

Structure of the 30S ribosomal subunit

Brian T. Wimberly^{*†}, Ditlev E. Brodersen^{*†}, William M. Clemons Jr^{*†‡}, Robert J. Morgan-Warren^{*†}, Andrew P. Carter^{*†}, Clemens Vornheim[§], Thomas Hartsch^{||} & V. Ramakrishnan[†]

[†]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[‡]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA

[§]Global Phasing Ltd., Sheraton House, Castle Park, Cambridge CB3 0AX, UK

^{||}Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany

* These authors contributed equally to this work

Genetic information encoded in messenger RNA is translated into protein by the ribosome, which is a large nucleoprotein complex comprising two subunits, denoted 30S and 50S in bacteria. Here we report the crystal structure of the 30S subunit from *Thermus thermophilus*, refined to 3 Å resolution. The final atomic model rationalizes over four decades of biochemical data on the ribosome, and provides a wealth of information about RNA and protein structure, protein–RNA interactions and ribosome assembly. It is also a structural basis for analysis of the functions of the 30S subunit, such as decoding, and for understanding the action of antibiotics. The structure will facilitate the interpretation in molecular terms of lower resolution structural data on several functional states of the ribosome from electron microscopy and crystallography.

Protein synthesis is a complex, multistep process that requires, in addition to the ribosome, several extrinsic GTP-hydrolysing protein factors during each of the main stages of initiation, elongation and termination. The 30S ribosomal subunit has a crucial role in decoding mRNA by monitoring base pairing between the codon on mRNA and the anticodon on transfer RNA; the 50S subunit catalyses peptide-bond formation. Despite several decades of work, the molecular details of the process are poorly understood, and the elucidation of the mechanism of translation is one of the fundamental problems in molecular biology today. A recent collection of articles summarizes the field¹.

An important contribution was made by Yonath and co-workers², who showed that structures as large as the 50S ribosomal subunit would form crystals that diffract beyond 3 Å resolution. Originally it was not clear that phase information from such a large asymmetric unit could be obtained to high resolution, but the development of bright, tunable synchrotron radiation sources, large and accurate area detectors, vastly improved crystallographic computing, and the advent of cryocrystallography have all contributed to making structural studies of the ribosome more tractable. In our work, the use of anomalous scattering from the LIII edges of lanthanides and osmium has also played a critical role in obtaining phases³.

The 30S ribosomal subunit (hereafter referred to as 30S) from *Thermus thermophilus* was originally crystallized by the Puschino group in 2-methyl-2,4-pentenediol (MPD)⁴, and by Yonath and co-workers⁵ in a mixture of ethyl-butanol and ethanol. The MPD crystal form originally diffracted to about 9–12 Å resolution^{6,7}. The diffraction limit of these crystals did not improve beyond 7 Å resolution for almost a decade, but more recently both Yonath *et al.*^{8,9} and we³ obtained crystals of the MPD form that exhibit significantly improved diffraction. However, unlike the crystals obtained by the Yonath group⁹, our crystals do not require soaking in tungsten clusters or heat treatment to obtain high-resolution diffraction.

Last year, we described the structure of the 30S at 5.5 Å resolution³. We placed all seven proteins whose structures were known at the time, inferred the structure of protein S20 to be a three-helix bundle, traced the fold of an entire domain of 16S RNA, and identified a long RNA helix at the interface that contains the decoding site of the 30S. Proteins S5 and S7 were also placed in electron density maps of the 30S obtained by Yonath *et al.*⁹. We have now solved and refined the structure of the 30S at 3 Å resolution. The structure contains all of the ordered regions of 16S RNA and 20 associated proteins, constituting over 99% of the total 16S RNA and

95% of the ribosomal proteins, with the missing parts being exclusively at the termini of RNA or polypeptide chains. Here we describe the overall structural organization of the 30S subunit, and in an accompanying paper¹⁰ we describe functional insights gleaned from the structure and the interaction of antibiotics bound to the 30S. A more detailed analysis of the structure will be presented elsewhere.

Results

Crystallographic statistics are presented in Table 1. Experimentally phased maps clearly showed main-chain density for RNA and protein, individual bases (of sufficient quality to distinguish purines from pyrimidines) and large well-ordered side chains of proteins (Fig. 1). The structure was initially built from these experimental maps and was rebuilt after a round of refinement. The current model consists of nucleotides 5–1,511 of *Thermus thermophilus* 16S RNA (corresponding to 5–1,534 of *Escherichia coli* 16S RNA)¹¹ and all of the ordered regions of the associated 20 proteins. These proteins correspond to *E. coli* proteins S2–S20 and a small 26-residue peptide, Thx (ref. 12). *Thermus* does not contain S21, and in our work S1 was removed from the 30S before crystallization. The model has been refined against 3.05 Å native data, resulting in an R/R_{free} of 0.208/0.252 with good geometry. For the proteins, 95.7% of the residues were in the core or allowed regions of the Ramachandran plot, 2.4% in the generously allowed region and 1.9% in the disallowed region.

Overview of the 30S

The secondary structure diagram for 16S RNA is shown in Fig. 2a, along with the definitions for the standard helix numbering H1–H45 (ref. 13). The sequence numbering for *E. coli* is used throughout; the main difference is that *Thermus* has a shorter H6 and H10, and insertions in H9 and H33a. Insertions in *Thermus* relative to *E. coli* are indicated in the coordinates by a letter code following the practice for transfer RNA.

The overall shape of the 30S is very similar to the model derived from negatively stained electron microscopy samples and to more recent cryo-electron microscopy reconstructions¹⁴, and appears to be closer to the 50S-bound form than the different free 30S form¹⁵. The shape is largely determined by the RNA component; none of the gross morphological features is all protein. In the canonical ‘front’ view from the 50S, the tertiary fold of 16S RNA (Fig. 2b) shows the head with a beak pointing leftwards, the body with the shoulder at top left and the spur at lower left, and the platform at top right.

Individual secondary structure domains (Fig. 2a) make up each of these morphological features (Fig. 2b), consistent with proposals made in previous modelling studies^{13,16,17}. The 5' domain makes up the bulk of the body; the central domain most of the platform, and the 3' major domain constitutes the bulk of the head. The 3' minor domain is the only significant exception to this rule, as it is part of the body at the subunit interface. The four domains of the 16S RNA secondary structure radiate from a central point in the neck region of the subunit, and are especially tightly associated in this area, which is functionally the most important region of the 30S ribosomal subunit.

The distribution of proteins and RNA in the 30S is asymmetric, as was predicted from neutron scattering¹⁸. The proteins are concentrated in the top, sides and back of the 30S (Fig. 2c, d). None of the proteins binds entirely inside an RNA domain, although S20 binds between two domains (the 3' minor domain and 5' domain). The 50S interface is largely free of protein, with the exception of S12 which lies near the decoding site at the top of the long H44 that runs down the interface. Other proteins lie at the periphery of the subunit interface, allowing them to make contact with the 50S subunit. A movie of the structure is available in the Supplementary Information.

The structure of the RNA. The secondary structure of 16S RNA contains over 50 regular double helices connected by irregular single-stranded loops (Fig. 2a). In the crystal structure, many of these formally single-stranded loop regions are in fact only slightly irregular double-stranded extensions of neighbouring regular

helices. Thus, most of 16S RNA may be described as helical or approximately helical, and it is useful to consider the RNA structure as a three-dimensional arrangement of helical elements. Interactions between helical elements include vertical co-axial stacking of helices neighbouring in sequence, and horizontal packing of helices, usually between their minor grooves. Co-axial stacking of helices is very common: the helices of 16S RNA are organized into 13 groups of co-axially stacked helices and 23 unstacked helices, for a total of 36 helical elements. The packing of these helical elements largely determines the overall fold of each of the four domains of 16S RNA. Short single-stranded RNA segments make idiosyncratic long-range interactions to stabilize the packing of helical elements. Proteins also help stabilize the RNA tertiary structure by binding to two or more RNA helical elements, as described below.

Helix packing interactions. There are three types of helix-helix packing in the structure, all of which use the wide and shallow minor groove as an interaction surface (Fig. 3). The most common packing mode is docking of the minor grooves of two helices, as shown for the interaction of H6 with H8 (Fig. 3a). Usually one or both helices are distorted from canonical A-form helical geometry in order to create a larger and more complementary interaction surface. Such helical distortions are caused by the base-pair geometries of noncanonical base pairs and sometimes by a bulged-out base. Noncanonical pairs involving adenines, especially sheared G·A, A·A, and reverse-Hoogsteen U·A and C·A pairs, are particularly common and are often adjacent to create a cross-strand adenine stack motif. This motif widens the minor groove to make it more

Table 1 Crystallographic statistics

Data collection

Data set*	Beamline	Resolution limit (Å)	No. of observations	No. of unique reflections	Completeness (%)		⟨I⟩/⟨σ⟩		R _{sym} (%)	
					Overall	Outer shell	Overall	Outer shell	Overall	Outer shell
Native	ID14-4+SBC 19ID	3.05	840101	254607	94.0	81.7	12.0	1.8	10.8	49.0
Os1	SBC 19IF	3.35	1355593	199999	98.4	90.3	16.4	2.4	13.9	59.2
Os2	ID14-4	4.00	362452	118494	97.4	97.1	8.5	2.3	10.9	38.9
Os3a	ID14-4	4.30	259175	93894	96.4	99.0	6.4	1.7	18.7	53.2
Os3b	SBC 19ID	3.50	714774	175049	97.3	94.7	10.4	2.5	16.7	66.5
Os4	X25	4.50	301216	79267	93.9	65.9	11.6	3.3	9.8	22.9
Lu1a	SBC 19ID	3.35	981691	180035	97.5	83.5	16.9	3.4	10.9	37.3
Lu1b	SBC 19ID	3.20	929154	230296	98.1	97.6	10.0	2.1	14.7	60.5

Phasing

Derivative*	SOLVE†			SHARP‡		
	No. of sites	R _{crit} §	Phasing power	No. of sites	R _{crit} §	Phasing power
Os1	56	0.69	0.84	93	0.55	1.15
Os2	6	0.87	0.20	–	–	–
Os3a	49	0.70	0.61	–	–	–
Os3b	–	–	–	23	0.77	0.69
Os4	4	0.94	0.10	–	–	–
Lu1a	18	0.78	0.33	–	–	–
Lu1b	14	0.77	0.39	19	0.83	0.60

Refinement

Resolution range	99–3.05 Å
Reflections excluded for cross-validation	5%
Number of non-hydrogen atoms	
Proteins	19,198
RNA	32,470
Metals	180
R factor (conventional)	0.208
R factor (free)	0.252
Cross-validated coordinate error	0.45 Å
Deviations from ideality	
r.m.s. deviations in bond lengths	0.006 Å
r.m.s. deviations in bond angles	1.24°

* Os1, Osmium hexammine chloride; Os2, Penta-ammine (dinitrogen) osmium(II) chloride; Os3a, b, Penta-ammine (trifluorosulphonato) osmium(III) trifluoromethanesulphonate; Os4, Osmium bipyridine; Lu1a, b, lutetium chloride.

