Nucleic Acids Research, 2002, Vol. 30, No. 4 **931–941**

Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein

Jeffrey F. Kuhn, Elizabeth J. Tran and E. Stuart Maxwell*

Department of Molecular and Structural Biochemistry, North Carolina State University, Box 7622, Raleigh, NC 27695-7622, USA

Received October 26, 2001; Revised and Accepted December 18, 2001

ABSTRACT

Recent investigations have identified homologs of eukaryotic box C/D small nucleolar RNAs (snoRNAs) in Archaea termed sRNAs. Archaeal homologs of the box C/D snoRNP core proteins fibrillarin and Nop56/58 have also been identified but a homolog for the eukaryotic 15.5kD snoRNP protein has not been described. Our sequence analysis of archaeal genomes reveals that the highly conserved ribosomal protein L7 exhibits extensive homology with the eukaryotic 15.5kD protein. Protein binding studies demonstrate that recombinant *Methanoccocus jannaschii* L7 protein binds the box C/D snoRNA core motif with the same specificity and affinity as the eukaryotic 15.5kD protein. Identical to the eukaryotic 15.5kD core protein, archaeal L7 requires a correctly folded box C/D core motif and intact boxes C and D. Mutational analysis demonstrates that critical features of the box C/D core motif essential for 15.5kD binding are also required for L7 interaction. These include stem I which juxtaposes boxes C and D, as well as the sheared G:A pairs and protruded pyrimidine nucleotide of the asymmetric bulge region. The demonstrated presence of L7Ae in the *Haloarcula marismortui* 50S ribosomal subunit, taken with our demonstration of the ability of L7 to bind to the box C/D snoRNA core motif, indicates that this protein serves a dual role in Archaea. L7 functioning as both an sRNP core protein and a ribosomal protein could potentially regulate and coordinate sRNP assembly with ribosome biogenesis.

INTRODUCTION

The small nucleolar RNAs (snoRNAs) are essential for ribosome biogenesis, facilitating the folding and cleavage of the pre-ribosomal RNA transcript and guiding the modification of targeted rRNA nucleotides (1–3). The snoRNAs are classified into two major families based on conserved sequence elements. All box C/D snoRNAs possess conserved nucleotide boxes C and D located at the 5′ and 3′ termini, respectively, whereas the box H/ACA snoRNAs contain a box H in the hinge region and an ACA triplet nucleotide sequence positioned near the 3′ end (4–6). The primary function of both box C/D and H/ACA snoRNAs is to guide nucleotide modification reactions by base pairing with the rRNA precursor and targeting specific bases for modification. The box C/D snoRNAs direct 2′-*O*-methylation of specific ribose sugars whereas the box H/ACA snoRNAs guide the conversion of designated uridine residues to pseudouridine (3,7,8).

The snoRNAs are found in the nucleolus as ribonucleoprotein (RNP) complexes and the common core proteins of each snoRNA family are highly conserved in eukaryotes. Investigations have identified four box H/ACA snoRNP core proteins including the putative pseudouridine synthase Cbf5p (9,10). Core proteins have also been identified for the box C/D snoRNAs. The nucleolar protein fibrillarin (Nop1p in yeast) has long been known to be associated with box C/D snoRNAs (11) and structural analysis has suggested that it is the methylase enzyme (12). Genetic and biochemical experiments have shown that a pair of highly related nucleolar proteins, Nop58p (Nop5p) and Nop56p, are also box C/D snoRNP core proteins (13–17).

Most recently, a fourth core protein designated the 15.5kD protein (Snu13p in yeast) has been identified (18). Eukaryotic 15.5kD snoRNP binds directly to the box C/D core motif and is a structural protein, initiating formation of the box C/D snoRNP core complex. Interestingly, 15.5kD protein also recognizes spliceosomal U4 snRNA and the crystal structure of the U4 snRNA–15.5kD protein complex has been solved (19,20). The 5 + 2 nt asymmetric bulge formed between flanking stems I and II is stabilized by tandem sheared G:A pairs and a protruded pyrimidine nucleotide is important for 15.5kD recognition. This RNA motif adopts a highly ordered tertiary structure including a sharp bend of approximately 65° between the helical axes.

Analysis of archaeal genomes has revealed the presence of box C/D RNAs (termed sRNAs) in both Crenarcheota and Euryarcheota (21–23). The archaeal box C/D sRNAs possess regions of complementarity with rRNA and correspond to sites of nucleotide 2′-*O*-methylation in rRNA. Not surprisingly, box C/D core protein coding sequences have been identified in archaeal genomes and include homologs of both fibrillarin and Nop56/58 (24,25). Notably absent has been a defined homolog of the 15.5kD core protein. However, our database searches reveal a high degree of sequence similarity between eukaryotic 15.5kD protein and the archaeal ribosomal protein L7. We have cloned the *Methanoccocus jannaschii* L7 gene, expressed recombinant protein in *Escherichia coli* and tested the ability

*To whom correspondence should be addressed. Tel: +1 919 515 5803; Fax: +1 919 515 2047; Email: maxwell@bchserver.bch.ncsu.edu

of this ribosomal protein to function as an archaeal homolog of the eukaryotic 15.5kD protein. Electrophoretic mobility-shift analysis demonstrates that L7 binds the snoRNA box C/D core motif with the same affinity and specificity as 15.5kD protein. L7 requires a highly folded RNA, as well as those structural features of the asymmetric bulge critical for 15.5kD binding. The demonstrated presence of L7 in the *Haloarcula marismortui* 50S ribosomal subunit (26) suggests that this protein may serve a dual role as both archaeal ribosomal and sRNP core protein.

MATERIALS AND METHODS

In vitro transcription and labeling of RNA

The U14 box C/D core motif and mutant RNAs were transcribed from linearized plasmid templates pSP64T7U14ΔAV and pSP64T7U14ΔAVΔCD, respectively (27). RNA transcription was carried out using the RiboMAX system (Promega) according to the manufacturer's protocol, at template concentrations of 0.1 μg/μl. RNA transcripts were purified by polyacrylamide gel electrophoresis. Wild-type and mutant box C/D core motif RNAs were also transcribed from short DNA templates generated from PCR amplification of the plasmid pSP64T7U14ΔAV using the DNA oligo pairs listed below. Primer pairs containing the desired mutations were used to produce DNA templates positioned downstream of a T7 promoter site. RNA synthesis from these DNA templates was carried out using the RiboMAX system at template concentrations of 0.2–0.4 μg/μl and RNA transcripts were purified by polyacrylamide gel electrophoresis. Internally radiolabeled RNAs were transcribed from linearized plasmids pSP64T7U14ΔAV and pSP64T7U14ΔAVΔCD in the presence of [α-^{32}P]CTP using T7 RNA polymerase (28). Gel-purified RNA transcripts were radiolabeled at the 5′ end and purified from free nucleotides by Sephadex G-25 spin column chromatography.

Primer pairs for PCR amplification of DNA templates

Wild-type U14 ΔAV snoRNA (A+F); U7c (M+F); U7g (L+F); G8c (C+F); A9c (D+F); U10a (N+F); U10c (U+F); G11c (S+F); C41g (A+T); G11c/C41g (S+T); G43c (A+H); A44c (A+I); stem I-disruption (stem I mut: E+F); stem I-restoration (stem I comp: E+P); stem II-disruption (stem II mut: K+F); stem II-restoration (stem II comp: K+J); Δ bulge-two nucleotide deletion (R+F).

