The structure and hydration of the A-DNA fragment d(GGGTACCC) at room temperature and low temperature

M.Eisenstein, F.Frolow, Z.Shakked and D.Rabinovich
Department of Structural Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

Received March 14, 1990; Revised and Accepted May 11, 1990

ABSTRACT

The DNA fragment d(GGGTACCC) was crystallized as an A-DNA duplex in the hexagonal space group $P6_1$. The structure was analyzed at room temperature and low temperature (100K) at a resolution of 2.5 Å. The helical conformations at the two temperatures are similar but the low-temperature structure is more economically hydrated than the room-temperature one. The structure of d(GGGTACCC) is compared to those of d(GGGTGCCC) and d(GGGCGCCC). This series of molecules, which consists of a mismatched duplex and its two Watson-Crick analogues, exhibits three conformational variants of the A-form of DNA, which are correlated with the specific intermolecular interactions observed in the various crystals. The largest differences in local conformation are displayed by the stacking geometries of the central pyrimidine-purine and the flanking purine-pyrimidine sites in each of the three duplexes. Stacking energy calculations performed on the crystal structures show that the mismatched duplex is destabilized with respect to each of the error-free duplexes, in accordance with helix-coil transition measurements.

INTRODUCTION

The crystal structure analysis of d(GGGTACCC) was carried out as a part of our studies on the mismatch-containing duplex d(GGGTGCCC) (I) and its two Watson-Crick analogues, d(GGGCGCCC) (II) and d(GGGTACCC) (III). The structures of (I) and (II) were published (1–4).

(I) was crystallized in the hexagonal space group $P6_1$, whereas (II) was crystallized in two forms: tetragonal ($P4_32_12$) and hexagonal ($P6_1$). All duplexes adopt the A-DNA-type conformation. The two G·T base-pairs of (I) displayed the wobble geometry in agreement with other crystal structures and solution studies (1). Each of the two crystal forms of (II) was studied at both room temperature (RT) and low temperature (LT). While the RT and LT structures of the tetragonal form were shown to be similar (2), a major conformational transition occurred on cooling the hexagonal crystals (3,4). The various structures of (II) revealed a range of helical conformations for the same base sequence (3,4), demonstrating that the DNA helix is highly flexible and should be analyzed in the context of its environment.

d(GGGTACCC) (III) has been crystallized as an A-DNA helix in the hexagonal form ($P6_1$). Here we present the RT and LT crystal structures of (III) and compare them to those of (I), (II) and other A-DNA structures in terms of global and local conformations and their relation to the crystalline environment.

EXPERIMENTAL

The deoxyoligonucleotide (III) was prepared by the solid-phase procedure using fast phosphoramidite chemistry (5) scaled up for the production of milligram quantities (6) and purified on a Sephadex G-50 column. Crystals of (III) were grown from a solution containing 1.5mM DNA, 5mM $MgCl_2$, 0.1mM spermine·4HCl, 10% MPD and cacodylate buffer (*p*H 7.0). Initially, tiny crystals formed; these were dissolved in buffered solutions and cooled to 4°C. The dissolved crystallites acted as seeds for the growth of larger crystals, suitable for X-ray analysis.

The crystals of (III) are hexagonal, space group $P6_1$. Cell dimensions at RT and LT are $a=46.80(0.02)$, $c=44.52(0.04)$ Å, and $a=46.06(0.01)$, $c=44.09(0.03)$ Å, respectively. The LT cell volume is smaller than the RT value (80865.6 *versus* 84445.7 Å^3) by 4%. A similar small shrinkage was noticed for tetragonal-II (2).

RT data were measured on a CAD4 diffractometer with a sealed-tube source to a resolution limit of 2.5 Å, yielding 1818 reflections of which 1582 were observed $\{F_o > 2\sigma(F_o)\}$. For these measurements the crystal was sealed in a glass capillary containing a droplet of mother liquor.

The specimen chosen for cooling was immersed in a droplet of glue (epoxy-resin part of the Master Ment Epoxy glue, Duro TM, Stock No. QM-50, Loctite Corporation, USA), mounted on a glass fiber and shock-cooled to 100K on a Rigaku AFC5 diffractometer equipped with a rotating anode source. The measured mosaic spread at low temperature was similar to the one at room temperature, 0.8°. Data were collected to a nominal resolution of 2.1 Å, yielding 3403 reflections (2008 observed ones). However, the effective resolution of the LT data was ~2.5 Å, because in the last shell $2.5 < d < 2.1$ only 35% of the reflections were observed (Table 1). Thus, no significant improvement in resolution was achieved by applying this cryotechnique to the crystal of (III), as noted before for the tetragonal form of (II) (2). The coordinates for both structures will be deposited in the Brookhaven data bank.

STRUCTURE SOLUTION AND REFINEMENT

The RT structure of (III) was solved by the multi-dimensional search method ULTIMA (7) using a fiber-derived A-type helix of the same sequence. The refinement proceeded in steps using the restrained least-squares program of Hendrickson and Konnert (8) modified for nucleic acids (9). First, only low-order reflections were included in the refinement, then the resolution limits were gradually extended as summarized in Table 2. Refining individual *B* parameters in the last step lowered the *R*-factor to 0.185.

Solvent peaks were located from F_o-F_c difference electron density maps by an automated search procedure and checked on an Evans and Sutherland PS390 computer-graphics system using program FRODO (10). Peaks which fulfilled hydrogen-bonding geometrical criteria were treated as oxygen atoms and were included in the refinement together with the DNA duplex. This procedure was repeated several times until the F_o-F_c map showed only a low, uniformly scattered noise. The scale factor was refined by a trial and error procedure. Thus, several refinement cycles were run to achieve convergence at different values of the scale factor, seeking the lowest *R*-factor. The RT structure refined to $R=0.119$, including 89 ordered water molecules.

The refinement of the LT structure of (III) started from its RT coordinates and converged to $R=0.219$ for 1502 observed reflections in the range $2.5<d<8.0$ Å. At this stage solvent molecules were introduced in the same manner as for the RT structure and the resolution was gradually increased to 2.1 Å. The LT structure refined to $R=0.16$ and contained 76 ordered water molecules.

Details of the geometrical restraints and deviations from ideal stereochemistry in the last RT and LT refinement cycles are listed in Table 3. The backbone torsion angles were not restrained. In the RT structure they displayed typical values for A-type duplexes (see below), whereas in the LT structure α and γ angles for residues G2 and A13 were close to 120°, an unfavorable eclipsed conformation. Restraining these angles to either *gauche*$^-$,*gauche*$^+$ or *trans*,*trans* changed them in the expected direction but did not produce a significant difference in the *R*-factor, which would allow discrimination between the various conformations. To resolve this problem, the LT structure was subjected to a molecular dynamics treatment using the program X-PLOR (11). It was heated to 3000K and slow cooled to room temperature in steps of 25°. This treatment resulted in α and γ torsion angles of the *trans*,*trans* conformation for G2 and the *gauche*$^-$,*gauche*$^+$ one for A13. These were the only significant changes in the structure of the DNA duplex. The root-mean-square (r.m.s.) deviation between the two sets of coordinates (before and after X-PLOR) was only 0.13 Å and the several least-squares refinement cycles preformed after X-PLOR did not alter the *R*-factor or the conformation. The electron density maps at the two residues were of similar quality for the various refined conformations. Therefore, we do not rule out conformational disorder of these nucleotides in the LT structure as also indicated by its elevated thermal parameters discussed below.

