RNA-GUIDED RNA MODIFICATION IN ARCHAEA: DYNAMIC ASSEMBLY OF ACTIVE H/ACA RNA/PROTEIN COMPLEXES AND A POTENTIAL ROLE OF SM4 PROTEIN IN RNA MODIFICATION

by

OSAMA A YOUSSEF

(Under the Direction of Michael P. Terns and Rebecca M. Terns)

ABSTRACT

Post-transcriptional RNA modifications play important roles in all organisms. In eukaryotes and archaea, RNA modifications are brought about by the action of two large families of RNA/protein complexes known as the H/ACA and C/D RNPs. Specifically, H/ACA and C/D RNPs catalyze the conversion of uridine to pseudouridine and 2'-*O*-ribose methylation in cellular RNAs, respectively. The first part of this thesis describes the insights gained into the organization of archaeal H/ACA RNPs. Archaeal H/ACA RNPs consist of a guide RNA and four proteins: Cbf5, Gar1, Nop10, and L7Ae. The guide RNA is required to identify the target RNA nucleotide by base-pairing. We used biochemical assays to detail the RNA/protein and protein/protein interactions that control the assembly and function of these complexes. Our results showed that not only L7Ae but also Cbf5 binds directly to H/ACA guide RNAs. The data presented here revealed that Cbf5 has another important role in addition to its function as the pseudouridine synthase. Cbf5 is required for the recruitment of Gar1 and Nop10 to the guide RNA. We found that all four proteins are required for the efficient catalytic activity of the complex. Unlike the C/D RNPs, H/ACA RNPs do not require L7Ae for the recruitment of the

other three proteins to the guide RNA. However, our RNA footprinting data showed that L7Ae plays a key role in the structural organization of H/ACA RNPs.

The Sm and Sm-like proteins are a large family of evolutionarily ancient RNA binding proteins that form ring-like structures and mediate important cellular activities. Specific functions for any archaeal Sm-like proteins are unknown. In the second part of this thesis, we report a novel archaeal Sm-like protein (Sm4) from the hyperthermophilic archaeon, *Pyrococcus furiosus* and its associated RNAs. A 2.8 Å X-ray structure indicates that Sm4 can form a stable octamer in the absence of the RNA. We found that Sm4 specifically and directly associates with members of C/D and H/ACA RNAs as well as two uncharacterized non-coding RNAs. We propose that Sm4 is important for the biogenesis or/and function of the modification guide RNAs and thus in the formation of active ribosomes.

INDEX WORDS: RNA modifications, pseudouridylation, ribose methylation, H/ACA RNPs, Sm/Lsm proteins, non-coding RNAs

RNA-GUIDED RNA MODIFICATION IN ARCHAE: DYNAMIC ASSEMBLY OF ACTIVE H/ACA RNA/PROTEIN COMPLEXES AND A POTENTIAL ROLE OF SM4 IN RNA MODIFICATION

by

OSAMA A YOUSSEF

B.S., Assiut University, Egypt, 1994

M.S., South Valley University, Egypt, 1999

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

RNA-GUIDED RNA MODIFICATION IN ARCHAEA: DYNAMIC ASSEMBLY OF ACTIVE H/ACA RNA/PROTEIN COMPLEXES AND A POTENTIAL ROLE OF SM4 PROTEIN IN RNA MODIFICATION

by

OSAMA A YOUSSEF

Major Professor:	Michael P. Terns Rebecca M. Terns
Committee:	Michael W. W. Adams Claiborne V. C. Glover

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Michael Terns and Dr. Rebecca Terns for their guidance throughout my time as a graduate student. I appreciate your giving me the opportunity to join your lab and to contribute to your research. I would also like to thank my committee members Dr. Michael Adams and Dr. Claiborne Glover for all of their help.

Additionally, I would like to thank my colleagues in the Terns Lab. Thank you to all of my fellow members of the archaeal RNA group: Jason Carte, Caryn Hale, Lindsay Jones, Kyle Kleppe, Sonali Majumdar, and Neil Pfister. I would also like to thank the other members of the Terns Lab, including Eladio Abreu, Matthew Belcher, Brad Culp, and Micah Hale. Some former members of the Terns Lab who have helped me with my research are Dan Baker and Sarah Finch. Working with all of you has contributed a lot to my educational experience.

Finally, I would like to thank my friends and family for their support. Thank you for congratulating me when I was doing well and for encouraging me when I was struggling. I would especially like to thank my parents for instilling in me from an early age the importance of education.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Overview of RNA Modification

Several chemically modified nucleotides are found in cellular RNAs. More than 100 different post-transcriptionally modified nucleotides are currently known in all types of cellular RNA in the three kingdoms [1-3]. Coding (messenger) and non-coding (e.g., transfer, ribosomal, small nuclear, and small nucleolar) RNAs undergo post-transcriptional nucleotide modifications that are shown or predicted to be important for their functions [4]. While the exact role of many of these modifications is still unclear, many are highly conserved and contribute to the biological activities of the organism. The most common modifications in cellular RNAs are the addition of a methyl group to the 2'-hydroxyl group of the ribose ring and the isomerization of uridine (1-ribosyluracil) to pseudouridine (5-ribosyluracil) [5-11].

Ribosomes are large ribonucleoprotein complexes that carry out protein synthesis. Synthesis of ribosomal RNAs (rRNAs) is not achieved by simple transcription, but requires a complex series of post-transcriptional processing and nucleotide modification steps. In eukaryotes, 18S, 5.8S, and 25S/28S rRNAs are transcribed by RNA polymerase I as one contiguous unit (35S/45S rRNA) in the nucleolus. As soon as transcription of the precursor ribosomal RNA (pre-rRNA) is completed, the transcript undergoes extensive covalent modifications. Mammalian rRNAs contain approximately 110 2'-*O*-methylated sugar residues, and approximately100 pseudouridines [8,12]. Yeast (*Saccharomyces cerevisiae*) ribosomes contain about 50 2'-*O*-methylated riboses and approximately 50 pseudouridines [13]. Following modification, the nascent transcript undergoes a series of endo- and exonucleolytic cleavages to remove the two external transcribed spacers (ETSs) and the two internal transcribed spacers (ITSs) to generate the mature rRNA species.

Prokaryotic and archaeal rRNAs also contain 2'-*O*-methylated nucleotides and pseudouridines. The level of 2'-*O*-methylation in rRNAs of at least one archaeal species appears similar to that present in eukaryotes. *Escherichia coli* rRNAs contain four 2'-*O*-methylated nucleotides whereas *Sulfolobus solfataricus* (theromophilic archaeon) rRNAs contain approximately 67 2'-*O*-methylated nucleotides [14-17]. In contrast, the number of pseudouridines in archaea appears close to that present in prokaryotes. *E. coli* rRNAs contain 10 pseudouridines whereas there are 9 and 6 pseudouridines in *S. solfataricus* and *S. acidocalodarius* rRNAs, respectively [14-18]. The hyperthermophilic archaeon *Pyrococcus abyssi* contains 17 pseudouridines in rRNAs [19].

Pseudouridines and 2'-*O*-methylated nucleotides are present in all transfer RNAs (tRNAs) in eukaryotes, prokaryotes and archaea. Ribose methylation comprises approximately 8% of all tRNA modifications [6,20]. Pseudouridine is also found in all elongator tRNAs notably as the nearly universal conserved pseudouridine 55 (ψ55) in the T loop. In eukaryotes, spliceosomal small nuclear RNAs (snRNAs) involved in pre-mRNA splicing also contain a substantial number of modified nucleotides. U1, U2, U4, U5, and U6 snRNAs have a total of 30 2'-*O*-methylated nucleotides and 24 pseudouridines [9]. Although the presence of pseudouridines or ribose-methylated nucleotides has not been reported to date, the possibility that some mRNA nucleotides are subject to nucleotide modification is considered (see below for more details).

Nucleotide modifications can take place by two different mechanisms: An RNA-independent mechanism that is present in all organisms, and an RNA-dependent mechanism that is so far discovered only in eukaryotes and archaea. 2'-*O*-methylation and pseudouridylation of bacterial rRNAs and tRNAs are catalyzed by protein-only enzymes (methyltransferase and

pseudouridine synthase, respectively) that are site- and region-specific. On the other hand, the same two modifications in eukaryotic rRNAs and snRNAs and in archaeal rRNAs and tRNAs are created by site-specific RNA-protein complexes. There are two main classes of site-specific RNA-protein complexes. H/ACA RNPs guide and catalyze the conversion of target uridine to pseudouridine by base rotation; C/D RNP complexes guide and catalyze the methylation of 2'-hydroxyl group of ribose ring in the target nucleotide.

Two Classes of RNA Modification Guides

The guide RNA identifies its target nucleotide by base-pairing with the substrate RNA(s). The guide RNA functions as part of RNA/protein (RNP) complexes. A set of core proteins associate with the mature form of each guide RNA to form functional RNP complexes. All guide RNAs to date fall into two major classes, C/D and H/ACA guide RNAs, based on the conserved secondary structure and the presence of short conserved elements. In eukaryotes, the two classes of guide RNAs (ranging from 60 to 150 nucleotides in length) are localized in the nucleolus (a sub-nuclear organelle where ribosome biogenesis occurs). Hence, they are termed small nucleolar RNAs (snoRNAs) [21]. The snoRNAs play a key role in the maturation of pre-rRNAs in the nucleolus. The snoRNAs are not only involved in the modification of cellular RNAs but are also required for the cleavage of pre-rRNA to produce the mature form of each rRNA. Several snoRNAs (U3, U8, U14, and U22 C/D RNAs and U17/snR30, snR10, and E2 and E3 H/ACA RNAs) are involved in the processing of pre-rRNAs [22]. However, the vast majority of snoRNAs guide site-specific nucleotide modifications of rRNAs in the nucleolus.

Several guide RNAs are also implicated in nucleotide modification of snRNAs and are localized within Cajal bodies, which are nuclear organelles that appear to be involved in the

biogenesis of snRNAs [23-26]. The guide RNAs which accumulate in Cajal bodies have been called small Cajal body-specific RNAs (scaRNAs) [25]. In some cases, one scaRNA molecule contains both H/ACA and C/D motifs [25,27]. U85 scaRNA guides 2'-*O*-methylation of cytosine 45 and also pseudouridylation of uridine 46 of human and *Drosophila* U5 snRNA [27].

The presence of several tissue-specific guide RNAs raised the possibility that these guide RNAs can target nucleotides in mRNAs. Three potential C/D guide RNAs and one H/ACA RNA are present in mouse brain tissues [28]. One potential C/D RNA (MBII-52) has 18 nucleotides complementary to a serotonin 2C receptor mRNA sequence specifically expressed in the brain [28]. However, it is not known yet if the target nucleotide is indeed a modified one. Moreover, several orphan guide-like RNAs showing no complementarity to known RNA substrates (rRNAs and snRNAs) have been identified in eukaryotes [28,29]. Recently, it was shown that HBII-52 is involved in serotonin 2C receptor mRNA splicing [30]. The RNA orphans may guide modifications of other cellular RNAs including mRNA or have other functions distinct from their predicted function in the modification.

Eukaryotic H/ACA RNP Complexes

Pseudouridine was the first modified nucleotide to be discovered, and it is the most common modified nucleotide found in cellular RNAs [31,32]. The pseudouridylation of eukaryotic rRNAs and snRNAs (and perhaps mRNAs) is carried out by RNA-dependent pseudouridine synthase [33,34]. The pseudouridylation is started by nucleophilic attack of the catalytic aspartate residue of the enzyme on the target uridine at C6. This leads to the breakage of the N1 glycosidic bond and rotation of the base 180^{o} around N3-C6 axis, followed by

formation of a covalent bond at C5 (figure 1.1C) [35]. The H/ACA RNP complexes are responsible for this modification in eukaryotes and also in archaea.

The conserved secondary structure of the guide H/ACA RNA is folded into two irregular hairpins connected by a short single-stranded region (figure 1.1A) [36,37]. The conserved H (Hinge) element (ANANNA, where N is any nucleotide) is located between the two hairpins. The conserved ACA elements are located exactly three nucleotides upstream of the mature 3' end of the guide RNA. An internal loop found in the 5'- and/or 3'-hairpin contains an antisense sequence 9-13 nucleotides in length that is complementary to the target RNA sequence. The target uridine and an adjacent nucleotide are unpaired and located at the base of the upper stem closing the recognition loop of the guide RNA (pseudouridylation pocket). The distance between the target uridine and the conserved H or ACA elements is always 14-16 nucleotides [37,38]. Most eukaryotic guide RNAs are characterized by two hairpins. However, in trypanosomatids (unicellular protozoan parasites), H/ACA RNAs are comprised of a single hairpin carrying an AGA (instead of ACA) sequence at the 3' end [39,40].

Two separate studies showed that the H/ACA guide RNA can bind to the substrate RNA in the absence of the core proteins [41,42]. However, the stable interaction requires high concentrations of the two RNAs. In the first study, the solution structure of the 5'-hairpin of human U65 bound to a substrate RNA showed that the substrate RNA adopts an omega (Ω)-shaped conformation [41]. The resulting guide RNA/substrate RNA is strongly stabilized by magnesium ions. The second solution structure of substrate RNA/guide RNA complex was obtained using 3'-hairpin of human U65 bound to a substrate RNA [42]. The RNA/RNA interaction is weak (Kd = 200–300 μM), and the addition of divalent ions or higher concentration of monovalent ions did not significantly stabilize the complex. In the complex, the substrate

RNA adopts a U shape and makes a 180° turn at the target U and the 3′ unpaired nucleotide. Both solution structures of substrate RNA/guide RNA indicate that the substrate RNA interacts with only one face of the pseudouridylation pocket [41,42]. This observation may explain how the longer substrate (e.g. rRNA) can be associated and dissociated without the requirement for helicase activity *in vivo*.

The H/ACA guide RNAs are associated with four core proteins in yeast: Cbf5p (dyskerin in human, NAP57 in rodents), Gar1p, Nop10p, and Nhp2p (L7Ae in archaea, see below) [21,33,43-47]. Immunoprecipitation experiments showed that all four core proteins are localized in the dense fibrillar component (DFC) of yeast nucleolus where pre-rRNAs are modified and then processed [43,46]. Each of these four core proteins is described in detail below.

Cbf5p was originally identified in yeast as a centromere binding factor [48]. Cbf5p is a highly charged protein that contains ten tandem KKE/D repeats near its carboxy-terminal domain [48]. All the available evidence indicates that Cbf5p is the putative pseudouridine synthase that catalyzes the modification. Sequence comparison between Cbf5p and all known pseudouridine synthases showed high homology between them [49]. Moreover, genetic depletion of Cbf5p or mutation of the conserved aspartate residue in Cbf5p completely eliminated *in vivo* pseudouridylation of yeast rRNAs [50,51]. This provides support for Cbf5p being the pseudouridine synthase. Mutations in the human Cbf5 homolog dyskerin have been demonstrated to cause the X-linked skin and bone marrow disease Dyskeratosis Congenita (DC) [52].