DNA oligonucleotide sequences

(A) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTGATGATGGATTCC-3′; (C) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTCATGATGGATTCC-3′; (D) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTGCTGATGGATTCC-3′; (E) 5′-CTAATACGACTCACTATAGGCCATTCCGTGTGATGATGGATTCC-3′; (F) 5′-ATTCGCTCAGACATCCAAGGAAGGAATTTTGG-3′; (H) 5′-ATTCGCTGAGACATCCAAGGAAGGAATTTTGG-3′; (I) 5′-ATTCGCGCAGACATCCAAGGAAGGAATTTTGG-3′; (J) 5′-ATTCGCTCAGTACTCCAAGGAAGGAATTTTGG-3′; (K) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTGATGTGTGATTCC-3′; (L) 5′-CTAATACGACTCACTATAGGCCATTCGCTGGGATGATGGATTCC-3′; (M) 5′-CTAATACGACTCACTATAGGCCATTCGCTGCGATGATGGATTCC-3′; (N) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTGAAGATGGATTCC-3′; (P) 5′-ATTCCGTCAGACATCCAAGGAAGGAATTTTGG-3′; (R) 5′-CTAATACGACTCACTATAGGCCATTCGCTGATGATGGATTCC-3′; (S) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTGATCATGGATTCC-3′; (T) 5′-ATTCGCTCACACATCCAAGGAAGGAATTTTGG-3′; (U) 5′-CTAATACGACTCACTATAGGCCATTCGCTGTGACGATGGATTCC-3′.

Cloning of *M.jannaschii* L7 and mouse 15.5kD protein genes: expression and purification of recombinant proteins

The mouse 15.5kD protein coding sequence was PCR amplified from I.M.A.G.E. Consortium clone no. 809008 using gene-specific primers. The *M.jannaschii* L7 protein coding sequence was amplified from genomic DNA. Primers possessed terminal *Bam*HI restriction sites and amplified DNA fragments were digested with *Bam*HI restriction endonuclease and inserted into the *Bam*HI restriction site of the pGEX-4T-1 plasmid (Amersham Pharmacia Biotech). *Escherichia coli* DH5α cells were transformed with plasmid constructs and selected clones sequenced to confirm that the correct coding sequences were in frame with the N-terminal glutathione *S*-transferase (GST) tag. Protein expression was accomplished using standard techniques. At 3 h post-induction, harvested cells were disrupted by sonication in buffer D [20 mM HEPES pH 7.0, 100 mM KCl, 3 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM dithiothreitol (DTT), 20% w/v glycerol] and insoluble protein removed from the sonicate by ultracentrifugation at 39 000 *g* for 25 min at 4°C. Soluble sonicate was then applied to glutathione–Sepharose 4B resin (Amersham Pharmacia Biotech) equilibrated in buffer D and bound GST-tagged protein was eluted in buffer D supplemented with 20 mM glutathione. Protein concentrations were determined by the Bradford Assay. N-terminal GST tags were removed by thrombin (Sigma) cleavage overnight. GST cleavage reactions for the 15.5kD recombinant protein were carried out at 4°C at a thrombin concentration of 1 U/50 μg protein. GST cleavage reactions for the L7 recombinant protein were carried out at 25°C at a thrombin concentration of 2 U/50 μg protein.

Thermal denaturation analysis

Thermal denaturation of the U14 box C/D core motif RNA transcripts was carried out using a CARY Varian model 3 UV-Vis spectrophotometer. Briefly, 5–10 μg of RNA, suspended in 400 μl phosphate buffer (10 mM phosphate buffer pH 7.2, 100 mM NaCl) was denatured and reannealed for six cycles at a ramp rate of 1°C/min from 5 to 90°C per cycle. UV absorbance profiles were collected at 260 nm wavelength using a quartz cuvette with a standard 1 cm path length. A_{260} values were normalized and plotted versus temperature.

Protein–RNA interaction analysis

Recombinant 15.5kD and L7 protein binding to the U14 core motif RNA was assessed by electrophoretic mobility-shift analysis as detailed previously (27) with the following modifications: binding reactions contained 0.1 nM ^{32}P-labeled RNA, 2.0 μM recombinant protein (GST–15.5kD, GST–L7 or L7) and assembly was carried out for 1 h at 4°C. For competition studies, non-radiolabeled competitor RNA was added at 4000-fold molar excess of radiolabeled RNA. RNP complexes were

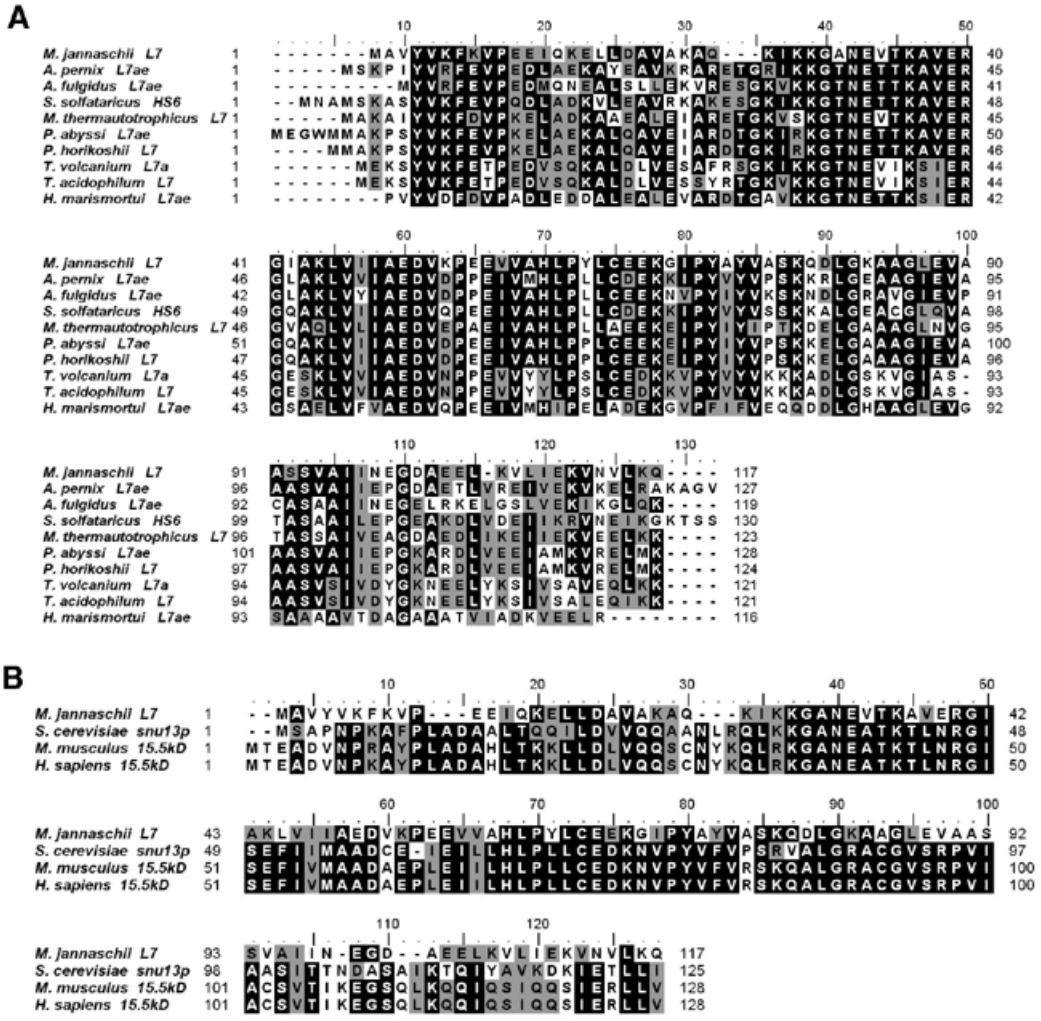

Figure 1. Archaeal ribosomal protein L7 is a homolog of the eukaryotic 15.5kD snoRNP protein. Amino acid sequence alignments of (**A**) archaeal L7 proteins and (**B**) the *M.jannaschii* L7 protein with eukaryotic 15.5kD/Snu13p box C/D snoRNP core proteins. Primary sequence alignments were accomplished using ClustalW (35). Identical residues are indicated by white letters highlighted in black and similar residues are indicated by gray shading. Dashes indicate inserted gaps to aid in the alignment.

resolved on a 4% non-denaturing polyacrylamide gel in TBE buffer containing 8% glycerol and visualized by autoradiography. Equilibrium RNA–protein binding was assessed using a nitrocellulose filter binding assay to measure affinities of recombinant L7 protein for the U14 box C/D core motif RNAs. Binding reactions were carried out as previously described (27) with the following modifications: 100 µl reactions were assembled at 4°C using 5′-labeled RNAs at a concentration of 2.5 nM with the recombinant L7 protein being titrated from 15 pM to 750 nM. Binding reactions were carried out for 1 h at 4°C and then the reaction mix was applied to a nitrocellulose membrane. Prior to loading, membranes were equilibrated in binding buffer (30 mM HEPES pH 7.0, 250 mM KCl, 4.5 mM $MgCl_2$, 1.0 mM DTT, 0.1 mM EDTA, 10% w/v glycerol) and were washed twice after sample application with binding buffer. Membranes were then dried and bound radioactivity was visualized using a Molecular Dynamics Model 425F PhosphorImager and quantified using the ImageQuant software v3.3.

