The solution structure of ribosomal protein L18 from *Thermus thermophilus* reveals a conserved RNA-binding fold

Esmeralda A. WOESTENENK*, George M. GONGADZE†, Dmitry V. SHCHERBAKOV†, Alexey V. RAK†, Maria B. GARBER†, Torleif HÄRD* and Helena BERGLUND*¹

*Royal Institute of Technology (KTH), SCFAB, Department of Biotechnology, S-106 91 Stockholm, Sweden, and †Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

We have determined the solution structure of ribosomal protein L18 from *Thermus thermophilus*. L18 is a 12.5 kDa protein of the large subunit of the ribosome and binds to both 5 S and 23 S rRNA. In the uncomplexed state L18 folds to a mixed α/β globular structure with a long disordered N-terminal region. We compared our high-resolution structure with RNA-complexed L18 from *Haloarcula marismortui* and *T. thermophilus* to examine RNA-induced as well as species-dependent structural differences. We also identified *T. thermophilus* S11 as a structural homologue

and found that the structures of the RNA-recognition sites are conserved. Important features, for instance a bulge in the RNA-contacting β -sheet, are conserved in both proteins. We suggest that the L18 fold recognizes a specific RNA motif and that the resulting RNA–protein-recognition module is tolerant to variations in sequence.

Key words: NMR spectroscopy, protein structure, ribosome, RNA-binding protein.

INTRODUCTION

Recently published crystal structures of ribosomal subunits and entire ribosomes at high resolution have increased our understanding of the structure and activity of the ribosome to a great extent. Structures of the large ribosomal subunit have been determined of the archaea *Haloarcula marismortui*, at 2.4 Å resolution [1], and recently of the Gram-positive bacterium *Deinococcus radiodurans*, at 3.1 Å [2]. Two high-resolution structures of the small subunit of the Gram-positive bacterium *Thermus thermophilus* have been reported: one at 3.0 Å [3], and one at 3.3 Å of a functionally activated subunit [4]. Furthermore, the structure of the whole ribosome of *T. thermophilus* has been solved at a resolution of 5.5 Å [5]. The structural data showed that the peptidyl transferase site is completely made up of rRNA and that the nearest distance to a protein residue is 18 Å. This observation led to the suggestion that the peptidyl transferase function lay solely with the rRNA, and that the ribosomal proteins functioned merely as cement, stabilizing the rRNA [1]. However, two recent studies [6,7] showed that correct placement of tRNA and rRNA with respect to each other is the very basis of peptidyl transferase activity. Since ribosomal proteins introduce the correct structure into the rRNA they contribute indirectly to catalytic activity. Apart from this more secondary function in peptide-bond formation, the primary rRNA-binding proteins are necessary for ribosome assembly by changing the secondary structure of the RNA sites they bind and thus introducing new binding sites for secondary and tertiary rRNA-binding proteins. Moreover, ribosome assembly is known to be linked to ribosome synthesis by the auto-regulatory function of the expression of several ribosomal proteins. Apart from their own expression these proteins control the expression of other ribosomal proteins in the same operon [8]. In addition to ribosome-related functions, many ribosomal proteins have additional extra-ribosomal functions [9]. Complete understanding of all tasks performed by the ribosomal proteins thus requires

information of the structure and dynamics of the ribosomal proteins in their RNA-bound and free states.

Ribosomal protein L18 from *T. thermophilus* (TthL18) contains 111 amino acids and has a molecular mass of 12.5 kDa. It is located in the central protuberance of the large subunit of the bacterial ribosome. In *T. thermophilus* the central protuberance consists of ribosomal proteins L5 and L18, together with the 5 S rRNA molecule and part of domain V of the 23 S rRNA molecule. L18 is a primary 5 S rRNA-binding protein, i.e. no binding of other components of the 5 S/23 S rRNA–protein complex is required for binding of L18 to 5 S rRNA. Binding of L18 to *Escherichia coli* 5 S rRNA induces a conformational change in the junction of helices 1, 2 and 4 (also called helices A, B and D, respectively) [10], and has been shown to stimulate binding of L5 to 5 S rRNA and formation of the 5 S/23 S rRNA–protein complex [11]. The N-terminal 20 residues of L18 are not required for binding of 5 S rRNA, but they are needed for formation of the complex with 23 S rRNA and L5 [11].

