SURVEY AND SUMMARY

RNA binding strategies of ribosomal proteins

David E. Draper* and Luis P. Reynaldo

Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA

Received July 29 1998; Revised October 28, 1998; Accepted November 16, 1998

ABSTRACT

Structures of a number of ribosomal proteins have now been determined by crystallography and NMR, though the complete structure of a ribosomal protein–rRNA complex has yet to be solved. However, some ribosomal protein structures show strong similarity to well-known families of DNA or RNA binding proteins for which structures in complex with cognate nucleic acids are available. Comparison of the known nucleic acid binding mechanisms of these non-ribosomal proteins with the most highly conserved surfaces of similar ribosomal proteins suggests ways in which the ribosomal proteins may be binding RNA. Three binding motifs, found in four ribosomal proteins so far, are considered here: homeodomain-like α-helical proteins (L11), OB fold proteins (S1 and S17) and RNP consensus proteins (S6). These comparisons suggest that ribosomal proteins combine a small number of fundamental strategies to develop highly specific RNA recognition sites.

INTRODUCTION

It is now common knowledge that ribosomal RNA sequences from all living organisms can be aligned, and their conservation has become a major tool in establishing phylogenetic relationships (1). It is less commonly appreciated that a number of ribosomal proteins are as highly conserved as the rRNAs: the sequences of these proteins are found in all phylogenetic domains and, in some cases, homologs in organisms as diverse as *Escherichia coli* and yeast have been shown by mutational analysis to have similar functions (2). Many of these conserved proteins are known to have direct interactions with ribosomal RNA. The ribosome appears to have a core of both protein and RNA that has not changed since the divergence of phylogenetic domains from the last common ancestor.

The fact that the ribosome has existed as a protein–RNA complex since very early times implies that ribosomal proteins may have been the first in the cell to devise ways of recognizing specific sites in nucleic acids. If this is the case, then the nucleic acid recognition strategies used by highly conserved ribosomal proteins might have been adapted during evolution for use in other proteins with DNA- or RNA-related functions. Conversely, ribosomal proteins unique to a phylogenetic domain or group of organisms may have arisen through modification of nucleic acid binding proteins already existing in the cell. When enough sequences of ribosomal proteins first became available to begin to notice highly conserved regions, no similarities of these proteins to other nucleic acid binding proteins could be detected (3). Either ribosomal proteins were too specialized to be adapted for recognition of other nucleic acids, or similarities with other proteins could not be detected at the sequence level. The latter possibility turns out to be the case: now that a number of ribosomal protein structures have been determined, it is clear that several of the most common RNA and DNA recognition motifs are represented in the ribosome. The adaptation of a nucleic acid recognition motif to different purposes (and structurally different nucleic acids) in different contexts is interesting from both functional and evolutionary perspectives.

This review summarizes structural and functional information for three different nucleic acid binding motifs that have been found among ribosomal proteins so far. In all three cases, structural work on a non-ribosomal instance of the motif has revealed the mechanism of RNA or DNA binding to the protein. Although more than a dozen ribosomal protein structures are now known, none of these structures yet includes the RNA. Thus an aim of this review is to ask whether any insight into RNA binding properties of ribosomal proteins can be gained by considering their similarities to non-ribosomal proteins whose properties are better understood. A second theme of this review is how a few fundamental strategies have been adapted by proteins for recognition of a wide variety of nucleic acids for different functional purposes.

For each of the ribosomal proteins considered here, we will first examine the pattern of conserved surface residues, assuming that a subset of these are involved in RNA recognition (others may be essential for protein–protein contacts). We will then compare these conserved residues to the known nucleic acid binding surface of non-ribosomal protein(s) with the same overall fold. In one case, that of ribosomal protein L11, NMR studies have confirmed that this comparative procedure predicts most of the RNA binding surface. In the other three ribosomal proteins considered here, there are striking correspondences between phylogenetically conserved surface residues and nucleic acid-binding residues of similarly folded proteins.

*To whom correspondence should be addressed. Tel: +1 410 516 7448; Fax: +1 410 516 8420; Email: draper@jhunix.hcf.jhu.edu

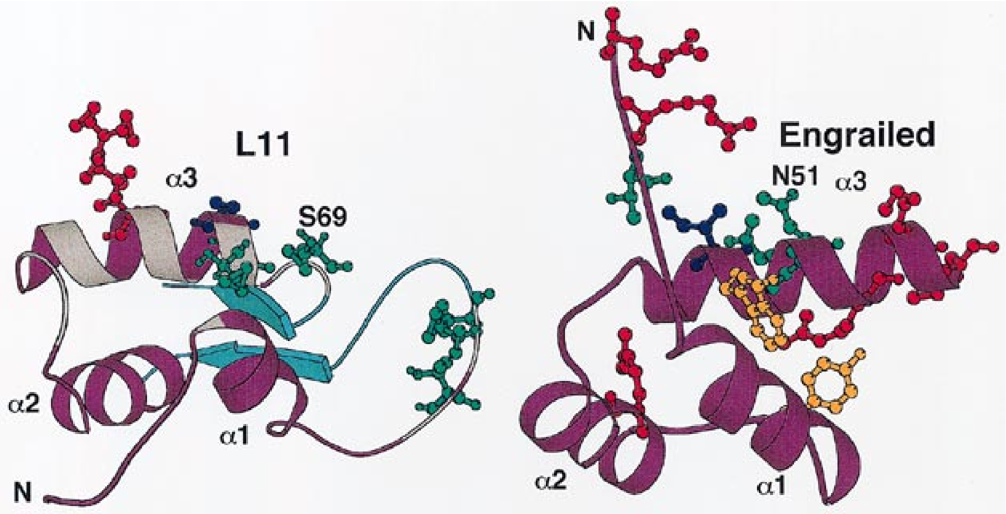

Figure 1. Homeodomain-like α-helical nucleic acid binding proteins. L11: the structure of the C-terminal RNA binding domain, determined in complex with RNA, is shown (L11-C76, 1foy). Positions of residues showing close proximity to RNA are in grey (14). Side chains of conserved surface residues (see Fig. 2) are shown. Engrailed: structure of the engrailed homeodomain protein, from a complex of the protein with cognate DNA, is shown (1hdd). Side chains are those making contact with DNA. Engrailed N-terminal residues wrap around the DNA and make contact with the minor groove, while the N-terminus of L11-C76 remains disordered when bound to RNA. In α-helix 3, ser 69 of L11-C76 and asn 51 of engrailed occupy similar positions when the two proteins are aligned (5). Side chains in this and subsequent figures are colored by type: red for basic, orange for aromatic, green for polar or acidic and blue for hydrophobic.

