Structure
with Folding & Design

Editors: Wayne A Hendrickson, Carl-Ivar Brändén and Alan R Fersht

Vol 8 No 5 15 May 2000

Current Biology Publications

RNA in signal recognition, and beyond

Form and function in modern biology

Crystal structure of the Ffh and EF-G binding sites in the conserved domain IV of *Escherichia coli* 4.5S RNA

Luca Jovine[1,2*†], Tobias Hainzl[1], Chris Oubridge[1], William G Scott[3], Jade Li[1], Titia K Sixma[2], Alan Wonacott[4], Tadeusz Skarzynski[4] and Kiyoshi Nagai[1*]

Background: Bacterial signal recognition particle (SRP), consisting of 4.5S RNA and Ffh protein, plays an essential role in targeting signal-peptide-containing proteins to the secretory apparatus in the cell membrane. The 4.5S RNA increases the affinity of Ffh for signal peptides and is essential for the interaction between SRP and its receptor, protein FtsY. The 4.5S RNA also interacts with elongation factor G (EF-G) in the ribosome and this interaction is required for efficient translation.

Results: We have determined by multiple anomalous dispersion (MAD) with Lu^{3+} the 2.7 Å crystal structure of a 4.5S RNA fragment containing binding sites for both Ffh and EF-G. This fragment consists of three helices connected by a symmetric and an asymmetric internal loop. In contrast to NMR-derived structures reported previously, the symmetric loop is entirely constituted by non-canonical base pairs. These pairs continuously stack and project unusual sets of hydrogen-bond donors and acceptors into the shallow minor groove. The structure can therefore be regarded as two double helical rods hinged by the asymmetric loop that protrudes from one strand.

Conclusions: Based on our crystal structure and results of chemical protection experiments reported previously, we predicted that Ffh binds to the minor groove of the symmetric loop. An identical decanucleotide sequence is found in the EF-G binding sites of both 4.5S RNA and 23S rRNA. The decanucleotide structure in the 4.5S RNA and the ribosomal protein L11–RNA complex crystals suggests how 4.5S RNA and 23S rRNA might interact with EF-G and function in translating ribosomes.

Addresses: [1]MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, England, [2]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands, [3]Department of Chemistry and Biochemistry and The Center for the Molecular Biology of RNA, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, CA 95064, USA and [4]Glaxo Wellcome Research and Development, Medicines Research Centre, Stevenage, Hertfordshire SG1 2NY, England.

[†]Present address: Department of Biochemistry and Molecular Biology, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029-6574, USA.

*Corresponding authors.
E-mails: jovinl02@doc.mssm.edu
kn@mrc-lmb.cam.ac.uk

Key words: EF-G, Ffh, non-canonical base pairs, 4.5S RNA, RNA–protein recognition, signal recognition particle

Received: **27 January 2000**
Revisions requested: **3 March 2000**
Revisions received: **13 March 2000**
Accepted: **15 March 2000**

Published: **2 May 2000**

Structure 2000, **8**:527–540

0969-2126/00/$ – see front matter

Introduction

The signal recognition particle (SRP) is a ubiquitous ribonucleoprotein complex that is essential for GTP-dependent protein translocation into the eukaryotic endoplasmic reticulum (ER) lumen or the prokaryotic periplasmic space [1,2]. Phylogenetic comparison of SRP RNAs has led to the division of the RNA into four structural domains (I–IV), of which only domain IV (also referred to as helix 8) is found in all homologues [3,4]. The eukaryotic SRP consists of a 7S RNA and six proteins, named SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72 after their predicted molecular weight. SRP54 is a GTPase [5] that binds to domain IV of 7S RNA and plays a key role by interacting with the ribosome, the emerging signal sequence and the SRP receptor (SR) at the ER membrane. The bacterial SRP, which represents the minimal evolutionary conserved particle, consists of 4.5S RNA (114 nucleotides in *Escherichia coli*) and a 48 kDa protein, Ffh or p48, which is homologous to SRP54 and binds to domain IV of 4.5S RNA [6,7]. The presence of 4.5S RNA increases the affinity of Ffh for signal peptides [8,9] and is required for the interaction of *E. coli* SRP with its receptor FtsY [10–12]. Because the 4.5S RNA/Ffh/FtsY and the 7S RNA/SRP54/SR systems are functionally interchangeable, the bacterial particle provides a good model for understanding SRP function and structure [13].

Evidence for a second, distinct role of 4.5S RNA in the cell originated from the observation that its gene is essential for viability in *E. coli* [14]. Depletion of 4.5S to lethal levels can be suppressed by mutations arising in the gene for EF-G, another GTPase, which lead to the co-sedimentation of 4.5S RNA with ribosomes [15,16]. The 4.5S RNA associates with the ribosome following translocation and prior to the release of uncharged tRNA from the E-site [17]. Partial depletion of 4.5S RNA has a general

inhibitory effect on peptide elongation before the processing of secreted proteins is compromised, and it is therefore likely that the lethal effects of 4.5S RNA depletion result from its role in protein synthesis [18]. There are four molecules of 4.5S RNA for every Ffh protein molecule in *E. coli*; three-quarters of the 4.5S RNA pool are thus available to perform a function distinct from its role in SRP [19]. Ribosomes exist in 25–100-fold excess over 4.5S RNA, so it is unlikely that 4.5S RNA takes part in every cycle of elongation [19]. EF-G footprints the 1067 and sarcin/ricin loops of 23S rRNA [20]. The 1067 loop includes a decanucleotide sequence shared with 4.5S RNA and almost invariantly conserved in both RNAs within the eubacteria [17]. A specific interaction between EF-G and 4.5S RNA was detected *in vivo* [21]. *In vitro*, an oligonucleotide representing the 1067 loop competes with 4.5S RNA for EF-G binding [21], and therefore EF-G can bind to either 4.5S RNA or the 1067 hairpin of 23S rRNA, but not to both simultaneously.

Complete understanding of the molecular mechanism of RNA–protein interactions requires structural knowledge not only of RNA–protein complexes, but also of their individual components. We have determined the crystal structure of the domain IV of 4.5S RNA at 2.7 Å resolution. The structures of 24-, 28- and 43-nucleotide fragments of 4.5S RNA have been studied using nuclear magnetic resonance (NMR) [22–24]. Structural information about RNA provided by NMR and crystallography is not redundant but complementary. Crystallography provides precise stereochemistry of RNA within the crystal lattice as well as its interaction with solvent molecules, whereas information provided by NMR is less precise but free from distortion by lattice forces [25,26]. More recently, the crystal structure of the complex between the RNA-binding domain of Ffh (M domain) and a fragment of 4.5S RNA has been determined at 1.8 Å resolution by Batey *et al.* [27]. Comparison of the free and protein-bound forms of 4.5S RNA gives important insights into the recognition mechanism between Ffh and 4.5S RNA. The EF-G binding sites within 4.5S RNA and 23S rRNA contain the same decanucleotide sequence [17]. This decamer RNA sequence is included in the recently reported crystal structure of ribosomal protein L11 in complex with a fragment of 23S rRNA [28,29]. We therefore compared the structure of the decamer sequence in 4.5S RNA with that in 23S rRNA to gain insights into the binding of EF-G to these two RNAs and the biological implications of these interactions.

Results

Structure determination

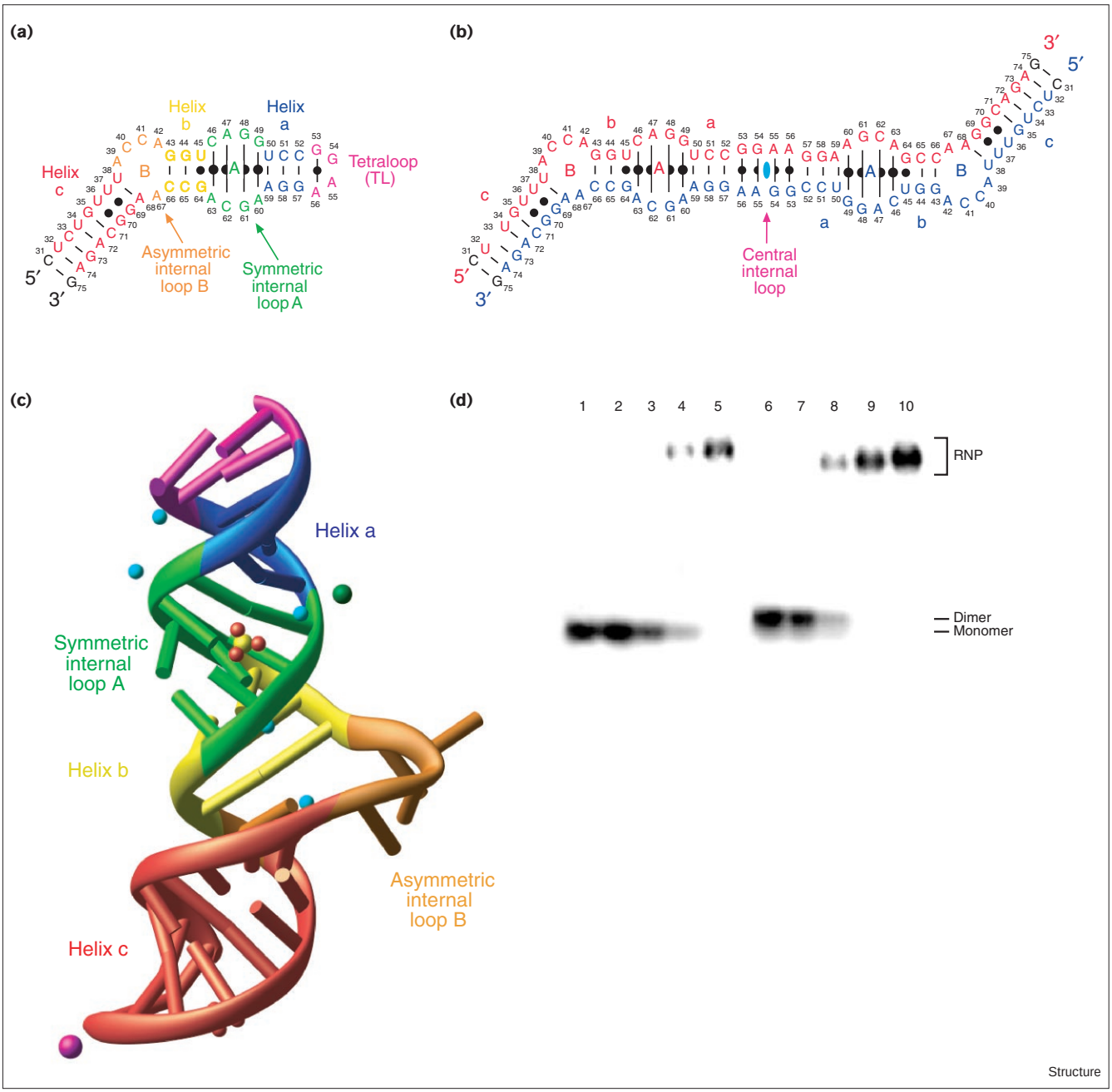
A 45-nucleotide fragment corresponding to domain IV of *E. coli* 4.5S RNA (nucleotides 31–75; henceforth referred to as '45 RNA'; Figure 1a) was prepared as described previously [30]. Trigonal crystals (space group $P3_221$) of this construct were initially obtained by hanging-drop vapour diffusion, using an ammonium sulphate condition of a sparse matrix screen based on that of Scott *et al.* [31]. After extensive optimisation, cryo-cooled native crystals diffracted anisotropically to 2.6–2.8 Å at synchrotron sources. Attempts to obtain derivatives by soaking heavy atoms into these crystals led, in all cases, to either non-isomorphous crystals or severe reduction or complete loss of diffraction. Similarly, diffraction-quality crystals could not be obtained with chemically synthesised RNA [32]. In order to solve these problems, the luminescence properties of Tb^{3+} ions [33,34] were exploited to estimate the minimal concentration of lanthanides that showed specific binding to domain IV RNA constructs in solution [32]. This measurement allowed us to produce a single-site Lu^{3+} derivative of 45 RNA which, although still not sufficiently isomorphous with native crystals to produce interpretable single isomorphous replacement (SIR) maps, has been used to solve the structure using multiple anomalous dispersion (MAD) methods. The structure has now been refined to a crystallographic R factor of 23.0% and an R_{free} of 24.5% (Tables 1 and 2).