† SOLVE (to 3.35 Å): mean figure of merit: 0.48 (overall); 0.24 (outer shell).

‡ SHARP (to 3.2 Å): mean figure of merit: 0.37 (overall); 0.03 (outer shell).

§ R_{crit} is for centric reflections.

|| Phasing power is for anomalous differences.

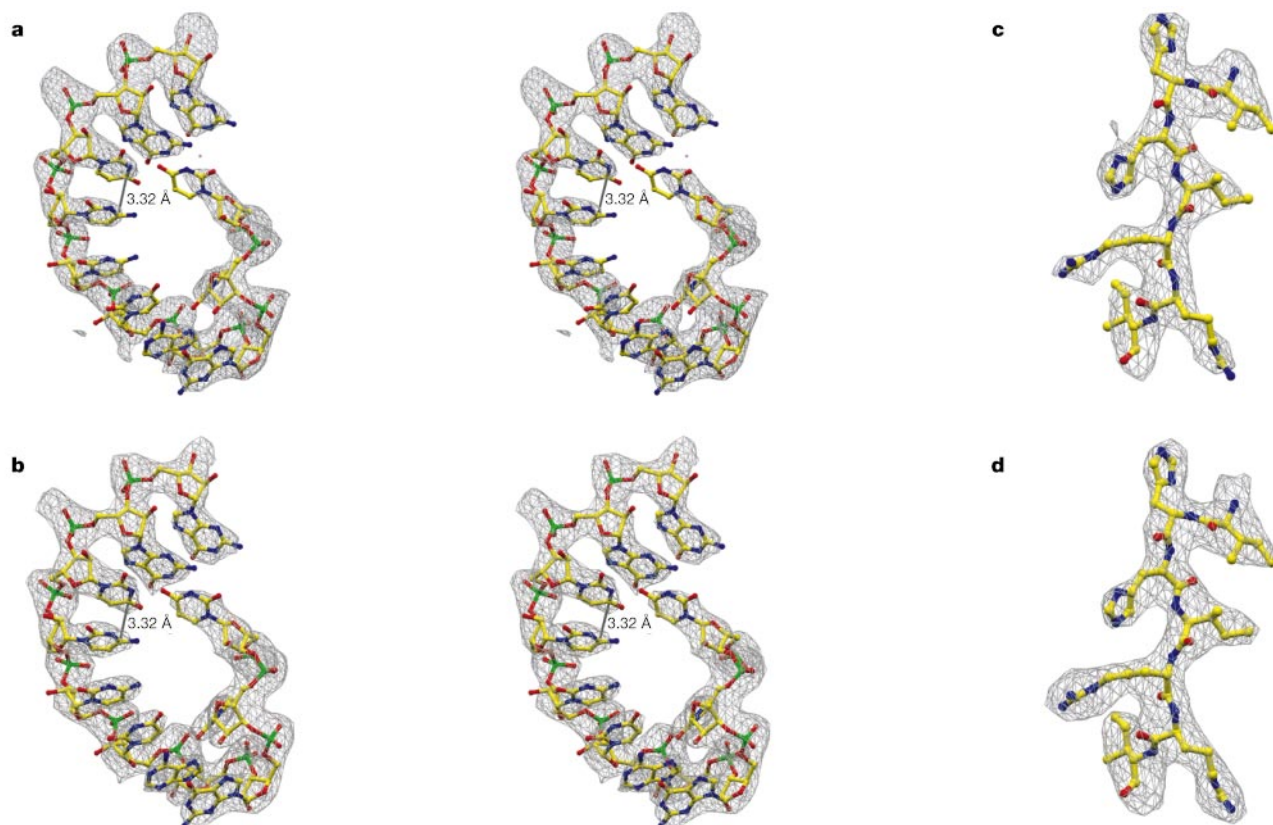


Figure 1 Electron density maps of the 30S. The experimental maps shown here were obtained by phasing with SHARP followed by density modification using SOLOMON and DM (see text); refined maps represent σ_A -weighted ($2mF_o - DF_c$) maps using the final

model. **a, b**, Experimental and refined maps respectively, of a loop of RNA. **c, d**, Experimental and refined maps of a strand of protein showing amino-acid side chains. These and other figures were made using RIBBONS³⁷.

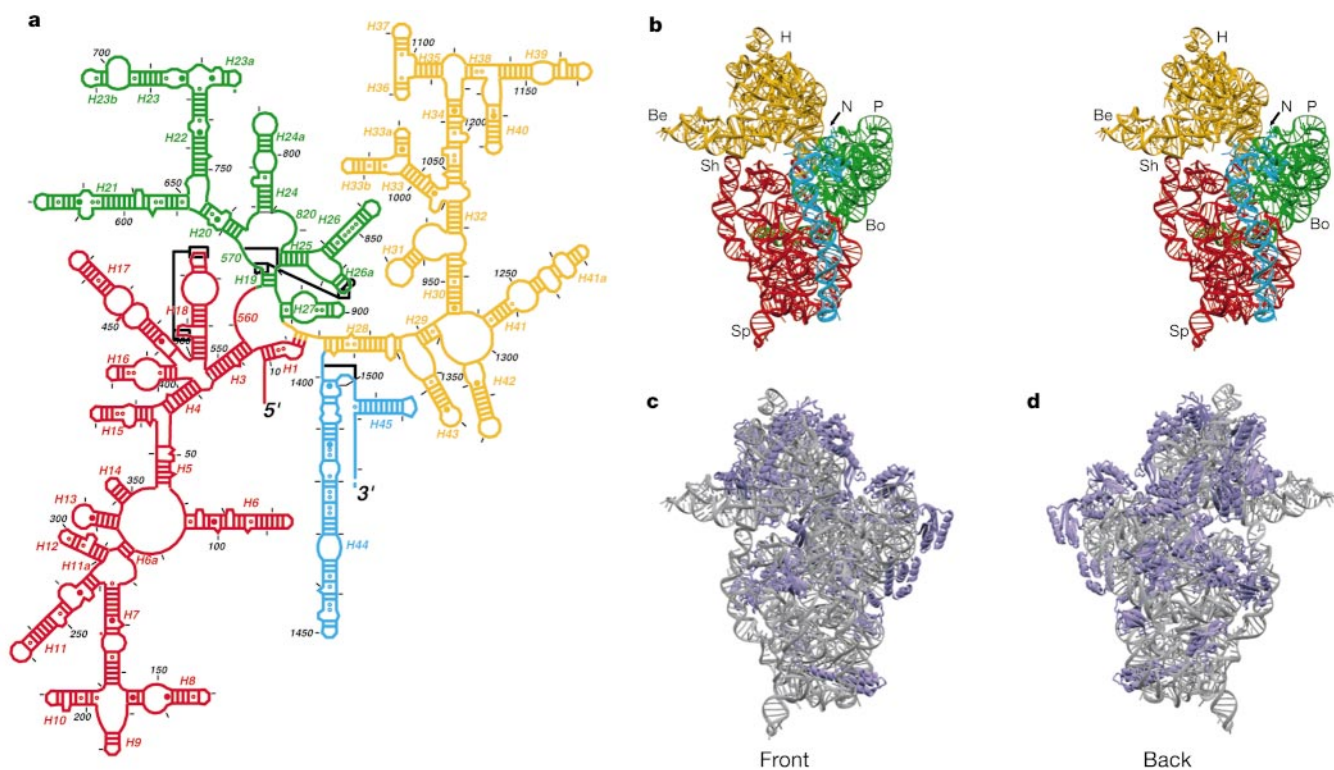


Figure 2 Overview of the 30S structure. **a**, Secondary structure diagram of 16S RNA (modified with permission from <http://www.rna.icmb.utexas.edu/CSI/2STR/Schematics/e.coli16s.27.5.5.schem.ps>; see also ref. 21), showing the definition of the various helical elements used throughout the text. The numbering and diagram correspond to the *E. coli* sequence. Red, 5' domain; green, central domain; orange, 3' major domain; cyan, 3'

minor domain. **b**, Stereo view of the tertiary structure of 16S RNA from our refined model, showing the 50S or 'front' view, with the same colouring for the domains. H, head; Be, beak; N, neck; P, platform; Sh, shoulder; Sp, spur; Bo, body. **c, d**, Front (50S) and back sides of the 30S. Grey, RNA; blue, proteins.

accessible, and it also pushes the adenine bases into the minor groove to facilitate hydrogen bonding to sugar and base functional groups from the packing partner helix. Helix–helix docking results in an extensive and intimate interaction stabilized by a dense network of hydrogen bonds from adenine base nitrogens and 2' OH moieties to the 2' OH, guanine NH₂ and pyrimidine O2 atoms from the partner helix (Fig. 3a, with adenines in red). Some of the distorted adenine-rich structures used to mediate this mode of helix–helix packing are recurrent structural motifs (for example, the common S-turn motif¹⁹). This minor-groove packing mode is not limited to helix–helix interactions: a similar mode is often seen between single-stranded adenines and a regular double helix, particularly the docking of the last three nucleotides of GNRA hairpin loops against the minor groove of a helix. Occasionally a single unpaired adenine base packs against the minor groove of a regular helix.