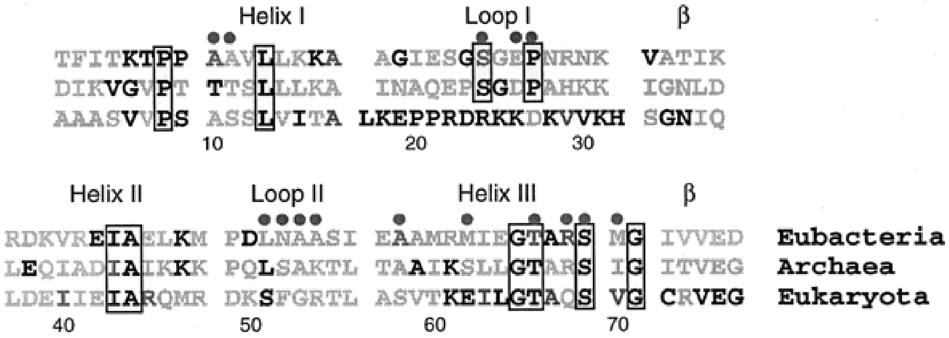
Helix I
Loop I
β
TFITKTPP AAVLLKKA AGIESGSGEPNRNK VATIK
DIKVGVPT TTSLLLKA INAQEPSGDPAHKK IGNLD
AAASVVPS ASSLVITA LKEPPRDRKKDKVVKH SGNIQ
10
20
30
Helix II
Loop II
Helix III
β
RDKVREIAELKM PDLNAASI EAAMRMIEGTARS MG IVVED Eubacteria
LEQIADIAIKKK PQLSAKTL TAAIKSLLGTARS IG ITVEG Archaea
LDEIIEIARQMR DKSFGRTL ASVTKEILGTAQS VG CRVEG Eukaryota
40
50
60
70

Figure 2. Seqeunce conservation within the L11 RNA-binding domain (L11-C76). An example sequence from each of the three phylogenetic domains is shown: eubacteria, *B.stearothermophilus*; archaea, *Sulfolobus acidocaldarius*; eukarya, *Saccharomyces cerevisiae*. Conservation within a domain is indicated by black face (90% identity) and dark grey (80% identity); residues conserved in all three domains are boxed, with the exception of loop 1, which has been substituted by an entirely different sequence in the eukarya. Dots above residues indicate detection of an NOE from the residue to RNA in NMR experiments (14). Residue numbering is that of the L11-C76 fragment (4).

L11 AND HOMEODOMAIN PROTEINS

Seventy-five residues at the C-terminus of ribosomal protein L11 are protected from trypsin digestion by the ribosomal RNA target of the protein (4). The structure of this RNA binding domain, termed L11-C76, was solved by NMR spectroscopy (5), and is shown in Figure 1. The protein has a core of three α-helices, a large disordered loop between helices 1 and 2, and a two-stranded β-sheet linking the termini of helices 2 and 3. Homologs of L11 have been found in all phylogenetic domains, and the protein is interchangeable between eubacteria, archaea and eukaryotes (6–9). It is therefore likely that L11 residues contacting rRNA have been conserved during evolution. In Figure 2, L11 residues whose identities are conserved within a phylogenetic domain are highlighted, and eight residues that are conserved between all domains are boxed. Three of these eight conserved residues are hydrophobic and buried within the core of the protein (leu 13, ile 43, ala 44). A fourth, gly 71, is at a sharp bend between helix 3 and strand 2 of the β-sheet, and may have been conserved to preserve this bend. This leaves three residues near the C-terminus of helix 3 as likely candidates for RNA recognition (gly 65, thr 66, ser 69). Loop 1 is unusual in that it has been conserved between prokaryotic domains but is a completely different sequence in eukaryotes. Loop 1 residues ser 24 and pro 27 are highly conserved in prokaryotes, and are thus additional candidates for the RNA recognition surface. These conserved side chains are shown in Figure 1.

The α-helix core of the L11 RNA binding domain is strikingly similar to the homeodomain class of DNA binding proteins, with an rmsd for helix backbone atoms of 1.2 Å (Fig. 1). A number of homeodomain–DNA crystal structures have defined the recognition mechanism in detail (10–13). Helix 3 sits in the DNA major groove and hydrogen bonds with bases, while the N-terminal amino acids wrap around the DNA into the minor groove. Additional contacts with the DNA backbone may come from the C-terminus of helix 1. The correspondence between the highly conserved residues in helix 3 of L11-C76 and the homeodomain helix 3 residues recognizing DNA bases is a strong argument that L11 uses helix 3 for RNA recognition.

Mutagenesis and NMR experiments confirm that the RNA contact surface of L11 includes helix 3. Mutation of any of the three highly conserved residues in helix 3 weakens RNA binding by at least 10-fold, as do mutations of the conserved loop 1 residues gly 23 and pro 27 (4; D.GuhaThakurta and D.E.Draper, unpublished observations). At a higher resolution level, NMR experiments have determined the structure of the RNA-bound protein and detected 40 NOEs between the RNA and protein (14). These NOEs, which indicate a protein–RNA distance of less than a few Ångstroms, originate from 15 L11 residues, as mapped onto the L11-C76 surface in Figure 1. The RNA contact surface is clearly centered on helix 3, and at least two of the NOEs within helix 3 suggest contact with base, rather than sugar, protons. The core structure of the protein remains unchanged in the presence of RNA. RNA binding dramatically changes Loop 1 from a highly disordered state in the free protein to a specific conformation as rigid as the rest of the protein, as indicated by ^{15}N relaxation experiments (15). Bound RNA may also induce a small movement of loop 2.

