Acta Cryst. (1996). D**52**, 655–667

Structure of the Purine–Pyrimidine Alternating RNA Double Helix, r(GUAUAUA)d(C), with a 3′-Terminal Deoxy Residue

MARKUS C. WAHL,[a] CHANGILL BAN,[b] CHANDRA SEKHARUDU,[b] BOOPATHY RAMAKRISHNAN[a,b]† AND MUTTAIYA SUNDARALINGAM[a,b]*

The Ohio State University, Laboratory of Biological Macromolecular Structure, Departments of [a]Biochemistry and [b]Chemistry, 1060 Carmack Road, Columbus, Ohio 43210-1002, USA. E-mail: sunda@biot.mps.ohio-state.edu

(*Received 3 October 1995; accepted 3 January 1996*)

Abstract

The crystal structure of the purine–pyrimidine alternating octameric RNA helix, r(GUAUAUA)d(C), carrying a 3′-terminal deoxycytidine residue, has been determined at 2.2 Å resolution. The molecule crystallizes in the rhombohedral space group $R3$ (hexagonal cell constants: $a = b = 43.07, c = 59.36$ Å; $\alpha = \beta = 90, \gamma = 120°$) with one duplex in an asymmetric unit. The structure was solved by molecular replacement and refined with 83 and 2/3 solvent molecules and 2/3 sodium ions to a final R factor of 15.6% using 1775 reflections (86%). The duplexes are approximately linear, their global helix axes are inclined by 10° with respect to the 3_2-screw axes, and they are stacked on top of each other in a head-to-tail fashion. The twist between the junction base pairs of the stacked duplexes is negligible resulting in a discontinuity of the helix backbones and grooves. The sodium ions on the threefold axis play a significant role in the organization of the packing network. The helical parameters, particularly the twist and the roll, of this alternating sequence are in accord with Calladine's rules. Almost all the 2′-hydroxyl groups are involved in specific hydrogen-bonding interactions, either directly to the sugar ring oxygens O4′ on the 3′ side, or, through water bridges, to the sugars, phosphates, or bases. This hydrogen bonding of the 2′-hydroxyl groups restrains the conformation of the sugar-phosphate backbone and the glycosidic torsion angles of this RNA fragment. The lack of intermolecular packing contacts in the grooves provides a clear picture of the groove solvation.

1. Introduction

There is tremendous interest in the principles which govern tertiary RNA structure to improve our understanding and modeling capabilities of large RNA folds (Michel & Westhof, 1990). Since the initial investigations on the tRNA structure (for a review see *e.g.* Sundaralingam, 1979), there was a pause in the elucidation of the tertiary structure of RNA, until the determination of the structures of the tRNA/aminoacyl-tRNA synthetase complexes (Rould, Perona, Söll & Steitz, 1989; Ruff *et al.*, 1991; Biou, Yaremchuk. Tukalo & Cusack, 1994; Arnez & Steitz, 1994), a RNA hairpin-spliceosomal protein complex (Oubridge, Ito, Evans, Teo & Nagai, 1994), a RNA–bacteriophage coat protein complex (Valegård, Murray, Stockley, Stonehouse & Liljas, 1994), and the recent work on the hammerhead ribozyme motifs (Pley, Flaherty & McKay, 1994; Scott, Finch & Klug, 1995). Only recently problems with automated RNA synthesis and product purification have been overcome which is expected to increase the number of RNA crystal structures in the same fashion as DNA (Wahl & Sundaralingam, 1995). Although several RNA duplex crystal structures are available (Dock-Bregeon *et al.*, 1988, 1989; Holbrook, Cheong, Tinoco & Kim, 1991; Leonard *et al.*, 1994; Betzel *et al.*, 1994; Cruse *et al.*, 1994; Portmann, Usman & Egli, 1995; Schindelin *et al.*, 1995; Baeyens, De Bondt & Holbrook, 1995), only one of them (Dock-Bregeon *et al.*, 1988, 1989) exhibits a long stretch of alternating sequence. The current octamer structure, r(GUAUAUA)d(C), is completely alternating, but the packing scheme is very different from the above.

2. Experimental

2.1. *Oligonucleotide synthesis and purification*

The oligonucleotide was synthesized by the solid-phase phosphoramidite protocol on an ABI nucleic acid synthesizer (Model 391). After cleavage from the synthesis column, using 3:1 ammonium hydroxide/ethanol, and deprotection of the bases in an overnight incubation at 328 K, the material was lyophilized and the 2′-hydroxyl groups deprotected at room temperature in 2 M tetrabutylammonium fluoride in THF (100 µl base^{-1}, 24 h). The sample was further purified by reverse-phase FPLC, employing a step gradient from 0 to 30% acetonitrile, and a subsequent ethanol precipitation. Aldrich chemicals were used without further purification. Residues for the duplex are numbered 1–8 in strand 1 and 9–16 in strand 2 in the 5′-to-3′ direction.

		1	2	3	4	5	6	7	8	
Strand 1	5′	rG	rU	rA	rU	rA	rU	rA	dC	3′
		|	|	|	|	|	|	|	|	
Strand 2	3′	dC	rA	rU	rA	rU	rA	rU	rG	5′
		16	15	14	13	12	11	10	9	

† Present address: Building Park Number 5, Room 410, LMMB-DCBDC-NCI-NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA.

ISSN 0907-4449

Table 1. *Refinement statistics*

Space group	*R*3 (rhombohedral)
Unit-cell dimensions (Å, °)	
Rhombohedral setting	$a = 31.78$
	$\alpha = 85.33$
Hexagonal setting	$a = b = 43.07$, $c = 59.36$
	$\alpha = \beta = 90$, $\gamma = 120$
Resolution range (Å)	10–2.2
Number of reflections	
Theoretical	2076
Observed ($\geq 2\sigma$)	1775
% Completeness	85.5
R_{sym} (%)	1.71
Asymmetric unit	1 duplex
Final *R* factor (%)	15.6
Parameter file	PARAM11.DNA
R.m.s. deviation from ideal geometry	
Bonds (Å)	0.017
Angles (°)	2.55
Final model	
Nucleic acid atoms	330
Water O atoms	83 2/3
Sodium ions	2/3

2.2. *Crystallization and data collection*

Crystals were grown by hanging-drop vapor diffusion from a solution which initially contained 2 µl oligonucleotide (2 m*M* double-stranded concentration), 0.5mL cobaltic hexamine trichloride (6 m*M*), 1.6 µl sodium cacodylate (pH 7.0, 100 m*M*), 5 µl water, and 1 µl MPD. By equilibration against 1 ml of 60% MPD, crystals formed in about one week and continued to grow for another week. A crystal measuring 0.6 × 0.2 × 0.2 mm was used in the data collection at 295 K. No crystal decay was observed after several days in the X-ray beam. Data up to 2.2 Å resolution were collected with two φ (180°) and one ω (70°) scans (0.25° frame^{-1}) on our in-house four-circle Siemens goniostat equipped with a multiwire area detector and a Macscience rotating anode, generating graphite-monochromated Cu $K\alpha$ radiation ($\lambda = 1.5418$ Å) at 40 kV and 100 mA. The crystal-to-detector distance was 12.0 cm. There were 1775 data above 2σ between 10 and 2.2 Å, constituting 86% of the theoretical value. Indexing of the reflections in the rhombohedral space group *R*3 with unit-cell dimensions $a = b = c = 31.78$ Å, $\alpha = \beta = \gamma = 85.33°$ yielded an R_{sym} of 1.71%. The rhombohedral cell was transformed to a hexagonal cell for the structure solution and refinement (Table 1). An asymmetric unit contained one RNA duplex.

