FEBS Letters 351 (1994) 159–164

FEBS 14447

Crystal structure of domain A of *Thermus flavus* 5S rRNA and the contribution of water molecules to its structure

Ch. Betzel[b], S. Lorenz[a], J.P. Fürste[a], R. Bald[a], M. Zhang[a], Th.R. Schneider[b], K.S. Wilson[b], V.A. Erdmann[a,*]

[a]*Institut für Biochemie, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany*
[b]*EMBL c/o DESY, Notkestraße 85, 22603 Hamburg, Germany*

Received 13 June 1994; revised version received 22 July 1994

Abstract

This is the first high resolution crystal structure of an RNA molecule made by solid phase chemical synthesis and representing a natural RNA. The structure of the domain A of *Thermus flavus* ribosomal 5S RNA is refined to $R = 18\%$ at 2.4 Å including 159 solvent molecules. Most of the 2′-hydroxyl groups as well as the phosphate oxygens are involved either in specific hydrogen bonds in intermolecular contacts or to solvent molecules. The two U-G and G-U base-pairs are stabilized by H-bonds supplied via three water molecules to compensate for the lack of base-pair hydrogen bonds. The structure shows for the first time in detail the importance of highly ordered internal water in stablizing an RNA structure.

Key words: rRNA; Solid phase synthesis; Synchrotron radiation; Solvent structure

1. Introduction

Although ribonucleic acids are involved in a large number of biological functions little is known about their structures at atomic resolution. The reason for this predicament is found in the fact that so far only a few RNA molecules could be crystallized for X-ray structural determinations and that the RNA molecules are usually too large for NMR analysis. The study presented here focuses on the structure of ribosomal 5S RNA, which is an essential component of large ribosomal subunits [1,2]. The fact that 5S rRNAs are known to specifically interact with several ribosomal proteins [3] make these structural studies also to a model system for other RNA–protein complexes. Our extensive attempts to crystallize ribosomal 5S RNAs have led to crystals which currently diffract to 8 Å [4], insufficient to warrant their structural determination. On the other hand developments in chemical synthesis of RNA domains give new possibilities in studies of RNA structure and function. So far only two X-ray structures with simplified sequences of the synthetic oligomer $[U(U\text{-}A)_6A]_2$ [5,6] and the GGAC-UUCGGUCC duplex [7] have been reported in detail. We have turned to the chemical synthesis of the various structural domains of 5S rRNAs in order to attempt their crystallization and to determine their structures at the atomic level. We report here results on domain A of *Thermus flavus* 5S rRNA, as shown schematically in Fig. 1, giving not only the RNA structure, but also the first comprehensive localization of a large number of solvent molecules associated with an RNA.

*Corresponding author. Fax: (49) (30) 838 6403.

2. Materials and methods

2.1. Chemical synthesis and crystallization

The domain A of 5S rRNA of *Thermus flavus* was prepared by solid phase chemical synthesis [8]. Crystals of the dodecamer suitable for X-ray analysis were obtained by vapour diffusion followed by repeated seeding as reported previously [9]. The space group was found to be $P4_3$ with unit cell parameters of $a = b = 30.10$ Å and $c = 86.80$ Å. The packing parameter V_M is 2.6 Å^3/Dalton for one helical fragment per asymmetric unit.