The *R*-factor for the LT data with $8.0>d>2.5$ Å was higher than that for the RT ones (0.143 *versus* 0.119), the number of ordered water molecules in the LT structure was smaller than in the RT one and the relative error in the final LT difference density map $\{(F_o\text{-}F_c)_{max}/(F_o)_{max}\}$ was larger than the corresponding RT value (0.09 *versus* 0.07), suggesting that the LT data were less accurate than the RT ones.

Table 1. Analysis of the number of reflections in small intervals in *d*

d_{min}(Å)	RT*			LT		
	N_{theory}	$N_{observed}$	%observed	N_{theory}	$N_{observed}$	%observed
5.0	255	249	98	243	233	96
3.5	480	459	96	450	429	95
2.8	662	555	84	659	568	86
2.7	172	107	62	149	104	70
2.6	175	99	57	169	110	65
2.5	220	112	51	215	117	54
2.1				1266	447	35

* RT data collection stopped at $d_{min}=2.5$ Å.

Table 2. Steps in the RT least-squares refinement*

Resolution limits (Å)	No. of reflections	R-factor†	R.M.S deviation from molecular symmetry (Å)
15-8	51	0.143	0.10
12-7	73	0.168	0.15
12-6	125	0.203	0.21
12-5	230	0.253	0.33
10-5	214	0.235	0.34
10-4	434	0.252	0.47
8-4	408	0.163	0.51
8-3	1021	0.180	0.60
8-2.5	1545	0.199	0.62

* Note the increasing deviation from molecular 2-fold symmetry on extending the resolution.
† $R=\Sigma|F_o\text{-}F_c| / \Sigma F_o$ where F_o and F_c are the observed and calculated structure factors.

Table 3. R.M.S. deviations from ideal stereochemistry in the last refinement cycle*

Geometrical restraint	RT	LT
Sugar-base bond distances (Å)	0.007/0.020	0.009/0.020
Sugar-base bond angles (Å)	0.018/0.030	0.027/0.040
Phosphate bond distances (Å)	0.025/0.030	0.021/0.030
Phosphate angles and H-bonds (Å)	0.037/0.050	0.045/0.050
Bases planarity (Å)	0.006/0.020	0.008/0.020
Sugars chiral volumes (Å^3)	0.012/0.040	0.008/0.020
Single-torsion contacts (Å)	0.062/0.100	0.075/0.100
Multiple-torsion contacts (Å)	0.064/0.100	0.083/0.100
B-parameter differences (Å^2) for		
sugar-base bonds	0.35 /1.50	0.45 /1.50
sugar-base angles	0.52 /2.00	0.68 /2.00
phosphate bonds	0.54 /1.50	0.59 /1.50
phosphate angles and H-bonds	0.54 /2.00	0.61 /2.00

* The number on the left is the r.m.s. deviation from ideality and that on the right is the s value used in the refinement ($1/s^2$ is the weight given to the corresponding restraint).

RESULTS AND DISCUSSION

Molecular packing

The packing motif of (III) (RT and LT) is similar to that of other A-DNA structures. Thus, each terminal base pair stacks at the shallow minor groove of a neighboring molecule related by either a 6-fold (6_1) or a 2-fold screw axis (2_1). Base pair C8·G9 is in van del Waals contact with a 6_1-related duplex. It is also involved in two direct intermolecular hydrogen bonds: O4′(G9)···N2(G2), N3(G9)···N2(G3), and three water-mediated interactions: N2(G9)··· W···O2(C14), O5′(G9)···W···O4′(G3) and O2(C8)···W···O2(T4) (Fig. 1;

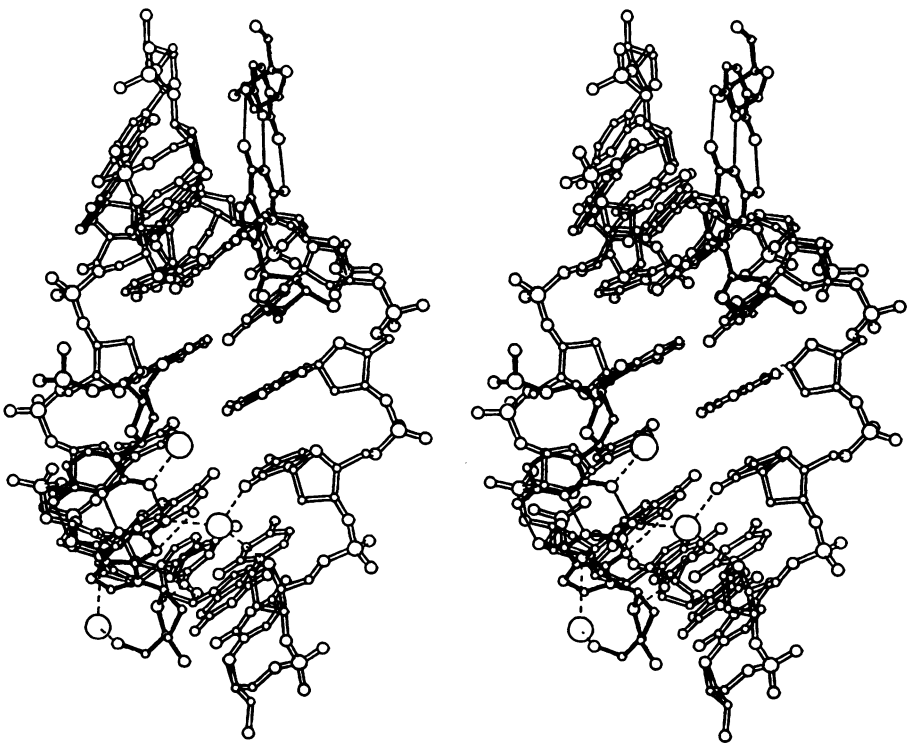

Fig. 1. Intermolecular interactions of base pair G1·C16 from the 2_1-related duplex and base pair C8·G9 from the 6_1-related neighbor.

Table 4). The direct hydrogen bonds and the first water-mediated interaction occur also in d(GGGTGCCC) (I), in the RT and LT hexagonal forms of d(GGGCGCCC) (II) (12) and in d(GGGATCCC) (13). For the last structure only direct interactions can be compared (note the interchanged strand notation compared to I,II,III).

Base-pair G1·C16 makes only van der Waals interactions with the 2_1-related molecule. However, O5′(G1) interacts with two other duplexes (Table 4): It binds to O3′(C8) of a 3_1-related neighbor through a water molecule as also observed in (I) (unpublished) and LT hexagonal-(II) (12), and to O1P(C16) of another DNA molecule, related to the first one by a second 6_1 axis. The latter interaction is mediated through two water molecules in (I) and (III), whereas in LT hexagonal-(II) and in d(GGGGCTCC) (14) direct hydrogen bonding between the duplexes is observed.

