

The structure of ribosomal protein S7 at 1.9 Å resolution reveals a β -hairpin motif that binds double-stranded nucleic acids

Brian T Wimberly¹, Stephen W White^{2*} and V Ramakrishnan^{1*}

Background: Ribosomal protein S7, a crucial RNA-binding component of the ribosome, is one of two proteins that initiates assembly of the 30S ribosomal subunit. It is required for proper folding of a large 3' domain of 16S ribosomal RNA. S7 regulates its own synthesis by binding to its own mRNA. This ability of S7 to bind both messenger and ribosomal RNAs makes determination of its mode of RNA recognition particularly interesting.

Results: The crystal structure of S7 from *Thermus thermophilus* was determined by a two-wavelength anomalous diffraction experiment using the LIII edge of mercury. The S7 structure consists of a bundle of six helices and an extended β hairpin between helices 3 and 4, with two or more RNA-binding sites on its surface. The hairpin, along with portions of helices 1, 4 and 6, forms a large, positively charged, concave surface that has the appropriate curvature and dimensions to bind double-stranded RNA. A second putative RNA-binding site comprises parts of loop 2 and the helix 4–loop 5 turn.

Conclusions: Structural similarity between S7 and the IHF/HU family of proteins strongly suggests that the β hairpin of S7 binds to a groove of double-stranded RNA. The β hairpin of S7 is also similar to those from other nucleic acid binding proteins, such as ribosomal protein L14 and BIV Tat, suggesting that it belongs to an extended family of such motifs, all of which bind to a groove of double-stranded nucleic acid. The residues in S7 loop 2 that belong to the second putative RNA-binding site may have a role analogous to the N-terminal residues of IHF/HU which grip an unbent portion of double helix.

Introduction

The ribosome is a large ribonucleoprotein complex that provides the framework for protein synthesis. Its many functions include peptidyl-transferase activity as well as recognition of mRNA and cognate tRNAs. The catalytic and recognition functions of the ribosome were originally ascribed to the protein components, but more recent studies point to the crucial role played by RNA in ribosome function [1,2]. Recent structural studies on ribosomes have therefore been directed increasingly towards understanding how the well-characterized secondary structures of the 16S and 23S RNA molecules are organized into their respective tertiary structures. Many ribosomal proteins are now thought to function primarily as architectural components that direct the proper folding of ribosomal RNA and maintain its functional three-dimensional structure.

The determination of structures of ribosomal proteins is important for several reasons. Since these proteins interact specifically with ribosomal RNA, their structures provide new insight into the poorly understood fundamental principles of RNA–protein interactions. High-resolution crystal and NMR structures of individual ribosomal proteins have revealed the nature of their RNA-binding

Addresses: ¹Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84132, USA and ²Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.

*Corresponding authors.
E-mail: v.ramakrishnan@m.cc.utah.edu
stephen.white@stjude.org

Key words: multiwavelength anomalous diffraction, protein–RNA interactions, ribosome, RNA binding, X-ray crystallography

Received: 25 June 1997
Revisions requested: 31 July 1997
Revisions received: 4 August 1997
Accepted: 4 August 1997

Structure 15 September 1997, 5:1187–1198
<http://biomednet.com/elecref/0969212600501187>

© Current Biology Ltd ISSN 0969-2126

regions [3–13]. Moreover, these structures were particularly useful in biochemical and mapping studies aimed at identifying sites on ribosomal RNA that interact with specific ribosomal proteins [14,15]. In conjunction with biochemical and electron-microscopic studies, the structures of ribosomal proteins have proved useful in constraining and clarifying models of the ribosome, particularly of the 30S subunit. Since the relative locations within the ribosome of the small-subunit proteins have been determined [16,17], as well as their respective sites of interaction with 16S RNA [18–20], each known protein structure with its putative RNA-binding surface(s) represents a high-resolution probe of its immediate environment. As new structures are determined, the possible folds of the 16S RNA molecule will become increasingly restricted and better determined. The structures of some 15 small-subunit ribosomal proteins remain to be elucidated, and it is now important to concentrate on those that can provide the most useful information on the architecture of the subunit.

Ribosomal protein S7 is an essential component of the bacterial ribosome that consists of some 155 amino acids and has a molecular weight of 17,500 Daltons. It is located in the 'head' of the 30S subunit close to proteins S9, S13, S14

and S19 [16,17,21]. These proteins organize a major part of the 3' domain of 16S RNA that constitutes the RNA component of the 'head' region [22–24]. S7 is a so-called primary RNA-binding ribosomal protein in that it recognizes and binds a defined region of 16S RNA in the absence of other ribosomal proteins. It is the only primary RNA-binding protein that binds to the 3' domain of 16S RNA [25], and it appears to nucleate the folding of this half of the RNA molecule [22,24]. Consequently, it plays a major role in the assembly of the 30S head region and, along with S4, actually initiates the entire 30S assembly process [26].

Consistent with its role in nucleating the RNA-folding process, protein–RNA crosslinking studies [27–33] and RNA footprinting experiments [19,24] have shown that S7 is very close to, and probably interacts with, several defined areas of the 3' end of 16S RNA. The site to which S7 binds obviously comprises a small substructure within the 16S RNA molecule that is created from non-contiguous regions by the folding process. This substructure has recently been identified as two adjacent multistem junctions from studies with RNA fragments synthesized *in vitro* [34,35].

The S7 region of the 30S subunit is known to be close to several well-characterized sites on 16S RNA of great functional importance. In several independent experiments [36–38], S7 has been crosslinked to the anticodon loop of tRNA, therefore placing it firmly at the decoding site of 16S RNA. This has been confirmed by crosslinking studies using probes that are complementary to the 1397–1405 region of 16S RNA [39], which is a well-established landmark of the decoding site [40,41]. Other crosslinking studies using synthetic mRNA molecules place S7 on the 5' or 'exiting' side of the decoding site [42–44]. The location of the decoding site relative to that of the proofreading site is unclear, and the structure of S7 should help resolve this issue. These sites were originally thought to be some 70 Å apart, but recent data suggest that they may be adjacent or even overlap [45]. Crosslinking data have shown that the '530-loop', an accepted proofreading component, is adjacent to S7, an accepted decoding site component [46]. Clearly, S7 has to be elongated if the two sites are indeed significantly separated.

Finally, in common with several other primary RNA-binding ribosomal proteins in *Escherichia coli* [47], S7 controls the synthesis of ribosomal proteins by a feedback mechanism at the level of translation [48,49]. It specifically controls translation of the *str* operon by binding to mRNA in an intercistronic region between the S12 and S7 coding sequences in the operon [50,51]. Surprisingly, *Thermus thermophilus* S7 can bind with high affinity to the intercistronic region in *E. coli*, although the *Thermus* operon itself lacks this region [52]. This suggests that the mRNA-binding site on S7 is highly conserved, and also that an RNA structure similar to the target intercistronic site may

be present elsewhere in the *Thermus* operon. This dual (ribosomal- and messenger-) RNA-binding property of S7 makes the determination of its mode of RNA recognition and binding particularly interesting. Here, we present the 1.9 Å crystal structure of S7 from *T. thermophilus*.

Results

Crystallization of S7

Crystals of S7 were obtained in a variety of forms using PEG 8000 as a precipitant in the presence of 5–10% methylpentanediol (MPD) at pH 6.5–8.0 and temperatures from 4–25°C. The crystals were tetragonal with space group $P4_32_12$, and they varied in dimensions along the *c* axis from 102 to 122 Å, typically diffracting to about 3 Å. Optimization of the crystallization conditions, specifically by addition of various anions, resulted in improved diffraction. Crystals of *Thermus thermophilus* S7 that diffracted to 1.9 Å were obtained in 15% PEG 8000, 20% glycerol, 6% MPD, 0.1 M NaHCO₃, 10 mM K₂Pt(CN)₄, 0.1 M Tris, pH 8.0 and 5 mM 2-mercaptoethanol, at 4°C. The crystals were harvested after 3 weeks, and they had dimensions of *a* = *b* = 55.6 Å and *c* = 103.6 Å. There was no sign that the platinum in the crystallization drop was ordered in the crystal, and isomorphous crystals that diffracted similarly were subsequently produced without platinum. Co-crystals of S7 with methyl-mercury nitrate were obtained in 25% glycerol, 20% PEG 8000, 6% MPD, 0.1 M NaHCO₃, 0.1 M Tris pH 8.0 and 6 mM CH₃HgNO₃, at 13°C. These crystals had slightly different cell dimensions of *a* = *b* = 55.5 Å and *c* = 102.3 Å, and they were also harvested after 3 weeks. For data collection, crystals were plunged into liquid nitrogen and then transferred to a cold stream at 100K.