2.2. Data collection and structure solution

Two crystals were used to collect data up to 2.38 Å resolution. On a conventional sealed tube X-ray source with MoK_α radiation and a graphite monochromator, data to 3.0 Å were collected using a MAR 180 mm image plate detector. A second data set was collected to 2.38 Å resolution with synchrotron radiation using a MAR 300 mm image plate detector at the EMBL beam line X11 at DORIS/DESY. The storage ring was operated in main user mode at 4.7 GeV and 20–40 mA. The wavelength was 0.92 Å. The images of the first data set were processed using the program DENZO [10]. The reduced data set contains 1,477 reflections and shows a completeness of 94%. The R_{merge} defined as $R(I) = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$ is 6.6% . The images collected using synchrotron radiation were processed using a modified version of the XDS program package [11]. The unique data up to 2.3 Å contain 2,170 reflections with an R_{merge} of 3.7%. Finally the two data sets were scaled and merged to 2,477 reflections up to 2.4 Å. The resulting completeness for all data up to 2.4 Å is 83.5% and for all data up to 2.3 Å is 77.3% due to the limited completeness of only 50% in the resolution shell between 2.3 and 2.4 Å caused by radiation damage. The structure was solved by molecular replacement using the coordinates of the synthetic RNA helix $[U(UA)_6A]_2$ [6] as starting model and program AMORE [12]. The rotation function gave a clear solution for the orientation of the molecule. In the following translation search the space group was assigned to be $P4_3$ using all data in the resolution range of 8.0–3.0 Å and giving a R-value of 41%. Refinement was started using the program X-PLOR [13], simulated annealing with a slow cool protocol using data up to 3.0 Å. The R-value dropped to 26%. The true sequence was fitted to the electron density and 64 solvent water molecules were introduced. In later stages the conventional restrained least- squares method using the program NUCLSQ [14] was applied. The refinement was carried out in steps of model building, adding solvent molecules and refinement. The final R value was 18% including 159 solvent molecules. The r.m.s. deviations from ideal values for bond distances, angle distances, base planarity, chiral volumes and bond angles were 0.025 Å, 0.068 Å,

SSDI 0014-5793(94)00834-5

0.014 Å, 0.036 Å and 3.7° respectively. The mean B value was 23.9 Å^2. Attempts to collect data to higher resolution are in progress.

3. Results and discussion

The final structure presented in Fig. 2 is a right-handed A-type double helix with two strands defined as A and B according to Fig. 1. Strand A consists of 245 non-hydrogen atoms and strand B of 263. The structure includes a total of 159 solvent molecules. 49 were internal (labelled I) located in the interior of the duplex; 68 were external (labelled E) involved in H-bonds forming the first hydration shell around the surface of the duplex and 42 were crystal waters filling the spaces between symmetry related molecules.

The helical parameters are reported in Table 1 in comparison to other known structures. The average helix rotation angle of the dodecamer duplex is 33.3° and the rise per residue is 2.43 Å compared to values of 32.7° and 2.81 Å in the canonical RNA-A helix determined by X-ray fibre diffraction [15] and to 33.5° and 2.41 Å in tRNA [16]. These values indicate that domain A has a higher similarity to the tRNA structure than to the synthetic fragments. The P-P intrastrand distances have mean values of 5.91 Å and 5.77 Å for strands A and B, respectively, with r.m.s values of 0.46 Å and 0.34 Å and with a maximum difference of 1.3 Å. This fluctuation further indicates the flexibility of this natural structure (sequence) and its ability to adjust and adapt to local geometric restraints.

The internal water molecules are shown in Fig. 2 using the space filling representation to demonstrate the compact filling of the interior of the helix with the bulk of solvent in the narrow and the deep major groove and a more regular water structure at both ends of the RNA. In contrast to the synthetic oligomer $[U(UA)_6A]_2$, analysis of the present helix shows all solvent positions in the interior. It may well be that the relatively short data collection time allows the localization of a more ordered cluster, which is susceptible on a longer time scale to exchange with the solvent in the crystal.

The specific involvement of solvent molecules in interstrand H-bonds is also summarized in Table 2. All H-bonds of the internal solvent molecules, labeled with I, are included. Most of the O-1-P oxygens are involved in this network and most of the centered waters form bridges between the two strands of the duplex. These interstrand H-bonds are mediated via two to three solvent molecules and form complex H-bond networks. At the two ends of the domain some of the external waters are included in the interaction scheme.