The strategy used by L11 to recognize RNA is clear: helix 3 associates with an RNA surface, perhaps a distorted helix major groove, while loops 1 and 2 'clamp' onto either side. [The narrow major groove of A-form RNA is unable to accommodate an α-helix in the same way as the major groove of B-form DNA. However, non-canonical pairs and bulged bases may dramatically widen the major groove of an RNA helix, as seen in the RRE hairpin complex with the α-helical Rev protein fragment (16,17).] The C-terminus of helix 1 may also contact RNA. It is interesting to note that among DNA binding proteins, a number have a homeodomain 'core' and have added β-sheets or loops ('wings') which increase the DNA contact surface (18). The L11 strategy of using an α-helix flanked by loops is thus imitated in spirit, if not in exact detail, among other nucleic acid recognition proteins.

An important point for this review is that a comparison between L11 conserved surface residues and the DNA binding surface of homeodomain proteins provided a reasonable first approximation as to the RNA binding mechanism of L11; only the involvement of loop 2, which is very poorly conserved, was not anticipated. It is remarkable that a simple examination of phylogenetic and structural databases was such a reliable guide to the functional surface of L11, especially in view of the major difference in the overall structures of the recognized nucleic acids.

THE OB FOLD FAMILY OF NUCLEIC ACID BINDING PROTEINS

A five-stranded β-barrel was first noted as a common structure among four proteins binding single-stranded nucleic acids (staphylococcal nuclease and aspartyl-tRNA synthetase) or oligosaccharides (B subunits of enterotoxin and verotoxin-1), and has been termed the oligonucleotide/oligosaccharide binding motif, or OB fold (19). Ten families of proteins, many having nucleic acids substrates, have since been identified as having the OB fold (see Structural Classification of Proteins, http://pdb.pdb.bnl.gov/scop/). Two ribosomal proteins, S17 and S1, are members of this class, and have different variations of the OB fold theme. Comparisons with other OB fold nucleic acid binding proteins suggest somewhat different mechanisms of RNA recognition in each case.

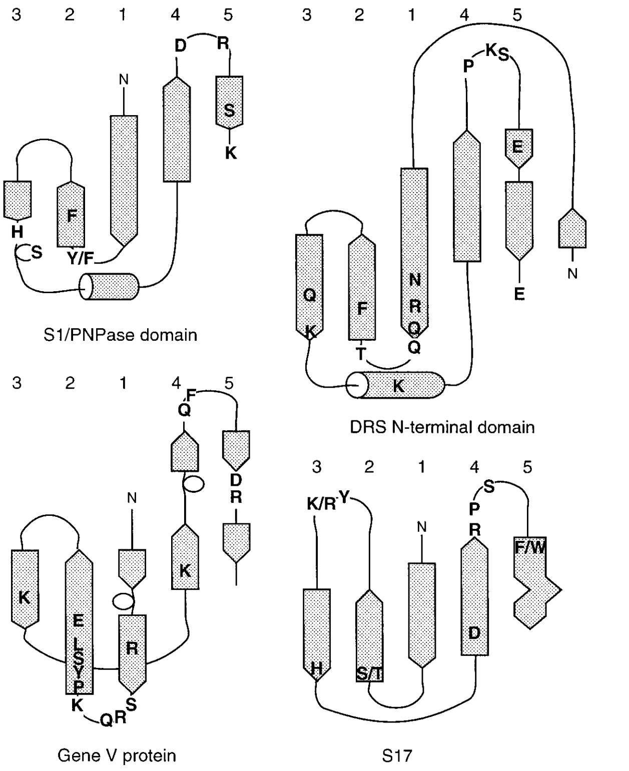

Figure 3. Secondary structures of four different proteins containing an OB fold. Residues that are >95% conserved among 28 different domains are shown on the S1/PNPase domain (see text for discussion). In ribosomal protein S17, residues that are conserved among 14 eubacterial, 11 eukaryotic and three archaeal sequences, with no more than one exception in eubacteria or eukarya and no archaeal exceptions, are shown. For asp-tRNA synthetase and gene V protein, residues in contact with nucleic acid are shown (see text for discussion). β-sheet strand numbering is shown above each structure.

S17 and aminoacyl-tRNA synthetase anticodon recognition domains

The structure of *Bacillus stearothermophilus* S17 has been solved by NMR methods (20,21), and has the β-barrel secondary structure topology and tertiary fold characteristic of the OB fold family of proteins. The barrel is open on one side, with strands 3 and 5 separated by about the width of a β-strand. The protein is well conserved in all three phylogenetic domains, and Figures 3 and 4 have highlighted the nine most highly conserved surface residues. Six of these have no more than one exception among a set of 28 aligned sequences, and three others are conserved as serine/threonine, basic or aromatic. These conserved residues potentially define the RNA binding surface of the protein. Before discussing this possibility further, it is instructive to consider the mechanism by which another family of OB fold proteins,

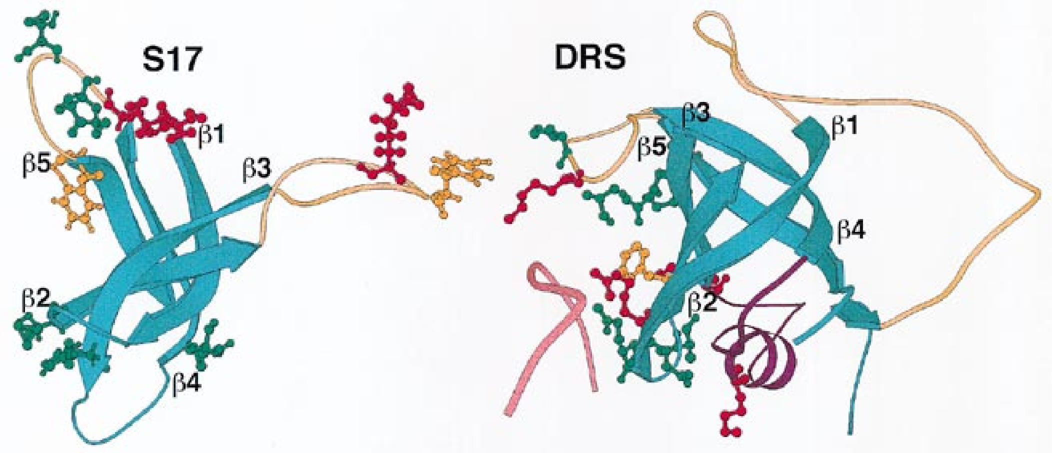

Figure 4. OB fold proteins. S17 (1rip): C-termini of β-strands are numbered as in Figure 3, and side chains of conserved residues noted in Figure 3 are shown. This structure is the first in a set of six NMR structures deposited in the PDB, and has a threading of β3 through the β2–β4 loop that is not found in the other structures. DRS: shown is the Asp-tRNA synthetase (1asz) domain binding the tRNA anticodon loop (glu68–thr200) with residues contacting RNA highlighted (Fig. 3). The backbone of the RNA anticodon loop is also shown as a pink ribbon.

anticodon-binding domains of certain aminoacyl-tRNA synthetases, recognizes RNA.