Nhp2p, a highly basic protein, was originally identified as a non-histone protein. It contains a putative RNA-binding domain that is shared with the ribosomal protein L30 and the C/D RNP core protein p15.5 kDa (see below) [46,53,54]. Consistent with its predicted role as an

RNA binding protein, selective mutations in Nhp2p negatively affect the stability of H/ACA RNAs [55]. However, it was later shown that the interaction of Nhp2p with yeast and mammalian H/ACA RNAs is not sequence-specific [55,56].

Gar1p was the first protein identified as an H/ACA RNA associated protein in yeast [57]. Gar1p comprises a conserved central domain flanked by two glycine-arginine rich (GAR) domains [57]. The long GAR domain is located at the C-terminus and the short one is located at the N-terminus. An *in vitro* study indicated that Gar1 can bind directly to snR10 and snR30 H/ACA RNAs in yeast [58]. It was shown that the central domain of yeast Gar1 is sufficient for nucleolar accumulation and binding [58,59]. However, other work failed to find a direct interaction of Gar1 with H/ACA RNA [56].

Nop10 (Nucleolar protein 10 kDa) is the smallest core protein and lacks any known motifs [46]. The crystal structure of yeast Nop10 alone showed that the protein is mostly unstructured [60]. It contains an N-terminal β-hairpin domain connected by a disordered linker to an unstructured C-terminal domain [60-62].

Electron micrographs of purified yeast H/ACA RNPs have been interpreted to reflect that each hairpin of an H/ACA RNA is associated with one set of the core proteins [44,54]. All four H/ACA RNP core proteins are required for cell growth and maturation of pre-rRNAs. Depletion of any of the four core proteins leads to impairment of cell growth [46,50,54,57]. With the exception of Gar1p, the core proteins are required for accumulation and stability of H/ACA RNPs in the nucleolus [50,54,55,63,64]. In mammals, it was shown that immunopurified H/ACA RNP particles are sufficient for the pseudouridylation activity *in vitro* [65]. No energy or helicase is required for pseudouridylation activity *in vitro*. With the exception of Cbf5p, the specific functional roles of the remaining three core proteins remain uncertain.

Reconstitution of mammalian H/ACA RNP complexes was reported using proteins produced by *in vitro* transcription/translation [56]. It was shown that the Nop10-Cbf5 interaction is required for Nhp2 binding. However, it is not clear whether Nhp2 binds to Cbf5 or Nop10, or both. Gar1 associates independently with Cbf5 and its association is stronger in the presence of Nop10 and Nhp2. Nhp2 binds nonspecifically to yeast and mammalian H/ACA RNA [55,56] whereas mammalian Cbf5/Nhp2/Nop10 core trimeric complex specifically recognizes H/ACA RNA *in vitro* [56]. Mammalian Gar1 and Cbf5 crosslink to the guide RNA near the catalytic site of RNPs [56]. In this study, no pseudouridylation activity was detected with mammalian reconstituted H/ACA RNPs.

In yeast, Henras and co-workers investigated the protein/protein interactions taking place between the four core proteins *in vivo* [66]. This study indicated that Cbf5p, Gar1p, and Nop10p can form a complex in the absence of Nhp2p and the guide RNA. The pseudouridylation activity of the purified yeast complexes was not reported in this study. In the two above-mentioned studies, the expression and purification of stable eukaryotic H/ACA RNP core proteins, especially Cbf5, were challenging [56,66].

Eukaryotic C/D RNP Complexes

The second class of guide RNPs are C/D RNP complexes that guide and catalyze the methylation of the 2'-hydroxyl group of the ribose ring in the target nucleotide (figure 1.2B) [67,68]. Eukaryotic C/D RNPs contain a guide RNA and four core proteins: fibrillarin (Nop1 in yeast), Nop56, Nop58, and p15.5 kDa (Snu13p in yeast) [69-74]. The secondary structure of the guide C/D RNA contains two highly conserved motifs: the C (RUGAUGA, where R stands for any purine) and D (CUGA) elements, which are located near the 5' and 3' ends, respectively

(figure 1.2A) [75]. These two motifs are brought together by a terminal stem. C/D RNAs can also contain two weakly conserved motifs called C' and D' elements located internally and brought together via a stem-loop [76-79]. Each C/D RNA contains 10-21 nucleotides, which are located upstream of the D and/or D' elements that are complementary to a sequence in the target RNA. The target RNA nucleotide to be modified is always located five nucleotides upstream of the D and/or D' elements [80-82]. Eukaryotic C/D RNA contains one kink (k)-turn motif located in the C/D elements and sometimes contains a second k-turn motif located in the C'/D' elements. Archaeal C/D and H/ACA RNAs contain a similar k-turn motif. The C'/D' elements are less conserved in eukaryotes and often lacks the k-turn motif. The k-turn is a helix-internal loop-helix motif containing ~15 nucleotides and a kink in the phosphodiester backbone that bends the RNA helix axis by 120^{o} (figure 1.1B) [83]. The first helix ends at the internal loop with a G-C base pair. The internal loop is purine-rich and contains three asymmetric unpaired nucleotides. The second helix starts with two non-Watson-Crick base pairs, which are always G-A base pairs.

The four core proteins of C/D RNPs are highly conserved and important for yeast viability. Fibrillarin (Nop1 in yeast) exhibits amino acids sequence motifs characteristic of SAM (*S*-adenosyl-L-methionine)-dependent methyltransferases and is likely the C/D RNA-guided modifying enzyme [84,85]. Point mutations in the methylase-like domain disrupt rRNA methylation [86]. The p15.5 kDa (Snu 13p in yeast) binds specifically to the C/D guide RNA in an interaction mediated by the kink (k)-turn motif. The p15.5 kDa is also a core protein of the splicesomal U4 snRNP where it binds to the conserved k-turn motif within the snRNA [87,88]. Nop56 and Nop58 are highly related to each other in sequence, having about 37% identity (in human) or 40% identity (in yeast) [72]. However, the precise function of these two proteins is unclear. In yeast Nop58 is required for C/D RNA stability [73,89]. In contrast, many studies

have revealed that Nop56 is not essential for C/D RNA stability *in vivo* [73,89,90]. All four of the C/D RNP core proteins are required for enzymatic activity and nuclear localization of C/D RNPs [91].

The structural organization of the eukaryotic C/D guide RNP complexes has been studied *in vivo* and *in vitro* [92,93]. UV-crosslinking experiments showed that Nop56 and Nop58 interact with C' and C elements respectively *in vivo*. Fibrillarin cross-linked to both C/D and C'/D' motifs [92]. An *in vitro* study showed that p15.5 kDa recombinant protein interacts only with the C/D motif in the absence of the other proteins and it initiates RNP assembly [90,93].

Interaction of Guide RNP Core Proteins with Other Proteins

In addition to the core proteins of the guide RNPs, there are several proteins that interact transiently with guide RNPs in eukaryotes. These proteins are associated with guide RNPs based on their ability to interact with one or more core proteins. However, none of these proteins appears to be part of the mature guide RNP complexes or are required for the modification. Instead, they are implicated in some aspect of the biogenesis of the guide RNPs.

Nopp140 is a highly phosphorylated protein localized in the nucleolus and Cajal bodies [43,94]. Nopp140 interacts with both classes of guide RNP complexes *in vitro* and *in vivo* [95,96]. Coimmunoprecipitation experiments showed that the association of Nopp140 with RNPs is phosphorylation-dependent [65]. An *in vitro* pseudouridylation assay revealed that pseudouridylation activity is independent of Nopp140 [65]. However, the finding that guide RNPs localize in the nucleolus and Cajal bodies and that Nopp140 shuttles between these two sub-nuclear organelles suggests the role of this protein in the RNP assembly or transport *in vivo*.

SMN (Survival of Motor Neurons) interacts transiently with newly made snRNPs but is not a component of spliceosomes [97,98]. *In vitro* protein-protein interactions showed that SMN interacts directly and independently with both Gar1 (of H/ACA RNPs) and fibrillarin (of C/D RNPs) [99-101]. Mutational analysis indicated that SMN/Gar1 and SMN/fibrillarin interactions are mediated by the Tudor domain of SMN and the GAR domains of Gar1 and fibrillarin [100,101]. It is predicted that SMN may play a role in the biogenesis of both C/D and H/ACA guide RNP complexes. However, the predicted function of SMN in RNPs biogenesis would be restricted to higher eukaryotes as yeast lacks an SMN homolog.

p50 and p55 are two highly related nucleoplasmic proteins required for production and localization of C/D and H/ACA RNPs [102]. They contain Walker A and B motifs that are responsible for ATP/GTP binding and ATP hydrolysis [103]. Depletion of p50 gene in yeast impairs the localization of Gar1p and Nop1 (yeast fibrillarin) suggesting a role of p50 in the RNPs assembly. Point mutations in p50 showed that the conserved Walker motifs are required for the accumulation of C/D and H/ACA RNPs [102].

In addition, proteins have been identified that are specific to H/ACA RNP complexes. For example, Naf1 and Shaq1 proteins have been shown to be involved in the assembly of H/ACA RNPs but not C/D RNPs. Genetic depletion of either Naf1 or Shaq1 causes specific loss of all H/ACA RNAs [104-106]. While Shaq1 binds to Nhp2p, Naf1p associates with all H/ACA RNP core proteins. Rnt1p is the yeast RNase involved in the cleavage of the 3' end of pre-rRNA and physically interacts with Gar1p. This protein is required for nuclear import of H/ACA RNP core proteins [107].

Mechanisms of RNA Modification in Archaea

Archaea are bacteria-like in cell structure, genomic organization, and the structure and function of enzymes involved in basic metabolism. On the other hand, archaea are eukaryote-like in DNA replication, transcription, translation and RNA modifications [108,109]. Although archaea lack a nucleolus and nucleus, their genomes encode homologs of proteins used in eukaryotic guide RNP complexes [109,110]. Archaea use an RNA-dependent mechanism to methylate the 2'-hydroxyl group of rRNAs and tRNAs ribose rings. Archaea also use an RNA-dependent mechanism to convert uridine to pseudouridine in rRNAs. On the other hand, the pseudouridylation of archaeal tRNAs takes place by a protein-only enzyme mechanism that is similar to its eukaryotic counterpart.

Archaeal H/ACA RNP Complexes

Archaeal H/ACA RNP complexes, which are the main focus of the first part of this thesis, guide and catalyze the pseudouridylation of archaeal rRNAs [68,109]. The RNA components of archaeal H/ACA RNPs share the same conserved secondary structure with the eukaryotic counterparts (figure 1.1B). However, there are some differences between the archaeal and eukaryotic H/ACA guide RNAs. Archaeal H/ACA RNAs are characterized by one, two, or three hairpin domains, which each contain a pseudouridylation pocket (eukaryotic counterparts have two hairpin domains) [111,112]. The distance between the modified nucleotide and ACA elements is maintained (14-16 nucleotides). Archaeal H/ACA RNAs (but not eukaryotic H/ACA RNAs) exhibit a k-turn motif located near or in the apical loop [113]. To date, the k-turn motif is not found in any known eukaryotic H/ACA RNAs. In archaeal H/ACA RNAs, ANA (N is any

nucleotide) sequence is located at the 3'-end of the guide RNAs and also between the two hairpin structures.

In a cDNA library from *Archaeoglobus fulgidus*, four non-coding RNAs exhibit H/ACA motifs and guide the isomerization of six uridines to pseudouridines in 16S and 23S rRNAs [114]. Screening GC-rich regions in *Pyrococcus furiosus* genome and computational searches in the three *Pyrococcus* genomes (*P. furiosus*, *P. abyssi,* and *P. horikoshii*) uncovered seven non-coding RNAs that exhibit homology to H/ACA RNAs [19,111,115]. It is predicted that these H/ACA RNAs can guide pseudouridylation of 15-17 uridine residues in 16S and 23S rRNAs [19,111,116].

Analyses of several archaeal genomes showed the presence of archaeal homologs to all eukaryotic H/ACA RNP core proteins except Nhp2 [112]. However, it was shown that archaeal ribosomal protein L7Ae exhibits sequence homology to Nhp2p and also to p15.5 KDa (eukaryotic C/D RNPs core protein) [111]. There is a 45% identity between human Nhp2 and L7Ae from *Methanococcus jannaschii* [46]. L7Ae is an RNA-binding protein that binds directly to archaeal H/ACA RNA (also C/D RNA, see below). This RNA/protein interaction is mediated by a k-turn motif [111,117].

In the past three years, a lot of progress has been made to better understand the function of each component in the H/ACA RNP complexes. Biochemical studies and three dimensional (3D) structures from our lab and others have unveiled the structural organization of H/ACA RNP complexes in detail. Our lab reported the first *in vitro* reconstitution of functionally active archaeal H/ACA RNP complexes [116]. Recombinant H/ACA RNP core proteins and, guide and substrate RNAs derived from *Pyrococcus furiosus* (hyperthermophilic archaeon) were used [116]. Information from biochemical studies indicated that Gar1 and Nop10 bind directly and

independently to Cbf5 [116,118]. Cbf5 and L7Ae interact with the guide RNA directly and independently of each other. In the absence of the guide RNA, L7Ae does not interact with any of the other three core proteins. Gar1 and Nop10 do not bind to the guide RNA in the absence of Cbf5. Archaeal Cbf5 forms a trimeric complex with Nop10 and Gar1 independent of L7Ae and the guide RNA. An *in vitro* functional assay indicated that Cbf5 and Nop10 are the minimal protein components required for the pseudouridylation activity [118]. However, all four core proteins are required for full pseudouridylation activity [116,118].

The crystal structure of archaeal Cbf5 has been solved in complex with Nop10 [60,119], with Nop10 and Gar1 [120], and with the entire H/ACA RNP complex [121]. Cbf5 belongs to the *E. coli* pseudouridine synthase TruB family which is responsible for pseudouridylation of uridine 55 in all elongator tRNAs [122]. There is about 30% sequence identity between archaeal Cbf5 and *E. coli* TruB [60]. The X-ray structure of Cbf5 from three different hyperthermophilic archaea, *Pyrococcus furiosus* [120,121,123], *Pyrococcus abssyi* [119], and *Methanococcus jannaschii* [60] are very similar. Cbf5 contains a catalytic domain at the N-terminus and a pseudouridine and archaeosine transglycosylase (PUA) domain at the C-terminus (figure 1.3). The catalytic domain is sub-divided by a central active site cleft into roughly two equal sub-domains. The PUA domain is conserved in all pseudouridine synthases and archaeosine transglycosylases (ArcTGT). ArcTGT catalyzes the first step in the conversion of guanine 15 (G15) to archaeosine in the D-stem of archaeal tRNA [124]. The ArcTGT PUA domain interacts with the acceptor stem and the terminal 3'-CCA sequence of the archaeal tRNA [125]. The crystal structures of archaeal Cbf5 and *E. coli* TruB superimpose very closely (figure 1.4). Analysis the structures of archaeal Cbf5 and *E. coli* TruB showed that most of the amino acid residues that participate in the modification are conserved in equivalent locations in the TruB and

Cbf5 active site [126]. However, there are two significant differences between the two structures. First, Cbf5 lacks a β5/β6 hairpin, α4, and β8/β9 hairpin structures (figure 1.4). These secondary structures in TruB are located in the thumb-loop where the substrate tRNA binds. Secondly, Cbf5 has a longer N-terminal domain that forms an additional β-strand and wraps around the PUA domain (see figure 3, N-terminal tail).