The crystal structure is of a dimer

Secondary structure analysis, biophysical and biochemical measurements of 4.5S RNA indicate an extensively base-paired hairpin fold for domain IV (Figure 1a) [3,35,36]. Two 45 RNA strands associate to form a dimeric RNA molecule in the crystal (Figure 1b). This dimer essentially represents a palindromic form of the monomeric structure. Each strand of the dimer, corresponding to a single RNA molecule, spans two adjacent asymmetric units with the molecular dyad of the dimer lying on a crystallographic dyad. The biologically relevant structure is the monomer represented by nucleotides 31–54 of one molecule paired with nucleotides 55–75 of the second (Figure 1c). It is clear from Figures 1a and 1b that the only expected difference between the monomer and dimer forms of 45 RNA is that the tetraloop nucleotides of the monomer (G53–A56) have become a central internal loop in the dimer structure.

Prior to crystallisation, the annealed 45 RNA was found (using native polyacrylamide gel electrophoresis) to be exclusively in the monomeric hairpin form [32]. Hence the conversion from the monomeric to dimeric form must have taken place during crystallisation, with the dimer selectively packed into the crystal lattice. This accounts for the long period and the rather high temperatures required for nucleation. All RNAs that fold back and form hairpin loop structures are potentially able to form dimers using the same base-pairing scheme. Such conversions from the monomeric to dimeric form during crystallisation have been reported for an RNA that was designed to form a tetraloop [37], as well as the dimerisation–initiation site of genomic human immunodeficiency virus-1 (HIV-1) RNA [38] and helix 6 of 7S RNA [39].

Figure 1



(a) Diagram of the 45 RNA monomer, with individual helices and loops depicted in different colours. Numbering corresponds to full-length *E. coli* 4.5S RNA, lines indicate Watson–Crick base pairs and dots non-canonical pairs. The terminal C31–G75 base pair is not found in the natural RNA sequence and is indicated in black. **(b)** Diagram of the 45 RNA dimer observed in the crystal, with the two chains coloured in red and blue and the crystallographic dyad axis indicated in cyan. Conventions are as in (a). **(c)** Ribbon diagram of the biologically relevant 45 RNA structure, colour-coded as in (a). The lutetium ion, magnesium ion, sulphate ions and water molecules are shown in magenta, dark green, yellow/red and cyan, respectively. **(d)** Ffh protein binds to monomers and dimers of domain IV RNA with similar affinity. Equimolar amounts of [^{32}P]-labelled *E. coli* 4.5S RNA domain IV were annealed to form either monomeric or dimeric species (see Materials and methods). 1.0 μM monomer (lanes 1–5) and 0.5 μM dimer (lanes 6–10) were incubated with increasing concentrations of *E. coli* Ffh C406S protein [μM]: 0 (lanes 1 and 5), 0.5 (lanes 2 and 6), 0.75 (lanes 3 and 8), 1.0 (lanes 4 and 9) and 2.0 (lanes 5 and 10). The resulting RNA–protein complexes were analysed by gel mobility shift assay on 1% agarose. The positions of the free RNA monomers and dimers and of the assembled ribonucleoprotein complexes (RNP) are indicated. The lack of a supershift in the assay shows that only one molecule of Ffh binds the dimer, the binding of a second molecule being presumably prevented by steric hindrance.

Table 1

Crystallographic data collection and phasing.

	λ_1 (1.3366 Å; peak 1)	λ_2 (1.3369 Å; inflection)	λ_3 (0.9968 Å; remote 1)	λ_4 (1.0302 Å; remote 2)	λ_5 (1.3359 Å; peak 2)
Crystallographic data					
d (Å)	18.0–2.7	18.0–2.7	18.0–2.7	22.5–2.7	35.0–2.7
Reflections*	154065 (12332)	154248 (12338)	156009 (12471)	154595 (12371)	95847 (12328)
Redundancy†	4.3 (4.2)	4.3 (4.2)	4.2 (3.5)	4.2 (4.1)	2.6 (2.6)
Completeness (%)†	98.3 (100.0)	98.4 (100.0)	99.4 (98.4)	98.5 (100.0)	98.0 (99.9)
I/σ†	10.1 (1.6)	10.0 (1.8)	9.7 (1.2)	14.5 (2.1)	9.2 (1.7)
R_{sym} (%)†‡	8.2 (22.3; 73.2)	8.0 (21.6; 69.0)	13.0 (30.5; ND)§	7.6 (17.1; 59.3)	7.2 (15.5; 47.0)
Phasing power and FOM					
Phasing power (18.00–2.69 Å) LOC#	0.92	1.14	–	0.92	1.05
Phasing power (18.00–2.69 Å) FP#	1.73	1.82	0.81	1.35	1.37
Phasing power (2.81–2.69 Å) LOC#	0.51	0.33	–	0.78	0.90
Phasing power (2.81–2.69 Å) FP#	0.47	0.39	0.27	0.69	0.64
FOM (18.00–2.69 Å) LOC#	0.20	0.22	–	0.18	0.18
FOM (18.00–2.69 Å) FP#	0.26	0.26	0.13	0.23	0.22
FOM (2.81–2.69 Å) LOC#	0.07	0.05	–	0.15	0.14
FOM (2.81–2.69 Å) FP#	0.06	0.06	0.03	0.14	0.09

Space group = $P3_221$. Cell dimensions (Å): a = b = 69.697 c = 84.102. FOM weighted R value after density modification = 0.2215. Mean FOM after density modification = 0.7705. *Values in parentheses are the number of unique reflections. †Values in parentheses are for the high-resolution shell (2.80–2.70 Å); in the case of R_{sym}, values for the 3.04–2.91 Å shell are also reported. ‡$R_{sym} = \sum_h\sum_i|I_i(h)-\langle I(h)\rangle|/\sum_h\sum_i I_i(h)$, where $I_i(h)$ is the ith measurement and $\langle I(h)\rangle$ is the mean of all measurements of I(h) for Miller indices h. §Because of a software read-out problem, the data collected at the third wavelength was underexposed by a factor of ~4 compared with the other datasets. This is reflected in its higher R_{sym} values, particularly in the high-resolution shell whose R_{sym} value is out of range (ND). In spite of this limitation, the λ3 dataset provided significant phase information (the structure could be solved using only datasets λ1, λ2 and λ3) and was therefore included during phase calculation. #Lack of closure values are reported both for the real part of the anomalous difference from the reference wavelength (LOC) and between the Friedel pairs for each wavelength (FP). Phasing power = sqrt $(\Sigma_{hkl}F_H^2/\Sigma_{hkl}(F_{PH,\,obs} - F_{PH,\,calc})^2)$. The overall values of FOM are 0.53 and 0.27 for FOM (18.00–2.69 Å) and FOM (2.81–2.69 Å), respectively. Number of sites = 1.

We and others have shown that RNA molecules as small as 43 nucleotides (within the 45 RNA fragment) are bound by Ffh protein with the same affinity as is full-length 4.5S RNA [22,27,32] and can stimulate the GTP-hydrolysing activity of Ffh and FtsY in a ternary complex (TH, unpublished observations) [27,32]. Furthermore, domain IV constructs can substitute full-length 4.5S RNA *in vivo* [27]. The Ffh binding assay of the monomeric and dimeric forms of these shortened constructs shows that they are bound with similar affinities (Figure 1d), demonstrating that the essential RNA features recognised by the protein are not affected by conversion to the dimeric form and thus the structure we determined is biologically significant.

Description of the overall structure

The overall structure of a half of the 45 RNA dimer is shown in Figure 1c. Within this structure, base stacking is continuous except for the four nucleotides on the 5′ strand of the asymmetric internal loop B, which protrude from the duplex with their bases stacking either against others in the loop or against those of a symmetry-related molecule. The continuous stacking results in a molecule that is essentially a straight rod. Close inspection of the structure nevertheless reveals that it is not a simple duplex; many non-Watson–Crick base pairs are observed of which the interactions with adjacent pairs give rise to an intriguing fold.