The N-terminal domains of aspartyl-, aspariginyl and lysyl-tRNA synthetases are typical OB folds (22), and recognize the tRNA anticodon. Details of this interaction are seen in a crystal structure of aspartyl-tRNA synthetase (DRS) (23). The secondary structure of the DRS OB motif is shown in Figure 3, and the residue side chains contacting RNA are explicitly shown in Figures 3 and 4. Anticodon bases are pulled into a cleft formed on one side of the barrel. Residues from strands 1, 2, 3 and 5 form the base of the binding pocket, while residues in two loops, between strands 4 and 5 and strands 1 and 2, line the edges. Particularly noteworthy is phe 127, conserved in all synthetises of this class, which stacks against the U35 common to tRNAs recognized by this group of enzymes.

In S17, the conserved residues thr 21, his 48 and phe 74 are in close proximity within the same cleft region used by DRS to bind RNA, and additional conserved residues in the loop between strands 4 and 5 could form part of the same RNA-binding region (Fig. 4). It is interesting that the pocket contains a conserved phenylalanine, as in the DRS domain (and also the S1/PNPase domain; see below), but it occupies a different position within the cleft. The correspondence between this conserved region and the known nucleic acid binding sites of the DRS N-terminal domain (and other OB fold proteins discussed below) suggests a similar nucleic acid recognition strategy for all the proteins of this class.

A striking difference between S17 and typical OB fold proteins is the large extension of the loop between strands 1 and 2, which is 14 residues (thr 28–tyr 41) compared to two residues in DRS (Fig. 4). Conserved phenylalanine and basic residues suggest that the middle of the loop sequence could be contacting RNA. A small number of NOE restraints at residues 30–37 (21) means that the structure of this loop is not well defined by the data. Though these conserved loop residues appear at some distance from the OB fold cleft, it is possible that the loop would move considerably from the position shown in Figure 4 when RNA binds. This situation is similar to the L11 RNA binding domain, in that a large and poorly structured loop has been added to the structural framework of a nucleic acid binding motif, and it is plausible that the S17 loop also ‘clamps’ the RNA bound in the OB fold site. Jaishree *et al*. (21) have suggested a different scenario, in which the loop binds to an RNA segment distinct from the RNA in the cleft of the molecule, and thereby ‘crosslinks’ two parts of the rRNA during ribosome assembly. Since RNA fragments binding S17 have not yet been closely defined, it is not yet possible to judge between these two different possibilities for the function of this loop.

S1/PNPase family of sequence homologs

S1 is an unusual ribosomal protein: at 557 residues, it is more than twice the size of the next largest ribosomal protein, and its sequence has four obvious repeats of ~70 amino acids (24). More sophisticated analysis has suggested the existence of two additional, highly diverged repeats of the same sequence (25). S1 thus consists of six repeated motifs, each separated by 10–15 residues. A number of other proteins share this same sequence motif, including translation factors (initiation factor 1 and eukaryotic initiation factor 2α), proteins involved in mRNA metabolism (polyribonucleotide phosphorylase, RNase E, RNase II and the yeast RNA helicase PRP22) and several potential transcription factors (25,26). S1 and fragments of S1 bind single-stranded RNA and DNA with relatively little intrinsic sequence specificity (27,28), and it has been thought likely that the 70 residue S1 repeat corresponds to an RNA binding domain.

The structure of the 76 residue S1-like domain from polynucleotide phosphorylase (PNPase) has been solved by NMR methods (26). The domain folds into a five-stranded β-barrel similar to OB fold proteins, as shown in Figures 3 and 5. We have aligned the three most similar domains from eight available eubacterial S1 homologs and four eubacterial PNPase sequences, and found seven surface positions that are highly conserved (>95%). An eighth position is conserved as serine in S1 domains but is arginine in three of the PNPase domains. These eight residues are highlighted in Figures 3 and 5. With the exception of the C-terminal lysine, the conserved residues are clustered on either side of a deep groove or cleft on one face of the protein; the groove is defined by β-strands 2 and 3 on one side and the loop between

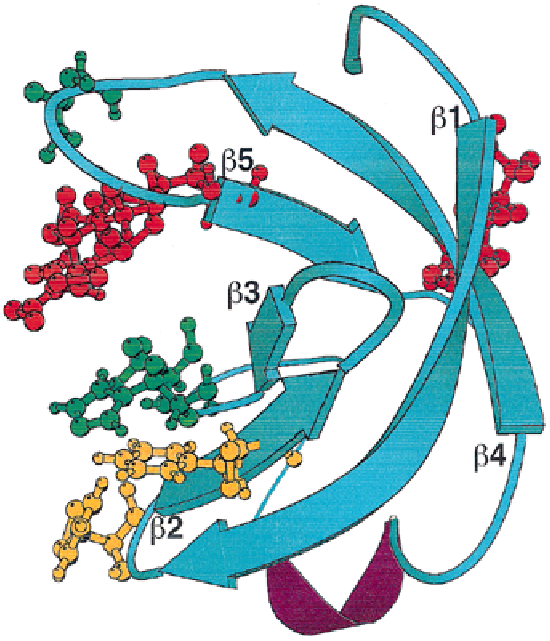

Figure 5. Ribbon diagram of the S1 domain in PNPase (1sro). C-termini of β-strands are numbered and conserved residue side chains shown, as in Figure 3.

strands 4 and 5 on the other. The two conserved aromatic resides (19 and 22) are particularly suggestive of single strand nucleic acid binding, as they are known to stack against bases in DRS (discussed above) and other OB fold proteins binding single-stranded nucleic acids.