Archaeal Gar1 lacks the two GAR (glycine-arginine rich) domains that are present in the eukaryotic homolog. To date, there is no three dimensional structure for Gar1 alone. The crystal structure of Gar1 bound to Cbf5 showed that Gar1 belongs to a superfamily of reductase, isomerase, and elongation factor fold [120]. This protein folds into a six-stranded β barrel and binds to one side of the Cbf5 catalytic domain [120,121]. Archaeal Gar1 shares the same fold with RNA-binding proteins (EF-Tu domain 2). The 3D structure of Gar1 in the Cbf5/Gar1/Nop10 complex showed that Gar1 utilizes this fold to bind to Cbf5 instead of the RNA [120].

Archaeal Nop10 proteins contain a zinc-binding consensus sequence (CX2CX8CX2C, where X is any amino-acid residue) that is not conserved in eukaryotic Nop10 [60]. The crystal structure of archaeal Nop10 in complex with Cbf5 alone [60,119] or with Cbf5 and Gar1 [120] showed that the protein contains a zinc-binding domain in the N-terminal region connected to an α-helix in the C-terminal region by a linker. However, the NMR structures of archaeal Nop10 alone showed that the protein is mostly unstructured, similar to eukaryotic Nop10 [60]. Consistent with the biochemical data, the 3D structure of the Cbf5/Gar1/Nop10 heterotrimeric complex showed that Gar1 and Nop10 bind independently to the catalytic domain of Cbf5 (figure 1.3) [120,121].

The crystal structure of H/ACA RNP holoenzyme was reported at 2.1 Å [121]. This structure reveals that Nop10 is located between Cbf5 and L7Ae, contacting the upper stem of the guide RNA, Cbf5, and L7Ae (figure 1.5) [121]. The pseudouridylation pocket is located at the center of Cbf5 catalytic domain. L7Ae binds to the k-turn motif. The observed interaction between Nop10 and L7Ae or between Nop10 and the guide RNA apparently only occurs in the context of RNP complexes. The structure of Cbf5/Gar1/Nop10 trimer alone [120] is very similar to that in the entire H/ACA RNPs [121]. This indicates that binding of the trimeric complex to the guide RNA has no apparent effect on the structure of the heterotrimeric complex. RNA footprinting experiments showed that binding of L7Ae induces a conformational change in the guide RNA [127]. This conformational change alters Cbf5 binding sites on the guide RNA.

Recently, the cocrystal structure of an H/ACA RNP (lacking the k-turn motif and L7Ae) bound to substrate RNA was reported at 2.87 Å (figure 1.6) [123]. The structures of Cbf5, Gar1 and Nop10 in this complex are very similar to those in H/ACA RNP complexes lacking the target RNA (compare between figure 1.5 and 1.6). This indicates that binding of the target RNA does not affect the structure of the trimeric complex. However, in the cocrystal structure, the target uridine is placed about 11 Å away from the catalytic site and consequently cannot be modified. Fluorescence analysis indicated that L7Ae is required for the correct placement of the target uridine at the active site [123].

Archaeal C/D RNP Complexes

Archaeal C/D guide RNAs have similar secondary structures to the eukaryotic counterparts (figure 1.2A). While only ~20% of eukaryotic C/D RNAs direct ribose methylation using both D and D' elements, the majority of archaeal C/D RNAs are able to use both elements

to direct ribose methylation [128]. Archaea also contain protein homologs to eukaryotic C/D RNP core proteins, but archaeal fibrillarin homologs lack the GAR domain present in eukaryotic fibrillarin [129]. The crystal structure of archaeal fibrillarin showed that the protein contains an *S*-adenosyl methionine-binding site [85]. Archaea posses a single protein homolog for both Nop56 and Nop58 called Nop56/58. The archaeal protein most closely related to p15.5 kDa is the ribosomal protein L7Ae [130,131].

The *in vitro* reconstitution of functionally active archaeal C/D guide RNP complexes was reported [131,132]. These studies suggest that L7Ae binds first and then Nop56/58 followed by fibrillarin. L7Ae interacts directly with C/D guide RNA in the absence of the other two proteins. This interaction is mediated by a k-turn motif found at each C/D and C'/D' elements. Binding of Nop56/58 to the guide RNA requires L7Ae, and association of fibrillarin with the RNA requires both L7Ae and Nop56/58.

Biogenesis of Guide RNP Complexes

In eukaryotes, biogenesis of snoRNPs (guide and processing RNPs) takes place in several steps: (i) Synthesis of snoRNAs in the nucleoplasm and their protein components in the cytoplasm, (ii) Assembly of RNPs in the nucleoplasm and (iii) Movement of functional RNPs to the nucleolus or Cajal bodies [133].

Eukaryotic cells use different strategies to generate snoRNAs. snoRNAs are found either within introns of mRNA genes or in independent transcription units. In vertebrates, all guide RNAs are transcribed from the introns of pre-mRNA genes by RNA polymerase II in the nucleoplasm [134,135]. In most cases, the host genes encode proteins involved in the biosynthesis, structure, or function of the ribosome [136]. However, some host genes encode no

proteins and appear to act as carriers for snoRNAs biosynthesis [135,137]. All vertebrate snoRNAs derived from host genes belong to the 5'-terminal oligopyrimmidine (5'-TOP) family, whose transcripts start with an oligoprymidine tract [138]. In mammals, introns can encode either single or multiple guide RNAs but their host genes do not encode proteins in the exonic sequences [139]. Intronic guide RNA molecules can be processed by two different pathways. The major pathway involves exonucleolytic cleavage of the debranched lariat [140,141]. In the minor pathway, guide RNAs are released directly from the pre-mRNA by endonucleolytic cleavages followed by exonucleolytic trimming [140-142]. The main feature of these RNAs is the presence 5' monophosphate group [136,143].

In vertebrates and yeast, a few snoRNAs are transcribed from independent genes by RNA polymerase II. In general, vertebrate guide RNA genes are found in monocistronic genes. In other organisms, guide RNA genes are found in polycistronic genes [144]. A characteristic feature of these RNAs is the presence of a trimethyl guanosine cap at their 5' ends. These guide RNAs can mature by endonucleolytic cleavage followed by exonucleolytic trimming.

The biogenesis of archaeal guide RNAs is still obscure [137]. Guide RNAs in archaea are located in the regions between protein coding genes. They sometimes slightly overlap the upstream and/or downstream ORFs (open reading frames) [145,146]. Analysis of many archaeal species showed that one C/D RNA is located within the intron of the pre-tRNA-Trp gene [146]. This C/D RNA guides the methylation of C34 and U39 within the mature form of tRNA-Trp by a cis- or trans-acting mechanism [110,147]. A recent study showed that C/D guide RNAs exist as circular RNAs in *P. furiosus,* raising the possibility of the existence of an uncharacterized biogenesis pathway for C/D RNAs in archaea [148].

Functional Roles of Nucleotide Modifications

One of the most exciting questions in RNA modification research is what role(s) modified nucleotides play in the cell. To date, the precise role of the nucleotide modification is still unclear. Some experimental evidence indicated that specific rRNA modifications are not essential [149]. In yeast, deletion of guide RNAs individually results only in the loss of the modifications [149]. However, some observations indicate their biological importance. Blocking the modifications in larger multiples affected the cell viability [150]. These modifications are located within conserved functionally important regions of rRNAs such as peptidyl transferase and decoding centers [15,76,151,152]. However, they are often located away from the protein-binding sites [153]. Some other observations indicate that modifications can also play a role in either ribosomal protein assembly or ribosomal RNA folding [10,150].

The addition of a methyl group to a ribose ring is known to stabilize RNA either by increasing the hydrophobicity or by preventing endonucleolytic cleavages. Pseudouridines are able to form an additional hydrogen bond, and this may increase the rigidity and stability of the RNA [154]. The modifications of U2 snRNA are required for the assembly of functional 17S U2 snRNPs supporting the idea that modifications are required for protein assembly [155]. Interestingly, the modified nucleotides in snRNA are mainly located in regions involved in RNA/RNA interactions or conformational switches during spliceosomal assembly and function. This observation suggests that the modifications play an important role in splicing control [9,155]. The idea that ribose methylation can enhance RNA stability is supported by the observation that the number of ribose methylations in *Sulfolobus solfataricus* (Theromophilic archaon) increase when the cell is grown at higher temperatures [156].

One possible role for guide RNAs that cannot be roled out is functioning as chaperones helping in the correct folding of the ribosomal RNA whereby the modification is just a by-product of or signal for guide RNA/rRNA dissociation [157]. Interestingly, the mRNA nucleotide potentially targeted for 2'-*O*-methylation by C/D RNA MBII-52 is also subject to adenosine to inosine editing. *In vitro* methylation of the target adenosine dramatically inhibits its deamination to inosine suggesting for a role of guide RNA in the regulation of mRNA editing [158].

Biological Functions of H/ACA RNP Complexes

H/ACA RNA can be subdivided into two subclasses: guide and non-guide RNAs. Guide H/ACA RNAs are required for the conversion of uridines to pseudourindines (as discussed above in detail). Non-guide RNAs include a processing H/ACA RNP and vertebrate telomerase RNA [159]. U17/E1 (snR30 in yeast) is required for processing of pre-rRNA to produce the mature form of 18S rRNA [160]. Human telomerase RNA is required for the synthesis of the telomere at the end of the chromosome. The 5' domain of telomerase RNA folds into a pseudoknot containing the template for reverse transcriptase. The 3' domain contains H/ACA RNA double-hairpin structure [159]. In yeast, stable expression of human telomerase depends on its association with Cbf5p, Nhp2p, and Nop10p [64]. The H/ACA RNPs domain is required for nuclear retention, accumulation, and stability of telomerase RNA [159,161].

In addition, one of the human H/ACA RNP core proteins, dyskerin (Cbf5 in yeast), is connected to a rare skin and bone marrow failure disease, Dyskeratosis Congenita (DC). DC is caused by mutations in the gene DKC1 encoding dyskersin [52]. Most of the mutations that could be mapped to archaeal Cbf5 occur in the PUA domain with three in the catalytic domain

[60,119,120]. The crystal structure of H/ACA RNP complexes revealed that these mutations cluster to Cbf5 PUA domain involved in interacting with the lower stem of the guide RNA [121]. Very recently, it was shown that mutations in Nhp2 can also cause DC [162].

Sm/Lsm/HfQ Family of Proteins

Sm/Lsm/HfQ proteins are a large family of evolutionarily ancient RNA binding proteins that are involved in the processing of a wide variety of cellular RNAs in all organisms.

Structure and Function of Sm Proteins

Sm proteins were originally identified as immune targets in autoimmune disease. Stephanie Smith was the first patient in which the systemic lupus erthematosus anti-Sm autoimmune antibodies were identified [163]. Seven different Sm proteins termed B/B', D1, D2, D3, E, F, and G are identified in human [164]. B and B' Sm proteins are highly related proteins generated by alternative splicing of a single transcript [165]. Sm proteins form three stable sub-complexes B/D3, D1/D2, and F/E/G in absence of the RNA [166,167]. The crystal structures of D1/D2 and B/D3 heterodimers revealed that Sm proteins have a conserved Sm fold that consists of an α-helix at the N-terminus followed by five twisted antiparallel β-sheet strands [168,169] (figure 1.7B). Sm proteins contain two conserved motifs termed Sm1 and Sm2 separated by a region of variable length and composition [166]. The Sm1 motif consists of the β1, β2 and β3 strands whereas, the Sm2 motif consists of β4 and β5 strands. Crystal structures of Sm sub-complexes showed that amino acid residues located in β4 of one Sm protein interacts with residues located in β5 of the other Sm protein [168]. Images obtained by electron microscopy

showed that the seven Sm proteins form a ring-shaped structure with a diameter of 8 nm bound to snRNAs [170].

The snRNP complexes are part of the spliceosome that is essential for pre-mRNA splicing in eukaryotes [168]. Sm proteins are required for the maturation and assembly of the spliceosome. In the cytoplasm, Sm proteins bind to snRNAs in a stepwise process and induce the hypermethylation of the N7-monomethyl guanosine at their 5' ends [171]. The hypermethylated snRNPs return to the nucleus where they will function [172,173]. The Sm proteins form a heteroheptameric complex bound to a conserved single-stranded region ($AU_{(4-6)}G$) termed Sm site of U1, U2, U4 and U5 snRNAs to form snRNP complexes. Sm site is usually located between two-hairpin structures required for proper functional binding [174]. In eukaryotes, D1/D2 and E/F/G Sm heteromers binds to snRNA to produce snRNP sub-complex [175,176]. B/D3 Sm heterodimer then joins the snRNP complexes. Electron micrographs of eukaryotic Sm complex bound to U1 snRNA showed that the RNA binds to the center of the Sm ring with one nucleotide binding to each monomer [177]. The loops between the β2 and β3 strands and between the β4 and β5 strands face the center of the ring and form an RNA-binding pocket in the Sm ring (figure 1.7A).

Structure and Function of Eukaryotic Lsm Proteins

Like-sm (Lsm) proteins were first identified as a conserved protein family that contains the Sm motifs found in the Sm proteins [166,178]. There are at least 16 different genes encoding Lsm proteins in eukaryotes which can form different complexes with different functional roles [179]. Lsm proteins share the same Sm fold with the conservation of the two Sm motifs. A multiple alignment showed that Lsm2-8 proteins are related to SmD1, D2, D3, E, F, G, and B,

respectively. Lsm1 and Lsm9 proteins do not appear clearly related to Sm proteins [180]. Sm and Lsm proteins are members of an RNA binding protein family. Unlike Sm proteins, Lsm proteins can form stable heptameric or hexameric complexes in the absence of the substrate RNA. Lsm proteins are involved in the metabolism of a variety of cellular RNAs. This includes pre-mRNA splicing [181,182], telomere synthesis [183], histone mRNA maturation [184-186], mRNA degradation [187,188], and rRNA and tRNA maturation [180,189-192]. Lsm proteins interact as polymeric complexes with various RNAs, including the spliceosomal snRNA, pre-RNase P RNA, snoRNAs and other non-coding RNAs (see below).