2.3. *Structure solution and refinement*

The structure of the oligonucleotide was solved by molecular replacement and refined using the program package *X-PLOR* 3.0 (Brünger, 1988). The central octamer sequence r(AU)$_4$, consisting of nucleotides (A3–U10)·(A19–U26) of r[U(UA)$_6$A] (Dock-Bregeon *et al.*, 1989), was the model in a rotation–translation search using reflections in the 8–4 Å range with intensities of at least 3σ (303 reflections). A solution was obtained with an *R* factor of 48%. At this point a rigid-body refinement was performed, gradually increasing the resolution in three steps to 10–2.4 Å, followed by a positional least-squares refinement to an *R* factor of 36%. Visual inspection with the program *FRODO* (Jones, 1985) of the model in $2F_o - F_c$ and $F_o - F_c$ difference maps confirmed the correctness of the solution. The A's and U's at the ends of the duplex were replaced by G's and dC's, respectively. One round of simulated annealing followed by further positional and temperature-factor refinement dropped the *R* factor to 22.3%. Placement of 46 water molecules, represented by O atoms, produced an *R* factor of 18.0%. Manual fitting in *FRODO* using difference and difference omit-maps followed by a positional and temperature-factor refinement yielded an *R* factor of 17.8% for 1481 reflections. Upon inspection of the molecular geometry with the program *NUPARM* (Battacharyya & Bansal, 1990), three α and γ backbone bonds were found in the *trans* configuration. These were switched to the more preferred *gauche*$^-$/*gauche*$^+$ form and the structure again refined. The *trans*/*trans* configurations appeared again and were therefore maintained. Subsequently, all data above 2σ up to 2.2 Å resolution were incorporated into the refinement and an additional 41 water molecules were located, giving a total of 87 water molecules which were confirmed in $F_o - F_c$ difference omit maps. Water *B* values were restricted to approximately three times the average phosphate group *B* value. Waters which refined with higher *B* values were removed from the structure.

Four water molecules refined directly on the threefold axis. Upon inspection, one of these had an octahedral coordination with coordination distances ranging from 2.34 to 2.42 Å and a very low *B* value (4 Å^2) indicating that it could be a metal ion instead of water. Sodium and cobalt were present in the crystallization and we tried both. Sodium was selected because it was consistent with the complete lack of color of the crystals and the peak heights and coordination distances in the difference Fourier maps. A second site on the threefold axis had a tetrahedral coordination with coordination distances around 2.4 Å and a low *B* value (2 Å^2) and was, after a similar examination, also considered a sodium ion. The *B* values did not change significantly upon replacement of the O atoms with sodium ions. The coordination distances of the presumed metal ions were not restrained during the refinement because a reasonable geometry was maintained. The other two water molecules on the threefold axis were kept as O atoms because they were coordinated to one and three neighbors, respectively, at distances of around 2.8 Å. The four residues on the threefold axis were given 1/3 occupancy and all other atoms were refined with full occupancy. The final *R* factor was 15.6% using 1775 reflections $\geq 2\sigma$ between 10 and 2.2 Å resolution for an asymmetric unit containing

330 nucleic acid atoms, 83 and 2/3 water O atoms, labeled 17–101, and 2/3 sodium ions, labeled I and II. The *R* value when all 83 2/3 solvent molecules and 2/3 sodium ions were omitted was 20.7%. The refinement statistics are summarized in Table 1. The atomic coordinates have been deposited.*

3. Results

3.1. *Asymmetry in the molecular structure*

The chimera, r(GUAUAUA)d(C), containing RNA residues throughout except for the 3′-terminal dC residue, crystallizes as a right-handed A-form double helix. The two independent strands in the duplex are conformationally different (r.m.s. deviation 0.8 Å) and are related by an approximate twofold axis (178.3° rotation, 0.07 Å translation). The differences in the two strands arise from differences in the conformation of the deoxyriboses 8 and 16 (sugar pucker), and the *trans/trans* α/γ backbone bonds of residues A3 and A7 on strand 1 which are complemented by U14 and U10 on strand 2, respectively, having the *gauche*$^-$/*gauche*$^+$ backbone. The duplex is compressed and deviates from the average fiber model for A-RNA (r.m.s. deviation 0.9 Å). The octamer helix is slightly bent at two positions, as revealed by analysis with the *CURVES* algorithm (Lavery & Sklenar, 1988). A 6° bend is seen between residues U2 and A3, and a 15° bend between residues U6 and A7, both of which are directed towards the deep (major) groove. The bending may result from the backbone assuming the extended conformation about α/γ (*trans/trans*) in one strand but not in the other.

3.2. *Conformational parameters of the alternating duplex*

The conformational parameters, calculated with the program *NUPARM* (Battacharyya & Bansal, 1990; Fig. 1) place the molecule into the A-form family where the major groove is deep and narrow (width 5.5 Å), while the minor groove is wide and shallow (average width 10.4 Å). The duplex has an average of 10.6 residues per turn, 34.0 (1.6)° twist and 2.8 (0.4) Å rise. All ribose rings are in the northern hemisphere (N-type) of the pseudorotation cycle (C3′-*endo*). However, one of the terminal deoxyriboses, dC16, displays C3′-*exo* sugar puckering (S-type) while the other, dC8, adopts the C3′-*endo* (N-type) pucker. When dC16 is manually forced into a C3′-*endo* conformation the *R* factor rises by 0.2% and after several cycles of refinement the sugar reverts to C3′-*exo*. The average intrastrand phosphate–phosphate distance is 5.84 Å. Roll angles alternate between large positive values for UpA steps (average 13.3°) and small positive or negative values for ApU steps (average 3.6°). The average propeller twist is −17.8 (3.4)° the inclination is 13.0 (9.1)°. A's seem to be associated with slightly smaller χ values (−157.9°) than U's (−155.5°) although this dependence is not very clear-cut. Excluding the terminal base pairs, the lowest values for the glycosidic torsions are correlated with the *trans/trans* (A3 and A7) and high *gauche*$^-$/*gauche*$^+$ (A13) conformations of backbone torsion angles α and γ. Stacking interactions are very similar for analogous steps in the duplex. Partial interstrand stacking of purines at UpA steps is observed while ApU steps exhibit pure intrastrand stacking. The stacking pattern of the UpA steps is prone to clashes in the shallow groove due to the base-pair propeller twist. Such clashes can be avoided through alternating rolls and twists (Calladine's rules), which is actually observed.

The helical parameters twist, roll, inclination, and *x* displacement oscillate throughout because of the alternating base sequence (Fig. 1). Alternation of the propeller twist and the rise is interrupted at the center of the sequence, the two halves of the molecule exhibiting the same trend, while the tilt values for the two halves are out-of-phase. Although the sequence is self-complementary, the alternation in the top half of the duplex is bigger than in the bottom half (Fig. 1).

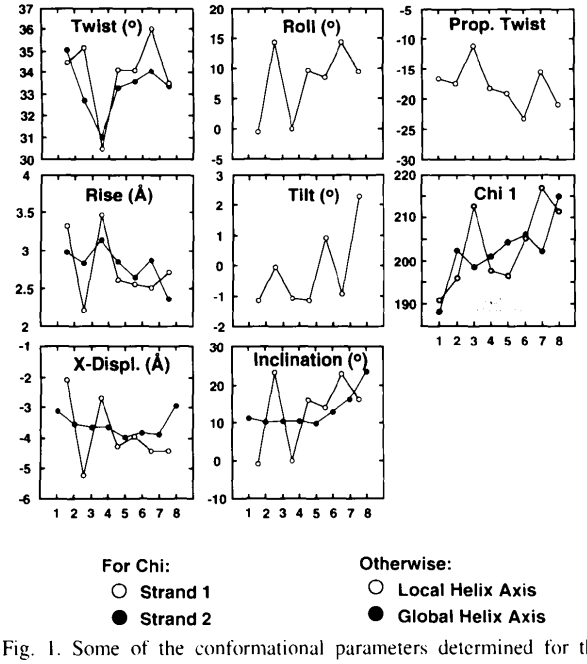

Fig. 1. Some of the conformational parameters determined for the current chimera show alternation between high and low values, which could reflect the alternation of purines and pyrimidines in the sequence. Particularly the twist, the roll, the inclination, and the *x* displacement show clear alternation. Also in the propeller twist and the rise some oscillation between high and low values is visible. The tilt alternates out-of-phase in the two molecular halves. The torsion angles around the glycosidic bonds show no clear dependence on the nature of the associated base. If alternating structural properties are associated with certain sequences it is conceivable that these could be recognized by ligand molecules.

* Atomic coordinates have been deposited with the Nucleic Acid Database (Reference: AHH071). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: gr0486).

3.3. *Solvent structure*

The duplex appears to be uniformly enveloped with solvent molecules. Of the 85 solvent molecules, 31 are in the deep groove, 24 in the shallow groove, and the remainder on the backbone and in the second hydration sphere. In the deep groove (Fig. 2*a*) waters form a continuous solvent column throughout the stacked duplexes, being anchored to the N7 atoms of purines, O4 of U's, and N4 of dC's without adopting an obvious pattern or network. In the shallow groove the water molecules form a discontinuous multiple spine (Fig. 2*b*). The waters mainly attach to the sugar ring O4′, the exocyclic O2′ sugar O atoms, the exocyclic N2 atoms of G's, O2 atoms of U's, and N3 atoms of A's. Unlike in A-tract B-DNA, which usually carries a single string of water molecules in the minor groove, the shallow groove here is flat and wide and easily accommodates short stretches of a double spine and more complicated networks of waters. At the backbone, single waters bridge some of the consecutive O1P phosphate O atoms but there is no continuous chain of such solvent molecules (not shown).