Here we present the solution structure of the uncomplexed TthL18, determined by NMR spectroscopy. From a comparison of the uncomplexed and the recently published RNA-complexed structures of L18 we see that RNA-contacting loops change their conformation when they reach out to interact with RNA. Furthermore, comparison of the structures of TthL18 and its homologue in *T. thermophilus*, S11, has resulted in the suggestion that the L18 scaffold recognizes a specific RNA motif where sequence variation is allowed in both protein and RNA sequences to ensure sequence-specific binding.

EXPERIMENTAL

NMR spectroscopy

NMR samples contained 0.8–1.6 mM L18 in a 50 mM KH_2PO_4 /200 mM LiCl NMR buffer (pH 5.9) and 10% (v/v) $^2\text{H}_2\text{O}$. All

Abbreviations used: NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; HSQC, heteronuclear single-quantum coherence spectroscopy; r.m.s.d., root mean square deviation; TthL18, *Thermus thermophilus* L18; HmaL18, *Haloarcula marismortui* L18.

¹ To whom correspondence should be addressed (e-mail helena.berglund@biotech.kth.se).

The coordinates for the presented ensemble of L18 structures have been deposited in the Protein Data Bank under accession code 1ILY.

Figure 1 Amino acid sequence alignment of L18 and homologous eukaryotic L5 sequences identified by their SWISS-PROT codes

Residues that are part of the hydrophobic core in TthL18 are shaded black. Grey shading indicates universally conserved or conservatively substituted amino acids that do not contribute to the hydrophobic core of TthL18. Of the archaeal and eukaryotic sequences only the regions homologous to L18 are included in the alignment. The secondary-structure elements of TthL18 are indicated above the alignment; open boxes indicate β -strands and filled boxes are α -helices; ht, helical turn.

NMR spectra were recorded at 30 °C on Bruker AVANCE spectrometers at 500 and 600 MHz unless indicated otherwise. Data were processed using NMRPipe [12] on SGI workstations. All spectra were analysed using the software ANSIG 3.3 [13] on SGI workstations or Ansig for Windows [14] on a PC. Expression and purification of TthL18, as well as resonance assignments, are described in [15]. Nuclear Overhauser enhancement (NOE) distance restraints were extracted from two-dimensional NOE spectroscopy (NOESY) [16] recorded at 800 MHz, ^{15}N -edited NOESY-heteronuclear single-quantum coherence spectroscopy (HSQC) [17] recorded at 700 MHz and ^{13}C -edited NOESY-HSQC [18] recorded at 800 MHz, all with cross-relaxation mixing times of 120 ms. The two-dimensional NOESY and the ^{15}N -edited NOESY-HSQC were recorded on Bruker AVANCE spectrometers, while the ^{13}C -edited NOESY-HSQC was recorded on a Varian Inova spectrometer.

NOE distance restraints

NOESY crosspeaks were integrated within ANSIG and distances were calibrated according to known distances in secondary-structure elements [19]. NOE restraints were grouped into four classes: very weak (1.8–6.0 Å), weak (1.8–4.3 Å), medium (1.8–

3.5 Å) and strong (1.8–2.5 Å). The software AQUA [20] was used to remove redundant restraints.

Dihedral angle restraints and stereospecific assignments

Backbone ϕ dihedral angle restraints were derived from $^3J_{\text{HNH}\alpha}$ couplings as described previously [15]. For $^3J_{\text{HNH}\alpha}$ coupling values lower than 5.0 Hz, the ϕ angle was restrained to -60° ($\pm 30^\circ$). When the $^3J_{\text{HNH}\alpha}$ value exceeded 8.0 Hz, the ϕ angle was restrained to -120° ($\pm 30^\circ$). Dihedral angles were also determined from chemical-shift data using the software TALOS [21]. When ϕ angles could not be determined from $^3J_{\text{HNH}\alpha}$ couplings and TALOS ϕ angles were classified as good, the TALOS angles were used in the structure calculation. All TALOS ψ angles classified as good were used in the structure calculation. ϕ angles of the remaining residues (except Gly and Pro) were restrained to be negative ($-90^\circ \pm 90^\circ$).

