed. The ensemble of helical axes along with the molecular structure is shown in figure 6. The RMS deviation between the refined structure and the starting fiber diffraction model is 1.2Å, figure 2. The RMS and mean deviations between the constant central ten base-pairs of the 5′-GC and 5′-CC dodecamers are 0.36Å and 0.33Å respectively.

The derived helical parameters for 5′-GC are compiled in tabular form in tables 4a and 4b, following the nomenclature and sign conventions adopted at the EMBO Workshop (1989). For the parameters the values with respect to the global and half-molecule helical axes are given. The helical rise per residue for the global axis shows a substantial amount of scatter, with large increases above the mean occurring at the extreme end steps and the steps closest to the dyad. Steps in between show local rise much less than the 2.56Å observed for fiber diffraction A-DNA. There is a substantial difference in rise at the ends of the duplexes, presumably reflecting the fine conformational adjustments of reversing the terminal base pairs against adjacent molecules in the crystal. There are also differences in the inner ten base pairs, particularly at the center of the molecule, where the sequence is identical. It is precisely at this region that symmetry related molecules interact with the reference dodecamer. Perhaps these changes are due to stacking adjustments needed to accommodate the reversed end base pairs on the sugar phosphate backbone of symmetry related duplexes.

When the helical rise is calculated for the half molecule axes, the largest difference in the two molecules is seen at the terminal step, with smaller shifts at the center of the molecule, which are in the same direction as seen in the half molecule. The large sag in helical rise for the half molecule not included in the axis calculations is an amplified version of the variation seen in both

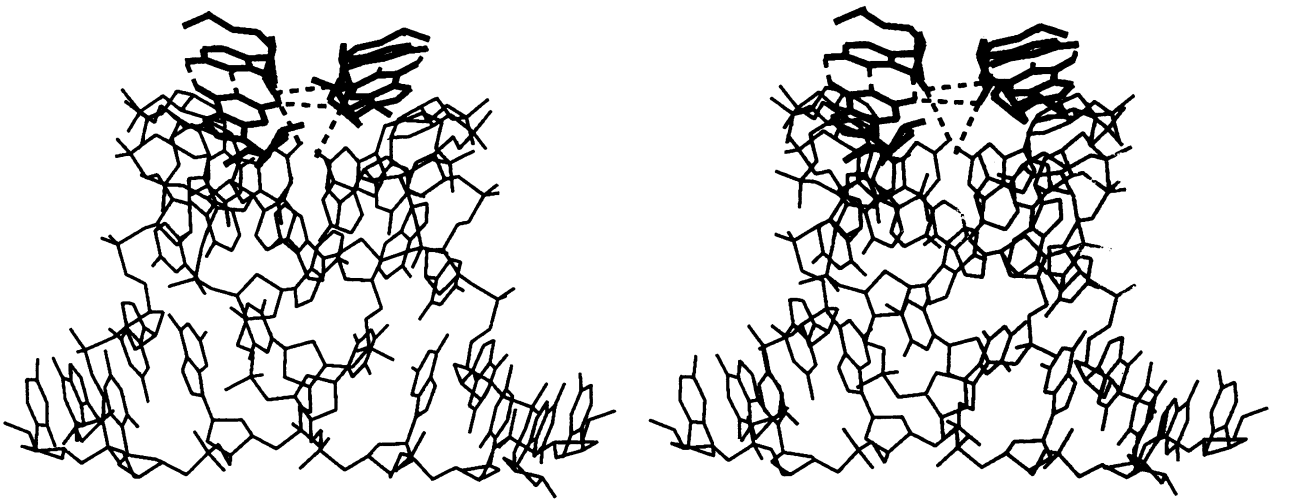

Figure 7. Crystalline environment and intermolecular hydrogen bonds. The reference duplex is shown in thin solid bonds, and the terminal base-pairs impinging on the two strands of the minor groove region are shown in thick solid bonds. The symmetry-related molecules are linked to each other and to the reference duplex by a network of hydrogen bonds. There is also an extensive van der Waal's interaction between the planar terminal base pairs and the aliphatic minor groove, as well as an interlocking, complementary interaction between the terminal O3'-hydroxyl group and the recess of the minor groove.