We are also seeing a number of intermolecular contacts involving 2′ hydroxyl groups, as previously reported by Wang et al. [17], (details will be presented elsewhere). This highly structured network of H-bonds

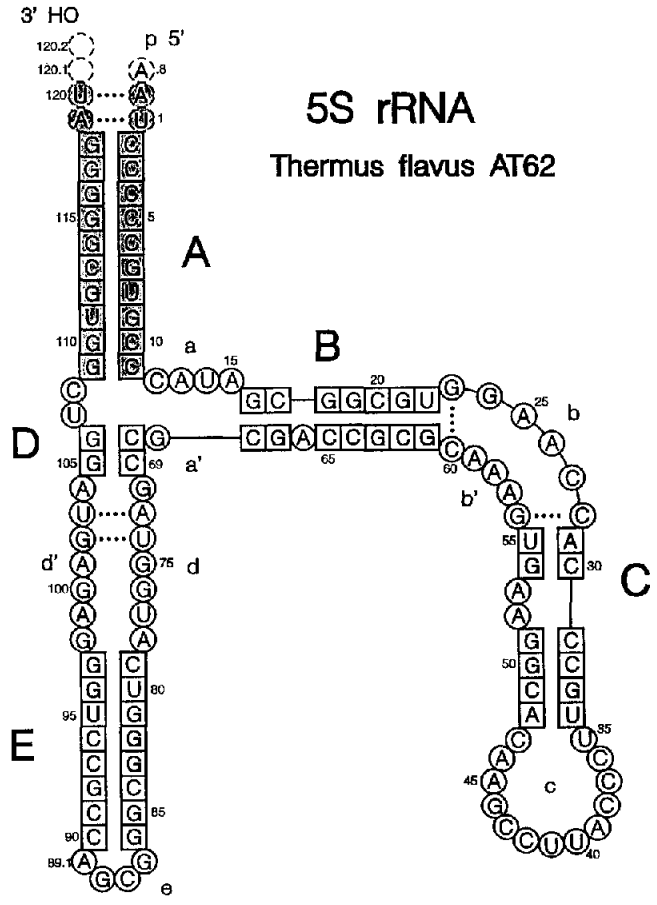

Fig. 1. Secondary structure of *Thermus flavus* 5S rRNA. The individual domains are marked A through E. The section of the helical domain A is boxed. The chemical synthesis of RNA was performed on a DNA synthesizer (Applied Biosystems 391) using self-prepared 2′-triisopropylsilyl-protected phosphoramidites [8].

is responsible for maintaining the three-dimensional structure. The packing is shown in Fig. 3. The molecules interact as a set of crossed cogwheels by direct intermolecular contacts or via crystal solvent molecules. The O2′ hydroxyls are important in supporting the packing interactions. A similar crystal packing has been found and described before for DNA [18]. Such specific contacts may also provide a model for RNA–RNA interactions.

Table 1
Average torsion angles and helical parameters; calculated using the program HELIX [23]

	Twist	Rise	Roll	X-Dsp	Prop
Domain A	33.3	2.43	−1.5	4.5	13.1
(S.D.)	(4.7)	(0.45)	(5.8)	(0.5)	(3.8)
Fibre [15]*	32.7	2.81	−9.0	4.4	13.8
$[U(UA)_6]_2$ [6]	33.2	2.78	9.4	3.6	19.1
GGACUUCGGUCC [7]	32.1	2.93			
tRNA (mono) [16]	33.2	2.50	−5.2	4.4	14.2

*Values taken from [14] for a canonical RNA-A helix as determined by fibre diffraction.

