Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of an RNA duplex with an unusual G·C pair in wobble-like conformation at 1.6 Å resolution

M. Perbandt,[a] M. Vallazza,[b] C. Lippmann,[b] Ch. Betzel[a]* and V. A. Erdmann[b]

[a]Institute of Medical Biochemistry and Molecular Biology, University Hospital Eppendorf, c/o DESY Building 22a, Notkestrasse 85, 22603 Hamburg, Germany, and [b]Institute of Biochemistry, Free University of Berlin, Thielallee 63, 14195 Berlin, Germany

Correspondence e-mail: betzel@unisgi1.desy.de

Received 19 July 2000
Accepted 10 November 2000

NDB Reference:
r(CUGGGCGG)·r(CCGC-CUGG), AR0012.

The structure of the RNA duplex r(CUGGGCGG)·r(CCGC-CUGG) has been determined at 1.6 Å resolution and refined to a final *R* factor of 18.3% ($R_{\rm free}$ = 24.1%). The sequence of the RNA fragment resembles domain E of *Thermus flavus* 5S rRNA. A previously undescribed wobble-like G·C base-pair formation is found. Owing to the observed hydrogen-bond network, it is proposed that the cytosine is protonated at position N3. The unusual base-pair formation is presumably strained by intermolecular interactions. In this context, crystal packing and particular intermolecular contacts may have direct influence on the three-dimensional structure. Furthermore, this structure includes two G·U wobble base pairs in tandem conformation, with the purines forming a so-called 'cross-strand G stack'.

1. Introduction

The 5S rRNA is located in the central protuberance of the large ribosomal subunit near the peptidyl transferase, but its precise function in protein biosynthesis remains unclear (Erdmann *et al.*, 1971; Hartmann *et al.*, 1988; Horne & Erdmann, 1972; Moore, 1995). X-ray structure analysis will support a more detailed understanding of the 5S rRNA structure–function relationship. The recently published structures of *Escherichia coli* 5S rRNA domains with and without binding proteins (Correll *et al.*, 1997; Dallas & Moore, 1997; Stoldt *et al.*, 1999; Lu & Steitz, 1999) have already provided some detailed insights about RNA–protein interactions in particular 5S rRNA regions. Moreover, the recently published atomic structure of the large ribosomal subunit from *Haloarcula marismortui* reveals the three-dimensional jigsaw puzzle of the ribosomal proteins and RNAs, including the 5S rRNA (Ban *et al.*, 2000), at 2.4 Å resolution. However, the sequences for *H. marismortui* and *T. flavus* 5S rRNA differ remarkably in the region which is examined and described in this report. Unfortunately, extensive attempts to crystallize the complete uncomplexed *T. flavus* 5S rRNA in our laboratory have resulted in crystals which diffract only to about 7.5 Å (Lorenz *et al.*, 2000). The difficulties in obtaining suitable crystals for high-resolution structures of large RNA molecules have encouraged us to consider synthesis of selected motifs of the 5S rRNA according to the secondary-structure model, as shown in Fig. 1 (Betzel *et al.*, 1994; Correll *et al.*, 1997; Perbandt *et al.*, 1998). The structure analysis and description we present here includes the formation of a wobble-like G·C base pair observed for the first time for an RNA duplex, which leads us to propose a protonated state of the involved cytosine. The protonation of cytosine at N3 in RNA oligonucleotides has already been observed at pH 6.8 (Leitner *et al.*, 1998) and it has also been shown that the protonation of cytosine

bases is essential in forming DNA triple helices (Leitner *et al.*, 1998). Before this, the only base pairs observed to have a single protonated cytosine were those of the form (C·C$^+$); the first was in the crystal structure of acetyl cytosine (Marsh *et al.*, 1962), but subsequently this form was seen in polycytidylic RNA (Akrinimski *et al.*, 1976; Hartmann & Rich, 1963). Single C·C$^+$ base pairs have also been observed in the structure of the complex of ribo- (CpA) and proflavin (Westhof *et al.*, 1980; Westhof & Sundaralingam, 1980), in a tetrameric DNA structure (Gehring *et al.*, 1993), and in the intercalated four-stranded cytosine-rich metazoan telomere (Kang *et al.*, 1995).

