Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome

V. Ramakrishnan and Stephen W. White

Models of the bacterial ribosome based on recent structural analyses are beginning to provide new insights into the protein synthetic machinery. Central to evolving models are the high-resolution structures of individual ribosomal proteins, which represent detailed probes of their local RNA and protein environments. Ribosomal proteins are extremely ancient molecules; the structures therefore also provide a unique window into early protein evolution. Many of the proteins contain domains that are present in more recently evolved families of RNA- and DNA-binding proteins. Such structural homology can be used to predict mechanisms by which proteins interact with RNA in the ribosome.

Ribosomal protein structures have now been determined for various bacterial species, and many of these structures have been solved using a variety of methods, including X-ray crystallography, neutron diffraction, and nuclear magnetic resonance spectroscopy. These structures have provided insights into the architecture, machinery, and evolution of the ribosome, including the elucidation of the role of specific amino acid residues in protein binding and catalysis.

Ribosomal proteins are classified according to the subunit in which they reside (L or S) and their mobility on a two-dimensional gel. The proteins range in molecular weight from 5500 (L34) to 29 500 (L2). E. coli S1, which has a molecular weight of 61 000, is also involved in other RNA-associated processes and is probably not a bona fide ribosomal protein. The ribosome contains one copy of each protein, apart from L7/L12, which is present as a dimer of dimers in the so-called L7/L12 ‘stalk’. Ribosomal proteins were first purified under denaturing conditions, but problems associated with renaturation made these preparations unsuitable for structural studies. These problems were overcome by the development of non-denaturing purification protocols, which were particularly successful when applied to Thermophilic bacteria. However, progress was hindered by poor yields of proteins, and it was still impossible to isolate the molecules from the core of the thermophilic ribosomes. The real breakthrough occurred with the development of PCR techniques that allowed cloning of the ribosomal protein genes and tightly regulated bacterial systems in which the proteins could be overexpressed.

A total of 15 ribosomal-protein structures have now been determined, using both X-ray crystallography and NMR spectroscopy. Pertinent information on each of these proteins is shown in Table I. Here, we review these structures and describe how they contribute to our understanding of the architecture and function of the ribosome.

Protein structures

The concern that ribosomal proteins might adopt artificial conformations outside their normal environment has often been expressed. This is not supported by either the structures or the biochemistry of these proteins. First, all of the proteins contain well-ordered secondary and tertiary structures that are consistent with the patterns of sequence conservation and follow the accepted ‘rules’ of protein folding. It is true that many of the loop regions appear to be flexible and relatively unstructured, but these are usually the sites of interaction with RNA. NMR studies of L11 show that this type of loop becomes structured when bound to RNA, which is likely to be the case for loop regions in the other proteins.

Second, the crystal structures of S7 and S8 have been determined from two different species and in different crystal forms. They are identical apart from small variations in the loop regions.

Finally, several of the isolated proteins have been shown to retain their RNA-binding specificities. The only minor exception thus far is S15, which comprises a four-helical bundle that appears to be somewhat unstable both in solution and in the crystal.

The S’ and L’ proteins are shown separately in Figs 1 and 2, respectively. The proteins are of a wide variety of structural types, and many of the common folding motifs are represented (Table I). Consistent with their proximity to RNA, the surfaces of the proteins are highly basic and contain few areas of negative potential. None of the proteins studied thus far contains a folded domain that has a molecular mass greater than 17 kDa, and all of the larger molecules comprise two subdomains joined together. This probably reflects the extreme age of the proteins.

Very early genetic events, including gene fusion (e.g. S5 and S1) and gene duplication (e.g. L6), appear to have pre-dated the emergence of the modern ribosome. Eukaryotic ribosomal proteins are generally much larger than those of bacteria; however, homologies between the two types are clearly evident from the amino acid sequences.

Structural homologies

The topology of the first ribosomal protein determined – the C-terminal half of L7/L12 (Ref. 15) – the fold...
was somewhat novel. It was therefore significant that the second structure – that of L30 – turned out to be very similar\(^\text{11}\). This suggested that the two molecules are evolutionarily related and that ribosomal proteins might be members of only a few structural families\(^\text{20}\). Although a number of the ribosomal proteins do belong to the L7/L12 family, more interestingly, many display significant homology to modern families and could be structural prototypes. Of the structures determined thus far, only three domains have not been seen in other proteins: the C-terminal halves of S8 and L9, and the L14 β-barrel. However, several of the ribosomal-protein folds appeared to be unique, when they were first discovered, but were subsequently found to be present in other proteins. We expect this will also prove to be the case for the three folds that at present appear to be unique.