3.4. *Solvation of the 2′-hydroxyl groups*

2′-hydroxyl groups make direct or water-mediated intraduplex contacts to various parts of the nucleotide residues, *i.e.* to the bases, phosphate groups, and ribose

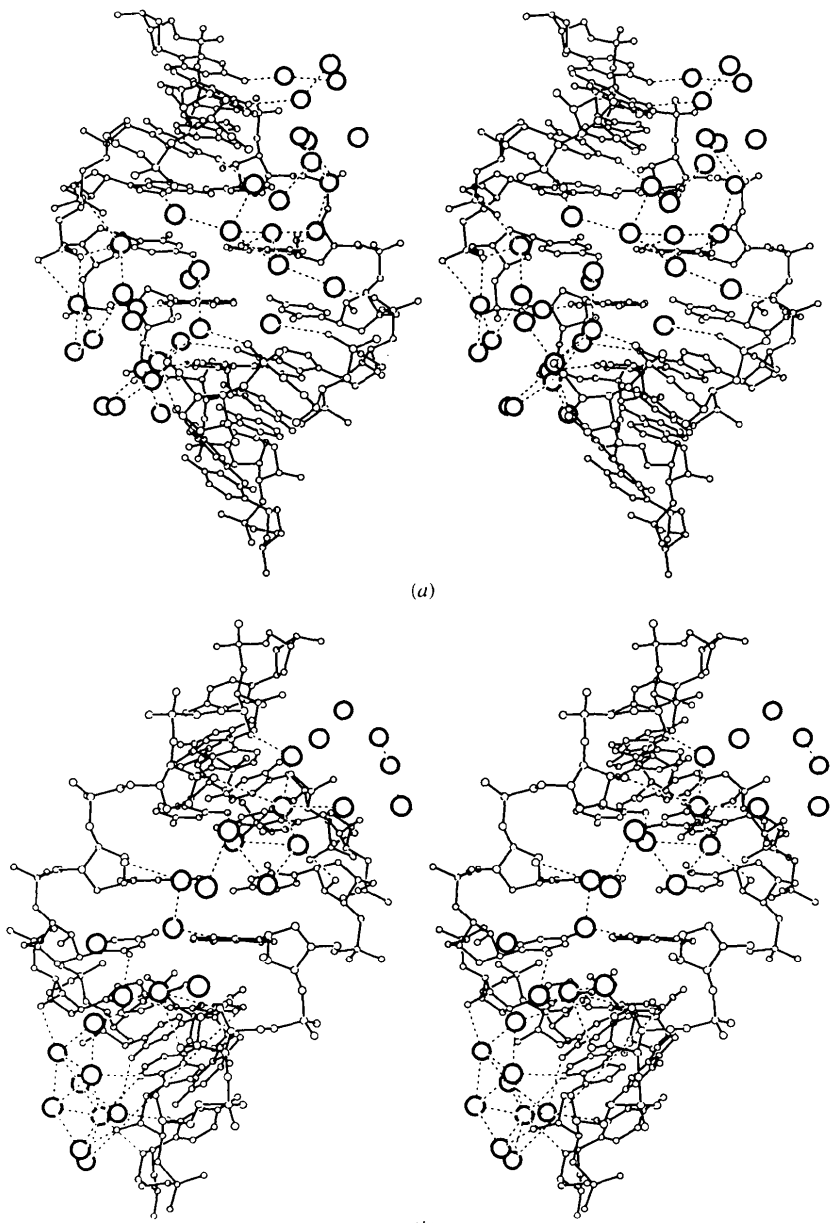

Fig. 2. Stereoviews of the hydration pattern of (*a*) the deep and (*b*) the shallow groove of the duplex. Water molecules are represented as larger spheres. Some hints of strings of water molecules are visible in the deep groove, but no clear organizational pattern is discernible. In the shallow groove a discontinuous double spine of hydration is observed. Note the asymmetry of the hydration, although the sequences of the top and bottom halves of the duplex are identical.

sugars. Contacts are both of the intranucleotide and internucleotide types. 11 of the 14 2′-hydroxyl groups are hydrated. Three of them, G1, A7 and U12, are involved in direct lattice interactions. In addition, these three 2′-hydroxyl groups and the 2′-hydroxyl group of A11 use water molecules or sodium ions to achieve indirect lattice contacts. The 2′-hydroxyl groups of residues A3 and U12 carry single hydrogen-bonded waters which make no further contacts.

A 2′-hydroxyl is hydrogen-bonded directly to the O4′ atom of the following 3′ residue four times and twice to the O5′ atom of the 3′-side residue. These types of interactions both occur between residues 15 and 16 and are lost when dC16 assumes the C3′-*endo* conformation. There are two examples in which an O2′ is connected *via* water molecules to the O4′ atom of the following residue. In one case the O2′ and O4′ atoms on the same residue are linked through a water. Water-bridged O2′–O3′ atoms are not frequent, but a unique feature of this structure is the hydrogen bonding of O2′ and O3′ of A7 to the exocyclic amino group of a symmetry-related A9.

Interestingly, some of the interactions which were previously seen with one bridging water, are mediated by two water molecules here. Among the 2′-hydroxyl group–base interactions, a two-water bridge between the 2′-hydroxyl and the O2 of the corresponding uracil appears particularly frequently (four times, Fig. 3*a*). Several other two-water-bridged patterns are seen: a double-water bridge between O2′ and N3 of residue