Table 2
H-bonds of internal solvent molecules

Atom	Atom	Strand	Dist.
OW I1	O4 URI	B 3	2.64
	OW I3		2.67
OW I2	OW I43		3.48
	O4 URI	A 9	2.89
	N7 GUA	A 10	3.32
OW I3	OW I1		2.67
	O4 URI	A 9	3.29
	O6 GUA	B 4	2.77
	OW I33		2.68
OW I4	N4 CYT	A 6	3.01
	OW I33		2.93
OW I5	O1P URI	B 3	3.24
	OW I9		2.63
	OW I19		3.25
	OW I36		2.66
OW I6	OW I7		3.57
	OW I21		2.98
	O1P CYT	A 4	3.31
OW I7	OW I22		2.92
	OW I6		3.57
OW I8	O1P GUA	B 4	3.08
	O2P GUA	B 4	2.87
	OW I20		3.03
	OW I37		3.33
OW I9	O1P CYT	A 5	3.11
	OW I10		2.99
	OW I5		2.63
OW I10	O1P CYT	A 6	3.21
	OW I9		2.99
OW I11	OW I22		3.58
OW I13	OW E63		3.13
	OW I43		3.43
	OW E33		3.50
OW I14	O1P CYT	A 11	3.44
	OW E33		2.82
	OW E55		2.93
OW I16	N7 GUA	B 8	3.42
	OW I29		3.38
OW I17	O1P GUA	B 9	2.80
	OW I41		3.49
OW I18	OW I25		3.11
	N7 GUA	B 10	2.85
	O1P GUA	B 10	2.62
OW I19	OW I5		3.25
	N7 GUA	B 4	3.44
	O1P GUA	B 4	3.55
OW I20	O1P CYT	A 3	3.37
	OW I8		3.03
	O1P URI	A 2	3.53
OW I21	OW E24		3.10
	OW I6		2.98
	O1P CYT	A 3	3.53
	OW I37		2.48
OW I22	OW I11		3.58
	OW I24		3.37
	OW I7		2.92
OW I23	O6 GUA	B 6	3.09
	N7 GUA	B 6	2.40
	OW I49		2.77
OW I24	O6 GUA	B 8	3.04
	OW I22		3.37

Table 2 *(continued)*
H-bonds of internal solvent molecules

Atom	Atom	Strand	Dist.
OW I25	OW I18		3.11
	N 6 ADE	B 11	3.04
	OW I42		3.07
	OW I26		2.50
OW I26	OW I25		2.50
	OW I28		3.24
	O4 URI	A 2	2.91
OW I28	OW I29		3.13
	OW I26		3.24
	OW I38		2.90
OW I29	OW I16		3.38
	OW I28		3.13
	OW I31		2.90
	OW I30		3.02
OW I30	OW I29		3.02
	O1P GUA	B 6	2.52
	OW I31		2.72
	O1P GUA	B 7	3.55
OW I31	OW I29		2.90
	O1P GUA	B 6	3.32
	OW I49		2.82
	OW I30		2.72
	OW I40		2.82
OW I32	O1P GUA	A 8	2.62
	OW I35		3.14
	O1P CYT	A 7	3.48
OW I33	OW I3		2.68
	OW I4		2.93
OW I34	OW E23		2.69
OW I35	O1P GUA	A 8	2.91
	OW I32		3.14
OW I36	O1P GUA	B 2	2.63
	O1P URI	B 3	2.79
	OW I5		2.66
OW I37	OW E24		3.14
	OW I21		2.48
	O1P GUA	B 4	2.69
	OW I8		3.33
OW I38	OW I28		2.90
	OW I40		2.64
	O1P URI	A 2	3.04
OW I40	OW I31		2.82
	OW E25		3.46
	OW I38		2.64
OW I41	OW I17		3.49
	OW I27		3.49
OW I42	OW I25		3.07
	OW E29		3.21
OW I43	OW I13		3.43
	OW I2		3.48
OW I49	OW I23		2.77
	O1P CYT	B 5	2.59
	OW I31		2.82

