The EMBO Journal vol.15 no.6 pp.1350–1359, 1996

Crystal structure of the RNA binding ribosomal protein L1 from *Thermus thermophilus*

S.Nikonov, N.Nevskaya, I.Eliseikina, N.Fomenkova, A.Nikulin, N.Ossina, M.Garber, B.-H.Jonsson[1], C.Briand[2,3], S.Al-Karadaghi[2], A.Svensson[2], A.Ævarsson[2] and A.Liljas[2,4]

Department of Structure and Function of Ribosomes, Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia, [1]Department of Biochemistry, University of Umeå, S-901 87 Umeå and [2]Department of Molecular Biophysics, University of Lund, PO Box 124, S-221 00 Lund, Sweden

[3]Present address: Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, Université Louis Pasteur, BP 163, 670404 Illkirch Cedex, France

[4]Corresponding author

L1 has a dual function as a ribosomal protein binding rRNA and as a translational repressor binding mRNA. The crystal structure of L1 from *Thermus thermophilus* has been determined at 1.85 Å resolution. The protein is composed of two domains with the N- and C-termini in domain I. The eight N-terminal residues are very flexible, as the quality of electron density map shows. Proteolysis experiments have shown that the N-terminal tail is accessible and important for 23S rRNA binding. Most of the conserved amino acids are situated at the interface between the two domains. They probably form the specific RNA binding site of L1. Limited non-covalent contacts between the domains indicate an unstable domain interaction in the present conformation. Domain flexibility and RNA binding by induced fit seems plausible.

Keywords: repressor/ribosomal protein/translation/X-ray crystallography

Introduction

The general outline of the ribosome structure is emerging from electron microscopy studies (Frank *et al.*, 1995; Stark *et al.*, 1995) and progress has also been made in crystallographic studies of ribosomes and ribosomal subunits (Trakhanov *et al.*, 1987; Yonath, 1992; Yonath and Franceschi, 1993), although the resolution remains too low to identify individual components. The elucidation of the spatial structure of ribosomes as well as detailed investigations of individual ribosomal components, including rRNA fragments, ribosomal proteins and their complexes, is a prerequisite to understanding the ribosomal functions. The bacterial ribosome contains >50 ribosomal proteins, several of which are primary binding proteins, i.e. they are able to bind tightly to rRNA independently of other ribosomal proteins. The ribosome is a rich source of information on protein–RNA interactions which, at present, are not well understood at a molecular level.

L1 is one of the largest ribosomal proteins and is located on a side protuberance (often referred to as the L1 protuberance) opposite the L7/L12 stalk of the 50S ribosomal subunit (Stöffler and Stöffler-Meilicke, 1986). L1 from *Escherichia coli* associates independently, specifically and strongly with the 23S rRNA and protects a fragment of ~100 nucleotides from nuclease digestion (Zimmermann, 1980; Branlant *et al.*, 1981). L1 from this bacteria can bind to the corresponding rRNA from phylogenetically distantly related species like the eukaryote *Dictyostelium discoideum* (Gourse *et al.*, 1981). In addition, L1 from the archaeon *Methanococcus vannielii* is able to functionally replace L1 in *E.coli* ribosomes *in vitro* (Baier *et al.*, 1990).

The RNA binding activity of L1 is of profound interest since the protein is also able to regulate gene expression by binding to its own mRNA, thereby acting as a translational repressor (Gourse *et al.*, 1986). L1 from *E.coli* mediates autogenous regulation of translation by binding to a region within the leader sequence, close to the Shine–Dalgarno sequence, of the polycistronic mRNA of the L11 operon coding for ribosomal proteins L1 and L11 (Gourse *et al.*, 1986). Such regulation at the translational level has also been shown recently for L1 from *M.vannielii*. It was shown that *E.coli* L1 can inhibit the *in vitro* translation of *M.vannielii* L1 polycistronic mRNA and, conversely, that L1 from *M.vannielii* can inhibit the synthesis of both L11 and L1 from *E.coli* (Hanner *et al.*, 1994). It is believed

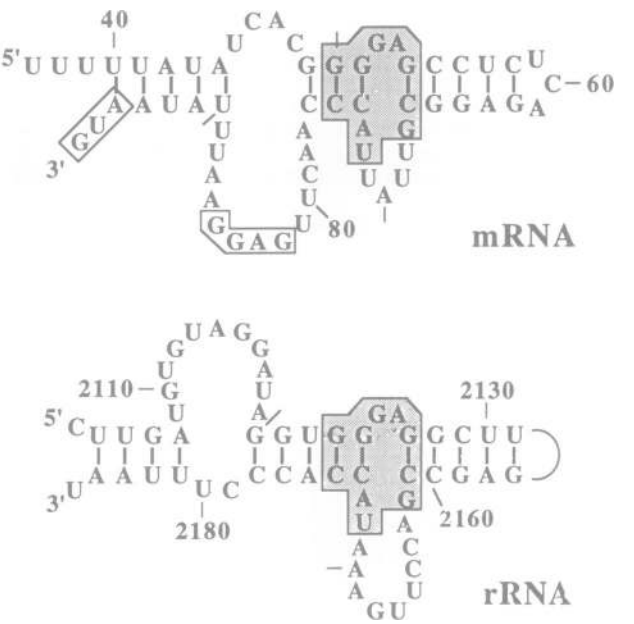

Fig. 1. The consensus structure (shaded boxes) of the L1 binding site on mRNA and rRNA. Also indicated by boxes are the Shine–Dalgarno sequence and the start codon in the mRNA. The sequences are from *E.coli* (adapted from Draper, 1989).

Table I. Data collection and phasing statistics

Compound	Resolution (Å)	Completeness (%)	*R*-merge (%)	Concentration (%)	Time of soak (days)	*R*-diff.	No. of sites	Phasing power		Cullis factor	
								cent.	acent.	cent.	acent.
Native (orig.)	2.2	95.3	6.8	–	–	–	–	–	–	–	–
Native (over.)	1.85	98.0	5.3	–	–	16.5	–	–	–	–	–
Red Pt[a]	2.8	92.9	4.6	10 mg/ml	6	14.0	7	1.2	1.7	0.67	0.69
Na mersalyl[a]	2.8	94.2	6.4	1 mM	4	14.6	7	0.9	1.3	0.74	0.79
Red Pt[b]	3.2	96.1	4.8	10 mg/ml	6	18.8	8	1.3	2.0	0.63	0.58
Na mersalyl[b]	3.0	93.2	6.5	1 mM	4	9.7	6	0.8	1.0	0.73	0.79
Orange Pt[b]	3.0	93.3	5.3	20 mM	0.8	13.7	6	0.8	1.0	0.75	0.76
Na_2WO_4[b]	3.6	94.5	6.9	1 mM	6	9.6	8	0.7	0.9	0.83	0.85

[a]Derivatives prepared by soaking crystals of overexpressed L1.
[b]Derivates prepared by soaking original L1 crystals.
Red Pt is thiolato(2,2′,2″-terpyridine)platinum(II) complex (Jennette *et al.*, 1976); Orange Pt is platinum(II)(2,2′:6′,2″-terpyridine)chloride dihydrate; Na mersalyl is sodium{2-[*N*-(3-hydroxymercuri-2-methoxipropyl)carbamoyl]phenoxyacetate}. Na_2WO_4 is sodium wolframate.
R-merge = $\Sigma|I_{obs}-\langle I\rangle|/\Sigma I_{obs}$; *R*-diff. = $2*\Sigma|F_i-F_j|/\Sigma(F_i+F_j)$, where the summation is over all reflections. Cullis factor is defined as <lack of closure>/<isomorphous difference>; phasing power is F_h/A where F_h is the heavy atom structure factor and *A* is the residual lack of closure. The overall figure of merit was 0.63 at 2.8 Å resolution.