In order to determine χ_1 angles, short-mixing-time ^{15}N -TOCSY-HSQC (40 ms isotropic mixing time) and ^{15}N -NOESY-HSQC (50 ms cross-relaxation mixing time) experiments [17,22], as well as a constant-time HNHB experiment [23], were recorded at 600 MHz. Angles were restrained to gauche gauche ($+60^\circ$), trans gauche (-60°) or gauche trans (180°) based on the $^3J_{\text{H}\alpha\text{H}\beta}$ coupling (TOCSY), the $^3J_{\text{H}/\text{HN}}$ coupling (HNHB) and

Charges were assigned to atoms by the simple charge definition of MOLMOL. Dielectric constants of 2 and 80 were used for protein and solvent respectively. The solvent radius was 1.4 Å, the salt concentration was 1 M and the salt radius was 2 Å. A boundary condition of zero potential at 10 Å from the surface of the molecule was applied.

RESULTS

NMR spectroscopy, assignments and structure calculations

Sequence homology was only found within the L18 family; the alignment of the amino acid sequences of TthL18 and other members of the L18 family is shown in Figure 1. A total of 101 of the 109 expected backbone resonances could be assigned in the $^1\text{H}, ^{15}\text{N}$ -HSQC spectrum of L18 recorded at 30 °C (Figure 2). The assignment and secondary-structure characterization of TthL18 has been published previously [15]. A total of 1925 NOE distance restraints were used for structure calculation, the statistics of which are shown in Table 1. The number of NOE distance restraints per residue was quite evenly distributed for residues 23–111 (results not shown). For the N-terminal 22 residues either no NOEs or only intra-residue and sequential NOEs could be observed. Similarly, a decrease in the number of NOE restraints was observed for loop regions around residues 31–32, 58–63 and 90–95. A notably high number of long-range NOEs (69) was assigned to Leu-25, which was found to be completely buried in the core of the protein.

A total number of 125 backbone dihedral angles (68 ϕ angles and 57 ψ angles), derived from $^3J_{\text{HNH}\alpha}$ couplings or predicted by TALOS [21], were included as restraints in the structure calculation (Table 1). For 12 residues, the side chain χ_1 angles were

determined. Additionally, 34 hydrogen bonds defined by 68 distance restraints were invoked in the calculation.

A total of 50 structures were calculated within the program CNS [24] using simulated annealing. Of these, 27 structures were selected for further analysis, based on their low energy. For this ensemble, the root mean square deviation (r.m.s.d.) of the backbone heavy atoms compared with the average structure was 0.48 Å (± 0.11 Å) for well-ordered regions. Structural statistics (Table 1) of the ensemble are based on structures calculated for residues 22–111. The N-terminal region was left out of the calculation because it was determined previously to be unstructured in solution [15]. In some structures four residues (Leu-31, Lys-61, Tyr-93 and His-94) have dihedral angles in the generously allowed region of the Ramachandran diagram. In 12 out of the 27 selected structures, Lys-56 has a disallowed backbone conformation. These residues reside in regions where there are little or no structural data. For example, no dihedral angle restraints are available for any of these five residues. Additionally, no NOE distance restraints are available for Leu-31 and Lys-61, and very few NOEs have been assigned to Lys-56, Tyr-93 and His-94.

Description of the structure

The final ensemble of L18 structures (residues 22–111) and the average structure with secondary-structure elements included are shown in Figure 3. The structured part of L18 consists of two α -helices, a β -sheet formed by three anti-parallel β -strands and one small parallel β -strand, and a helical turn. The unstructured N-terminus is followed by the first β -strand (β_1 ; Leu-23–Arg-29), loop 1 (Ser-30–His-33), the second β -strand (β_2 ; Ile-34–Asp-40) anti-parallel to β_1 , loop 2 (Asp-41–Val-45), the third β -strand (β_3 ; Thr-46–Ser-51) anti-parallel to β_2 and with a bulge at Leu-47, a short helical turn (Ser-52–Leu-55), loop 3 (Lys-56–Thr-62), the first helix (α_1 ; Glu-63–Leu-79), loop 4 (Gly-80–Gln-83), the fourth β -strand (β_4 ; Val-84–Phe-86) parallel to β_1 , loop 5 (Asp-87–Gly-95), the second helix (α_2 ; Arg-96–Gly-107) and the extended C-terminus (Gly-108–Phe-111). The C-terminal three residues of L18 are well ordered. The dihedral angles and overall organization of the C-terminus suggested the presence of an additional or transient β -strand parallel to the beginning of β_4 . However, since $^3J_{\text{HNH}\alpha}$ coupling constants for Gln-110 and Phe-111 are lower than in a well ordered β -strand, the region was designated as a coil. Slow amide hydrogen-exchange rates suggest that the amide hydrogen of Phe-111 is hydrogen-bonded, and the only hydrogen-bond acceptor in the vicinity of the amide hydrogen is the aromatic ring of the same residue. In the structure the amide hydrogen is located above the ring plane, approx. 2.5 Å from the centre. This position is also manifested in an upfield shift of the amide hydrogen chemical shift (Figure 2). The N–H bond vector has a parallel orientation with respect to the side-chain ring, which is the most frequently observed relative positioning in intra-residue backbone amide–aromatic ring interactions [26]. Aromatic–backbone amide interactions have been found to be stabilizing for local structure [26] and the conservation of Phe-111 in bacterial L18 indicates that its presence is required for stabilization of the C-terminus.