2. Material and methods

2.1. Synthesis and crystallization

The solid-phase synthesis of oligonucleotides was performed on a PCR-Mate EP model 391 DNA synthesizer (Applied Biosystems) with 2′-*O*-triisopropylsilyl-protected phosphoramidite synthons. Incubation of the controlled-pore glass beads in a 1 ml moisture-free ammonium–ethanol solution [3:1(*v*/*v*)] at 328 K for 24 h removed the protecting groups on the phosphates and on the exocyclic amino groups of the bases. Subsequent incubation at room temperature for 72 h in 1.1 *M* tetrabutylammonium fluoride (TBAF) in tetrahydrofuran–absolute ethanol [10:1(*v*/*v*)] removed the 2′-hydroxyl protecting group. The deprotected oligoribonucleotides were desalted by ion-exchange chromatography on a Qiagen Tip 500 column (QIAGEN) and the RNA was eluted with 2 *M* triethylammonium acetate (TEAAc) pH 7.0. Samples were evaporated to dryness and resuspended in distilled water. The full-length product was purified by reversed-phase HPLC on an ODS C18 Beckman Ultrasphere column (4.6 × 250 mm) heated at 318 K. The nucleotide composition was reanalysed by mass spectrometry and it was verified unambiguously that each strand of the duplex contains only one uridine base. The exact concentrations of the single-stranded RNA oligonucleotides were determined by means of the Lambert–Beer equation, applying the molar extinction coefficients: oligo1 nt79–86 = 62 167 and oligo2 nt90–97 = 60 171 l mol^{-1} cm^{-1}). A 1:1 ratio of both strands at a single-strand concentration of 1 m*M* was incubated at 363 K for 10 min and subsequently cooled to room temperature. The crystals were grown by the hanging-drop vapour-diffusion method using 0.25 m*M* RNA, 12 m*M* spermine tetrahydrochloride, 80 m*M* KCl, 20 m*M* $BaCl_2$ in 40 m*M* sodium cacodylate buffer pH 6.0 and 10%(*v*/*v*) MPD and equilibrating against 1 ml 40% MPD in the reservoir at room temperature. Data were collected at 100 K with synchrotron radiation (λ = 1.0 Å) at ELETTRA (Trieste, Italy). The diffraction data were indexed and integrated using the program *DENZO* (Otwinowski & Minor, 1997). The space group was assigned to be the rhombohedral space group *R*32 with one duplex in the asymmetric unit. The data-collection statistics are summarized in Table 1.

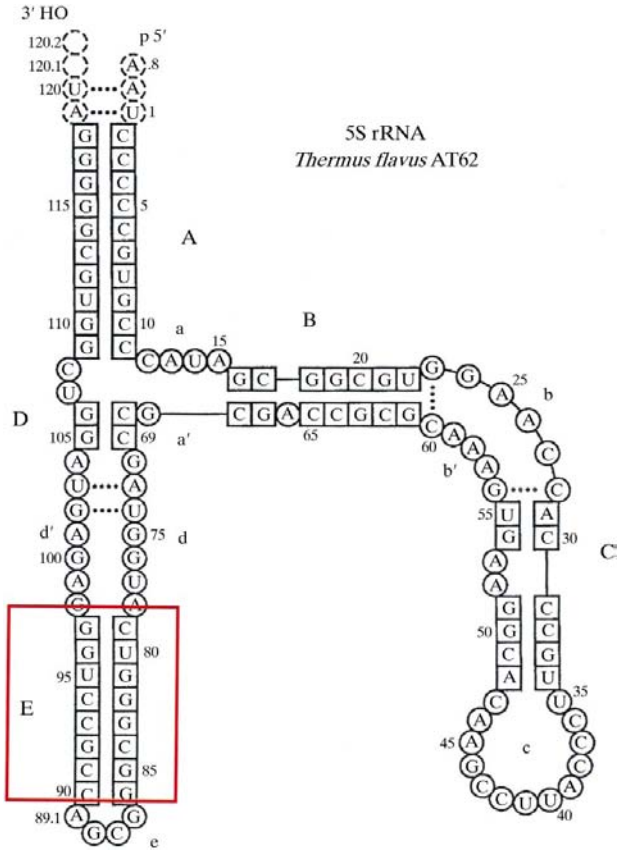

Figure 1
Schematic diagram of the 5S rRNA. Domain E is additionally indicated by a box. Roman numerals designate helical domain regions.

Table 1
Crystal data and refinement statistics.