Table II. Refinement statistics

Resolution range	8–1.85 Å
Residues included	5–228
Number of waters	146
R-factor	18.1%
R_{free}	23.2%
r.m.s. deviations	
Bond lengths	0.015 Å
Bond angles	1.9°

that L1 binds rRNA and mRNA in a very similar manner, since both corresponding binding sites on the RNA share a conserved consensus structure (Figure 1; Draper, 1989).

L1 from the bacterial thermophile *Thermus thermophilus* contains 228 amino acids (49% sequence identity with *E.coli* L1) and has an M_r of 24 694 Da (Amons *et al.*, 1993). Here, we describe its three-dimensional structure at 1.85 Å resolution and discuss some aspects of the RNA–protein interactions.

Results

Structure determination

L1 from ribosomes of *T.thermophilus* was purified under non-denaturing conditions and crystallized as described previously (Sedelnikova *et al.*, 1987; Agalarov *et al.*, 1990). Native data and data for a large number of heavy atom derivatives were collected (Table I). Analysis of these data has shown that the conformation of the protein in the crystals is unusually sensitive to different chemicals. This is observed as a lack of isomorphism, and no high quality heavy atom isomorphous derivatives could be obtained. Even native crystals did not give a reproducible diffraction pattern.

The gene encoding L1 from *T.thermophilus* has been cloned and overexpressed in *E.coli*. The overexpressed protein crystallized in the same space group and with the same unit cell dimensions as L1 isolated from ribosomes (original L1). A comparison of the two corresponding native data sets showed that d*F*/*F* increased linearly from 11.0% at low resolution to 25.9% at 2.2 Å resolution. Two heavy atom derivatives, prepared from crystals with overexpressed L1, showed a considerable lack of isomorphism compared with the corresponding native data set, but a tolerable non-isomorphism compared with the data set for crystals of original L1. Therefore, these L1 derivative data were scaled to the original L1 native data set and used to solve the phase problem.

Isomorphous phases were calculated at 2.8 Å resolution and subsequently modified by density modification and phase extension to 2.2 Å resolution. The final map was of sufficient quality to trace the polypeptide chain unambiguously, except the region of the N-terminal tail. The quality of the electron density map decreased markedly in the N-terminal area. The model was built and refined at 2.2 Å resolution. The original L1 phases were then applied to the overexpressed L1 structure amplitudes. The final refined model at 1.85 Å resolution incorporates residues 5–228. The statistics of the refinement are shown in Table II.

The models derived from the two native data sets have a small relative displacement with an overall r.m.s. deviation of 0.26 Å for the Cα atoms. This deviation could explain the observed difference between the intensity data in the two native data sets, but does not seem significant considering the difference in resolution and the error level of the coordinates.

Structural description

The L1 molecule consists of two domains. The approximate dimensions of the molecule are 52×39×35 Å. The overall structure and the notation of the secondary structure elements are presented in Figure 2. The location of secondary structure elements within the sequence is shown in Figure 3. Both the N- and C-termini of the protein are located in domain I. This domain includes residues 1–67 and 160–228, and domain II (residues 68–159) thus appears as an insert in domain I.

The connectivity scheme of domain I is α1–α2–β1–β2–β7–β8–α7–β9–β10 (Figure 2B). This domain contains a well known structural motif, the split β–α–β motif (Orengo and Thornton, 1993) or abc/ad unit (Efimov, 1994). This motif consists of a three-stranded (β1, β8, β9) antiparallel β-sheet and one α-helix (α7). Two α-helices (α1 and α2) are found N-terminal to the motif. The polypeptide chain