In yeast and vertebrates, several Lsm complexes have been identified. The protein composition of the ring seems to determine the function of the complex. Lsm2-8 complex is localized in the nucleus and is associated with U6 snRNA that is part of the spliceosome. This complex binds to the uridine-rich sequence at the 3' end of U6 snRNA and is necessary for the stability of U6 snRNA [180,193,194]. Lsm2-8 complex remains associated with U6 in U4/U6 di-snRNPs and U4/U6.U5 tri-snRNPs [193-195]. One proposed role for Lsm complex in the binding to U6 snRNA is to facilitate RNA/RNA interaction between U6 and U4 snRNAs [193]. In yeast, depletion of any of Lsm2 to Lsm8 proteins reduced the level of U6 snRNA and caused pre-mRNA defect [194].

Lsm proteins are also required for the maturation of rRNAs and tRNAs in yeast. Depletion of any of Lsm2-5 or Lsm8 resulted in a delay of pre-rRNA processing and accumulation of aberrant processing intermediates [196]. These observations suggest a role of Lsm proteins in the maturation of rRNAs. In another study, depletion of any of Lsm2-5 or Lsm8 resulted in strong accumulation of pre-tRNA species [192]. In yeast, Lsm2-7 proteins associate with pre-RNase P RNA but not with the mature RNase P RNA suggesting a role for Lsm in the

maturation of RNase P RNA and in tRNA processing [180] (RNase P is required for the maturation of tRNAs).

Lsm proteins are also involved in the maturation and function of snoRNAs. Analysis of the maturation of U3 snoRNA revealed that Lsm proteins are essential for the normal processing of the U3 3'-end [190]. In yeast, maturation of U3 snoRNA is generated from the extended precursors by endonucleolytic cleavage followed by exonucleolytic trimming. Depletion of any of Lsm2-5 or Lsm8 resulted in losing the normal 3' extended precursor and accumulation of truncated forms of both mature and pre-U3. Moreover, Lsm2-4 and Lsm6-8 complex binds specifically to U8 C/D snoRNA in *Xenopus laevis* [191]. U8 snoRNA is required for the maturation of 5.8S and 28S rRNAs [197]. Mutational analysis showed that the binding of the Lsm complex to U8 snoRNA requires a conserved eight-nucleotide sequence in the third loop of the RNA. This eight-nucleotide sequence (GCUGAUUA) is well conserved among vertebrate U8 snoRNAs [198]. There is also a hexameric Lsm complex that can bind to yeast snR5 (an H/ACA guide RNA) and contains Lsm2-7 proteins [189]. Mutational analysis indicated that the last 9 nucleotides (including the conserved ACA sequence) in snR5 are critical for the binding. In this work, it was shown that Lsm2-7 complex is not required for pseudouridylation activity of the snR5. It is predicted that the Lsm2-7 hexameric complex may also play a role in rRNA maturation as the complex localizes to nucleoli where the maturation of rRNA occurs.

The Lsm1-7 complex is cytoplasmic, and its binding to mRNA is required for the deadenylation-dependent decapping step and subsequent 5'- to 3'- exonucleolytic cleavage [187,188]. Mutations in Lsm1-7 proteins resulted in accumulation of capped mRNA degradation intermediates in yeast [187,188]. The Lsm1-7 complex also protects the 3'-end of mRNA from premature degradation by 3'- to 5'- exonuclease [199]. Mutations in Lsm1-7 proteins resulted in

accumulation of 3'-shortened mRNAs [187,199]. Lsm10-11 proteins form a heptameric complex with five different Sm proteins (B, D3, E, F, and G) on U7 snRNA [200]. The resulting snRNP complex is essential for processing the 3' ends of histone mRNAs. Very recently, the crystal structure of a yeast homo-octameric Lsm3 complex was solved at 2.5 Å [201]. Pulldown (protein/protein interaction) assay showed that this octameric Lsm3 complex can select Lsm2, Lsm5, and Lsm6 from yeast lysate. It is predicted that Lsm3 complex binds to other complexes containing Lsm2, Lsm5 and Lsm6 or the octameric Lsm3 complex is disrupted and then forms a heteromeric complex containing the four different Lsm proteins [201].

Structure and Functions of Bacterial Lsm (HfQ) Proteins

A eubacterial Lsm protein, HfQ, was originally identified as a host factor required for bacteriophage Qβ replication [202]. HfQ protein forms a stable homohexameric complex in the absence of the substrate RNA. The crystal structures of HfQ complexes from different prokaryotic species shows that the protein shares the same Sm fold [203-205].

Several studies suggested that HfQ is a post-transcriptional riboregulator. HfQ regulates the decay of some mRNAs by binding to their poly (A) tails, stimulating poly (A) adenylation and protecting these mRNAs from enzymes involved in mRNA degradation [206-209]. HfQ can also regulate mRNA decay by interfering with the ribosome binding site [210]. HfQ facilitates the binding of small non-coding regulatory RNAs to their mRNA targets. These regulatory RNAs function as post-transcriptional regulators, controlling the stability of some mRNAs and affecting the translation of others [211]. For example, HfQ binds to oxyS RNA, an oxidative stress response regulator, facilitating its interaction with *rpoS* mRNA to inhibit *rpoS* translation [212]. The cocrystal structure of HfQ bound to a short sequence (5'-AU_5G-3') showed that the

RNA winds around the central basic charged hole of the complex. The binding site can accommodate either poly A or poly U but not poly C or poly G [205]. This is consistent with the finding that HfQ binds AU-rich regions and poly (A) tails with high affinity.

Structure and Function of Archael Lsm Proteins

Archaea also contain genes that encode Lsm proteins [180,194]. The crystal structures of Lsm proteins became available from several distinct archaeal species. Archaeal Lsm proteins can form homomeric complexes that have structural similarity to other Sm/Lsm complexes. The crystal structures of two Lsm proteins from *Archaeoglobus fulgidis* were solved. AF-Sm1 forms a homoheptameric complex *in vitro* in the absence of RNA similar to Lsm and HfQ proteins [213]. On the other hand, AF-Sm2 protein forms a heptameric complex only in the presence of RNA (similar to eukaryotic Sm proteins) [214] or hexameric complex at low pH in the absence of RNA [215]. *A. fulgidis* Lsm complexes (AF-Sm1 and AF-Sm2) associate with RNase P RNA *in vivo* indicating a potential role in pre-tRNA processing [213]. The 3D structure of a *P. abyssi* Sm1 (PA-Sm1) complex revealed that the protein can form a heptameric complex that binds to a seven-nucleotide long oligo uridine (U) [216]. Oligo Us contact the complex at two different sites: inside the center of the complex and on the surface of the complex close to N-terminal α-helix. Archaeal Lsm protein (Sm3) from *Pyrobaculum aerophilum* contains an additional C-terminal domain [217]. Sm3 proteins form a dimer of heptamers (14-mer) that bound to 14 cadmium ions.

A recent crystal structure showed that archaeal species also contain a homolog of bacterial HfQ. *Methanococcus jannashii* encodes a protein homolog of HfQ (Mja-HfQ) [218]. Mja-HfQ protein forms a homohexameric complex that shows a functional and structural

relationship to bacterial HfQ. Mja-HfQ facilitates *in vitro* RNA/RNA interaction between mRNA and its regulatory small RNA. Moreover, an *in vivo* studies indicated that Mja-HfQ can bind to the small regulatory RNA and protect it from endonucleolytic cleavage [218].

Focus of this Dissertation

The primary objective of this thesis was to characterize the assembly and structural organization of functionally active archaeal H/ACA RNP complexes. Including is the detailed structural characterization of H/ACA RNP sub-complexes. As previously described in detail, there are four core proteins associated with H/ACA RNA. Each one of these core proteins is essential for the pseudouridylation activity of the complex. One of these proteins is the modifying enzyme (Cbf5). *In vitro* RNA/protein and protein/protein interaction assays were performed to investigate in detail the architecture of H/ACA guide RNP complexes. We demonstrated that efficient pseudouridylation activity requires the four core proteins and the full-length guide RNA molecule. The results in this thesis elucidated the functional role of L7Ae in the formation of the complex.

A second objective of this thesis was to identify RNAs associated with a novel archaeal Lsm protein (Sm4) *in vivo*. The direct interaction between archaeal the Lsm protein and its associated RNAs was tested. The results in this thesis showed that Sm4 associates with archaeal guide RNAs in addition to novel non-coding RNAs *in vivo*. *In vitro* studies demonstrated a direct interaction between Sm4 and the guide RNAs.

The work in this thesis is presented in four additional chapters. Chapter 2 presents the *in vitro* reconstitution of archaeal H/ACA RNP complexes (RNA/protein and protein/protein interactions). A detailed structural analysis of H/ACA RNP sub-complexes is presented in

chapter 3. Chapter 4 presents the identification of archaeal Sm4 associated RNAs. Chapter 5 briefly discusses the presented work and potential future experiments.

References

1. Rozenski J, Crain PF, McCloskey JA: **The RNA Modification Database: 1999 update**. *Nucleic Acids Res* 1999, **27**:196-197.

2. Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM: **MODOMICS: a database of RNA modification pathways**. *Nucleic Acids Res* 2006, **34**:D145-149.

3. Motorin Y, Grosjean H: **Chemical structures and classification of posttranscriptionally modified nucleotides in RNA**. In *Modification and editing of RNA*. Edited by Grosjean H, Benne R: ASM Press; 1998:543-549.

4. Limbach PA, Crain PF, McCloskey JA: **Summary: the modified nucleosides of RNA**. *Nucleic Acids Res* 1994, **22**:2183-2196.

5. Bachellerie JP, Cavaille J: **Guiding ribose methylation of rRNA**. *Trends Biochem Sci* 1997, **22**:257-261.

6. Grosjean H, Sprinzl M, Steinberg S: **Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies**. *Biochimie* 1995, **77**:139-141.

7. Hopper AK, Phizicky EM: **tRNA transfers to the limelight**. *Genes Dev* 2003, **17**:162-180.

8. Maden BE: **The numerous modified nucleotides in eukaryotic ribosomal RNA**. *Prog Nucleic Acid Res Mol Biol* 1990, **39**:241-303.

9. Massenet S, Mougin A, Branlant C: **Posttrnascriptional modifications in the U snRNAs**. In *Modification and Editing of RNA*. Edited by Grosjean H, Benne B: ASM Press; 1998:201-228.

10. Ofengand J: **Ribosomal RNA pseudouridines and pseudouridine synthases**. *FEBS Lett* 2002, **514**:17-25.

11. Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S: **Compilation of tRNA sequences and sequences of tRNA genes**. *Nucleic Acids Res* 1998, **26**:148-153.

12. Eichler DC, Craig N: **Processing of eukaryotic ribosomal RNA**. *Prog Nucleic Acid Res Mol Biol* 1994, **49**:197-239.

13. Venema J, Tollervey D: **Ribosome synthesis in Saccharomyces cerevisiae**. *Annu Rev Genet* 1999, **33**:261-311.

14. Bachellerie J-P, Cavaille J: **Small nucleolar RNAs guide the ribose methylations of eukaryotic rRNAs**. In *Modification and Editing of RNA*. Edited by Grosjean H, Benne B: ASM Press; 1998:255-272.

15. Ofengand J, Fournier MJ: **The pseudouridine residues of rRNA: Number, location, biosynthesis, and function**. In *Modification and Editing of RNA*. Edited by Grosjean H, Benne B: ASM Press; 1998:229-253.

16. Noon KR, Bruenger E, McCloskey JA: **Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus**. *J Bacteriol* 1998, **180**:2883-2888.

17. Omer AD, Ziesche S, Decatur WA, Fournier MJ, Dennis PP: **RNA-modifying machines in archaea**. *Mol Microbiol* 2003, **48**:617-629.

18. Massenet S, Ansmant I, Motorin Y, Branlant C: **The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea Sulfolobus acidocaldarius**. *FEBS Lett* 1999, **462**:94-100.

19. Muller S, Leclerc F, Behm-Ansmant I, Fourmann JB, Charpentier B, Branlant C: **Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs**. *Nucleic Acids Res* 2008.

20. Sprinzl M, Vassilenko KS: **Compilation of tRNA sequences and sequences of tRNA genes**. *Nucleic Acids Res* 2005, **33**:D139-140.

21. Balakin AG, Smith L, Fournier MJ: **The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions**. *Cell* 1996, **86**:823-834.

22. Venema J, Tollervey D: **Processing of pre-ribosomal RNA in Saccharomyces cerevisiae**. *Yeast* 1995, **11**:1629-1650.

23. Tycowski KT, You ZH, Graham PJ, Steitz JA: **Modification of U6 spliceosomal RNA is guided by other small RNAs**. *Mol Cell* 1998, **2**:629-638.

24. Ganot P, Jady BE, Bortolin ML, Darzacq X, Kiss T: **Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA**. *Mol Cell Biol* 1999, **19**:6906-6917.

25. Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T: **Cajal body-specific small nuclear RNAs: a novel class of 2'-O- methylation and pseudouridylation guide RNAs**. *EMBO J* 2002, **21**:2746-2756.

26. Kiss AM, Jady BE, Darzacq X, Verheggen C, Bertrand E, Kiss T: **A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains**. *Nucleic Acids Res* 2002, **30**:4643-4649.

27. Jady BE, Kiss T: **A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA**. *EMBO J* 2001, **20**:541-551.

28. Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A: **Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization**. *Proc Natl Acad Sci U S A* 2000, **97**:14311-14316.

29. Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J: **RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse**. *EMBO J* 2001, **20**:2943-2953.

30. Kishore S, Stamm S: **The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C**. *Science* 2006, **311**:230-232.

31. Cohn WE, Volkin E: **Nucleoside-5'-phosphates from ribonucleic acid**. *Nature* 1951, **167**:483-484.

32. Lane BG, Ofengand J, Gray MW: **Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins**. *Biochimie* 1995, **77**:7-15.

33. Terns MP, Terns RM: **Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin**. *Gene Expr* 2002, **10**:17-39.

34. Meier UT: **The many facets of H/ACA ribonucleoproteins**. *Chromosoma* 2005, **114**:1-14.

35. Gu X, Liu Y, Santi DV: **The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA**. *Proc Natl Acad Sci U S A* 1999, **96**:14270-14275.

36. Balakin AG, Smith L, Fournier MJ: **The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions**. *Cell* 1996, **86**:823-834.

37. Ganot P, Caizergues-Ferrer M, Kiss T: **The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation**. *Genes Dev* 1997, **11**:941-956.

38. Ni J, Tien AL, Fournier MJ: **Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA**. *Cell* 1997, **89**:565-573.

39. Liang XH, Liu L, Michaeli S: **Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA**. *J Biol Chem* 2001, **276**:40313-40318.

40. Liang XH, Xu YX, Michaeli S: **The spliced leader-associated RNA is a trypanosome-specific sn(o) RNA that has the potential to guide pseudouridine formation on the SL RNA**. *RNA* 2002, **8**:237-246.