RESULTS

Archaeal ribosomal protein L7 is a functional homolog of eukaryotic 15.5kD snoRNP protein and binds the box C/D snoRNA core motif

Genes for box C/D sRNAs and homologs for the core proteins fibrillarin and Nop56/58 have been identified in numerous archaeal genomes (21,24,25). These findings suggest that a gene(s) encoding a homolog of the eukaryotic 15.5kD snoRNP core protein is also present. However, genome sequence analysis of archaeal organisms has not revealed an obvious 15.5kD homolog. The lone archaeal coding sequence that exhibits significant sequence homology with this eukaryotic snoRNP core protein is the well conserved ribosomal protein L7 (29) (Fig. 1A). Indeed, the *M.jannaschii* L7 sequence is 33% identical/60% similar to human 15.5kD protein and 36% identical/58% similar to yeast Snu13p (Fig. 1B).

Based upon the sequence similarity of the eukaryotic 15.5kD and archaeal L7 proteins, we reasoned that L7 could function

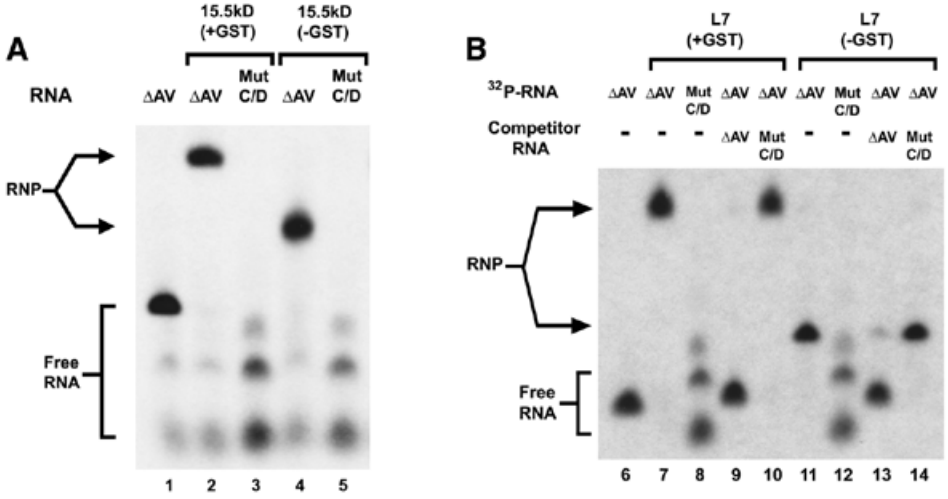

Figure 2. Archaeal ribosomal protein L7 binds the box C/D snoRNA core motif. (**A**) Radiolabeled, wild-type mouse U14 box C/D core motif (ΔAV construct) and mutant U14 box C/D RNA (mut C/D with altered box C and D sequences) were incubated with purified mouse 15.5kD protein (+/– GST tag) and RNP formation assessed by electrophoretic mobility-shift analysis (lanes 1–5). The relative positions of free RNA and assembled RNP complex are indicated at the side. (**B**) Radiolabeled, wild-type mouse U14 box C/D core motif and mutant U14 box C/D RNA were incubated with purified *M.jannaschii* L7 protein (+/– GST tag) in the presence or absence of excess, non-radiolabeled competitor RNA as designated (lanes 6–14). The relative positions of free RNA and assembled RNP complex are indicated at the side. Note: wild-type U14 RNA used for 15.5kD binding possesses a 12 nt extension of its 3′ terminus and thus migrates more slowly on the gel than wild-type U14 RNA used for L7 binding. Also, free mut C/D RNA migrates on the polyacrylamide gel in two different RNA conformations due to altered box C and D sequences.

as a homolog of the 15.5kD protein and bind the box C/D sRNAs. To test this hypothesis, we assessed the ability of recombinant *M.jannaschii* L7 protein to recognize the box C/D core motif and form an RNP complex. The RNA used for these electrophoretic mobility-shift analyses was the box C/D core motif derived from mouse U14 snoRNA consisting of base paired 5′ and 3′ termini with flanking boxes C and D, respectively. This RNA substrate also possesses the rRNA-complementary sequence immediately upstream of box D but lacks the internal box C/D motif with associated rRNA-complementary sequence as well as the adjacent U14 variable region. Recombinant mouse 15.5kD protein (+/– GST tag) binds this box C/D RNA and mutagenesis of boxes C and D demonstrates the essentiality of these conserved nucleotide sequences for protein recognition (Fig. 2A). Similarly, the *M.jannaschii* L7 protein (+/– GST tag) binds the U14 box C/D core motif with the same specificity (Fig. 2B). Competition experiments with non-radiolabeled U14 snoRNA also demonstrate the requirement of boxes C and D for L7 binding. Addition of excess non-radiolabeled wild-type box C/D RNA disrupts RNP formation while added RNA possessing point mutations in boxes C and D is not an effective competitor.

Archaeal L7 protein binds the box C/D core motif with the same affinity as eukaryotic 15.5kD protein

The snoRNA box C/D core motif can be folded into a secondary structure consisting of helical stems I and II flanking an asymmetric 5 + 2 nt bulge (Fig. 3A). This structure is based upon the crystal structure of the 15.5kD protein-binding site observed in the U4 snRNA–15.5kD RNP complex (20) which has an essentially identical box C/D-like motif (18). Eukaryotic 15.5kD protein binds the snoRNA box C/D core motif with high affinity and a K_d of 8 nM has been estimated from electrophoretic mobility-shift analysis of the 15.5kD–box C/D core motif complex (18). To compare the binding affinity of ribosomal protein L7 with the 15.5kD protein, the K_d of the L7–box C/D core motif was determined in equilibrium binding analysis using a filter-binding assay. L7 protein binds the box C/D core motif with the same affinity as the 15.5kD protein with a K_d of 5 nM (Table 1). The mouse 15.5kD protein exhibits a similar K_d value of 19 nM when its binding to the box C/D core motif is assessed using the filter binding assay (data not shown). However, we believe the determined K_d values for both 15.5kD and L7 binding are underestimates of their affinity for the box C/D core motif. Both the electrophoretic mobility-shift and filter-binding assays are limited in determining an actual binding constant when analyzing RNA-binding proteins that exhibit such high affinity for the RNA substrate. The concentrations of radiolabeled RNA in our equilibrium RNA–protein binding assay are nearly equivalent to the L7 protein concentrations at the determined K_d value of 5 nM. Therefore, the filter-binding assay is being carried out under stoichiometric conditions rather than under conditions of limiting RNA concentrations. Consistent with stoichiometric conditions, the binding curve for RNP formation exhibits a steep shape (data not shown). Unfortunately, the exceptionally high affinity of L7 for the box C/D core motif precludes the possibility of measuring radioactively labeled RNP formation at sufficiently low RNA concentrations where small amounts of RNA are bound to protein. With sufficiently dilute RNA concentrations, it is impossible to detect RNP formation since the L7–box C/D core motif complex contains too little radioactive RNA to detect. Therefore, the nanomolar K_d values determined for both 15.5kD (electrophoretic mobility shift) and L7 (filter-binding assay) binding are most likely upper