Space group	*R*32
Unit-cell parameters (Å)	
a = *b*	41.9
c	127.1
Resolution range (Å)	20–1.6
No. of unique reflections	5994
Total No. of reflections	43171
Overall redundancy	4.9
R_{symm} (final shell)	0.066 (0.291)
Completeness (%)	96.1
R factor (R_{free})	0.183 (0.241)
R.m.s. deviation from ideality	
Bond lengths (Å)	0.023
Bond angles (°)	1.4
Torsion angles (°)	4.6
Improper angles (°)	1.9
Average *B* factors (Å^2)	
Nucleic acid atoms	27.9
Water O atoms	38.7

2.2. Structure solution and refinement

The structure was solved by molecular-replacement methods using the program *AMoRe* (Navaza, 1987). A rotation–translation solution was obtained for one molecule in the asymmetric unit using a 8 bp RNA fragment of the crystal structure of the synthetic r[U(UA)$_6$A]$_2$ oligoribonucleotide (Dock-Bregeon *et al.*, 1988) as a search model. For the refinement, the programs *REFMAC* (Collaborative Compu-

tational Project, Number 4, 1994) and *X-PLOR* 3.1 (Brünger, 1992*a*) were applied and a subset of the reflections (5%) was utilized for the R_{free} calculations (Brünger, 1992*b*). A rigid-body refinement of the initial solution reduced the R factor to 37.8% and R_{free} to 44.7% (from initial values of 47.2 and 55.4%, respectively) for all data in the resolution range 10–2.5 Å. A few cycles of restrained refinement reduced the R and R_{free} values further to 32.1 and 41.9%, respectively. At this stage, all bases were subsequently corrected according to the real duplex sequence and further cycles of refinement were performed. A total of 80 water molecules and one Ba^{2+} ion were located as peaks above the 3σ level in a $F_o - F_c$ difference map which had peaks above 1σ in the $2F_o - F_c$ density map. For regions of particular interest (*e.g.* the wobble-like G·C base pair), simulated-annealing omit maps were calculated to verify the structure without bias. The following cycles of individual isotropic B-factor refinement reduced the R value to 22.4% and R_{free} to 27.1%. Finally, anisotropic B-factor refinement reduced the R and R_{free} values to 18.3 and 24.1%, respectively, for all data in the resolution range 20–1.6 Å. The entire structure is well ordered as shown in Fig. 2. The refinement statistics are also summarized in Table 1.

3. Results and discussion

3.1. Overall structure description

The RNA duplex adopts a standard A-form conformation, with all the helical parameters resembling those of an A-RNA as summarized in Table 2. The program *NEWHEL* (Dickerson, 1989) was used to calculate the local helical parameters. There are 11 residues per helical turn, with an average helical twist of 32.8°. The helical twist angle for the G·U wobble step is 44.8° and for the step after the 'unusual' G84·C92 base pair 35.7°, which are both significantly higher than the average twist angle. The average rise per residue is 2.6 Å. The average propeller twist is −10.7° and the displacement of the base pairs towards the minor groove is 4.7 Å. All of the ribose rings of the present structure are either in C3′-*endo* or 2′-*exo*-3′-*endo* puckering conformations and all α torsion angles are in the g^- conformation and correlated with γ torsion angles which are in the g^+ conformation (data not shown).

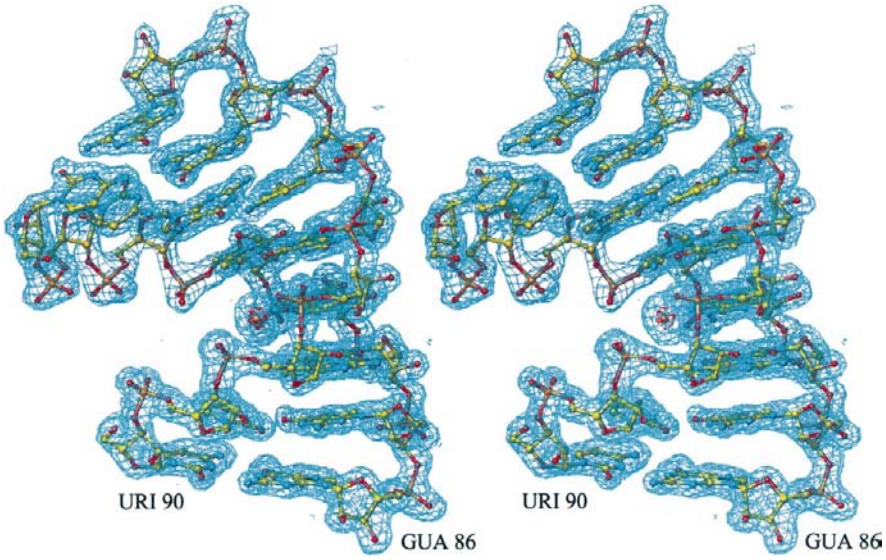

Figure 2
Stereoview of the final structure superimposed on a $2F_o - F_c$ difference electron-density map contoured at the 1σ level. In a deep pocket of the major groove the bound Ba^{2+} ion is shown with electron density contoured at the 10σ level.