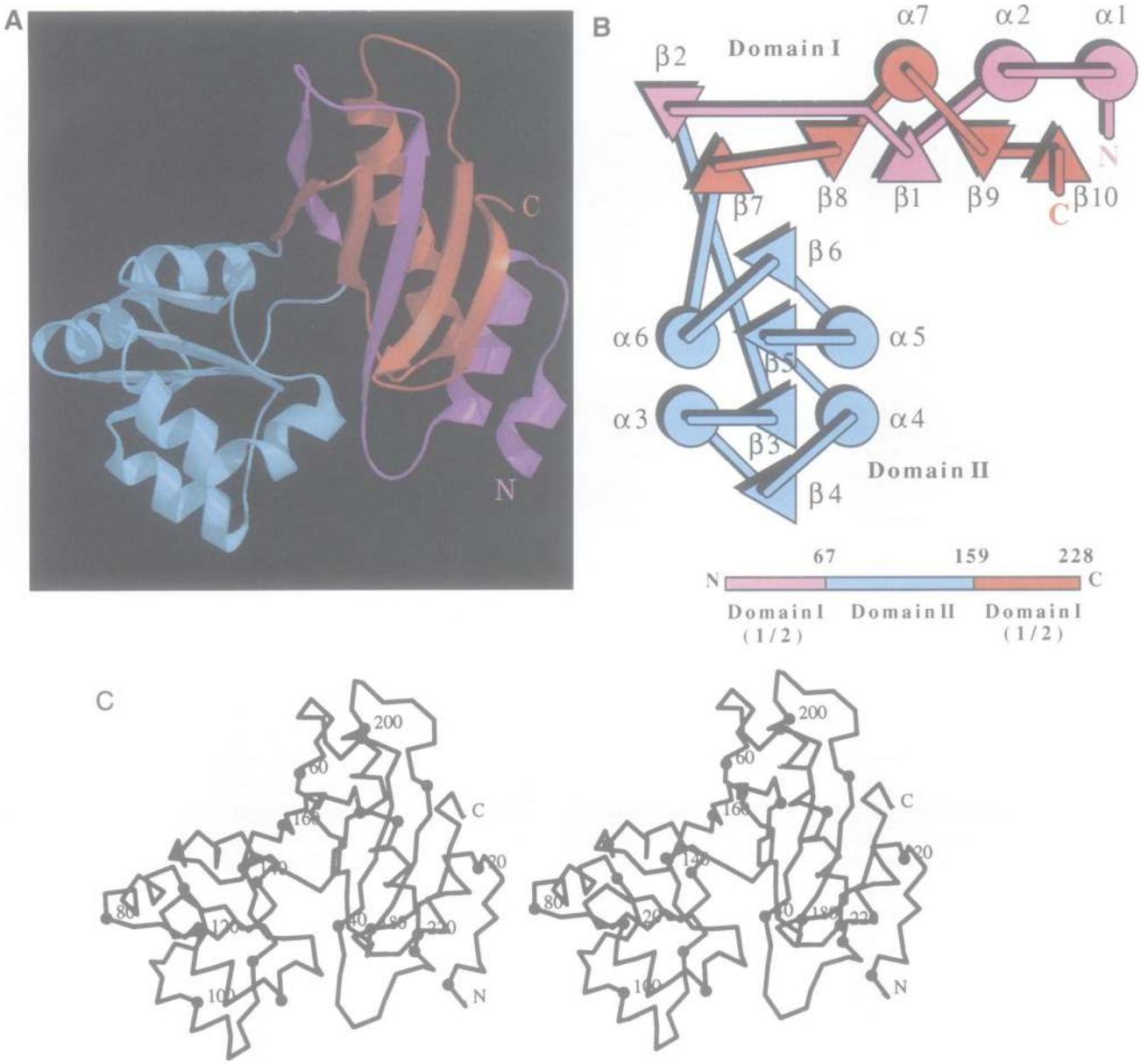

Fig. 2. (**A**) Schematic representation of the structure of ribosomal protein L1 from *T.thermophilus*. Domain I is coloured purple and red representing the N- and C-terminal parts, respectively. Domain II is shown in blue. Figures 2A, 4A and 4B were made with MOLSCRIPT (Kraulis, 1991) and rendered with Raster 3D (Merrit and Murphy, 1994). (**B**) Topological diagram and designation of secondary structure elements in L1. The colour coding is the same as in (A). (**C**) Stereo view of a Cα trace of L1. Every 10th Cα atom is represented by a sphere (starting at number 10). The first N-terminal residues of the model (residues 5–7) have very high temperature factors and were omitted in all structure illustrations.

runs through the first strand of the motif (β1), then strand β2, the whole domain II and strand β7 and subsequently returns back to the second and third strands (β8 and β9) of the motif (Figure 2B). An additional C-terminal strand (β10) extends the β-sheet of the motif.

Helix α1 is quite separated from the globular part of the molecule and is associated with helix α2 and strand β10 by hydrophobic interactions as well as by a salt bridge between Lys13 and Glu31. The loop at positions 216–219 contains highly conserved residues (Figure 3) and protrudes into the interdomain region. Residues 58–63 and 160–165 form a double-stranded (β2, β7) antiparallel β-sheet which is separated from the main sheet of domain I and makes the covalent connections to domain II. This sheet, together with domain II, is equivalent to the previously identified small tryptic fragment (10 kDa) of L1 from *T.thermophilus* (Amons *et al.*, 1993).

Domain II contains two helices on each side of a four-stranded parallel β-sheet with an overall 'Rossmann fold' topology. The connectivity scheme is β3–α3–β4–α4–β5–α5–β6–α6 (Figure 2B). Two protrusions in the interdomain space at positions 131–136 and 138–144 are stabilized by a number of hydrogen bonds. The first of these protrusions includes a part of helix α5 and is the most highly conserved part of domain II (Figure 3). Helix α5 has an irregular conformation due to a kink at residue 127.

In the crystals, two molecules related by the crystallographic 2-fold axis form a dimer stabilized by two salt bridges between Lys36 in one molecule and Asp111 in the second and vice versa.

Fig. 3. Alignment of L1 sequences (see Materials and methods for details on alignment procedure). The upper 10 sequences are from bacteria, the others from Archaea. The numbering and the location of secondary structure elements (same colours as in Figure 2A and B) correspond to the *T.thermophilus* structure (THETH). The coloured bars indicate strictly conserved residues or very conservative substitutions (V/I/L, T/S, D/E, N/Q, Y/F, R/K) with red colour where this applies for all the sequences or pink colour where this applies at least for all the bacterial sequences. The sequences are THETH: *Thermus thermophilus* [accession no. (Acc.) P27150, Amons *et al.*, 1993]; ECOLI: *Escherichia coli* (Acc. P02384, Post *et al.*, 1979); SERMA: *Serratia marcescens* (Acc. P09764, Sor and Nomura, 1987); PROVU: *Proteus vulgaris* (Acc. P10054, Sor and Nomura, 1987); CGDAB: unclassified bacterium-like organism (Acc. P36248), STRSQ: *Streptomyces* sp. (Acc. Q07976, Ruengjitchatchawalya *et al.*, 1993); STRGR: *Streptomyces griseus* (Acc. P36256); BACST: *Bacillus stearothermophilus* (Acc. P04447, Kimura *et al.*, 1985); THEMA: *Thermotoga maritima* (Acc. P29393, Liao and Dennis, 1992); SYNY3: *Synechocystis* sp. (Acc. P36236, Schmidt *et al.*, 1993); HALHA: *Halobacterium halobium*, (Acc. P13575, Itoh, 1988); HALCU: *Halobacterium cutirubrium*, (Acc. P05966, Shimmin and Dennis, 1989); HALMA: *Haloarcula marismortui* (*Halobacterium marismortui*) (Acc. P12738, Arndt and Weigel, 1990); HALVO: *Halobacterium volcanii* (*Haloferax volcanii*) (Acc. P41199); METVA: *Methanococcus vannielii* (Acc. P15824, Baier *et al.*, 1990); SULSO: *Sulfolobus solfataricus* (Acc. P35024, Ramirez *et al.*, 1989).

Discussion

Structural similarities

The structures of a number of ribosomal proteins have been determined previously: L6 (Golden *et al.*, 1993b), the C-terminal domain of L7/L12 (Leijonmarck and Liljas, 1987), L9 (Hoffman *et al.*, 1994), L14 (C.Davies, S.W.White and V.Ramakrishnan, personal communication), L30 (Wilson *et al.*, 1986), S5 (Ramakrishnan and White, 1992), S6 (Lindahl *et al.*, 1994) and S17 (Golden