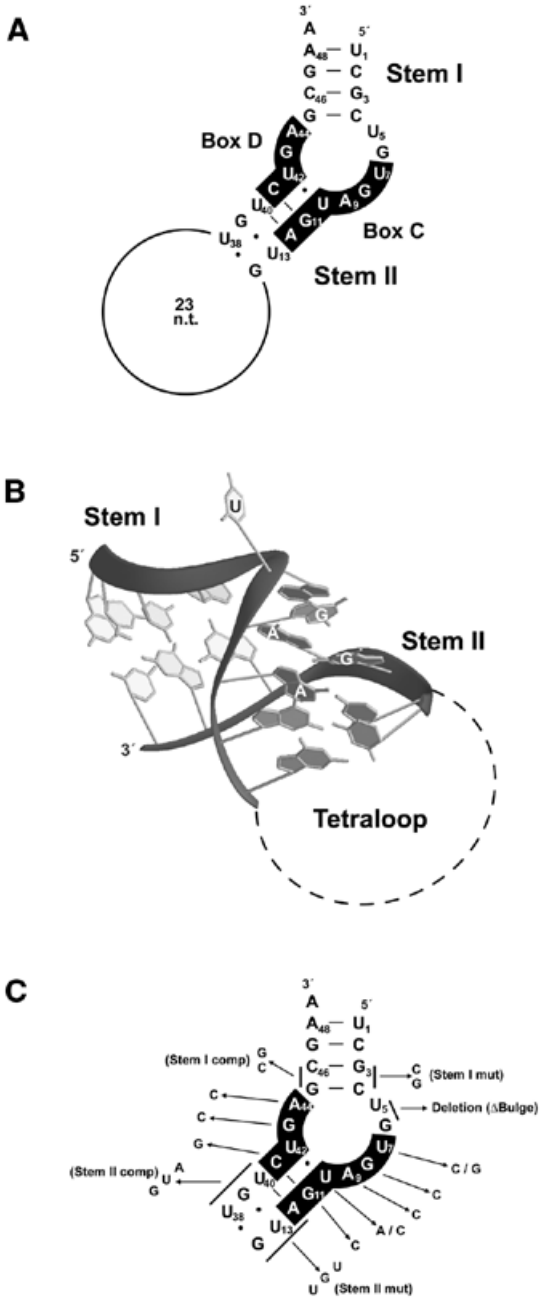

Figure 3. Folded structure of the box C/D snoRNA core motif. (**A**) Secondary structure of the box C/D snoRNA core motif. Folding of the box C/D core motif is based upon the structure of the 15.5kD protein binding site (18). Conserved boxes C and D are indicated by white letters on black backgrounds. (**B**) Folded tertiary structure of the 15.5kD protein-binding site on U4 snRNA as determined in the crystal structure of the 15.5kD–U4 snRNP complex (20). (**C**) Specific box C/D core motif mutations created to study *M.jannaschii* L7 binding.

limits of binding affinity and the actual K_d values are probably an order of magnitude or more lower.

L7 also exhibits low affinity, non-specific binding with ribosomal RNA (Table 1). This is typical of many RNA-binding proteins that interact with RNA in a sequence-independent manner via charge–charge interactions. An estimated K_d value of ≈400 nM therefore establishes a relative baseline of non-specific interaction to which L7 binding to the box C/D core motif mutant RNAs is compared (see below).

Table 1. Relative binding affinity of L7 protein for selected box C/D snoRNA mutants

RNA	K_d value (nM)
Wild-type	≈5
Non-specific rRNA	≈400
Stem I region	
Stem I mutant	≈400
Stem I compensatory	53
Stem II region	
Stem II mutant	51
Stem II compensatory	67
C41g	137
G11c	122
G11c / C41g	≈5
U10a	≈5
U10c	115
Asymmetric bulge region	
U7c	73
U7g	≈400
G8c	≈400
ΔBulge	118

The relative affinities of L7 protein for various box C/D snoRNA mutants were determined by RNA–protein equilibrium binding as described in Materials and Methods. RNP complex formation was measured and plotted versus protein concentration to obtain a binding curve from which a K_d was determined for each RNA.

L7 requires the same box C/D sequence and structural elements for binding as does the 15.5kD protein

X-ray crystallographic analysis of the 15.5kD–U4 snRNP complex has revealed that the 15.5kD protein requires specific sequence and structural elements of the folded RNA for protein binding (20). In particular, the juxtaposition of box C and D-like sequences by flanking stems I and/or II to form the asymmetric bulge is critical. This 5 + 2 nt asymmetric bulge establishes the bulk of the protein binding site where tandem, sheared G:A pairs and a protruded pyrimidine nucleotide contact the bound protein (Fig. 3B). These same structural features are presumed to be present in the snoRNA box C/D core motif and the observed requirements for box C/D snoRNP assembly *in vivo* are consistent with this RNA folded structure (4). We therefore assessed the importance of specific RNA elements of the box C/D core motif for L7 protein binding to determine if this archaeal ribosomal protein interacts with the RNA in the same manner as the eukaryotic 15.5kD protein.

The folded structure of box C/D core motif RNA and various mutants was monitored in thermal denaturation/renaturation analysis (Fig. 4) while RNP assembly (Fig. 5) and L7 affinity