The symmetric loop A

Within the stem regions a, b and c, as indicated in Figure 1b, base–base interactions are either of the standard Watson–Crick type or of the G•U wobble type [40–43]. All the bases in loop A form non-canonical base pairs (Figure 2). G64 and U45 form a standard wobble G•U base pair (Figure 3a) whereas A63 and C46 form an unusual sheared base pair involving the 2′OH group of C46. The exocyclic amino group (N6) of A63 hydrogen bonds with the ribose 2′OH as well as O2 of C46 (Figure 3a). As a result, this base pair has a large shear and shows extensive stacking with the G64•U45 base pair. The Watson–Crick face of A63 and the 2′OH group of C46 are displayed in the minor groove, with the N4 of the latter nucleotide hydrogen-bonded to a phosphate oxygen of C62 (Figure 3b). A47 and C62 form an A•C reverse Hoogsteen base pair, which has not been observed before (Figure 3b) [41–43]. The exocyclic amino group of A47 forms hydrogen bonds with O2 and N3 of C62; N7 of A47 also forms a hydrogen bond with N4 of C62. This base pair lies above A63 and no stacking interaction is seen with C46. The N1 and N2 groups of G61 form hydrogen

Table 2

Refinement statistics.

Resolution (Å)	22.5–2.7
Number of atoms	988 (45 nt, 1 Lu^{3+}, 1 Mg^{2+}, 2 SO_4^{2-}, 6 H_2O)
R_{value}*(%)	23.0 (36.6; 50.5)
R_{free}*(%)	24.5 (33.2; 51.5)
Bond length deviation (Å)	0.0057
Bond angle deviation (°)	1.0
Improper angle deviation (°)	1.2
Dihedrals (°)	10.6
Average B factor (Å^2)	80.6†
Minimum B factor (Å^2)	51.3
Maximum B factor (Å^2)	148.7
B rmsd for bonded mainchain atoms (Å^2)	2.1
B rmsd for bonded sidechain atoms (Å^2)	2.4

*Overall (22.5–2.7 Å) as well as 3.20–2.91 Å shell and 2.91–2.70 Å shell (in parentheses) values are reported. R_{free} is the R value obtained for a test set of reflections, consisting of a randomly selected 8.6% subset (909 reflections) of the diffraction data (10,539 reflections) not used during refinement. †A high average B factor value was expected on the basis of Wilson statistics. This value is partly because of the contribution of the highly flexible residues G35–A42 (average B factor ~106 Å^2).

bonds with a phosphate oxygen of A47 in addition to a single hydrogen bond between G61 N2 and G48 O6 (Figure 3c). The distance between N2(G61) and N7(G48) is 3.2 Å, which is slightly too long to form a stable hydrogen bond. G48 stacks with the C62•A47 base pair whereas G61 does not. G48 protrudes into the minor groove and its Watson–Crick face is exposed. A60 and G49 form a G•A imino base pair (Figure 3d). The two purine bases show extensive stacking interactions with G61 and G48.

As predicted by Leontis and Westhof [44,45], the overall structure of 4.5S RNA loop A resembles that of the 5S RNA loop E submotifs (root mean square deviation [rmsd] ~2 Å), although the peculiar stacking of base pairs A47•C62 and G48•G61 on the preceding pairs within loop A is substantially different from that of the corresponding pairs of loop E.

An electron-density peak in the proximity of the O3′ of A59 and one of the phosphate oxygens of A60 has been interpreted as a magnesium ion (Figure 3e). By stabilising the unusually compressed backbone region of nucleotides A59 and A60 (P–P distance = 5.2 Å), this ion could be responsible for the well documented Mg^{2+}-specific stabilisation of 4.5S RNA [22,35,36]. In particular, its position is consistent with the significant sharpening of the spectral peak of A59 in the presence of 5 mM $MgCl_2$ reported by Schmitz *et al.* [22]. A crystal contact between the phosphate group of A60 and an arginine sidechain from a symmetry-related Ffh molecule explains why this Mg^{2+} ion is not observed in the structure of the complex [27].

A second large density peak (4.9σ in the $|2F_o–F_c|$ map; Figure 2) was found in the minor groove of loop A and on the basis of its shape, chemical environment and behaviour in refinement it was interpreted to be a sulphate ion. At this position, this ion could form hydrogen bonds with N1 and N2 of G48 (Figure 3c) and with N2 and N3 of G49 (Figure 3d), therefore further stabilising the stacking interaction between these bases. At the 5′ terminus of the molecule, another sulphate ion makes a long-range interaction with N4 of C31. After a recent report by Masquida *et al.* [46], this is the second case in which specific binding of sulphate ions to nitrogen atoms of RNA bases has been observed in a crystal structure. Although it is unlikely that binding of sulphate to RNA is physiologically relevant, it is possible that more abundant ions such as PO_4^{2-}, Cl^- or HCO_3^- might bind to the same sites *in vivo*. This might be functionally important in the case of loop A because, in the structure of the

Figure 2

Stereo ball-and-stick representation of the symmetric loop A region of the refined 45 RNA crystal structure, with combined, sigmaa-weighted $|2F_o–F_c|$ electron-density map contoured at 1.0σ.

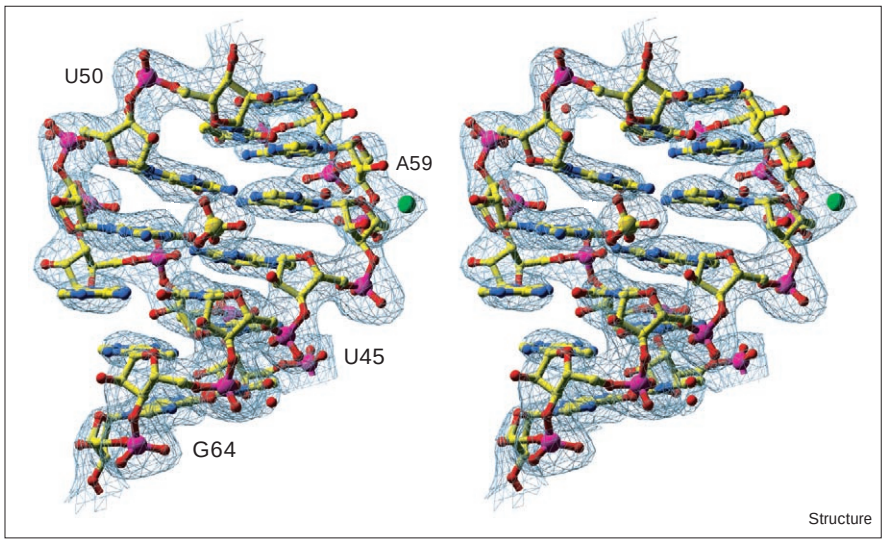
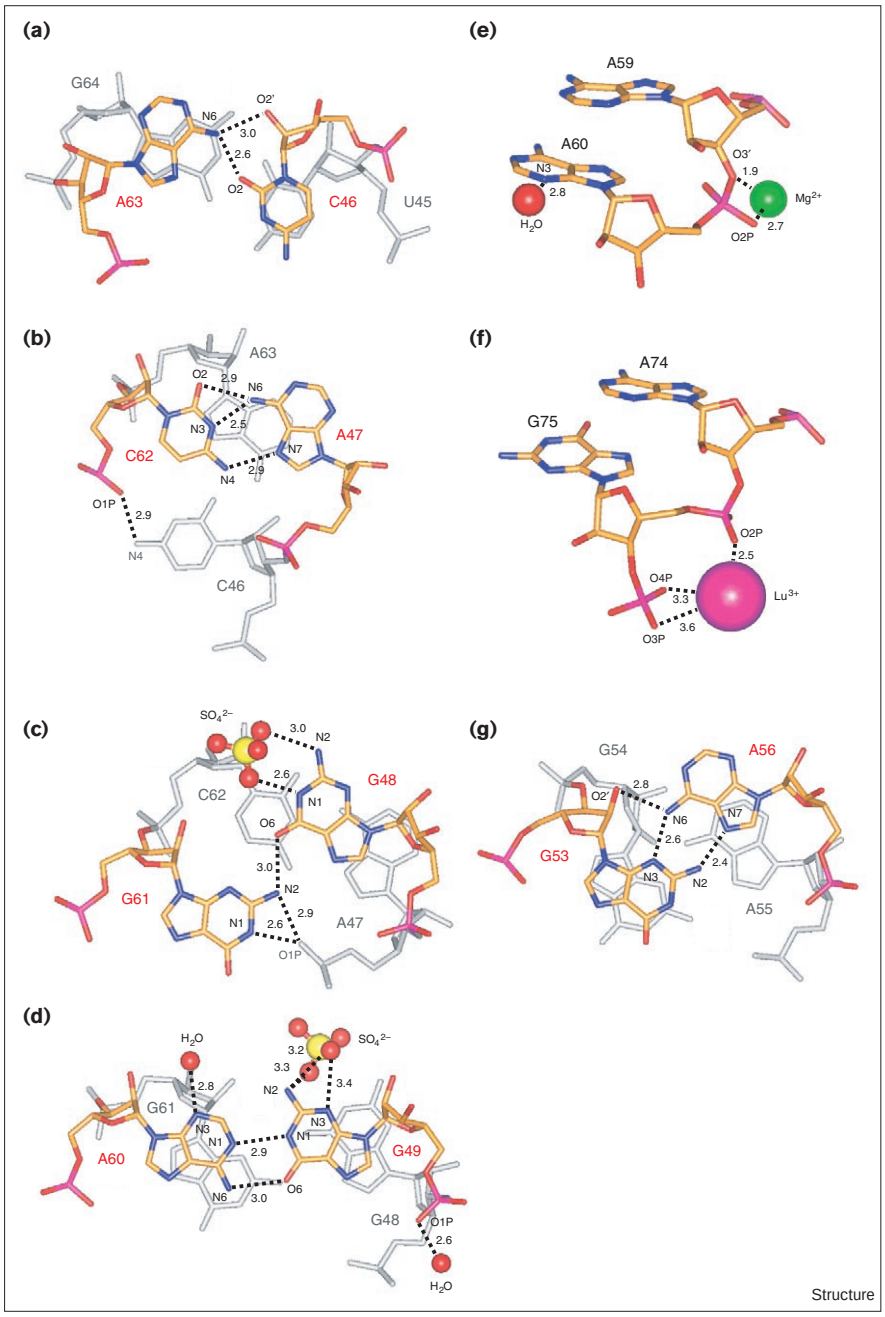


Figure 3



Details of RNA structure. Dashed lines indicate hydrogen bonds, with distances in Å. **(a–d)** Non-canonical base pairs within the symmetric loop A. (a) C46 and A63 interact so that the N6 of A63 not only contacts the O2 of the C46 base, but also makes a hydrogen bond with its 2′OH. (b) The A47•C62 reverse Hoogsteen base pair, with a cross-strand hydrogen bond between C62 O1P and C46 N4. Note the lack of stacking between A47 and the preceding nucleotide C46. (c) G48 and the phosphate of A47 interact in a base mismatch with G61. (d) The G49•A60 imino base pair. **(e,f)** Interaction of ions with the RNA. (e) An ordered Mg^{2+} ion stabilises the compressed backbone conformation of nucleotides A59 and A60 within loop A. (f) The Lu^{3+} binding site. **(g)** The two consecutive sheared G•A base pairs within the central internal loop.

complex between 4.5S RNA domain IV and the M domain of Ffh [27], protein residues make hydrogen bonds with both N1 and N2 of G48. Binding of a negatively charged ion to loop A of the free RNA might thus stabilise a nucleic acid conformation that favours its recognition by Ffh protein.

