Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the dodecamer r(GAUCACUUCGGU) with four 5′-overhang nucleotides

Subramaniam Eswaramoorthy, Sambhorao T. Rao, Baocheng Pan and Muttaiya Sundaralingam*

Department of Chemistry and Biochemistry, The Ohio State University, 200 Johnston Laboratory, 176 West 19th Avenue, Columbus, OH 43210, USA

Correspondence e-mail: sundaral@chemistry.ohio-state.edu

Received 23 September 2003
Accepted 2 December 2003

NDB Reference: r(GAUCACUUCGGU), AR0007.

The crystal structure of an RNA dodecamer, r(GAUCACUUCGGU), was solved at 2.6 Å resolution by the molecular-replacement method and refined to an R_{work} of 18.8% (R_{free} = 22.8%) using 2494 reflections. The dodecamer crystallized in the monoclinic space group $C2$, with unit-cell parameters $a = 71.34$, $b = 39.98$, $c = 32.47$ Å, $\beta = 104.7^\circ$ and two independent strands in the asymmetric unit. The dodecamer adopts an octamer duplex structure with four 5′-overhang residues (G1A2U3C4), which form Watson–Crick base pairs with another four 5′-overhang residues of a symmetry-related duplex. The octamer duplex (ACUUCGGU) contains at its center four mismatched base pairs flanked by two Watson–Crick base pairs. The mismatched bases form two G·U wobble base pairs at the ends and two U·C base pairs at the center, with one base–base hydrogen bond N4(C)···O4(U) and a water bridge connecting the N(3) of the cytosine and uridine. The present study reinforces the concept of the stability of the conformation of UUCG in RNA double-helical structures.

1. Introduction

RNA molecules display quite a variety of three-dimensional structures. The important structural motifs include double helices, hairpin loops, bulges and pseudo-knots. The double helices usually contain both Watson–Crick base pairs and mismatched base pairs. Mismatches are often found in conserved regions of RNA, suggesting that they have biological importance. Tetraloops are rampant in the architecture of ribosomal RNA molecules (Woese *et al.*, 1990), where the UUCG loop is most commonly found. In crystals, oligonucleotides containing tetraloops form a two-stranded blunt-end duplex with complementary and mismatched base pairs (Holbrook *et al.*, 1991; Cruse *et al.*, 1994; Baeyens *et al.*, 1995, 1996), in contrast to the single-stranded hairpin observed in solution studies (Cheong *et al.*, 1990; Varani *et al.*, 1991). All these sequences contain tetraloops in the center and form blunt-end stems. In this paper, we designed a dodecamer r(GAUCACUUCGGU) with the tetraloop asymmetrically disposed toward the 3′-end with four 5′-terminal overhang residues. This dodecamer may form a UUCG tetraloop with its overhang nucleotides forming Watson–Crick base pairs with their counterpart from another molecule (Fig. 1*a*). Such a constructed homodimer may enhance the stability of the formation of the tetraloop in the crystalline state. However, the present study shows that the tetraloop sequence UUCG in the dodecamer still forms four consecutive mismatches inside a duplex.

Table 1
Crystal data and refinement statistics for r(GAUCACUUCGGU).

Space group	*C*2
Unit-cell parameters (Å, °)	a = 71.34, b = 39.98, c = 32.47, β = 104.7
Resolution range (Å)	10.0–2.6
No. of observed reflections [$>2.0\sigma(F)$]	2494
Completeness (overall) (%)	91
Completeness (outermost shell, 2.76–2.6 Å) (%)	71
$R_{\rm merge}$ (%)	4.5
Asymmetric unit contents	2 strands
$R_{\rm work}$	0.188
$R_{\rm free}$	0.228
Parameter file used	param_nd.dna
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.01
Bond angles (°)	1.2
Torsion angles (°)	7.1
Improper angles (°)	1.5
Final model	
Nucleic acid atoms	500
Water atoms	27