The ribosomal proteins that resemble L7/L12 are now considered to be related to the ribonucleoprotein (RNP) family of RNA-binding proteins, members of which control many aspects of RNA processing in eukaryotic cells. RNP-like domains are also present in a number of other proteins, including the ‘palm’ domain of DNA polymerase. Their basic structure is exemplified by U1A (Ref. 21) and comprises two split βαβ motifs. The four βαβ-strands form an antiparallel β-sheet, and the α-helices are packed on one side. This exact topology is found in S6 (see Fig. 1), and variations packed on one side. This exact topology that at present appear to be unique. We expect this will also prove to be unique, when they were first discovered, but were subsequently found to be present in other proteins. We expect this will also prove to be the case for the three folds that at present appear to be unique.

![Table I. Ribosomal proteins whose structures have been determined](https://example.com/table.jpg)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Species</th>
<th>X-ray/NMR Domains</th>
<th>Structural type</th>
<th>Structural similarity</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4*</td>
<td>B.s.</td>
<td>X-ray, NMR 2</td>
<td>Domain 1 – α-helical bundle</td>
<td>Tet repressor binding domain</td>
<td>Ref 4</td>
</tr>
<tr>
<td>S5</td>
<td>B.s.</td>
<td>X-ray 2 (fusion)</td>
<td>N-terminal domain – αβ</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>T1</td>
<td>B.s. + T.c.</td>
<td>X-ray 1</td>
<td>α-helical bundle + β-ribbon</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>S1</td>
<td>B.s.</td>
<td>X-ray 2 (fusion)</td>
<td>N-terminal domain – αβ</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>T1</td>
<td>B.s.</td>
<td>X-ray 2 (insertion)</td>
<td>Domain 1 – αβ</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>L6</td>
<td>B.s.</td>
<td>X-ray 2 (duplication)</td>
<td>Domain 1 – αβ</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>L7</td>
<td>B.s.</td>
<td>X-ray + NMR 2 (fusion)</td>
<td>N-terminal domain – αβ</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>L8</td>
<td>B.s.</td>
<td>X-ray 2 (insertion)</td>
<td>Domain 1 – αβ</td>
<td>L7/L12b E.c. X-ray 1</td>
<td>Ref 4</td>
</tr>
<tr>
<td>L11**</td>
<td>N.M.</td>
<td>X-ray 1</td>
<td>αβββ</td>
<td>None</td>
<td>Ref 4</td>
</tr>
<tr>
<td>L12**</td>
<td>T.c.</td>
<td>X-ray 1</td>
<td>αββββββββββββββ</td>
<td>None</td>
<td>Ref 4</td>
</tr>
<tr>
<td>L30**</td>
<td>B.s.</td>
<td>X-ray 1</td>
<td>αββββββββββββββ</td>
<td>None</td>
<td>Ref 4</td>
</tr>
</tbody>
</table>

*Only the structures of C-terminal fragments of these proteins have been solved, but those retain their RNP-binding properties. S7-L12 has two domains connected by a flexible hinge. The structure of the C-terminal half has been solved.
C. Davies et al., unpublished; M. A. Markus et al., unpublished; Abbreviations used: B.s., Bacillus stearothermophilus; X-ray, double-stranded RNP-binding domain; cT.e., E. coli RNA polymerase; EF-G, elongation factor G; G, GTP; OHA, gnas B; IF3, initiation factor 3; IF1, integration factor; MT, DNA methyltransferase; OB, oligonucleotide/oligosaccharide; RNP, ribonucleoprotein; T.c., Thermus thermophilus.