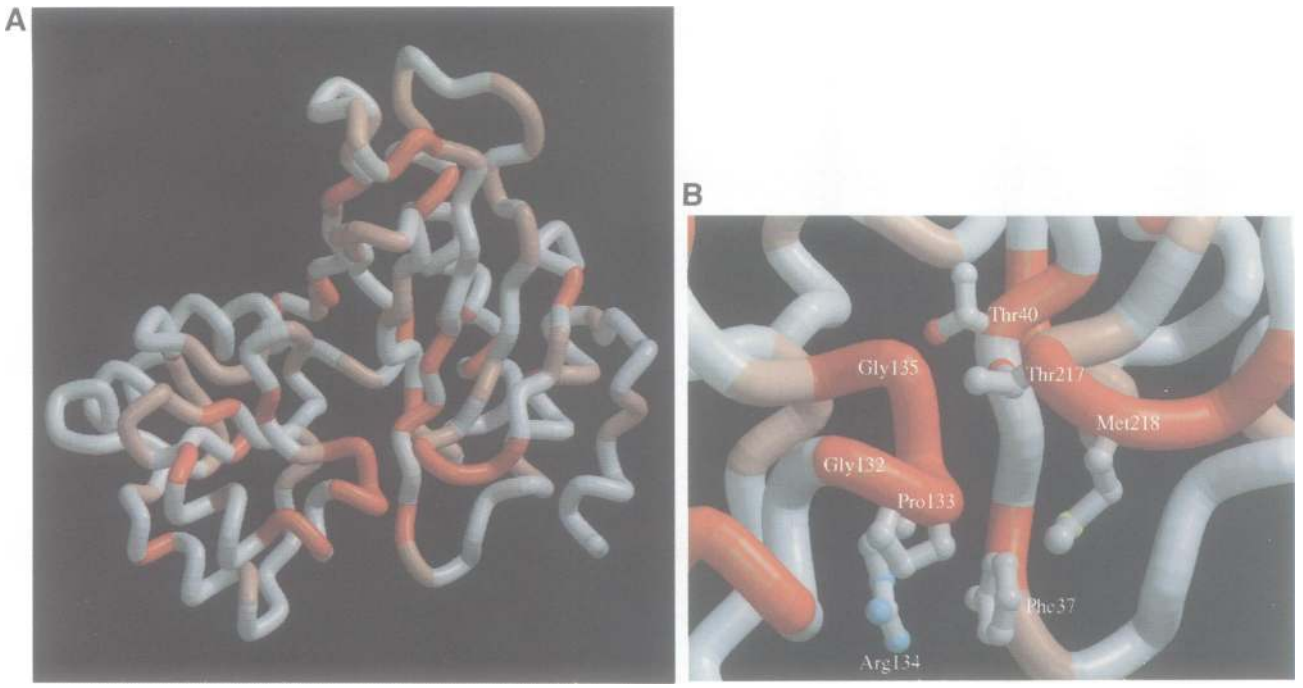

Fig. 4. (**A**) The location of conserved residues in the structure of L1. The colour coding is the same as in Figure 3. (**B**) The interdomain region with a putative RNA binding site. This is an enlargement of the corresponding region in (A) but, in addition, showing well conserved side chains which may be involved in the RNA binding. This figure also illustrates the very limited domain contacts observed in the present structure.

et al., 1993a). Curiously, the majority of these proteins (L6, L7/L12, L9, L30, S6), in addition to L1, contain the split β–α–β motif. Among this group of proteins, the connection between the first and the second strand of the common motif is very variable and in L1 it constitutes a whole separate domain. The split β–α–β motif is also found in another translational repressor, regA from bacteriophage T4 (Kang *et al.*, 1995), and in a large family of RNA binding proteins including the spliceosomal protein U1A (Burd and Dreyfuss, 1994). The crystal structure of the RNA binding domain of U1A in complex with a fragment of RNA has been deter-mined and it was shown that residues on the surface of the β-sheet and the connecting loops of the split β–α–β motif are involved in RNA binding (Oubridge *et al.*, 1994). This motif has also been observed in a large number of other unrelated proteins with entirely different functions. Efimov (1994) treated the split β–α–β motif as a nucleus in protein folding. However, the high frequency of the split β–α–β motif among ribosomal proteins and other RNA binding proteins may suggest that the common structural motif is associated with their common function as RNA interacting proteins. Furthermore, the split β–α–β motif has been found in the 'palm' domain of DNA and RNA polymerases where it is involved in interaction with nucleic acid (Arnold *et al.*, 1995). Parts of domain I in L1 and the palm domain in the Klenow fragment of DNA polymerase I (Ollis *et al.*, 1985) show rather good structural similarity. When these domains are superimposed, domain II of L1 is in a similar location as part of the 'thumb' domain in the DNA polymerase. Interestingly, DNA binding occurs in a cleft lined by the 'palm' and the 'thumb' (Arnold *et al.*, 1995).

The α/β topology of domain II, with a parallel β-sheet and helices on both sides, is the first example of this structural class among ribosomal proteins. This topology has been found in numerous proteins including the nucleotide binding domain of many dehydrogenases and kinases. However, the specific features involved in nucleotide binding of these proteins have not been observed in L1.

RNA binding

L1 has been investigated intensively for its RNA binding properties and its dual function as a ribosomal protein and translational repressor (Gourse *et al.*, 1986). *E.coli* L1 has the ability to bind specifically not only to a wide variety of bacterial and archaeal 23S rRNA (Stanley *et al.*, 1978; Zimmermann *et al.*, 1980; Baier *et al.*, 1989) but also to eukaryal 26S/28S rRNA (Gourse *et al.*, 1981). It is also known that the primary and secondary structure of the L1 binding site on the 23S/28S rRNA is highly conserved in Bacteria, Archaea and Eukarya (Hanner *et al.*, 1994). The L1 binding regions of the mRNAs coding for L1 from different organisms exhibit a significant structural similarity to the binding site in 23S rRNA (Kearney and Nomura, 1987; Baier *et al.*, 1990). Furthermore, it has been shown that the *E.coli* and *M.vannielii* L1 are functionally interchangeable in the ribosome as well as in repression of translation (Baier *et al.*, 1990; Hanner *et al.*, 1994). These data suggest that both the RNA binding site of L1 and the L1 binding site of the respective RNA molecules are structurally highly conserved (Sor and Nomura, 1987; Hanner *et al.*, 1994). Thus, conserved residues in L1 should form the binding site(s) for the consensus structure of the RNA. The location of conserved residues in the structure of L1 (according to the alignment in Figure 3) is shown in Figure 4A. Figure 5A (left) and B (left) shows that the main contiguous region of conserved residues on the surface is found at the interface between the domains. Clusters of conserved residues form two protrusions in close contact with each other at the domain interface