The hexagonal A-DNA structures investigated so far can be divided into three groups according to their inter-duplex interactions: The first group includes the RT and LT structures of (III), RT hexagonal-(II) and d(GGGATCCC) (13), which form hydrogen bonds mainly between 6_1-related duplexes. (I) and LT hexagonal-(II) (second group) form intermolecular hydrogen bonds through both 6_1 and 2_1 operations, and in the third group, which consists of d(GGGGCTCC) (14), d(GGTATACC) (15), d(GGGGCCCC) (16) and d(GGGGTCCC) (17), direct hydrogen bonds occur mainly between 2_1-related molecules. One notes that hydrogen bonds in the 2_1 interaction exist only in structures with a relatively short *c* axis. The values for groups 2 and 3 are 41.07–42.97 Å compared to 44.09–44.54 Å for group 1. The individual cell parameters of all members of group 1 are almost identical.

Table 4. Intermolecular hydrogen bonds

			Distances (Å)		Symmetry operation
			RT	LT	
O4′(G9)	–	N2 (G2)	2.6	2.6	$x-y,x,1/6+z$
N3 (G9)	–	N2 (G3)	3.1	3.1	$x-y,x,1/6+z$
N2 (G9)	– W –	O2 (C14)	2.6,3.0	2.7,3.1	$x-y,x,1/6+z$
O5′(G9)	– W –	4′ (G3)	3.3,3.3	3.3,3.3	$x-y,x,1/6+z$
O2 (C8)	– W –	O2 (T4)	2.8,2.8	3.3,2.5	$x-y,x,1/6+z$
O5′(G1)	– W –	O3′ (C8)	3.1,2.5	3.2,2.6	$-y,1+x-y,-2/3+z$
O5′(G1)	–W–W–	O1P(C16)	2.4,2.4,3.3	2.9,2.6,3.3	$1+x-y,1+x,1/6+z$

Conformational features of the DNA duplex

The overall shape of duplex (III) at room temperature is shown in Fig. 2 and the global and local conformational parameters are listed in Tables 5–7. (III) forms a slightly underwound helix with 12 residues per turn. The same feature was shown by d(GGGATCCC) (13).

The RT and LT structures of (III) are very similar as demonstrated by their average helix parameters (Table 5) and by the small r.m.s. difference of 0.27 Å calculated by a best molecular fit procedure. Some local parameters, *i.e.* helix twist, rise and slide (Tables 6,7), are also very close in the two structures; others show similar trends along the helix. The base-pair displacements from the helix axis are systematically smaller in the LT than in the RT structure (*dx* in Table 7), suggesting a small transverse shrinkage of the duplex upon cooling. Similar effects were noted for the tetragonal structures of (II) (2), whereas a major structural change was observed when the hexagonal crystals of (II) were shock cooled to 100K (3).

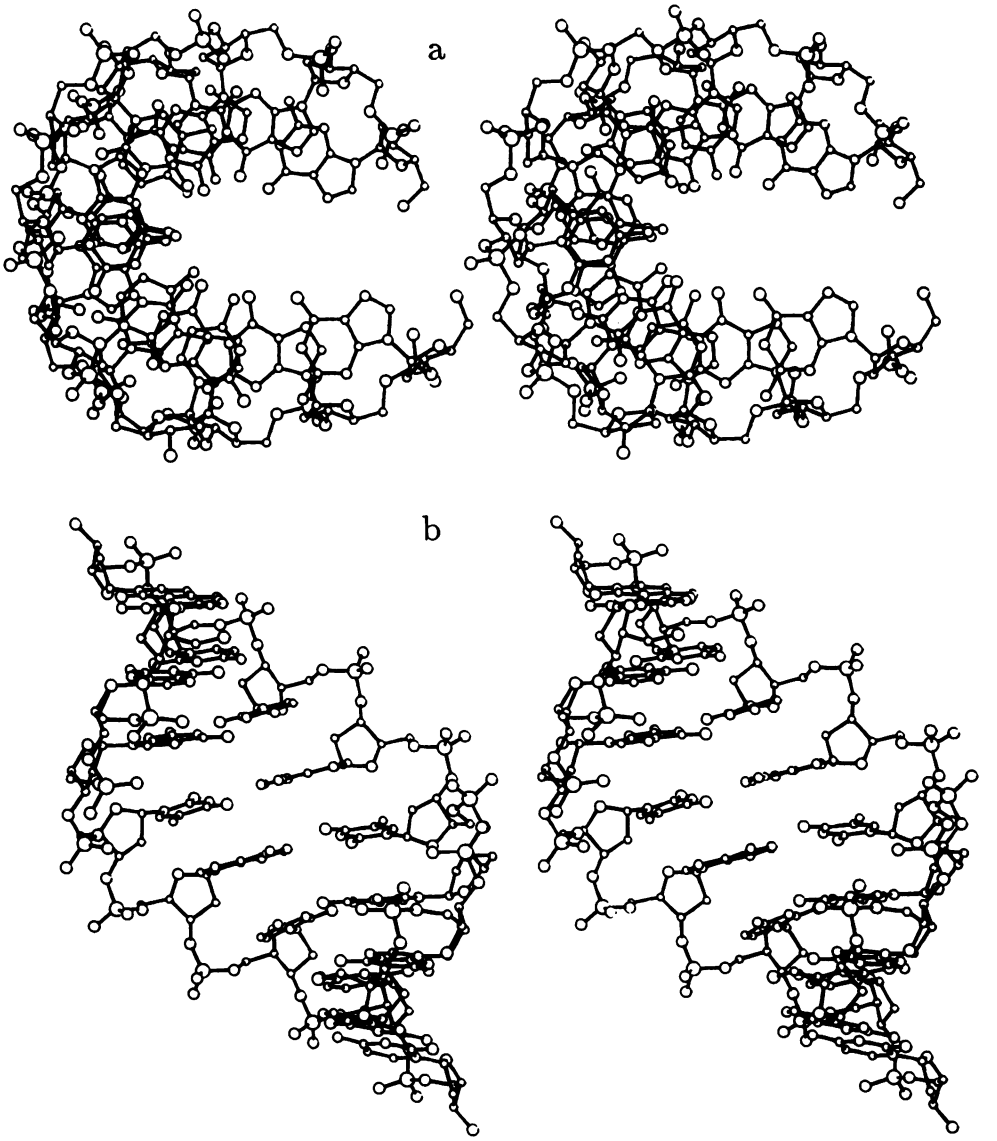

Fig. 2. Stereoscopic drawings of RT d(GGGTACCC). The numbering of the residues is G1 to C8 for the first strand and G9 to C16 for the second one. (a) View down the helix axis. G1·C16 is the uppermost base pair. (b) View into the major groove. G1·C16 is at the top.

Table 5 Average helix parameters*

	GGTGCCC		GGCGCCC†		GGTACCC	
		IIa	IIb	IIc	RT	LT
Helix twist (°)	32(6)	32(4)	31(5)	32(2)	30(2)	30(2)
Rise/base pair (Å)	2.9(0.2)	3.3(0.1)	2.9(0.3)	2.9(0.3)	2.9(0.3)	2.9(0.2)
Base-pair inclination (°)	13(2)	7(1)	13(2)	14(2)	13(2)	12(2)
Propeller twist (°)	−9(5)	−9(3)	−6(6)	−9(5)	−8(3)	−9(5)
Base-pair displacement (Å)	−3.8(0.3)	−3.8(0.2)	−4.9(0.3)	−3.5(0.2)	−4.4(0.3)	−4.1(0.4)
Base-pair slide (Å)	−1.4(0.4)	−1.6(0.3)	−1.9(0.4)	−1.2(0.4)	−1.7(0.4)	−1.7(0.4)
Major groove width (Å)	5.7	10.1	8.2	5.0	8.1	8.1
Minor groove width (Å)	10.0	9.4	9.7	10.3	9.8	9.7
Helix diameter (Å)	18.1	19.5	19.7	18.1	19.3	18.9

* Helix parameters in Tables 5−7 are as defined in references 18−20. The major groove width was estimated from the distance between the O5′ terminal phosphate groups. Estimated standard deviations are given in parentheses.