Table 5. Intermolecular contacts (<3.6Å) with the terminal bases in 5'-GC and 5'-CC dodecamers

Res/Atom	Res/atom	Dist	Res/Atom	Res/atom	Dist
5-GC dodecamer			**5-CC dodecamer**		
Interaction with G1			*Interaction with C1*		
N1(G1)	C4'(G19)	3.42Å	None with base C1		
	O4'(G19)	3.36			
C2(G1)	O4'(G19)	3.57			
N2(G1)	**O3'(C12)**	**3.27**			
C4(G1)	C2'(C18)	3.58			
C6(G1)	C4'(G19)	3.42			
	C5'(G19)	3.51			
O6(G1)	C4'(G19)	3.58			
	C5'(G19)	3.42			
Interaction with C24			*Interaction with G24*		
N3(C24)	C4'(G19)	3.45	N1(G24)	O4'(G19)	3.18
C4(C24)	O3'(G19)	3.41		C4'(G19)	3.39
N4(C24)	O3'(G19)	3.41	N2(G24)	O3'(G12)	3.72*
			C2(G24)	O4'(G19)	3.47
			C6(G24)	C4'(G19)	3.15
				O4'(G19)	3.47

Hydrogen bonds are shown in **bold**
* favorable interaction

halves of the molecule with respect to the overall axis. This proves that this poor correspondence between local helical axes and the axis used in the calculation can give rise to precisely the sort of systematic variations seen in the overall axis case.

The helical twist values for this dodecamer clearly show a sequence induced variation, with 5'-RY-3 steps having low helical twist angles, and 5'-YR-3 steps exhibiting higher helical twists. The general variation in the baseline of helical twists in both molecules is quite similar, with these sequence induced effects being more noticeable for 5'-GC. As with 5'-CC, choice of helical axis has less impact on the precise values of this parameter, although they are clearly distorted in the right half of the plot (Figure 10 of Bingman *et al.*, 1992a)

The overall inclination angles for both molecules are highly conserved, with the largest variations at the central two base pairs. The maximum in this plot is centered around residue 4, which is closely related to the minimum seen in the helical rise per residue around this position. The base stacking distance is approximately 3.4Å. Low helical rises such as seen in fiber diffraction A-DNA must be associated with large inclinations, so that only a fraction of the inter-base-pair distance is directed along the helical axis. It has been observed that in general, there is an inverse relationship between helical rise and base pair inclination angles in A-DNA structures as a group (Heinemann *et al*, 1987; Fairall *et al*, 1989). Within the dodecamer structure this correlation also holds with the maximum inclination