Structure determination

The non-isomorphism between the mercury co-crystals and native crystals was expected to limit the resolution of useful phase information, so we decided to use the technique of multiwavelength anomalous diffraction, or MAD [53], using the L-III edge of mercury for anomalous scattering. MAD analysis normally uses at least three wavelengths. Data at the inflection point and a remote point give rise to the largest dispersive differences, whereas data at an *f''* peak (the 'white line') near the transition contribute the maximal Bijvoet differences. Since the fluorescence spectrum at the L-III edge of most forms of mercury is rather broad with no obvious white line near the edge, it was decided that there would be no particular benefit to collecting data at a third wavelength. Accordingly, one of the wavelengths was chosen to be the inflection point at 1.01 Å, and the second wavelength was chosen to be 0.98 Å, which is sufficiently remote from the edge but still results in a large *f''* for mercury. The broad transition of mercury means that the dispersive difference term $\Delta f'$ between the wavelengths is small, which is reflected in the lower phasing power of the isomorphous contribution from the $\lambda_2 - \lambda_1$ difference (Tables 1 and 2).

Table 1

Data collection statistics.			
	Native	CH3-Hg (λ_1)	CH3-Hg (λ_2)
Source	Cu-K α	X12-C (NSLS)	X12-C (NSLS)
Wavelength (Å)	1.54	1.01	0.98
f' of Hg (electrons)	–	–12.0	–9.0
f'' of Hg (electrons)	–	7.5	9.7
Resolution (Å)	1.9	2.5	2.5
Total Reflections	103,923	26,422	44,878
Unique Reflections	13,428	5748	5721
Completeness (%)	99.8	94.9	94.9
R _{sym} (total/outer shell) [†]	0.040/0.307	0.067/0.138	0.078/0.138

[†]R_{sym} = $\sum_h \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean of the observations I_{ih} of reflection h .

Nevertheless, a two-wavelength MAD experiment was sufficient to produce excellent experimental maps. The structure was solved by treating MAD as a special case of multiple isomorphous replacement (MIR) [54,55]. The solvent-flattened experimental map (Figure 1a) yielded continuous mainchain density for the entire molecule, except near the ends (residues 7–11 and 147–150) which are disordered.

Refinement was first done against data from the mercury cocrystal and then against native data as described in Materials and methods. The current model contains S7 residues 12–146 and 152–156, and 130 water molecules, with an R_{free} of 0.284 (R = 0.229) using all data from 6.0–1.9 Å. The geometry of the current model is ‘better than average’, as judged by a G factor of 0.45 in the program PROCHECK [56] with deviations from ideality shown in Table 3. This ‘Lake Wobegon effect’ [57] is to be expected, however, given the model-building and refinement protocols used [58]. More importantly, a Ramachandran plot of the structure showed 96% of the residues in the most favored regions and the remaining 4% in the additional allowed regions, with no residues in the generously allowed or forbidden regions.

The results of the crystallographic data collection, phasing and refinement are summarized in Tables 1–3. The original

experimental MAD map and the final 2F_o–F_c map of the refined structure are shown in Figure 1.

Description of the structure

S7 consists of a bundle of six α helices, with an extended and highly twisted β hairpin between helices 3 and 4. This helical architecture appears to constitute a new fold, at least as judged by its classification in SCOP [59]. A C α trace and ribbon diagram of S7 are shown in Figure 2. The overall dimensions of S7 are 30 × 35 × 40 Å not including the β hairpin and 30 × 35 × 55 Å if the hairpin is included. The secondary structure elements are shown in Figure 3, along with the aligned sequences of S7 from several widely divergent species.

Helices

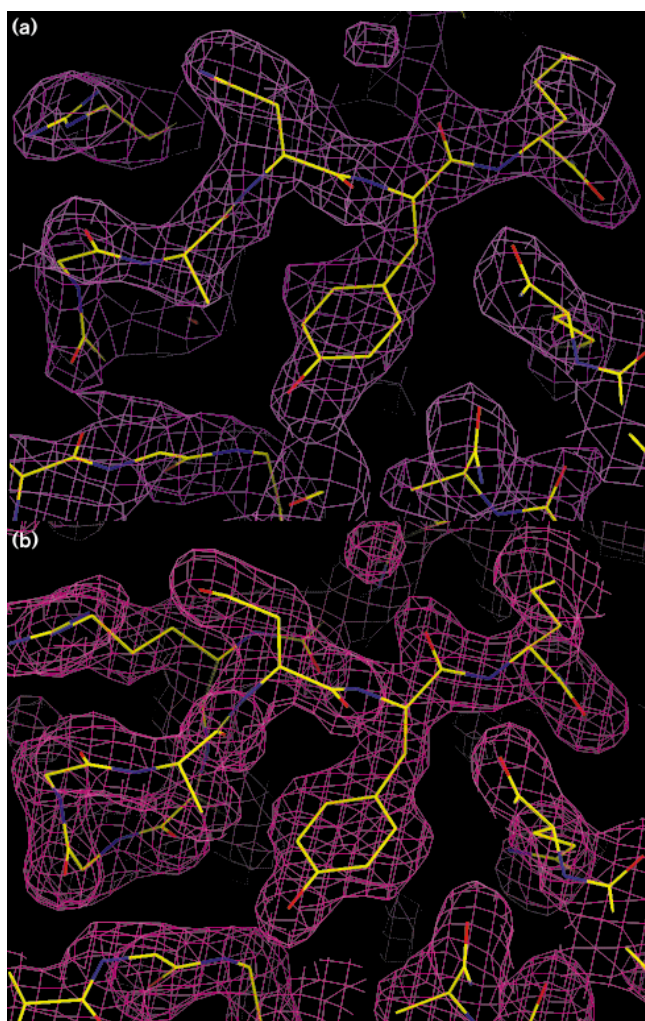
A cluster of helices 1–5 form the body of the protein. Helix 6 is not as well-packed onto the other helices; it bridges the main body of the protein and the β hairpin, and thereby helps determine the orientation of the β hairpin relative to the rest of the protein. The relative orientations of helices 1–3 are stabilized in part by an aromatic cluster consisting of Y18, F26, F43, Y44 and F62. In known S7 sequences, all of these residues are strictly conserved hydrophobic residues except Y44, which is partly exposed in our structure. Because none of these residues is strictly conserved as aromatic, part of the thermostability of S7 may arise from the

Table 2

	λ_1 – λ_2 Isomorphous		λ_1 Anomalous	λ_2 Anomalous
	Centrics	Acentrics	Acentrics	Acentrics
R _{cullis} (centrics) [†]	0.52	–	–	–
R _{kraut} [‡]	0.075	0.019	0.036	0.036
Phasing power [§]	1.2	1.7	2.9	3.2

[†]R_{cullis} = $\sum |F_{PH} - F_P| - F_H | / \sum |F_{PH} - F_P|$; [‡]R_{kraut} = $\sum |F_{PH}(\text{obs}) - F_{PH}(\text{calc})| / (F_{PH}(\text{obs}))$; [§]Phasing power = F_H / E_{RMS} , where F_P and F_{PH} are the structure factors for the ‘native’ and ‘derivative’, respectively, and E_{RMS} is the residual lack of closure. The overall figure of merit to 2.5 Å was 0.58 and 0.37 for acentric and centric reflections, respectively.

Figure 1



Experimental (a) and $2F_o - F_c$ (b) maps of S7. The region around Y85 is shown, and the maps are contoured at 1σ using the program O. The $2F_o - F_c$ map was generated from a model that had been refined to 1.9 Å.

aromatic nature of this cluster. The close approach of some of the five helices (1–5) requires the conserved alanines found at positions 39, 65, 100, 108, 117, 121 and 134, where longer residues would be sterically unfavorable. Possibly because it is more solvent-exposed than the other helices, helix 6 contains a highly conserved network of stabilizing salt bridges (E142/K138/E139/K136/D140/R143). There is a large number of additional salt bridges throughout the S7 structure that may stabilize individual helices as well as the relative orientations of helices. Many of these salt bridges are not well-conserved among S7 species, and may therefore help rationalize the thermostability of the *T. thermophilus* S7. Others are highly conserved, notably the R111–E123 and E74–R95 interactions that orient the R95 and R111 sidechains, probably for RNA binding.