The OB fold protein most closely resembling the S1 fold is the bacterial cold shock protein, a transcription factor that may also bind mRNA (29,30). Both the S1 fold and cold shock protein include a turn of 3_{10} helix that follows strand 3 and is not seen in other OB fold proteins (26). The cold shock protein will be discussed further in reference to RNP consensus proteins below. For defining the RNA binding surface of S1, the more relevant OB fold proteins here are two for which the nucleic acid binding surface has been defined, DRS (discussed above), and gene V protein from filamentous phage. A crystal structure of gene V protein is available, and a detailed model of the protein interaction with DNA has been proposed, based on extensive experimental evidence (31). A secondary structure diagram of this protein is also shown in Figure 3, with residues proposed to be in contact with DNA shown; it utilizes the same protein surface as DRS for nucleic acid interactions. (Note that the protein is a tight dimer, with an interface formed between strand 4 of two monomers rotated by 180° relative to each other. Each of the two putative DNA binding sites in the dimer is composed of a β-barrel from one subunit and the loop between strands 4 and 5 from the other subunit.) An aspect of gene V protein potentially relevant to S1 is its high cooperativity of DNA binding. This is probably mediated by the loop between strands 3 and 4 binding strands 4 and 5, as seen in the packing of dimers in one crystal structure (31), and bringing the N-terminal region of one protein near the C-terminal region of another. The dimers are thus able to line up in a way that creates a continuous DNA binding surface. The OB fold repeats in intact S1 protein may similarly be able to form a lengthened nucleic acid binding site. [Since it has been possible to distinguish two distinct nucleic acid binding regions in the protein (27), the six OB folds may be arranged in two or more nucleic acid binding regions.] S1 protein also exhibits significant cooperativity in binding to polypyrimidine RNAs, which may also be mediated by protein–protein contacts similar to those in the gene V protein.

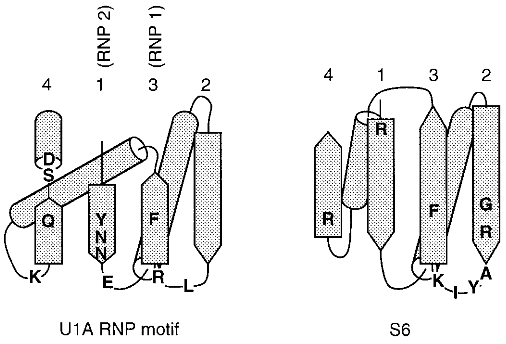

Figure 6. Secondary structures of proteins containing the RNP consensus sequence. β-strands are numbered above each structure, and the two strands approximately corresponding to the RNP1 and RNP2 consensus sequences are labeled. For human U1A protein, residues contacting RNA are shown (37). Residues shown in S6 are those conserved in at least 9 of 10 available eubacterial sequences.

S6 AND RNP MOTIF PROTEINS

Perhaps the best described and most widespread RNA binding motif is the so-called RNP motif, first recognized as two conserved sequences, eight and six amino acids long, separated by ~30 residues and referred to as RNP1 and RNP2 (32). From a crystal structure of the prototypical RNP domain of the human U1A protein (from U1 snRNP) and NMR studies of U1A and other proteins (33–36) it is now known that the RNP1 and RNP2 sequences correspond approximately to the middle two strands of a four-stranded β-sheet (Figs 6 and 7). Two α-helices run behind the sheet to link strands 1 and 2 and strands 3 and 4. Each of the two RNP sequences contains a conserved tyrosine or phenylalanine residue which is exposed on the solvent surface of the β-sheet, suggestive of stacking with single-stranded bases. A high resolution crystal structure of U1A complexed with its cognate RNA hairpin (from U1 snRNA) showed that the two aromatic residues are indeed stacked against two bases of the hairpin loop (37). The seven underlined bases of the hairpin loop sequence, AUUGCACUCC, make close contacts with the protein; most of these contacts are within the β-sheet. Many of the hydrogen bonds are to backbone amide groups; those side chains making specific contacts (including the two aromatic residues stacking with bases) are shown in Figure 7. The β2–β3 loop inserts into the RNA hairpin loop, making contacts with the closing base pair of the stem and helping to orient the loop nucleotides on the β-sheet surface. These contacts are an essential feature of the complex, as mutation of arg 52 (within the β2–β3 loop) severely weakens binding and the seven nucleotide single-stranded sequence cannot itself bind the protein (33). This protein loop is disordered in one of the two chains in the asymmetric unit of the free protein crystals (33), and amides from this loop were disordered even in an NMR study of the protein–RNA hairpin complex (38). Thus the β2–β3 loop is another instance of a poorly structured loop participating in nucleic acid recognition.

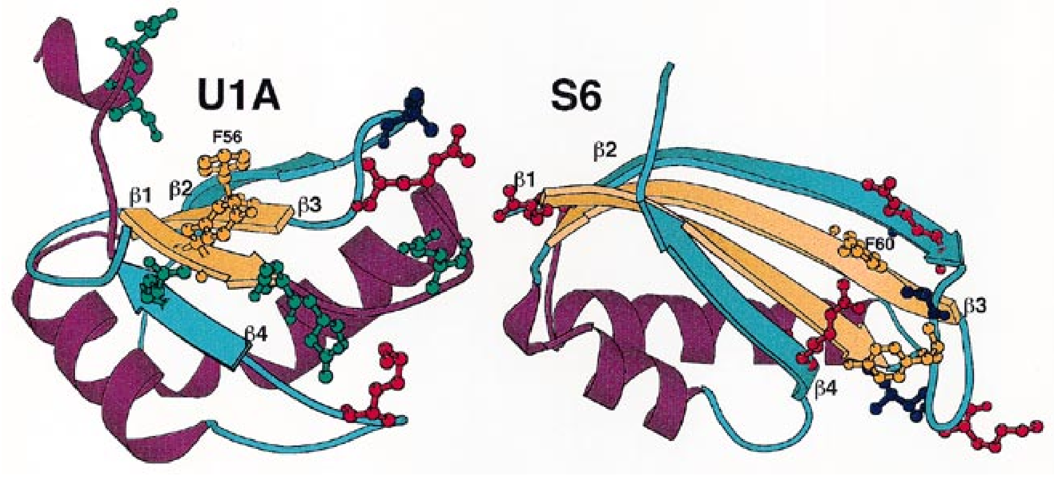

Figure 7. Ribbon structures of RNP motif proteins. C-termini of β-sheets are labeled as in Figure 6, and the middle two strands of the sheet (β1 and β3) are colored gold. Residue side chains are those contacting RNA (U1A, 1urn) or those conserved among S6 (1ris) eubacterial sequences, as in Figure 6.