The asymmetric loop B

A series of structurally interesting packing interactions is found in the region of the crystal lattice where the asymmetric loop B is found (Figure 4a). The loops of two symmetry-related molecules form a continuous five-base stack that includes nucleotides A39, C40 and A42, the last being in the C2′-endo conformation (Figure 4b). The base of A39 stacks between those of residues C40 from both molecules. This positions it on the crystallographic twofold so that it would clash against its equivalent in the symmetry-related molecule. Because no density is visible for alternative conformations of this base, we deduced that A39 is found on the twofold in only half of the molecules within the crystal and is disordered in the

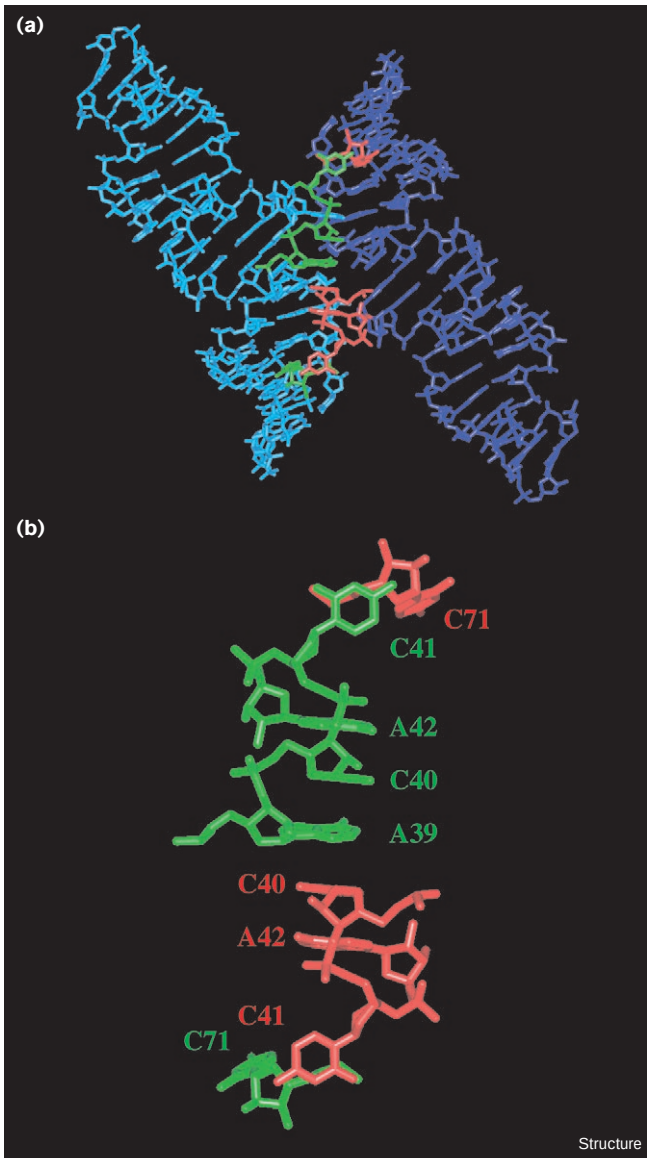
other half. The remaining nucleotide C41, which is also in C2′-endo conformation and has very unusual pseudo-torsion angles ($\eta = 323°$, $\theta = 313°$) [47], does not take part in the stacked interface, but instead packs against the backbone of C71 from the other molecule.

On the 3′ side of the loop, the single unpaired nucleotide A67 continues the stacking of helices b and c. This shows the important role of base stacking in RNA structure because the positioning of this nucleotide maintains the straightness and rigidity of the molecule. Although the structure of internal loop B is significantly affected by crystal contacts, the high B factors of all the nucleotides on its 5′ side as well as those of G35–U38 that precede them (average B factor ~106 Å^2) clearly indicate that this part of the molecule is likely to be highly flexible in solution, in agreement with the NMR study of Schmitz *et al.* [24].

The central internal loop

In the central internal loop of the 45 RNA dimer (equivalent to the tetraloop nucleotides in the monomer form of the RNA; Figure 3g), G53 pairs with A56 by a sheared G•A base pair [41,42], as observed for closing G•A base pairs in GNRA tetraloops (where N and R denote any nucleotide and purine, respectively) [48]. The backbone distortions in this region of the structure also allow the N6 of A56 to hydrogen bond with the 2′OH of the G53 ribose. The same pairing geometry is seen for the bases of nucleotides G54 and A55, although in this case G54 is in C2′-endo conformation and a hydrogen bond is made between its N2 and the 2′OH of G54 from the opposite strand. The geometry of both base pairs results in stacking between G53 and G54 that is much more extensive than between A55 and A56. Strikingly, the arrangement of the two consecutive G•A pairs is extremely similar to that of G163•A140/G164•A139 within the structure of the P4–P6 fragment of group I intron (rmsd = 0.86 Å) [49], which are found in a completely different structural environment. Hydrogen bonding between the guanine 2′OH and the adenine N6 has also been reported for isolated G•A base pairs in the sarcin/ricin loop (G2664•A2657) [50], in a fragment of 5S rRNA (G72•A104, G98•A78) [51] and in the hammerhead ribozyme (G120•A90) [52,53]. In the 5S rRNA G72•A104 pair, on the opposite side of the pair in which these interactions are found, a water molecule mediates a hydrogen bond between the N2 of the guanine and a phosphate oxygen of the adenine. Similarly, in the case of the two consecutive G•A pairs found in the 45 RNA structure, there is evidence in the difference map for a water mediating a cross-pair hydrogen bond between the N1 of G53 and the O1P of A55 (data not shown). It is clear that these recurring features of G•A pairs have the function of increasing their stability, and it will be interesting to see whether or not they will also be observed in future high-resolution structures of ribosomal RNAs, in which G•A tandem mismatches are often found [54].

Figure 4



(a) Two 45 RNA molecules, coloured in cyan and blue, interact in the crystal via nucleotides A39–A42 of their asymmetric internal loops (shown in green and red, respectively). **(b)** Detail of the central region of (a), showing the five-membered intermolecular stack at the interface, with the A39 base lying on the crystallographic twofold axis. The equivalent base of the symmetry-related molecule is disordered and thus generates a local symmetry violation. Stacking of nucleotides C41 on C71 of the symmetry-related molecules is also observed.

The lutetium ion binds to a cleaved cyclic phosphate

The RNA fragment that we crystallised was generated by cleavage of a larger transcription product by both hammerhead (at the 5′ end) and hepatitis δ virus (at the 3′ end) ribozymes [30,32]. Consequently, we expected to find a 2′,3′-cyclic phosphate at the 3′ end of the RNA as seen in the crystal of helix 6 of human 7S RNA [39]. However, there is no evidence for a cyclic species in the map, which

in contrast shows clear density for only a 3′-phosphate of G75. In Lu^{3+}-soaked crystals, the heavy atom contacts both the 5′-phosphate of G75 and its 3′-phosphate, holding the latter in a well-defined conformation (Figure 3f). Although it remains to be determined at which stage the cyclic phosphate was cleaved, crystallographic analysis of native crystals also confirmed its absence (data not shown), suggesting that it was either broken during RNA preparation or during the long period required for crystals to nucleate. The proximity of the 2′-OH group of the 3′-terminal nucleotide to a symmetry-related molecule suggests that molecules with an O2′-phosphate would have been excluded from the crystal, even if equal amounts of the 2′- and 3′-phosphate species were present as a result of the random cleavage of the cyclic precursor. Alternatively, ion-mediated selective cleavage of the cyclic phosphate could have occurred prior to crystallisation. Given that a cleaved cyclic phosphate has large conformational freedom around the ε and ζ angles [41], the Lu^{3+} binding mode observed in this structure could be exploited to produce lanthanide derivatives of other RNA molecules generated by ribozyme cleavage and subsequent breaking of the cyclic species.

Discussion

Comparison of the loop A structures determined by crystallography and NMR

The crystal structure of the 45-nucleotide fragment of 4.5S RNA consists of two double helical rods connected by a hinge corresponding to the asymmetric loop. Our crystal structure is significantly different from the NMR structures of 28- and 43-nucleotide RNA fragments reported recently [23,24]. Loop A forms continuously stacked non-canonical base pairs in the crystal (Figure 3), whereas in the NMR structures A47 and A63 in the symmetric loop A show a cross-strand stack. Furthermore, the imino group and exocyclic keto oxygen of G48 are hydrogen-bonded to the 2′OH group of G61 and the phosphate of C62, respectively, inducing severe bending of the backbone. In our crystal structure, A47 and A63 are paired with C62 and C46 respectively. The crystal structure with continuous stacking of base pairs seems thermodynamically more stable than the NMR structures, in which C62 and C46 are neither involved in base pairing nor in stacking. The same argument applies to the G48•G61 pair, which we observe in the crystal structure but it is not found in the NMR structures. Schmitz *et al.* showed that both the 24- and 28-nucleotide fragments of 4.5S RNA undergo a structural change upon addition of Mg^{2+} ions, with nucleotide A59 being the most affected [22,23]. In their preliminary report, the conformation of the 24-nucleotide fragment in the presence of 5 mM Mg^{2+} was interpreted as an extensively paired structure (Figure 6 in [22]) similar to our crystal structure; in contrast, the structure of loop A proposed in their more recent studies [23,24] is significantly different from both the crystal structure and the preliminary NMR structure.