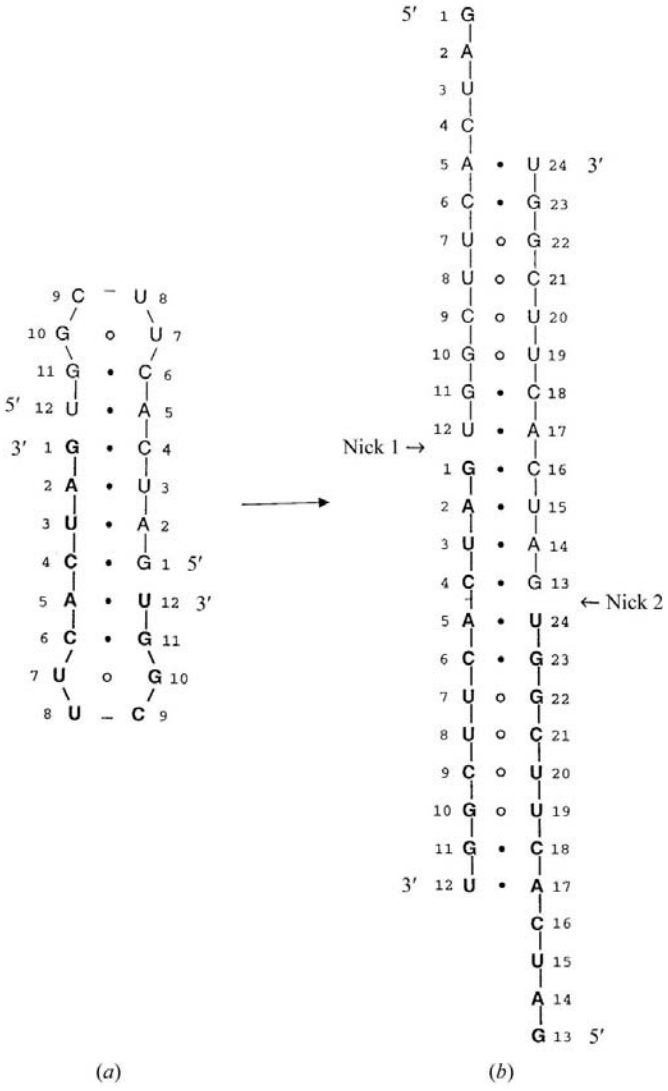

Figure 1
The schematic diagrams of two different possible conformations of the dodecamer r(GAUCACUUCGGU). (*a*) The targeted conformation of homodimer, containing a UUCG tetraloop. (*b*) The conformation observed in crystal structures.

2. Materials and methods

2.1. Synthesis, crystallization and data collection

The oligonucleotide was synthesized using an ABI nucleic acid synthesizer, employing the solid-phase phosphoramidite method. The solid-supported oligonucleotide was cleaved using ammonium hydroxide with 30% ethanol. The bases were deprotected in an overnight incubation at 328 K. The 2′-OH groups were deprotected with 2 *M* tetrabutylammonium fluoride in THF at a concentration of 100 µl per base. The sample was purified by ethanol precipitation and FPLC in 30% acetonitrile with a linear gradient of $LiCl_2$ as the eluant. The pure RNA single-strand sample was concentrated to 4 m*M*. 2 µl of RNA sample was dissolved in 4 µl of a Natrix solution (Scott *et al.*, 1995) containing 0.2 *M* KCl, 0.1 *M* magnesium acetate, 2 µl 0.6 m*M* spermine, 0.05 *M* sodium cacodylate pH 6.5 and 10% PEG 8000. The hanging-drop technique was utilized for crystallization by equilibrating the solution against 20% PEG 400. Crystals grew slowly to maximum dimensions of 0.2 × 0.2 × 0.1 mm in 60 d.

A crystal was mounted in a glass capillary with mother liquor at one end and sealed. The data were collected on our R-AXIS IIc imaging-plate system with a Rigaku X-ray generator and graphite monochromator ($\lambda_{\rm Cu\,K\alpha}$ = 1.5418 Å) operating at 50 kV and 100 mA. The space group is *C*2 and the unit-cell parameters are a = 71.34, b = 39.98, c = 32.47 Å, β = 104.7°, corresponding to two dodecamer strands in the asymmetric unit. A total of 42 frames with 4° oscillation per frame were recorded and the data were processed using *DENZO* (Otwinowski & Minor, 1997). 2494 reflections (91% complete) with $F > 2\sigma(F)$ were collected to 2.6 Å resolution. The $R_{\rm merge}$ on intensity was 4.5%. In the outermost resolution shell (2.76–2.6 Å) the data completeness was 71%. Data-collection and refinement statistics are given in Table 1.