† IIa=RT-tetragonal; IIb=RT-hexagonal; IIc=LT-hexagonal. The LT parameters of tetragonal-II are very similar to the RT ones, except for base-pair displacement and helix diameter whose absolute values are smaller by 0.4 Å.

The global helix parameters of (III) are compared to those of (I) and the three conformations of (II) (a,b,c) in Table 5. The tetragonal structure of (II) is the odd-one out in this series. It is elongated, with a large rise per base pair (3.3 Å), a small base pair inclination and a widely open major groove. All these features were observed in the other A-type octamers crystallizing in the tetragonal space group (4). Within the five duplexes crystallizing in the hexagonal space group (I) and (IIc) display one conformational type whereas (IIb) and (III) display another one, which is common to other hexagonal structures (4). The two conformational variants differ in the width (average 5.3, 8.1 Å) and depth (average −3.6, −4.6 Å) of their major grooves and in the helix diameter (average 18.1, 19.5 Å). This is further established by the small r.m.s. differences between the first two structures (0.5 Å) and last two ones (0.4 Å). The r.m.s. deviations between members of different hexagonal groups are close to 1.0 Å and those between the tetragonal structure and any of the hexagonal ones is greater than 1.3 Å. The two conformational subclasses observed in the hexagonal structures are directly correlated with the ones based on intermolecular contacts discussed earlier. This demonstrates the mutual dependence between the global conformation and crystal packing interactions.

As discussed previously (3,4) it is not clear whether the helical conformation is determined by crystal-packing forces or a particular conformation adopts the appropriate crystal packing. It appears however, that water molecules play an important role in both conformational transitions (3,4) and the formation of intermolecular contacts discussed in the previous section.

Local parameters of the RT and LT structures of (III) are given in Tables 6 and 7. Significant deviations from molecular 2-fold symmetry are shown by several parameters (Tables 6,7). The r.m.s. differences between the two strands (0.54 and 0.65 Å for the RT and LT structures respectively) are larger than the r.m.s. variation between the two duplexes. Some of these differences appear to result from intermolecular interactions. For example, the large slide of the G1-G2 step, which interacts with a 2_1-related duplex, is also observed in (IIb) (3) and d(GGGATCCC) (13), which are involved in similar intermolecular contacts to (III) (see above).

The helix twist of the T-A step in (III) is small compared to the other ones (Table 6). This feature is common to T-A sites in all A-DNA structures (21−23), and so is the secondary effect of increased twist for the adjacent steps. Large roll angles (average 13.1°) distinguish T-A steps in the two hexagonal structures, (III) and d(GGTATACC) (21), from central T-A steps in the tetragonal ones, d(GTGTACAC) (22) and d(CTCTAGAG)

Table 6. Local stacking parameters

Helical step	Twist(°) RT	Twist(°) LT	Roll(°) RT	Roll(°) LT	Rise(Å) RT	Rise(Å) LT	Slide(Å) RT	Slide(Å) LT
G-G/C-C	28.0	27.6	2.2	0.1	3.1	3.1	−2.2	−2.0
G-G/C-C	31.7	29.2	7.9	7.9	2.6	2.7	−2.0	−2.1
G-T/A-C	30.6	32.5	2.7	1.6	3.0	3.0	−1.3	−1.3
T-A/T-A	27.5	27.4	14.1	7.8	3.3	3.0	−1.3	−1.3
A-C/G-T	33.3	35.0	−3.8	2.5	2.8	2.8	−1.7	−1.5
C-C/G-G	32.0	31.2	12.9	13.3	2.6	2.7	−2.1	−2.3
C-C/G-G	30.3	29.6	11.6	7.4	3.2	3.1	−1.7	−1.5
Average	30.5	30.4	6.8	5.8	2.9	2.9	−1.7	−1.7
s.d.	2.1	2.8	6.6	4.6	0.3	0.2	0.4	0.4

Table 7. Local base-pair properties

Base-pair	Propeller twist(°) RT	Propeller twist(°) LT	Buckle(°) RT	Buckle(°) LT	*dx* (Å) RT	*dx* (Å) LT
G1 · C16	−3.9	−1.2	7.6	12.9	−4.8	−4.6
G2 · C15	−4.6	−9.7	4.3	4.0	−4.2	−3.8
G3 · C14	−9.3	−13.3	8.3	7.3	−4.3	−4.0
T4 · A13	−9.1	−7.8	−2.1	−5.6	−4.8	−4.6
A5 · T12	−9.0	−14.2	−8.8	−11.4	−4.9	−4.6
C6 · G11	−10.5	−13.4	−17.6	−15.8	−4.0	−3.6
C7 · G10	−11.0	−9.4	−7.0	−2.0	−4.3	−4.0
C8 · G9	−4.1	−0.6	−6.7	−5.4	−4.3	−4.0
Average	−7.7	−8.7	−2.7	−2.0	−4.4	−4.1
s.d.	3.0	5.3	9.1	9.6	0.3	0.4

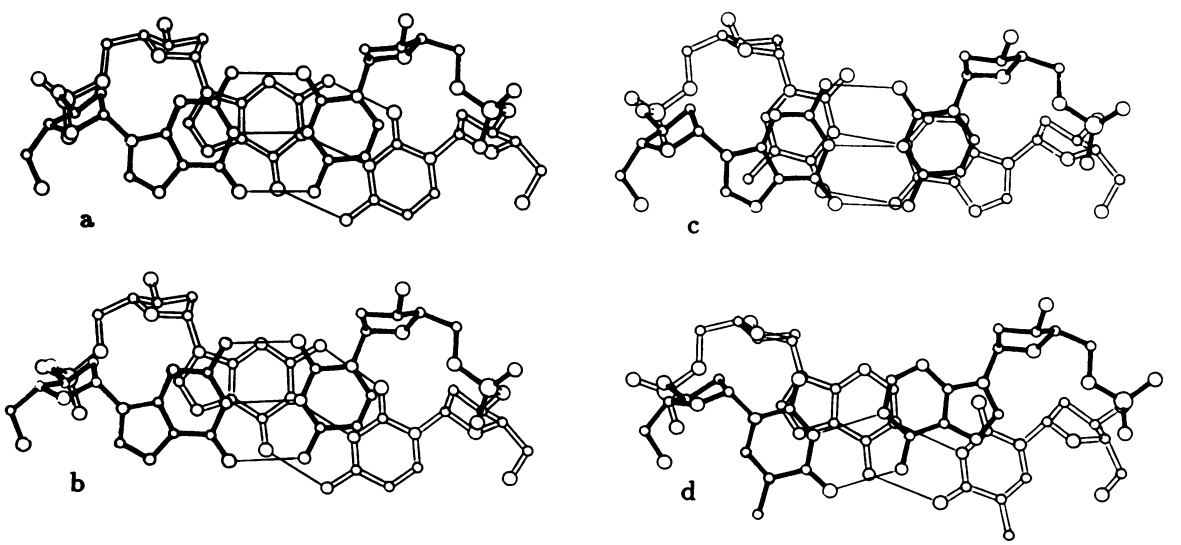

Fig 3. Base-stacking diagrams for four steps in RT d(GGGTACCC). Each view is perpendicular to the mean plane of the upper base pair shown with dark bonds. The 5′ to 3′ direction is downwards for the right strand and upwards for the left strand.
(a) G1-G2/C15-C16; (b) G2-G3/C14-C15; (c) G3-T4/A13-C14; (d) T4-A5/T12-A13.