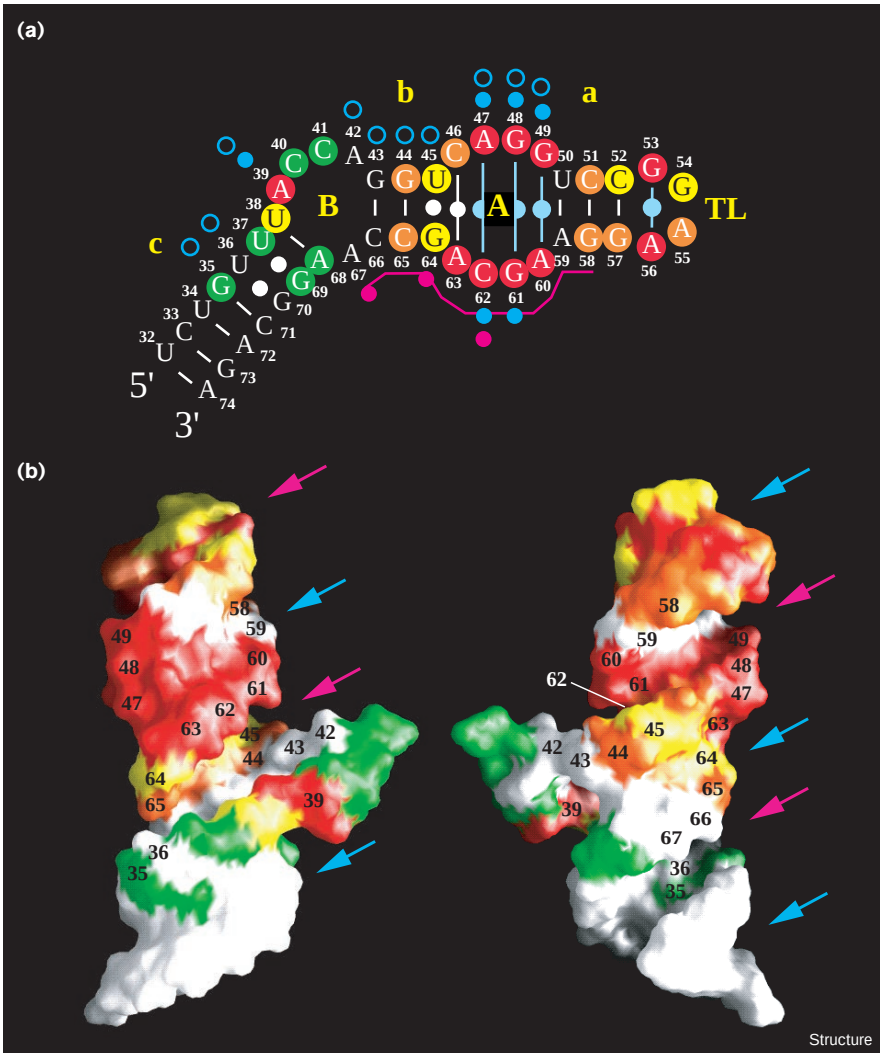
The crystal structure is consistent with chemical probing studies

Our structure truly represents the structure of the RNA in the crystal lattice but we cannot prove directly that it also represents the structure in solution. Chemical probing can be used to narrow this gap. The chemical probe dimethylsulphate (DMS) modifies the N1 of adenine and the N3 of cytidine, whereas kethoxal modifies the N1 and N2 of guanine [55]. The 4.5S RNA was probed with these compounds and its modification determined using reverse transcriptase primer extension [36]. The nitrogens of bases in Watson–Crick base pairs are not susceptible to modification by DMS or kethoxal, but the N2 of G in G•U wobble base pairs is exposed and likely to be reactive. These agents can therefore be used to study the base pairing scheme of mismatched bases. Loop A bases C46 and A63 are both modified. Their base–base interaction in the crystal structure (Figure 3a) leaves both the cytidine N3 and the adenine N1 exposed to solvent and their chemical modification by DMS is thus consistent with the structure. A47 and C62, in contrast, form an A•C reverse Hoogsteen pair that protects the cytosine N3 while leaving the adenine N1 exposed (Figure 3b). Our structure is consistent with the observed modification of A47. G48 and a phosphate oxygen of A47 interact with the G61 base so as to protect both its N2 and N1 groups, but for the G48 base both these groups are exposed (Figure 3c). Kethoxal modifies G48 but not G61. The G•A imino base pair of G49 and A60 protects the N1 groups of both bases, but leaves the N2 of G49 exposed (Figure 3d). Chemical modification only affects G49. In conclusion, our crystal structure is consistent with the chemical probing experiments of Lentzen *et al.* [36] carried out in solution.

Interaction of 4.5S RNA with Ffh protein

Sequence analysis ([56,57]; Figure 5a) and mutagenesis studies [58,59] showed that Ffh protein binds primarily to the symmetric loop of domain IV. Nucleotide A39 in the asymmetric loop increases the complex affinity whereas the apical tetraloop is not required for binding ([22,56,58]; Figure 1d). *In vitro* chemical protection of 4.5S RNA by Ffh shows that the protein mainly protects the 5′ side of domain IV, with highly conserved nucleotides A39, A47, G48 and G49 being the most protected [36]. Mapping of these data (Figure 5a) onto the crystal structure shows that the groups protected from chemical modification by Ffh in the last three residues are on the minor groove side of the symmetric loop A (Figure 5b). It is generally believed that the shallow and wide minor groove of the A form RNA is more accessible for protein than the deep major groove; on the other hand, the patterns of hydrogen-bond donors and acceptors displayed by canonical base pairs in the minor groove are limited, making it less suitable for recognition by proteins. However, the relative arrangement of the five consecutive non-canonical base pairs introduce substantial irregularity into the helical structure

Figure 5

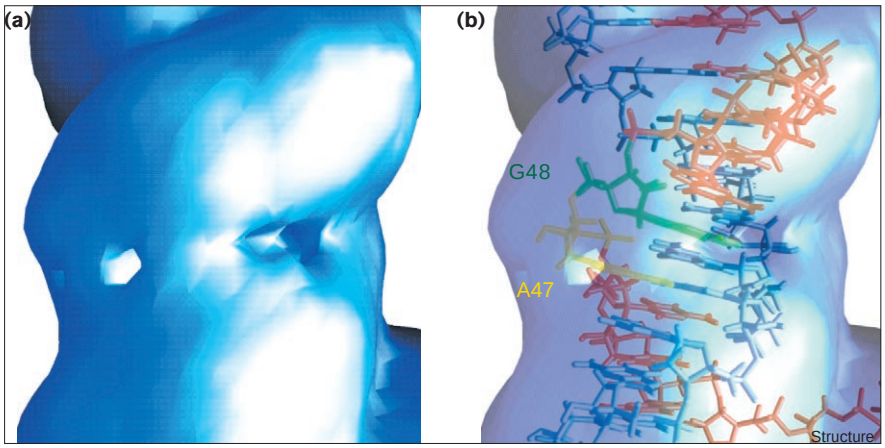


(a) Diagram of 4.5S RNA domain IV, summarising sequence and secondary structure conservation, mutagenesis and protection data. Positions conserved in 100%, >90%, >80% and >70% of bacterial SRP RNA sequences [87] are encircled in red, orange, yellow and green, respectively. Phylogenetically conserved and invariant base pairs are indicated in white and cyan, respectively. Closed circles above nucleotides indicate positions essential for binding of Ffh/SRP54 (blue) [58,59] and EF-G (magenta) [16]. A magenta line marks the decanucleotide sequence identical in 4.5S and 23S RNAs [17] and required for EF-G binding [21]. Open blue circles above nucleotides indicate positions protected from modification upon Ffh binding *in vitro* [36]. **(b)** Front and back surface representations of the 45 RNA crystal structure, with nucleotides important for Ffh/SRP54 and EF-G recognition indicated by their number and coloured according to their conservation as in (a). Major and minor grooves are indicated by magenta and blue arrows, respectively.

of symmetric loop A. In this region, the C1′–C1′ distances range from 9.2 Å (C46•A63) to 12.6 Å (G49•A60). As a result, the minor groove of loop A is particularly flat and exposes its peculiar stack of non-canonical base pairs, displaying a larger and unique set of hydrogen-bond donors and acceptors for specific protein recognition. On the basis of these considerations, we expected some important RNA–protein interactions to be made at the 5′ side of the symmetric loop minor groove. Cavities in isopotential contours of RNA have been shown to occur at positions important for functional interaction [60]. In support of the data discussed above, deep potential holes are observed in our structure at nucleotides A47 and G48, the latter occurring in the middle of the minor groove (Figure 6).

Recently, Batey *et al.* [27] solved the structure of 4.5S RNA domain IV bound to the M domain of Ffh at 1.8 Å resolution. The helix–turn–helix motif of the M domain recognises the non-canonical base pairs in the minor groove of the symmetric loop. Three of the asymmetric loop bases are stacked, and their phosphate backbone wraps around the outside of the RNA helix forming a platform, which positions α helix 3 of the M domain in place. We exchanged coordinates with Batey *et al.* to compare the structure of 4.5S RNA in the free and unbound forms. This analysis showed that the double helical region consisting of helix a, helix b and the symmetric internal loop A has nearly identical structures in the free RNA and complex crystals (rmsd = 0.72 Å), with all hydrogen-bonding interactions between symmetric loop nucleotides conserved between the two structures. Helix c also has a nearly identical structure in the free RNA and the complex crystals (rmsd = 0.76 Å), but its orientation relative to the former region is different in the two structures. This suggests that the asymmetric loop functions as a flexible hinge between the two double

Figure 6



(a) Detail of the three-dimensional isopotential contour of 45 RNA at –1.0 kT/e in the symmetric loop A region. Two deep potential holes are clearly visible. **(b)** Stick representation of part of 45 RNA, superimposed on the same isopotential contour region as in (a). RNA strands are coloured as in Figure1b, with the exception of nucleotides A47 (yellow) and G48 (green) the bases of which are located in positions corresponding to the potential holes shown in (a).

helical segments [24], of which the relative orientation becomes fixed upon binding to the M domain because of the interaction of the asymmetric loop with both the protein and the symmetric loop [27].