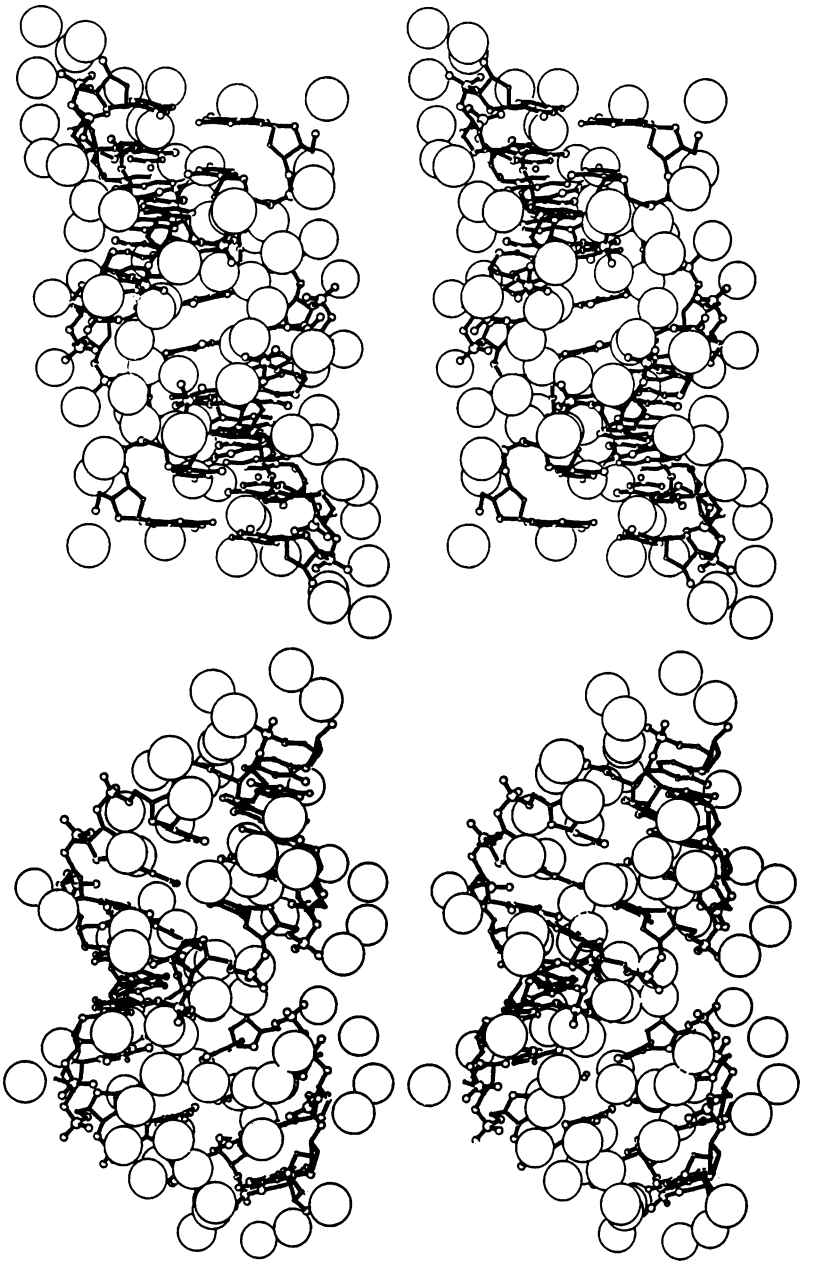

Figure 8. Hydration of d(GCGTACGTACGC) **Top:** The view is into the major groove. **Bottom:** The view is rotated 90° with respect to the top figure.

corresponding with the minimum in local helical rise, although in this case the variation is clearly an artifact of the poor match of the local and global helical axes.

The inclination angles calculated with respect to the half-molecule axes are more different, this difference is primarily due to the 5° discrepancy in the half-molecule axes for the two molecules. The largest variations in the two molecules directly reflect this difference in helical axes. While there appears to be some residual systematic rise in the inclinations with 5′-CC, such effects seem to be almost entirely eliminated in 5′-GC. It may be that the half-molecule axes calculated for 5′-CC still do not adequately describe the local geometry of the molecule, although this description greatly reduces the systematic variation in this parameter in both structures.

The tilt angles for the two molecules calculated with respect to the overall helical axes are also almost precisely preserved in the two dodecamers. The pattern of tip angles for the half-molecule axes is slightly different, with the largest changes occurring over two outermost steps. Presumably, this difference is caused by the sequence reversal at the duplex termini, although as has been noted above, it may not represent a sequence-dependence in the true sense, since these base pairs with the changed sequence are also involved in inter-molecular interactions.

In both dodecamers, the roll angles calculated with respect to the overall helical axis show a sharp increase at the central step, the locus of most of the bending. The roll angles are shifted to slightly more positive values when calculated for the half-molecule only, with the largest changes observed at the ends of the molecules.

Crystal packing

The 5'-GC and 5'-CC dodecamers crystallize isomorphously with each other and show very similar crystal packing arrangements. In both cases, the duplex is surrounded by six nearest neighbors in the crystal lattice, making two terminal base-pair to minor groove interactions, minor groove to terminal base-pair interactions with two other symmetry related molecules, as well as displaying the novel minor groove to minor groove packing interaction similar to that described for the 5'-CC dodecamer (Bingman *et al*, 1992a). As with the 5'-CC dodecamer, there is a large solvent channel running along the c-axis of the crystal, which is sufficiently large to accommodate disordered *B*-DNA of the same sequence. Figure 7 shows the crystalline environment near the center of the 5'-GC dodecamer. As can be seen, the two symmetry related molecules impinging on the central minor groove region of the reference duplex are linked to each other by hydrogen bonds between the O3-hydroxyl of the 3-terminal molecule and the N2 group of the 5'-terminal guanine (3.3Å). The symmetry related terminal-O3 groups are also hydrogen bonded to the N2 group of the seventh guanine in the reference duplex. The hydrogen bond involving the terminal O3 and the −5 terminal G is not present in the 5'-CC dodecamer. Instead there appears to be a favorable interaction between the terminal 3-hydroxyl group and N2 of G7 (3.5Å).

In 5'-CC, the electron density for the 5'-ultimate cytidine was substantially weaker than for its Watson-Crick partner, G24. The vdW contacts involving G24 were substantially more numerous than for C1. In the present dodecamer, the terminal base pair is reversed. The electron density for G1 seems more robust as expected, and it makes a number of vdW contacts with the symmetry related molecule (Table 5). This need for improved stacking with the 5'-terminal base may be responsible for the increase in rise at the terminal steps of 5'-GC relative to 5'-CC and for the differences in the bending of the two DNA duplexes.

Solvent structure

The pattern of water molecules surrounding the dodecamer is illustrated in figure 8. As was the case with 5'-CC, the potential major groove sites are not all associated with water molecules. Only very strong peaks were included in the model, and the number of waters included in the model is probably an underestimate. When the solvent sites of the 5'-GC and 5'-CC are compared, 10 sites agree to within 0.7Å, 12 to within 0.8Å and 20 to within 1.0Å. Generally, it seems that in both dodecamers, the anionic phosphate oxygens have a large number of bound solvents, especially O2P, and that the ester O3 and O5 oxygens are less highly hydrated, with 5 of 12 O3 atoms having an attached water and only 2 of 12 O5 atoms with attached waters. The major and minor grooves have few waters associated with them.