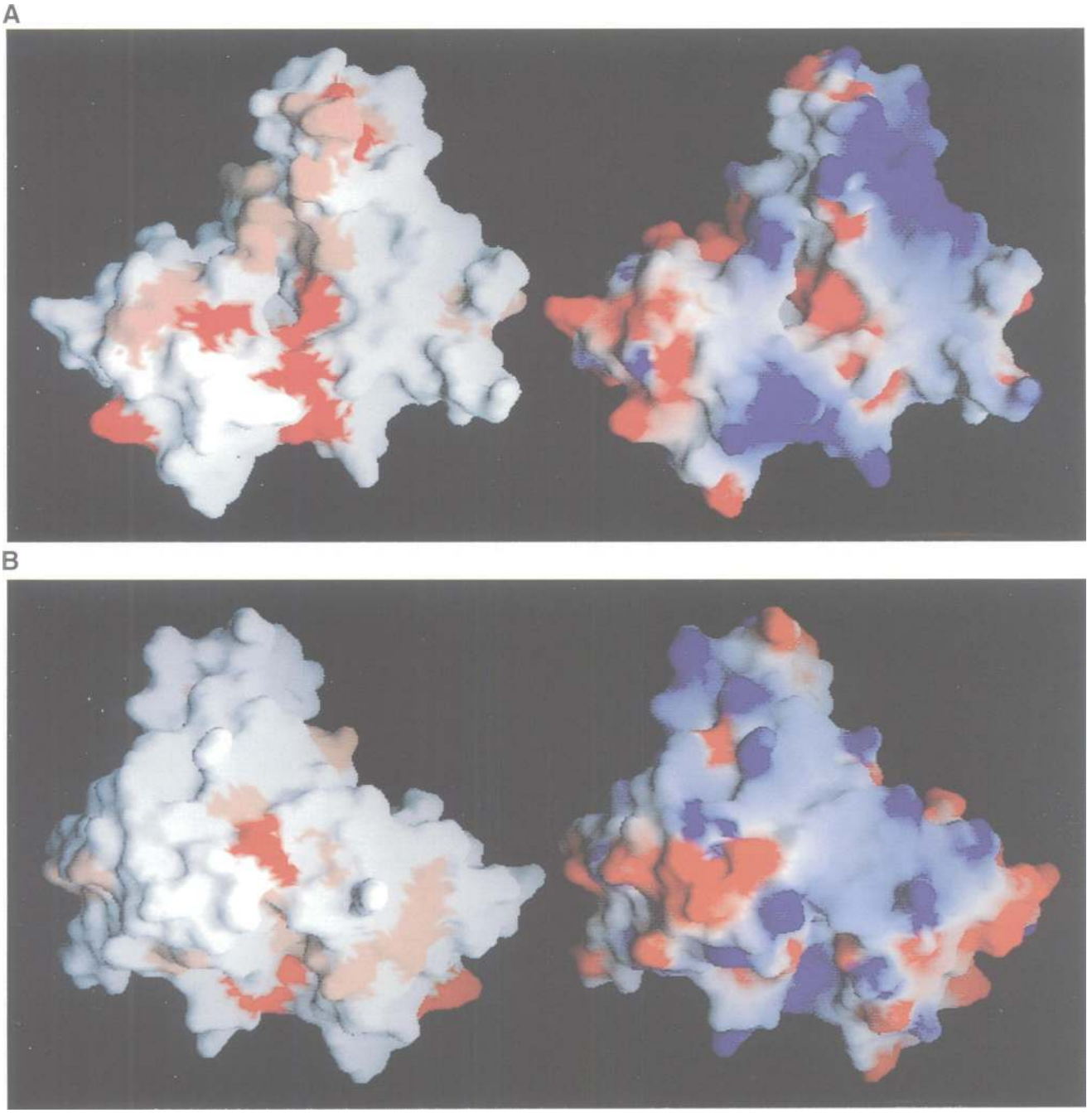

Fig. 5. The location of conserved residues (left side, colours as in Figures 3 and 4) and electrostatic potential (right side, blue for positive and red for negative) on the surface of L1. (**A**) Front view (with orientation the same as in Figures 2 and 4). (**B**) Rear view [rotated 180° relative to (A)]. Note that the most conserved area is the domain interface and that this interface is associated with basic residues viewed from both directions.

(Figure 4B). One of these protrusions is located at the C-terminal end of helix α5 in domain II. It is interesting to note that the RNA binding loop, connecting two β-strands in U1A, contains an α-helical turn just like this L1 protrusion. The other protrusion is formed by the loop connecting strand β9 of the split β–α–β motif and strand β10 in domain I. In close proximity to these protrusions is the strictly conserved Phe37 belonging to the loop between helix α2 and strand β1, the first strand of the split β–α–β motif (Figure 4B).

At present, few structures of protein–RNA complexes are available, and the knowledge of the details of protein–RNA interactions in general is limited. Stacking of bases with side chains, preferably exposed aromatic side chains, and hydrogen bonds between bases and amino acid residues of various types have been found to be important for the specific RNA binding by U1A (Oubridge *et al.*, 1994). For L1, potential side chains for RNA interaction in the conserved interdomain region include Phe37, Thr40, Arg134, Thr217 and Met218 (Figure 4B). The structure of the U1A–RNA complex shows few direct contacts between the protein and the phosphates of the RNA. However, long-range electrostatic interactions between the RNA and regions of positive potential on the protein

surface seem important to position the RNA and stabilize the interaction (Nagai *et al.*, 1995). The electrostatic potential of the surface of L1 (Figure 5, right) shows some patches of positive electrostatic potential located close to the domain interface with certain overlap with the regions of conserved residues. Although the conservation of residues in these regions is limited, similar positive surface areas may still be conserved for a similar long-range interaction with the RNA.

We have found that L1 from *T.thermophilus* lacking eight amino acids from its N-terminus (obtained by spontaneous proteolysis) exhibits a reduced 23S rRNA binding capacity (data not shown). An additional break of the polypeptide chain between residues Arg36 and Phe37 (obtained by limited trypsinolysis) completely abolishes RNA binding. Phe37 is part of the conserved cluster of residues at the interdomain region (Figure 4B). The first eight N-terminal residues probably form a flexible tail, as indicated by their disorder in the crystals. The first four residues have not been included in the current model but they may be located close to the interdomain region. The residues in this tail, half of which are positively charged, could take part in electrostatic interactions with RNA and play a key role in orientating the protein with respect to the RNA. An analogous observation was made for the C-terminal extension of the RNA binding domain of the U1A spliceosomal protein (Oubridge *et al.*, 1994), the RNP domain in hnRNP C protein (Görlach *et al.*, 1992) and the U1 70K protein (Query *et al.*, 1989). According to the sequence alignment shown in Figure 3, the archaeal L1 proteins are missing a corresponding N-terminal tail. The proposed involvement of these residues in RNA binding seems unique to bacteria and the details of the protein–RNA interaction thus differ in Bacteria and Archaea. However, the archaeal L1 proteins have extended loop regions (loop α2/β1 and/or loop α4/β5) close to the interdomain region which may have a similar role.

Domain flexibility

Some of the conserved residues in the interdomain region proposed to participate in the RNA binding are also involved in domain–domain contacts. However, these contacts are not extensive: only one hydrogen bond and limited van der Waals contacts between a few residues. There is a small but a distinct hole between the domains. The calculated area of water-accessible surface lost upon domain interaction is small, ~250 Å^2 or ~4% of the total area of each domain. The interaction between domains of similar size in other proteins involves 18–29% of the total area (Argos, 1988), corresponding to ~1000 Å^2 or more. Proteolysis experiments with L1 (data not shown) indicate that the 10 kDa fragment including domain II dissociates from the rest of the molecule after cleavage. Large changes of up to 8% in cell dimensions of some derivative crystals (not used for phasing) and the observed lack of isomorphism also indicate a certain conformational flexibility of the molecule. All these observations make us suspect that the domain interaction observed in the current structure is very unstable and the domains may become separated to some extent under normal conditions *in vivo*. The putative hinge region at the border of the domains contains three Gly residues, Gly67, Gly69 and Gly159, which are strictly conserved in all bacterial L1 sequences. However, the archaeal L1 proteins probably have a somewhat different conformation in this region since they have a one residue deletion close to the position corresponding to Gly159 (Figure 3).

Some of the conserved residues are buried between the domains and line the hole. They make no contacts with residues in the opposite domain. It seems likely that domain movements leading to further exposure of these conserved residues are needed for a proper binding of the RNA between the domains rather than on the surface of the protein in the present conformation. Thus, the two domains are probably flexible in the free protein and may become stabilized by an induced fit mechanism upon RNA binding.

Structure–function relationships

The binding of RNA by L1, proposed to occur at least in part between its two domains, is fundamental to its function as a ribosomal protein and as a translational repressor. The function of some of the numerous ribosomal proteins in the ribosome may be to assist the proper folding of the rRNA, i.e. to function as RNA chaperones. Others may have different or additional functions. L1 is not an essential protein, since viable mutants lacking L1 have been obtained. However, these mutants grow poorly and the ribosomes show only about half the rate of protein synthesis *in vitro* compared with wild-type. Further studies could not demonstrate any other defects in these ribosomes besides the reduced rate of protein synthesis (Subramanian and Dabbs, 1980). L1 has been cross-linked to tRNA in the P-site of the ribosome, indicating a possible functional interaction with tRNAs (Rosen *et al.*, 1993). These observations indicate that L1 may have another activity, distinct from rRNA/mRNA binding, and thus another functional site separate from the rRNA/mRNA binding site.