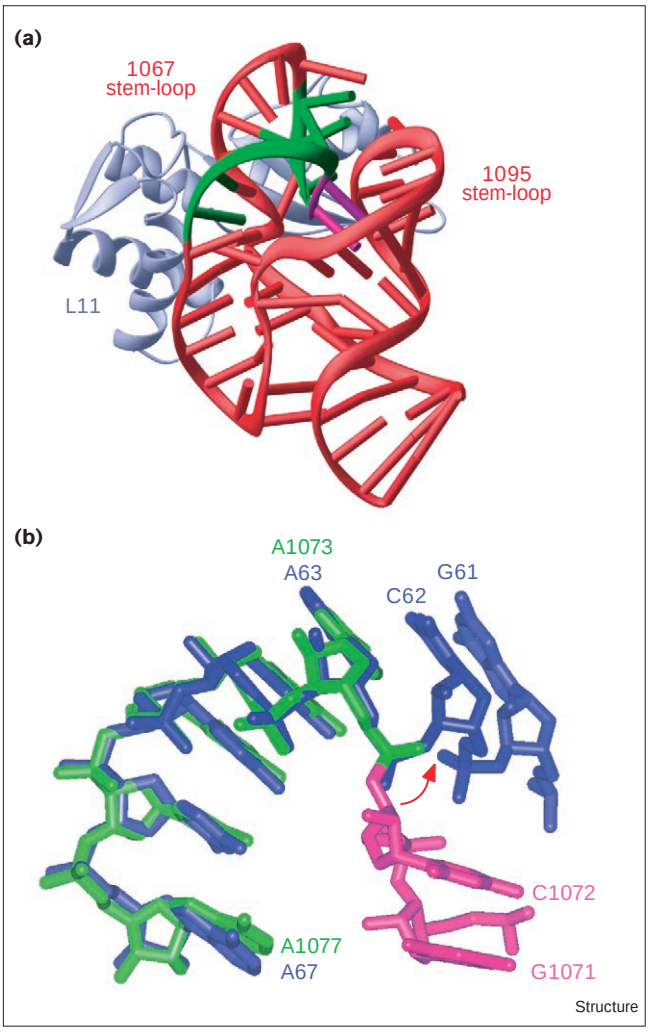
In contrast, the agreement of the two double helical segments of the 43mer NMR structure [24] with the corresponding segments of either the free RNA or the complex crystal structure is not as close (rmsd >2 Å). This might not be surprising as the NMR structure is based on short range distance constraints [25,26], but it is worth noting that the hydrogen-bonding interactions, within the symmetric loop deduced from the NMR constraints, are different from those seen in both free RNA and complex crystals. We cannot exclude the possibility that the symmetric loop A structure in our crystal might be distorted by the high ammonium sulphate concentrations used for crystallisation or by the crystal lattice. Nevertheless, it is surprising that the structure of helix a, helix b and the symmetric loop shows excellent agreement with that in the complex despite the extensive interaction between the M domain and the symmetric loop. It is therefore possible that the structure of the symmetric loop is highly stable and little affected either by protein binding or the crystal lattice. Our free RNA structure would in this case be a true representation of the solution structure.

The fragment of 4.5S RNA consisting of helix a, helix b and the symmetric loop A alone binds weakly but specifically to Ffh [23]. Schmitz *et al.* [24] proposed that the M domain first binds to this region followed by the interaction with the asymmetric loop B. The crystal structure of the free and bound RNA suggests that the initial step of binding between the M domain and 4.5S RNA consists of a rigid-body docking of the loop A region and the M domain involving no structural rearrangements of either component. This mechanism is distinct from predominantly induced fit mechanisms described previously [61–66]. The second step of binding involves the freezing of the relatively flexible asymmetric loop B between helix 3 of the M domain and the loop A region of the RNA. The crystal structure of the complex [27] shows that the RNA–protein binding creates an extensive network of interactions involving the asymmetric loop, which compensate for a loss of entropy.

Interaction of 4.5S RNA with elongation factor G

The crystal structure of 4.5S RNA presented here includes the decanucleotide sequence proposed to bind EF-G [21,67]. The ten residues are found on the opposite strand to those important for Ffh binding, from the 3′ side of helix a to the unpaired nucleotide A67 of the asymmetric loop (Figure 5). EF-G not only binds to 4.5S RNA, but also to a region of 23S rRNA (loop 1067) that contains the same decanucleotide sequence [16,20]. Comparison of the EF-G binding sites in 4.5S RNA and 23S rRNA within the L11 complex structure (Figure 7a) [28,29], shows that the last five decanucleotide residues have almost identical conformations (rmsd = 0.96 Å; Figure 7b). In contrast, the first five decanucleotide residues have a distinct orientation. This partial structural identity can be interpreted in two ways: firstly, EF-G binds to the 23S conformation of RNA in its GTP-bound state, but binds preferentially to the 4.5S conformation in its GDP-bound state or, secondly, 23S rRNA undergoes a conformational change in its 1067 loop concomitant with translocation and/or GTP hydrolysis and the structure seen in the L11 complex represents a different affinity state, most likely a lower affinity one. At present, there is insufficient data to distinguish between these two possibilities, although evidence exists for a conformational change of EF-G in the ribosome upon GTP hydrolysis. This is concomitant with translocation [68] and involves a change in 23S rRNA conformation in the 1067 loop during the same step [69]. The 4.5S RNA can also associate with EF-G in the absence of ribosomes, and has more than twice the affinity for EF-G•GDP than

Figure 7



(a) Ribbon representation of the L11–23S rRNA complex structure (adapted from [28]). Ribosomal protein L11 is coloured in grey and 23S rRNA in red, with the conserved decanucleotide sequence in the 1067 loop highlighted in green (G1068–A1070, A1073–A1077) and magenta (G1071, C1072). Nucleotides G1071 and C1072 bring together stem-loops 1067 and 1095 of the rRNA by formation of base triples. Interaction of EF-G with the decanucleotide sequence would not be hindered by the presence of L11 protein, which binds to the opposite face of 23S rRNA. **(b)** Superposition of nucleotides A63–A67 of 45 RNA (blue) and nucleotides A1073–A1077 of the 1067 loop of 23S rRNA (green) [28], showing common structural features likely to be important for recognition of both RNAs by EF-G. A red arrow indicates the rotation that would be required to also bring nucleotides G1071 and C1072 of 23S rRNA (magenta) in the same conformation as their equivalents in 4.5S RNA. This putative conformational change would require breaking of the base triples that involve G1071 and C1072 in the 23S rRNA crystal structure.

for EF-G•GTP [21]. For the decanucleotide residues in 23S rRNA to adopt a conformation more similar to their equivalents in 4.5S RNA, breaking of the two base triple ternary interactions made by G1071 and C1072 [28,29] would be required.

The biological function of the interaction between 4.5S RNA and the ribosome remains to be determined [16,67]. On the basis of all previous data and on the structural comparison discussed above, we propose that 4.5S RNA might interact with the ribosome to prevent 'stalling' caused by either uncharged tRNAs or EF-G•GDP not being expelled from the ribosome following translocation. This would be consistent with an essential role in protein synthesis that is nevertheless not required for every round of elongation.

Biological implications

4.5S RNA binds to both Ffh and EF-G proteins. The crystal structure of domain IV of 4.5S RNA has revealed that the asymmetric loop B acts as a hinge between the flanking helical regions. In the symmetric loop A, stacking of five consecutive non-canonical base pairs gives rise to a unique helical structure, with the shallow minor groove projecting an unusual set of functional groups. This makes the symmetric loop ideal for a specific protein-binding site. The structure of the symmetric loop is identical in the free and complexed structures, suggesting that the initial binding might be a rigid-body docking followed by induced ordering of the asymmetric loop. The decanucleotide sequence found in the EF-G binding sites within 4.5S RNA and 23S rRNA suggests that binding of EF-G to 23S rRNA consists of rigid-body binding of part of the decanucleotide and induced fit of the rest. The combination of rigid-body docking and induced fit might be a common mechanism of RNA–protein interactions, balancing entropy and enthalpy terms to achieve high specificity and binding energy.

Materials and methods

RNA synthesis and purification

After *in vitro* transcription and purification as described in [30], RNA constructs were concentrated to 5–15 mg/ml, in 5 mM Na cacodylate pH 6.5, and stored at –70°C.

Electrophoretic mobility shift and GTPase assays

Electrophoretic mobility shift assays with purified *E. coli* Ffh C406S protein [32] were performed in a buffer containing 20 mM Tris-HCl pH 7.4, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and 1 mg/ml purified BSA. RNA monomers were prepared as described below; 1 mM spermine was added to the annealing reactions to favour dimer formation. GTPase assays were carried out as detailed in [10].

RNA annealing, crystallisation and derivatisation

Samples of 45 RNA were quickly thawed in water at room temperature and kept on ice. After dilution to 0.370 mM in 20 mM Na cacodylate pH 6.5, the RNA was incubated on ice for 1 h, heated at 65°C for 10 min, immediately centrifuged at 13,000 rpm for 10 s and then snap-cooled on ice to prevent formation of dimeric species (as assessed by native gel electrophoresis). After a further hour, 2 µl of the annealed RNA solution were mixed with an equal volume of crystallisation solution (1.6–1.8 M $(NH_4)_2SO_4$, 90 mM $Mg(OAc)_2$, 50 mM Na cacodylate pH 6.0, 0–2 mM $Co(NH_3)_6$) at 19°C, and equilibriated against 750 µl of the same solution by sitting drop vapour diffusion at 30°C. After ~2 months, conical crystals appeared that grew to average dimensions of ~200 × 200 × 200 μm^3. Following slow adjustment of the incubation temperature to 19°C (–1°C/h), crystals were stabilised overnight in

2.2 M $(NH_4)_2SO_4$, 90 mM $Mg(OAc)_2$, 50 mM Na cacodylate pH 6.0. Finally, they were cryoprotected by stepwise (+5% (v/v)/2 min) addition of glycerol to a final concentration of 20% (v/v) in stabilisation solution and flash-frozen in liquid N_2 for storage and data collection.