A second and less common form of helix packing involves the insertion of a ridge of phosphates into the minor groove of another helix. This packing mode is stabilized by hydrogen bonds between the ridge of phosphate oxygens and a layer of 2' OH and guanine base NH₂ groups, as is shown for the interaction between the phosphate backbone of H7 with the minor groove of H21 (Fig. 3b). These guanine NH₂ groups (red in Fig. 3b) are often made more accessible by the geometry of G·U wobble pairs, which places this moiety further into the minor groove compared with

Watson–Crick base pairs. This phosphate-ridge packing mode creates fewer hydrogen bonds and buries less surface area than the minor-groove mode described above, and it may be less stable.

The third and least common mode of helix packing uses an unpaired purine base to mediate the perpendicular packing of one helix against the minor groove of another helix. Three examples of this mode are present in 16S RNA: H27 against the minor groove of a helical stack made up of H1 and H28 (Fig. 3c); H34 against the H35–H34–H38 stack; and H44 against H28. Like the phosphate-ridge mode, this end-on packing may be less stable than the minor-groove packing. Significantly, all of these end-on packing interactions involve functionally important helices, and in one case (H27 against H1/H28) it appears likely that the packing interaction may change as a result of a conformational switch in H27 (ref. 20).

Overview of the domains of 16S RNA

Stacking and packing of the helical elements of 16S RNA generates three compact domains (5' domain, central domain and 3' major domain) and one extended domain (3' minor domain). Packing interactions between the domains create the functionally important 50S and transfer RNA/messenger RNA binding sites. Here we provide an overview of the structure of each domain; details will be published elsewhere.

The 5' domain. The 5' domain is the RNA component of the body. It contains 19 double helices packed as a wedge-shaped mass of RNA

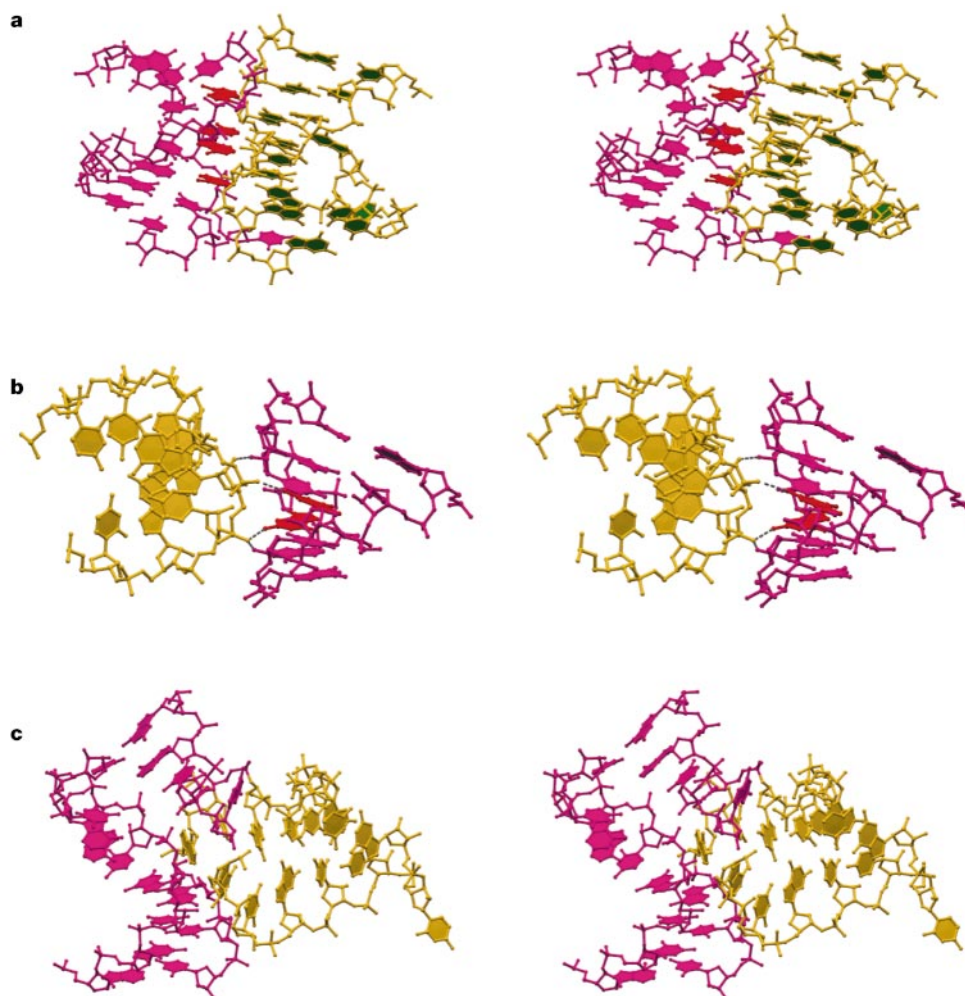


Figure 3 Different modes of interhelical packing in 16S RNA. **a**, The common minor-groove to minor-groove packing mode is often stabilized by a layer of adenosines (red), which mediate most of the hydrogen bonds between two helices (magenta and yellow). **b**, The phosphate ridge to minor groove mode. Usually this mode is stabilized by hydrogen

bonds between guanine N2 groups (red) and phosphate oxygens. **c**, The rare end-on mode of packing uses an unpaired purine (leftmost yellow base) to mediate packing of two helices at right angles to each other.

that tapers to a single layer of double helices near the top (Fig. 4). Like the other domains, it is rather longer along the subunit interface than in the perpendicular direction. The 5' domain can be divided into three subdomains, roughly corresponding to the upper, lower and middle thirds of the secondary structure of the

domain (Fig. 4b–d). These subdomains make up the top and left-hand, the middle and the lower right-hand sides of the body, respectively, in the view from 50S. The spur at the bottom of the 30S is formed by H6, which is known to vary in length across species²¹.

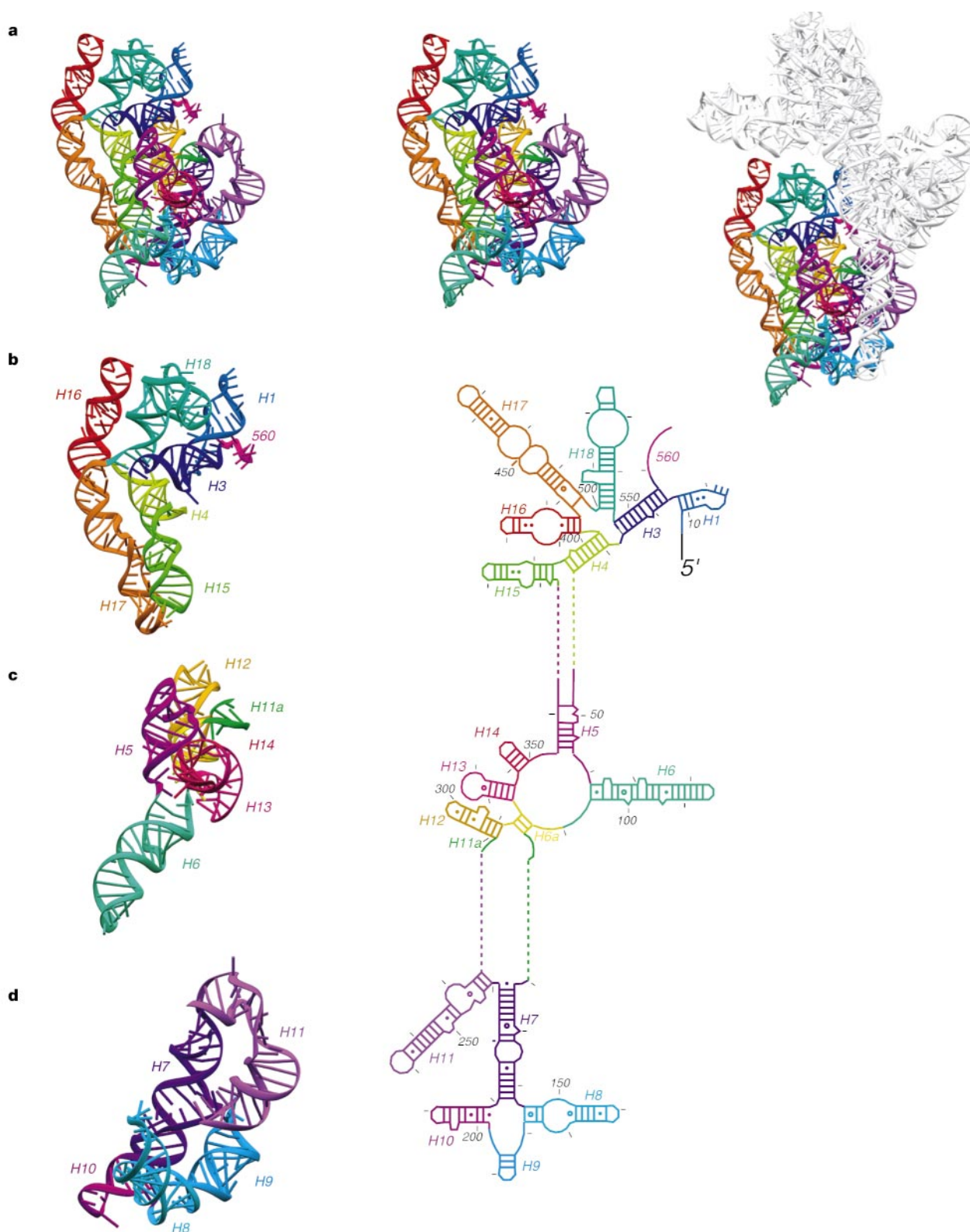


Figure 4 Structure of the 5' domain of 16S RNA. **a**, Stereo view of the entire 5' domain, with an inset on the right showing its location in the 30S subunit. The upper **(b)**, middle **(c)** and lower **(d)** subdomains are shown separately next to corresponding parts of the

secondary structure diagrams. The colours in the secondary structure diagrams match those in the structure in this and Figs 5 and 6.