Apart from the flexible N-terminus, there is some mobility in loop 3 (between the helical turn and helix α_1), as judged by attenuated (Asn-60) and/or broadened (Leu-53, Lys-56 and Gly-59) amide resonances (Figure 2), and lowered $\{^1\text{H}\}$ - ^{15}N steady-state NOE values [15]. The top panel of Figure 3 shows that the backbone r.m.s.d. values are relatively large for loop 3.

The solvent-accessible surface of TthL18 was calculated in MOLMOL [25] using a probe of 1.4 Å. Residues less than 5% accessible were defined as buried. A large part of the hydrophobic

Table 1 Structural statistics of the L18 ensemble

Total number of NOE-derived distance restraints	1925
Intra-residue	459
Sequential ($ i-j = 1$)	504
Medium-range ($1 < i-j \leq 5$)	429
Long-range ($ i-j > 5$)	533
Total number of dihedral angle restraints	137
Backbone ϕ angle restraints	68
Backbone ψ angle restraints	57
Side-chain χ_1 angle restraints	12
Number of hydrogen bonds	34
Number of violations	
Distance restraints (> 0.5 Å)	0
Dihedral angle restraints ($> 5^\circ$)	0
Deviations from idealized covalent geometry	
Bonds (Å)	0.0037 ± 0.00017
Angles ($^\circ$)	0.5769 ± 0.0184
Impropers ($^\circ$)	0.5148 ± 0.023
Atomic r.m.s.d. to the average structure (L23–R29, H33–D40, V45–L55, A65–G89, R96–F111)	
Backbone heavy atoms (Å)	0.48 ± 0.11
All heavy atoms (Å)	1.26 ± 0.23
Ramachandran plot analysis (residues 22–111)	
Residues in most favoured regions (%)	74.5
Residues in additional allowed regions (%)	23.1
Residues in generously allowed regions (%)	1.9
Residues in disallowed regions (%)	0.5
Ramachandran plot analysis (L23–R29, H33–D40, V45–L55, A65–G89, R96–F111)	
Residues in most favoured regions (%)	84.2
Residues in additional allowed regions (%)	15.8
Residues in generously allowed regions (%)	0.0
Residues in disallowed regions (%)	0.0

Figure 3 Solution structure of L18

Top panel: stereo view of the final ensemble of 27 structures superimposed on ordered regions (residues 23–29, 33–40, 45–55, 65–89 and 96–111). Residues 1–21 are not depicted. Bottom panel: ribbon representation of the secondary-structure elements of L18. The structure used for this figure is the average structure of the ensemble shown in the top panel. All structures shown were made using MOLMOL [25].

core is formed by residues of the β -sheet and helix $\alpha 1$ (Figure 1). The second helix, $\alpha 2$, lies anti-parallel to $\alpha 1$ in a slightly tilted way and is more exposed to the solvent than $\alpha 1$.

The electrostatic surface of L18 consists of distinct patches with positive or negative potential, as is shown in Figures 4(B) and 4(C). Of the 22 unstructured N-terminal residues, 10 are positively charged. Most of these charges are conserved, in both bacterial and archaeal L18. In the globular part of L18 most positive charges cluster on the surface formed by residues of the

β -sheet, loops 1, 3 and 5, and the two helices. Most of the surface-exposed residues of the β -sheet are neutral.