41. Jin H, Loria JP, Moore PB: **Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA**. *Mol Cell* 2007, **26**:205-215.

42. Wu H, Feigon J: **H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification**. *Proc Natl Acad Sci U S A* 2007, **104**:6655-6660.

43. Meier UT, Blobel G: **NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria**. *J Cell Biol* 1994, **127**:1505-1514.

44. Lubben B, Fabrizio P, Kastner B, Luhrmann R: **Isolation and characterization of the small nucleolar ribonucleoprotein particle snR30 from Saccharomyces cerevisiae**. *J Biol Chem* 1995, **270**:11549-11554.

45. Ganot P, Bortolin ML, Kiss T: **Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs**. *Cell* 1997, **89**:799-809.

46. Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gelugne JP, Caizergues-Ferrer M: **Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs**. *EMBO J* 1998, **17**:7078-7090.

47. Dragon F, Pogacic V, Filipowicz W: **In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs**. *Mol Cell Biol* 2000, **20**:3037-3048.

48. Jiang W, Middleton K, Yoon HJ, Fouquet C, Carbon J: **An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules**. *Mol Cell Biol* 1993, **13**:4884-4893.

49. Koonin EV: **Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases**. *Nucleic Acids Res* 1996, **24**:2411-2415.

50. Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D: **The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase**. *Genes Dev* 1998, **12**:527-537.

51. Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J: **Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA**. *Mol Cell Biol* 1999, **19**:7461-7472.

52. Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I: **X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions**. *Nat Genet* 1998, **19**:32-38.

53. Leung AK, Lamond AI: **In vivo analysis of NHPX reveals a novel nucleolar localization pathway involving a transient accumulation in splicing speckles**. *J Cell Biol* 2002, **157**:615-629.

54. Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Luhrmann R: **Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure**. *RNA* 1998, **4**:1549-1568.

55. Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M: **Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p**. *Nucleic Acids Res* 2001, **29**:2733-2746.

56. Wang C, Meier UT: **Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins**. *EMBO J* 2004, **23**:1857-1867.

57. Girard JP, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D, Lapeyre B: **GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast**. *EMBO J* 1992, **11**:673-682.

58. Bagni C, Lapeyre B: **Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element**. *J Biol Chem* 1998, **273**:10868-10873.

59. Girard JP, Bagni C, Caizergues-Ferrer M, Amalric F, Lapeyre B: **Identification of a segment of the small nucleolar ribonucleoprotein-associated protein GAR1 that is sufficient for nucleolar accumulation**. *J Biol Chem* 1994, **269**:18499-18506.

60. Hamma T, Reichow SL, Varani G, Ferre-D'Amare AR: **The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs**. *Nat Struct Mol Biol* 2005, **12**:1101-1107.

61. Reichow SL, Varani G: **Nop10 is a conserved H/ACA snoRNP molecular adaptor**. *Biochemistry* 2008, **47**:6148-6156.

62. Khanna M, Wu H, Johansson C, Caizergues-Ferrer M, Feigon J: **Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA**. *RNA* 2006, **12**:40-52.

63. Bousquet-Antonelli C, Henry Y, G'Elugne J P, Caizergues-Ferrer M, Kiss T: **A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs**. *EMBO J* 1997, **16**:4770-4776.

64. Dez C, Henras A, Faucon B, Lafontaine D, Caizergues-Ferrer M, Henry Y: **Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p**. *Nucleic Acids Res* 2001, **29**:598-603.

65. Wang C, Query CC, Meier UT: **Immunopurified small nucleolar ribonucleoprotein particles pseudouridylate rRNA independently of their association with phosphorylated Nopp140**. *Mol Cell Biol* 2002, **22**:8457-8466.

66. Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M: **Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs**. *RNA* 2004, **10**:1704-1712.

67. Tollervey D: **Small nucleolar RNAs guide ribosomal RNA methylation**. *Science* 1996, **273**:1056-1057.

68. Matera AG, Terns RM, Terns MP: **Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs**. *Nat Rev Mol Cell Biol* 2007, **8**:209-220.

69. Schimmang T, Tollervey, D., Kern, H., Frank, R. and Hurt, E.C.: **A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability**. *EMBO J.* 1989, **8**:4015-4024.

70. Tyc K, and Steitz, J.A.: **U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus**. *EMBO J.* 1989, **8**:3113-3119.

71. Gautier T, Berges T, Tollervey D, Hurt E: **Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis**. *Mol Cell Biol* 1997, **17**:7088-7098.

72. Lyman SK, Gerace, L., and Baserga, S.J.: **Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins**. *RNA* 1999, **5**:1597-1604.

73. Lafontaine DL, Tollervey D: **Synthesis and assembly of the box C+D small nucleolar RNPs**. *Mol Cell Biol* 2000, **20**:2650-2659.

74. Newman DR, Kuhn JF, Shanab GM, Maxwell ES: **Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription**. *RNA* 2000, **6**:861-879.

75. Caffarelli E, Fatica A, Prislei S, De Gregorio E, Fragapane P, Bozzoni I: **Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA**. *EMBO J* 1996, **15**:1121-1131.

76. Bachellerie J-P, and Cavaille, J.: **Small nucleolar RNAs guide ribose methylation of eukaryotic rRNAs**. In *Modification and Editing of RNA*. Edited by Grosjean H, and Benne, R.: American Society for Microbiology; 1998:255-272.

77. Kiss T: **Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs**. *EMBO J* 2001, **20**:3617-3622.

78. Bachellerie JP, Cavaille J, Huttenhofer A: **The expanding snoRNA world**. *Biochimie* 2002, **84**:775-790.

79. Terns MP, Terns RM: **Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin**. *Gene Expr* 2002, **10**:17-39.

80. Cavaille J, Nicoloso M, Bachellerie JP: **Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides**. *Nature* 1996, **383**:732-735.

81. Caffarelli E, Losito, M., Giorgi, C., Fatica, A., and Bozzoni, I.: ***In vivo* identification of factors interacting with the conserved elements of box C/D small nucleolar RNAs**. *J. Cell. Biol.* 1998, **18**:1023–1028.

82. Samarsky DA, Fournier MJ, Singer RH, Bertrand E: **The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization**. *EMBO J* 1998, **17**:3747-3757.

83. Klein DJ, Schmeing TM, Moore PB, Steitz TA: **The kink-turn: a new RNA secondary structure motif**. *EMBO J* 2001, **20**:4214-4221.

84. Niewmierzycka A, Clarke S: **S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase**. *J Biol Chem* 1999, **274**:814-824.

85. Wang H, Boisvert D, Kim KK, Kim R, Kim SH: **Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution**. *EMBO J* 2000, **19**:317-323.

86. Tollervey D, Lehtonen, H., Jansen, R., Kern, H., and Hurt, E.C.: **Temperature sensitive mutations demonestrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly**. *Cell* 1993, **72**:443-457.

87. Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, Ficner R: **Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment**. *Mol Cell* 2000, **6**:1331-1342.

88. Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Luhrmann R: **Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA**. *EMBO J* 1999, **18**:6119-6133.

89. Lafontaine DL, Tollervey D: **Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability**. *RNA* 1999, **5**:455-467.

90. Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R: **A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP**. *Cell* 2000, **103**:457-466.

91. Verheggen C, Mouaikel J, Thiry M, Blanchard JM, Tollervey D, Bordonne R, Lafontaine DL, Bertrand E: **Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain**. *EMBO J* 2001, **20**:5480-5490.

92. Cahill NM, Friend K, Speckmann W, Li ZH, Terns RM, Terns MP, Steitz JA: **Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP**. *EMBO J* 2002, **21**:3816-3828.

93. Szewczak LB, DeGregorio SJ, Strobel SA, Steitz JA: **Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP**. *Chem Biol* 2002, **9**:1095-1107.

94. Meier UT, Blobel G: **Nopp140 shuttles on tracks between nucleolus and cytoplasm**. *Cell* 1992, **70**:127-138.

95. Isaac C, Yang Y, Meier UT: **Nopp140 functions as a molecular link between the nucleolus and the coiled bodies**. *J Cell Biol* 1998, **142**:319-329.

96. Yang Y, Isaac C, Wang C, Dragon F, Pogacic V, Meier UT: **Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140**. *Mol Biol Cell* 2000, **11**:567-577.

97. Pellizzoni L, Charroux B, Dreyfuss G: **SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins**. *Proc Natl Acad Sci U S A* 1999, **96**:11167-11172.

98. Pellizzoni L, Kataoka N, Charroux B, Dreyfuss G: **A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing**. *Cell* 1998, **95**:615-624.

99. Pellizzoni L, Baccon J, Charroux B, Dreyfuss G: **The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1**. *Curr Biol* 2001, **11**:1079-1088.

100. Jones KW, Gorzynski K, Hales CM, Fischer U, Badbanchi F, Terns RM, Terns MP: **Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin**. *J Biol Chem* 2001, **276**:38645-38651.

101. Whitehead SE, Jones KW, Zhang X, Cheng X, Terns RM, Terns MP: **Determinants of the interaction of the spinal muscular atrophy disease protein SMN with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GAR1**. *J Biol Chem* 2002, **277**:48087-48093.

102. King TH, Decatur WA, Bertrand E, Maxwell ES, Fournier MJ: **A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins**. *Mol Cell Biol* 2001, **21**:7731-7746.

103. Kanemaki M, Makino Y, Yoshida T, Kishimoto T, Koga A, Yamamoto K, Yamamoto M, Moncollin V, Egly JM, Muramatsu M, et al.: **Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB**. *Biochem Biophys Res Commun* 1997, **235**:64-68.

104. Dez C, Noaillac-Depeyre J, Caizergues-Ferrer M, Henry Y: **Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs**. *Mol Cell Biol* 2002, **22**:7053-7065.

105. Fatica A, Dlakic M, Tollervey D: **Naf1 p is a box H/ACA snoRNP assembly factor**. *RNA* 2002, **8**:1502-1514.

106. Yang PK, Rotondo G, Porras T, Legrain P, Chanfreau G: **The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis**. *J Biol Chem* 2002, **277**:45235-45242.

107. Tremblay A, Lamontagne B, Catala M, Yam Y, Larose S, Good L, Elela SA: **A physical interaction between Gar1p and Rnt1pi is required for the nuclear import of H/ACA small nucleolar RNA-associated proteins**. *Mol Cell Biol* 2002, **22**:4792-4802.

108. Olsen GJ, Woese CR: **Archaeal genomics: an overview**. *Cell* 1997, **89**:991-994.

109. Dennis PP, Omer A: **Small non-coding RNAs in Archaea**. *Curr Opin Microbiol* 2005, **8**:685-694.

110. Clouet d'Orval B, Bortolin ML, Gaspin C, Bachellerie JP: **Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of**

two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* 2001, **29**:4518-4529.

111. Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A: **Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea**. *Nucleic Acids Res* 2003, **31**:869-877.

112. Watanabe Y, Gray MW: **Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria**. *Nucleic Acids Res* 2000, **28**:2342-2352.

113. Ferre-D'Amare AR: **RNA-modifying enzymes**. *Curr Opin Struct Biol* 2003, **13**:49-55.

114. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A: **Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus**. *Proc Natl Acad Sci U S A* 2002, **99**:7536-7541.

115. Klein RJ, Misulovin Z, Eddy SR: **Noncoding RNA genes identified in AT-rich hyperthermophiles**. *Proc Natl Acad Sci U S A* 2002, **99**:7542-7547.

116. Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP: **RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP**. *Genes Dev* 2005, **19**:1238-1248.

117. Hamma T, Ferre-D'Amare AR: **Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution**. *Structure (Camb)* 2004, **12**:893-903.

118. Charpentier B, Muller S, Branlant C: **Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation**. *Nucleic Acids Res* 2005, **33**:3133-3144.

119. Manival X, Charron C, Fourmann JB, Godard F, Charpentier B, Branlant C: **Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity**. *Nucleic Acids Res* 2006, **34**:826-839.

120. Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H: **Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita**. *Mol Cell* 2006, **21**:249-260.

121. Li L, Ye K: **Crystal structure of an H/ACA box ribonucleoprotein particle**. *Nature* 2006, **443**:302-307.

122. Hoang C, Ferre-D'Amare AR: **Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme**. *Cell* 2001, **107**:929-939.

123. Liang B, Xue S, Terns RM, Terns MP, Li H: **Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex**. *Nat Struct Mol Biol* 2007.

124. Watanabe M, Nameki N, Matsuo-Takasaki M, Nishimura S, Okada N: **tRNA recognition of tRNA-guanine transglycosylase from a hyperthermophilic archaeon, Pyrococcus horikoshii**. *J Biol Chem* 2001, **276**:2387-2394.

125. Ishitani R, Nureki O, Nameki N, Okada N, Nishimura S, Yokoyama S: **Alternative tertiary sturcture of tRNA for recognition by a posttranscriptional modification enzyme**. *Cell* 2003, **113**:383-394.

126. Hamma T, Ferre-D'Amare AR: **Pseudouridine synthases**. *Chem Biol* 2006, **13**:1125-1135.

127. Youssef OA, Terns RM, Terns MP: **Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP**. *Nucleic Acids Res* 2007, **35**:6196-6206.

128. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea**. *Science* 2000, **288**:517-522.

129. Amiri KA: **Fibrillarin-like proteins occur in the domain Archaea**. *J Bacteriol* 1994, **176**:2124-2127.

130. Kuhn JF, Tran EJ, Maxwell ES: **Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein**. *Nucleic Acids Res* 2002, **30**:931-941.

131. Omer AD, Ziesche S, Ebhardt H, Dennis PP: **In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex**. *Proc Natl Acad Sci U S A* 2002, **99**:5289-5294.

132. Tran EJ, Zhang X, Maxwell ES: **Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs**. *EMBO J.* 2003, **22**:3930-3940.

133. Filipowicz W, Pogacic V: **Biogenesis of small nucleolar ribonucleoproteins**. *Curr Opin Cell Biol* 2002, **14**:319-327.

134. Tollervey D, Kiss T: **Function and synthesis of small nucleolar RNAs**. *Curr Opin Cell Biol* 1997, **9**:337-342.

135. Weinstein LB, Steitz JA: **Guided tours: from precursor snoRNA to functional snoRNP**. *Curr Opin Cell Biol* 1999, **11**:378-384.

136. Maxwell ES, Fournier MJ: **The small nucleolar RNAs**. *Annu Rev Biochem* 1995, **64**:897-934.

137. Yu YT, Terns RM, Terns MP: **Mechanisms and functions of RNA-guided RNA modification**. In *Fine-tuning of RNA functions by modification and editing*. Edited by Grosjean H: Topics in Current Genetics; 2005:223-262. vol 12.

138. Tycowski KT, Steitz JA: **Non-coding snoRNA host genes in Drosophila: expression strategies for modification guide snoRNAs**. *Eur J Cell Biol* 2001, **80**:119-125.