2.2. Structure solution and refinement

The structure was solved by the molecular-replacement method and refined using the program *X-PLOR* (Brünger, 1992). An $(AU)_4$ octamer segment from the known crystal structure of an r[A$(UA)_6$U] duplex (Dock-Bregeon *et al.*, 1989) was used as the search model with 649 reflections in the 10–4 Å resolution range. The best orientation gave a clear maximum in the translation search. Rigid-body refinement of this solution reduced the R value to 0.442. Including all 2494 reflections between 10 and 2.6 Å, $R_{\rm work}$ was 0.480 ($R_{\rm free}$ was 0.505 for 6% or 163 randomly selected reflections). Powell minimization reduced $R_{\rm work}$ and $R_{\rm free}$ to 0.392 and 0.461, respectively. $2F_o - F_c$ and $F_o - F_c$ electron-density maps were calculated by omitting a base pair at a time and the octamer sequence was identified as ACUUCGGU at the 3′-end. Further refinement reduced $R_{\rm work}$ and $R_{\rm free}$ to 0.354 and 0.403, respectively. $2F_o - F_c$ and $F_o - F_c$ electron-density maps now clearly revealed the expected density for the four 5′-overhang residues of both strands. These residues were included in the model and simulated annealing was performed by heating the system to 3000 K and then slow-cooling to 300 K. Individual B-factor refinement in the resolution range 10–2.6 Å lowered

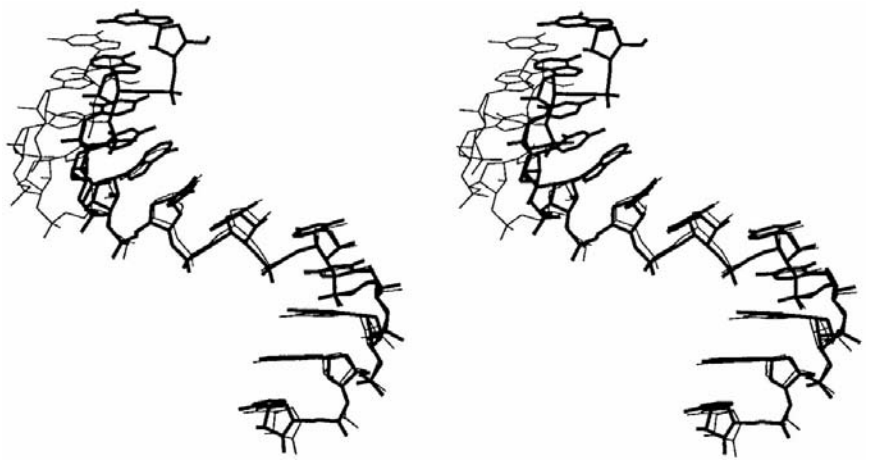

Figure 2
A stereoview diagram showing the different orientation of the four 5′-terminus overhang residues in two independent strands by superposing the octamer segment r(ACUUCGGU) in strand 1 (dark lines) and strand 2 (light lines).