For preparation of the lanthanide derivative, the $Mg(OAc)_2$ concentration of the stabilisation solution was approximately halved every 90 min to a final value of 10 mM. $Mg(OAc)_2$ and Na cacodylate were then exchanged for $MgCl_2$ and Bis-Tris-HCl pH 6.0, respectively, to maximise binding of the heavy-atom compound. $LuCl_3{\bullet}6H_2O$ was then added to a final concentration of 2 mM in a stepwise manner (0.1–0.25–0.5–1–2 mM, with 2 h incubation/step) and crystals were cryoprotected as described above, with the difference that 2 mM $LuCl_3{\bullet}6H_2O$ was included in the cryoprotection solutions.

Data reduction and phase determination

Datasets were collected at 100K at beamline 5.2R of Elettra synchrotron (Trieste, Italy) and processed with the programs Denzo and Scalepack [70]. Merging and scaling were performed with the Crystallography and NMR System (CNS) [71]. The same program suite was used to refine the parameters for the Lu^{3+} site, initially located in the anomalous Patterson maps, using a maximum-likelihood target function [72], and to calculate both initial and solvent-flattened phases. For MAD phasing, the first remote wavelength (λ_3) served as reference. Forty cycles of solvent flipping [73] and histogram matching [74] were performed for the density modification, using an automatic solvent mask generated with starting and finishing radii of 3.50 and 2.98 Å, respectively. Scaling and phasing statistics are reported in Table 1.

Model building

Both self-rotation and self-translation functions with the program GLRF [75] strongly suggested the presence of noncrystallographic symmetry within the 45 RNA crystals, which was eventually attributed to the intrinsic symmetry of the molecule. The first experimental map was therefore calculated using a solvent content of 52%, consistent with the presence of two molecules of RNA per asymmetric unit. This map was of high quality and its extensive continuity at 1.0σ allowed us to build most of the molecule with the program O [76], using nucleotides derived from a model of 4.5S RNA domain IV generated with the program MC-SYM [77]. Although the initial density of the map region that was subsequently assigned to internal loop B nucleotides A39–A42 was rather ambiguous and discontinuous, its clear asymmetry provided a first landmark in the sequence. This was combined with both the constraint imposed by the crystallographic dyad running through the middle of the RNA dimer and the different density of purines and pyrimidines to establish the correct sequence register. At this stage, the density for the sulphate ion in internal loop A was already clearly defined. As soon as it became evident that there was only a single molecule in the asymmetric unit, a second map was calculated by applying a solvent content of 70% during flattening. The new map allowed tracing of nucleotides C41 and A42 and introduction of additional ions and water molecules.

Refinement

The model was refined against the remote λ_4 wavelength data, using CNS. Simulated annealing [78] was used in the first rounds of refinement, while energy minimisation followed by Lu^{3+} occupancy refinement and restrained individual B-factor refinement of all atoms were performed at later stages. A correction was applied to account for the high anisotropy of the data (B11 = B22 = 26.514, B33 = –53.028); in addition, to take advantage of the unusually high solvent content (76.9%), individual B-factor refinement was performed using a low resolution limit of 8.0 Å for initial B factor and bulk-solvent correction. The A39 nucleotide was refined with an occupancy of 0.5 to account for its alternatively ordered and disordered conformation in adjacent molecules. Refinement statistics may be found in Table 2.

Analysis of RNA structure conformation and electrostatics

Visualisation of RNA structures was performed with RIBBONS (Figures 1c,7a) [79], Insight II (Figures 3,4 and 7b) (Molecular Simulations Inc., San Diego, CA) and GRASP (Figure 5b and 6) [80]. Figure 2 was made with Swiss-PdbViewer [81] and POV-Ray™ (http://www.povray.org/). Conformational analyses were carried out using the programs CURVES [82] and AMIGOS [47]. Superpositions and rmsd calculations were performed using the McLachlan algorithm [83], as implemented in the program ProFit (Martin, ACR, http://www.biochem.ucl.ac.uk/~martin/programs/#profit1).

Electrostatic potentials were calculated using the non-linear Poisson-Boltzmann equation [84] in the program Qnifft, as described in [60]; isopotential contours at –1.0 kT/e were visualised with the program GRASP [80]. Parameters for calculations were chosen to match the crystal stabilisation conditions, as described in [84,85]: dielectric constants were set to $\varepsilon_{RNA} = 2$ and $\varepsilon_{solvent} = 80$, ionic strength was set to 2.3 M and ions and water molecules were excluded from the calculations.

Phylogenetic conservation analysis of SRP RNA domain IV

Potential pairing, strict covariation and mutual information analyses were performed with the program Covariation [86], using a database of non-redundant SRP RNA sequences from all organisms (71 sequences). This was derived from all aligned RNA sequences in the SRP database (SRPDB) [87] and all new non-aligned entries within the same database, with the exception of the highly divergent *Saccharomyces cerevisiae* and *Coturnix coturnix* sequences. Sequence manipulations were performed using the programs SeqPup (Don Gilbert, Indiana University; Bloomington, Indiana; http://iubio.bio.indiana.edu/soft/molbio/seqpup/), Belvu (Eric Sohnhammer, Karolinska Institute, Stockholm, Sweden; http://kisac.cgr.ki.se/cgr/groups/sonnhammer/Belvu.html) and Analyse Conservation (LJ, unpublished).

Accession numbers

Atomic coordinates and structure factors have been deposited in the Nucleic Acid Database and the Protein Data Bank with the ID codes UR0009 and 1DUH, respectively. Coordinates for the biologically relevant structure can also be downloaded at http://www2.mrc-lmb.cam.ac.uk/personal/kn/kiyoshi.html.

Acknowledgements

The authors thank all the Protein Structure Group of Glaxo Wellcome for continued help and support during the course of the project; the staff at beamline 5.2R of Elettra synchrotron, Italy, in particular Luca Olivi, Edoardo Busetto and Marino Miculin, for excellent assistance during MAD data collection; Christian Kambach, Dean Derbyshire, Katjusa Brejc and Meindert Lamers for help during data collection; Daniela Rhodes, Ben Luisi, Phil Evans, Randy Read, Sandra Searles, James Murray, Chris Johnson, Kevin Chin, Brian Wimberly, William Clemons, Gertie van Pouderoyen, David Earnshaw and all the members of the Nagai and Sixma groups for many helpful discussions and comments. We thank Rob Batey and Jennifer Doudna for communicating their results prior to publication.

LJ thanks Glaxo Wellcome Research & Development UK for a studentship and the Netherlands Cancer Society for a fellowship; TH thanks EU for a fellowship. This work was supported by the Medical Research Council.

Note added in proof

A third case of binding of a sulphate ion to RNA has been recently observed in the complex between a mutant Gln-tRNA and its cognate aminoacyl-tRNA synthetase (Tim Bullock and John Perona, *Nat. Struct. Biol.,* in press). In this structure, a sulphate ion binds to the major groove of tRNA via hydrogen bonds to exocyclic nitrogens of adenine and cytosine bases.

References

1. Walter, P. & Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87-119.
2. Lütcke, H. (1995). Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *Eur. J. Biochem.* **228**, 531-550.
3. Poritz, M.A., Strub, K. & Walter, P. (1988). Human SRP–RNA and *E. coli* 4.5S RNA contain a highly homologous structural domain. *Cell* **55**, 4-6.