Loops

Given the positions of the secondary structure elements, most of the loops are rather short and several are stabilized by conserved idiosyncratic interactions. Conserved proline residues occur at the ends of helices and may also help stabilize loops 1, 3, and 4. D15, conserved in prokaryotic sequences of S7, is buried and makes hydrogen bonds to the amide groups of residues 19 and 20 and to the –OH of Y44, thereby stabilizing loop 1. The sidechain of D20 also stabilizes loop 1 and the N terminus of helix 1 by forming a salt bridge with K63 and by hydrogen bonding to the amide group of V23. Loop 4 is stabilized by a hydrogen bond from the amide group of R72 to the sidechain of E142 (semiconserved as E/H). Loops 1 and 3 are stabilized by a hydrogen bond between the sidechains of the semiconserved residues Y18 and E57. The sidechain of R119 stabilizes loop 5 by hydrogen bonding to mainchain carbonyls of residues 113 and 114. The conformation of loop 6 is stabilized by an integral buried water molecule that hydrogen bonds to the amide protons of A134 and V135, the carbonyl of K131 and the sidechain of the conserved residue E123.

Several of the loops exhibit significant disorder. Loop 2 is quite exposed on the protein surface, and it has mainchain B factors of 55–80 Å². The C-terminal tail is also exposed and disordered — no experimental electron density is visible for residues 147–150, and residues 151–156 are visible only because they pack loosely against a symmetry-related molecule in the crystal. Five residues at the N terminus (A7–Q11) are also not visible.

The β hairpin

Proline residues that are well-conserved among S7 sequences play a crucial role in the architecture of the β hairpin. P71 and P93 start and end the β hairpin, respectively, and P88 causes the hairpin to twist. In prokaryotes, the inbound strand of the hairpin contains a preponderance of conserved hydrophobic sidechains, whereas the outbound strand contains conserved basic residues. The sequence of the hairpin is much less conserved in eukaryotic organisms.

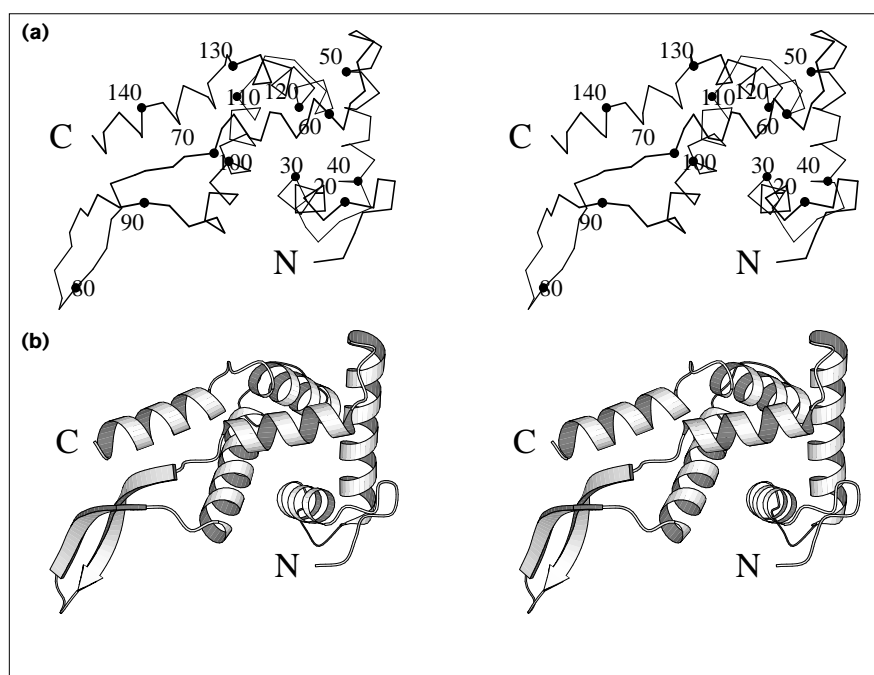
Table 3

Refinement statistics.

R factor	0.229
R_{free}	0.284
Percent of reflections used to calculate R_{free}	10
Resolution range (Å)	6.0–1.9
Number of reflections	12890
σ cut-off	0.0
Number of non-hydrogen protein atoms	1143
Number of water molecules	130
Rms deviation from ideal geometry	
bond lengths (Å)	0.007
bond angles (°)	1.2

Figure 2

The overall structure of S7. **(a)** Stereoview of a C α trace of S7, with every tenth residue shown as a small black sphere and labelled. **(b)** Stereoview of a ribbon diagram from the same view as in (a). The figure was produced using the program MOLSCRIPT [82].



Functional residues

We have shown previously that RNA-binding sites in ribosomal proteins often include a preponderance of conserved basic residues, conserved and solvent-exposed hydrophobic residues, and sites of mutations that confer antibiotic resistance [3–5,9,10,60]. Our predictions of RNA-binding regions, based on these criteria, were subsequently confirmed by biochemical experiments involving protein–RNA cross-linking [32,33] or hydroxyl-radical cleavage from tethered Fe [15]. Analysis of the S7 surface using these criteria reveals the existence of two potential RNA-binding regions. The larger of these is a large concave surface with dimensions and curvature appropriate for wrapping around double-stranded nucleic acid. This surface has a large positive electrostatic potential (Figure 4a) and a large number of conserved basic and exposed hydrophobic residues (Figure 4b). This region includes all of the β hairpin and portions of helices 1, 4 and 6. Residue K75 in *E. coli* S7 (homologous to hairpin residue R76 in *Thermus* S7) has recently been crosslinked to cytosine 1378 of 16S RNA [33], in strong support of an RNA-binding role for the β hairpin (Figure 4b).

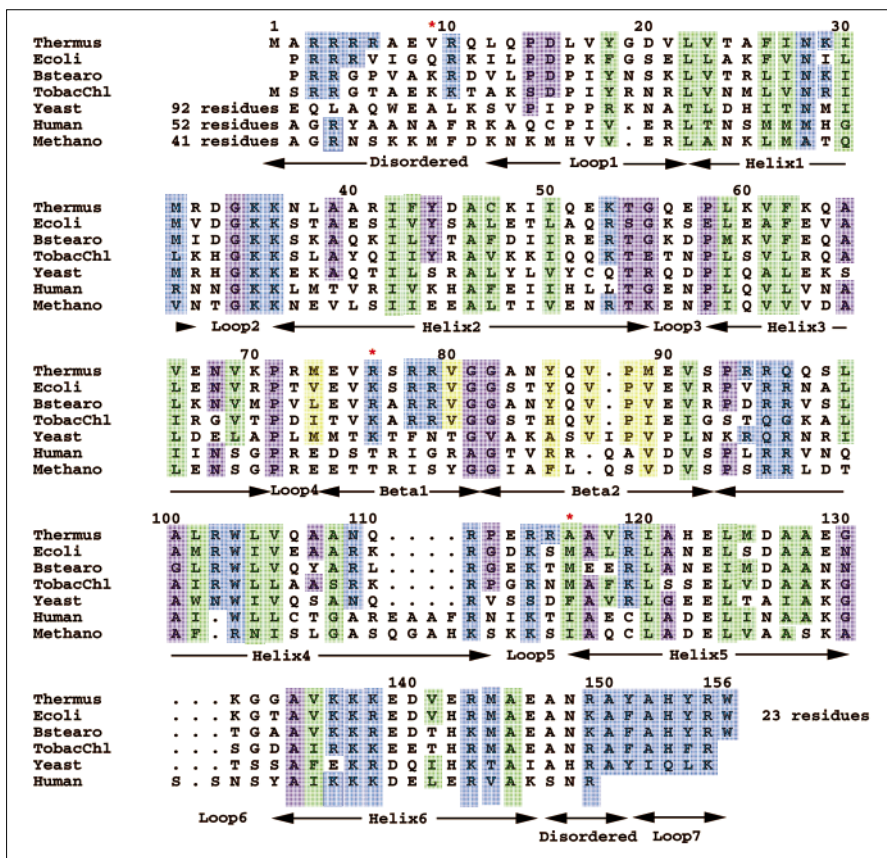
Highly conserved basic and exposed aromatic residues near the N and C termini of S7 probably form extensions of this large concave RNA-binding surface. Residues R2, R3 and R10, although not visible in the crystal structure, must be located near the concave surface, and their conservation in S7 sequences suggests that they are probably involved in RNA binding. There is experimental evidence that the N terminus of S7 is involved in binding RNA — the region