CONCLUSIONS

The 5'-GC and 5'-CC A-DNA dodecamer structures represent the most distorted of all the A-DNA crystal structures reported to date. Although they do not have any unusual torsion angles, the helical axes of these structures are deflected by 30° and 20° in the direction of the major groove. It is difficult to avoid the conclusion that this is at least partially due to crystal packing interactions, since in the octamer d(GTACGTAC), whose sequence is identical to the inner 8 base pairs of both dodecamers, the fifth residue has the *all-trans* backbone conformation and is deflected by 15° in the opposite direction (Takusagawa, 1990). It is clear that this sequence has substantial conformational flexibility, since the helical axis varies by up to 45° in the different crystal structures. It seems that part of this flexibility may be due to the strictly alternating purine-pyrimidine sequence, which leads to the very unusual base stacking arrangement shown in figure 5, in which the bases are arranged in vertical blocks with only hydrogen bonding connecting them. Further studies will identify the basis of this conformational flexibility.

The structures are also directly relevant to the ongoing attempt to map the variation possible in A-form helices. Although both dodecamers have limiting major groove dimensions quite similar to fiber diffraction A-DNA; 2.9Å for 5'-GC, 2.7Å for 5'-CC and 2.4Å for fiber diffraction A-DNA, the width of the major groove expands quite rapidly, being nearly 2Å wider at the next position. This rapid variation is consistent with the fact that the helix axis of the dodecamers is deflected at the center of the molecule. This deflection is immediately opposite the phosphate groups forming the limiting aperture of the major groove. As one moves radially around the molecule, it is expected that the width of the major groove should increase. With such wide swings in the width of the major groove, where the dodecamer is placed within the highly variable ensemble of A-DNA major groove dimensions depends on precisely where the measurement is taken.

ACKNOWLEDGEMENTS

We thank the National Institutes of Health for grant GM-17378 in support of this research.

REFERENCES

Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography*. North Holland Publ., Amsterdam.
Arnott, S. & Hukins, D. W. L. (1972) *Biochemical and Biophysical Research Communications*, **48**, 1392−1398
Bingman, C. A., Zon, G. & Sundaralingam, M. (1992a) *J. Mol. Biol.*, **227**, 738−756.
Bingman, C., Li, X., Zon, G. & Sundaralingam, M. (1992b) *Biochemistry*, in press.
Brennan, R. G. & Sundaralingam, S. (1985) *J. Mol. Biol.*, **181**, 561−563.
Brennan, R. G., Westhof, E. & Sundaralingam, M. (1986) *J. Biomol. Struct. Dynam.*, **3**, 649−665.
Brünger, A. (1988) *X-PLOR Manual, Version 1.5*, Yale University, New Haven, USA.
Chandrasekaran, R., Wang, M., He, R.-G., Pugianer, L. C., Byler, M. A., Millane, R. P. & Arnott, S. (1989) *Journal of Biomolecular Structure and Dynamics*, **6**, 1189−1202.
EMBO Workshop (1989) *EMBO J.*, **8**, 1−4.
Fairall, L., Martin, S. & Rhodes, D. (1989) *The EMBO Journal*, **8**, 1809−1817
Fitzgerald, P.M.D. (1988) *J. Appl. Crystallogr.* **21**, 273−278.
Heinemann, U., Lauble, H., Frank, R. & Blocker, H. (1987) *Nucleic Acids Research* **15**, 9531−9550.
Hendrickson, W. A. & Konnert, J. (1981). In *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., ed.), pp. 43−57, Pergammon Press, Oxford.
Jain, S. & Sundaralingam, M. (1987) *J. Mol. Biol.* **197**, 141−145
Jain, S., Zon, G. & Sundaralingam, M. (1989) *Biochemistry* **28**, 2360−2364
Jain, S. & Sundaralingam, M. (1989) *J. Biol. Chem.* **264**, 12780−12784
Otwinowski, Z. Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R., Luisi, B. F. & Sigler, P. B. (1988) *Nature* **335**, 321−9.
Quadrifoglio, F., Manzini, G. & Yathindra, N. (1984) *J. Mol. Biol.* **175**, 419−423
Takusagawa, F. (1990). *J. Biol. Struct. Dynam.*, **7**, 795−809.
Westhof, E. Dumas, P. & Moras, D. (1985). *J. Mol. Biol.*, **184**, 119−145.
Zon, G. & Thompson, J. A. (1986) *Biochromatography*, **1**, 22−31.