4. Larsen, N. & Zwieb, C. (1991). SRP–RNA sequence alignment and secondary structure. *Nuc. Acids Res.* **19**, 209-215.
5. Connolly, T., Rapiejko, P.J. & Gilmore, R. (1991). Requirement of GTP hydrolysis for dissociation of the signal recognition particle from its receptor. *Science* **252**, 1171-1173.
6. Poritz, M.A., *et al.*, & Walter, P. (1990). An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science* **250**, 1111-1117.
7. Ribes, V., Römisch, K., Giner, A., Dobberstein, B. & Tollervey, D. (1990). *E. coli* 4.5S RNA is part of a ribonucleoprotein particle that has properties related to signal recognition particle. *Cell* **63**, 591-600.
8. Luirink, J., *et al.*, & Dobberstein, B. (1992). Signal-sequence recognition by an *Escherichia coli* ribonucleoprotein complex. *Nature* **359**, 741-743.
9. Luirink, J. & Dobberstein, B. (1994). Mammalian and *Escherichia coli* signal recognition particles. *Mol. Microbiol.* **11**, 9-13.
10. Miller, J.D., Bernstein, H.D. & Walter, P. (1994). Interaction of *E. coli* Ffh/4.5S ribonucleoprotein and FtsY mimics that of mammalian signal recognition particle and its receptor. *Nature* **367**, 657-659.
11. Powers, T. & Walter, P. (1995). Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases. *Science* **269**, 1422-1424.
12. Macao, B., Luirink, J. & Samuelsson, T. (1997). Ffh and FtsY in a *Mycoplasma mycoides* signal-recognition particle pathway: SRP RNA and M domain of Ffh are not required for stimulation of GTPase activity *in vitro*. *Mol. Microbiol.* **24**, 523-534.
13. Powers, T. & Walter, P. (1997). Co-translational protein targeting catalyzed by the *Escherichia coli* signal recognition particle and its receptor. *EMBO J.* **16**, 4880-4886.
14. Brown, S. & Fournier, M.J. (1984). The 4.5S RNA gene of *Escherichia coli* is essential for cell growth. *J. Mol. Biol.* **178**, 533-550.
15. Brown, S. (1987). Mutations in the gene for EF-G reduce the requirement for 4.5S RNA in the growth of *E. coli*. *Cell* **49**, 825-833.
16. Nakamura, K., Fujii, Y., Shibata, T. & Yamane, K. (1999). Depletion of *Escherichia coli* 4.5S RNA leads to an increase in the amount of protein elongation factor EF-G associated with ribosomes. *Eur. J. Biochem.* **259**, 543-550.
17. Brown, S. (1989). Time of action of 4.5 S RNA in *Escherichia coli* translation. *J. Mol. Biol.* **209**, 79-90.
18. Jensen, C.G., Brown, S. & Pedersen, S. (1994). Effect of 4.5S RNA depletion on *Escherichia coli* protein synthesis and secretion. *J. Bacteriol.* **176**, 2502-2506.
19. Jensen, C.G. & Pedersen, S. (1994). Concentrations of 4.5S RNA and Ffh protein in *Escherichia coli*: the stability of Ffh protein is dependent on the concentration of 4.5S RNA. *J. Bacteriol.* **176**, 7148-7154.
20. Moazed, D., Robertson, J.M. & Noller, H.F. (1988). Interaction of elongation factor EF-G and EF-Tu with a conserved loop in 23S RNA. *Nature* **334**, 362-364.
21. Shibata, T., Fujii, Y., Nakamura, Y., Nakamura, K. & Yamane, K. (1996). Identification of protein synthesis elongation factor G as a 4.5 S RNA-binding protein in *Escherichia coli*. *J. Biol. Chem.* **271**, 13162-13168.
22. Schmitz, U., *et al.*, & Walter, P. (1996). NMR studies of the most conserved RNA domain of the mammalian signal recognition particle (SRP). *RNA* **2**, 1213-1227.
23. Schmitz, U., James, T.L., Lukavsky, P. & Walter, P. (1999). Structure of the most conserved internal loop in SRP RNA. *Nat. Struct. Biol.* **6**, 634-638.
24. Schmitz, U., *et al.*, & James, T.L. (1999). Structure of the phylogenetically most conserved domain of SRP RNA. *RNA* **5**, 1419-1429.
25. Rife, J.P., *et al.*, & Moore, P.B. (1999). Comparison of the crystal and solution structures of two RNA oligonucleotides. *Biophys. J.* **7 6**, 65-75.
26. Allain, F.H. & Varani, G. (1997). How accurately and precisely can RNA structure be determined by NMR? *J. Mol. Biol.* **267**, 338-351.
27. Batey, R.T., Rambo, R.P., Lucast, L., Rha, B. & Doudna, J.A. (2000). Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**, 1232-1239.
28. Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11–RNA complex. *Cell* **97**, 491-502.
29. Conn, G.L., Draper, D.E., Lattman, E.E. & Gittis, A.G. (1999). Crystal structure of a conserved ribosomal protein–RNA complex. *Science* **284**, 1171-1174.
30. Price, S.R., Oubridge, C., Varani, G. & Nagai, K. (1998). Preparation of RNA: protein complexes for X-ray crystallography and NMR. In *RNA: protein interactions: a practical approach,* (Smith, C., ed.), pp. 37-74, Oxford University Press, Oxford.
31. Scott, W.G., *et al.*, & Klug, A. (1995). Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.* **250**, 327-332.
32. Jovine, L. (1998). Structural studies of the U1A-3′UTR complex and of signal recognition particle. Ph.D. thesis, University of Cambridge, England.
33. Horrocks, W.D., Jr. (1993). Luminescence spectroscopy. *Methods Enzymol.* **226**, 495-538.
34. Feig, A.L., Scott, W.G. & Uhlenbeck, O.C. (1998). Inhibition of the hammerhead ribozyme cleavage reaction by site-specific binding of Tb(III). *Science* **279**, 81-84.
35. Bourgaize, D.B., Farrell, C., Langley, K.H. & Fournier, M.J. (1984). Physical properties of the *E. coli* 4.5S RNA: first results suggest a hairpin helix of unusual thermal stability. *Nucleic Acids Res.* **12**, 2019-2034.
36. Lentzen, G., Moine, H., Ehresmann, C., Ehresmann, B. & Wintermeyer, W. (1996). Structure of 4.5S RNA in the signal recognition particle of *Escherichia coli* as studied by enzymatic and chemical probing. *RNA* **2**, 244-253.
37. Holbrook, S.R., Cheong, C., Tinoco, I., Jr. & Kim, S.H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature* **353**, 579-581.
38. Ennifar, E., *et al.*, & Dumas, P. (1999). The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure* **7**, 1439-1449.
39. Wild, K., Weichenrieder, O., Leonard, G.A. & Cusack, S. (1999). The 2 Å structure of helix 6 of the human signal recognition particle RNA. *Structure* **7**, 1345-1352.
40. Watson, J. & Crick, F. (1953). A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.
41. Saenger, W. (1983). *Principles of nucleic acid structure.* Springer-Verlag, Berlin.
42. Burkard, M., Turner, D. & Tinoco, I. (1999). Appendix 1: Structures of base pairs involving at least two hydrogen bonds. In *The RNA world, 2nd edition*, (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds), pp. 675-680, CSHL Press, Cold Spring Harbor, NY.
43. Nagaswamy, U., Voss, N., Zhang, Z. & Fox, G.E. (2000). Database of non-canonical base pairs found in known RNA structures. *Nucleic Acids Res.* **28**, 375-376.
44. Leontis, N.B. & Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.* **283**, 571-583.
45. Leontis, N. & Westhof, E. (1998). The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* **4**, 1134-1153.
46. Masquida, B., Sauter, C. & Westhof, E. (1999). A sulfate pocket formed by three G•U pairs in the 0.97 Å resolution X-ray structure of a nonameric RNA. *RNA* **5**, 1384-1395.
47. Duarte, C.M. & Pyle, A.M. (1998). Stepping through an RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.* **284**, 1465-1478.
48. Heus, H.A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191-194.
49. Cate, J.H., *et al.*, & Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685.
50. Correll, C.C., Wool, I.G. & Munishkin, A. (1999). The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J. Mol. Biol.* **292**, 275-287.
51. Correll, C.C., Freeborn, B., Moore, P.B. & Steitz, T.A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705-712.
52. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74.
53. Scott, W.G., Finch, J.T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991-1002.
54. Gautheret, D., Konings, D. & Gutell, R.R. (1994). A major family of motifs involving G•A mismatches in ribosomal RNA. *J. Mol. Biol.* **242**, 1-8.
55. Ehresmann, C., *et al.*, & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109-9128.
56. Althoff, S.M., Selinger, D. & Wise, J.A. (1994). Molecular evolution of SRP cycle components: functional implications. *Nucleic Acids Res.* **22**, 1933-1947.
57. Zwieb, C., Müller, F. & Larsen, N. (1996). Comparative analysis of tertiary structure elements in signal recognition particle RNA. *Fold. Des.* **1**, 315-324.

58. Wood, H., Luirink, J. & Tollervey, D. (1992). Evolutionary conserved nucleotides within the *E. coli* 4.5S RNA are required for association with P48 *in vitro* and for optimal function *in vivo. Nuc. Acids Res.* **20**, 5919-5925.
59. Selinger, D., Brennwald, P., Liao, X. & Wise, J.A. (1993). Identification of RNA sequences and structural elements required for assembly of fission yeast SRP54 protein with signal recognition particle RNA. *Mol. Cell. Biol.* **13**, 1353-1362.
60. Chin, K., Sharp, K.A., Honig, B. & Pyle, A.M. (1999). Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* **6**, 1055-1061.
61. Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.
62. Jovine, L., Oubridge, C., Avis, J.M. & Nagai, K. (1996). Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy. *Structure* **4**, 621-631.
63. Allain, H.F.-T., Howe, P.W.A., Neuhaus, D. & Varani, G. (1997). Structural basis of the RNA binding specificity of human U1A protein. *EMBO J.* 16, 5764-5774.
64. Price, S.R., Evans, P.R. & Nagai, K. (1998). Crystal structure of the spliceosomal U2B′′ -U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645-650.
65. Handa, N., *et al.*, & Yokoyama, S. (1999). Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579-585.
66. Deo, R.C., Bonanno, J.B., Sonenberg, N. & Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-845.
67. Brown, S. (1991). 4.5S RNA: does form predict function? *New Biol.* **3**, 430-438.
68. Rodnina, M.V., Savelsbergh, A., Katunin, V.I. & Wintermeyer, W. (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* **385**, 37-41.
69. Rodnina, M.V., *et al.*, & Wintermeyer, W. (1999). Thiostrepton inhibits the turnover but not the GTPase of elongation factor G on the ribosome. *Proc. Natl. Acad. Sci. USA* **96**, 9586-9590.
70. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
71. Brünger, A.T., *et al.*, & Warren, G.L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
72. Burling, F.T., Weis, W.I., Flaherty, K.M. & Brünger, A.T. (1996). Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **271**, 72-77.
73. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F_1 ATPase. *Acta Crystallogr. D* **52**, 30-42.
74. Zhang, K. & Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Crystallogr. A* **46**, 41-46.
75. Tong, L. & Rossmann, M.G. (1997). Rotation function calculations with GLRF program. *Methods Enzymol.* **276**, 594-611.
76. Jones, T.A. & Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277**, 173-208.
77. Major, F., *et al.*, & Cedergren, R. (1991). The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science* **253**, 1255-1260.
78. Brünger, A.T. & Rice, L.M. (1997). Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol.* **277**, 243-269.
79. Carson, M. (1997). Ribbons. *Methods Enzymol.* **277**, 493-505.
80. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
81. Guex, N. & Peitsch, M. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis* **18**, 2714-2723.
82. Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63-91.
83. McLachlan, A. (1982). Rapid comparison of protein structures. *Acta Crystallogr. A* **38**, 871-873.
84. Sharp, K.A., Honig, B. & Harvey, S.C. (1990). Electrical potential of transfer RNAs: codon-anticodon recognition. *Biochemistry* **29**, 340-346.
85. Cate, J.H. & Doudna, J.A. (1996). Metal-binding sites in the major groove of a large ribozyme domain. *Structure* **4**, 1221-1229.
86. Brown, J.W. (1991). Phylogenetic comparative analysis of RNA structure on Macintosh computers. *Comp. Appl. Biosci.* **7**, 391-393.
87. Samuelsson, T. & Zwieb, C. (1999). The Signal Recognition Particle Database (SRPDB). *Nucleic Acids Res.* **27**, 169-170.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.

Supplementary material

Crystal structure of the Ffh and EF-G binding sites in the conserved domain IV of *Escherichia coli* 4.5S RNA

Luca Jovine, Tobias Hainzl, Chris Oubridge, William G Scott, Jade Li, Titia K Sixma, Alan Wonacott, Tadeusz Skarzynski and Kiyoshi Nagai

Structure 2000, **8**:527–540

Table S1

Observed diffraction ratios*.

	λ_1	λ_2	λ_3	λ_4	λ_5
λ1	0.087	0.050	0.078	0.061	0.064
λ2	–	0.077	0.079	0.063	0.065
λ3	–	–	0.080	0.074	0.075
λ4	–	–	–	0.062	0.057
λ5	–	–	–	–	0.071

*Values are $<(\Delta|F|)^2>^{1/2}/<|F|^2>^{1/2}$, where $\Delta|F|$ is the dispersive/isomorphous (off-diagonal elements) difference, or Bijvoet difference (diagonal elements), computed between 18.0 and 2.7 Å resolution.