on the protein crosslinked to 16S RNA nucleotides 1377–1378 [29] has been localized to K8 in *Bacillus stearothermophilus* S7 [32]. This is the same RNA site that crosslinks to K75 in *E. coli* S7 [33], so strengthening the argument that the N terminus is close to the large concave surface that includes the β hairpin. The C-terminal tail of bacterial S7 sequences consists of highly conserved basic and aromatic residues (YAHYRW). They are separated from the rest of S7 by a disordered loop, and pack loosely against a symmetry related molecule in the crystal structure. In solution, the C-terminal tail is almost certainly disordered in the absence of RNA. In the crystal structure, the C-terminal tail constitutes another extension of the concave surface shown in Figure 4. However, given the flexibility of the intervening residues, the C terminus need not be contiguous with the large concave RNA-binding surface in the RNA-bound state.

The other region of S7 that appears to be involved in RNA binding includes parts of loop 2 and the helix 4–loop 5 turn (Figure 4b). In support of an RNA-binding role for this region, M114 of *E. coli* S7 has been cross-linked to nucleotide 1240 on 16S RNA [28,33]. The residue corresponding to M114 in *Thermus* S7 is A116, which is found in the crystal structure in the vicinity of highly conserved basic residues K35, K36 and R114, and is significantly separated from the concave RNA-binding surface described above.

Thus, it appears that S7 has at least two different regions that bind RNA. It is not clear whether these different protein regions bind distinct portions of RNA, or whether

Figure 3



Alignment of sequences of S7 from widely divergent species: *Thermus thermophilus*; *E. coli*, *Escherichia coli*; *B. stearo*, *Bacillus stearothermophilus*; *TobacChl*, *Nicotiana tabacum* chloroplast; *Yeast*, *Saccharomyces cerevisiae*; *Human*, *Homo sapiens*; and *Methano*, *Methanococcus vannielii*. Residues are shaded as follows: blue, putative RNA-binding residues; green, conserved hydrophobic core; purple, other conserved structural residues; and yellow, exposed hydrophobics in β hairpin. Red asterisks indicate the sites of crosslinking to 16S RNA. The sequences were aligned using the GCG program [83].

a contiguous RNA structure wraps itself around a substantial part of the S7 molecule.

Discussion

S7 in the 30S ribosomal subunit

Several biochemical experiments place S7 near two crucial functional locations in the 30S ribosomal subunit, the decoding and proofreading sites. Photochemical labelling using a deoxyribonucleotide probe complementary to nucleotides 518–526 of 16S RNA (the '530 loop') resulted in cross-links to proteins S3, S4, S7 and S12 [46], which suggests that all of these proteins are close to the 530 loop and within 35 Å of the probe. These proteins are not clustered tightly enough in the neutron map of ribosomal proteins in the 30S subunit to explain the S7 crosslink, however. This apparent discrepancy suggests that S7 is either highly elongated or located in a flexible region of the 30S subunit. We have shown here that the structured part of S7 is not particularly elongated, and we suggest that flexibility of the 30S subunit is the more likely explanation for the apparent discrepancy between the crosslinking data and the neutron map. Furthermore, the location of S7 near both the 530 loop, which is part of the proofreading site [46], and the anticodon of tRNA [36,37] favors the notion

that the decoding and proofreading sites may not be very far apart in the ribosome [45], given the relatively compact size of S7.

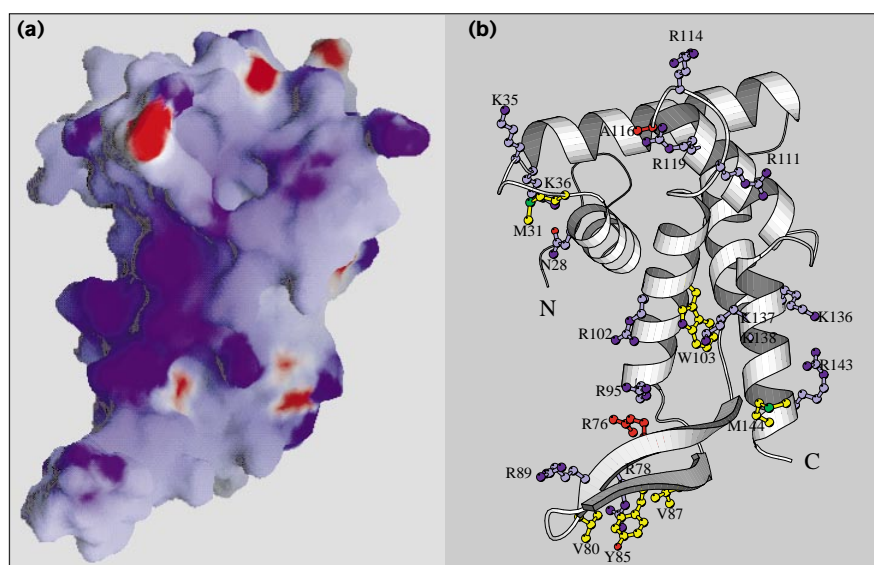
In the 30S ribosomal subunit, S7 has two near neighbors, S9 and S13 [16]. Of the two, S9 is significantly closer to S7 as judged by the center-of-mass separation. On the S7 surface, there is a cluster of exposed hydrophobic residues, conserved among S7 sequences, which includes L16, V17, Y18, L59, Y44 and possibly L12 and P14. Most of these residues (L16, V17, Y18 and L59) are well-conserved hydrophobic residues in S7 sequences from all organisms, whereas others (L12, P14 and Y44) are conserved hydrophobic residues only in prokaryotic S7 sequences (Figure 3). The universally conserved portion of the hydrophobic patch is on the side of S7 opposite its concave putative RNA-binding surface (Figure 4), and it is a likely candidate for part of the S7–S9 interface.

Structural similarity of S7 with the bacterial HU family of proteins

Despite the novelty of the six-helix architecture of S7, as judged by SCOP [59], we have found a structural similarity between S7 and the bacterial HU family of DNA-binding

Figure 4

RNA-binding regions of S7. **(a)** Electrostatic surface potential of S7. The potential displayed represents a range from -12 to $+12$ $k_B T$, shown with red as negative and blue as positive. The surface potential calculation and display was done using the program GRASP [84]. **(b)** Ribbon diagram of a similar view, showing residues that are likely to be involved in RNA-binding. Basic residues are shown in blue and solvent-exposed hydrophobic residues are shown in yellow. The red residues R76 and A116 correspond to the sites of crosslinks to 16S RNA. The figure was produced using the program MOLSCRIPT [82].



proteins. This family, which includes the non-specific DNA-binding protein HU and the specific DNA-binding protein integration host factor (IHF), control DNA architecture by introducing bends into double-stranded DNA [61]. The proteins are dimeric and each monomer has a long β hairpin that has been predicted [62,63] and recently shown [64] to interact with the DNA minor groove. Helices 2 and 3 of S7 superimpose very well with helices 1 and 2 of the HU monomer, and the β hairpin of S7 aligns closely with the HU arm (Figure 5a). Although none of the proteins' other secondary structural elements can be aligned, these three are topologically equivalent, with the rest differing only in terms of insertions and deletions (Figure 5b). In particular, the HU/IHF proteins have an insertion that functions as a dimerization surface that is lacking in the monomeric S7. A detailed analysis reveals other common features that support the significance of these gross similarities. Both protein types have distinctive aromatic clusters near the N terminus, both have a long β hairpin with an adjacent large basic surface, clearly involved in binding nucleic acid, and both have a similar β -hairpin architecture with conserved structural prolines as well as basic residues on the outward strand and conserved hydrophobic residues on the return strand. Superposition of the S7 and IHF–DNA cocrystal structures suggests that putative RNA-binding residues in S7 loop 2 may play a role analogous to the N-terminal residues of IHF and HU, which grip an unbent portion of the nucleic acid double helix [64].