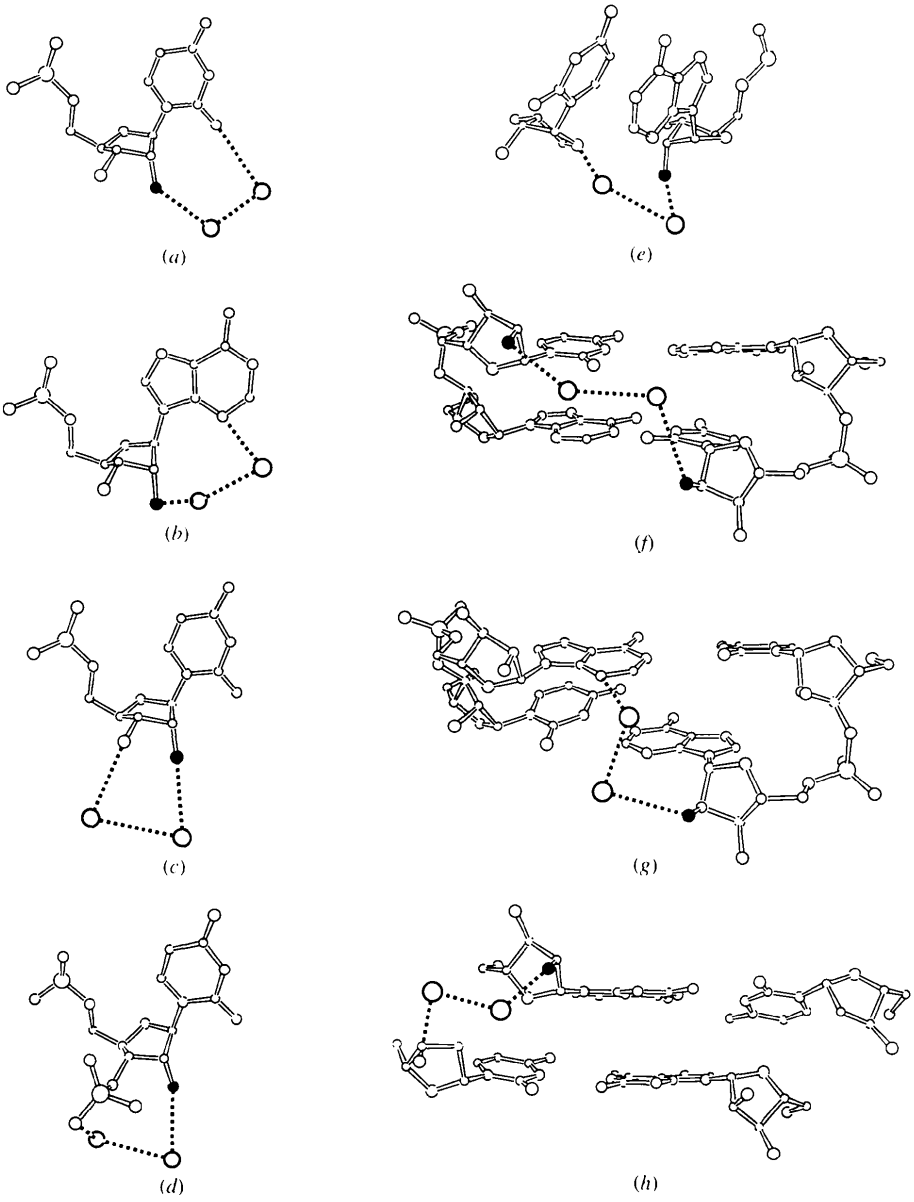

Fig. 3. Ball-and-stick figures illustrating long-range two-water bridges involving the 2′-hydroxyl groups. One-water analogs of the bridges shown in (*a*)–(*c*) are known and the insertion of an additional water does not seem to change any of the nucleotide conformational parameters by a large measure. From the figures it is obvious that the 2′-hydroxyl groups are important handles in RNA, involved in long-range contacts to various parts of the double helix. The bridging of 2′-hydroxyl groups across the shallow groove by two water molecules as in (*f*) accounts for the stretches of double-spine hydration seen in the shallow groove.

A11 (Fig. 3*b*); between O2′ and O3′ of U10 (Fig. 3*c*); between O2′ of residue U10 and O2P of residue A11 (Fig. 3*d*); and between O2′ of A11 and O4′ of U12 (Fig. 3*e*). A portion of the electron density showing the two-water bridge between O2 of U10 and its 2′-hydroxyl group is shown in Fig. 4. In the shallow groove there are indications for double and triple waters connecting 2′-hydroxyl groups between the two strands; *e.g.* the 2′-hydroxyl groups of U4 and U14 are connected *via* two water molecules (Fig. 3*f*) while three waters mediate the interaction between U6 and A13. These connections give the impression of a multiple spine of hydration in the shallow groove. There are other interstrand interactions across the shallow groove, *e.g.* between N3 of A's and the 2′-hydroxyl groups (Fig. 3*g*). The coaxial stacking of the duplexes is reinforced by double water bridges between O2′ and O3′ atoms (Fig. 3*h*).

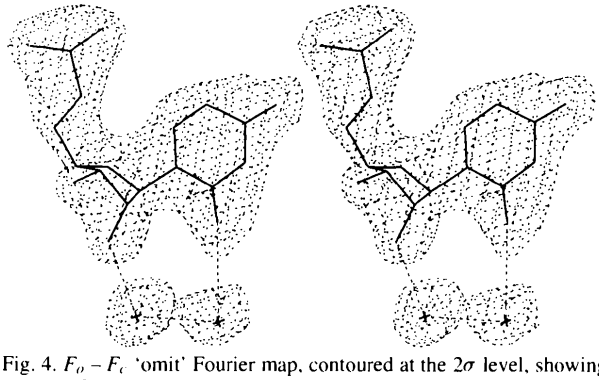

Fig. 4. $F_o - F_c$ 'omit' Fourier map, contoured at the 2σ level, showing part of the hydration of the U10 nucleotide. A two-water bridge between the base O2 atom and the sugar 2′-hydroxyl group is clearly visible.

3.5. *Interplay of solvation and helical parameters*

Early fiber-diffraction studies have demonstrated that the conformation of DNA is highly dependent on the humidity of the environment (see *e.g.* Saenger, 1984). The solvation patterns of nucleic acids have been studied previously (Westhof, 1988; Schneider *et al.*, 1993). Interestingly, although the two halves of r(GUAUAUA)d(C) have identical sequence, the solvent structure is quite different (Fig. 5). It is possible that this asymmetry may transiently exist in solution and is trapped in the crystal lattice. The solvation can, in part, explain the observed variations in conformation. The exocyclic O atoms of pyrimidine residues are linked *via* water molecules to exocyclic O atoms of pyrimidines in the opposite strand. This interaction is observed in three of the four possible cases and stabilizes the propeller

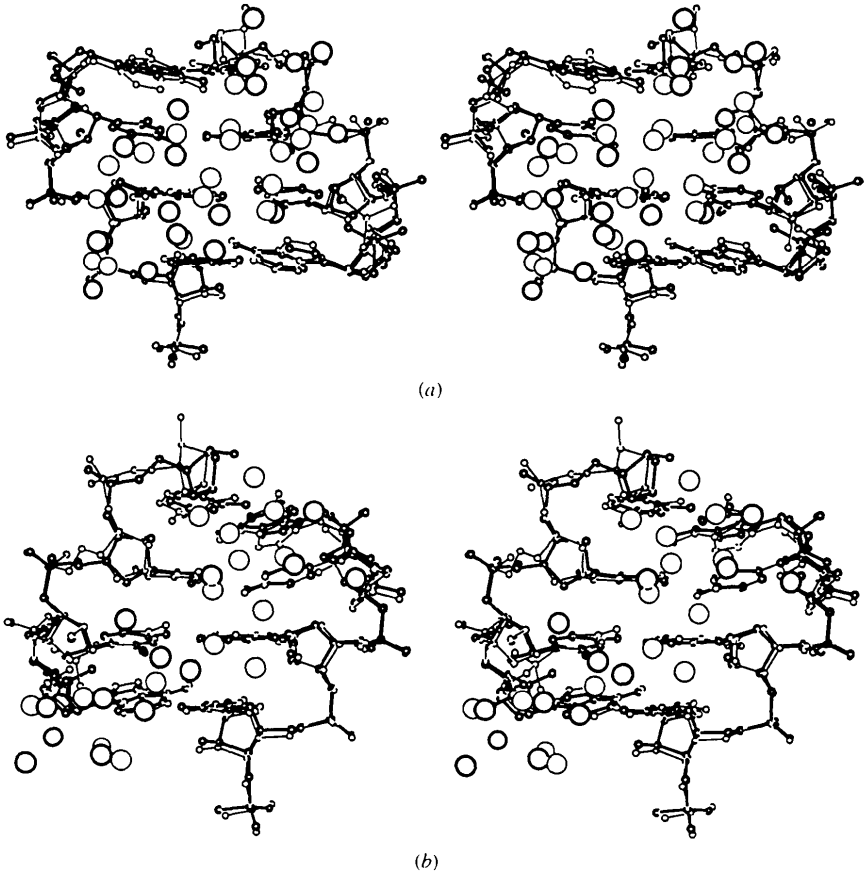

Fig. 5. Stereoviews of the superposition of the two molecular halves including the waters. Although the sequences of the two half-molecules are identical, no common hydration patterns are observed in either (*a*) the deep groove or (*b*) the shallow groove. Differential hydration may have an influence on the different conformations found in the two halves of the duplex or the two strands, or could be the result of different packing forces.

twist and possibly the roll angle. In the deep groove, the O6 of the terminal G's are in direct hydrogen-bonding contact with the exocyclic amino groups of A's in the complementary strand, further reinforcing the observed propeller twist. In addition, there is a water-mediated interaction between O4 of U6 and O4 of U10 which nicely complements the bridging of the O2 atoms of U6 (with O2 of U12) and U10 (with O2 of dC8) in the shallow groove. Analogous hydration patterns for the bases have been observed in alternating A-DNA structures (Kennard *et al.*, 1986). Because of the A-conformation of the duplex, the 2′-hydroxyl groups point into the shallow groove and are instrumental in organizing the water structure. The intra- and interstrand water involving the 2′-hydroxyl groups can be expected to lend additional rigidity to the RNA helix compared with A-DNA.