**Interactions within the ribosome**

Ribosomal proteins represent a unique opportunity for studying protein-RNA interactions at the molecular level. Thus far, no structure of a protein–rRNA complex has been determined. This is in part because it has been difficult to determine the minimal cognate fragment of RNA that interacts with each protein. During ribosome assembly, the ‘primary’ binding proteins are the first to be incorporated and recognize specific RNA sequences in order to initiate the global folding\(^\text{22}\). These primary binding proteins are followed by ‘secondary’ and ‘tertiary’ proteins that presumably bind to sites that are created by this global folding\(^\text{22}\). The primary proteins are the best candidates for studying complexes with RNA, and a number of their minimal RNA sites have now been identified. This will facilitate structural analyses of the complexes.

In the meantime, a variety of biochemical and structural data have been used to locate the putative RNA-binding sites within each protein structure. The mapping of conserved amino acid residues onto the structures has revealed that exposed basic and aromatic amino acids are frequently clustered in surface patches that probably reflect the RNA-binding sites. The structures of the U1A–RNA\(^\text{23}\) and MS2–RNA\(^\text{24}\) complexes have revealed that basic residues interact with the sugar-phosphate backbone of RNA, and aromatic residues stack onto the unpaired bases. Many of the binding sites in ribosomal RNA are characterized by single-stranded loops and
bulges that could mediate these interactions. Similar nucleotide-binding sites are also found in dsDNA-binding proteins. Because a number of ribosomal proteins are homologous to nucleic-acid-binding proteins for which structures of nucleic-acid-DNA complexes have been solved, another approach has been to use these complexes as models for ribosomal-protein-rRNA interactions. In addition, protein mutations that result in defined ribosomal phenotypes have been investigated. These mutations appear to alter the local rRNA structure and are therefore likely to be within the RNA-binding sites. Finally, NMR spectroscopy and protein–RNA crosslinking are now being used to probe the complexes directly. Putative RNA-binding sites on each of the proteins identified by combining the above approaches are shown in Figs 1 and 2.

The prediction that the N-terminal half of S5 binds to a region of dsDNA represents a good example of this combined approach. The structure of the homologous dsRBD bound to RNA has now been determined (J. Ryter and S. Schultz, pers. commun.) and, in the complex, the unstructured basic loop at the tip of the domain extends into a β-ribbon, which binds to the minor groove. Although unstructured in the absence of RNA, the corresponding loop in S5 can clearly adopt the same conformation. This loop is also the site of mutations that result in resistance to spectinomycin, an antibiotic known to bind to a region of dsDNA in 16S rRNA.

S7 is another protein that probably binds dsDNA. It is homologous to the HU/HHF family of dimeric proteins, which use a pair of β-ribbon ‘arms’ to bind DNA in the minor groove. The arms and the intervening surface are highly basic, in order to accommodate dsDNA. The surface in S7 that includes the β-ribbon has a very similar architecture, and an S7–RNA crosslink has been identified within this region.

Our knowledge of the mechanism by which the RNP family of proteins bind to RNA provides a model for the interactions between S6, L1, L6, L9, L7, L12, L22 and L30, and rRNA. The typical RNP has conserved aromatic side chains on the exposed surface of the β-sheet, which are involved in base-stacking interactions, and a ‘recognition loop’ that is encircled by, and specifically interacts with, the cognate nucleotides on the RNA. None of the ribosomal proteins has these exact RNA-binding elements, but there is evidence that they use the same region of the motif. For example, S6 shares the most structural homology with U1A and has a potential RNA-binding loop in the same region of the molecule. In L6, we have shown that gentamicin-resistant mutants contain a deletion within the C-terminal RNP-like region and a protein–RNA crosslink can also form in this part of the molecule.