The Protein Data Bank was searched for structurally homologous proteins using the programs TOP [27] and DALI [28]. Apart from *H. marismortui* L18 (HmaL18; with a DALI z score of 8.4), ribosomal protein S11 from *T. thermophilus* was identified as a structural homologue by both programs, with a DALI z score of 7.3. Sequence alignment showed that 21.6% of the residues of S11 and TthL18 were identical.

Figure 4 Interaction of L18 with RNA

(A) Superposition in stereo of HmaL18 (residues 29–129) in blue, RNA-bound TthL18 of the 5.5 Å crystal structure as modelled from HmaL18 (residues 29–129) in green, and the free TthL18 NMR structure (residues 18–111) in red, positioned on the L18-binding part of *H. marismortui* 5S rRNA (nucleotides 1–70 and 111–122), which is shown in surface mode. The N- and C-termini of the proteins are indicated. (B) Surface electrostatic potential of L18. Red and blue represent negative and positive electrostatic potentials, respectively. L18 is in the same orientation as in (A), while in (C) the molecule is rotated 90° upwards in order to show the RNA-binding surface. (D) Left-hand side: alignment of S11, TthL18 and HmaL18 with secondary-structure elements of each protein indicated (β -strand, blue shading; α -helix, red shading). Right-hand side: alignment of the 5S and 16S rRNA nucleotides of which the backbone phosphate atoms are displayed as spheres in (E); sequence identity is shown as red shading. (E) Superposition of HmaL18 (residues 30–130) in orange and TthS11 (residues 14–108) in blue with the parts of their respective substrates, 5S rRNA nucleotides 1–12 and 116–122 (yellow ribbon), and 16S rRNA nucleotides 676–690 and 701–713 (light-blue ribbon), that have comparable structures and orientations in the two complexes.

DISCUSSION

We have determined the high-resolution solution structure of ribosomal protein TthL18. Comparable with many other ribosomal proteins, L18 contains a globular domain and an extended region. The globular domain folds to a mixed α/β structure while the N-terminus is totally unstructured and flexible in solution [15]. By comparison of our structure of uncomplexed L18 with the RNA-complexed L18 structures from *H. marismortui* and *T. thermophilus* we were able to investigate both species-specific and RNA-induced structural differences. Despite low sequence similarities we also found a pronounced similarity between L18 and ribosomal protein S11, both in terms of structure and in their modes of RNA binding. Finally we were able to map the regions important for extra-ribosomal functions of the L18-like domain of an eukaryotic ribosomal protein on to the uncomplexed L18 structure.

Comparison with RNA-complexed L18

We compared our solution structure of TthL18 with the 2.4 Å crystal structure of HmaL18 determined within the large subunit of the archaeal ribosome [1]. We also compared our solution structure with the HmaL18-based C α model of TthL18 from the 5.5 Å crystal structure of the complete *T. thermophilus* ribosome [5]. HmaL18 is larger than TthL18 but most buried residues in the globular domain are universally conserved or conservatively substituted (Figure 1). The two helix-breaking glycines (Gly-80 and Gly-108) are also conserved. The structure of their globular domains is essentially the same (Figure 4A). Like TthL18, HmaL18 contains a bulge in β 3, with Leu-58, corresponding to Leu-47 in TthL18, as the bulge-forming residue. Superposition of 48 C α atoms (22–30, 33–40, 50–52, 66–79 and 97–110 of TthL18) of the secondary-structure elements of the two structures results in a r.m.s.d. of 1.16 Å between RNA-complexed HmaL18 and uncomplexed TthL18.

In RNA-bound L18, the surface of the four-stranded β -sheet forms a large interface with 5 S RNA. Seven of the eleven residues that face the RNA are conserved between HmaL18 and TthL18, and the side-chain distribution allows for a close fit of the β -sheet with the groove of the 5 S RNA helix. The structure of the β -sheet does not change upon RNA interaction, apart from a small translocation of β 3 in the *T. thermophilus* ribosome L18 model. Structural differences between the uncomplexed and the RNA-complexed protein are instead found in the loops that reach out to contact RNA. Loop 2, together with the top part of the β -sheet, stretches out over helix 1 towards loop A of 5 S rRNA, whereas the short helical turn following β 3 together with loops 3 and 5 contact 5 S rRNA loops B and C and helix 3. Loop 2 is longer in *H. marismortui* and, as a result of this and its interaction with RNA, its conformation differs in all three L18 structures (Figure 4A). In HmaL18 the helical turn is extended to a short helix. Since the same extension is not seen in TthL18 complexed with RNA we interpret this as a species-dependent difference. The conformation of loop 3, which displays length variation within the L18 family (Figure 1), differs in all three L18 structures. In the uncomplexed state loop 3 undergoes conformational dynamics, some of which could remain in the RNA-complexed state. Loop 5 possesses very similar conformations in HmaL18 and uncomplexed L18 but as a result of interaction with RNA it is differently oriented in the two proteins. Finally, loop 4 has an identical conformation in all three proteins; it faces out from the globular domain of L18 and lacks RNA contacts.