139. Tycowski KT, Shu MD, Steitz JA: **A mammalian gene with introns instead of exons generating stable RNA products**. *Nature* 1996, **379**:464-466.

140. Petfalski E, Dandekar T, Henry Y, Tollervey D: **Processing of the precursors to small nucleolar RNAs and rRNAs requires common components**. *Mol Cell Biol* 1998, **18**:1181-1189.

141. Ooi SL, Samarsky DA, Fournier MJ, Boeke JD: **Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA**. *RNA* 1998, **4**:1096-1110.

142. Villa T, Ceradini F, Presutti C, Bozzoni I: **Processing of the intron-encoded U18 small nucleolar RNA in the yeast Saccharomyces cerevisiae relies on both exo- and endonucleolytic activities**. *Mol Cell Biol* 1998, **18**:3376-3383.

143. Yu YT, Scharl EC, Smith CM, Steitz JA: **The growing world of small nuclear riobnucleoproteins**. In *The RNA world, 2nd edn.* Edited by Gersteland RF, Cech TR, Atkins JF: Cold Spring Harbor laboratory Press, Cold Spring Harbor; 1999:487-524.

144. Huang ZP, Zhou H, Liang D, Qu LH: **Different expression strategy: multiple intronic gene clusters of box H/ACA snoRNA in Drosophila melanogaster**. *J Mol Biol* 2004, **341**:669-683.

145. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea**. *Science* 2000, **288**:517-522.

146. Gaspin C, Cavaille J, Erauso G, Bachellerie JP: **Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes**. *J Mol Biol* 2000, **297**:895-906.

147. Singh SK, Gurha P, Tran EJ, Maxwell ES, Gupta R: **Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA**. *J Biol Chem* 2004, **279**:47661-47671.

148. Starostina NG, Marshburn S, Johnson LS, Eddy SR, Terns RM, Terns MP: **Circular box C/D RNAs in Pyrococcus furiosus**. *Proc Natl Acad Sci U S A* 2004, **101**:14097-14101.

149. Venema J, and Tollervey, D.: **Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae***. *Yeast* 1995, **11**:1629-1650.

150. King TH, Liu B, McCully RR, Fournier MJ: **Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center**. *Mol Cell* 2003, **11**:425-435.

151. Brimacombe R, Mitchell P, Osswald M, Stade K, Bochkariov D: **Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA**. *FASEB J* 1993, **7**:161-167.

152. Maden BE: **Eukaryotic rRNA methylation: the calm before the Sno storm**. *Trends Biochem Sci* 1998, **23**:447-450.

153. Decatur WA, Fournier MJ: **rRNA modifications and ribosome function**. *Trends Biochem Sci* 2002, **27**:344-351.

154. Davis DR: **Biophysical and conformation properties of modified nucleotides in RNA**. In *In modification and editing of RNA*. Edited by Grosjean H, Benne B: ASM Press; 1998:85-102.

155. Yu YT, Shu MD, Steitz JA: **Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing**. *EMBO J* 1998, **17**:5783-5795.

156. Noon KR, Bruenger, E., and McCloskey, J.A.: **Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus***. *J. Bacteriol.* 1998, **180**:2883-2888.

157. Yu YT, Scharl EC, Smith CM, Steitz JA: **The growing world of small nuclear ribonucleoproteins**. In *The RNA world, 2nd edn*. Edited by Gesteland RF, Cech TR, Atkins JF: Gold Spring Harbor Laboratory Press, Cold Spring Harbor; 1999:467-524.

158. Yi-Brunozzi HY, Easterwood LM, Kamilar GM, Beal PA: **Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2**. *Nucleic Acids Res* 1999, **27**:2912-2917.

159. Mitchell JR, Cheng J, Collins K: **A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end**. *Mol Cell Biol* 1999, **19**:567-576.

160. Morrissey JP, Tollervey D: **Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis**. *Mol Cell Biol* 1993, **13**:2469-2477.

161. Speckmann W, Narayanan A, Terns R, Terns MP: **Nuclear retention elements of U3 small nucleolar RNA**. *Mol Cell Biol* 1999, **19**:8412-8421.

162. Vulliamy T, Beswick R, Kirwan M, Marrone A, Digweed M, Walne A, Dokal I: **Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita**. *Proc Natl Acad Sci U S A* 2008, **105**:8073-8078.

163. Tan EM, Kunkel HG: **Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus**. *J Immunol* 1966, **96**:464-471.

164. Luhrmann R, Kastner B, Bach M: **Structure of spliceosomal snRNPs and their role in pre-mRNA splicing**. *Biochim Biophys Acta* 1990, **1087**:265-292.

165. Chu JL, Elkon KB: **The small nuclear ribonucleoproteins, SmB and B', are products of a single gene**. *Gene* 1991, **97**:311-312.

166. Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Luhrmann R: **snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions**. *EMBO J* 1995, **14**:2076-2088.

167. Camasses A, Bragado-Nilsson E, Martin R, Seraphin B, Bordonne R: **Interactions within the yeast Sm core complex: from proteins to amino acids**. *Mol Cell Biol* 1998, **18**:1956-1966.

168. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Luhrmann R, Li J, Nagai K: **Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs**. *Cell* 1999, **96**:375-387.

169. Kambach C, Walke S, Nagai K: **Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles**. *Curr Opin Struct Biol* 1999, **9**:222-230.

170. Kastner B, Bach M, Luhrmann R: **Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family**. *Proc Natl Acad Sci U S A* 1990, **87**:1710-1714.

171. Mattaj IW: **Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding**. *Cell* 1986, **46**:905-911.

172. Hamm J, Darzynkiewicz E, Tahara SM, Mattaj IW: **The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal**. *Cell* 1990, **62**:569-577.

173. Palacios I, Hetzer M, Adam SA, Mattaj IW: **Nuclear import of U snRNPs requires importin beta**. *EMBO J* 1997, **16**:6783-6792.

174. Jarmolowski A, Mattaj IW: **The determinants for Sm protein binding to Xenopus U1 and U5 snRNAs are complex and non-identical**. *EMBO J* 1993, **12**:223-232.

175. Raker VA, Hartmuth K, Kastner B, Luhrmann R: **Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner**. *Mol Cell Biol* 1999, **19**:6554-6565.

176. Will CL, Luhrmann R: **Spliceosomal UsnRNP biogenesis, structure and function**. *Curr Opin Cell Biol* 2001, **13**:290-301.

177. Stark H, Dube P, Luhrmann R, Kastner B: **Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle**. *Nature* 2001, **409**:539-542.

178. Seraphin B: **Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs**. *EMBO J* 1995, **14**:2089-2098.

179. Albrecht M, Lengauer T: **Novel Sm-like proteins with long C-terminal tails and associated methyltransferases**. *FEBS Lett* 2004, **569**:18-26.

180. Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Seraphin B: **Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin**. *EMBO J* 1999, **18**:3451-3462.

181. Newman A: **RNA splicing**. *Curr Biol* 1998, **8**:R903-905.

182. Staley JP, Guthrie C: **Mechanical devices of the spliceosome: motors, clocks, springs, and things**. *Cell* 1998, **92**:315-326.

183. Seto AG, Zaug AJ, Sobel SG, Wolin SL, Cech TR: **Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle**. *Nature* 1999, **401**:177-180.

184. Galli G, Hofstetter H, Stunnenberg HG, Birnstiel ML: **Biochemical complementation with RNA in the Xenopus oocyte: a small RNA is required for the generation of 3' histone mRNA termini**. *Cell* 1983, **34**:823-828.

185. Schaufele F, Gilmartin GM, Bannwarth W, Birnstiel ML: **Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA**. *Nature* 1986, **323**:777-781.

186. Strub K, Galli G, Busslinger M, Birnstiel ML: **The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing**. *EMBO J* 1984, **3**:2801-2807.

187. Bouveret E, Rigaut G, Shevchenko A, Wilm M, Seraphin B: **A Sm-like protein complex that participates in mRNA degradation**. *EMBO J* 2000, **19**:1661-1671.

188. Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R: **Yeast Sm-like proteins function in mRNA decapping and decay**. *Nature* 2000, **404**:515-518.

189. Fernandez CF, Pannone BK, Chen X, Fuchs G, Wolin SL: **An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5**. *Mol Biol Cell* 2004, **15**:2842-2852.

190. Kufel J, Allmang C, Verdone L, Beggs J, Tollervey D: **A complex pathway for 3' processing of the yeast U3 snoRNA**. *Nucleic Acids Res* 2003, **31**:6788-6797.

191. Tomasevic N, Peculis BA: **Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence**. *Mol Cell Biol* 2002, **22**:4101-4112.

192. Kufel J, Allmang C, Verdone L, Beggs JD, Tollervey D: **Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p**. *Mol Cell Biol* 2002, **22**:5248-5256.

193. Achsel T, Brahms H, Kastner B, Bachi A, Wilm M, Luhrmann R: **A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro**. *EMBO J* 1999, **18**:5789-5802.

194. Mayes AE, Verdone L, Legrain P, Beggs JD: **Characterization of Sm-like proteins in yeast and their association with U6 snRNA**. *EMBO J* 1999, **18**:4321-4331.

195. Vidal VP, Verdone L, Mayes AE, Beggs JD: **Characterization of U6 snRNA-protein interactions**. *RNA* 1999, **5**:1470-1481.

196. Kufel J, Allmang C, Petfalski E, Beggs J, Tollervey D: **Lsm Proteins are required for normal processing and stability of ribosomal RNAs**. *J Biol Chem* 2003, **278**:2147-2156.

197. Peculis BA, Steitz JA: **Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte**. *Cell* 1993, **73**:1233-1245.

198. Peculis BA: **snoRNA nuclear import and potential for cotranscriptional function in pre-rRNA processing**. *RNA* 2001, **7**:207-219.

199. He W, Parker R: **The yeast cytoplasmic LsmI/Pat1p complex protects mRNA 3' termini from partial degradation**. *Genetics* 2001, **158**:1445-1455.

200. Pillai RS, Will CL, Luhrmann R, Schumperli D, Muller B: **Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein**. *EMBO J* 2001, **20**:5470-5479.

201. Naidoo N, Harrop SJ, Sobti M, Haynes PA, Szymczyna BR, Williamson JR, Curmi PM, Mabbutt BC: **Crystal structure of Lsm3 octamer from Saccharomyces cerevisiae:**

implications for Lsm ring organisation and recruitment. *J Mol Biol* 2008, **377**:1357-1371.

202. Franze de Fernandez MT, Eoyang L, August JT: **Factor fraction required for the synthesis of bacteriophage Qbeta-RNA**. *Nature* 1968, **219**:588-590.

203. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, et al.: **Structure of Pseudomonas aeruginosa Hfq protein**. *Acta Crystallogr D Biol Crystallogr* 2005, **61**:141-146.

204. Sauter C, Basquin J, Suck D: **Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli**. *Nucleic Acids Res* 2003, **31**:4091-4098.

205. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG: **Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein**. *EMBO J* 2002, **21**:3546-3556.

206. Folichon M, Allemand F, Regnier P, Hajnsdorf E: **Stimulation of poly(A) synthesis by Escherichia coli poly(A)polymerase I is correlated with Hfq binding to poly(A) tails**. *FEBS J* 2005, **272**:454-463.

207. Folichon M, Arluison V, Pellegrini O, Huntzinger E, Regnier P, Hajnsdorf E: **The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation**. *Nucleic Acids Res* 2003, **31**:7302-7310.

208. Hajnsdorf E, Regnier P: **Host factor Hfq of Escherichia coli stimulates elongation of poly(A) tails by poly(A) polymerase I**. *Proc Natl Acad Sci U S A* 2000, **97**:1501-1505.

209. Mohanty BK, Maples VF, Kushner SR: **The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli**. *Mol Microbiol* 2004, **54**:905-920.

210. Vytvytska O, Moll I, Kaberdin VR, von Gabain A, Blasi U: **Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding**. *Genes Dev* 2000, **14**:1109-1118.

211. Gottesman S: **The small RNA regulators of Escherichia coli: roles and mechanisms***. *Annu Rev Microbiol* 2004, **58**:303-328.

212. Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G: **The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein**. *EMBO J* 1998, **17**:6061-6068.

213. Toro I, Thore S, Mayer C, Basquin J, Seraphin B, Suck D: **RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex**. *EMBO J* 2001, **20**:2293-2303.

214. Achsel T, Stark H, Luhrmann R: **The Sm domain is an ancient RNA-binding motif with oligo(U) specificity**. *Proc Natl Acad Sci U S A* 2001, **98**:3685-3689.

215. Toro I, Basquin J, Teo-Dreher H, Suck D: **Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus**. *J Mol Biol* 2002, **320**:129-142.

216. Thore S, Mayer C, Sauter C, Weeks S, Suck D: **Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya**. *J Biol Chem* 2003, **278**:1239-1247.

217. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D: **Structure and assembly of an augmented Sm-like archaeal protein 14-mer**. *Proc Natl Acad Sci U S A* 2003, **100**:4539-4544.

218. Nielsen JS, Boggild A, Andersen CB, Nielsen G, Boysen A, Brodersen DE, Valentin-Hansen P: **An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii**. *RNA* 2007, **13**:2213-2223.

219. Collins BM, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC: **Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs**. *J Mol Biol* 2001, **309**:915-923.