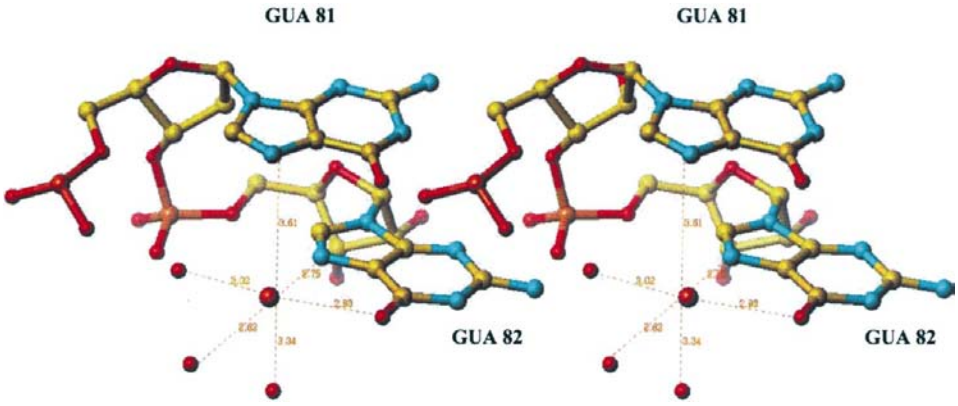

Figure 3
Stereoview of the bound Ba^{2+} ion in octahedral coordination geometry. The Ba^{2+} ion is complexed by two guanines and three water molecules; in particular, the N7 and O6 of guanine are involved. All distances are included.

3.2. Hydration and ions

A total of 80 independent solvent molecules and one Ba^{2+} ion were identified in the crystal structure. 19 solvent molecules are located in the minor groove, 47 in the major groove and 14 are located along the sugar backbone. The backbone and the 2′-hydroxyl groups are particularly well hydrated. The 2′-hydroxyl groups form direct or water-mediated contacts with various regions within the duplex and are also involved in extensive intermolecular interactions, with a preference for guanine N2 atoms and backbone O atoms. The Ba^{2+} ion has a octahedral coordination geometry and three direct contacts towards two purines. The other three coordination sites are occupied by water molecules inside the

Table 2
Helical parameters.

Calculated with the program *NEWHEL* (Dickerson, 1989).

	Twist (°)	Rise (Å)	*X*-Disp (Å)	Prop. (°)
C79·G97			−3.8	−6.7
	30.8	3.3		
U80·G96			−3.8	−17.7
	44.8	2.0		
G81·U95			−4.9	−7.8
	28.7	2.8		
G82·C94			−5.2	−7.9
	29.4	2.4		
G83·C93			−5.5	−10.2
	27.9	3.1		
C84·G92			−5.3	−10.7
	35.7	2.0		
G85·C91			−4.7	−13.0
	32.7	2.7		
G86·C90			−4.7	−12.4
Average	32.8	2.6	−4.7	−10.8
SD	5.5	0.5	0.6	3.3

major groove. As observed in other RNA crystal structures for the coordination of Mg^{2+} ions (Correll *et al.*, 1997; Pan *et al.*, 1993; Pley *et al.*, 1994; Scott *et al.*, 1995; Cate & Dounda, 1996), the Ba^{2+} is also coordinated by the N7 and O6 of a guanine (Fig. 3).

The position of the Ba^{2+} ion does not however appear to be fully occupied; it refined well with an occupancy of 0.5. Moreover, the coordination contacts are relative lengthy, as shown in Fig. 3, which indicates the prospect of an easy substitution by solvent water. K^+ and Na^+ were also refined for this position. However, they were ruled out because even with full occupancy and the lowest *B* value the subsequently calculated $F_o - F_c$ difference electron density showed significant residual density. Consequently, a half-occupied Ba^{2+} was confirmed.