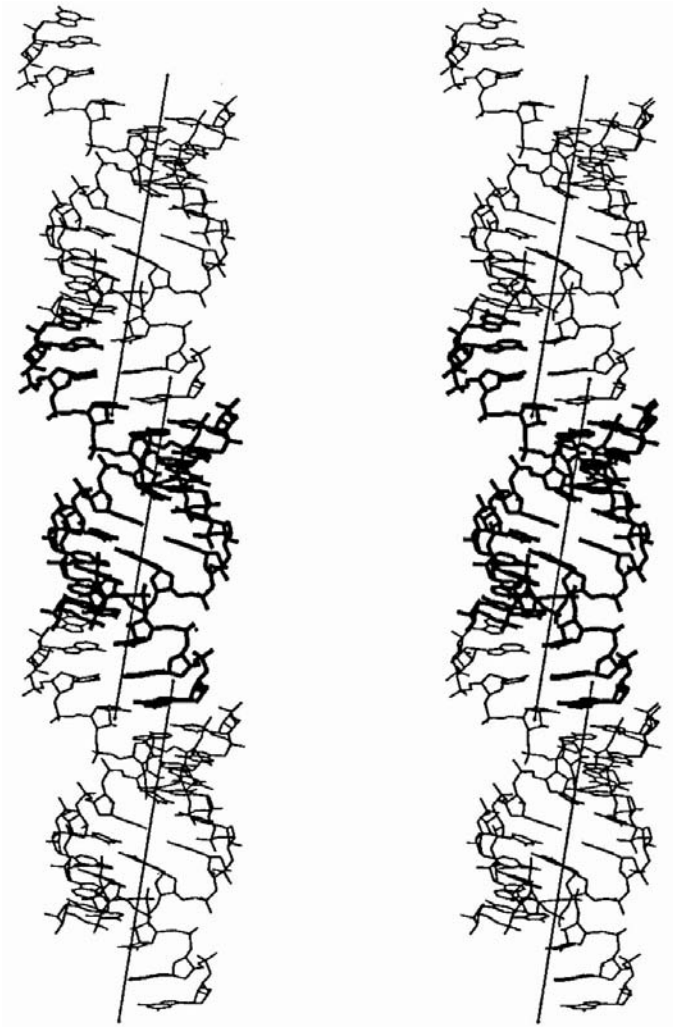

Figure 3
Stereo diagram of three duplexes of r(GAUCACUUCGGU) related by lattice translation along the *c* axis, which runs vertical. The helix axes of the segments are also shown, which are inclined at an angle of 9.8° to the *c* axis, resulting in the displacement of adjacent helix axes by 5.5 Å.

R_{work} and R_{free} to 0.214 and 0.261, respectively. 27 water molecules were identified from the difference electron-density maps. The final R_{work} and R_{free} for the 500 nucleotide atoms and 27 water O atoms were 0.188 and 0.228, respectively.

3. Results and discussion

3.1. Structural features

The RNA dodecamer r(GAUCACUUCGGU)$_2$ adopts an octamer duplex segment with four overhanging nucleotides in the 5′-terminus. The octamer segment at the 3′-end, r(ACUUCGGU), belongs to an A-form duplex and contains four central mispairs U·G, U·C, C·U and G·U flanked by two Watson–Crick base pairs. All the four overhang nucleotides are engaged in Watson–Crick base pairs with their counterparts from a symmetry-related duplex (Fig. 1*b*). These duplexes form a pseudo-continuous helical column by sharing the common tetrameric duplexes with nicks four and eight nucleotides apart. The two independent strands in the asymmetric unit are distinct structurally, with an r.m.s. deviation of 1.58 Å on superposition of all the atoms in the two strands. The major difference lies in the 5′-overhang residues (Fig. 2). However, the tetramer and octamer segments of the two independent strands are quite similar, with r.m.s. deviations of 0.53 and 0.60 Å, respectively. Despite the four consecutive mismatches at the center of the octamer, the helical axis is almost straight. However, the helical axis of the adjacent octamer duplexes

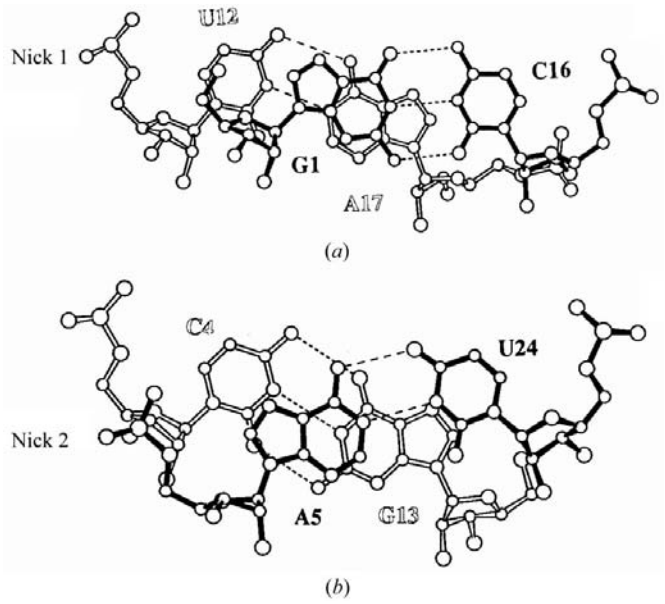

Figure 4
Stacking of the base pairs across the two nicks. (*a*) Nick 1, at which the helix axes are displaced by 5.5 Å. (*b*) Nick 2, at which the helix axes of the tetramer and the octamer are continuous; the stacking is more reminiscent of canonical duplexes.

deviates by a translation of about 5.5 Å (Fig. 3). The average values of the helical twist, rise, inclination and propeller twist are 32.4° (11.2 residues per turn), 2.63 Å, 15.6° and −9.8°, respectively.