3.6. *Crystal packing – the role of water and sodium ions*

A deoxy residue was chosen for the 3′ end of the oligomer because we thought that the duplex might crystallize in a pattern analogous to octamer A-DNA's with the terminal base pair of one duplex stacking on the flat shallow groove of a symmetry-related molecule (Shakked *et al.*, 1983; Jain & Sundaralingam, 1989). Although such a stacking pattern has been observed for RNA (Dock-Bregeon *et al.*, 1988), there is the possibility that the 2′-hydroxyl group of the 3′-terminal residue could interfere with the packing. However, the packing scheme for the present octamer turned out to be quite different from what we had expected and the duplexes were stacked on top of each other. So far we have not obtained diffraction-quality crystals of the all-RNA analog. We have no explanation as to why crystals are obtained only when the terminal residue is d(C) but not r(C). Preparation and crystallization procedures were exactly the same for the two molecules. Each duplex is surrounded by three threefold axes and three 3_1-screw axes (Fig. 6*a*) so that the central duplex is surrounded by six neighboring duplexes. Corresponding to the three threefold axes, a duplex laterally faces three different crystal environments. These three lateral regions can be clearly seen for any pair of stacks of three duplexes (Fig. 6*b*). Direct and water-mediated contacts are observed only in two of the three regions (contact regions 1 and 2; Fig. 7). Many of these interactions involve 2′-hydroxyl groups. In contact region 1 (Fig. 7*a*) the duplexes interact with each other through the backbone of one duplex and a short stretch of the shallow groove of the other. In the upper part of contact region 1, O2′ of A7 hydrogen bonds to both N2 of G1^s and O2 of U2^s (s denotes the symmetry-related duplex). In the lower part, O2′ of U12 contacts O2P of A3^s. In addition, waters mediate contacts between N3 of A11 and O2′ of U2^s, O2 of U6 and O3′ of U2^s, and O2P of U12 and O2P of U4^s. In contact region 2 (Fig. 7*b*), O3′ of dC16^s is hydrogen bonded to both O3′ of U14 and O2P of A15. Waters

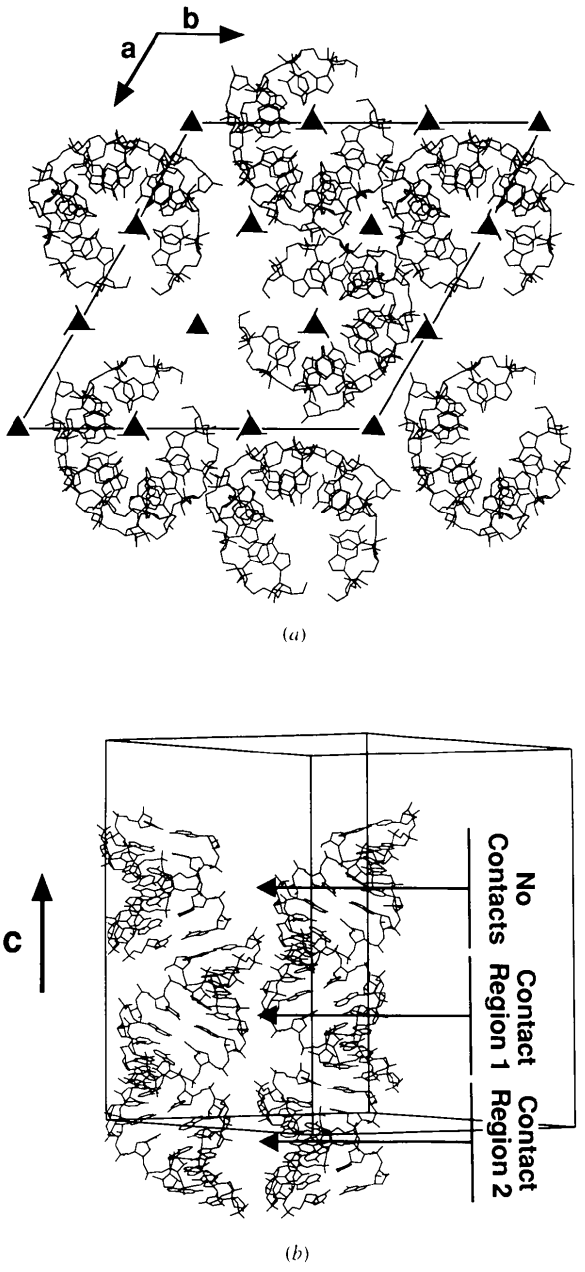

Fig. 6. (*a*) Top and (*b*) side views of the arrangement of the duplexes in the crystal lattice. As seen from the neighborhood of the duplex situated in the lower right corner of the unit cell in (*a*), there are six nearest neighbors to each duplex related by three threefold axes (two into the negative *a* direction, two approximately into the negative *b* direction and two in the positive *a* direction with respect to the reference duplex). The three groups of lateral nearest neighbors account for three different close-approach areas around each helix as shown in (*b*). Only in two of the groups we observe direct and water-mediated contacts. The duplexes themselves are positioned on 3_2-screw axes which accounts for their head-to-tail pseudocontinuous stacking along the *c* direction (*b*). Because the helix axes are slightly inclined with respect to the *c* axis, the screw symbols do not appear centered in the deep groove (*a*).

connect O2′ and O3′ of A5 to O3′ of A3^{s} and O2P of U4^{s}, respectively. In both these cases the water bridges are weak with water distances of ~3.8 Å to one of the contact partners. RNA helices may be arranged in this fashion in large RNA molecules. The 2′-hydroxyl groups may be instrumental in this packing scheme.

The global helix axis of the octamer duplex is inclined by ~10° with respect to the 3_2-screw axis of the crystal. The screw axis relates molecules which are stacked on top of each other in a head-to-tail fashion (Figs. 6*b* and 8). Interestingly, the approximate 11-fold helix of the RNA is almost perfectly superimposed on the crystallographic 3_2-screw axis. The helical twist is such that the octamer repeats approximately every 238°, fortuitously matching closely the 240° repeat of the 3_2-screw.

Two water molecules (100 and 101) and two sodium ions on the threefold axis make direct or water-mediated contacts to the three residues, dC8, U10, and A11. Waters 20, 21, 58, 60 and 75 together with their symmetry-related equivalents are involved in an elaborate hydrogen-bonding network, which can be grouped into three layers centered around sodium I, sodium II and water 101, respectively (Fig. 9). Sodium I is tetrahedrally coordinated (2.4–2.5 Å) by the 2′-hydroxyl groups of A11 and symmetry equivalents and the water 100 (Fig. 9*a*). Below this sodium on the threefold axis is sodium II, which is octahedrally coordinated by waters 21 and 75 and their symmetry-related equivalents, with distances of 2.4 and 2.3 Å, respectively (Fig. 9*b*). The hydrated sodium II is connected to the 2′-hydroxyl of U10 through waters 21 and 60 (2.8 Å). Further along the axis, water 101 is associated with water 58 and its equivalents (Fig. 9*c*). Water 58 is connected to water 21 and O2 of dC8 which completes the connectivity. Water 101 would have to be disordered to engage in three identical contacts. The 2′-hydroxyl groups, which are RNA specific, are intimately involved in the hydrogen-bonding network.

3.7. *Helical junction and duplex stacking*

There are clear discontinuities across the gaps between the stacked helices. The helical twist and the rise across the gap are ~1° and 2 Å, respectively. Pseudo-interstrand purine ring stacking is emphasized at the junction instead of pseudo-intrastrand purine/pyrimidine stacking. The stacking is reminiscent of the O4′ sugar-ring O atoms of the guanines on top of the deoxycytosine bases as observed in crystal structures of nucleotide building blocks (Bugg, Thomas, Sundaralingam & Rao, 1971). Similar stacking behavior is found for the dC

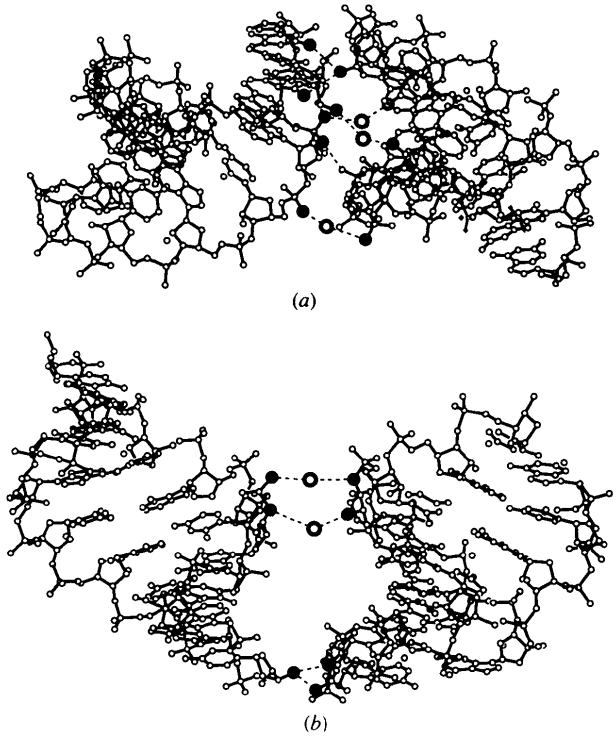

Fig. 7. Close-up of the two contact regions (see Fig. 5*b*). Water molecules which mediate contacts are shown as open spheres. Note the reciprocal fit of shallow groove and backbone in (*a*). In this contact region I most of the direct contacts are encountered, with a more loose association in region 2. Atoms involved in the contacts are shown as larger solid spheres.