A particularly striking feature is the H16–H17 co-axial stack, which is almost 120 Å long and forms the left-hand border of the body (Fig. 4b), with H16 reaching out to the head. There are also two examples of sharp bends in helices. H18 is sharply bent to accommodate the functionally important 530 pseudoknot (Fig. 4b), which packs against the central pseudoknot at the H18–H1 interface. H11 contains two sharp bends that allow its conserved terminal hairpin loop to pack against H7 (Fig. 4d). Both bends are stabilized by short-range minor-groove to minor-groove packing contacts. Finally, there is an unusual packing interaction between the highly conserved UACG and GAAA tetraloops at the ends of H8 and H14 near the subunit interface.

The central domain. The central domain is the RNA component of the platform. Its fold based on our previous 5.5 Å model³, and the high-resolution structures of parts of it^{22,23} are in excellent agreement with our current structure. It contains nine helical elements folded into a W-shape in the 50S view (Fig. 5). Two long single-stranded segments of RNA, the 570 and 820 loops, are also important structural elements. The central domain is dominated by the long stack of H21–H22–H23, which forms the outer arms of the W. At one end, H21 wraps around the back of the 5' domain; at the other, H22, H23 and the roughly parallel H24 form the bulk of the platform. The tip of the platform consists of H23B and H24A, whose conserved and functionally important hairpin loops (the 690 and 790 loops) are tightly packed. This arrangement requires sharp bends between H23 and H23B, and between H24 and H24A. The H23–H23B bend is stabilized by short-range minor groove/minor

groove packing interactions. The H24–H24A bend is more unusual in that the bend is towards the major groove, which places a ridge of H24A phosphates in the major groove of H24. This major-groove bend is stabilized partly by short-range base–base and base–backbone interactions in the major groove of the bend, and partly by long-range interactions between the bent H24/H24A minor groove and the minor groove of H23.

The 3' major domain. The 3' major domain is the RNA component of the head. The left-hand side of the head tapers to a beak made of RNA on the 50S side and protein on the solvent side (Fig. 6a). The 3' major domain consists of 15 helical elements, which can be roughly divided into three subdomains (Figs 6b–d). The upper subdomain is an extended structure in the part of the head farthest from the 50S subunit, and makes relatively few packing contacts with other RNA. The lower and middle subdomains are more globular and are more intimately packed together, and make up the front-right and front-left portions of the head, respectively. The middle subdomain includes the RNA portion of the beak.

In contrast to the extensive stacking of neighbouring helices seen in the 5' domain and the central domain, most of the helices in the 3' major domain do not stack on a neighbouring helix. An exception is the H35–H36–H38–H39 stack that dominates the upper subdomain (Fig. 6b) and stretches from the top to the bottom of the head. Significantly, the functionally important helices H31 and H34 are quite irregular and make only rather weak packing interactions with other RNA helices (Fig. 6a, c, d).

The 3' minor domain. The 3' minor domain consists of just two

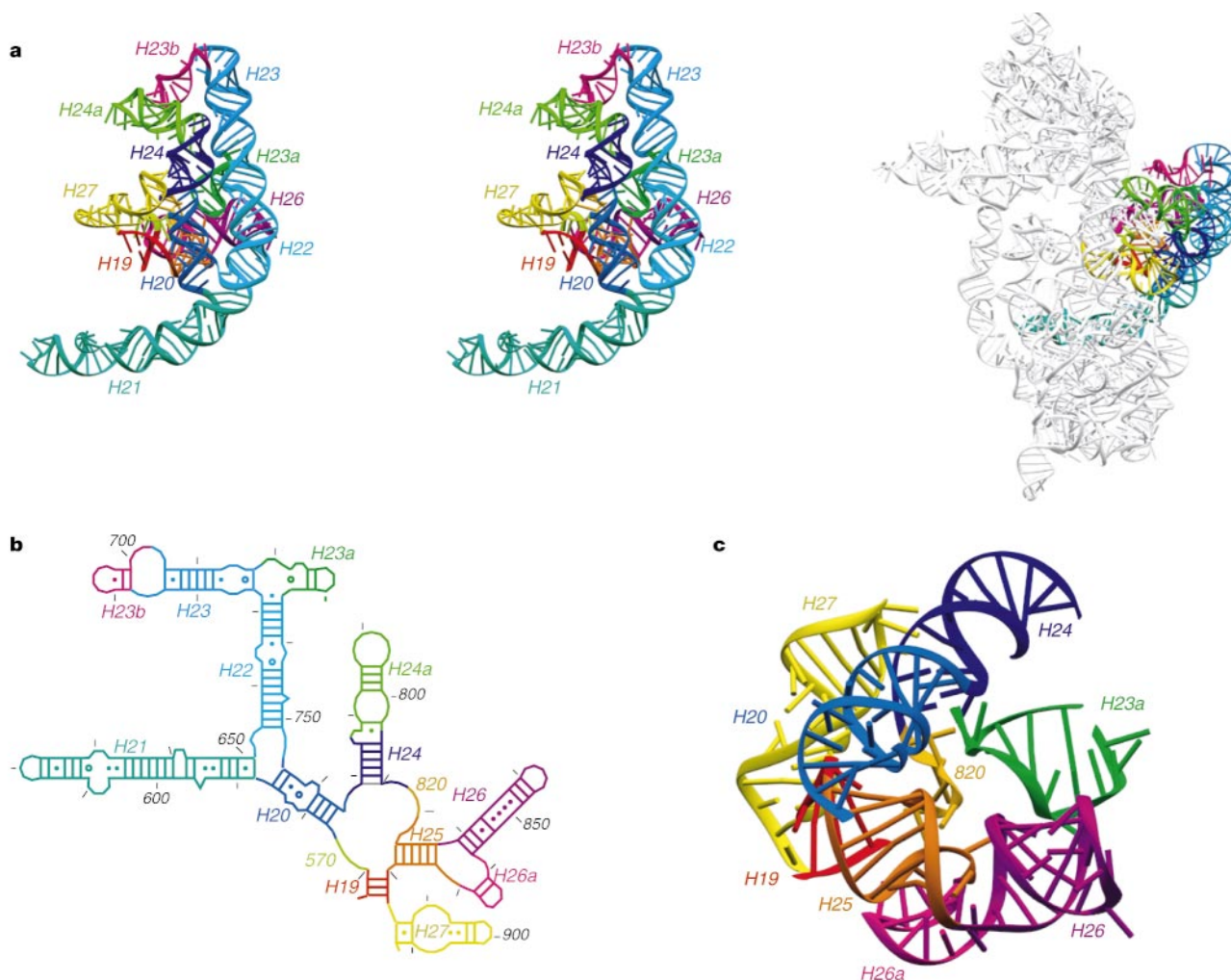


Figure 5 Structure of the central domain of 16S RNA. **a**, Stereo view of the domain with secondary structure diagram and inset showing its location in the 30S. **b**, Secondary

structure diagram for the central domain. **c**, Central portion of the domain, rotated relative to the other domains for clarity.