U1A protein also binds the same seven nucleotide AUUGCAC sequence in the context of an internal loop found in the 3′ UTR of its own pre-mRNA. An NMR-derived structure of this complex has appeared (38,39). Contacts between the β-sheet surface and the seven recognized nucleotides are largely the same as in the U1A–hairpin crystal structure. As in the hairpin complex, the β2–β3 loop inserts into the RNA loop and helps orient the recognized sequence on the β-sheet surface; contacts between the β2–β3 loop and the 3′ UTR structure are more extensive than in the complex with an snRNA hairpin.

U1A protein recognizes single-stranded RNA, in the sense that there is extensive hydrogen bonding to bases that are largely unstacked. However, contacts between the β2–β3 loop and the stem of the hairpin or internal loop, many of which are with backbone atoms, are also crucial for binding and require a specific RNA structure. Thus the RNP motif binding strategy is a combination of unstructured, base-specific contacts (largely with β-sheet) and shape-dependent contacts with secondary structure (largely by a flexible loop). The RNA binding mechanism common to all the RNP motif proteins is likely to be stacking of the conserved aromatic residues in strands 1 and 3 with single-stranded bases.

Ribosomal protein S6 contains essentially the same fold as the U1A protein (Figs 6 and 7B). A striking difference from U1A protein is the extension of the β2 and β3 strands and the loop between them; the twist of the extension places the β2–β3 loop over the β-sheet surface.

S6 has not been extensively conserved during evolution. Sequences homologous to S6 have not been detected among eukarya or archaea, and only 10 examples from eubacteria are currently known. Even among this limited set of sequences, the degree of homology is not high. The most highly conserved positions are indicated in Figure 6, which makes clear a striking correspondence between residues making RNA contacts in U1A protein and conserved S6 residues: S6 has a conserved aromatic residue (phenylalanine in *Thermus thermophilus* S6, but tyrosine in all other eubacterial sequences) on the solvent surface of β3, as characteristic of RNP motif proteins, and the β2–β3 loop sequences is strikingly conserved, including a basic residue at the C-terminus and hydrophobic residues in the middle, as in U1A. This suggests a similarity of RNA binding mechanism, with the β2–β3 loop orienting the backbone of an RNA loop for stacking and hydrogen bonding interactions with the β-sheet surface. [By analogy with U1A protein–RNA interactions, it might be expected that there would be additional base-specific hydrogen bonds from residues on the β-sheet. Since the rRNA sequences that are the most likely contact sites for S6 are poorly conserved among eubacteria (40), S6 residues making such contacts would probably not be phylogenetically conserved.]

The NMR-derived solution structure of U1A protein (41) differs from the crystal structure of a U1A–RNA complex (37) in that aromatic residues of the U1A β-sheet are contacting the C-terminal α-helix in the free protein; the helix must rotate out of the way for RNA to bind. Although thermodynamic studies have failed to detect any strong interaction between the C-terminal tail and the β-sheet surface (42), the comparison of structures still cautions that a protein may undergo some structural rearrangements in binding RNA. S6 is a candidate for such a change. Phe 60 of S6 is located in a loose pocket formed by val 1 (β1), leu 43 (β2) and arg 46 and leu 48 (β2–β3 loop). Some conformational change, particularly in the β2–β3 loop, would be required before phe 60 could stack against a base, as seen in other RNP motif proteins.

Some time ago it was realized that the RNP1 sequence of RNP motif proteins is also found in a class of DNA binding proteins exemplified by the *E.coli* cold shock protein (43). The structure of the cold shock domain was subsequently determined by both NMR and X-ray crystallography and found to be a β-barrel similar to OB fold proteins (29,30). Thus there is an underlying similarity between RNP motif and OB fold proteins, in that each class of proteins uses aromatic residues on the surface of a twisted β-sheet surface to stack with single-stranded bases.

THEMES IN RNA–PROTEIN RECOGNITION

Among the ribosomal proteins discussed here, three basic strategies for recognizing specific RNA sites can be discerned. First, as exemplified by L11, an α-helix provides a surface suitable for hydrogen bonding to RNA bases, presumably in the distorted groove of a helix as seen in the Rev–RRE complex (16). Second, aromatic residues on the twisted surface of a β-sheet are a means to bind unstacked bases; additional polar groups can then hydrogen bond to the bases to provide sequence specificity. Third, extended loops, frequently poorly structured in the absence of nucleic acid, provide additional free energy of binding by interactions with backbone moieties. These backbone interactions may contribute specificity, to the degree that they depend on a particular conformation of the backbone. It is remarkable that all three themes are seen among proteins binding DNA, either in sequence-specific recognition (as homeodomain–DNA complexes) or in non-specific binding to single-stranded DNA (as phage gene V protein). One might have thought that the idiosyncratic structures of ribosomal RNAs, which differ so dramatically from B-form or single-stranded DNA, might have engendered an equally exotic array of protein structures for their recognition, but it appears that rRNA and DNA are similar enough to utilize similar protein surfaces.

Other recognition motifs besides those discussed here are represented among the ribosomal proteins. For instance, S7 and L14 both have an extended β-hairpin with roughly similar distributions of basic and hydrophobic residues (44–46). The hairpin is similar to one found in the HU family of DNA binding proteins, which place it in the DNA minor groove (47), and to the BIV Tat hairpin, which lies in the distorted major groove of an RNA hairpin (48). The N-terminal domain of S8 has an α–β–α–β–β fold resembling portions of DNase I and *Hae*III methyltransferase that bind DNA (49). S5 closely resembles a structure known to bind double-stranded RNA (50); this case is particularly intriguing as a class of DNA binding proteins contains the same protein fold but apparently utilizes a different surface of the protein to contact DNA (51). About half of the known ribosomal protein structures seem to correspond to other proteins that bind nucleic acids, suggesting that there are a small number of very useful strategies for designing a DNA or RNA binding site in a protein. As the RNA binding surfaces of ribosomal proteins become known in detail, it will be interesting to see if any of them adopt a unique method for binding, and whether there are RNA structures that demand an entirely novel protein fold.