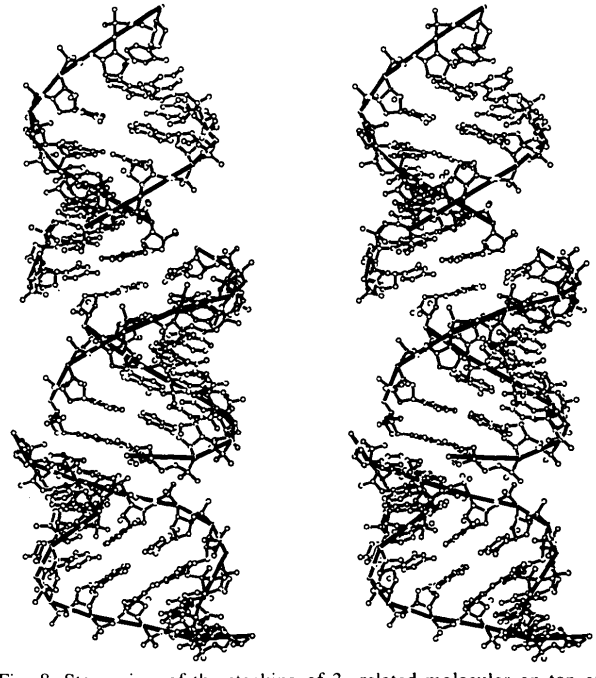

Fig. 8. Stereoview of the stacking of 3_2-related molecules on top of each other. The large purine rings maximize their interactions while the pyrimidine rings stack above the sugar rings of a symmetry-related molecule. There is virtually no twist between the duplexes illustrating the flexibility available to the coaxial stacking of RNA helices.

O4′ sugar ring O atom lying on top of the guanine base in the CpG step of Z-DNA (Egli & Gessner, 1995). The stacking is different to the stacking between terminal base pairs in r(CGCGAAUUAGCG) (Fig. 10; Leonard *et al.*, 1994) where pseudo-intrastrand stacking is observed. RNA can, therefore, modulate the coaxial stacking through the chemical identity of the helix termini and/or the interactions in distal parts of the duplexes, which may be exploited in coaxially stacked helices in large RNA molecules.

Although the groove widths in short oligonucleotides can be measured at only a small number of points (in octamers, *e.g.*, only one measurement of the deep groove is possible), grooves formed between two stacked duplexes can be evaluated. In the current structure both grooves are contracted between duplexes (Figs. 8 and 11). Four motions of the stacked duplexes relative to each other may influence the width of the grooves across the junction: the rise and the twist across the gap, and sliding and inclination of the two helices with respect to each other. In the current example, the sliding of the duplexes across one another in order to allow the purine–purine interstrand stacking, the decreased rise across the junction, and the inclination of the helix axes lead to the observed groove contraction across the gap. In order to satisfy the preference of the purine rings to efficiently stack upon each other across the junction, a small twist angle, as observed, may be favored.

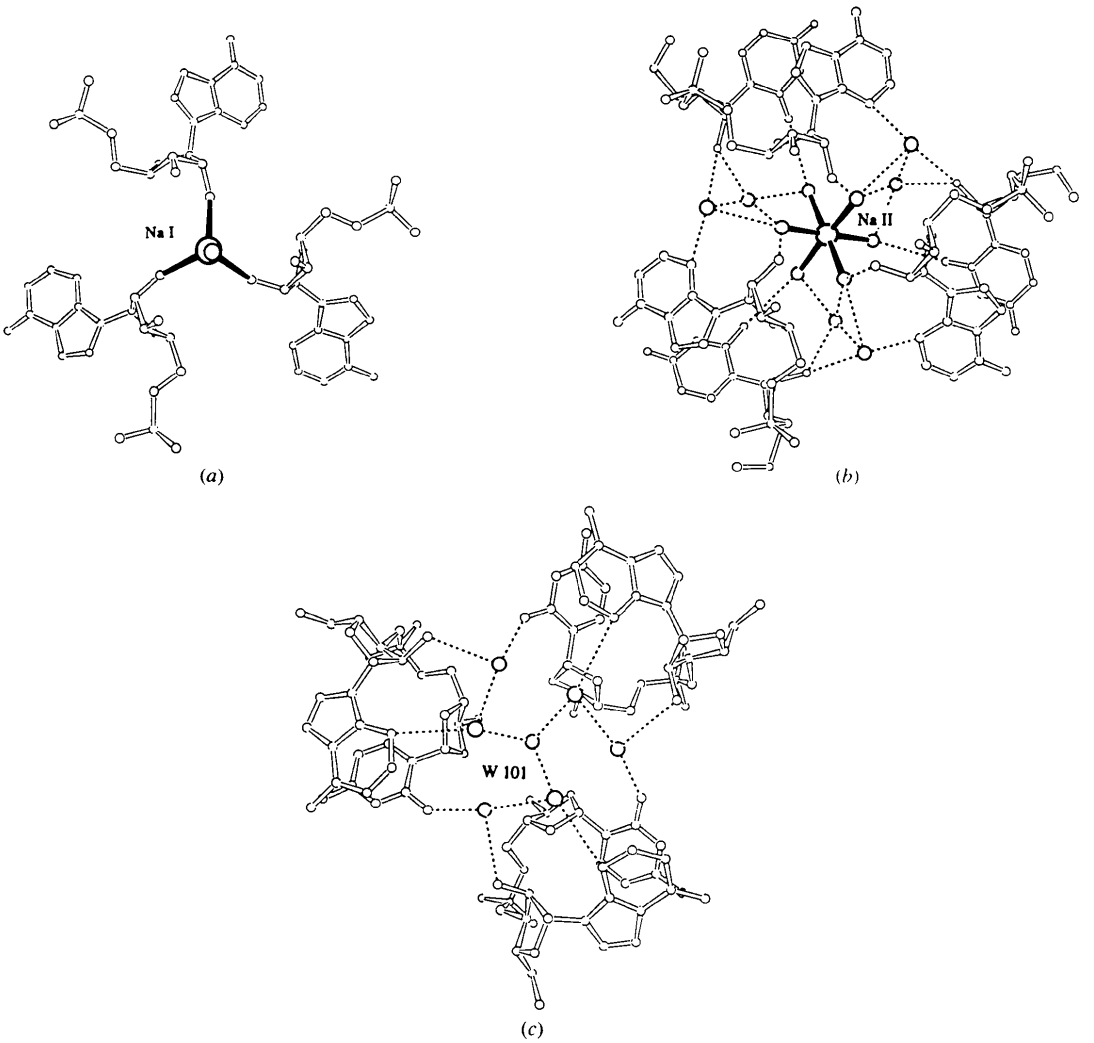

Fig. 9. Two sodium ions and two water molecules on the threefold axis mediate the approach of three duplexes. Large parts of the duplexes are omitted for clarity. (*a*)–(*c*) are consecutive views along the threefold axis. Note the involvement of 2′-hydroxyl groups in the organizations. (*a*) 2′-hydroxyl groups of residue A11 and its symmetry-related counterparts and water 100 tetrahedrally coordinate a sodium ion, sodium I. Water 94 is depicted on top of the sodium ion on the threefold axis. (*b*) Sodium II is octahedrally coordinated by waters which are in contact with 2′-hydroxyl groups of residues U10 and A11. A possible hydrogen-bonding network is depicted. (*c*) Water 101 is associated with a hydrogen-bonding network involving the 2′-hydroxyl group of A9 and O2 of dC8.

4. Discussion

4.1. *Alternating RNA structures*

Alternating structural parameters, as observed in the current octamer, could be exploited by ligand molecules, such as proteins, which could also modulate the alternation upon binding. In the RNA duplex, r[U(UA)6A], having a long stretch of alternating base sequence determined at a similar resolution (Dock-Bregeon *et al.*, 1989), alternation of the roll angles was also observed, while alternation in other helical parameters was lacking, possibly due to the use of an analysis program which employed a linear global helix axis. This is consistent with the theoretical models of Spôner and Kypr (Spôner & Kypr, 1991) who predicted that Calladine-type clashes in the minor groove at pyrimidine–purine steps can be avoided by a positive roll. Although the predicted clashes are less severe for UpA steps than for CpG steps, due to the lack of the guanine 2-amino group, the sequences are still observed to have clear roll alternation.