Figure 1.1: Conserved secondary structure of H/ACA guide RNA and target modification. The secondary structures of A) Eukaryotic and B) Archaeal H/ACA guide RNAs are shown. Conserved H and ACA (ANA in archaeal RNA) elements are boxed and colored red. Target RNAs (blue) are base paired to H/ACA guide sequences. The unpaired nucleotide (N, any nucleotide) and pseudouridine (ψ) of the target RNA are colored blue and magenta, respectively. Consensus secondary structure and sequence of the kink (k)-turn motif characteristic of archaeal H/ACA RNA is shown in a green box. C) Isomerization of uridine to pseudouridine

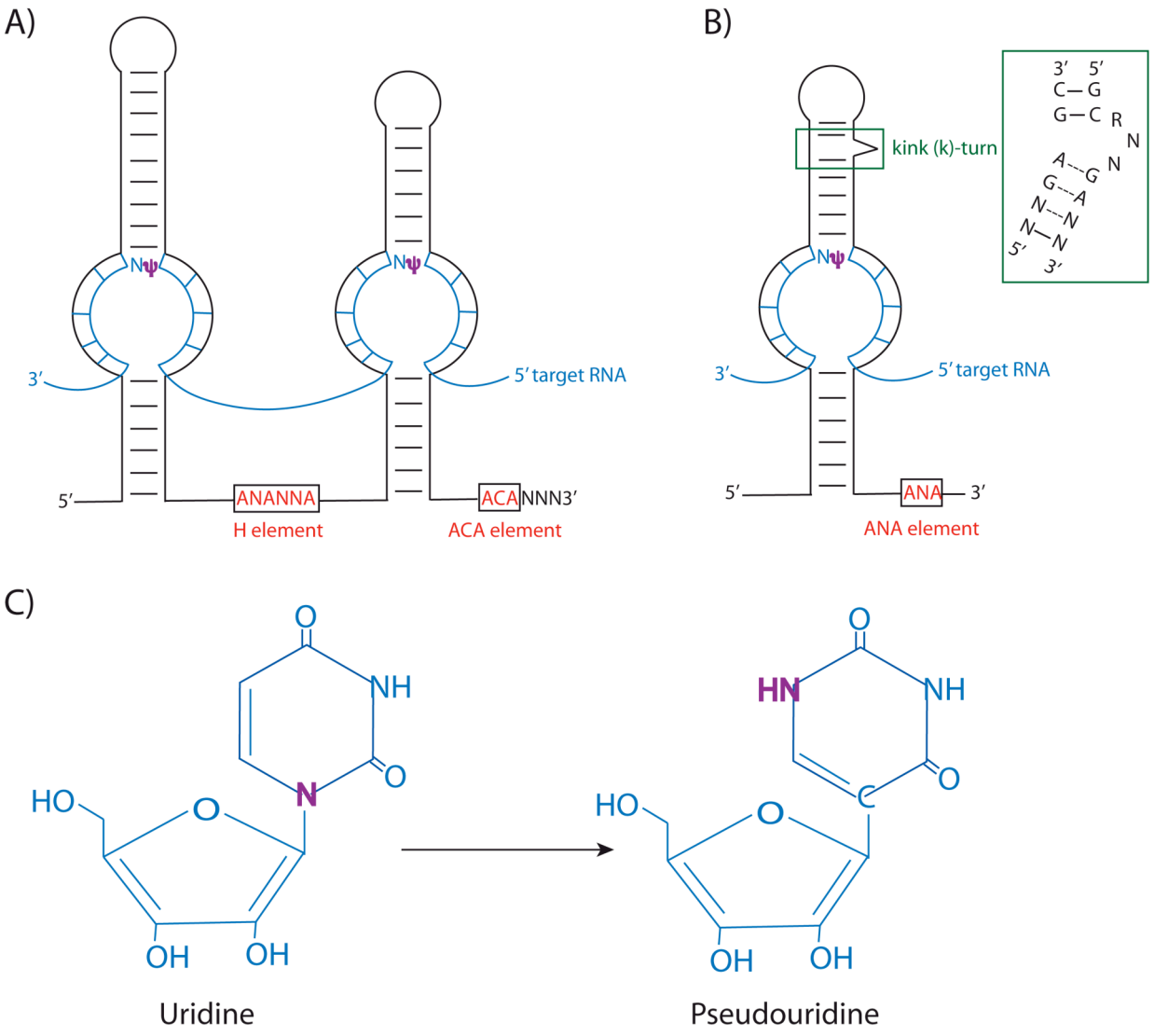
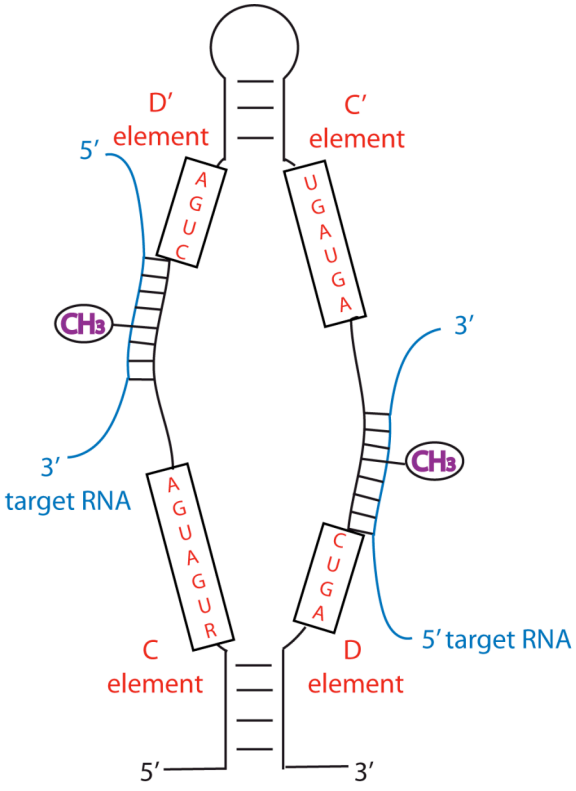
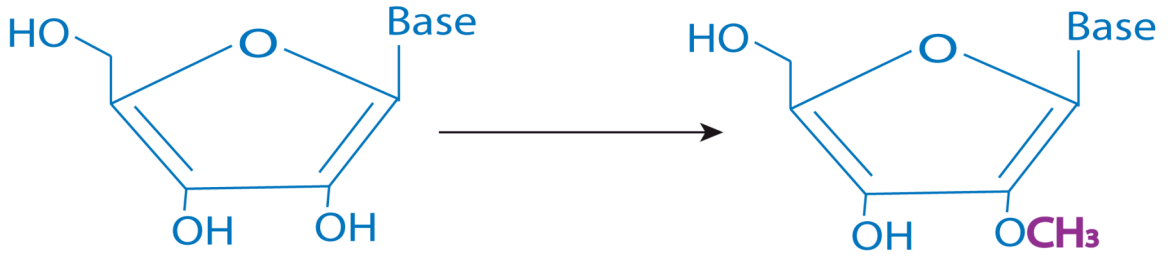

A)
Nψ
Nψ
3′
5′ target RNA
5′
ANANNA
ACA NNN3′
H element
ACA element
B)
kink (k)-turn
3′ 5′
C—G
G—C
R
N
A
G
N
G
A
N
N
N
N
5′
3′
Nψ
3′
5′ target RNA
5′
ANA
3′
ANA element
C)
O
NH
O
N
HO
O
OH
OH
Uridine
O
HN
NH
O
C
HO
O
OH
OH
Pseudouridine

Figure 1.2: Conserved secondary structure and sequence elements of C/D guide RNA. Secondary structure of eukaryotic and archaeal C/D guide RNA (same structure) is shown. Terminal C and D elements and internal C' and D' elements are colored in red. The target RNAs (blue) are base paired to C/D guide sequences located upstream of D or D' elements. The target nucleotide in the target RNA is methylated (CH_3) shown in magenta. B) Methylation of the 2'-hydroxyl group of the target nucleotide.

A)



B)



2′-*O*-methylated nucleotide

Figure 1.3: Ribbon representation of archaeal Cbf5/Gar1/Nop10 heterotrimeric structure (PDB # 2EY4) [120]. Gar1 and Nop10 bind to the catalytic domain of Cbf5 independently of one another. Different colors are used for the Cbf5 catalytic domain (blue), Cbf5 PUA domain (cyan), Cbf5 N-terminal tail (brown), Gar1 (orange), and Nop10 (yellow). The Zn^{2+} ion is represented by a red ball.

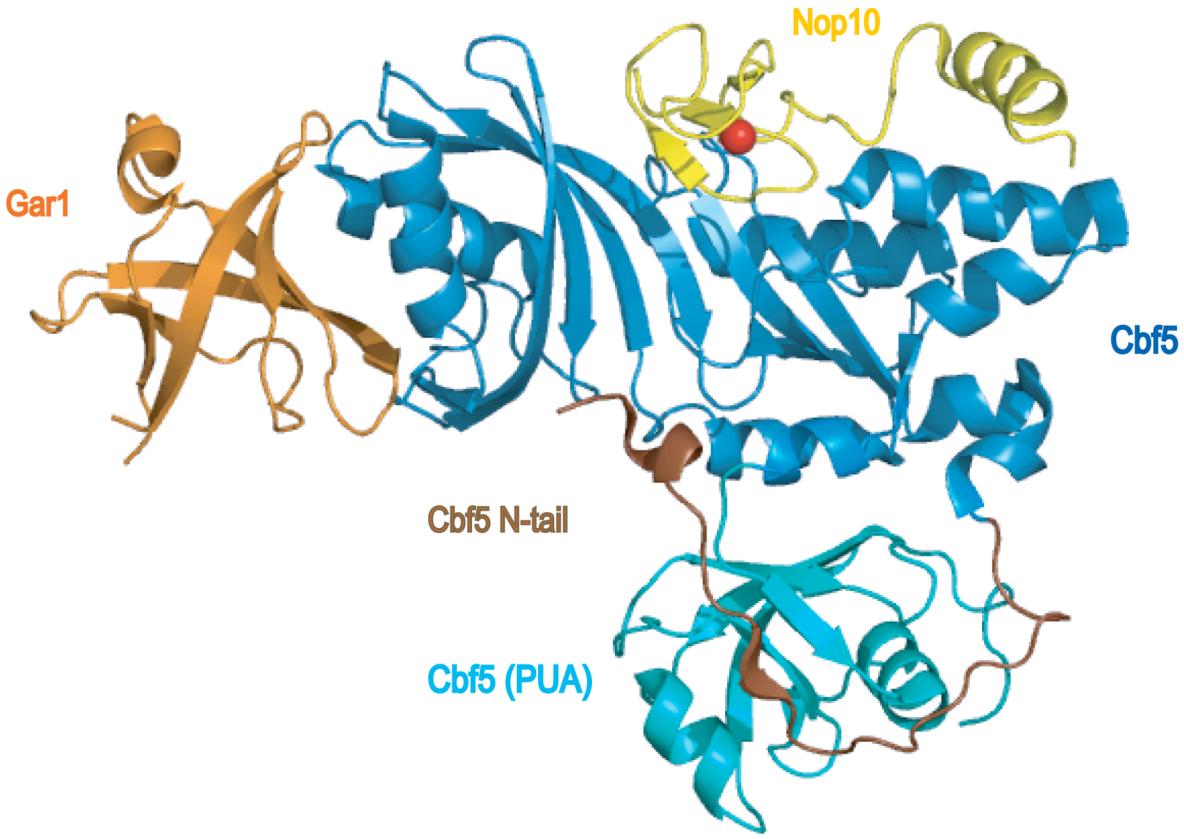

Nop10
Gar1
Cbf5
Cbf5 N-tail
Cbf5 (PUA)

Figure 1.4: Structural comparison of Cbf5 and TruB. *P. furiosus* Cbf5 (green) and *E. coli* TruB (grey) superimpose closely. Secondary structure elements found in TruB but not Cbf5 are shown in black. β7 and β10 of Cbf5 are marked by blue arrows. The catalytic aspartate residues of Cbf5 (magenta) and TruB (red) are shown. Figure adapted from [120].

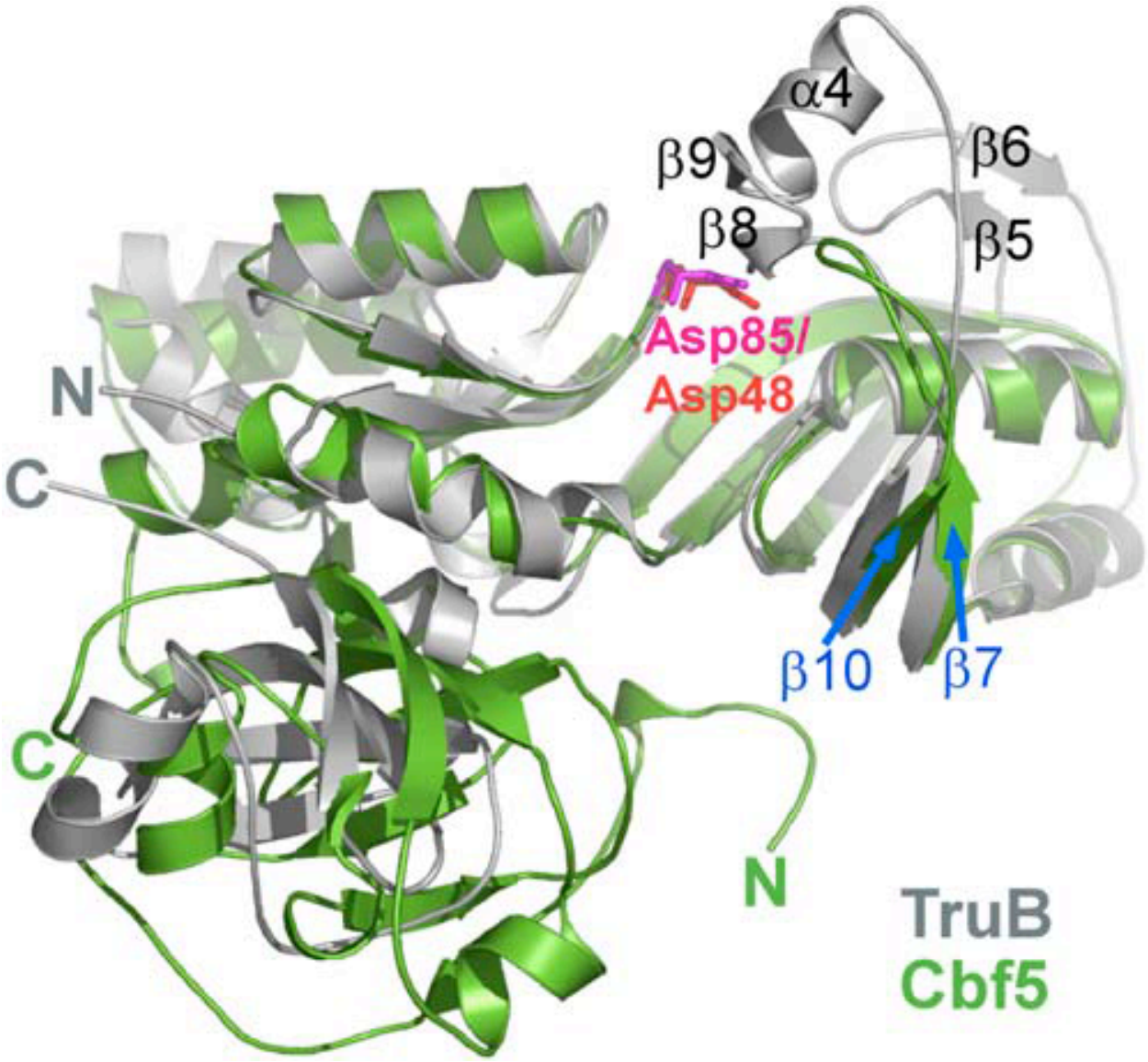

α4
β9
β6
β8
β5
Asp85/
Asp48
N
C
β10
β7
C
N
TruB
Cbf5

Figure 1.5: Crystal structure of H/ACA RNP holoenzyme (PDB # 2HVY) [121]. The catalytic domain of Cbf5 (blue) binds to the pseudouridylation pocket and the upper stem of the guide RNA (grey). Cbf5 PUA domain (cyan) binds to the lower stem and ACA (brown) of the guide RNA. Gar1 (orange) binds to the Cbf5 catalytic domain. Nop10 (yellow) binds to the Cbf5 catalytic domain, L7Ae (green), and the upper stem of the guide RNA. L7Ae binds to the k-turn motif of the guide RNA. For clarity the cartoon model is shown with the same colors.