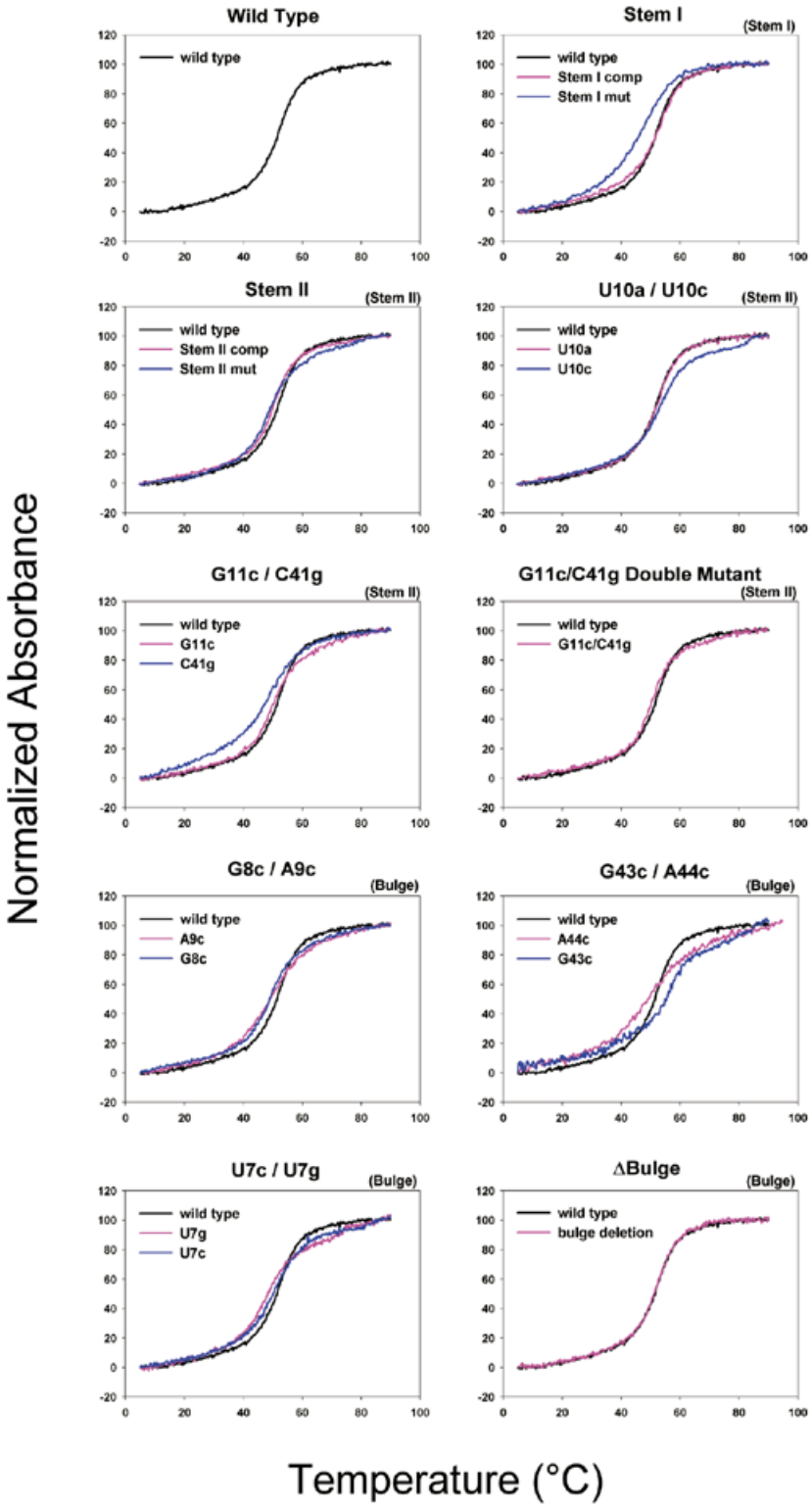

Figure 4. The box C/D core motif exhibits a highly ordered RNA structure. Wild-type and mutant box C/D core motif snoRNAs were subjected to thermal denaturation and renaturation and UV absorbance values at 260 nm were collected over a temperature range of 5–90°C as described in Materials and Methods. Absorbance values were normalized and plotted versus temperature. Specific nucleotide alterations for the individual box C/D core motif mutants indicated in each panel are detailed in Figure 3C. All RNAs were diluted and subjected to additional thermal denaturation (data not shown) to ensure that a unimolecular event was being measured.

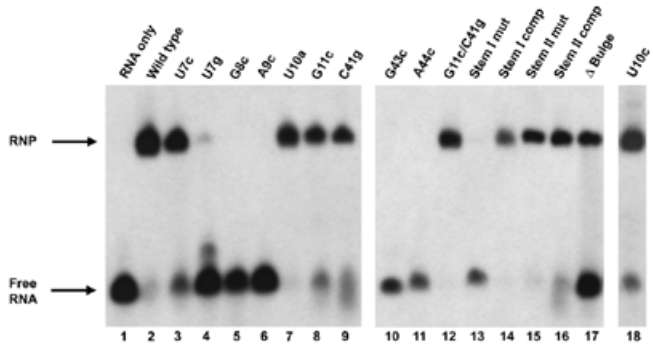

Figure 5. L7 requires specific sequence and structural elements of the box C/D core motif for RNA–protein interaction. Radiolabeled wild-type and mutant box C/D core motif RNAs were incubated with purified *M.jannaschii* L7 protein. L7 binding to the various RNAs was then assessed by electrophoretic mobility-shift analysis. The positions of free radiolabeled RNA and assembled RNP complex are indicated at the side. Specific box C/D snoRNA core motif mutants are indicated above each lane and detailed in Figure 3C.

(Table 1) were determined in electrophoretic mobility-shift and equilibrium binding analyses, respectively. Mutations were created in stem I, stem II and the asymmetric bulge (Fig. 3C) to assess the contribution of each region for both RNA folding and L7 binding. In contrast to the box C/D core motif RNA, the mutant RNAs exhibited lower binding affinities for L7 allowing radiolabeled RNP formation to be measured over a range of protein concentrations sufficient to determine equilibrium binding constants. Calculated K_d values provide relative L7 binding strengths with respect to the upper limit of 5 nM estimated for the box C/D core motif RNA and the baseline, non-specific protein binding to ribosomal RNA at ≈400 nM.

Stem I or external stem. Thermal denaturation of the folded box C/D snoRNA core motif results in a sharp melting profile (Fig. 4) that is typical of a highly structured RNA. The large increase and decrease in hyperchromicity over a narrow temperature range around the T_m is characteristic of an RNA that folds in a highly cooperative manner. Such a melting profile is consistent with the proposed folded structure of the box C/D motif as illustrated for U4 snRNA in Figure 3B. Mutating the 5′ terminal sequence of U14 to disrupt stem I base pairing (stem I mut) significantly alters the folded structure of the box C/D core motif as evidenced by the reduction in cooperativity during thermal denaturation (Fig. 4). As a consequence, L7 binding is disrupted (Fig. 5) and affinity for the core motif is reduced to a level equivalent to non-specific association (Table 1). Compensatory changes in the 3′ terminal sequence that re-establish stem I base pairing (stem I comp) restore the wild-type RNA melting profile and protein binding, albeit at a lower affinity. Thus, folding of the box C/D core motif for L7 binding is dependent on the formation of this RNA helix.

Stem II or internal stem. Two types of mutants were constructed to examine the importance of base pairing in stem II for RNA folding and L7 binding. Mutations included disruption of the base pairing between conserved nucleotides of boxes C and D as well as disruption of hydrogen bonding of bases in the extended stem II structure. Elimination of base pairing between nucleotides extending the stem II helix (stem II mut) has only slight effects upon the RNA motif melting profile and L7 affinity. Restoration of hydrogen bonding through compensatory mutations did not completely restore the original melting profile or L7 affinity. The fact that both RNA mutants substitute the highly conserved A12 nucleotide of box C suggests the potential importance of an adenosine residue at this position in box C for core motif structure.

More pronounced effects on both the melting profile and L7 affinity are observed when nucleotides of boxes C and D that establish stem II structure are altered. Mutation of either G11 in box C or C41 in box D alters the structure of the folded core motif as evidenced by changes in the RNA melting profiles. L7 binding to the core motif is also reduced. Restoration of base pairing via compensatory mutations restores the melting profile and L7 affinity indicating that base pairing of these two bases is important for L7 interaction. Thus, hydrogen bonding of this nucleotide pair within the core motif contributes to the stability of the protein–RNA interaction, but specific nucleotides at these particular positions appear less important for L7 binding. Interestingly, nucleotides G11 and C41 are conserved in boxes C and D, respectively, at >97% of all box C/D RNAs in eukaryotes and Archaea (21,23). Covariation of this G:C base pair is not evident, suggesting that these nucleotides positioned specifically within boxes C and D may be important for the binding of other proteins. Recent work has indeed indicated that this region of the snoRNA core motif is important for fibrillarin binding (30).

Equally interesting is the U10:U42 base pair of stem II that is adjacent to the asymmetric bulge. Mutant U10a replaces uridine with adenosine and preserves base pairing. This nucleotide substitution has no effect upon the melting profile or L7 affinity for the core motif RNA (Fig. 5 and Table 1). In contrast, substitution of the uridine with a cytidine alters core motif folded structure, as evidenced by changes in the melting profile, and L7 binds with reduced affinity. These results suggest that hydrogen bonding between these two nucleotide positions is important. U10 is >90% conserved in eukaryotic snoRNAs but is <80% conserved in archaeal sRNAs (4,21). Covariation analysis of archaeal sRNAs indicates that hydrogen bonding between this nucleotide pair is most often maintained with alternative nucleotide pairs (data not shown).