A difficult question to answer at this point is whether ribosomal proteins undergo major conformational changes upon binding rRNA; the answer is of considerable importance to attempts to incorporate protein structures into models of ribosome subunits. There are few precedents in which structures of a protein in the presence and absence of RNA are known. In the case of L11, the disorder → order transition in loop 1 of L11 is not surprising in view of similar observations in DNA binding proteins, and it is reassuring that the ordered part of the free protein changes very little upon RNA binding (14). However, the apparent change in the orientation of the C-terminal α-helix of U1A protein (41) suggests that speculations about the RNA binding surfaces of ribosomal proteins should proceed with some caution (see discussion of S6 protein, above). A conserved region of L14, for instance, has α-helices covering an aromatic residue on the surface of a β-barrel (44); it is not evident from inspection whether the free protein structure represents the RNA binding surface, or whether the helices may move to allow the aromatic residue to contact RNA, as in other β-barrel proteins. The existence of RNA-induced protein conformational changes also mean that protein mutations that affect RNA binding may be found in residues not directly contacting the RNA; this introduces additional uncertainty in experimentally defining the RNA binding surface of a protein. Structures of ribosomal protein–RNA complexes are clearly needed to resolve these questions.

FUNCTIONS OF RIBOSOMAL PROTEIN–RNA COMPLEXES

It is commonly assumed that the function of ribosomal proteins is to stabilize specific RNA structures and to promote a compact folding of the large rRNAs. This assumption has been accepted largely by default, as no other specific role (e.g., substrate binding or catalysis) has been unambiguously assigned to a ribosomal protein. (An exception is the mRNA binding activity of S1 protein.) As information about ribosomal proteins and their RNA targets accumulates, we can ask whether stabilization of specific or unusual RNA folds is a plausible function for any of the proteins.

The L11 RNA binding domain is a simple structure strongly resembling homeodomain proteins that recognize less than one turn of DNA helix. It may therefore seem an unlikely candidate for special stabilization of RNA tertiary structures. Yet the RNA that is recognized by L11 is large [a minimum of 58 nucleotides (52)] and compactly folded by a triple base and other tertiary interactions (53,54). The entire 58 nt domain is prevented from unfolding as long as L11 is bound (55). How is such a simple protein domain able to stabilize such an extensive RNA structure? The answer must lie with the RNA: although L11 can only contact a limited region of the RNA, it presumably is a region whose conformation strongly depends on the overall RNA tertiary structure. It has been hypothesized that helix 3 of L11 binds in a helix major groove that has been widened by tertiary contacts (5); this hypothesis is consistent with the L11 ‘footprint’ around an internal loop known to be an important component of tertiary structure (56).

The tertiary structure recognized by L11 is only marginally stable under physiological conditions in *E.coli*, supporting the idea that L11 function is to stabilize RNA structure. However, a single base mutation in the RNA stabilizes the same tertiary structure about as effectively as L11 (53), demonstrating that the RNA alone is capable of achieving a stable tertiary fold in the absence of protein. (The stabilizing RNA mutation also promotes L11 binding.) A better hypothesis for L11 function awaits elucidation of contacts made within the ribosome by the L11 N-terminal domain and experiments with intact ribosomes.

Many other ribosomal proteins, including those discussed in this review, appear to contact even fewer nucleotides than L11. By analogy with the U1A RNP motif protein, S6 probably binds a few single-stranded nucleotides, and it seems unlikely that such an interaction could, by itself, stabilize RNA tertiary structure in any substantial way. But S6 can bind the ribosome only in conjunction with S18 (40), so there is the possibility that strong protein–protein contacts may enable an S6–S18 complex to crosslink two parts of the ribosomal RNA. ‘Crosslinking’ functions are plausible for other ribosomal proteins; L9, for instance, has two well-separated domains, each of which contributes to contacts with a different region of RNA (57).

Investigations of ribosomes by cryo-electron microscopy have advanced to the point that an individual ribosomal protein (L1) could be distinguished and correlated with its crystal structure (58), and recent dramatic improvements in the resolution of 50S subunits by X-ray crystallography (59) hold out the prospect that many other proteins may be located in their ribosomal context in the future. A number of laboratories are also turning their attention to high resolution structural studies of ribosomal proteins in complex with rRNA fragments. In the near future it may be possible to make more detailed proposals about the ways ribosomal proteins contribute to ribosome function.

ACKNOWLEDGMENT

We thank S. Bausch for preparing the Molscript figures.