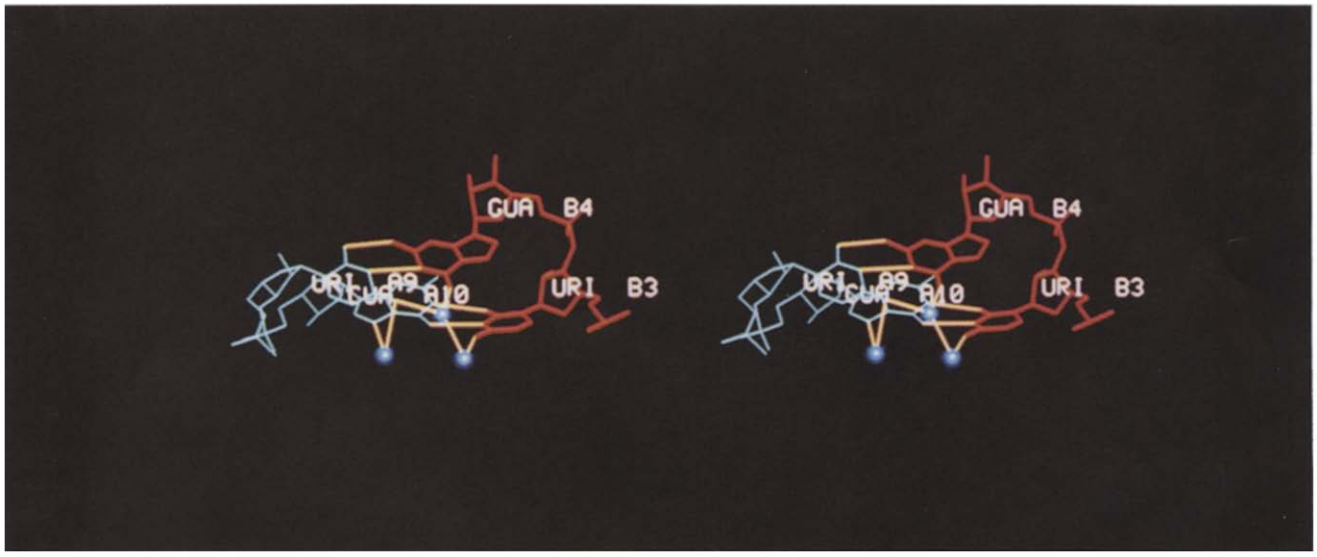

Fig. 2. The two strands of the helical domain A are shown in green and red. Internal waters are represented in blue using their Van der Waals radii to indicate the way they fill the interior of the helix.

Three internal water molecules OW1, OW2 and OW3 cluster around the unusual base pairs 9A, 10A and 3B, 4B, as shown in Fig. 4, to compensate the loss in base to base hydrogen bonds. This kind of solvent arrange-