Asymmetric bulge. Formation of the 5 + 2 nt asymmetric bulge with tandem, sheared G:A pairs and protruded pyrimidine nucleotide (Fig. 3B) is critical for 15.5kD binding to U4 snRNA (20) and presumably the snoRNA box C/D core motif. If L7 is a functional homolog of 15.5kD, then the sheared G:A pairs and protruded nucleotide should also be critical for L7 recognition. Indeed, mutations within the asymmetric bulge of the core motif have a profound impact on L7 binding. Point mutations in each of the four G/A residues that constitute the two sheared base pairs (mutants G8, A9, G43 and A44) substantially alter RNA folded structure as seen in the melting profiles with alteration of G43 and A44 in box D having a particularly severe effect. All four nucleotides are critical for protein binding as each mutation completely disrupts L7 binding. Analysis of the G8c point mutation demonstrates that the L7 binding affinity is reduced to non-specific levels.

The protruded nucleotide, which makes important base-specific contacts with the 15.5kD protein, is also critical. Mutations made at this critical nucleotide include substitution

of both pyrimidine (mutant U7c) and purine (mutant U7g) nucleotides. These substitutions affect the melting profile of the folded RNA similarly but have significantly different effects upon L7 binding. Replacement of uridine with another pyrimidine affects RNP formation and L7 affinity only slightly. In contrast, replacement of the wild-type pyrimidine nucleotide with the purine guanosine eliminates L7 binding and reduces L7 affinity to non-specific levels. These observations are consistent with the central role of U7 in sequence-specific protein recognition (see Discussion).

Finally, the two bases immediately upstream of box C on the 5′ side of the bulge region were removed and L7 binding was assessed. This 2 nt deletion completely disrupts intronic snoRNA processing in eukaryotes, suggesting that the loss of processing reflects a disruption in RNP complex formation (4). Surprisingly, deletion of these nucleotides has no effect upon the RNA melting profile, but does impact RNP formation and L7 binding affinity. The essentiality of these 2 nt for snoRNP formation in *Xenopus* oocyte nuclei suggests that while the 15.5kD protein may bind at reduced efficiency, the binding of additional snoRNP core proteins may be disrupted and snoRNP assembly is incomplete.

DISCUSSION

While genes for box C/D sRNAs and homologs for the sRNP core proteins fibrillarin and Nop56/58 have been identified in archaeal genomes, a defined homolog for the 15.5kD protein has been notably absent. Database analysis of available archaeal sequences reveals the strong sequence similarity of ribosomal protein L7 with mouse and human 15.5kD protein as well as yeast Snu13p. Other eukaryotic proteins exhibiting homology to archaeal L7 include eukaryotic ribosomal proteins L7, L30, S12 and the H/ACA snoRNP protein Nhp2p (19). Our studies now demonstrate that L7 binds the snoRNA box C/D core motif with the same specificity and affinity as the eukaryotic 15.5kD snoRNP core protein. Because of the high affinity of this RNA-binding protein for the box C/D core motif, it is difficult to determine an accurate K_d value. Our estimate in the nanomolar range is consistent with that made for eukaryotic 15.5kD protein and both values most likely represent lower estimates of the true strength of these RNA–protein interactions.

The highly conserved L7 gene was first identified from the sequencing of various archaeal genomes and designated L7 because of its relatedness to ribosomal proteins (29). Recent crystallization of the *H.marismortui* ribosome has demonstrated the presence of this protein (L7Ae) in the 50S large subunit, thus verifying its function as an archaeal ribosomal protein (26). An experimental demonstration of ribosomal protein L7's presence in the archaeal sRNP complex awaits further characterization of the sRNP core proteins. However, the ability of L7 to bind the box C/D core motif indicates that L7 serves a dual role in Archaea as both ribosomal and sRNP core proteins. The binding of L7 to both 23S rRNA and the sRNAs could potentially play important regulatory roles, coordinating ribosome assembly with sRNP assembly and sRNA stabilization. Strikingly, 15.5kD protein also serves a dual role in eukaryotes, binding to two different RNA substrates to function as both a snoRNP core protein and a U4-specific snRNP protein (18,19). Again, a regulatory role to coordinate snoRNP biogenesis with pre-mRNA processing has been suggested (18). The fact that this protein homolog binds both identical (snoRNA/sRNA) and unique (rRNA versus U4 snRNA) substrates in Archaea versus Eukaryotes is noteworthy, suggesting a possible evolution of protein function (see below).

L7 binds to an RNA structural motif recently defined as a kink-turn or K-turn (31). This RNA motif is widespread in evolution and is found in Eubacteria as well as Archaea and Eukaryotes (32). The consensus structure of the K-turn possesses two RNA helices (canonical and non-canonical stems corresponding to stems I and II, respectively) which flank an asymmetric bulge containing tandem, sheared G:A pairs. The G:A pairs play a central role in establishing the folded structure of this motif and bending the RNA at a sharp angle with a single nucleotide protruding from the asymmetric bulge. RNAs which possess demonstrated or presumed K-turn motifs include ribosomal RNAs, box C/D snoRNAs, U4 snRNA, human RNase P and L30 mRNA. Six K-turns have been described in the *H.marismortui* large ribosomal subunit, including KT-15 which binds the L7Ae protein (31). A K-turn motif has been revealed with the crystal structure of the U4–15.5kD complex and it is presumed that the box C/D motif exhibits the same RNA folded structure (20). The close packing of the canonical (stem I) and non-canonical (stem II) stems observed in all K-turns is important for stabilizing this motif's folded structure and is consistent with the highly cooperative melting profile observed for the box C/D core motif in our studies.

Individual K-turn motifs often exhibit unique structural features including variations in asymmetric loop size and protruded nucleotide identity while still exhibiting the overall features of the motif. The L7Ae binding site in 23S rRNA (KT-15) and the U4 (presumably box C/D core motif) share similarities in folded structure (Fig. 6A and B). Both exhibit helical regions flanking an asymmetric bulge and the protruded pyrimidine nucleotide is critical for L7 binding. In particular, Klein *et al*. noted that the bulged pyrimidine nucleotide of KT-15 fits the hydrophobic binding pocket of L7 and a uridine nucleotide afforded better interactions with critical amino acids than a cytidine residue (31). In contrast, a purine ring was too large for insertion into this hydrophobic pocket. Our mutational studies of base substitution at this protruded nucleotide within the box C/D core motif are consistent with their structural predictions of L7 binding at this nucleotide position in the KT-15 turn.

While similar in overall structure, the folded K-turns of U4 snRNA and KT-15 exhibit distinct differences when compared. Most notably, the U4 and box C/D core motif present two tandem, sheared G:A pairs characteristic of most K-turns (Fig. 6A). In contrast, the KT-15 K-turn determined from the *H.marismortui* 50S ribosome subunit crystal structure reveals a distinctly different nucleotide configuration in this region (Fig. 6B). Specifically, a single sheared G:A pair is followed by a G:U:A base triple where the adenine residue or third nucleotide is derived from the 'short side' of the asymmetric loop. This nucleotide arrangement causes a more pronounced bend in the K-turn motif and results in a sharp kink in the phosphate backbone of both RNA strands. These differences exhibited by individual K-turn motifs typically account for the recognition of the structurally similar K-turns by different