REFERENCES

1 Woese,C. (1987) *Microbiol. Rev.*, **50**, 221–271.
2 Alksne,L.E., Anthony,R.A., Liebman,S.W. and Warner,J.R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 9538–9541.
3 Draper,D.E. (1996) In Dahlberg,A. and Zimmermann,R. (eds), *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Synthesis*. CRC Press, Caldwell, NJ, pp. 171–197.
4 Xing,Y. and Draper,D.E. (1996) *Biochemistry*, **35**, 1581–1588.
5 Xing,Y., GuhaThakurta,D. and Draper,D.E. (1997) *Nature Struct. Biol*., **4**, 24–27.
6 Beauclerk,A.A.D., Hummel,H., Holmes,D.J., Böck,A. and Cundliffe,E. (1985) *Eur. J. Biochem*., **151**, 245–255.
7 El-Baradi,T.T.A.L., Regt,V.H.C.F.d., Einerhand,S.W.C., Teixido,J., Planta,R.J., Ballesta,J.P.G. and Raué,H.A. (1987) *J. Mol. Biol*., **195**, 909–917.
8 Musters,W., Gonçalves,P.M., Boon,K., Raué,H.A., van Heerikhuizen,H. and Planta,R.J. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 1469–1473.
9 Thompson,J., Musters,W., Cundliffe,E. and Dahlberg,A.E. (1993) *Eur. J. Biochem*., **12**, 1499–1504.
10 Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
11 Klemm,J., Rould,M.A., Aurora,R., Herr,W. and Pabo,C.O. (1994) *Cell*, **77**, 21–32.
12 Li,T., Stark,M.R., Johnson,A.D. and Wohlberger,C. (1995) *Science*, **270**, 262–269.
13 Hirsch,J.A. and Aggarwal,A.K. (1995) *EMBO J*., **14**, 6280–6291.
14 Hinck,A.P., Markus,M.A., Huang,S., Gresiek,S., Kustonovich,I., Draper,D.E. and Torchia,D.A. (1997) *J. Mol. Biol*., **274**, 101–113.
15 Markus,M., Hinck,A., Huang,S., Draper,D.E. and Torchia,D.A. (1997) *Nature Struct. Biol*., **4**, 70–77.
16 Battiste,J.L., Mao,H., Rao,N.S., Tan,R., Muhandiram,D.R., Kay,L.E., Frankel,A.D. and Williamson,J.R. (1996) *Science*, **273**, 1547–1551.
17 Peterson,R.D. and Feigon,J. (1996) *J. Mol. Biol*., **264**, 863–877.
18 Brennan,R.G. (1993) *Cell*, **74**, 773–776.
19 Murzin,A.G. (1993) *EMBO J*., **12**, 861–867.
20 Golden,B., Hoffman,D.W., Ramakrishnan,V. and White,S.W. (1993) *Biochemistry*, **32**, 12812–12820.
21 Jaishree,T.N., Ramakrishnan,V. and White,S.W. (1996) *Biochemistry*, **35**, 2845–2853.
22 Eriani,G., Dirheimer,G. and Gangloff,J. (1990) *Nucleic Acids Res*., **18**, 7109–7117.
23 Cavarelli,J., Rees,B., Ruff,M., Thierry,J.-C. and Moras,D. (1993) *Nature*, **362**, 181–184.
24 Schnier,J., Kimura,M., Foulaki,K., Subramanian,A.R., Isono,K. and Wittmann-Liebold,B. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 1008–1011.
25 Gribskov,M. (1992) *Gene*, **119**, 107–111.
26 Bycroft,M., Hubbard,T.J.P., Proctor,M., Freund,S.M.V. and Murzin,A.G. (1997) *Cell*, **88**, 235–242.
27 Draper,D.E., Pratt,C.W. and von Hippel,P.H. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 4786–4790.
28 Subramanian,A.R., Reinhardt,P., Kimura,M. and Suryanarayana,T. (1981) *Eur. J. Biochem*., **119**, 245–249.
29 Schindelin,H., Marahiel,M.A. and Heinemann,U. (1993) *Nature*, **364**, 164–168.
30 Schnuchel,A., Wiltscheck,R., Czisch,M., Herrier,M., Willimsky,G., Grauman,P., Marahile,M.A. and Holak,T.A. (1993) *Nature*, **364**, 169–171.
31 Guan,Y., Zhang,H. and Wang,A.H.-J. (1995) *Protein Sci*., **4**, 187–197.
32 Dreyfuss,G., Swanson,M.S. and Piñol-Roma,S. (1988) *Trends Biochem. Sci*., **13**, 86–91.
33 Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515–520.
34 Howe,P.W.A., Nagai,K., Neuhaus,D. and Varani,G. (1994) *EMBO J*., **13**, 3873–3881.
35 Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.
36 Wittekind,M., Gölach,M., Friedrichs,M., Dreyfuss,G. and Mueller,L. (1992) *Biochemistry*, **31**, 6254–6265.
37 Oubridge,C., Ito,N., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) *Nature*, **372**, 432–438.
38 Allain,F.H.-T., Howe,P.W.A., Neuhaus,D. and Varani,G. (1997) *EMBO J*., **16**, 5764–5772.
39 Allain,F.H.-T., Gubser,C.C., Howe,P.W.A., Nagai,K., Neuhaus,D. and Varani,G. (1996) *Nature*, **380**, 646–650.
40 Powers,T. and Noller,H.F. (1995) *RNA*, **1**, 194–209.
41 Avis,J.M., Allain,F.H.-T., Howe,P.W.A., Varani,G., Nagai,K. and Neuhaus,D. (1996) *J. Mol. Biol*., **257**, 398–411.
42 Zeng,Q. and Hall,K.B. (1997) *RNA*, **3**, 303–314.
43 Landsman,D. (1992) *Nucleic Acids Res*., **20**, 2861–2864.
44 Davies,C., White,S.W. and Ramakrishnan,V. (1996) *Structure*, **4**, 55–65.
45 Hosaka,H., Nakagawa,A., Tanaka,I., Harada,N., Sano,K., Kimura,M., Yao,M. and Wakatsuki,S. (1997) *Structure*, **5**, 1199–1208.
46 Wimberly,B.T., White,S.W. and Ramakrishnan,V. (1997) *Structure*, **5**, 1187–1198.
47 Rice,P.A., Yang,S.-W., Mizuuchi,K. and Nash,H.A. (1996) *Cell*, **87**, 1295–1306.
48 Puglisi,J.D., Chen,L., Blanchard,S. and Frankel,A.D. (1995) *Science*, **270**, 1200–1203.
49 Davies,C., Ramakrishnan,V. and White,S.W. (1996) *Structure*, **4**, 1093–1104.
50 Bycroft,M., Grunert,S., Murzin,A.G., Proctor,M. and St. Johnston,D. (1995) *EMBO J*., **14**, 3563–3571.
51 Connolly,K.M., Wojciak,J.M. and Clubb,R.T. (1998) *Nature Struct. Biol*., **5**, 546–550.
52 Ryan,P.C. and Draper,D.E. (1989) *Biochemistry*, **28**, 9949–9956.
53 Lu,M. and Draper,D.E. (1994) *J. Mol. Biol*., **244**, 572–585.
54 Conn,G.L., Gutell,R.R. and Draper,D.E. (1998) *Biochemistry*, **37**, 11980–11988.
55 Xing,Y. and Draper,D.E. (1995) *J. Mol. Biol*., **249**, 319–331.
56 Rosendahl,G. and Douthwaite,S. (1993) *J. Mol. Biol*., **234**, 1013–1020.
57 Adamski,F.M., Atkins,J.F. and Gesteland,R.F. (1996) *J. Mol. Biol*., **261**, 357–371.
58 Malhotra,A., Penczek,P., Agrawal,R.K., Gabashvili,I.S., Grassucci,R.A., Jünemann,R., Burkhardt,N., Nierhaus,K.H. and Franck,J. (1998) *J. Mol. Biol*., **280**, 103–116.
59 Ban,N., Freeborn,B., Nissen,P., Penczek,P., Grassucci,R.A., Sweet,R., Frank,J., Moore,P.B. and Steitz,T.A. (1998) *Cell*, **93**, 1105–1115.