3.3. Wobble-like G·C$^+$ base pair

The crystal structure exposed an unusual G·C base pair in non-Watson–Crick formation. The three hydrogen bonds expected for a standard G·C base pair could not be observed for the base pair C84·G92. According to distance criteria, two hydrogen bonds GuaO6·CytN3 (2.87 Å) and GuaN1·CytO2 (3.01 Å) are formed. As a result, the base-pair step C84·G92/G85·C91 stacks with a high twist angle of 35.7° and a low rise angle of 2.0 Å (Table 2). The hydrogen-bonding pattern for this region is unusual, but the atoms involved, which are clearly verified by simulated-annealing omit difference electron-density maps, show typical hydrogen-bond distances (Fig. 4). In particular, the GuaO6·CytN3 distance supports the evidence that CytN3 is protonated. There are two possibilities to explain this special hydrogen-bond distance: either the cytosine is in the imino-form, which is known to be quite unlikely (Saenger, 1987), or the cytosine has an additional proton at N3. The protonation of cytosine bases in oligonucleotides is not unusual and under weakly acidic conditions more than two pH units above the pK_a of monomeric cytosine bases (about 4.2) is feasible in principle (Leitner *et al.*, 1998). For deoxycytidylic acid polymers, the protonated cytosine structure was stable up to pH 7.0. The stability of the crystal lattice may have raised the p*K* for haemoprotonation even higher than pH 7.0 (Kang *et al.*, 1995). As mentioned before, protonated cytosines have been previously observed in crystal structures, but this is the first time it has been observed in a cytosine-to-guanine (G·C$^+$) base pair. Furthermore, this unusual base-pair formation is additionally stabilized by neighbouring water molecules and intermolecular RNA–RNA interactions: CytN4 is stabilized by a hydrogen bond to a water molecule and GuaN2 by a hydrogen bond to a 2′-hydroxyl group of a symmetry-related molecule, as shown in Fig. 4. The figure also indicates that the crystal packing obviously has a significant influence on the base-pair formation of these two nucleotides.

Unfortunately, it is not possible to compare this part of the *T. flavus* 5S rRNA with the *H. marismortui* 5S rRNA (Ban *et al.*, 2000), as the base-pairing scheme and the sequence are different for both organisms. Furthermore, at this region the *H. marismortui* 5S rRNA has one additional and bulged uridine.

3.4. Tandem G·U base pairs

The tandem G·U base-pair formation in the structure is the most frequent 'mismatch formation' found in the ribosomal RNA (Wu *et al.*, 1995). Three types of adjacent G·U pairs have been described (Gautheret *et al.*, 1995). The present G·U motif belongs to type 1, which is seven times more frequently observed than types 2 and 3 (Wu *et al.*, 1995). G·U base pairs in principle provide three potential hydrogen-bond acceptors, therefore offering more recognition diversity for protein–RNA interaction. The crystal structure of this RNA duplex reveals the conformation of a type 1 tandem wobble G·U base pair (Fig. 5). In contrast to the three hydrogen bonds expected in G·C Watson–Crick base pairs, the G·U wobble base pairs have only two hydrogen bonds: GuaN1·UriO2 and GuaO6·UriN3. A type 1 tandem of G·U base pairs, which is flanked on both sides by Watson–Crick bases, was first observed in the X-ray structure of domain A of *T. flavus* 5S rRNA (Betzel *et al.*, 1994). Subsequently, two other X-ray structures (Correll *et al.*, 1997; Biswas *et al.*, 1997) and one NMR structure (Dallas & Moore, 1997) confirmed this structural motif. All of these structures show identical conformation and stabilization by solvent waters and therefore this motif can be regarded as structurally conserved. Also in this region, the U·G/G·U step has a relatively high twist angle of 44.8° and a low rise of 2.0 Å (Table 2), resulting in a so-called 'cross-strand G stack', because the G of one G·U base pair stacks with the G of the neighbouring G·U base pair, which appears from the opposite strand. The hydration pattern of the wobble G·U base pair is highly conserved and can be compared with those found in other X-ray structures including single G·U pairs (Auffinger & Westhof, 1998; Müller *et al.*, 1999).