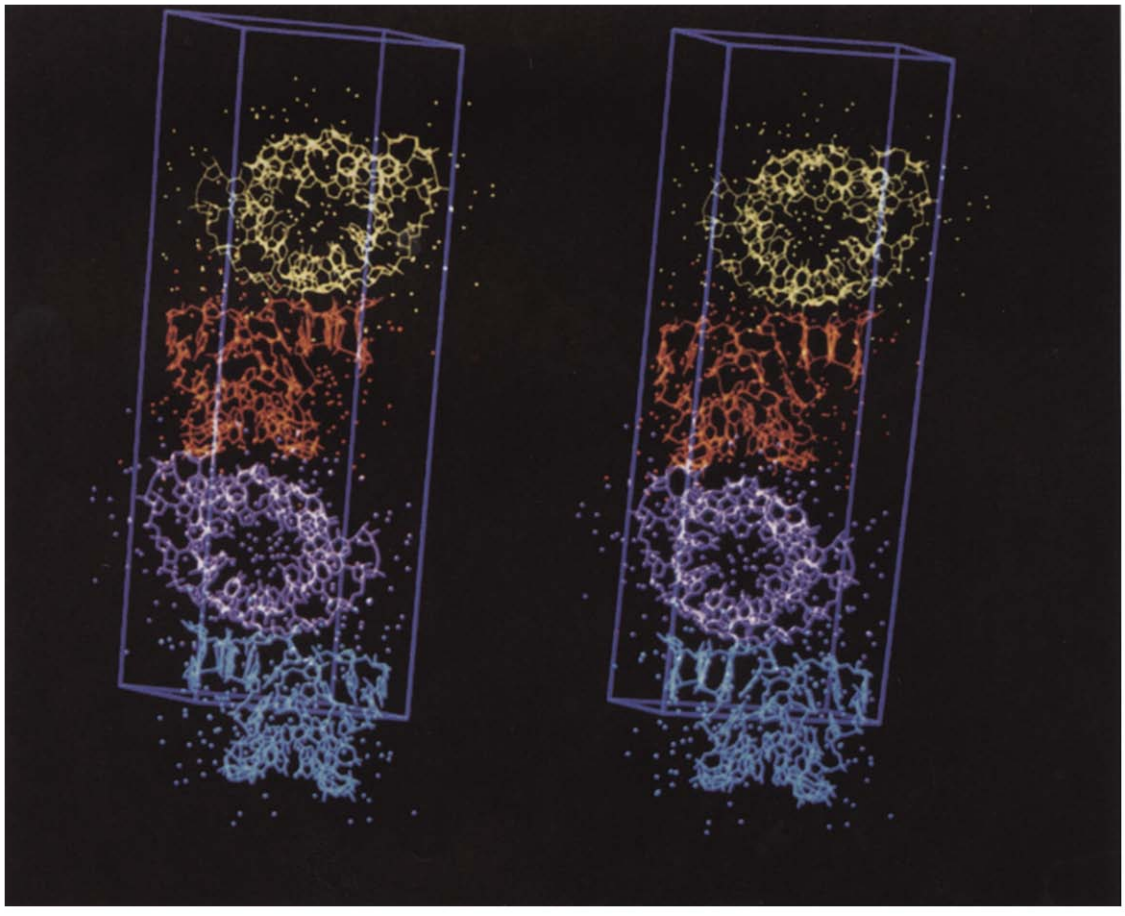

Fig. 3. Stereo view of the arrangement of molecules in the crystal lattice showing four symmetry related molecules. The individual molecules of domain A (in different colours) form some direct intermolecular packing contacts or via water molecules. The space between the molecules is filled by crystal solvent molecules.