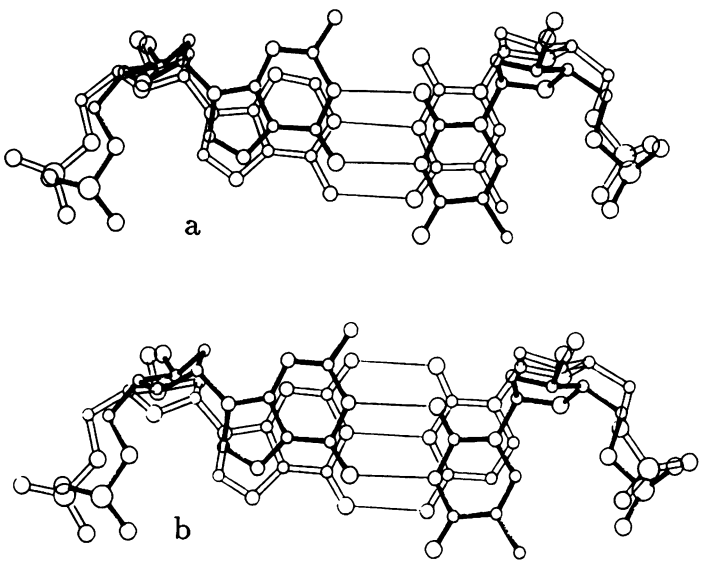

Fig. 4. Superposition of the mismatched base pair T4·G13 of (I) with its two Watson-Crick analogues: T4·A13 of (III) and C4·G13 of (IIb) designated by a and b respectively. These were obtained by best molecular fits of (I) and (III) or (I) and (IIb) applying zero weights to the two central base pairs of each duplex. A similar effect was observed in the comparison of (I) and (IIc).

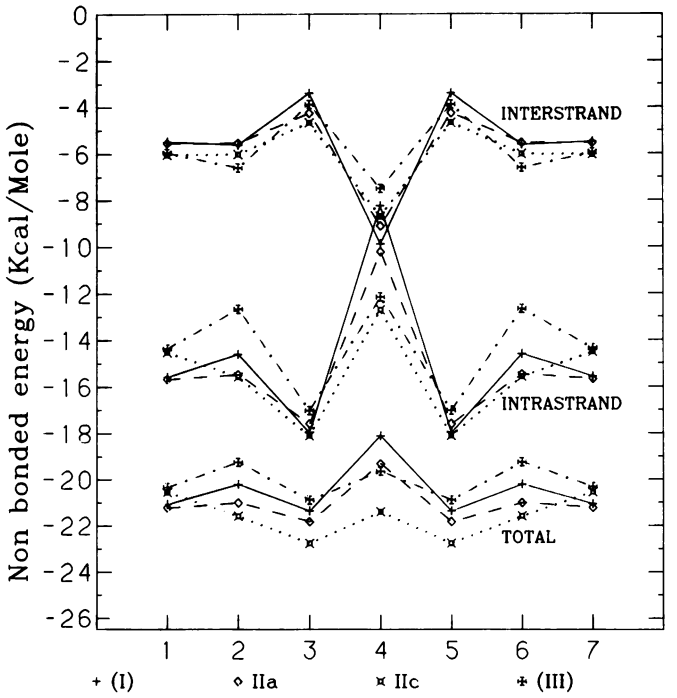

Fig. 5. Base-stacking energies for RT d(GGGTACCC) (III), d(GGGTGCCC) (I) and two variants of d(GGGCGCCC) (IIa and IIc). The corresponding diagram for LT d(GGGTACCC) is very similar to the RT one.

(23), which are characterized by small roll angles (6 and 5.6°).

The base-stacking pattern for the T-A step (Fig. 3) is similar to the corresponding ones in d(GGTATACC) but differs significantly from those in d(CTCTAGAG) (23) and in d(GTGTACAC) in its two crystal forms (24). This is due to the large relative sliding of the central T·A base pairs in the latter

Table 8. Backbone parameters*

RT	α	β	γ	δ	ε	ζ	χ	P	P_i-P_{i+1}
G1	–	–	82	79	225	281	183	14	–
G2	294	161	63	71	207	271	190	24	6.0
G3	315	166	58	80	208	287	190	12	6.0
T4	280	191	66	83	207	280	207	17	5.8
A5	298	163	49	78	199	281	197	21	5.9
C6	310	159	51	77	195	291	204	19	6.1
C7	283	182	57	80	211	281	202	17	5.8
C8	271	179	76	79	–	–	203	37	–
G9	–	–	72	79	235	265	185	6	–
G10	306	156	57	78	221	278	182	16	5.8
G11	292	163	73	75	205	288	175	30	6.4
T12	300	199	43	89	203	302	207	5	5.4
A13	258	166	87	81	215	279	199	28	6.4
C14	310	158	51	73	196	295	200	25	6.0
C15	271	183	74	82	200	283	196	29	6.2
C16	282	185	67	84	–	–	199	35	–
Av.	291	172	64	79	209	283	195	21	6.0
LT	**α**	**β**	**γ**	**δ**	**ε**	**ζ**	**χ**	***P***	**P_i-P_{i+1}**
G1	–	–	67	73	204	282	185	18	–
G2	184	203	145	67	214	281	183	14	6.5
G3	301	170	65	74	205	291	194	13	5.8
T4	277	199	58	82	199	282	211	9	5.7
A5	308	163	43	76	211	280	200	8	5.8
C6	307	156	41	81	204	281	200	11	6.0
C7	306	171	39	80	212	286	197	14	5.5
C8	295	171	54	85	–	–	209	25	–
G9	–	–	75	77	209	283	193	13	–
G10	271	170	80	80	240	261	180	13	5.8
G11	314	146	67	70	193	296	174	40	6.3
T12	280	215	57	86	183	309	211	14	5.5
A13	252	173	83	79	216	273	196	26	6.4
C14	316	152	44	74	211	290	201	24	5.9
C15	278	178	74	79	221	270	188	31	6.2
C16	293	167	53	76	–	–	197	29	–
Av.	284	174	65	77	209	283	195	19	6.0

*Torsion angles(°) defined as $P^{\alpha}O5'^{\beta}C5'^{\gamma}C4'^{\delta}C3'^{\epsilon}O3'^{\zeta}P$. Glycosyl torsion angle(°), χ, defined by O4'-C1'-N1-C2 or O4'-C1'-N9-C4. *P* is the pseudorotation parameter (27) and P_i-P_{i+1} indicates phosphate distances (Å).

structures, which leads to an extensive overlapping of the adenine bases. Similar variations in stacking were observed for the C-G sites in the several crystal forms of (II) and were attributed to the local water structure (3,4). It may well be that the variations in the stacking geometry of T-A doublets contained in different sequences, are also influenced by hydration in addition to the effects caused by the flanking base pairs.