An interesting relationship between structure and function has been proposed for proteins belonging to the α/β structural class with an open β-sheet topology (Rossmann fold). Domain II of L1 belongs to this class of proteins which have a topological switch point at the C-terminal edge of the β-sheet, where the connections between the strands and the helices change direction. At least a part of a functional binding site is found at these switch points in practically all proteins of this type (Brändén, 1980; Brändén and Tooze, 1991). The functional residues forming these sites are located in the loops between the β-strands and the α-helices. However, the residues at this location in domain II of L1 are not well conserved. On the other hand, the most conserved region is found at the other end of domain II, opposite to the topological switch point. Domain II of L1 may be an intriguing exception to the proposed universal rule of the structure–function relationship in the α/β structural class of proteins.

Materials and methods

Preparation of T.thermophilus L1

The ribosomal protein L1 was initially purified from *T.thermophilus* ribosomes as described by Sedelnikova *et al.* (1987). Subsequently, the gene encoding *T.thermophilus* L1 has been cloned and sequenced (accession no. X81375). *E.coli* BL21(DE3) transformed with pACA-L1 plasmid was grown at 37°C with shaking until a cell density of 0.5 A_{600} was reached. The production of the protein L1 was then induced by addition of isopropylthio-β-D-galactoside (0.5 mM) and, after that, the

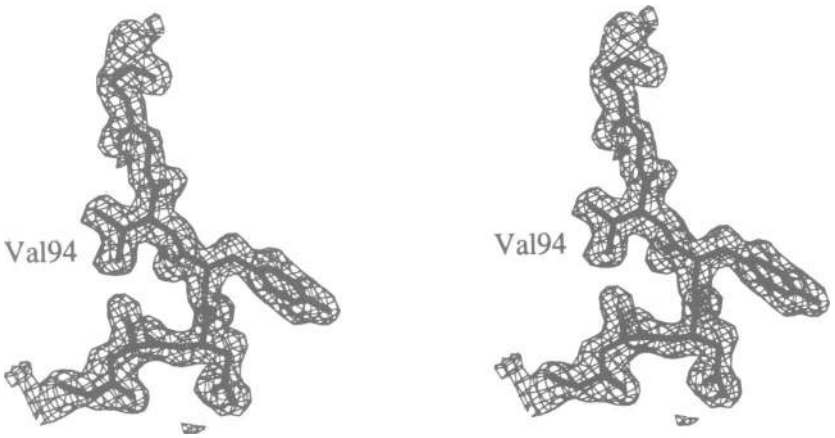

Fig. 6. A portion of the $3F_{obs}-2F_{calc}$ electron density map around residues 90–96. The map is contoured at the 1σ level.

cells were incubated for 3 h at the same temperature. Three grams of cells (obtained from a 1 l culture) were disrupted by grinding with alumina, and the cell lysate was supplemented with NaCl to a final concentra-tion of 0.8 M. The cell debris and the ribosomes were removed from the cell lysate by two successive centrifugations (13 000 *g* for 10 min and 100 000 *g* for 3 h). The supernatant was heated at 70°C for 20 min and denatured *E.coli* proteins were removed by centrifugation. The clarified supernatant was dialysed against 40 mM Na acetate, pH 5.6, buffer, containing 150 mM NaCl and 1 mM dithiothreitol, and loaded onto a 6 ml CM–Sepharose CL-6B column equilibrated with the same buffer. The protein was eluted from the column with a linear gradient of 150–600 mM NaCl (2×40 ml) in the same buffer. Fractions containing overexpressed protein were pooled. Ten to fifteen mg of purified L1 could be obtained from 3 g of cells. The protein purity, as tested by SDS electrophoresis (Laemmli, 1970), was at least 98%.

Crystallization

The procedure used for crystallization was as described earlier (Agalarov *et al.*, 1990) with some modifications. Crystals were obtained by the hanging drop vapour diffusion technique. Ten μl drops containing 12–15 mg protein/ml, 50 mM glycine-NaOH (pH 10.0), 5% (v/v) methane pentanediol and 30% saturated ammonium sulphate were equilibrated against 60% saturated ammonium sulphate with 7% methane pentanediol at room temperature. Crystals up to 0.8×0.4×0.4 mm were obtained in a few days. Crystals with overexpressed L1 belong to the same space group and have the same cell dimensions as the crystals of the original L1: P21212, a = 82.40 Å, b = 63.36 Å, c = 44.64 Å.

Preparation of heavy atom derivatives

The heavy atom derivatives were prepared by soaking crystals in a standard solution to which heavy metal compounds were added (Table I). The standard solution was prepared by mixing 5 ml of 60% saturated ammonium sulphate and 0.3 ml of methane pentanediol.

X-ray data collection

The native data set at 2.2 Å resolution and the derivative data sets were collected with a Siemens area detector using graphite monochromatized CuKα radiation from a Rigaku RU200 rotating anode generator operated at 50 kV and 80 mA. The native data set at 1.8 Å resolution was collected using synchrotron radiation at station X31 at the EMBL outstation, DESY, Hamburg. The frames were processed using either the XENGEN program system (Howard *et al.*, 1987) or the XDS program (Kabsch, 1988a,b). Each data set was obtained from one crystal. Crystals with overexpressed L1, soaked in the standard solution containing Na mersalyl or thiolato(2,2′,2″-terpyridine)platinum (II) complex (Jennette *et al.*, 1976), were aligned to record the Bijvoet pairs within the same frame to minimize the error in measuring the anomalous differences. The derivative data sets were scaled to native data using the program CTAB (S.Nikonov, unpublished).

Structure determination

The structure was solved using the standard method of multiple isomorphous replacement (MIR). Large changes in cell dimensions of some derivative crystals were observed, but only those having cell parameters close to the native crystals (<1% change) were used for phasing. The major heavy atom sites of Na mersalyl and thiolato(2,2′,2″-terpyridine)platinum (II) complex derivatives of overexpressed L1 were obtained from difference Patterson maps and refined at 3.7 Å resolution. Phases were calculated from each derivative, and cross-difference Fourier maps confirmed the heavy atom locations. Positional parameters for minor sites and for other suitable derivatives obtained from original L1 crystals were determined from difference Fourier maps. Phases improved by SQUASH (Zhang, 1993) were used for these calculations. MIR statistics are given in Table I. The 2.8 Å resolution MIR phases based on all six derivatives yielded an electron density map with many interpretable features. The phases were improved further and extended to 2.2 Å resolution by solvent flattening, histogram matching and application of Sayre's equation with the program DM (Cowtan, 1994). The map interpretation was based on skeletonized electron density (Greer, 1974). Model building was performed using the program 'O' (Jones *et al.*, 1991) and the model was refined using the program X-PLOR (Brünger, 1992). The map allowed ~60% of the molecule to be built unambiguously. This initial structure was refined by simulated annealing. The remaining parts of the molecule were built in several iterative steps involving a combined usage of the initial extended map, $3F_{obs}-2F_{calc}$, $F_{obs}-F_{calc}$ and phase combined maps, as well as simulating annealing. The final electron density map was of sufficient quality to trace the polypeptide chain unambiguously (Figure 6). The final model was refined against the synchrotron data to 1.85 Å resolution (Table II). Crystallographic calculations were performed using the CCP4 package (SERC, 1979).