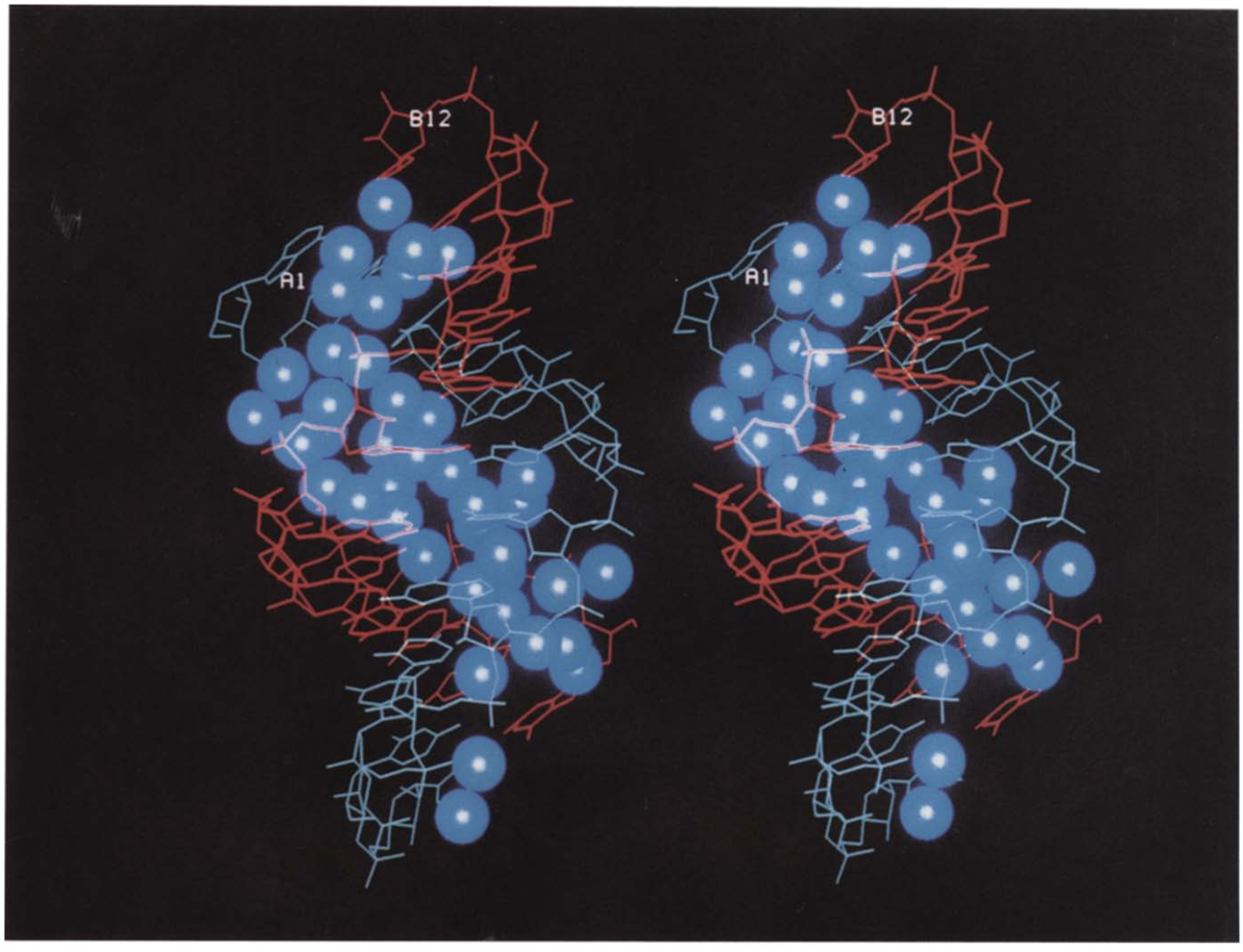

Fig. 4. Stereo view of the two wobble G–U base pairs including the internal waters OW1, OW2 and OW3. They rearrange the loss of H-bonds around the mismatched base pairs as found in crystal structures of DNA oligomers and tRNA structures [19,20]. The two sections of strands A and B are coloured in blue and red. The base to base and the additional H-bonds to the solvent molecules are shown as dashed lines.

ments have also been found in tRNA structures [19,20]. OW3 forms a triad-like H-bond bridge between opposite base pairs 9A and 4B and towards OW1 which forms an additional H-bond to O4 of Uracil 3B. OW2 supplies one intrastrand H-bond between the bases of 9A O4 and 10A N7. Thus this structural region, close to the bend of the intact 5S rRNA molecule, Fig. 1, is stabilized and the helical twist is close to the average value of 33.3°.

The mean *B*-value is 30.7 Å^2 for the internal water molecules. The external water molecules and the crystal lattice waters have slightly higher *B*-values. The mean for both groups is $B = 33.8$ Å^2. The highly structured internal waters show that they play a key role in RNA structure and are likely to be of prime importance for the interaction with proteins, structural changes of RNAs in connection with functions and in general biological functions.

In conclusion it is quite apparent that with the recent developments of the methodology to synthesize defined RNA molecules it will be possible to crystallize the molecules to determine their atomic structures by X-ray analysis [21,22].

Acknowledgements: The authors thank the Deutsche Forschungsgemeinschaft, the Bundesministerium für Forschung und Technologie and the Fonds der Chemischen Industrie E.V. for financial support. The X-ray crystallography work was performed at the EMBL c/o DESY in Hamburg, Germany.