The G-T steps in (III) are characterized by a very small roll, meaning that the base pairs involved are almost parallel to one another. This facilitates a large intrastrand overlap, with the six-membered ring of the purine base stacking on the pyrimidine base (Fig. 3). A very similar pattern was observed for G-T steps in d(GGTATACC) (21) and in the mismatched duplex d(GGGTGCCC) (I). Considerable intrastrand stacking characterizes all other A-type purine-pyrimidine steps: G-C (4,25), A-T (13) and A-U (26). The stacking parameters of the homopolymer steps of (III) are within the range observed in other A-DNA structures.

By comparing the local parameters of (III) to those of (I) and the three variants of (II), one notes major changes in the stacking geometries of the central pyrimidine-purine step and the flanking purine-pyrimidine sites when going from one duplex to the next. The common feature observed in the Watson-Crick helices is the relatively small helix twist at C-G or T-A sites (24 to 29°) and

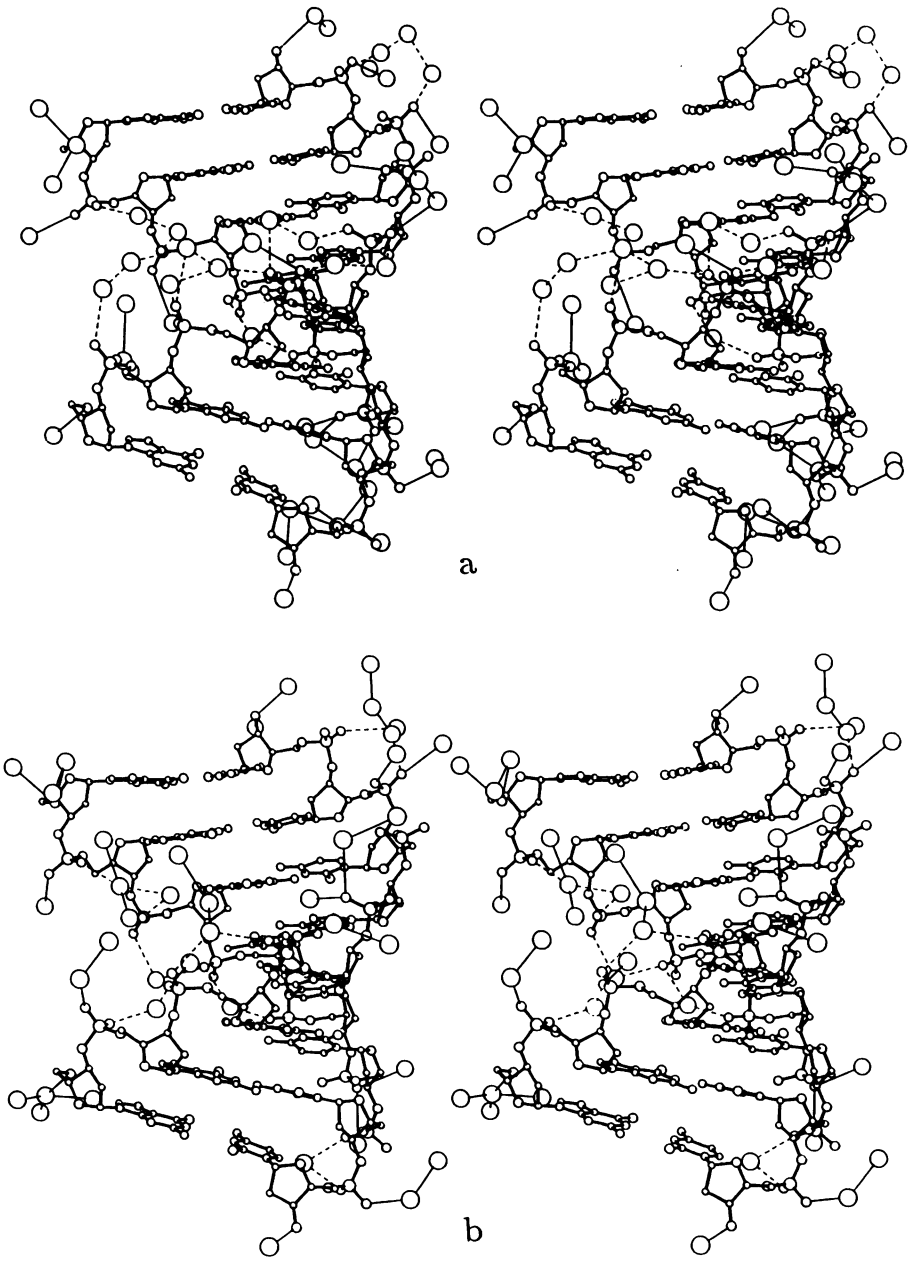

Fig. 6. Hydration of the phosphate groups. Dashed lines denote phosphate-phosphate solvent bridges. (a) RT structure; (b) LT structure.

larger than average values in the adjacent steps (33 to 36°). A reversed pattern is shown by the mismatched duplex: a very large helix twist at the central site (44°) and a significantly smaller than average values in the neighboring steps (26°). These distinctive features arise from the opposite displacements of the guanine and thymine bases of the mismatched pairs with respect to the equivalent bases in the error-free duplexes. Figure 4 presents a superposition of the wobble base pair T4·G13 of (I) and each of the two Watson-Crick base pairs, T4·A13 of (III) and C4·G13 of (IIb). In each of these figures, the thymine is displaced into the major groove and the guanine into the minor groove, in a direction perpendicular to the C1′-C1′ vector of the normal base pair (T·A or C·G). The same pattern was noticed before by comparing the central base pairs of (I) with those of (IIa), and the changes in local twist angles were rationalized on the basis of geometrical considerations (1).

The base stacking energy diagrams of (I), (III) and two structural variants of (II) are shown in Figure 5. The force field and the method for the energy calculations are described in references 1 and 30. The major contribution to the stabilization of homopolymer and purine-pyrimidine doublets comes from intrastrand stacking interactions, in agreement with the qualitative description of the stacking geometry. However, in the pyrimidine-purine sites the energy pattern is more complex and cannot always be anticipated from qualitative considerations. Whereas in the Watson-Crick sites the energy level of the intrastrand component is systematically lower than the interstrand one, a reversed pattern is shown by the mismatched doublet, resulting in a destabilization