The groove widths of a duplex can be measured by the shortest phosphorus–phosphorus distances across the groove, less 5.8 Å for the radii of the P atoms. The major-groove and minor-groove widths of the present structure are listed in Table 2. The major groove is narrow in the center, with a minimum of 3.8 Å for the groove width connecting the two phosphates of nucleotides in the octamer duplex (A5 and C18). The groove increases at both ends, with the 3′-end reaching a maximum of about 10.0 Å. This result indicates that the overhang nucleotides have a tremendous effect on the major-groove width. The minor-groove widths also show variations, ranging from 8.8 to 12.3 Å. The maximum value (12.3 Å) is near the two U·C mispairs, compared with the average value of 10.5 Å. The major-groove width and minor-groove width for fiber RNA are 4.1 and 11.3 Å, respectively.

Table 2
Major-groove widths and minor-groove widths of the present structure.

Phosphorus	Phosphorus	Groove widths (Å)
Major groove		
P(A2)	P(C21)	6.1
P(U3)	P(U20)	6.8
P(C4)	P(U19)	4.6
P(A5)	P(C18)	3.8
P(C6)	P(A17)	4.8
P(U7)	P(C16)	7.5
P(U8)	P(U15)	10.3
P(C9)	P(A14)	9.9
Minor groove		
P(U12)	P(U20)	8.8
P(G11)	P(C21)	10.5
P(G10)	P(G22)	12.3
P(C9)	P(G23)	10.4

All the sugars adopt the C3′-*endo* puckering conformation and the backbone torsion angles are in the favored range for the A-RNA duplex except for the residues G10, A17 and G22, which have α/γ as *trans*/*trans* (Sundaralingam, 1973). The difference between the torsion angles α/γ in A5 (*gauche*$^-$/*gauche*$^+$) and A17 (*trans*/*trans*) may be related to the different orientation of the two tetramer segments in the two independent strands. The junction twist angle across nick 1 is only 8° (Fig. 4*a*) where the terminal O(5′) of G1 stacks over U12 of a symmetry-related strand. The twist angle at nick 2 is 28° and the helix also unwinds (Fig. 4*b*).