The β hairpin as a nucleic acid binding motif

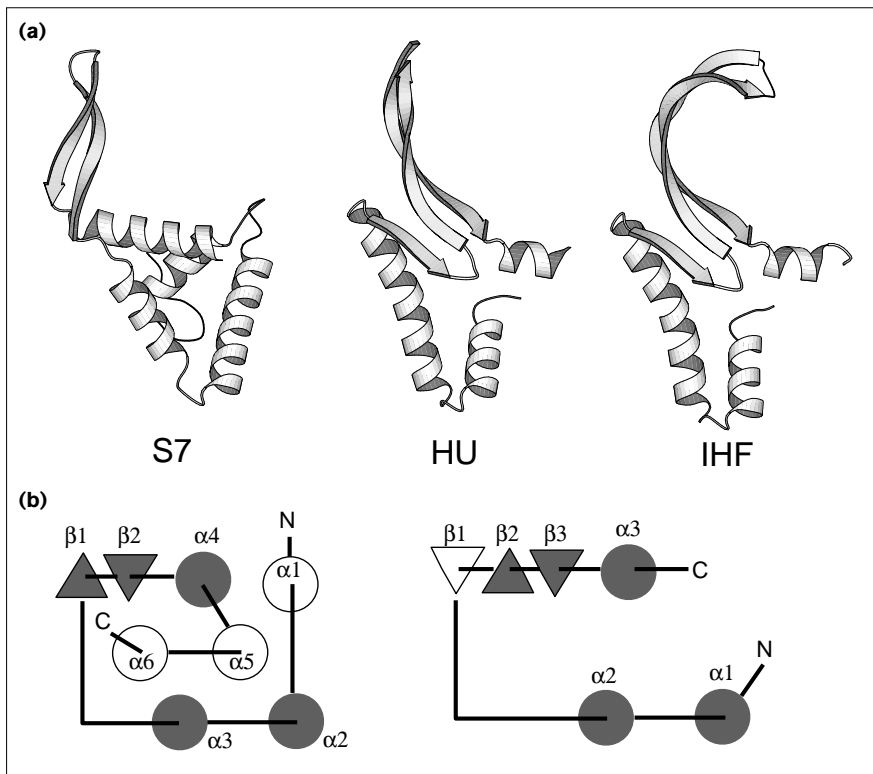
The idea that a protein β hairpin can bind a groove of double-stranded nucleic acid was proposed over twenty

years ago [65,66], but only recently has experimental evidence accumulated for its various roles in nucleic acid recognition.

In the crystal structure of S7, the β hairpin is well-ordered because it makes extensive contacts with symmetry-related neighbors. It is likely that the hairpin is highly flexible in solution. Indeed, the β hairpin of HU was found to be partly disordered in the crystal structure [62], and was subsequently shown by NMR spectroscopy to have a variable conformation in solution [67]. The flexibility of the HU arm can also be seen in the difference in the conformation of the hairpins of HU (without DNA) and IHF (with DNA) (Figure 5b). Clearly, the detailed conformation of the S7 hairpin, as well as its orientation with respect to the main body of S7, may change substantially upon binding RNA.

Nevertheless, the extensive similarities between the S7 and HU/IHF family of proteins, together with the recent structure of IHF bound to DNA, suggests a qualitative model for how the S7 β hairpin may bind RNA. In IHF/HU, conserved basic residues on the outward β hairpin strand make hydrogen bonds to DNA bases and sugars, and a pair of conserved hydrophobic residues on the return strand pry open the minor groove, thereby promoting DNA bending [63,64]. In addition, a conserved proline residue at the tip of the β hairpin intercalates between DNA bases, causing a large localized kink. Significantly, the S7 hairpin lacks the intercalating proline and also contains an arginine where steric constraints require a glycine in HU/IHF. Clearly, the details of how S7 and HU/IHF bind nucleic acid must differ. Nevertheless, the S7 hairpin does contain a pattern

Figure 5

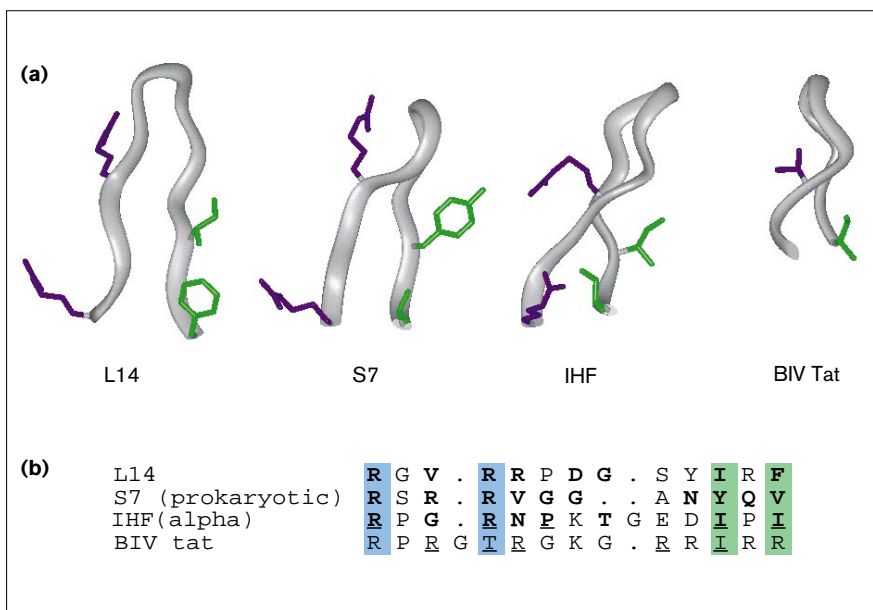


Structural homology of S7 with the HU/IHF family of DNA-binding proteins. **(a)** Ribbon diagrams showing the folds of S7 (this work), HU and IHF. Helices 5 and 6 of S7 are not shown for the sake of clarity. Nucleic acids are absent from the S7 and HU structures, whereas the IHF structure was solved in complex with DNA. The figure was produced using the program MOLSCRIPT [82]. **(b)** Topology of S7 (left) compared with HU/IHF (right), showing that the topologies are similar (shaded elements) except for the deletion or insertion of secondary structure elements.

of conserved basic and hydrophobic residues very similar to that seen in the HU/IHF arms (Figure 6). It thus appears likely that the β hairpin of S7 binds in a groove of probably somewhat irregular double-stranded RNA.

In addition to S7, HU and IHF, the β hairpin motif occurs in other proteins that have already been implicated in nucleic acid binding. In the complex of the BIV Tat peptide with Tar RNA [68,69], the Tat peptide forms a β

Figure 6



A comparison of the S7 β hairpin with those from other nucleic acid binding proteins. **(a)** Comparison of β hairpins from ribosomal protein L14, S7 (this work), IHF and BIV Tat, all shown in a similar orientation. Basic residues (blue) and exposed hydrophobics (green) that occur in a similar orientation and that are probably involved in nucleic acid binding are highlighted. Other sidechains are omitted for the sake of clarity. The figure was produced using Insight II (Biosym/MSI). **(b)** Alignment of the hairpin sequences based on the structural similarity shown in (a). Underlined residues are those with sidechains that make contact with nucleic acid in the BIV Tat and IHF structures. Conserved residues are in bold except for the BIV Tat sequence, for which too few sequences are available for a reliable determination of conservation. In defining the sequence conservation of S7, only prokaryotic sequences were considered because of considerable variation between prokaryotic and eukaryotic S7 sequences in this region. Residues are shaded with the same coloring scheme as in (a).

hairpin that sits in the major groove of an irregular RNA double strand containing a bulged-out residue. In the case of the ribosomal protein L14, a β hairpin has been implicated in RNA binding on the basis of the properties of residues in the hairpin, but it is not known whether the interaction is with the major or minor groove. In fact, the hairpins from S7, HU/IHF, BIV Tat and L14 form a continuum of structures with varying lengths and twists (Figure 6). There are common elements, such as conserved hydrophobic and basic residues in analogous locations, as well as significant differences, especially in loop lengths. The structural similarities are greatest for the S7–IHF pair, whereas the BIV Tat hairpin is significantly shorter than the others and contains some idiosyncratic elements, possibly because it recognizes some non-helical RNA features, such as the bulged-out residue. Most significantly, in the two cases for which the nucleic acid binding mode is known, one hairpin (that from IHF) interacts with the minor groove, whereas the other (that from BIV Tat) interacts with the major groove. This difference in binding mode, in spite of common structural elements with sequence conservation, suggests caution should be taken in predicting the details of how the S7 β hairpin binds RNA. More experimental data are needed to determine definitively to which groove the S7 β hairpin binds. Nevertheless, on the basis of the K75–cytosine 1378 crosslink [33] and the 5' stagger of protection in hydroxyl radical footprinting of S7 on RNA residues adjacent to this crosslink [19], a major-groove mode of binding appears more likely at present.