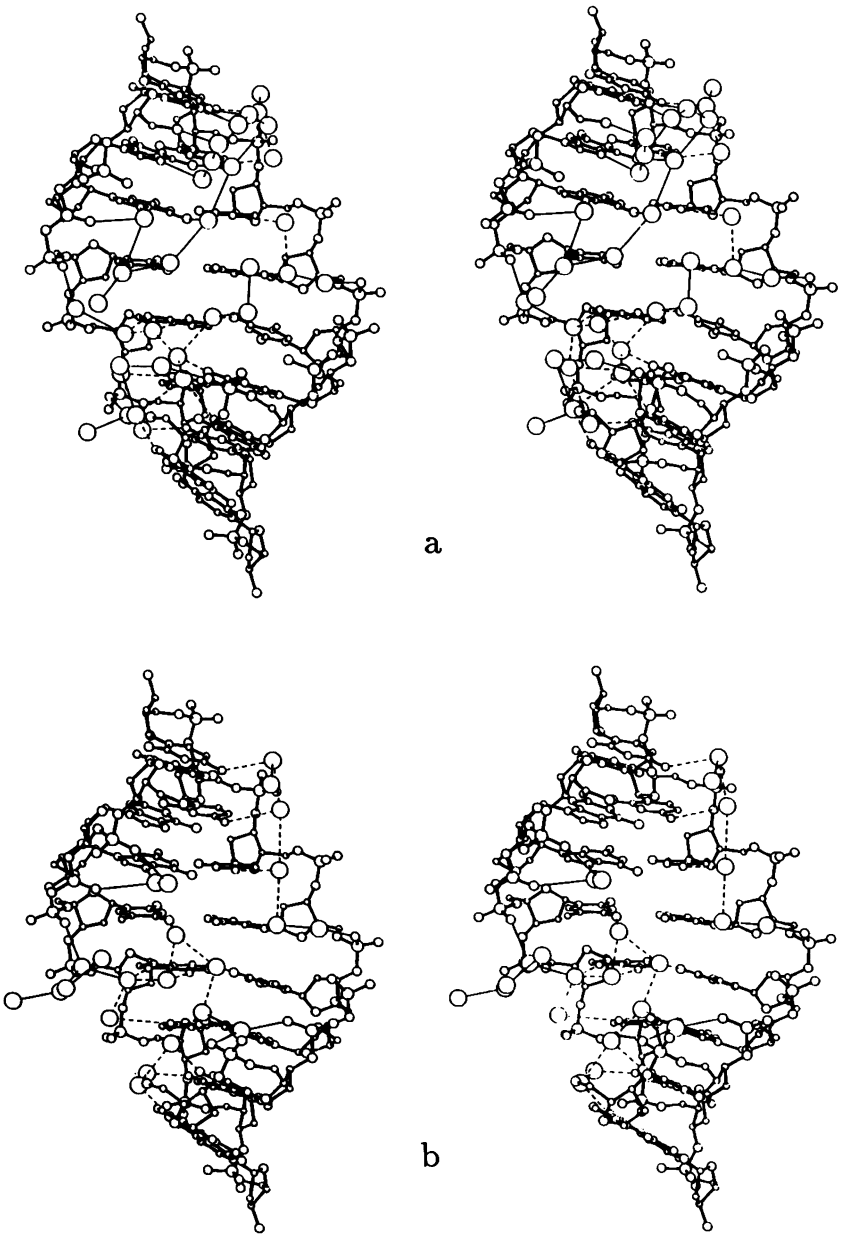

Fig. 7. Major groove hydration. Dashed lines mark hydration chains which are common to the RT and LT structures. Some of the common water molecules are connected in one structure and not in the other due to slight positional changes resulting in interatomic distances exceeding 3.5 Å. Links to the phosphates are also indicated. (a) RT structure; (b) LT structure.

with respect to T-A or C-G doublets. This finding is in agreement with our measurements of helix-coil transitions of (I), (II) and (III) in solution (1).

Backbone parameters are listed in Table 8. The average torsion angles in the RT and LT structures of (III) are very similar. However, individual values differ considerably. In particular G2 adopts the usual *gauche*$^-$,*gauche*$^+$ conformation around the P-O5′ and C5′-C4′ bonds in the RT structure and a *trans,trans* conformation in the LT one. The same nucleotide displayed a semi-extended conformation in (I) and an extended one in d(GGGATCCC) (13). Such changes are probably related to the different hydration schemes in the two crystals.

Residues A5 and A13 of the central T-A step in (III) display the usual backbone conformation, unlike the equivalent ones in d(GTGTACAC) (22) and d(CTCTAGAG) (23) where α, β and γ are all *trans*. This probably arises from the different crystalline and water environments. Extended backbone conformations occur in the central steps of all the A-octamers which crystallize in the tetragonal space group (22,23,25).

The α and γ angles in (III) are anticorrelated, as in other A-DNA structures (25). The conformation of the deoxyribose rings is C3′-*endo*, as indicated by the pseudorotation values (range 5−40) and δ angles (range 67−89°).

Hydration

Only a fraction of the solvent molecules in the crystal can be located in an X-ray study. These are mainly the waters belonging to the first and second hydration shells. The first shell comprises

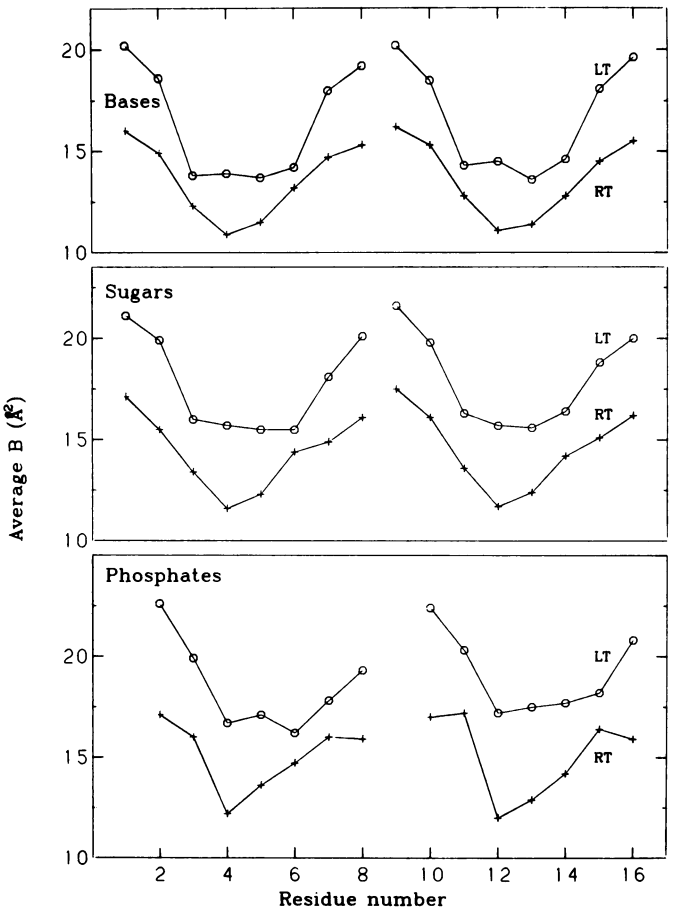

Fig. 8. Averaged *B* values for bases, sugars and phosphates.

55 of the 89 ordered water molecules in the RT structure and 42 of the 76 ones in the LT structure.

Figs. 6,7 show details of the water arrangement around the RT and LT duplexes. All the phosphate groups except C14 of LT are hydrated and several are linked to each other by water bridges (seven bridges for each duplex; Fig. 6). In the RT structure there are, on the average, 2.1 first-shell waters per phosphate (excluding O5′ and O3′ at the ends of each strand) and 2.1 waters per phosphate-phosphate bridge. The corresponding numbers for the LT structure are smaller: 1.7 and 1.3 molecules. The phosphates of residues 11–13 in the RT duplex are bridged twice, through their O1P and O2P oxygens.