By measuring intermolecular distances less than 4 Å between heavy atoms in the protein and on the RNA bases, we identified putative protein–rRNA base interactions in the HmaL18–5 S

rRNA complex. Such interactions were found for eight residues, four of which are located in loop regions (Asn-40 and Lys-41 in loop 1, Asn-53 in loop 2 and Gly-112 in loop 5). The other four are β -sheet residues (Arg-44 and Gln-46 in β 2, Asp-55 and Thr-57 close to the bulge in β 3). Of these eight residues, only Gln-46 is completely conserved in the L18 family (Figure 1). Asp-40, Arg-44, Asn-53 and Asp-55 are not conserved, whereas Lys-41 is mostly conserved and Gly-112 is totally conserved in archaea and eukarya. The position of Thr-57, corresponding to Thr-46 in TthL18, is mostly occupied by threonine and valine residues (Figure 1). The 5 S rRNA nucleotides participating in base–protein interactions are divided over helix 1 (contacting residues of β 2, loop 2 and the bulge region) and helix 3/loop C (contacting residues of loops 1 and 5). Whereas the protein-contacting nucleotides of helix 1 vary in sequence between *H. marismortui*, *T. thermophilus* and *E. coli*, the helix 3/loop C nucleotides are completely conserved. Additionally, 20 residues are located within the same short distance from the RNA backbone and could participate in sugar and phosphate backbone interactions. This group includes the majority of the RNA-facing residues on the surface of β 1 and β 2 as well as most residues of loop 2 and the region around the bulge in β 3, and some residues in loops 3 and 5. Most of the residues available for RNA-backbone contacts are not conserved between different kingdoms or even within the same kingdom. The exception is Asn-77 in loop 3, which is almost universally conserved.

The largest RNA-induced conformational change occurs in the basic and extended N-terminus. From being completely unstructured in the uncomplexed state it forms a small helical turn (Arg-2–Tyr-6) and a helix (Asp-19–Ser-29) upon complex formation with RNA. The N-terminus is rich in contacts with RNA and is positioned like a rod between 5 S rRNA (helix 2 and loop A) and 23 S rRNA. The N-terminus of HmaL18 also contacts neighbouring ribosomal protein L5. It is not known whether a helix is formed in TthL18 upon interaction with 23 S rRNA. However, the N-terminus of the recently published structure of *D. radiodurans* L18 [2] does contain a helix in this position, which suggests that L18 from the closely related *T. thermophilus* may have a comparable structure.

Comparison with RNA-complexed S11

The crystal structure of the structural homologue S11 has been determined within the small subunit of the *T. thermophilus* ribosome to a resolution of 3.0 Å [3]. With 129 amino acids, S11 is slightly larger than L18. While the N-terminus of S11 is six residues shorter than the N-terminus of TthL18, the C-terminus is longer and includes an extra β -strand. Apart from the N- and C-termini, the structures of L18 and S11 are highly similar, despite low sequence similarity (Figure 4D). The solvent-accessible surface was determined for S11 in the same way as for TthL18 (see the Results section). It was found that the distribution of core residues throughout both structures is extremely similar. Although only 6 of the 21 core residues in TthL18 are identical in S11, 10 additional residues are substituted conservatively. The likeness between the TthL18 and S11 cores agrees well with the high structural similarity that we have found. Superposition of S11 and TthL18 on 45 residues in secondary-structure elements according to a sequence alignment generated by the GeneStream server [29] yields a r.m.s.d. of 3.6 Å, while superposition of the same residues based on the structural alignment by DALI [28] results in a r.m.s.d. of 1.4 Å. The sequence alignment between HmaL18, TthL18 and S11 shown in Figure 4(D) is therefore based on the structural alignment by DALI. From the alignment it can be observed that the RNA-contacting

loops show substantial variation in amino acid and charge composition. TthL18 contains more positive charges than HmaL18 and S11, and most of the extra charges reside in loop regions. In both L18 and S11 most of the positive charges are clustered in the N-terminal region.