4. Conclusions

The principles of RNA folding and conformation are important features determining RNA structure and function. Therefore, it is of great interest to indicate and recognize structural RNA motifs that are responsible for the mediation of intermolecular contacts and are essential for RNA folding and function. The frequent appearance of G·U wobble pairs in ribosomal RNA can be partly rationalized on the basis of structural information revealed by several crystal structures of RNA during the last few years (Betzel *et al.*, 1994; Correll *et al.*, 1997; Dallas & Moore, 1997; Biswas *et al.*, 1997). The stacking of two guanine bases and the presence of three potential hydrogen-bond acceptors in each G·U pair may be a distinguishing structural feature recognized by proteins prior to their interaction with rRNA. The overall conformation of those motifs is found to be identical in different RNA structures. Direct biochemical evidence for the involvement of water molecules in the immediate vicinity of G·U in minor-groove interactions has been also presented (Henderson *et al.*, 1998). For linear DNA molecules, it was recently shown that the additional protonation of cytosine bases is essential for the formation of triple helices (Leitner *et al.*, 1998) and for that reason it should be expected to be an element of major importance in folding of nucleic acids. In the structure presented here, only the protonated state of Cyt84 provides a sufficient explanation for the unusual conformation of the G·C base pair. As mentioned before, this particular formation is probably achieved by intermolecular interactions arising from the crystal packing effects. It will be interesting to see whether this novel base pair will be also observed in other structures, including for example the structure of the complete *T. flavus* 5S RNA.

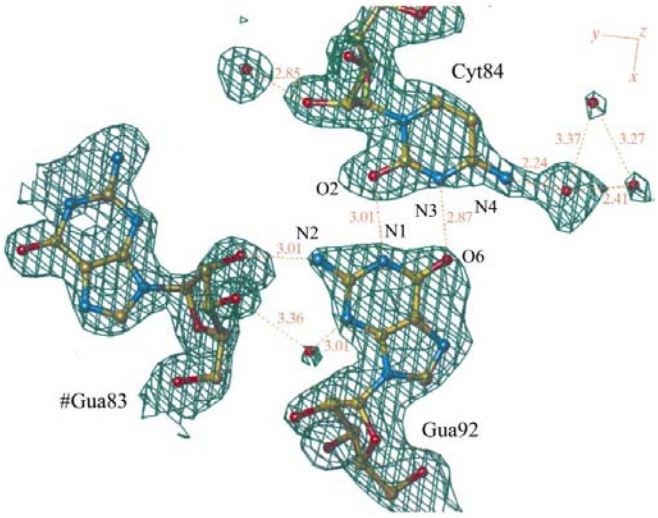

Figure 4
The G92·C84 base pair in wobble-like arrangement is superimposed by the corresponding $F_o - F_c$ simulated-annealing omit difference electron density shown in green at a level of 2.0σ. The displayed water molecules were also omitted for calculating the difference electron density. Instead of the three expected hydrogen bonds, only two [GuaO6·CytN3 (2.9 Å) and GuaN1·CytO2 (3.0 Å)] are identified according to distance criteria. This unusual wobble-like base pairing is further stabilized by RNA–RNA interactions with a symmetry mate (indicated by #) as well as surrounding water molecules.

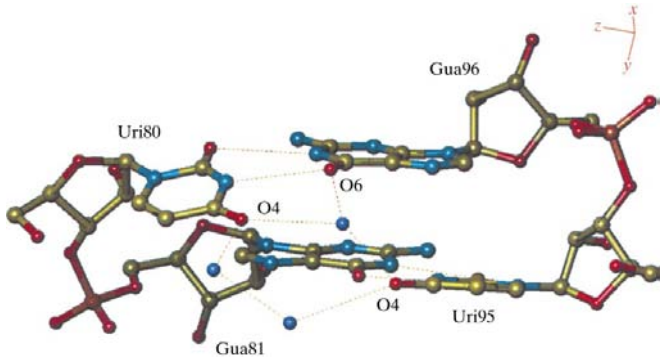

Figure 5
The wobble formation of two adjacent G·U base pairs is stabilized by three 'structural' water molecules.

This work has been supported by the BMBF *via* the network of RNA Technology (RiNA), the Deutsche Forschungsgemeinschaft (SFB 344-D6), the Deutsche Agentur für Raumfahrtangelegenheiten and the Fonds der Chemischen Industrie eV.