In summary, the structures determined so far establish the β hairpin as a family of structures that are involved in binding in a groove of double-stranded nucleic acid. Future biochemical experiments based on the structure of S7 should reveal which regions of RNA interact with the S7 hairpin, whether it binds in the major or minor groove, and whether bending occurs.

Implications of the S7 structure for the design of novel antibiotics

Because of the substantial differences between eukaryotic and prokaryotic ribosomes, many naturally occurring antibiotics work by interfering with bacterial ribosome function. Indeed, there are systematic differences between bacterial and eukaryotic S7 sequences, especially near the termini and in the β hairpin. In particular, the sequences of bacterial S7 β hairpins are highly conserved in many positions, whereas the corresponding eukaryotic residues are poorly conserved. In addition, eukaryotic S7 contains an N-terminal extension and lacks the C-terminal tail, which is highly conserved in prokaryotic S7 sequences. These differences in conservation suggest that there may be significant differences in the details of RNA binding, and that it may be possible to disrupt bacterial S7–RNA interactions without affecting the corresponding eukaryotic ones. Because S7–RNA interactions are crucial for

assembly and function of the ribosome, specific disruption of these interactions in bacterial ribosomes would seem to be an attractive target for novel antibacterial therapeutics. The structure of S7 provides a rational basis for such work.

Biological implications

The ribosome is the site of protein synthesis in all organisms. It is a large, two-subunit, complex of three RNAs and some fifty proteins. In the well-studied prokaryotic ribosome, the primary function of the ribosomal proteins appears to be stabilization of the correct three-dimensional structure of ribosomal RNA. During ribosomal assembly, a few key ribosomal proteins initiate proper folding of the ribosomal RNA by binding to conserved regions of RNA. S7 is the only such protein that interacts with the large 3' domain of 16S RNA. The binding of S7 to this domain nucleates assembly of the functionally important 'head' region of the small (30S) subunit. In addition to this role in assembly, S7 has also been identified as a component of the decoding site, a crucial region of the ribosome that monitors codon–anticodon complementarity. Finally, bacterial S7 regulates its own synthesis by binding to a site in its own mRNA.

Here, we describe the crystal structure of ribosomal protein S7 at 1.9 Å resolution. S7 consists of a bundle of six α helices and one long, twisted, β hairpin. The protein contains at least two putative RNA-binding regions. One of these is a large, concave, basic surface suitable for binding double-stranded RNA. The β hairpin is a prominent part of this surface. Striking structural similarities between S7 and the HU/IHF family of DNA-binding proteins suggest that the S7 β hairpin binds to a groove of an RNA double helix. Although HU and IHF recognize the minor groove of DNA, crosslinking and footprinting data from other workers suggest that the S7 β hairpin protects the major groove of a double-helical region of 16S rRNA. The other putative RNA-binding site comprises parts of loop 2 and the helix 4–loop 5 turn; this may have an analogous role to the N-terminal residues of IHF/HU proteins, which grip an unbent region of the nucleic acid double helix.

The crystal structure of S7 will be used to direct biochemical experiments that will improve models of the structure of the 30S ribosomal subunit. In particular, the structure will allow high-resolution biochemical mapping of the S7–rRNA environment. Such experiments have already had a major impact on evolving models of the structure of the 30S subunit; the availability of the S7 structure should allow refinement of a functionally very important part of the subunit. S7–RNA interactions are crucial for ribosome assembly and function, so specific disruption of such interactions

in bacterial ribosomes may be an attractive antibacterial target. The S7 structure may therefore provide the basis for the design of novel antibacterial agents.

Materials and methods

Purification and crystallization of S7 from *Thermus thermophilus*
The gene for S7 from *Thermus thermophilus* [70] was introduced into the T7-based expression vector pET-13a [71,72] and overexpressed in the *E. coli* strain BL21(DE3). In this construct, residues 2–6, which code for ARRNR, were deleted on the assumption that this region would be disordered in the absence of RNA and might interfere with crystallization. Cells were induced with 0.4 mM IPTG at an OD₆₀₀ of 0.8 and harvested three hours after induction. For purification, all buffers contained 0.05 mM PMSF and 6 mM 2-mercaptoethanol and all procedures were carried out at 0–4°C, unless noted. The cells were resuspended in lysis buffer (50 mM Tris, 1 mM EDTA, pH 8.0, with 0.5 mM PMSF added just before use). Although a significant amount of S7 appeared to be insoluble at low ionic strength, nearly all of the S7 was extracted from the cell lysate in soluble form by bringing the cell lysate NaCl concentration to 0.7 M. The supernatant from a low-speed spin was diluted with lysis buffer to bring the NaCl concentration to 0.5 M, and then loaded onto a cation-exchange column (Fractogel SO₃⁻, Merck) and eluted with an NaCl gradient. Fractions containing S7 were pooled, diluted with buffer containing no NaCl, and loaded onto a hydroxyapatite column (Biorad macro-prep ceramic hydroxyapatite). Elution was done with an NaCl gradient in 50 mM phosphate buffer at pH 7.0. Fractions containing S7 were pooled and concentrated and then loaded onto a gel-filtration column (Superdex-75, Pharmacia). The peak containing S7 was essentially pure as judged by a Coomassie stained gel. It was concentrated to 7 mg/ml and dialyzed into 5 mM HEPES, 0.1 M NaCl, 5 mM 2-mercaptoethanol, pH 7.5, for use in crystallization studies. Crystallization trials were done at 4 or 13°C using the hanging-drop method by mixing 3 µl of protein solution with 3 µl of the reservoir.

Data collection

Multiwavelength anomalous diffraction (MAD) data on a single methylmercury cocrystal of S7 were collected on beamline X12-C at the National Synchrotron Light Source at Brookhaven National Laboratory. The data were collected in one-degree rotations using a CCD detector with 1024 × 1024 pixels [73]. Data were collected at two wavelengths for MAD phasing using the mercury atom. The first wavelength, λ₁, was at the inflection point of the LIII-edge of mercury at 1.01 Å, whereas the second, λ₂, was at a 'remote' wavelength of 0.98 Å. Data were collected in a single continuous sweep for each wavelength. Additionally, for the second wavelength, a second sweep was collected 180° away in phi, to optimize the measurement of Bijvoet differences. Data to 1.9 Å were collected on a single crystal of native S7, using Cu-Kα radiation from a Rigaku RU-200 rotating anode with focusing mirrors. Data were collected in 0.5° rotations on an MSC RAXIS-IV detector. All data were integrated and scaled using the programs Denzo and Scalepack [74]. The integrated, scaled intensities were reduced to structure factors and scaled relative to each other using the CCP4 programs Truncate and Scaleit [75]. All data were collected at 100K using an Oxford Cryostream on crystals flash-cooled by plunging into liquid nitrogen directly from the crystallization drop.

Phasing and refinement

Phasing of multiwavelength data was done by treating data at one of the wavelengths (λ₂ = 0.98 Å) as 'native' and data at the other wavelength (λ₁ = 1.01 Å) as a 'derivative', treating MAD data as a special case of MIR [54,55]. In this procedure, data at each wavelength contributes an anomalous scattering term through Bijvoet differences, whereas the dispersive difference between wavelengths plays the role of isomorphous differences. The maximum-likelihood program Sharp [76] was used for heavy-atom refinement and phasing, and solvent flattening was carried out using the program Solomon [77,78] by launching a script from within

Sharp assuming a solvent content of 45%. The solvent-flattened map was used for model-building using the program O [79].