Structure analysis

The quality of the structure has been assessed using the program PROCHECK (Laskowski *et al.*, 1993). Amino acid sequences were obtained from the SWISS-PROT data bank (Bairoch and Boeckmann, 1993) and aligned with program Pileup of the GCG program suite (Genetic Computer Group, 1994). Minor adjustments of the alignment were made in regions with gaps. A protein data bank search for structural similarities to the L1 structure was carried out with program Dejavu (G.Kleywegt, Uppsala University, Sweden). Analysis of the molecular surface, including calculations of electrostatic surface potentials and accessible surface area, was done with program GRASP (Nicholls *et al.*, 1991).

Acknowledgements

We thank S.Sedelnikova, N.Avliyakulov and I.Gryshkovskaya for the help in the work on limited proteolysis of L1. The work was supported by the Russian Foundation for Fundamental Investigations (95-04-11850-a), The Russian Academy of Sciences, The Swedish Research Council, The Swedish National Board for Industrial and Technical Development, The Swedish Royal Academy of Sciences and the International Science Foundation (MTV000). The research of M.Garber was supported in part by an International Research Scholar's award from the Howard Hughes Medical Institute.

References

Agalarov,S.Ch., Eliseikina,I.A., Sedelnikova,S.E., Fomenkova,N.P., Nikonov,S.V. and Garber,M.B. (1990) Crystals of protein L1 from the 50S ribosomal subunit of *Thermus thermophilus*. *J. Mol. Biol.*, **216**, 501–502.

Amons,R., Muranova,T.A., Rykunova,A.I., Eliseikina,I.A. and Sedelnikova,S.E. (1993) The complete primary structure of ribosomal protein L1 from *Thermus thermophilus*. *J. Prot. Chem.*, **12**, 725–734.

Argos,P. (1988) An investigation of protein subunit and domain interfaces. *Prot. Engng*, **2**, 101–113.

Arndt,E. and Weigel,C. (1990) Nucleotide sequence of the genes encoding the L11, L1, L10 and L12 equivalent ribosomal proteins from the archaebacterium *Halobacterium marismortui*. *Nucleic Acids Res.*, **18**, 1285–1285.

Arnold,E., Ding,J., Hughes,S.H. and Hostomsky,Z. (1995) Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. *Curr. Opin. Struct. Biol.*, **5**, 27–38.

Baier,G., Hohenwarter,O., Hofbauer,C., Hummel,H., Stöffler-Meilicke,M. and Stöffler,G. (1989) Structural and functional equivalence between ribosomal proteins of *Escherichia coli* L1 and *Methanococcus vannielii* L6. *Syst. Appl. Microbiol.*, **12**, 119–126.

Baier,G., Piendl,W., Redl,B. and Stöffler,G. (1990) Structure, organization and evolution of the L1 equivalent ribosomal protein gene of the archaebacterium *Methanococcus vannielii*. *Nucleic Acids Res.*, **18**, 719–724.

Bairoch,A. and Boeckmann,B. (1993) The SWISS-PROT protein sequence data bank, recent developments. *Nucleic Acids Res.*, **21**, 3093–3096.

Brändén,C.-I. (1980) Relation between structure and function of α/β proteins. *Q. Rev. Biophys.*, **13**, 317–338.

Brändén,C. and Tooze, J. (1991) *Introduction to Protein Structure*. Garland Publishing, New York.

Branlant,C., Krol,A., Machatt,A. and Ebel,J.-P. (1981) The secondary structure of the protein L1 binding region of the ribosomal 23S RNA: homologies with putative secondary structures of the L11 mRNA and of a region of mitochondrial 16S RNA. *Nucleic Acids Res.*, **9**, 293–307.

Brünger,A.T. (1992) *X-PLOR Version 3.1. A System for Crystallography and NMR*. Yale University, New Haven, CT.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Cowtan,K. (1994) 'dm': an automated procedure for phase improvement by density modification. In Bailey,S. and Wilson,K.S. (eds), *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. Darebury Laboratory, Warrington, UK, pp. 34–38.

Draper,D.E. (1989) How do proteins recognize specific RNA sites? New clues from autogenously regulated ribosomal proteins. *Trends Biochem. Sci.*, **14**, 335–338.

Efimov,A.V. (1994) Common structural motifs in small proteins and domains. *FEBS Lett.*, **355**, 213–219.

Frank,J. *et al.* (1995) A model of protein synthesis based on cryo-electron microscopy of the *E.coli* ribosome. *Nature*, **376**, 441–444.

Genetic Computer Group (1994) *Program Manual for the Wisconsin Package, Version 8*. Genetic Computer Group, 575 Science Drive, Madison, WI.

Golden,B.L., Hoffman,D.W., Ramakrishnan,V. and White,S.W. (1993a) Ribosomal protein-S17—characterization of the 3-dimensional structure by H-1-NMR and N-15-NMR. *Biochemistry*, **32**, 12812–12820.

Golden,B.L., Ramakrishnan,V. and White,S.W. (1993b) Ribosomal protein-L6—structural evidence of gene duplication from a primitive RNA binding protein. *EMBO J.*, **12**, 4901–4908.

Gourse,R.L., Thurlow,D.L., Gerbi,S.A. and Zimmermann,R.A. (1981) Specific binding of a prokaryotic ribosomal protein to a eukaryotic ribosomal RNA: implications for evolution and autoregulation. *Proc. Natl Acad. Sci. USA*, **78**, 2722–2726.

Gourse,R.L., Sharrock,R.A. and Nomura,M. (1986) Control of ribosome synthesis in *Escherichia coli*. In Hardesty,B. and Kramer,G. (eds), *Structure, Function, and Genetics of Ribosomes*. Springer-Verlag, New York, pp. 766–788.

Greer,J. (1974) Three-dimensional pattern recognition: an approach to automated interpretation of electron density maps of proteins. *J. Mol. Biol.*, **82**, 279–301.

Görlach,M., Wittekind,M., Beckman,R.A., Mueller,L. and Dreyfuss,G. (1992) Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.*, **11**, 3289–3295.

Hanner,M., Mayer,C., Kohrer,C., Golderer,G., Grobner,P. and Piendl,W. (1994) Autogenous translational regulation of the ribosomal MvaL1 operon in the archaebacterium *Methanococcus vannielii*. *J. Bacteriol.*, **176**, 409–418.

Hoffman,D.W., Davies,C., Gerchman,S.E., Kycia,J.H., Porter,S.J., White,S.W. and Ramakrishnan,V. (1994) Crystal structure of prokaryotic ribosomal protein L9—a bi-lobed RNA-binding protein. *EMBO J.*, **13**, 205–212.

Howard,A.J., Gilliland,G.L., Finzel,B.C., Poulos T.L., Ohlendorf,D.H. and Salemme,F.R. (1987) The use of an imaginary proportional counter in macromolecular crystallography. *J. Appl. Crystallogr.*, **20**, 383–387.

Itoh,T. (1988) Complete nucleotide sequence of the ribosomal 'A' protein operon from the archaebacterium, *Halobacterium halobium*. *Eur. J. Biochem.*, **176**, 297–393.

Jennette,K.W., Gill,J.T., Sadownick,J.A. and Lippard,S.J. (1976) Metallointercalation reagents. Synthesis, characterisation, and structural properties of thiolato(2,2′,2″-terpyridin)platinum(II) complex. *J. Am. Chem. Soc.*, **98**, 6159–6168.

Jones,T.A., Zou,J., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in the models. *Acta Crystallogr.*, **A47**, 110–119.