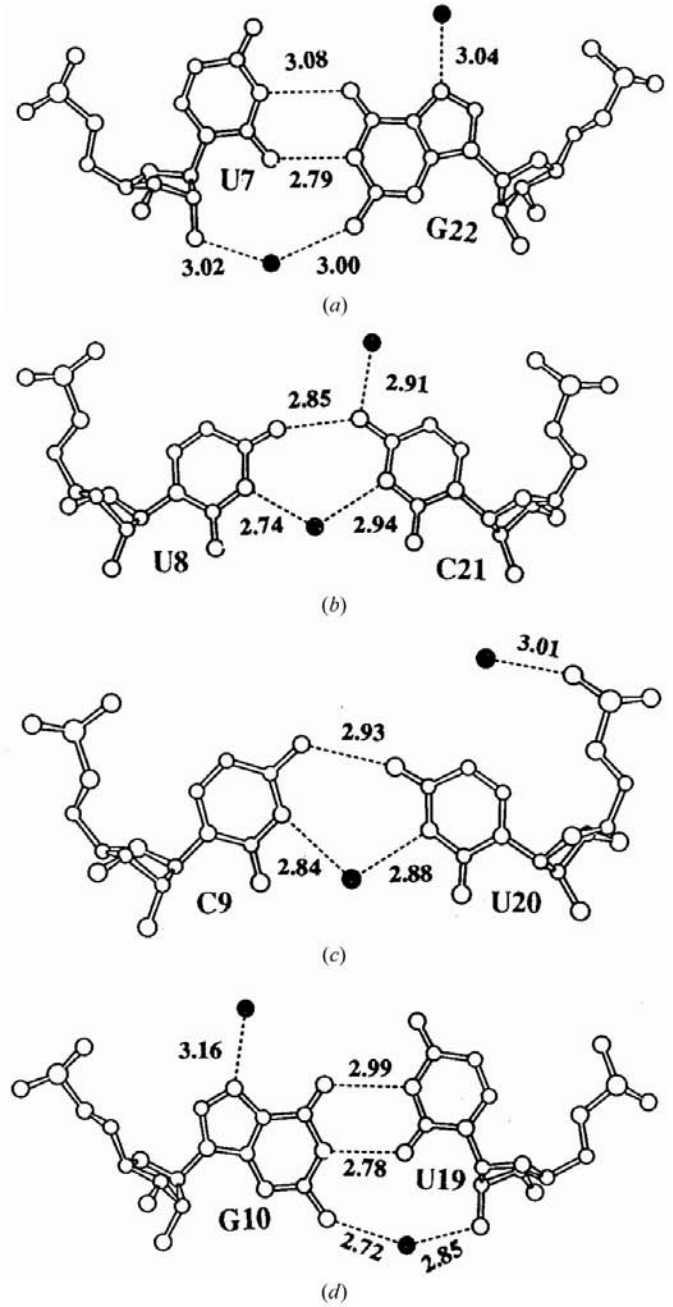

Figure 5
The geometry and hydration of the four mismatches. Water molecules are shown as filled circles and the numbers indicate the hydrogen-bonding distance between the two atoms connected by dashed lines. (*a*) U7·G22, (*b*) U8·C21, (*c*) C9·U20 and (*d*) G10·U19.

3.2. Mismatches

There is a stretch of four consecutive base mismatches at the center of the octamer segment (ACUUCGGU). The two G·U mismatches adopt the usual wobble conformation with two hydrogen bonds, N1(G)···O2(U) and O6(G)···N3(U) (Figs. 5*a* and 5*d*). The conformation is further stabilized by a water bridge connecting N2 of G and O2′ of U. The C1′—C1′ distance is 10.4 Å for the G·U wobble pairs. The purine is displaced towards the minor groove and the pyrimidine towards the major groove, which has also been observed in earlier crystal structures (Biswas *et al.*, 1997; Biswas & Sundaralingam, 1997; Holbrook *et al.*, 1991; Cruse *et al.*, 1994). The translation, sometimes together with rotation, of the bases is often observed for other wobble pairs such as I·m5C, I·U and A·C (Ramakrishnan & Sundaralingam, 1995; Pan, Mitra, Sun *et al.*, 1998; Pan, Mitra & Sundaralingam, 1998). The two tandem U·C mismatches have one base–base hydrogen bond N4(C)—O4(U) and a water bridge connecting N3(U) and

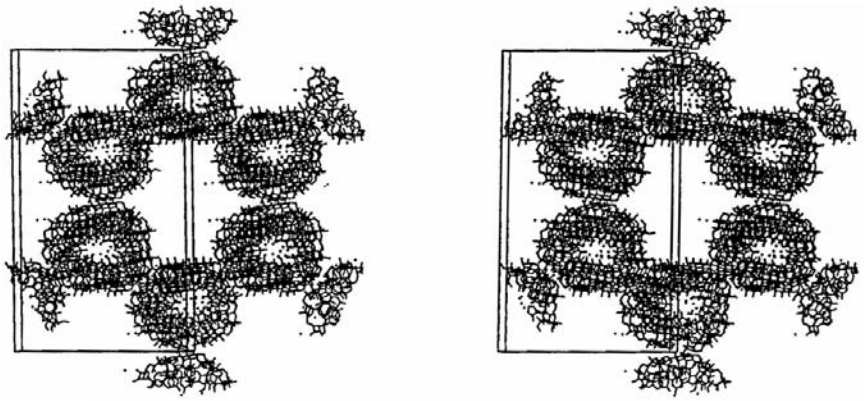

Figure 6
Packing diagram viewed down the *c* axis. There are hardly any ordered water molecules in the channel.