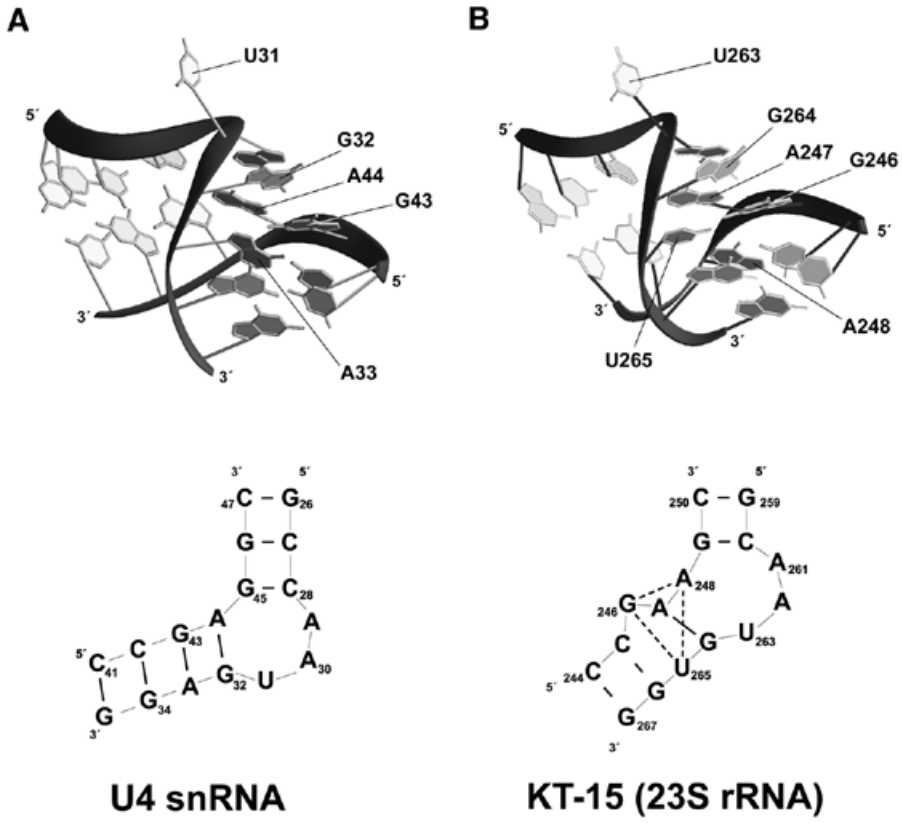

Figure 6. Comparison of the U4 snRNA 5′ stem loop and the KT-15 K-turn of 23S rRNA. (**A**) Structure of the U4 snRNA 5′ stem loop. The tertiary structure was determined from the crystal structure of the U4–15.5kD complex (20). The sheared G:A pairs and protruded nucleotide are indicated. Shown below is the folded secondary structure. (**B**) Structure of the KT-15 K-turn of *H.marismortui* 23S rRNA. The tertiary structure was determined from the crystal structure of the *H.marismortui* 50S ribosome subunit (26). The sheared G:A pair and base triple are indicated. Shown below is the folded secondary structure of KT-15. Dashed lines indicate those nucleotides involved in the base triple.

proteins. For example, the six K-turns of the *H.marismortui* 50S ribosome subunit bind nine different ribosomal proteins. However, despite these structural differences between the KT-15 and the U4 (box C/D core motif) K-turns, each is recognized by L7. Strikingly, L7 binds both motifs utilizing identical regions of the K-turn, including the NC or stem II region which exhibits differences in sheared G:A pair configuration.

Examination of the L7 crystal structure reveals that this ribosomal protein folds into the same folded conformation as eukaryotic 15.5kD protein and utilizes the same structural elements to bind their RNA substrates. The crystal structures of both eukaryotic 15.5kD (human) and archaeal L7 (*H.marismortui*) bound to their respective RNA substrates demonstrate essentially identical folded protein structures (Fig. 7). Overlay of the two proteins reveal that those residues critical for binding to U4 snRNA and 23S rRNA are found in corresponding positions within each folded protein. Highly conserved amino acid residues within α-helix 2 of both proteins contact the region of sheared G:A pairs while amino acid residues of the α-helix 4-β sheet 4 loops form a hydrophobic pocket and contact the RNAs in a base-specific manner. Two amino acids (E61/K86 in 15.5kD and Q55/Q80 in L7Ae) interact with the nucleotide of the asymmetric bulge that is rotated away from the motif. Not surprisingly, modeling of the *M.jannaschii* L7 protein reveals an identical folding pattern with these same residues critical for RNA binding similarly positioned in the folded protein structure (data not shown).

Particularly striking is the homology exhibited between snoRNP and ribosomal proteins and the similar RNP structures formed between these proteins and their cognate RNAs. Archaeal protein L7 and eukaryotic 15.5kD protein share sequence similarity with ribosomal protein L30, H/ACA snoRNP core protein Nhp2p and ribosomal protein S12 (19). All are members of a family of RNA-binding proteins originally identified by Koonin *et al*. (33). The homology exhibited by these ribosomal proteins and snoRNP core proteins indicate a common ancestral coding sequence. The 15.5kD–box C/D snoRNA and L7Ae–rRNA complexes are also strikingly similar in RNP structure, perhaps indicating a common ancestral RNP structure. The similarity in RNP organization suggests to us that the snoRNPs and sRNPs could well have their evolutionary origins in primordial ribosomes. Such a progenitor ribosomal RNP structure could have evolved to acquire *cis*-acting sequences complementary to rRNA that are important for RNA folding. Additional proteins bound to the RNP motif might assist in rRNA precursor folding and/or bring new function to the complex such as nucleotide modification activities. With time, these RNP motifs would have become independent from the ribosome itself, yet still retained their critical functions in ribosome biogenesis. The occurrence of two box C/D motifs within the intron of the *Archaeoglobus fulgidus*