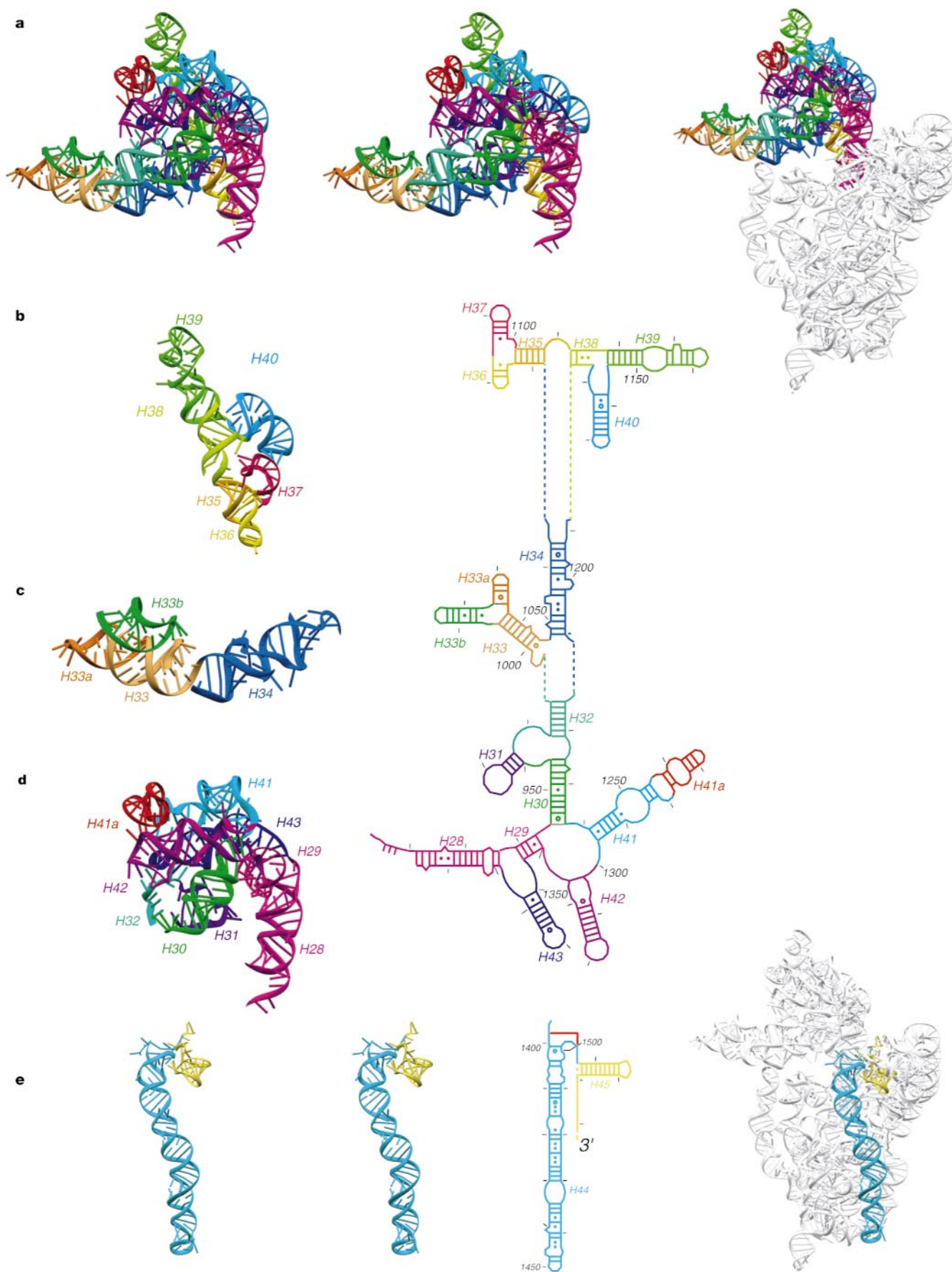


Figure 6 Structure of the 3' major and 3' minor domains of 16S RNA. **a**, Stereo view of the 3' major domain with inset showing its location in the 30S. **b–d**, The upper, middle and lower parts of the 3' major domain, with corresponding secondary structure diagrams. **e**, Stereo view of the 3' minor domain, with secondary structure diagram and inset showing its location in the 30S.

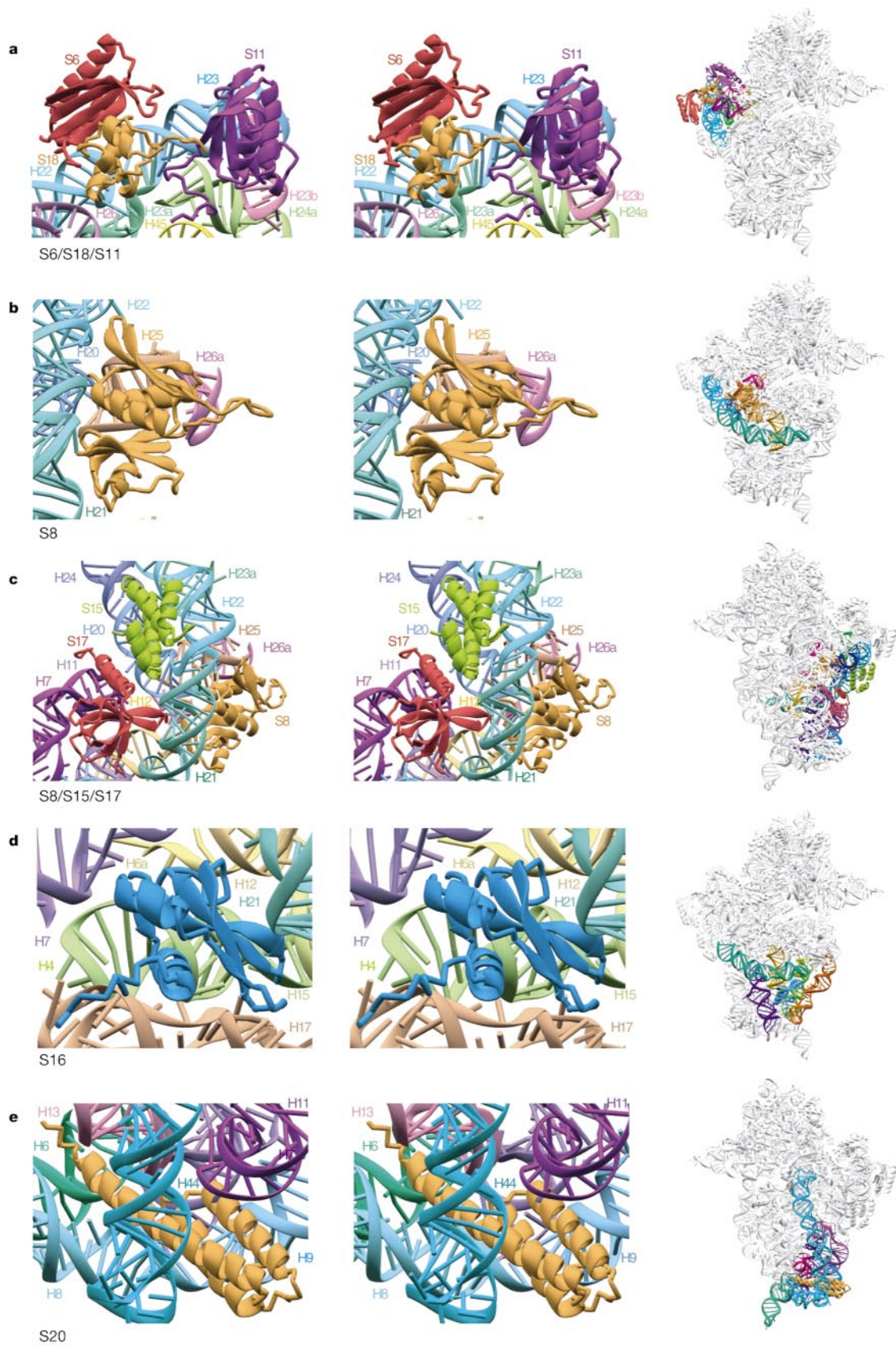


Figure 7 Proteins from the central and 5' domains. The colouring for the various RNA elements in this and subsequent figures is the same as in Figs 3–6. **a**, The S6–S18–S11 complex; **b**, S8; **c**, the S15–S17–S8 complex; **d**, S16; **e**, S20. In Figs 7–9 the stereo views have been rotated relative to the insets for clarity.

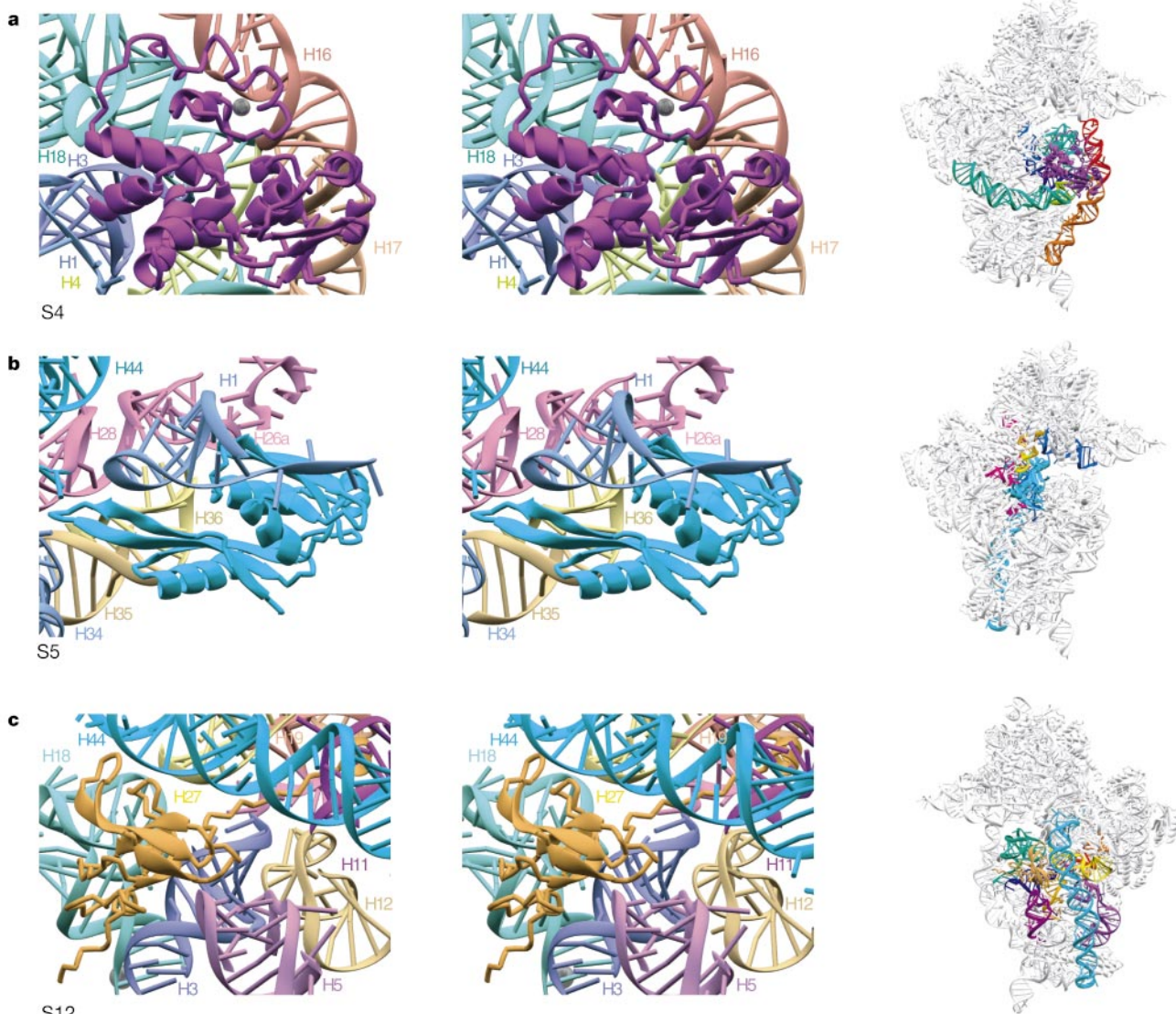
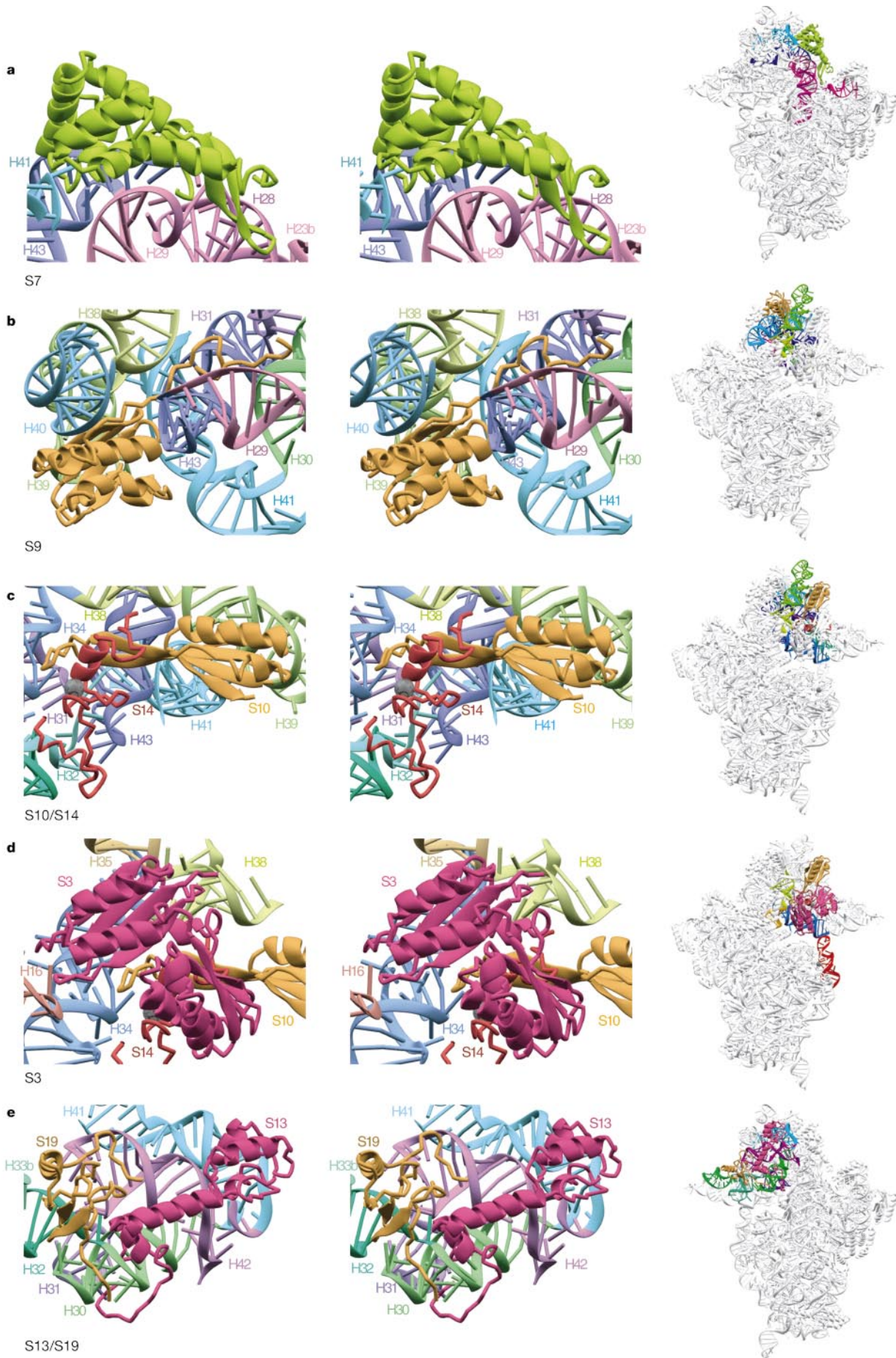