References

Akrinimski, E. O., Chandrasekaran, R. & Leslie, A. G. W. (1976). *J. Mol. Biol.* **106**, 735–748.
Auffinger, P. & Westhof, E. (1998). *J. Biomol. Struct. Dynam.* **16**(3), 693–707.
Ban, N., Nissen, P., Hansen, J., Morre, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–920.
Betzel, Ch., Lorenz, S., Fürste, J. P., Bald, R., Zhang, M., Schneider, T. R., Wilson, K. & Erdmann, V. A. (1994). *FEBS Lett.* **351**, 159–164.
Biswas, R., Wahl, M. C., Ban, C. & Sundaralingam, M. J. (1997). *J. Mol. Biol.* **267**, 11449–11456.
Brünger, A. T. (1992*a*). *X-PLOR. Version* 3.1. *A System for Crystallography and NMR.* New Haven: Yale University Press.
Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–474.
Cate, J. H. & Doudna, J. A. (1996). *Structure*, **4**, 1221–1229.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). *Cell*, **91**, 705–712.
Dallas, A. & Moore, P. B. (1997). *Structure*, **5**, 1639–1653.
Dickerson, R. E. (1989). *Nucleic Acids Res.* **17**, 1797–1803.
Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Moras, D., deBear, J. S., Gough, G. R., Gilham, P. T. & Johnson, J. E. (1988). *Nature (London)*, **335**, 375–378.
Erdmann, V. A., Fahnestock, K. H. & Nomura, M. (1971). *Proc. Natl Acad. Sci. USA*, **68**, 2932–2936.
Gautheret, R. R., Konings, D. & Gutell, R. (1995). *RNA*, **1**, 807–814.
Gehring, K., Leroy, J. L. & Gueron, M. (1993). *Nature (London)*, **363**, 561–565.
Hartmann, K. A. Jr & Rich, A. (1963). *Nature (London)*, **198**, 725–728.
Hartmann, R. K., Vogel, D. W., Walker, R. T. & Erdmann, V. A. (1988). *Nucleic Acids. Res.* **16**, 3511–3524.
Henderson, B. S., Beuning, P. J., Shi, J. P., Bald, R., Fürste, J. P., Erdmann, V. A., Musier-Forsyth, K. & Schimmel, P. (1998). *J. Am. Chem. Soc.* **120**, 9110–9111.
Horne, J. & Erdmann, V. A. (1972). *Mol. Genet.* **119**, 337–344.

Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 3874–3878.
Leitner, D., Schröder, W. & Weis, K. (1998). *J. Am. Chem. Soc.* **20**, 7123–7124.
Lorenz, S., Perbandt, M., Lippmann, C., Moore, K., DeLucas, L. J., Betzel, Ch. & Erdmann, V. A. (2000). *Acta Cryst.* D**56**, 498–500.
Lu, M. & Steitz, T. A. (1999). *Proc. Natl Acad. Sci. USA*, **97**, 2023–2028.
Marsh, R. E., Bierstedt, R. & Eichhorn, E. L. (1962). *Acta Cryst.* **15**, 310–316.
Moore, P. B. (1995). *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Synthesis*, edited by R. A. Zimmermann & A. E. Dahlberg, pp. 199–236. Boca Raton, FL, USA: CRC Press.
Müller, U., Schübel, H., Sprinzl, M. & Heinemann, U. (1999). *RNA*, **5**, 670–677.
Navaza, J. (1987). *Acta Cryst.* A**43**, 645–653.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pan, T., Long, D. M. & Uhlenbeck, O. C. (1993). *The RNA World*, edited by R. F. Gesteland & J. F. Atkins, pp. 271–302. New York: Cold Spring Harbour Laboratory Press.
Perbandt, M., Nolte, A., Lorenz, S., Bald, R., Betzel, Ch. & Erdmann, V. A. (1998). *FEBS Lett.* **429**, 211–215.
Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). *Nature (London)*, **372**, 68–74.
Saenger, W. (1987). *Principles of Nucleic Acids*, pp. 113–114. Berlin: Springer.
Scott, W. G., Finch, J. T. & Klug, A. (1995). *Cell*, **81**, 991–1002.
Stoldt, M., Wöhnert, J., Ohlenschläger, O., Görlach, M. & Brown, L. R. (1999). *EMBO J.* **81**(22), 6508–6521.
Westhof, E., Roa, S. T. & Sundaralingam, M. (1980). *J. Mol. Biol.* **142**, 331–361.
Westhof, E. & Sundaralingam, M. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 1852–1856.
Wu, M., McDowell, J. A. & Turner, D. H. (1995). *Biochemistry*, **32**, 3204–3211.