N3(C) (Figs. 5*b* and 5*c*). The average C1′—C1′ distance is 11.7 Å, nearly 1 Å longer than for canonical RNA. The hydrogen-bonding interactions observed in the present four consecutive mismatches are similar to those in the dodecamer r(GGACUUCGGUCC) (Holbrook *et al.*, 1991) and the nonamer r(GCUUCGGC)d^{Br}U (Cruse *et al.*, 1994), illustrating the stability of the conformations for a stretch of these four mismatches in RNA double-helical structures.

The two G·U wobbles exhibit high twist angles (35.2 and 39.0°) with the adjacent C·U mismatches, while the twist angle between the two C·U pairs is low (29.6°). Both G·U and C·U mismatches show very small propeller twists. The bridging water molecules in the four mismatches are well ordered with relatively low thermal parameters and appear to be an integral part of the mismatched base pairs (Holbrook *et al.*, 1991).

3.3. Crystal packing and hydration

Duplexes in the present structure form helical columns running throughout the crystal parallel to the *c* axis. Six such helical columns with pseudo-sixfold symmetry are shown in Fig. 6. The hexagonal columns enclose a channel running through the crystal, with an elliptical cross-section of 20 × 27 Å. Intercolumn hydrogen bonds occur between the O2′ atoms of A2, U7 and G11 in one column and O2′(C21), O3′(U24) and O2P(A17), respectively, in an adjacent column. In addition, other interactions between 2′-OH groups and water molecules stabilize the packing. In summary, the four consecutive mismatches are accommodated in the double helix with minimal distortion and the helix axis is nearly straight.

There are 27 water molecules in total located in the present structure. The major groove and the minor groove are equally hydrated, with ten water molecules in the groove. In the two grooves, water molecules cluster around the four consecutive mismatches, as observed in many mismatches in RNA duplexes (Pan & Sundaralingam, 1999). This result implies the importance of the interactions of water molecules in stabilizing the conformation of mismatches. There are seven water molecules associated with phosphate groups, with four interacting with O2P and three connecting with O1P. However, there are no regular patterns observed for the hydration in phosphate groups. There is only one water bridge connecting the two O2Ps of phosphate groups.

We gratefully acknowledge support by NIH grant GM-17378 and Ohio Regents Hayes Development Fund for partial support for the purchase of the R-AXIS IIc. We also thank the Board of Regents of Ohio for an Ohio Eminent Scholar Chair and Endowment to MS.

References

Baeyens, K. J., DeBondt, H. L. & Holbrook, S. R. (1995). *Nature Struct. Biol.* **2**, 56–62.
Baeyens, K. J., DeBondt, H. L., Pardi, A. & Holbrook, S. R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12851–12855.
Biswas, R. & Sundaralingam, M. (1997). *J. Mol. Biol.* **270**, 511–519.
Biswas, R., Wahl, M. C., Ban, C. & Sundaralingam, M. (1997). *J. Mol. Biol.* **267**, 1149–1156.
Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.
Cheong, C., Varani, G. & Tinoco, I. Jr (1990). *Nature (London)*, **346**, 680–682.
Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prangé, T. & Kennard, O. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4160–4164.
Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., deBear, J. S., Gough, G. R., Gilham, P. T. & Moras, D. (1989). *J. Mol. Biol.* **209**, 459–474.
Holbrook, S. R., Cheong, C., Tinoco, I. Jr & Kim, S. (1991). *Nature (London)*, **353**, 579–581.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pan, B., Mitra, S. N., Sun, L., Hart, D. & Sundaralingam, M. (1998). *Nucleic Acids Res.* **26**, 5699–5706.
Pan, B., Mitra, S. N. & Sundaralingam, M. (1998). *J. Mol. Biol.* **283**, 977–984.
Pan, B. & Sundaralingam, M. (1999). *Int. J. Quantum Chem.* **75**, 275–287.
Ramakrishnan, B. & Sundaralingam, M. (1995). *Biophys. J.* **69**, 553–558.
Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995). *J. Mol. Biol.* **250**, 327–332.
Sundaralingam, M. (1973). *Conformations of Biological Molecules and Polymers*, edited by E. Bergmann & B. Pullman, pp. 417–455. New York: Academic Press.
Varani, G., Cheong, C. & Tinoco, I. Jr (1991). *Biochemistry*, **30**, 3280–3289.
Woese, C. R., Winker, S. & Gutell, R. R. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.