Kabsch,W. (1988a) Automatic indexing of rotation diffraction pattern. *J. Appl. Crystallogr.*, **21**, 67–71.

Kabsch,W. (1988b) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.*, **21**, 916–924.

Kang,C.H., Chan,R., Berger,I., Lockshin,C., Green,L., Gold,L. and Rich,A. (1995) Crystal structure of the T4 regA translational regulator protein at 1.9 Å resolution. *Science*, **268**, 1170–1173.

Kearney,K.R. and Nomura,M. (1987) Secondary structure of the autoregulatory mRNA binding site of ribosomal protein L1. *Mol. Gen. Genet.*, **210**, 60–68.

Kimura,M., Kimura,J. and Ashman,K. (1985) The complete primary structure of ribosomal proteins L1, L14, L15, L23, L24 and L29 from *Bacillus stearothermophilus*. *Eur. J. Biochem.*, **150**, 491–497.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.

Laskowski,R.A., MacArthur,M.V., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Leijonmarck,M. and Liljas,A. (1987) Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. *J. Mol. Biol.*, **195**, 555–580.

Liao,D. and Dennis,P.P. (1992) Molecular phylogenies based on ribosomal protein L11, L1, L10 and L12 sequences. *J. Biol. Chem.*, **267**, 22787–22797.

Lindahl,M. *et al.* (1994) Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.*, **13**, 1249–1254.

Merrit,E.A. and Murphy,M.E.P. (1994) Raster 3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.

Nagai,K., Oubridge,C., Ito,N., Avis,J. and Evans,P. (1995) The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.*, **20**, 235–240.

Nicholls,A., Sharp,K. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Ollis,D.L., Brick,P., Hamlin,R., Xuong,N.G. and Steitz,T.A. (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762–766.

Orengo,C.A. and Thornton,J.M. (1993) Alpha plus beta folds revisited: some favored motifs. *Structure*, **1**, 105–120.

Oubridge,C., Ito,N., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

Post,L.E., Strycharz,G.D., Nomura,M., Lewis,H. and Dennis,P.P. (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **76**, 1697–1701.

Query,C.C., Bentley,R.C. and Keene,J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1snRNP protein. *Cell*, **57**, 89–101.

Ramakrishnan,V. and White,S.W. (1992) The structure of ribosomal protein-S5 reveals sites of interaction with 16S rRNA. *Nature*, **358**, 768–771.

Ramirez,C., Shimmin,L.C., Newton,C.H., Matheson,A.T. and Dennis,P.P.

(1989) Structure and evolution of the L11, L1, L10, and L12 equivalent ribosomal proteins in eubacteria, archaebacteria, and eucaryotes. *Can. J. Microbiol.*, **35**, 234–244.

Rosen,K.V., Alexander,R.W., Wower,J. and Zimmermann,R.A. (1993) Mapping the central fold of the tRNA/2 fMet* in the P site of the *Escherichia coli* ribosome. *Biochemistry*, **32**, 12802–12811.

Ruengjitchatchawalya,M., Okamato,S., Nihira,T. and Yamada,M. (1993) Nucleotide sequence of the genes encoding L11 and L1 equivalent ribosomal protein from *Streptomyces* Sp FRI-5. *Nucleic Acids Res.*, **21**, 2524.

Schmidt,J., Bubunenko,M. and Subramanian,A.R. (1993) A novel operon organization involving the genes for chorismate synthase (aromatic biosynthesis pathway) and ribosomal GTPase center proteins (L11, L1, L10, L12: rplKAJL) in *Cyanobacterium synechocystis* PCC 6803. *J. Biol. Chem.*, **268**, 27447–27457.

Sedelnikova,S.E., Agalarov,S.Ch., Garber,M.B. and Yusupov,M.M. (1987) Proteins of the *Thermus thermophilus* ribosome. Purification of several individual proteins and crystallization of protein TL7. *FEBS Lett.*, **220**, 227–230.

SERC (1979) *CCP4. Collaborative Computing Project No. 4.* A Suite of Programs for Protein Crystallography. Daresbury Laboratory, Warrington, UK.

Shimmin,L.C. and Dennis,P.P. (1989) Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. *EMBO J.*, **8**, 1225–1235.

Sor,F. and Nomura,M. (1987) Cloning and DNA sequence determination of the L11 ribosomal protein of *Serratia marcescens* and *Proteus vulgaris*. Translation feedback regulation of *Escherichia coli* L11 operon by heterologous L1 proteins. *Mol. Gen. Genet.*, **210**, 52–59.

Stanley,J., Sloof,P. and Ebel,J.-P. (1978) The binding site of ribosomal protein L1 from *Escherichia coli* on the 23S ribosomal RNA from *Bacillus stearothermophilus*. A possible base-paring scheme differing from that proposed for *Escherichia coli*. *Eur. J. Biochem.*, **85**, 309–316.

Stark,H., Mueller,F., Orlova,E.V., Schatz,M., Dube,P., Erdemir,T., Zemlin,F., Brimacombe,R. and van Heel,M. (1995) The 70S *Escherichia coli* ribosome at 23 Å resolution: fitting the ribosomal RNA. *Structure*, **3**, 815–821.

Stöffler,G. and Stöffler-Meilicke,M. (1986) Immuno electron microscopy on *Escherichia coli* ribosomes. In Hardesty,B. and Kramer,G. (eds), *Structure, Function, and Genetics of Ribosomes*. Springer-Verlag, New York, pp. 28–46.

Subramanian,A.R. and Dabbs,E.R. (1980) Functional studies on ribosomes lacking protein L1 from mutant *Escherichia coli*. *Eur. J. Biochem.*, **112**, 425–430.

Trakhanov,S.D., Yusupov,M.M., Agalarov,S.C., Garber,M.B., Ryazantsev, S.N., Tischenko,S.V. and Shirokov,V.A. (1987) Crystallization of 70S ribosomes and 30S ribosomal subunits from *Thermus thermophilus*. *FEBS Lett.*, **220**, 319–322.

Wilson,K.S., Appelt,K., Badger,J., Tanaka,I. and White,S.W (1986) Crystal structure of a procaryotic ribosomal protein. *Proc. Natl Acad. Sci. USA*, **83**, 7251–7255.

Yonath,A. (1992) Approaching atomic resolution in crystallography of ribosomes. *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 77–93.

Yonath,A and Franceschi,F. (1993) Structural aspects of ribonucleoprotein interactions in ribosomes. *Curr. Opin. Struct. Biol.*, **3**, 45–49.

Zhang,K.Y. (1993) SQUASH-combining constraints for macromolecular phase refinement and extension. *Acta Crystallogr.*, **D49**, 213–222.

Zimmermann,R.A. (1980) Interactions among protein and RNA components of the ribosome. In Chambliss,G., Craven,G.R., Davies,J., Davies,K., Kahan,L. and Nomura,M. (eds), *Ribosomes. Structure, Function and Genetics*. University Park Press, Baltimore, MD, pp. 135–169.

Zimmermann,R.A., Thurlow,D.L., Finn,R.S., Marsh,T.L. and Ferrett,L.K. (1980) Conservation of specific protein–RNA interactions in ribosome evolution. In Osava,S., Ozeki,H., Uchida,H. and Yura,T. (eds), *Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes*. University of Tokyo Press, Tokyo, pp. 569–584.

Received on September 20, 1995; revised on October 30, 1995