Figure 8 Proteins near the functional centre of the 30S. **a**, S4, with the Zn ion shown as a sphere; **b**, S5; **c**, S12.

Table 2 Summary of 30S protein structures

Protein	Figure no.	No. of residues (with N-terminal Met)	Visible residues	Previously known structures	No. of domains	Secondary structure in domains	RNA contacts	Protein interaction	Unusual features
S2	9f	256	V7–S235	No	2	$\alpha_2, \alpha_1\beta_5\alpha_3$	H26, H35, H36, H37, H40	None	Extended α -hairpin
S3	9d	239	G2–Q240	No	2	$\alpha_2\beta_3, \alpha_2\beta_4$	H16, H34, H35, H38, 530	S10, S14	N-terminal tail
S4	8a	209	G2–V207	Partial (refs 38, 39)	3	Zn finger, $\alpha_4, \alpha_3\beta_4$	H1, H3, H4, H16, H17, H18, H21	S5	N-terminal Zn finger
S5	8b	162	D5–G154	Yes (ref. 40)	2	$\alpha_1\beta_3, \alpha_2\beta_4$	H1, H2, H26a, H28, H34, H35, H36, H44, 560	S4, S8	extended β -hairpin
S6	7a	101	M1–A101	Yes (ref. 41)	1	$\alpha_2\beta_4$	H22, H23	S18	C-terminal tail
S7	9a	156	A2–W156	Yes (refs 42, 43)	1	$\alpha_6\beta_2$	H23b, H28, H29, H41, H43	S9, S11	extended β -hairpin
S8	7b,c	138	M1–W138	Yes (refs 44, 45)	2	$\beta_2\alpha_3, \alpha_1\beta_3$	H12, H20, H21, H22, H25, H26a	S5, S12, S17	
S9	9b	128	E2–R128	No	1	$\beta_3\alpha_3\beta_4$	H29, H30, H31, H38, H39, H40, H41, H43	S7	Long C-terminal tail
S10	9c, d	105	K3–T100	No	1	$\alpha_2\beta_4$	H31, H34, H38, H39, H41, H43	S3, S14	Long β -hairpin
S11	7a	129	K11–S129	No	1	$\alpha_2\beta_5$	H23, H23a, H23b, H24a, H45	S18, S7	Long N-terminal tail
S12	8c	135	P5–A128	No	1	$\alpha_1\beta_{/c5}$	H3, H5, H11, H12, H18, H19, H25, H27, 560, H44	S8, S17	Long N-terminal tail + extended β -hairpin loops
S13	9e	126	A2–K126	No	1	α_3	H30, H31, H41, H42	S19	Long C-terminal tail
S14	9c, d	61	A2–W61	No	1	none	H31, H32, H34, H38, 980, H42, H43	S3, S10	Zn module mostly extended
S15	7c	89	P2–G89	Yes (refs 22, 23, 46, 47)	1	α_4	H20, H22, H23a, H24	None	
S16	7d	88	M1–E83	Yes (ref. 48)	1	$\beta_1\alpha_2\beta_4$	H4, H6a, H7, H12, H15, H17, H21	None	C-terminal tail
S17	7c	105	P2–A105	Yes (ref. 49)	1	β_5	H7, H9, H11, H20, 560	None	C-terminal helix + β -hairpin loops
S18	7a	88	P16–K88	Yes (ref. 23)	1	α_4	H22, H23, H23a, H26	S6; β -Strand extends S11 sheet	Registry and loop different from ref. 23
S19	9e	93	P2–R81	Yes (ref. 50)	1	$\alpha_1\beta_3$	H30, H32, H33b, H42	S13	
S20	7e	105	R8–A106	No	1	α_3	H6, H7, H8, H11, H13, H44	None	
Thx	9g	27	G2–K25	No	None	α_1	H30, H41, H41a, H42, H43	None	Mostly extended



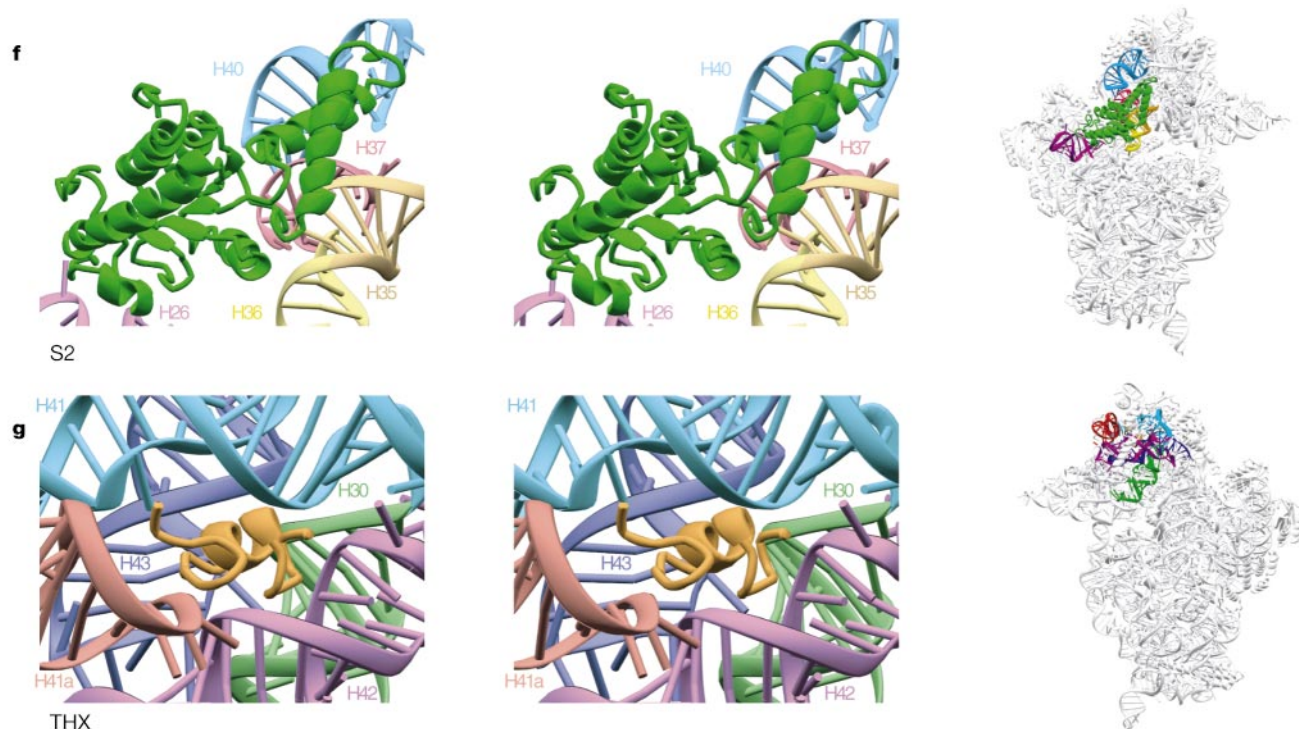


Figure 9 Proteins from the head. **a**, S7; **b**, S9; **c**, the S10–S14 complex, with the Zn ion in S14 shown as a sphere; **d**, S3 on top of S10 and S14; **e**, the S13–S19 complex; **f**, S2; **g**, Thx.

helices at the subunit interface (Fig. 6e). H44 is the longest single helix in the subunit, and stretches from the bottom of the head to the bottom of the body. It projects prominently from the body for interaction with the 50S subunit. H45 is roughly perpendicular to H44, with its conserved GGAA hairpin loop packed against H44 and available for interaction with the large subunit.