4.2. *Hydration of RNA at A·U base pairs*

The present crystal structure and that of r[U(UA)$_6$A] (Dock-Bregeon *et al.*, 1989) contain extended stretches of A·U base pairs. Analysis of the two structures reveals some common hydration motifs for the A·U base pairs, starting to outline a consensus hydration pattern. The tendency to link the base through one or two water molecules to the 2′-hydroxyl group seems to be more pronounced for U's than for A's, O2 of U's being, in general, more often hydrated than N3′s of A. The number of waters does not significantly influence the glycosyl χ angle or the sugar pucker, the two parameters which could most influence the distance between the 2′-hydroxyl group and the base hetero atom. In some instances the bases are bridged in the shallow groove between O2 of U and N3 of A by a water molecule but more commonly the two atoms are separately hydrated with no hydrogen bond between the water molecules. In a recent report (Wahl, Rao & Sundaralingam, 1996)

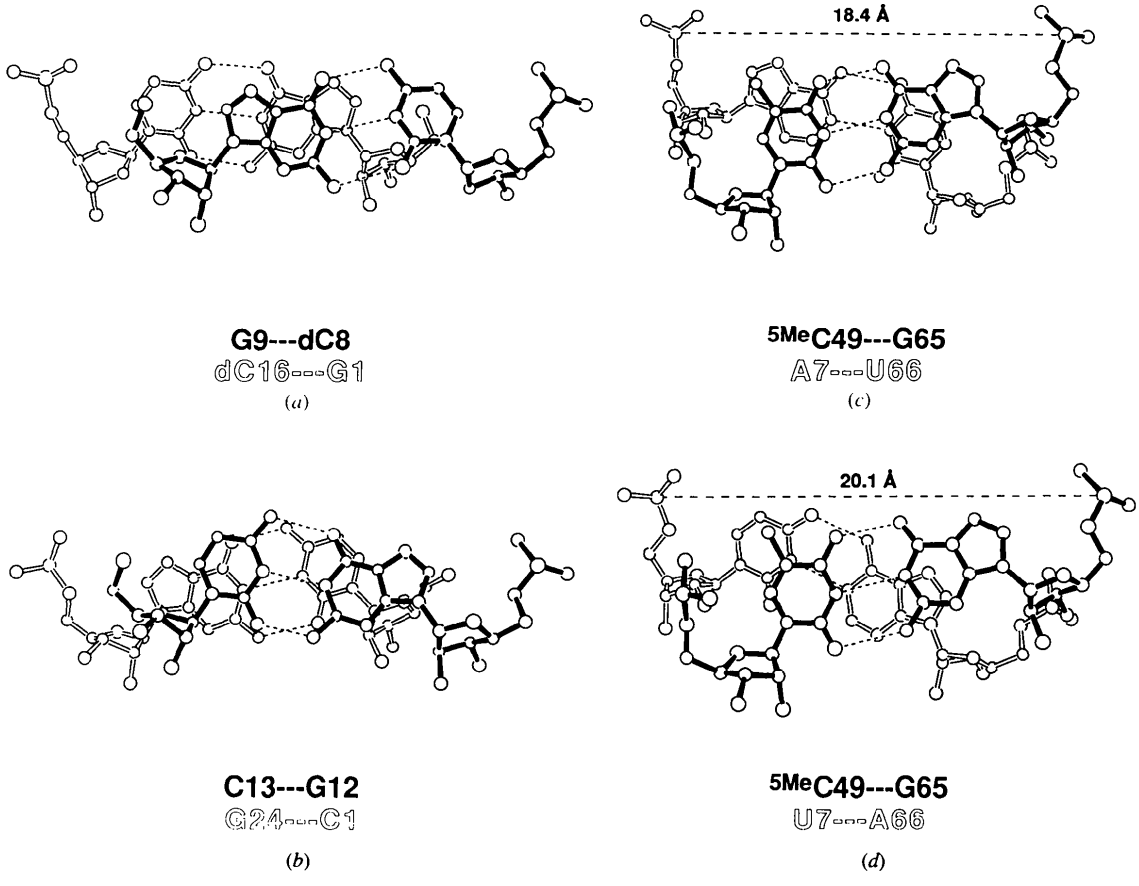

Fig. 10. Stacking of the terminal base pairs in (*a*) the current structure and (*b*) in the crystal structure of r(CGCGAAUUCAGCG) (Leonard *et al.*, 1994) in which the sequence alternation is changed from 5′-C-3′-G to 5′-G-3′-C. The stacking behavior is analogous to the alternating steps within the continuous duplex. (*c*) and (*d*) show the pseudocontinuous stacking of the pseudouridine stem on the amino-acid acceptor stem of tRNA^{Asp} and tRNA^{Phe}, respectively. Because of the change of the sequence the deep groove widths are different by almost 2 Å.

it was found that G·C pairs in RNA often contain a string of three water molecules in the shallow groove connecting the 2′-hydroxyl groups and hydrating each shallow groove-exposed hetero atom of the bases. Such a string of waters may be interrupted in A·U pairs because of the lack of an exocyclic substituent at the C2 position of A. In the deep groove the N7 atoms of A's are almost always hydrated but the water molecule does not seem to make further contacts. In some cases a single water molecule bridges U (O4) and A (N6) in analogy to that found in the G C pairs (Wahl, Rao & Sundaralingam, 1996). Comparison of the three UpA steps in the present structure and the six UpA steps in r[U(UA)$_6$A] shows that the consensus hydration is restricted to the base pair and reveals no new features and bridges for the step. The observed hydration pattern involving the 2′-hydroxyl groups probably reinforces the A-form conformation of the duplexes.

4.3. *Helix asymmetry*

Although the two halves of the current duplex have an identical base sequence, they are hydrated differently.

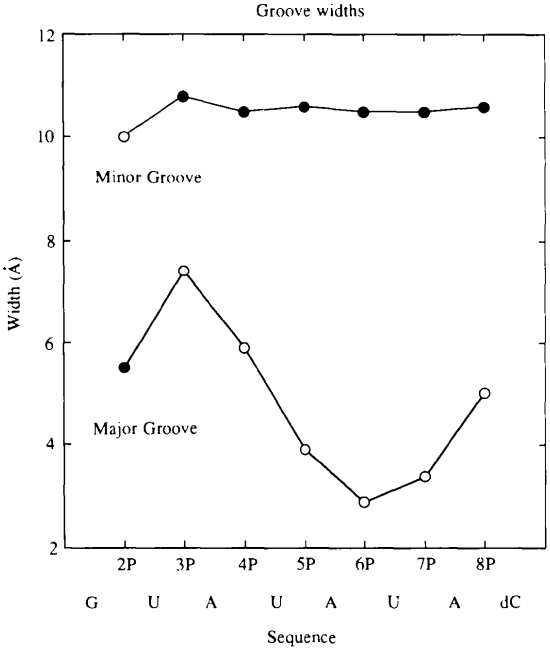

Fig. 11. Across the stacking junction, where bridging phosphate groups are missing, the widths of the grooves formed by two stacked duplexes can be determined. The groove widths were measured as the closest cross-strand P—P distance (minus 5.8 Å for two phosphate van der Waals radii) across the respective groove. Distances measured between phosphates of the same duplex are shown as solid circles and distances measured between phosphates of two different duplexes as open circles. While the shallow groove is mainly made up of residues from the same duplex the deep groove is mainly formed by residues from different duplexes. Both grooves are contracted when extended into a coaxially stacked molecules. The geometry of the stacking junction determines this modulation of the groove widths for stacked helices.

There is no extended common hydration network in the two halves in either groove. The asymmetry of the hydration is matched by the asymmetry of the structural parameters (different degree and phase of alternation). It is possible that the differential hydration is to some degree due to the crystal packing because the duplex does not face the same crystal environment in each direction. The water molecules that mediate the interactions between the helices could lead to an increased flexibility of their conformation and of their geometrical arrangement. A similar asymmetry in the hydration and the geometry can be seen in the structure of the self-complementary sequence r[U(UA)$_6$A]. However, the duplexes in the r[U(UA)$_6$A] structure pack in the common A-DNA fashion with the terminal base pair of one contacting the shallow groove of a symmetry-related molecule. This type of crystal packing, with extensive contacts in the grooves, is likely to locally perturb the RNA structure and be the source of the structural asymmetry of the two molecular halves.