References

[1] Erdmann, V.A., Fahnestock, K.H. and Nomura, M. (1971) Proc. Natl. Acad. Sci. USA 68, 2932–2936.
[2] Hartmann, R.K., Vogel, D.W., Walker, R.T. and Erdmann, V.A. (1988) Nucleic Acids Res. 16, 3511–3524.
[3] Horne, J. and Erdmann, V.A. (1972) Mol. Gen. Genet. 119, 337–344.
[4] Lorenz, S., Betzel, Ch., Raderschall, E., Dauter, Z., Wilson, K.S. and Erdmann V.A. (1991) J. Mol. Biol. 219, 399–402.
[5] Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Moras, D., deBear, J.S., Gough, G.R., Gilham, P.T. and Johnson, J.E. (1988) Nature 335, 375–378.

[6] Dock-Bregon, A.C., Chevrier, B., Podjanry, A., Johnson, J., deBear, J.S., Gough, G.R., Gilham, P.T. and Moras, D. (1989) J. Mol. Biol. 209, 459–474.
[7] Holbrook, S.R., Cheong, C., Tinoco, L. and Kim, S.H. (1991) Nature 353, 579 581.
[8] Bald, R., Brumm, K., Buchholz, B., Fürste, J.P., Hartmann, R.K., Jäschke, A., Kretschmer-Kazemi, F., Lorenz, S., Raderschall, E., Schlegl, J., Specht, T., Zhang, M., Cech, D. and Erdmann, V.A. (1992) in: Structural Tools for the Analysis of Protein-Nucleic Acid Complexes (Lilley, D.J., Heumann, H. and Suck, D., eds.) pp. 449–466, Basel-Birkhäuser-Verlag.
[9] Lorenz, S., Betzel, Ch., Fürste, J.P., Bald, R., Zhang, M., Raderschall, E., Dauter, Z., Wilson, K.S. and Erdmann, V.A. (1993) Acta Crystallogr. D49, 418–420.
[10] Otwinowski, Z. (1991) DENZO: A Film Processing Program for Macromolecular Crystallography, Yale University, New Haven.
[11] Kabsch, W. (1988) J. Appl. Crystallogr. 21, 916–918.
[12] Navaza, J. (1992) Proceedings of the Daresbury Study Weekend on Molecular Replacement.
[13] Brünger, A.T., Karplus, M. and Petsko, G.A. (1989) Acta Crystallogr. A45, 50–61.
[14] Westhof, E., Dumas, P. and Moras, D. (1985) J. Mol. Biol. 184, 119–145.
[15] Arnott, S., Hukins, D.W.L., Dover, S.D., Fuller, W. and Hodgson, A.R. (1973) J. Mol. Biol. 81, 107–122.
[16] Westhof, E. and Sundaralingam, M. (1986) Biochemistry 25, 4868–4878.
[17] Wang, A.H.J., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boeckel, S.A.A. and Rich, A. (1992) Nature 299, 601–604.
[18] Timsit, Y. and Moras, D. (1991) J. Mol. Biol. 221, 919–940.
[19] Westhof, E. (1987) Int. J. Biol. Macromol. 9, 186–192.
[20] Westhof, E. (1988) Annu. Rev. Biophys. Chem. 17, 125–144.
[21] Doudna, J.A., Grosshans, Ch., Gooding, A. and Kundrot, G.E. (1993) Proc. Natl. Acad Sci. USA 90, 7829–7833.
[22] Kim, R., Holbrook, E.L., Jancarik, J., Pandit, J., Weng, X., Bohm, A. and Kim, S.H. (1994) Acta Crystallogr. D50 290–292
[23] Fratini, A.V., Kopka, M.L., Drews, H.E. and Dickerson, R.E. (1982) J. Biol. Chem. 257, 14686–14707.