In both structures the major-groove hydration sites are linked by chains of water molecules (Fig. 7). At room temperature a chain of eight waters links residues 1–4. A second one (eight molecules) connects residues 5–7, and crosses over to the other strand linking residues 9–12. A short chain interacts with residues 13,14 and another one with residues 15,16. The two water chains in the LT structure are similar though not identical to the RT ones: a chain of nine molecules links residues 4–7 and 9–12 and another one, including four molecules, connects residues 13–16. In both structures the major groove hydration chains link to the phosphate hydration shell either directly or through additional water molecules.

Only a few solvent peaks were located at the minor grooves. Most of them are involved in inter-duplex interactions and are common to both structures.

The LT duplex is less hydrated than the RT one. In particular the difference in the hydration of the phosphate groups is significant. While the LT structure displays the economy-type hydration proposed for A-DNA (28), the RT structure does not. In a previous study we have shown that different helical conformations of the same base sequence (II) are associated with large variations in hydration (3,4). In this study we show that similar conformations of the same base sequence exhibit significant variations in the water structure. A similar trend was observed for the RT and LT structures of d(GGGCGCCC), whereas two independent refinements for LT d(GGGCGCCC) against data measured for different crystals produced very close water arrangements (to be published). Thus, changes in the water structure are brought about by the cooling process and probably, also by slight differences in the crystallization conditions.

Thermal parameters

The average *B*'s for phosphates, sugars and bases are 15.1, 14.5 and 13.7 Å^2 respectively in the RT refinement, and 18.8, 17.9 and 16.6 Å^2 in the LT one. The decrease from phosphates through sugars to bases reflects the rigid-body-motion character of the mean atomic displacements, as discussed for tetragonal-II (2). This is further established in Fig. 8, which presents the variation of average *B* parameters for bases, sugars and phosphates along the strands. In both structures the *B*'s shown by the central nucleotides are the lowest and rise toward the ends of the helix.

Contrary to the expected, the LT *B*'s are slightly higher than the RT ones. At room temperature dynamic disorder is likely to occur (29). Thus, the *B* parameters represent mean-square displacements from an average conformation. On cooling the crystal, a static disorder of the same order of magnitude as the RT dynamic one may have formed, thus leading to the high *B* values. This effect is probably a real one, but we must be careful in drawing quantitative conclusions from the *B* parameters, because of their strong correlation with the scale factor which cannot be fully resolved at 2.5Å resolution, even in a careful refinement as described above.

ACKNOWLEDGEMENTS

We thank the Minerva Foundation, Munich, Germany, and the US/Israel Binational Science Foundation for financial support. We thank Mr. E.A. Slatis for helping to prepare the figures. Z.S. holds the Helena Rubinstein Professorial chair in Structural Biology.

REFERENCES

1. Rabinovich,D., Haran,T., Eisenstein,M. and Shakked,Z. (1988) J. Mol. Biol. **200**, 151–161.
2. Eisenstein,M., Hope,H., Haran,T.E., Frolow,F., Shakked,Z. and Rabinovich,D. (1988) Acta Cryst. **B44**, 625–628.
3. Shakked,Z., Guerstein-Guzikevich,G., Eisenstein,M., Frolow,F. and Rabinovich,D. (1989) Nature **342**, 456–460
4. Shakked,Z., Guerstein-Guzikevich,G., Zaytzev,A., Eisenstein,M., Frolow,F., and Rabinovich,D. (1990) In Sarma,R.H and Sarma,M.H. (eds.), Proceedings of the Sixth Conversation in Biomolecular Stereodynamics. Adenine Press, New York, Vol. 3, DNA and RNA, pp. 55–72.
5. Beaucage,S.L. and Caruthers,M.H. (1981) Tet. Letters **22**, 1859–1862.
6. Dorman,M.A., Noble,S.A., McBride,L.J. and Caruthers,M.H. (1984) Tetrahedron **40**, 95–102.
7. Rabinovich,D. and Shakked,Z. (1984) Acta Cryst. **A40**, 195–200.

8. Hendrickson,W.A. and Konnert,J.H. (1981) In Srinivasan,R. (ed.), Biomolecular Structure, Conformation, Function and Evolution. Pergamon Press, Oxford, Vol. 1, pp. 43–47.
9. Westhof,E., Dumas,P. and Moras,D. (1985) J. Mol. Biol. **184**, 119–145.
10. Jones,T.A. (1978) J. Appl. Crystallog. **11**, 268–272.
11. Brunger,A.T., Kuriyan,J. and Karplus,M. (1987) Science **235**, 458–460.
12. Guerstein-Guzikevich,G. M.Sc. thesis (1989). The Feinberg Graduate School of the Weizmann Institute of Science.
13. Lauble,H., Frank,R., Blocker,H. and Heinemann,U. (1988) Nucleic Acid Research **16**, 7799–7816.
14. Hunter,W.N., Kneale,G., Brown,T., Rabinovich,D. and Kennard,O. (1986) **190**, 605–618.
15. Shakked,Z., Rabinovich,D., Cruse,W.B.T., Egert,E., Kennard,O., Sala,G., Salisbury,S.A. and Viswamitra,M.A. (1981) Proc. R. Soc. Lond. B **213**, 479–487.
16. McCall,M., Brown,T. and Kennard,O. (1985) J. Mol. Biol. **183**, 385–396.
17. Kneale,G., Brown,T., Kennard,O. and Rabinovich,R. (1985) J. Mol. Biol. **186**, 805–814.
18. Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) J. Biol. Chem. **257**, 14686–14707.
19. Dickerson,R.E., Kopka,M.L. and Pjura,P. (1985) In McPherson,A. and Jurnak,F. (eds.), Structures of Biological Macromolecules and Assemblies. Wiley Press, New York, Vol II: Nucleic Acids and Interactive Proteins, pp. 38–126.
20. Dickerson,R.E. (1989) The EMBO Journal **8**, 1–4.
21. Shakked,Z., Rabinovich,D., Kennard,O., Cruse,W.B.T., Salisbury,S.A.,and Viswamitra,M.A. (1983) J. Mol. Biol. **166**, 183–201.
22. Jain,S., Zon,G. and Sundaralingam,M. (1989) Biochemistry **28**, 2360–2364.
23. Hunter,W.N., D'Estaintot,B.L. and Kennard,O. (1989) Biochemistry **28**, 2444–2451.
24. Jain,S., Sundaralingam,M. (1989) J. Biol. Chem. **264**, 12780–12784.
25. Haran,T.E. and Shakked,Z. (1988) J. Mol. Struc. (Theochem) **179**, 367–391.
26. Dock-Bregeon,A.C., Chevrier,B., Podjarny,A., Johnson,J., de Bear,J.S., Gough,G.R., Gilham,P.T., and Moras,D. (1989) J. Mol. Biol. **209**, 459–474.
27. Altona,C. and Sundaralingam,M. (1972) J. Am. Chem. Soc. **94**, 8205–8212.
28. Saenger,W., Hunter,W.N. and Kennard,O. (1986) Nature **324**, 385–388.
29. Doster,W., Cusack,S. and Petry,W. (1989) Nature **337**, 754–756.
30. Haran,T.E., Berkovich-Yellin,Z. and Shakked,Z. (1984) J. Biomol. Struct. Dyn. **2**, 397–412.