The 30S proteins and their interaction with 16S RNA

The structures of the proteins S2–S20 and Thx are summarized in Table 2 and shown with surrounding RNA and proteins in Figs 7–9. The structures of proteins solved in isolation (see refs in Table 2) were very useful in initial interpretation of the map³. In general, previous biochemical data on hydroxyl-radical footprinting²⁴ and ultraviolet-induced crosslinking (summarized in ref. 25 and http://www.mpimg-berlin-dahlem.mpg.de/~ag_ribo/ag_brimacombe/drc) agree well with the structure, and were useful as a guide to interpreting the fold at lower resolution. Finally, the structure is in good agreement with the neutron map²⁶, with S13, S14, S16, S19 and S20 being outliers. Of these, only S20 is located in a completely different region of the 30S from that predicted by the neutron map.

The proteins generally contain one or more globular domains. The same types of folds are frequently found in different proteins, such as the β -barrel in S12 and S17, and α -helices packed against a β -sheet as observed in S3, S10, S6 and S11. It is interesting that these domains often interact with RNA in very different ways. The β -sheet of S11 packs flat against the minor groove of RNA, but in S6 the edge of the sheet interacts with RNA. In S3, one of the α/β -domains makes no contact with RNA at all.

In addition to a globular domain nearly all the proteins contain long extensions. These can be helical such as the α -hairpin in S2 or the carboxy-terminal helices in S13; long β -hairpins or loops protruding from S10 or S17; or extended carboxy- or amino-terminal tails seen in proteins S4, S9, S11, S12, S13 and S19. These extensions make intimate contact with RNA, and were

generally not visible in isolated structures because they were disordered in the absence of RNA. The extensions reach far into the surrounding RNA, and allow single proteins to contact several RNA elements, which is probably important for the stabilization of the RNA tertiary fold. The extensions are particularly well suited for this task because they are narrow, allowing close approach of different RNA elements, and they have the basic residues required to neutralize the charge repulsion of the RNA backbone. An extension of this principle is seen in Thx. This small peptide fits into a cavity between multiple RNA elements in the top of the head, and its positive charge stabilizes the organization of these elements.

Proteins are often found bound to junctions between helices. S4 binds to the 5-way junction between H3, H4, H16, H17, H18 in the 5' domain, and S7 binds tightly to the junction of H29, H30, H41, H42 in the 3' major domain. Both proteins are important in early stages of 30S assembly of the body and head respectively²⁷. Similarly, proteins S8, S15 and S17 bind to the three-way junction formed by H20, H21 and H22, and are important for the assembly of the central domain^{23,28}. Thus, protein binding to helical junctions is important for initiating the correct tertiary fold of RNA. Many of the proteins also contact RNA elements in different domains, helping to organize the overall structure of the 30S. For example, S17 contacts H7 and H11 in the 5' domain, and H21 in the central domain, while S20 mediates the contact between H44 in the 3' minor domain with the 5' domain at the base of the body. Much of the contact between the head and body is mediated by proteins, such as S2 and S5. In general the structure rationalizes biochemical studies on 30S assembly^{28,29}, although there are also discrepancies. For example, it is unlikely that the binding of S20 in the lower part of the 30S could influence the binding of S13 in the head. The implications of the structure for ribosome assembly will be published elsewhere.

In addition to the intimate contacts with RNA, a number of protein–protein interactions are seen. S3, S10 and S14 form a tight cluster held together by hydrophobic interactions. They are wedged

into a V-shaped gap between two RNA domains in the head, thus helping to stabilize its structure. In contrast, other protein clusters such as S4–S5–S8 are held together by predominantly electrostatic and hydrogen-bonding interactions, and the S4–S5 interface may undergo some rearrangement during 30S function, as discussed in the accompanying paper¹⁰. The globular domain of S12 on the interface side is connected by a long extension which snakes through the body of the RNA, and then folds into a short helix which makes contacts with S8 and S17 on the back side. A similar (although less extensive) interaction is observed between the C-terminal tail of S9 and the proteins S14 and S10 on the other side of the head. The tethering together of proteins that are on opposite sides of an RNA region may help hold the RNA structure together.

The proteins contain a number of special features that are probably required for the stability of *Thermus* 30S at high temperatures. An obvious example is the extra C-terminal helix in S17, which increases the RNA contacts made by this protein. Other examples are the Zn-binding motifs found in S14 and the N terminus of S4. The Zn-binding cysteine residues are not conserved, but the surrounding residues are, indicating that metal-ion coordination is used to give extra stability to a domain that is held together by hydrophobic contacts in mesophiles such as *E. coli*.

Conclusions

This high-resolution structure of the 30S ribosomal subunit is significant for several reasons. It will allow the rationalization in structural terms of four decades of biochemical efforts to elucidate the mechanism of protein synthesis. As a first step, functional insights from the structure and a description of 30S interactions with antibiotics are presented in the accompanying paper¹⁰. The structure will facilitate the design of testable models for various aspects of ribosome function, and it will also be a basis for the interpretation of lower resolution electron-microscopic or X-ray crystallographic maps of ribosomes in different functional states. The structure of the 30S also greatly increases our current database of RNA structure and protein–RNA interactions, and may provide general rules and improve the accuracy of prediction methods for RNA tertiary structure. □

Methods

Crystallization of the 30S subunit

We obtained crystals by a straightforward optimization of Trakhanov *et al.*⁴ with respect to pH, and concentrations of Mg²⁺ ions and MPD. The final conditions were 250 mM KCl, 75 mM NH₄Cl, 25 mM MgCl₂ and 6 mM 2-mercaptoethanol in 0.1 M potassium cacodylate or 0.1 M MES (2-*N*-morpholino-ethanesulphonic acid) at pH 6.5 with 13–17% MPD as the precipitant. We noticed initially that the 30S crystals completely lacked ribosomal protein S1 (J. L. C. May and V.R., unpublished results), so we removed S1 selectively from the 30S before crystallization, which improved both size and reproducibility. The crystals grew to the maximum size in about 6 weeks at 4 °C. The largest crystals, which were required for high-resolution data collection, grew to a size of 80–100 × 80–100 × 200–300 μm. The activity of redissolved crystals in poly(U)-directed protein synthesis was comparable to that of freshly isolated 30S subunits (data not shown).

Data collection

Crystals were transferred to 26% MPD by vapour diffusion in two steps over a period of six days. All solutions (except for those containing osmium hexamine or osmium pentamine) also contained 1 mM cobalt hexamine in the cryoprotectant. Crystals were flash-cooled by plunging into liquid nitrogen, and data collection was done in a cryostream at 90–100 K. A large fraction of crystals was screened at beamlines 9.6 or 14.1 at the SRS at Daresbury Laboratories, using two short exposures at least 40 degrees apart. These crystals were then analysed for diffraction limits, cell dimensions and mosaic spread. Only crystals of similar cell dimensions and with reasonable mosaic spread were used for data collection.

Potential derivatives were screened on beamlines X25 at the NSLS at Brookhaven National Laboratory and BM-14 at the ESRF (Grenoble). Data to about 4.5 Å were obtained from X25. High-resolution data were collected at SBC 19ID at the APS in Argonne National Laboratory, and ID14-4 at the ESRF. In all cases, derivative data were collected at the peak of the fluorescence at the LIII edge to maximize anomalous differences. At X25 and SBC 19ID, the kappa goniostat was used to rotate precisely about a mirror plane so that small anomalous differences could be measured accurately despite radiation decay and the use of multiple crystals. Each crystal typically yielded 3–10 degrees of data. Data were integrated and scaled using HKL-2000 (ref. 30).

Structure determination

Previously determined phases at 5.5 Å (ref. 3) were used to locate heavy atom sites using anomalous difference Fourier maps. Initially, these sites were used for phasing to 3.35 Å using the program SOLVE³¹, followed by density modification with SOLOMON³², using the procedure implemented in SHARP³³. Optimization of the various parameters in the procedure was required to obtain interpretable maps. The RNA and some of the proteins were built using the SOLVE maps. The sequence of *Thermus thermophilus* 16S RNA¹¹ and both previously available and unpublished (Göttingen *Thermus* Genome Sequencing Project) sequences for the proteins were used to build the structure. Improved maps were obtained by calculating experimental phases using SHARP³³ followed by density modification and phase extension to 3.05 Å with SOLOMON³² and DM³⁴. The improved maps allowed us to build all the remaining ordered parts of the structure. The model was built using O (ref. 35), and refined using the program CNS³⁶. Maximum likelihood refinement was used, initially with both amplitudes and experimental phase probability distributions to 3.35 Å, and subsequently with amplitudes to 3.05 Å.

Details of the experimental protocols will be published elsewhere.

Received 14 July; accepted 10 August 2000.