Structural alignment of HmaL18 and S11 with part of their respective RNA substrates shows how similarly the β -sheets of L18 and S11 bind their RNA-recognition sites (Figure 4E). The sequences of 5 S and 16 S rRNA used in this Figure are displayed in Figure 4(D). The RNA-contacting residues of the β -sheet are not very conserved between L18 and S11, yet the β -sheet binds to a very similar RNA motif. In parallel to this, the sequence similarity between the two (5 S and 16 S) RNA motifs is limited to two GC pairs (Figure 4D). As with HmaL18 and TthL18, S11 contains a bulge at the beginning of β 3, with Ile-40 as the bulge-forming residue. Although only the hydrophobic character of the bulge-forming residue is conserved, the structure of the bulge is remarkably similar in the three proteins. In both L18 and S11 the presence of the bulge causes a bending of the preceding loop 2 residues towards the RNA. Based on the same 4 Å condition as used for identification of interactions between HmaL18 and 5 S rRNA bases, residues Gly-37–Pro-39 (in loop 2 and β 3) were found to be available for interaction with the bases of 16 S rRNA base pair C⁷⁰⁷–G⁶⁸³. This is one of the two base pairs in this RNA motif identical with the corresponding base pair in 5 S rRNA (C⁶–G¹¹⁷). The bases of these 5 S rRNA nucleotides are close to loop 2 and β 3 residues of HmaL18. Apart from Gly-37–Pro-39, there are five other residues in S11 close enough to 16 S rRNA to form base contacts. Two of these, Ile-29 and Thr-31, are located in β 2. Interestingly, the corresponding residues in HmaL18 are also close enough to interact with RNA bases, but both the character of the protein residues (Ile-29 and Thr-31 correspond to Arg-44 and Gln-46, respectively, in HmaL18) and that of the interacting bases (UA in 16 S rRNA versus GG in 5 S rRNA) are different. This suggests that protein–RNA contacts in this region are sequence-specific.

When comparing the residues in HmaL18 and S11 involved in RNA backbone contacts, it can be concluded that mostly the same protein regions contribute to these contacts. Protein–RNA backbone contacts are plentiful in the bulge region, in both S11 and HmaL18. We suggest that these contacts serve to fit the local protein and RNA structures together, whereas protein–base contacts account for specific recognition. Most RNA-facing residues of strands β 1 and β 2, as well as those of loop 1, are close enough to contact the RNA backbone. Finally, it may be significant for RNA recognition that the RNA-binding surface made up of the β -sheet with its characteristic bulge is a well-defined structure in solution.

Based on the observations above, we suggest that the L18 fold is a scaffold that recognizes a specific RNA structure. The majority of contacts between protein and RNA are unspecific, but sequence-specific recognition may make the interaction specific for a protein and its rRNA-recognition element. Specific interactions may be necessary for recognition when the rRNA substrate has not yet reached its native structure. This view is supported by the knowledge that certain ribosomal proteins, including L18, introduce the correct structure in their rRNA-binding partner.

L18-like domain in eukaryotic L5

The eukaryotic homologue of L18 is L5. Analysis of the bacterial and archaeal L18 structures and homologous structure S11 suggests that the L18 domain of eukaryotic L5 is likely to have a similar structure, since most buried residues of TthL18 are

conserved or substituted conservatively in the eukaryotic members of the L18 family. In addition to its role in the eukaryotic ribosome, L5 is actively involved in 5 S rRNA nucleocytoplasmic transport. The L18 domain of eukaryotic L5 contains two signal sequences required for this function: a nucleolar-localization signal and a nuclear-export signal [30,31]. The nucleolar-localization signal is located in the region corresponding to the extended N-terminus and covers the helical loop and most of the helix in the N-terminal region of L18 that is unstructured prior to incorporation in the ribosome. The nuclear-export signal is formed by most of helix α 2 and contains a leucine-rich motif important for interaction with an export receptor [30]. The structure of the region corresponding to the nuclear-export signal in L18 suggests that these leucines are structurally important and that the recognition element for the export receptor is a helical structure.