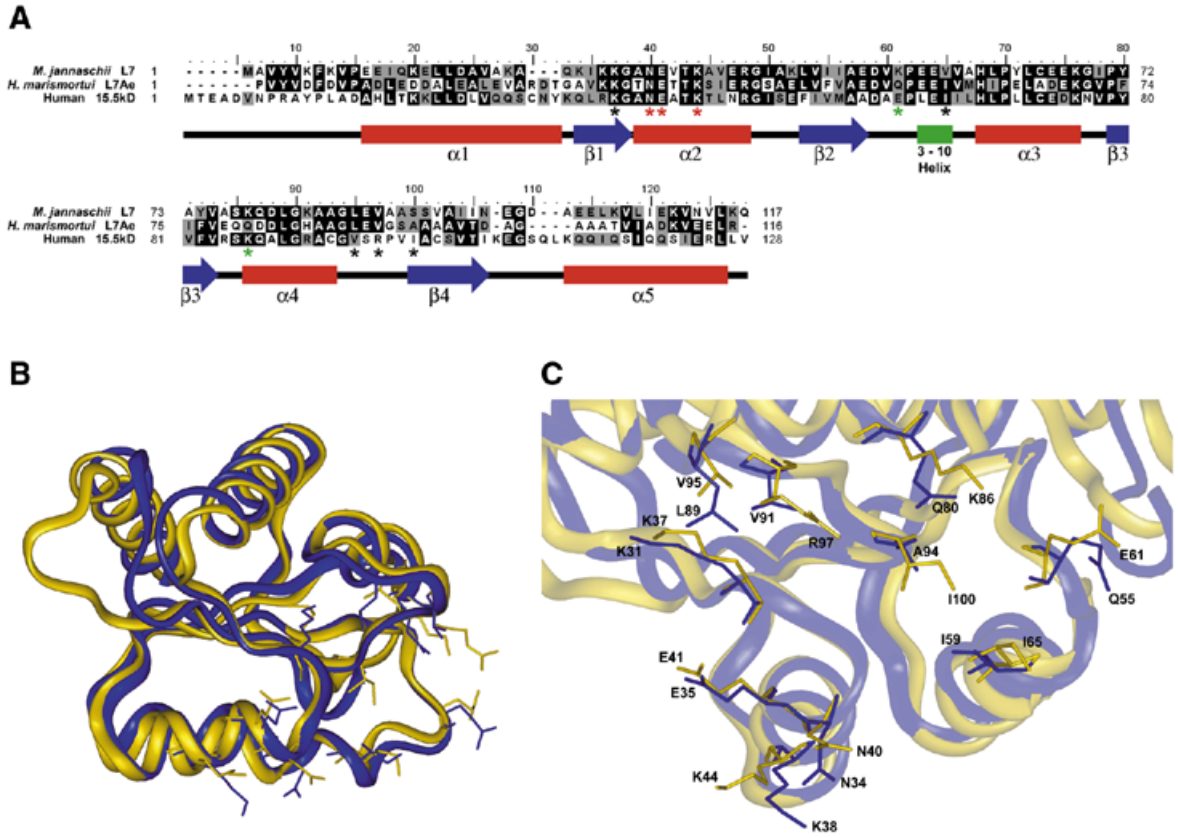

Figure 7. Comparison of human 15.5kD, *H.marismortui* L7Ae and *M.jannaschii* L7 proteins. (**A**) Amino acid sequence alignment of the *M.jannaschii* L7 protein with the human 15.5kD box C/D snoRNP protein and the *H.marismortui* L7Ae ribosomal protein. Alignments were accomplished using ClustalW (35). Identical residues are indicated by white letters on a black background and similar residues are shaded in gray. Shown below the sequence alignment is a schematic representation of the 15.5kD/L7Ae secondary structure. α-Helices are indicated by red bars and β-sheets are indicated by blue arrows with the short 3-10 helix shown in green. Secondary structure elements were assigned according to the Protein Data Bank (PDB) file entry. Asterisks denote the position of residues making sequence specific contacts with nucleotide bases in the bound RNA according to the U4–15.5kD crystal structure. Black asterisks indicate residues making hydrophobic contacts, red asterisks indicate residues contacting the conserved G:A pairs in the kink-turn motif, with green asterisks indicating residues contacting the protruding nucleotide. (**B**) Overlay of the 15.5kD protein (gold) and *H.marismortui* L7Ae protein (blue) tertiary structures. Coordinates for the 15.5kD and L7Ae proteins were obtained from the PDB (entries 1E7K and 1JJF, respectively) and the overlay was accomplished using the SWISS-model program (36). The side chains of amino acid residues denoted by asterisks in (A) are shown. (**C**) Close-up view of the RNA binding pocket from the overlayed proteins shown in (B). Amino acids are numbered according to their respective sequences and colors correspond to (B). (The presented graphic images were generated using the InsightII software module of Molecular Simulations, Inc., release 2000.)

Trp-tRNA that apparently function in *cis* to guide 2′-*O*-methylation of specific nucleotides in the unspliced tRNA is consistent with such a scenario (22). A similar model for the evolution of the snRNP splicing complexes has been previously suggested where folded elements of group II introns evolved to become independent snRNP complexes while maintaining their functional roles in pre-mRNA splicing (34). Thus, present day RNP complexes critical for post-transcriptional RNA processing and modification may have their origins in *cis*-acting RNP elements of the ancient RNA world.

ACKNOWLEDGEMENTS

We would like to thank J. Liu for technical assistance, R. Guenther and C. Yarian for assistance in the RNA thermal denaturation studies, C. Clark for his advice in the analysis of L7 binding data, J. Brown for *M.jannaschii* genomic DNA and help with covariation analysis and M. J. Fournier and J. Brown for critical comments of the manuscript. This work was supported by NSF Grant MCB9727945 to E.S.M.

REFERENCES

1. Maxwell,E.S. and Fournier,M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **35**, 897–934.
2. Kiss,T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617–3622.
3. Smith,C. and Steitz,J. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669–672.
4. Xia,L., Watkins,N. and Maxwell,E.S. (1997) Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA*, **3**, 17–26.
5. Balakin,A.G., Smith,L. and Fournier,M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823–834.
6. Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997) The family of box H/ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941–956.

7. Tollervey,D. (1996) Small nucleolar RNAs guide ribosomal RNA methylation. *Science*, **273**, 1056–1057.
8. Bachellerie,J.-P., Cavaille,J. and Qu,L.-H. (2000) Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNA guides revisited. In Garret,R., Douthwaite,R., Liljas,A., Matheson,A., Moore,P. and Noller,H. (eds), *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*. ASM Press, Washington, DC, pp. 191–203.
9. Watkins,N., Gottschalk,A., Neubauer,G., Kastner,B., Fabrizio,P., Mann,M. and Luhrmann,R. (1998) Cbf5p, a potential pseudouridine synthase and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all box H/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549–1568.
10. Henras,A., Henry,H., Bousquet-Antonelli,C., Noaillac-Depeyre,J., Gelunge,J.-P. and Caizergues-Ferrer,M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.*, **17**, 7078–7090.
11. Tyc,K. and Steitz,J. (1989) U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J.*, **8**, 3113–3119.
12. Wang,H., Boisvert,D., Kim,K., Kim,R. and Kim,S.-H. (2000) Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J.*, **19**, 7–17.
13. Newman,D., Kuhn,J., Shanab,G. and Maxwell,E.S. (2000) Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA*, **6**, 861–879.
14. Lafontaine,D. and Tollervey,D. (1999) Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA*, **5**, 455–467.
15. Lafontaine,D. and Tollervey,D. (2000) Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol. Cell. Biol.*, **20**, 2650–2659.
16. Caffarelli,E., Losito,M., Giorgi,C., Fatica,A. and Bozzoni,I. (1998) *In vivo* identification of nuclear factors interacting with the conserved elements of box C/D small nucleolar RNAs. *Mol. Cell. Biol.*, **18**, 1023–1028.
17. Lyman,S., Gerace,L. and Baserga,S. (1999) Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins. *RNA*, **5**, 1597–1604.
18. Watkins,N., Segault,V., Charpentier,B., Nottrott,S., Fabrizio,P., Bachi,A., Wilm,M., Roshbash,M., Branlant,C. and Luhrmann,R. (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457–466.
19. Nottrott,S., Hartmuth,K., Fabrizio,P., Urlaub,H., Vidovic,I., Ficner,R. and Luhrmann,R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5′ stem-loop of U4 snRNA. *EMBO J.*, **18**, 6119–6133.
20. Vidovic,I., Nottrott,S., Hartmuth,K., Luhrmann,R. and Ficner,R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331–1342.
21. Omer,A., Lowe,T., Russell,A., Ebhardt,H., Eddy,S. and Dennis,P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517–522.
22. Dennis,P., Omer,A. and Lowe,T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.*, **40**, 509–519.
23. Gaspin,C., Cavaille,J., Erauso,G. and Bachellerie,J.-P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.*, **297**, 895–906.
24. Amiri,K. (1994) Fibrillarin-like proteins occur in the domain *Archaea*. *J. Bacteriol.*, **176**, 2124–2127.
25. Lafontaine,D. and Tollervey,D. (1998) Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.*, **23**, 383–388.
26. Ban,N., Nissan,P., Hansen,J., Moore,P. and Steitz,T. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
27. Watkins,N., Newman,D., Kuhn,J. and Maxwell,E.S. (1998) *In vitro* assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA*, **4**, 582–593.
28. Watkins,N., Leverette,R., Xia,L., Andrews,M. and Maxwell,E.S. (1996) Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA*, **2**, 118–133.
29. Bult,C., White,O., Olsen,G., Zhou,L., Fleischmann,R., Sutton,G., Blake,J., Fitzgerald,L., Clayton,R., Gocayne,J. *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1072.
30. Fatica,A., Galardi,S., Altieri,F. and Bozzonni,I. (2000) Fibrillarin binds directly and specifically to U16 box C/D snoRNA. *RNA*, **6**, 88–95.
31. Klein,D., Schmeing,T., Moore,P. and Steitz,T. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
32. Winkler,W., Grundy,F., Murphy,B. and Henkin,T. (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA and eukaryotic RNAs. *RNA*, **7**, 1165–1172.
33. Koonin,E., Bork,P. and Sander,C. (1994) A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res.*, **22**, 2166–2167.
34. Sharp,P. (1991) Perspective: five easy pieces. *Science*, **254**, 663.
35. Thompson,J., Higgins,D. and Gibson,T. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **11**, 4673–4680.
36. Peitsch,M., (1996) ProMod and Swiss-Model: internet-based tools for automated comparative protein modeling. *Biochem. Soc. Trans*, **24**, 274–279.