- Garrett, R. A. *et al.* (eds) *The Ribosome. Structure, Function, Antibiotics and Cellular Interactions* (ASM, Washington DC, 2000).
- von Böhlen, K. *et al.* Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from *Haloarcula marismortui* diffracting to 3 Å resolution. *J. Mol. Biol.* **222**, 11–15 (1991).
- Clemons, W. M. *et al.* Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* **400**, 833–840 (1999).
- Trakhanov, S. D. Crystallization of 70S ribosomes and 30S ribosomal subunits from *Thermus thermophilus*. *FEBS Lett.* **220**, 319–322 (1987).
- Glötz, C. *et al.* Three-dimensional crystals of ribosomes and their subunits from eu- and archaeobacteria. *Biochem. Int.* **15**, 953–960 (1987).
- Yonath, A. *et al.* Characterization of crystals of small ribosomal subunits. *J. Mol. Biol.* **203**, 831–834 (1988).
- Yusupov, M. M., Tischenko, S. V., Trakhanov, S. D., Ryazantsev, S. N. & Garber, M. B. A new crystalline form of 30S ribosomal subunits from *Thermus thermophilus*. *FEBS Lett.* **238**, 113–115 (1988).
- Yonath, A. Crystallographic studies on the ribosome, a large macromolecular assembly exhibiting severe nonisomorphism, extreme beam sensitivity and no internal symmetry. *Acta Crystallogr. A* **54**, 945–955 (1998).
- Tocilj, A. *et al.* The small ribosomal subunit from *Thermus thermophilus* at 4.5 Å resolution: pattern fittings and the identification of a functional site. *Proc. Natl Acad. Sci. USA* **96**, 14252–14257 (1999).
- Carter, A. P. *et al.* Functional insights from the structure of the 30S ribosomal subunit and its interaction with antibiotics. *Nature* **407**, 340–348 (2000).
- Hartmann, R. K. & Erdmann, V. A. *Thermus thermophilus* 16S rRNA is transcribed from an isolated transcription unit. *J. Bacteriol.* **171**, 2933–2941 (1989).
- Choli, T., Franceschi, F., Yonath, A. & Wittmann-Liebold, B. Isolation and characterization of a new ribosomal protein from the thermophilic eubacteria, *Thermus thermophilus*, *T. aquaticus* and *T. flavus*. *Biol. Chem. Hoppe Seyler* **374**, 377–383 (1993).
- Mueller, F. & Brimacombe, R. A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. I. Fitting the RNA to a 3D electron microscopic map at 20 Å. *J. Mol. Biol.* **271**, 524–544 (1997).
- Gabashvili, I. S., Agrawal, R. K., Grassucci, R. & Frank, J. Structure and structural variations of the *Escherichia coli* 30S ribosomal subunit as revealed by three-dimensional cryo-electron microscopy. *J. Mol. Biol.* **286**, 1285–1291 (1999).
- Gabashvili, I. S. *et al.* Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* **100**, 537–549 (2000).
- Stern, S., Weiser, B. & Noller, H. F. Model for the three-dimensional folding of 16S ribosomal RNA. *J. Mol. Biol.* **204**, 447–481 (1988).
- Malhotra, A. & Harvey, S. C. A quantitative model of the *Escherichia coli* 16S RNA in the 30S ribosomal subunit. *J. Mol. Biol.* **240**, 308–340 (1994).
- Ramakrishnan, V. Distribution of protein and RNA in the 30S ribosomal subunit. *Science* **231**, 1562–1564 (1986).
- Wimberly, B., Varani, G. & Tinoco, I. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* **32**, 1078–1087 (1993).
- Lodmell, J. S. & Dahlberg, A. E. A conformational switch in *Escherichia coli* 16S ribosomal RNA during decoding of messenger RNA. *Science* **277**, 1262–1267 (1997).
- Gutell, R. R. in *Ribosomal RNA Structure, Evolution, Processing and Function in Protein Biosynthesis* (eds Dahlberg, A. E. & Zimmermann, R. A.) 111–128 (CRC, Boca Raton, 1996).
- Nikulin, A. *et al.* Crystal structure of the S15-rRNA complex. *Nature Struct. Biol.* **7**, 273–277 (2000).
- Agalarov, S. C., Sridhar Prasad, G., Funke, P. M., Stout, C. D. & Williamson, J. R. Structure of the S15₅₆,S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* **288**, 107–113 (2000).
- Powers, T. & Noller, H. F. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* **1**, 194–209 (1995).
- Mueller, F. & Brimacombe, R. A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. II. The RNA-protein interaction data. *J. Mol. Biol.* **271**, 545–565 (1997).
- Capel, M. S. *et al.* A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science* **238**, 1403–1406 (1987).
- Nowotny, V. & Nierhaus, K. H. Assembly of the 30S subunit from *Escherichia coli* ribosomes occurs via two assembly domains which are initiated by S4 and S7. *Biochemistry* **27**, 7051–7055 (1988).
- Stern, S., Powers, T., Changchien, L.-M. & Noller, H. F. RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* **244**, 783–790 (1989).
- Held, W. A., Ballou, B., Mizushima, S. & Nomura, M. Assembly mapping of 30S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.* **249**, 3103–3111 (1974).
- Otwinowski, Z. & Minor, W. in *Methods in Enzymology* (eds Carter, C. W. & Sweet, R. M.) 307–325 (Academic, New York, 1997).
- Terwilliger, T. & Berendzen, J. Automated MAD and MIR structure determination. *Acta Crystallogr. D* **55**, 849–861 (1999).



32. Abrahams, J. P. Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. *Acta Crystallogr. D* **53**, 371–376 (1997).
33. de la Fortelle, E. & Bricogne, G. in *Methods in Enzymology* (eds Carter, C. W. & Sweet, R. M.) 472–493 (Academic, New York, 1997).
34. Cowtan, K. & Main, P. Miscellaneous algorithms for density modification. *Acta Crystallogr. D* **54**, 487–493 (1998).
35. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277B**, 173–207 (1997).
36. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
37. Carson, M. Ribbons 2.0. *J. Appl. Cryst.* **24**, 958–961 (1991).
38. Davies, C., Gerstner, R. B., Draper, D. E., Ramakrishnan, V. & White, S. W. The crystal structure of ribosomal protein S4 reveals a two-domain molecule with an extensive RNA-binding surface: one domain shows structural homology to the ETS DNA-binding motif. *EMBO J.* **17**, 4545–4558 (1998).
39. Markus, M. A., Gerstner, R. B., Draper, D. E. & Torchia, D. A. The solution structure of ribosomal protein S4 delta41 reveals two subdomains and a positively charged surface that may interact with RNA. *EMBO J.* **17**, 4559–4571 (1998).
40. Ramakrishnan, V. & White, S. W. Structure of ribosomal protein S5 reveals sites of interaction with 16S RNA. *Nature* **358**, 768–771 (1992).
41. Lindahl, M. *et al.* Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249–1254 (1994).
42. Wimberly, B. T., White, S. W. & Ramakrishnan, V. The structure of ribosomal protein S7 at 1.9 Å resolution reveals a beta-hairpin motif that binds double-stranded nucleic acids. *Structure* **5**, 1187–1198 (1997).
43. Hosaka, H. *et al.* Ribosomal protein S7: a new RNA-binding motif with structural similarities to a DNA architectural factor. *Structure* **5**, 1199–1208 (1997).
44. Davies, C., Ramakrishnan, V. & White, S. W. Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from *Bacillus stearothermophilus* at 1.9 Å resolution. *Structure* **4**, 1093–1104 (1996).
45. Nevskaya, N. *et al.* Crystal structure of ribosomal protein S8 from *Thermus thermophilus* reveals a high degree of structural conservation of a specific RNA binding site. *J. Mol. Biol.* **279**, 233–244 (1998).
46. Berglund, H., Rak, A., Serganov, A., Garber, M. & Härd, T. Solution structure of the ribosomal RNA binding protein S15 from *Thermus thermophilus*. *Nature Struct. Biol.* **4**, 20–23 (1997).
47. Clemons, W. M., Davies, C., White, S. W. & Ramakrishnan, V. Conformational variability of the N-terminal helix in the structure of ribosomal protein S15. *Structure* **6**, 429–438 (1998).
48. Allard, P. Another piece of the ribosome: Solution structure of S16 and its location in the 30S subunit. *Structure* **8**, 875–882 (2000).
49. Golden, B. L., Hoffman, D. W., Ramakrishnan, V. & White, S. W. Ribosomal protein S17: characterization of the three-dimensional structure by ¹H- and ¹⁵N-NMR. *Biochemistry* **32**, 12812–12820 (1993).
50. Helgstrand, M. *et al.* Solution structure of the ribosomal protein S19 from *Thermus thermophilus*. *J. Mol. Biol.* **292**, 1071–1081 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements

This work was supported by the Medical Research Council (UK) and a US National Institutes of Health grant to V.R. and S. W. White. Beamlines at Argonne and Brookhaven were supported by the US Department of Energy. D.E.B. was supported by an EMBO long-term postdoctoral fellowship, and W.M.C. by an NIH predoctoral fellowship. We thank B. S. Brunschwig and M. H. Chou for gifts of osmium hexamine and osmium bipyridine; T. Terwilliger for help with phasing using SOLVE; T. A. Leaf-Jones for providing us a version of O with RNA tools; and our colleagues at the LMB for their advice and encouragement. We are indebted to A. Joachimiak, S. L. Ginell, R. Ravelli, S. McSweeney, G. Leonard, A. Thompson, H. Lewis, L. Berman, M. Papiz, S. Girdwood and M. MacDonald for help and advice on synchrotron beamlines.

Correspondence and requests for materials should be addressed to V.R. (e-mail: ramak@mrc-lmb.cam.ac.uk). Coordinates have been deposited in the Protein Data Bank, accession number 1FJF. Coordinates of individual components will be made available on <http://alf1.mrc-mb.cam.ac.uk/~ramak/30S>.