4.4. *Helix stacking*

Long continuous helices are seldom encountered in naturally occurring RNA's. Short helical stretches, however, can stack pseudocontinuously on top of each other; *e.g.* stacking of helices is known as an organizing motif in tRNA's. It was also observed in the recent structures of the hammerhead ribozymes (Pley, Flaherty & McKay, 1994; Scott, Finch & Klug, 1995). In addition, Cech's group introduced a general method to test for coaxially stacked helices and identified them in the catalytic core of the tetrahymena ribozyme (Murphy, Wang, Griffith & Cech, 1994). In almost all of the published RNA crystal structures, duplexes in the crystal lattice stack on top of each other forming pseudo-infinite helices. Again in the current example, helix stacking seems to be a major organizing theme in the crystal, particularly in light of the lack of extensive inter-duplex hydrogen-bonding interactions. The frequency of its occurrence, pointing to the effectiveness of this type of interaction, matches the high degree of ordering achieved when helices stack in an RNA molecule. Helical stacking, therefore, should be a guiding motif in attempts to predict RNA tertiary structure. From the comparison of the current octamer to the non-alternating structure of r(CGCGAAUUAGCG) (Leonard *et al.*, 1994) it is obvious that the pseudocontinuous stacking of the terminal base pairs is analogous to the stacking behavior of base pairs within the continuous duplexes, with 5′-pyrimidine-3′-purine steps showing extensive interstrand stacking and 5′-purine-3′-pyrimidine steps exhibiting exclusive intrastrand stacking (Fig. 10).

4.5. *Groove accessibility*

Although the deep groove of A-form RNA contains a larger number of potential hydrogen donor and ac-

ceptor atoms than the shallow groove, which could be used for specific interactions with ligands, it is rather narrow and, therefore, inaccessible. Weeks and Crothers demonstrated with chemical acylation experiments that structural perturbations, afforded *e.g.* by internal loops, which lead to a loss of stacking enthalpy, can serve as efficient helix terminators and result in more accessible deep grooves, while coaxially stacked helices could efficiently maintain helix continuity and therefore inaccessibility of the deep groove (Weeks & Crothers, 1992). The latter notion is clearly confirmed by the current X-ray structure. However, comparison with other RNA duplex structures (Fig. 10) suggests that the stacking geometry can be modulated by the chemical nature of the terminal bases and/or by interactions in distal parts of the duplex. A similar situation may be found with coaxially stacked helices in natural RNA molecules which could lead to differentially accessible grooves. The coaxially stacked helices found in natural RNA's usually display one consecutive strand and a nick in the opposite strand, as opposed to the double-strand nicks seen between duplexes in crystals. In the crystal structures of $tRNA^{Phe}$ (Suddath *et al.*, 1974; Robertus *et al.*, 1974; Stout, Mizuno, Rao, Swaminathan & Sundaralingam, 1978) and $tRNA^{Asp}$ (Moras *et al.*, 1980) the pseudouridine stem is stacked on the amino-acid acceptor stem. The base pairs involved are (U7/^{5Me}C49)·(G65–A66) in $tRNA^{Phe}$ and (A7/^{5Me}C49)·(G65–U66) in $tRNA^{Asp}$. Although the consecutive strand may be expected to somewhat restrain the stacking interaction, the change from a non-alternating step ($tRNA^{Phe}$) to a purine–pyrimidine alternating step ($tRNA^{Asp}$) changes the deep groove width by about 2 Å, estimated from the phosphate–phosphate distances between residues 7 and 65 (Fig. 10). Such modulations in the groove widths may allow or deny access for ligands, such as proteins or the third strand in an RNA triplex.

This work was supported by NIH grant GM-17378, an OSU Presidential Fellowship to MCW, and an Ohio Regents Eminent Scholar Chair to MS.

References

Arnez, J. G. & Steitz, T. A. (1994). *Biochemistry*, **33**, 7560–7567.

Baeyens, K. J., De Bondt, H. L. & Holbrook, S. R. (1995). *Nature Struct. Biol.* **2**, 56–62.

Bhattacharyya, D. & Bansal, M. (1990). *J. Biomol. Struct. Dynam.* **8**, 539–572.

Betzel, C., Lorenz, S., Forste, J. P., Bald, R., Zhang, M., Schneider, T. R., Wilson, K. S. & Erdmann, V. A. (1994). *FEBS Lett.* **351**, 159–164.

Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994). *Science*, **263**, 1404–1410.

Brünger, A. T. (1988). In *Crystallographic Computing 4: Techniques and New Technologies*, edited by N. W. Isaacs, & M. R. Taylor, pp. 126–140. Oxford: Clarendon Press.

Bugg, C. E., Thomas, J. M., Sundaralingam, M. & Rao, S. T. (1971). *Biopolymers* **10**, 175–219.

Cruse, W., Saludjian, P., Biala, E., Strazewski, P., Prange, T. & Kennard, O. (1994). *Proc. Natl Acad. Sci USA*, **91**, 4160–4164.

Dock-Bregeon, A. C., Cheerier, B., Podjarny, A., Moras, D., deBear, J. S., Gough, G. R., Gilham, P. T. & Johnson, J. (1988). *Nature (London)*, **335**, 375–378.

Dock-Bregeon, A. C., Cheerier, B., Podjarny, A., Johnson, J., deBear, J. S., Gough, G. R., Gilham, P. T. & Moras, D. (1989). *J. Mol. Biol.* **209**, 459–474.

Egli, M. & Gessner, R. V. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 180–184.

Holbrook, S. R., Cheong, C., Tinoco, I. Jr & Kim, S. H. (1991). *Nature (London)*, **353**, 579–581.

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

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

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

Lavery, R. & Sklenar, H. (1988). *J. Biomol. Struct. Dynam.* **6**, 63–91.

Leonard, G. A., McAuley-Hecht, K. E., Ebel, S., Lough, D. M., Brown, T. & Hunter, W. N. (1994). *Structure*, **2**, 483–494.

Michel, F. & Westhof, E. (1990). *J. Mol. Biol.* **216**, 585–610

Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P. & Giegé, R. (1980). *Nature (London)*, **288**, 669–674.

Murphy, F. L., Wang, Y. H., Griffith, J. D. & Cech, T. R. (1994). *Science*, **265**, 1709–1712.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). *Nature (London)*, **372**, 432–438.

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

Portmann, S., Usman, N. & Egli, M. (1995). *Biochemistry*, **34**, 7569–7575.

Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989). *Science*, **246**, 1135–1142.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. & Klug, A. (1974). *Nature (London)*, **250**, 546–551.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). *Science*, **252**, 1682–1689.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*, pp. 331–349. New York: Springer Verlag.

Schindelin, H., Zhang, M., Bald, R., Forste, J. P., Erdmann, V. A. Heinemann, U. (1995). *J. Mol. Biol.* **249**, 595–603.

Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K. & Berman H. M. (1993). *Biophys. J.* **65**, 2291–2303.

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

Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A. & Viswamitra, M. A. (1983). *J. Mol. Biol.* **166**, 183–201.

Spôner, J. & Kypr, J. (1991). *J. Mol. Biol.* **221**, 761–764.

Stout, C. D., Mizuno, H., Rao, S. T., Swaminathan, P. & Sundaralingam, M. (1978). *Acta Cryst.* B**34**, 1529–1544.

Suddath, F. L., Quigley, G. J., McPherson, A., Sneden, D., Kim, S. H. & Rich, A. (1974). *Nature (London)*, **248**, 20–24.

Sundaralingam, M. (1979). *Transfer RNA: Structure, Properties and Recognition*, edited by P. R. Schimmel, D. Söll & J. N. Abelson, pp. 115–132. New York: Cold Spring Harbor Laboratory Press.

Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). *Nature (London)*, **371**, 623–626.

Wahl, M. C. & Sundaralingam, M. (1995). *Curr. Opin. Struct. Biol.* **5**, 282–295.

Wahl, M. C., Rao, S. T. & Sundaralingam, M. (1996). *Nature Struct. Biol.* **3**. In the press.

Weeks, K. M. & Crothers, D. M. (1992). *Science*, **261**, 1574–1577.

Westhof, E. (1988). *Int. J. Biol. Macromol.* **9**, 185–192.