We thank Susanne van den Berg for producing the ¹³C/¹⁵N-labelled L18 sample and the Swedish NMR Centre, in Göteborg, Sweden, for use of their 800 MHz spectrometer. This work was supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

REFERENCES

- Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920.
- Harms, J., Schlutzenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679–688.
- Wimberly, B. T., Brodersen, D. E., Clemons, Jr, W. M., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature (London)* **407**, 327–339.
- Schlutzenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluermann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623.
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D. and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896.
- Polacek, N., Gaynor, M., Yassin, A. and Mankin, A. (2001) Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature (London)* **411**, 498–501.
- Thompson, J., Kim, D. F., O'Connor, M., Lieberman, K. R., Bayfield, M. A., Gregory, S. T., Green, R., Noller, H. F. and Dahlberg, A. E. (2001) Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc. Nat. Acad. Sci. U.S.A.* **98**, 9002–9007.
- Zengel, J. M. and Lindahl, L. (1994) Diverse mechanisms for regulating ribosomal protein synthesis in *Escherichia coli*. *Progr. Nucleic Acids Res. Mol. Biol.* **47**, 331–370.
- Wool, I. G. (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **21**, 164–16519.
- Ciesiolka, J., Lorenz, S. and Erdmann, V. A. (1992) Structural analysis of three prokaryotic 5S rRNA species and selected 5S rRNA-ribosomal-protein complexes by means of Pb(II)-induced hydrolysis. *Eur. J. Biochem.* **204**, 575–581.
- Newberry, V. and Garrett, R. A. (1980) The role of the basic N-terminal region of protein L18 in 5S RNA–23S RNA complex formation. *Nucleic Acids Res.* **8**, 4131–4142.
- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293.
- Kraulis, P. J. (1989) ANSIG: a program for the assignment of protein ¹H 2D NMR spectra by interactive graphics. *J. Magn. Reson.* **24**, 627–633.
- Helgstrand, M., Kraulis, P., Allard, P. and Härd, T. (2000) Ansig for Windows: an interactive computer program for semiautomatic assignment of protein NMR spectra. *J. Biomol. NMR* **18**, 329–336.
- Woestenenk, E. A., Allard, P., Gongadze, G., Moskalenko, S. E., Shcherbakov, D. V., Rak, A. V., Garber, M. B., Härd, T. and Berglund, H. (2000) Assignment and secondary structure identification of the ribosomal protein L18 from *Thermus thermophilus*. *J. Biomol. NMR* **17**, 273–274.

- 16 Macura, S. and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. *Mol. Phys.* **41**, 95–117
- 17 Jahnke, W., Baur, M., Gemmecker, G. and Kessler, H. (1995) Improved accuracy of NMR structures by a modified NOESY-HSQC experiment. *J. Magn. Reson. B* **106**, 86–88
- 18 Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., Smallcombe, S. H. and Kay, L. E. (1993) A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C -labeled proteins dissolved in H_2O . *J. Magn. Reson. B* **102**, 317–321
- 19 Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York
- 20 Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486
- 21 Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302
- 22 Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H - ^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1β . *Biochemistry* **28**, 6150–6156
- 23 Archer, S. J., Ikura, M., Torchia, D. A. and Bax, A. (1991) An alternative 3D NMR technique for correlating backbone ^{15}N with side chain $\text{H}\beta$ resonances in larger proteins. *J. Magn. Reson.* **95**, 636–641
- 24 Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S. et al. (1998) Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–912
- 25 Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55
- 26 Tóth, G., Watts, C. R., Murphy, R. F. and Lovas, S. (2001) Significance of aromatic-backbone amide interactions in protein structure. *Proteins* **43**, 373–381
- 27 Lu, G. (2000) *TOP*: a new method for protein structure comparisons and similarity searches. *J. Appl. Crystallogr.* **33**, 176–183
- 28 Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138
- 29 Person, W. R., Wood, T., Zhang, Z. and Miller, W. (1997) Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24–36
- 30 Rosorius, O., Fries, B., Stauber, R. H., Hirschmann, N., Bevec, D. and Hauber, J. (2000) Human ribosomal protein L5 contains defined nuclear localization and export signals. *J. Biol. Chem.* **275**, 12061–12068
- 31 Pieler, T. and Rudt, F. (1997) Nucleocytoplasmic transport of 5S ribosomal RNA. *Semin. Cell Dev. Biol.* **8**, 79–82

Received 5 November 2001/21 January 2002; accepted 15 February 2002