Finally, the exposed surfaces of the β-sheets of S6, L1, L6, L22 and L30 all contain conserved amino acid residues suitable for interacting with RNA. The structures of homeodomain-DNA and Ets-DNA complexes provide models for the L11–RNA and S4–RNA interactions, respectively. In both protein–DNA complexes, the principal recognition element is an α-helix that inserts into the
major groove; corresponding α-helices are present in L11 and S4. In the case of L11, NMR data confirm that, in the protein–RNA complex, this α-helix is positioned at the protein–RNA interface. These data also show that an adjacent flexible loop becomes ordered upon interaction with RNA17. In S4, the RNA-binding region appears to encompass a single surface that is highly conserved and almost entirely basic (C. Davies et al., unpublished; M. A. Markus et al., unpublished). This surface is adjacent to, but does not include, the α-helix mentioned above. The putative RNA-binding region in S17 appears to comprise two, possibly flexible, loop regions that cannot be structurally defined by NMR data. The location of these loops is equivalent to that of the known nucleotide-binding sites in members of the OB family of proteins, to which S17 is homologous25. Mutations in these loops result in neamine resistance and defective ribosome assembly, and one loop can be crosslinked to rRNA30. Finally, the N-terminal domain of S8 is homologous to several DNA-binding proteins that contain an important DNA-binding loop. Protein–RNA crosslinking can occur between the corresponding loop of S8 and rRNA30. We recognize that any detailed modeling of a protein–rRNA complex is unlikely to be useful, because both components can adopt unpredictable conformations in the complex. Although the proteins are stably folded, the RNA-binding sites are frequently located in flexible loop regions that probably become structured in the complex. This is apparently the case for the N-terminal domain of S5 and has been shown directly by NMR relaxation measurements for L11 (Ref. 17). The complexity of the problem is demonstrated by the U1A–rRNA complex, where both the protein and the RNA can adopt multiple structures. Also, many of the primary ribosomal proteins can bind to rRNA as part of a translational feedback mechanism32; in several cases, it is difficult to identify any similarities between the rRNA and mRNA binding sites. However, as discussed below, the identification of putative RNA-binding sites on the proteins does provide crucial information about the structure and machinery of the ribosome. Interactions between ribosomal proteins – in particular the secondary and tertiary proteins – are likely to be important for stability of the ribosome. A number of conserved hydrophobic patches have been identified on ribosomal proteins, in particular S5, S8, L14 and L30 (see Figs 1 and 2). Like the protein–RNA interactions, the details of these interprotein interactions are impossible to predict with any degree of confidence; however, they do provide additional constraints on the models of the complete subunits. Ribosomal proteins and ribosome architecture An important role of ribosomal proteins is to direct the folding and stabilize the tertiary structure of rRNA. Consistent

Figure 2
Structures of ribosomal proteins of the large, 50S subunit. Eight structures are shown arranged around the map of the protein locations determined by protein–protein crosslinking techniques47. N- and C-termini are indicated. Homology to the ribonucleoprotein (RNP) motif is shown in yellow or red: the smallest RNP-like motif, which consists of a three-stranded β-barrel and a single α-helix, is shown in yellow. Homology to the homeodomain is shown in purple. The Rossmann fold is shown in green. Regions of no known homology are shown in gray. The locations of putative sites of interaction with RNA (R) and other ribosomal proteins (P) are indicated. The RNA sites are labeled R1 and R2 according to whether they appear to be major or minor sites, respectively.
The structures of the individual ribosomal proteins are invaluable to these modeling efforts, because each represents a high-resolution probe and measuring stick of its immediate environment. An excellent example of this is S5: an understanding of the structure of the protein allowed significant changes and improvements to be made to earlier models of the 30S subunit (circa 1989). The two putative RNA-binding sites on S5 were shown to be separated by ~30 Å (Fig. 1). Consequently, the two corresponding regions of 16S rRNA to which S5 is thought to bind had to be moved to accommodate this distance. The two 16S rRNA sites are located in the ‘head’ and ‘body’ of the 30S subunit. There is now evidence that S5 secures these major domains late in ribosome assembly and is involved in conformational changes in the 30S subunit during translation.

Future directions

Within the next few years, increasingly detailed models of the ribosome, based on images from electron microscopy and X-ray crystallography, will provide new insights into the mechanism of protein synthesis. The structures of individual ribosomal proteins will be essential for the construction of these models – providing both constraints and modules that can be fitted into the electron density. The structures will also provide tools for testing the models using methods such as site-directed hydroxyl-radical footprinting and site-directed mutagenesis of the putative functional sites on the proteins. Even more powerful will be structures of protein–rRNA complexes. No such structures are currently known, but problems in identifying and synthesizing suitable RNA fragments for analysis have largely been overcome, and structures of complexes should soon be available.

Acknowledgments

We thank A. Liljas, S. Al-Karadaghi, S. Schultz and H. Noller for unpublished data, and C. Davies, B. Wimberly and R. Kriwacki for helpful comments. We apologize to the authors of the many important articles on the ribosome that space precludes to the authors of the many important articles on the ribosome that space precludes from citing in this review.

References