Refinement was initially carried out using the λ₂ data set of the mercury cocrystal and data from 6.0–2.5 Å in the program X-PLOR [80]. In the calculations, Bijvoet pairs were kept separate, and the anomalous scattering term of mercury was included in the calculation of structure factors. After initial rounds of positional, grouped B factor and torsional angle refinement, the resulting model was used to refine against data from 6.0–1.9 Å collected on a crystal of native S7. Because the dimension along the c axis of the native crystal differed from that of the mercury co-crystal, an initial rigid-body refinement was done, followed by refinement using standard protocols in X-PLOR [80]. The free R factor (R_{free}) [81] was used as a guide throughout refinement and no reflections were omitted during the refinement based on the signal-to-noise ratio.

Accession numbers

Coordinates of the S7 structure have been deposited with the Brookhaven Protein Data Bank, with accession code 1rss.

Acknowledgements

We thank RM Sweet, JM Skinner and SJ Sclafani for help and advice with MAD data collection on beamline X12-C, SE Gerchman for constructing the expression clone of S7, E de la Fortelle and J Irvin for help with the installation and use of the Sharp program, and A Kopylov for insightful comments on the manuscript. This work was supported by NIH Grant GM 44973 (to SWW and VR) and by a grant from the Lucille B Markey Charitable Trust. The diffraction facility at beamline X12-C in the National Synchrotron Light Source at Brookhaven is supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences, and by the National Science Foundation.

References

- Dahlberg, A.E. (1989). The functional role of ribosomal RNA in protein synthesis. *Cell* **57**, 525–529.
- Noller, H.F., et al., & Weiser, B. (1995). Structure and function of ribosomal RNA. *Biochem. Cell Biol.* **73**, 997–1009.
- Ramakrishnan, V. & White, S.W. (1992). Structure of ribosomal protein S5 reveals sites of interaction with 16S RNA. *Nature* **358**, 768–771.
- Golden, B.L., Ramakrishnan, V. & White, S.W. (1993). The structure of ribosomal protein L6: structural evidence of gene duplication from a primitive RNA binding protein. *EMBO J.* **12**, 4901–4908.
- Golden, B.L., Hoffman, D.W., Ramakrishnan, V. & White, S.W. (1993). Ribosomal protein S17: characterization of the three-dimensional structure by ¹H- and ¹⁵N-NMR. *Biochemistry* **32**, 12812–12820.
- Lindahl, M., et al., & Amons, R. (1994). Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249–1254.
- Nikonov, S., et al., & Liljas, A. (1996). Crystal structure of the RNA binding ribosomal protein L1 from *Thermus thermophilus*. *EMBO J.* **15**, 1350–1359.
- Jaishree, T.N., Ramakrishnan, V. & White, S.W. (1996). Solution structure of prokaryotic ribosomal protein S17 by high-resolution NMR spectroscopy. *Biochemistry* **35**, 2845–2853.
- Davies, C., White, S.W. & Ramakrishnan, V. (1996). The crystal structure of ribosomal protein L14 reveals an important organizational component of the translational apparatus. *Structure* **4**, 55–66.
- Davies, C., Ramakrishnan, V. & White, S.W. (1996). 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.
- Markus, M.A., Hinck, A.P., Huang, S., Draper, D.E. & Torchia, D.A. (1997). High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nat. Struct. Biol.* **4**, 70–77.
- Xing, Y., Guha Thakurta, D. & Draper, D.E. (1997). The RNA binding domain of ribosomal protein L11 is structurally similar to homeodomains [letter]. *Nat. Struct. Biol.* **4**, 24–27.
- Berglund, H., Rak, A., Serganov, A., Garber, M. & Härd, T. (1997). Solution structure of the ribosomal RNA binding protein S15 from *Thermus thermophilus*. *Nat. Struct. Biol.* **4**, 20–23.

14. Adamski, F.M., Atkins, J.F. & Gesteland, R.F. (1996). Ribosomal protein L9 interactions with 23 S rRNA: the use of a translational bypass assay to study the effect of amino acid substitutions. *J. Mol. Biol.* **261**, 357–371.
15. Heilek, G.M. & Noller, H.F. (1996). Site-directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S5. *Science* **272**, 1659–1662.
16. Capel, M.S., et al., & Moore, P.B. (1987). A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science* **238**, 1403–1406.
17. Stöffler-Meilicke, M. & Stöffler, G. (1987). The topography of ribosomal proteins on the surface of the 30S subunit of *Escherichia coli*. *Biochimie* **69**, 1049–1064.
18. Stern, S., Powers, T., Changchien, L.-M. & Noller, H.F. (1989). RNA–protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* **244**, 783–790.
19. Powers, T. & Noller, H.F. (1995). Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* **1**, 194–209.
20. Brimacombe, R. (1995). The structure of ribosomal RNA: a three-dimensional jigsaw puzzle. *Eur. J. Biochem.* **230**, 365–383.
21. Lambert, J.M., Boileau, G., Cover, J.A. & Traut, R.R. (1983). Cross-links between ribosomal proteins of 30S subunits in 70S tight couples and in 30S subunits. *Biochemistry* **22**, 3913–3920.
22. Wiener, L. & Brimacombe, R. (1987). Protein-binding sites on *Escherichia coli* 16S rRNA: RNA regions that are protected by proteins S7, S14 and S19 in the presence or absence of protein S9. *Nucleic Acids Res.* **15**, 3653–3670.
23. Wiener, L., Schüler, D. & Brimacombe, R. (1988). Protein-binding sites on *Escherichia coli* 16S ribosomal RNA: RNA regions that are protected by proteins S7, S9 and S19, and by proteins S8, S15 and S17. *Nucleic Acids Res.* **16**, 1233–1250.
24. Powers, T., Changchien, L.-M., Craven, G.R. & Noller, H.F. (1988). Probing the assembly of the 3' major domain of 16 S ribosomal RNA. Quaternary interactions involving ribosomal proteins S7, S9 and S19. *J. Mol. Biol.* **200**, 309–319.
25. Held, W.A., Ballou, B., Mizushima, S. & Nomura, M. (1974). Assembly mapping of 30S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.* **249**, 3103–3111.
26. Nowotny, V. & Nierhaus, K.H. (1988). 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.
27. Ehresmann, B., Reinbolt, J., Backendorf, C., Tritsch, D. & Ebel, J. (1976). Studies of the binding sites of *Escherichia coli* ribosomal protein S7 with 16S rRNA by ultraviolet irradiation. *FEBS Lett.* **67**, 316–319.
28. Möller, K., Zwieb, C. & Brimacombe, R. (1978). Identification of the oligonucleotide and oligopeptide involved in an RNA–protein crosslink induced by ultraviolet irradiation of *Escherichia coli* 30S ribosomal subunits. *J. Mol. Biol.* **126**, 489–506.
29. Wower, I. & Brimacombe, R. (1983). The localization of multiple sites on 16S rRNA which are cross-linked to proteins S7 and S8 in *Escherichia coli* 30S ribosomal subunits by treatment with 2-iminothiolane. *Nucleic Acids Res.* **11**, 1419–1437.
30. Osswald, M., et al., & Fasold, H. (1987). RNA–protein cross-linking in *Escherichia coli* 30S ribosomal subunits; determination of sites on 16S rRNA that are cross-linked to proteins S3, S4, S5, S7, S8, S9, S11, S13, S19 and S21 by treatment with methyl p-azidophenyl acetimidate. *Nucleic Acids Res.* **15**, 3221–3240.
31. Greuer, B., Osswald, M., Brimacombe, R. & Stöffler, G. (1987). RNA–protein cross-linking in *Escherichia coli* 30S ribosomal subunits: determination of sites on 16S rRNA that are cross-linked to proteins S3, S4, S7, S9, S10, S11, S17, S18 and S21 by treatment with bis-(2-chloroethyl)-methylamine. *Nucleic Acids Res.* **15**, 3241–3255.
32. Urlaub, H., Kruff, V., Bischof, O., Müller, E.C. & Wittmann-Liebold, B. (1995). Protein–rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J.* **14**, 4578–4588.
33. Urlaub, H., Thiede, B., Müller, E.C., Brimacombe, R. & Wittmann-Liebold, B. (1997). Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. *J. Biol. Chem.* **272**, 14547–14555.
34. Dragon, F. & Brakier-Gingras, L. (1993). Interaction of *Escherichia coli* ribosomal protein S7 with 16S rRNA. *Nucleic Acids Res.* **21**, 1199–1203.
35. Dragon, F., Payant, C. & Brakier-Gingras, L. (1994). Mutational and structural analysis of the RNA binding site for *Escherichia coli* ribosomal protein S7. *J. Mol. Biol.* **244**, 74–85.
36. Podkowinski, J. & Gornicki, P. (1989). Ribosomal proteins S7 and L1 are located close to the decoding site of *E. coli* ribosome – affinity labeling studies with modified tRNAs carrying photoreactive probes attached adjacent to the 3'-end of the anticodon. *Nucleic Acids Res.* **17**, 8767–8782.
37. Sylvers, L.A., Kopylov, A.M., Wower, J., Hixson, S.S. & Zimmermann, R.A. (1992). Photochemical cross-linking of the anticodon loop of yeast tRNA^{Phe} to 30S-subunit protein S7 at the ribosomal A and P sites. *Biochimie* **74**, 381–389.
38. Döring, T., Mitchell, P., Osswald, M., Bochkariov, D. & Brimacombe, R. (1994). The decoding region of 16S rRNA: a cross-linking study of the ribosomal A, P and E sites using tRNA derivatized at position 32 in the anticodon loop. *EMBO J.* **13**, 2677–2685.
39. Muralikrishna, P. & Cooperman, B.S. (1994). A photolabile oligodeoxyribonucleotide probe of the decoding site in the small subunit of the *Escherichia coli* ribosome: identification of neighboring ribosomal components. *Biochemistry* **33**, 1392–1398.
40. Oakes, M.I., Clark, M.W., Henderson, E. & Lake, J.A. (1986). DNA hybridization electron microscopy: ribosomal RNA nucleotides 1392–1407 are exposed in the cleft of the small subunit. *Proc. Natl. Acad. Sci. USA* **83**, 275–279.
41. Prince, J.B., Taylor, B.H., Thurlow, D.L., Ofengand, J. & Zimmermann, R.A. (1982). Covalent crosslinking of tRNA^{Val} to 16S rRNA at the ribosomal P site: identification of crosslinked residues. *Proc. Natl. Acad. Sci. USA* **79**, 5450–5454.
42. Stade, K., Rinke-Appel, J. & Brimacombe, R. (1989). Site-directed cross-linking of mRNA analogues to the *Escherichia coli* ribosome: identification of 30S ribosomal components that can be cross-linked to the mRNA at various points 5' with respect to the decoding site. *Nucleic Acids Res.* **17**, 9889–9908.
43. Dontsova, O., et al., & Brimacombe, R. (1992). Three widely separated positions in the 16S rRNA lie in or close to the ribosomal decoding region: a site-directed cross-linking study with mRNA analogues. *EMBO J.* **11**, 3105–3116.
44. Dontsova, O., et al., & Bogdanov, A. (1992). Identification of the *Escherichia coli* 30S ribosomal subunit protein neighboring messenger RNA during initiation of translation. *Biochimie* **74**, 363–371.
45. Brimacombe, R. (1992). Structure–function correlations (and discrepancies) in the 16S ribosomal RNA from *Escherichia coli*. *Biochimie* **74**, 319–326.
46. Alexander, R.W., Muralikrishna, P. & Cooperman, B.S. (1994). Ribosomal components neighboring the conserved 518–533 loop of 16S rRNA in 30S subunits. *Biochemistry* **33**, 12109–12118.
47. Nomura, M., Gourse, R. & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Ann. Rev. Biochem.* **53**, 75–117.
48. Nomura, M., Yates, J.L., Dean, D. & Post, L.E. (1980). Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc. Natl. Acad. Sci. USA* **77**, 7084–7088.
49. Dean, D., Yates, J.L. & Nomura, M. (1981). Identification of ribosomal protein S7 as a repressor of translation within the *str* operon of *E. coli*. *Cell* **24**, 413–419.
50. Saito, K., Mattheakis, L.C. & Nomura, M. (1994). Post-transcriptional regulation of the *str* operon in *Escherichia coli*. Ribosomal protein S7 inhibits coupled translation of S7 but not its independent translation. *J. Mol. Biol.* **235**, 111–124.
51. Saito, K. & Nomura, M. (1994). Post-transcriptional regulation of the *str* operon in *Escherichia coli*. Structural and mutational analysis of the target site for translational repressor S7. *J. Mol. Biol.* **235**, 125–139.
52. Spiridonova, V.A., Golovin, A.V., Drygin, D.Y. & Kopylov, A.M. (1997). An extremely high conservation of RNA–protein S7 interactions in prokaryotic *str* operon regulation. *FEBS Lett.*, in press.
53. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58.
54. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. & Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219–223.
55. Ramakrishnan, V. & Biou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. In *Methods of Enzymology* (Carter, C.W., Jr. & Sweet, R.M., eds), vol. **276**, pp. 538–557, Academic Press, New York.

56. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
57. Keillor, G.B. (1986). *Lake Wobegon Days*. Viking Press, New York, USA.
58. Kleywegt, G.J. & Jones, T.A. (1998). Good model-building and refinement practice. In *Methods of Enzymology* (Carter, C.W., Jr. & Sweet, R.M., eds), vol. **277**, in press, Academic Press, New York.
59. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
60. Hoffman, D.W., et al., & Ramakrishnan, V. (1994). Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein. *EMBO J.* **13**, 205–212.
61. Nash, H.A. (1996). The HU and IHF proteins: accessory factors for complex protein–DNA assemblies. In *Regulation of gene expression in Escherichia coli*. (Lin, E.C.C. & Simon Lynch, A., eds), pp. 149–179, R.G. Landes Company, Austin, Texas, USA.
62. Tanaka, I., Appelt, K., Dijk, J., White, S.W. & Wilson, K.S. (1984). 3 Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature* **310**, 376–381.
63. White, S.W., Appelt, K., Wilson, K.S. & Tanaka, I. (1989). A protein structural motif that bends DNA. *Proteins* **5**, 281–288.
64. Rice, P.A., Yang, S., Mizuuchi, K. & Nash, H.A. (1996). Crystal structure of an IHF–DNA complex: a protein-induced U-turn. *Cell* **87**, 1295–1306.
65. Carter, C.W., Jr. & Kraut, J. (1974). A proposed model for interaction of polypeptides with RNA. *Proc. Natl. Acad. Sci. USA* **71**, 283–287.
66. Church, G.M., Sussman, J.L. & Kim, S.H. (1977). Secondary structural complementarity between DNA and proteins. *Proc. Natl. Acad. Sci. USA* **74**, 1458–1462.
67. Boelens, R., Vis, H., Vorgias, C.E., Wilson, K.S. & Kaptein, R. (1996). Structure and dynamics of the DNA-binding protein HU from *Bacillus stearothermophilus* by NMR spectroscopy. *Biopolymers* **40**, 553–559.
68. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat–TAR peptide–RNA complex. *Science* **270**, 1200–1203.
69. Ye, X., Kumar, R.A. & Patel, D.J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide–TAR RNA complex. *Chem. Biol.* **2**, 827–840.
70. Yakhnin, A.V., Vorozheykina, D.P. & Matvienko, N.I. (1990). Nucleotide sequence of the *Thermus thermophilus* HB8 rps12 and rps7 genes coding for the ribosomal proteins S12 and S7. *Nucleic Acids Res.* **18**, 3659.
71. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 61–89.
72. Gerchman, S.E., Graziano, V. & Ramakrishnan, V. (1994). Expression of chicken linker histones in *E. coli*: sources of problems and methods for overcoming some of the difficulties. *Protein Express. Purif.* **5**, 242–251.
73. Stanton, M., Phillips, W.C. & O'Mara, D. (1994). CCD-based detector for X-ray crystallography. *Proc. Soc. Photo-Opt. Instr. Eng.* **2278**, 16–20.
74. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C.W.J. & Sweet, R.M., eds), vol. **276**, pp. 307–325, Academic Press, New York.
75. Collaborative Computing Project 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
76. de la Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In *Methods in Enzymology* (Carter, C.W., Jr. & Sweet, R.M., eds), vol. **276**, pp. 472–493, Academic Press, New York.
77. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Cryst. D* **52**, 30–42.
78. Abrahams, J.P. (1997). Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. *Acta Cryst. D* **53**, in press.
79. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
80. Brünger, A.T. (1988). Crystallographic refinement by simulated annealing. Application to a 2.8 Å structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.
81. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.
82. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
83. Dolz, R. (1994). GCG: production of multiple sequence alignment. *Methods Mol. Biol.* **24**, 83–99.
84. 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.