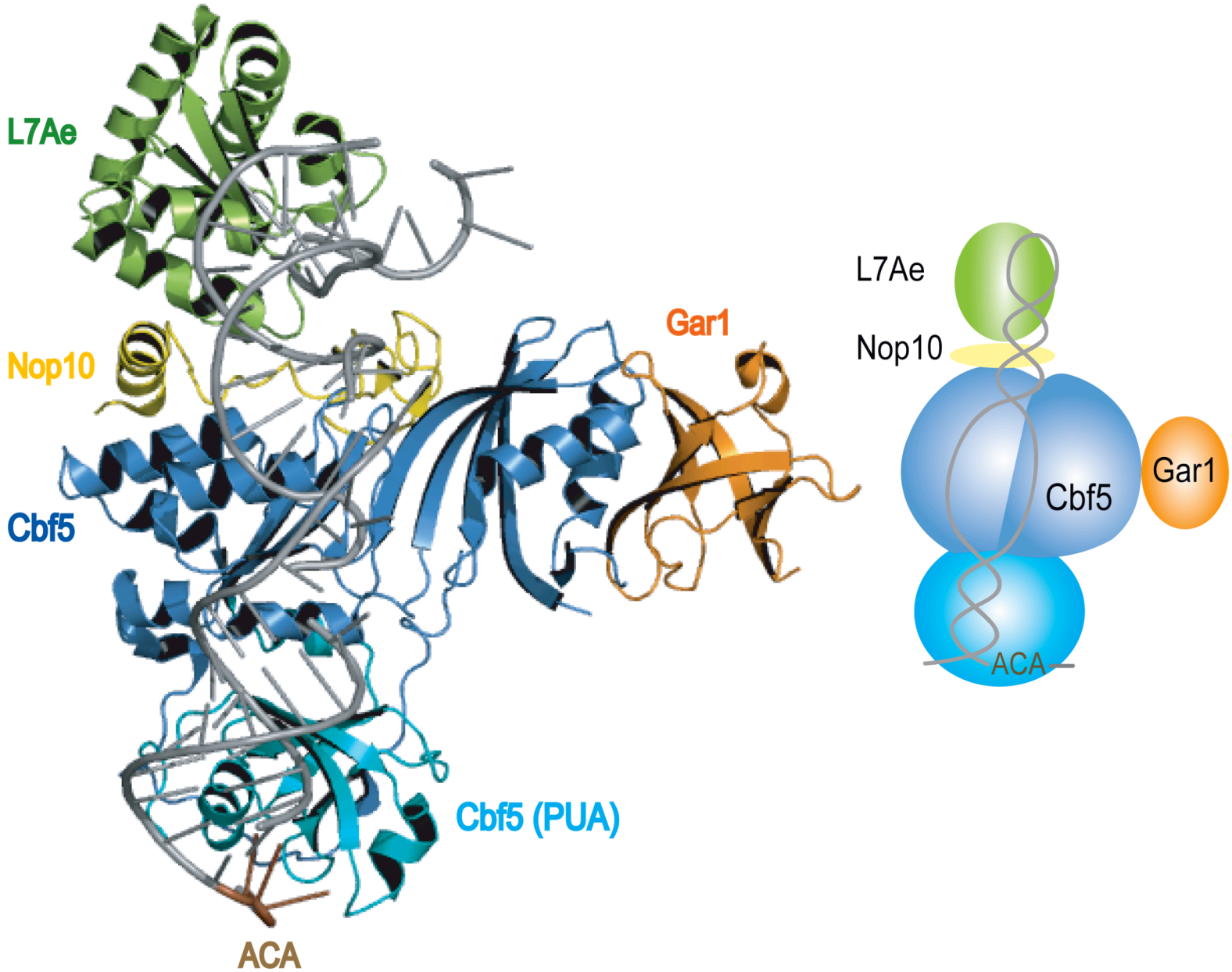

L7Ae
Nop10
Cbf5
Gar1
Cbf5 (PUA)
ACA
L7Ae
Nop10
Cbf5
Gar1
ACA

Figure 1.6: Crystal structure of H/ACA guide RNP/substrate RNA sub-complex (PDB # 2RFK) [123]. The target RNA is shown in magenta with the target uridine in red. Proteins and guide RNA are colored as in figure 1.5.

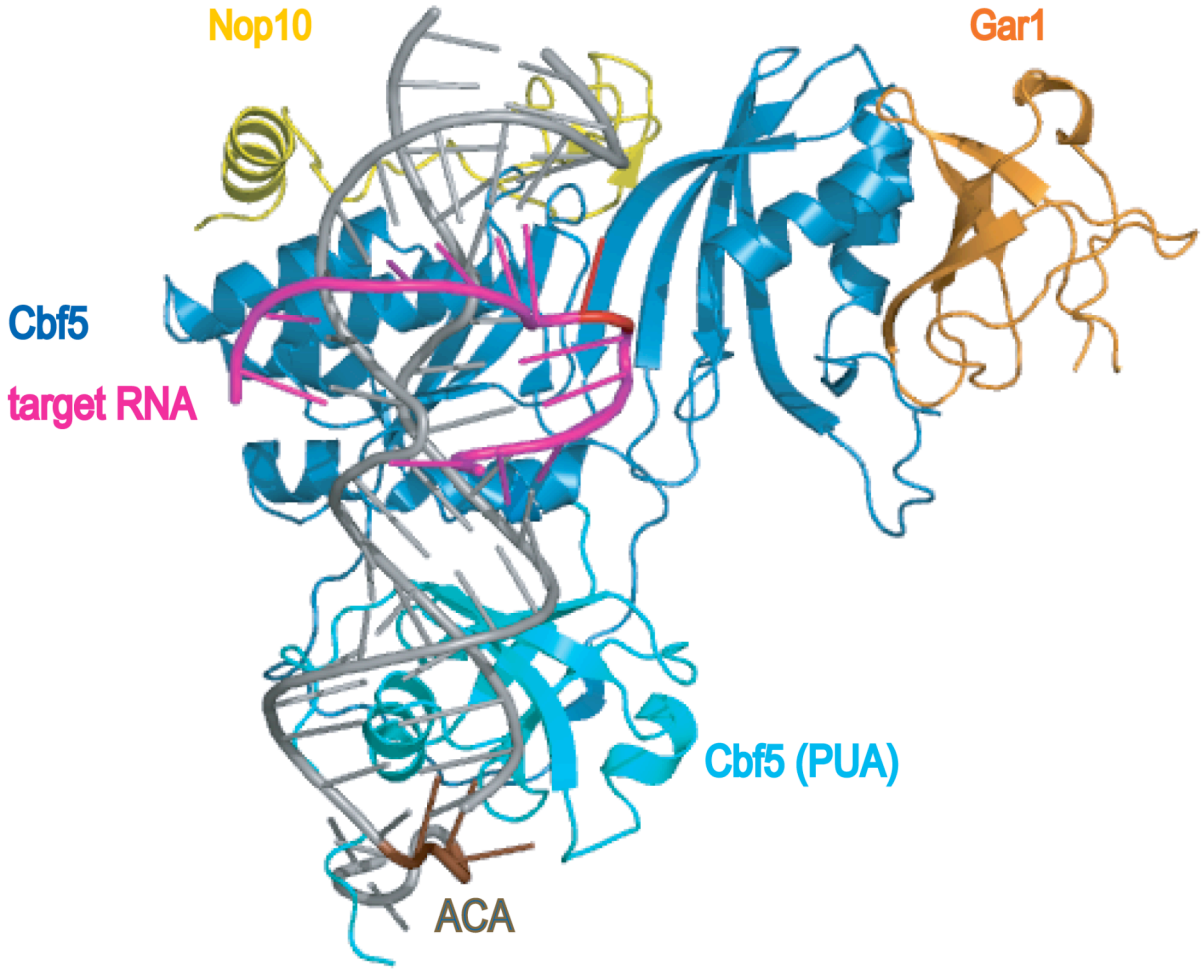
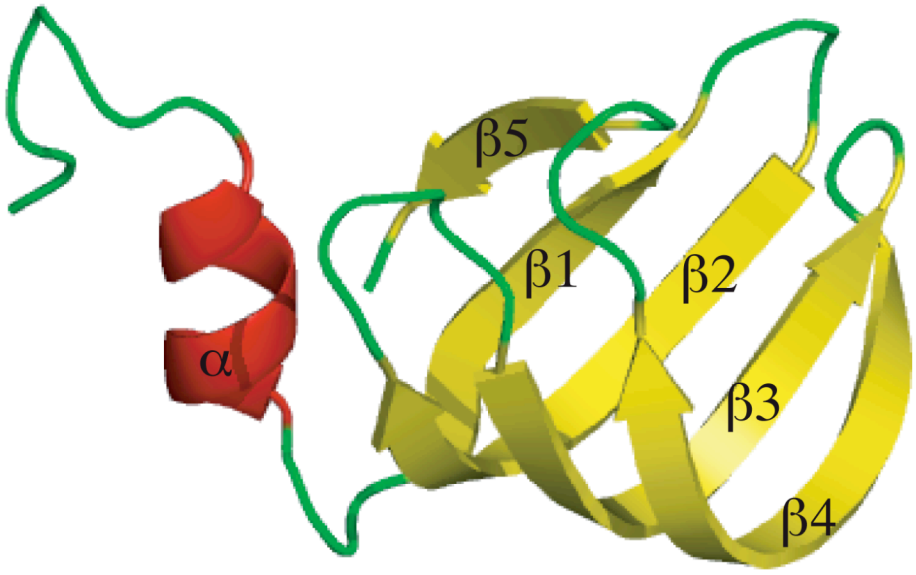
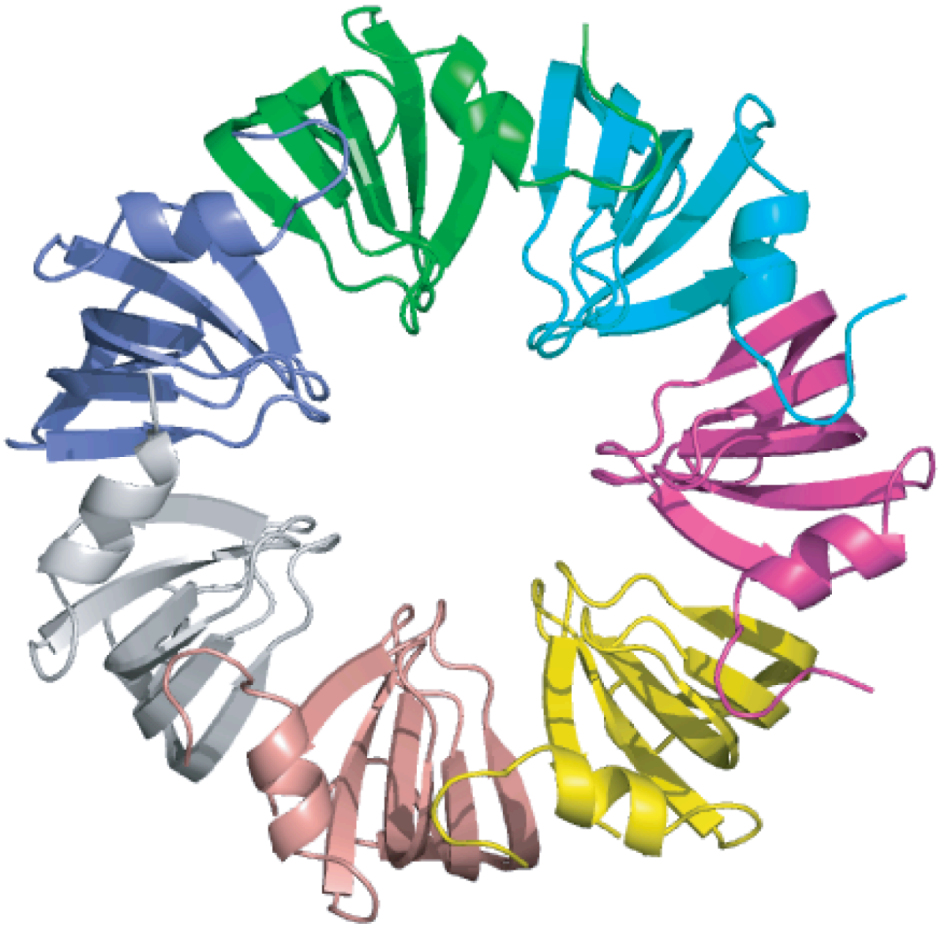

Nop10
Gar1
Cbf5
target RNA
Cbf5 (PUA)
ACA

Figure 1.7: Crystal structure of archaeal Lsm protein (PDB # 1I81) [219]. Ribbon representations of; (A) Lsm monomer with α-helix, β-strands, and loops colored red, yellow, and green, respectively, and (B) Lsm heptamer with each subunit colored differently.

A)



B)



CHAPTER 2

RNA-GUIDED RNA MODIFICATION:

FUNCTIONAL ORGANIZATION OF THE ARCHAEAL H/ACA RNP[1]

[1]Osama A. Youssef*, Daniel L. Baker*, Michael I.R. Chastkofsky, David A. Dy, Rebecca M. Terns and Michael P. Terns, Genes and Development (2005), 15:1238-1248
*These authors contributed equally to this work
Reprinted here with permission of publisher

Abstract

In eukaryotes and archaea, uridines in various RNAs are converted to pseudouridines by RNA-guided RNA modification complexes termed H/ACA RNPs. Guide RNAs within the complexes basepair with target RNAs to direct modification of specific ribonucleotides. Cbf5, a protein component of the complex, likely catalyzes the modification. However, little is known about the organization of H/ACA RNPs and the roles of the multiple proteins thought to comprise the complexes. We have reconstituted functional archaeal H/ACA RNPs from recombinant components, defined the components necessary and sufficient for function, and determined the direct RNA-protein and protein-protein interactions that occur between the components. The results provide substantial insight into the functional organization of this RNP. The functional complex requires a guide RNA and each of four proteins: Cbf5, Gar1, L7Ae and Nop10. Two proteins interact directly with the guide RNA: L7Ae and Cbf5. L7Ae does not interact with other H/ACA RNP proteins in the absence of the RNA. We have defined two novel functions for Cbf5. Cbf5 is the protein that specifically recognizes and binds H/ACA guide RNAs. In addition, Cbf5 recruits the two other essential proteins, Gar1 and Nop10, to the pseudouridylation guide complex.

Introduction

The two most extensive classes of non-coding RNAs are microRNAs (or siRNAs) and modification guide RNAs. MicroRNAs regulate protein production by basepairing with target mRNAs, and triggering destruction or inhibition of translation [1-3]. Similarly, modification guide RNAs basepair to target RNAs, in this case effecting modification of targeted nucleotides [4,5]. The RNA-guided RNA modification system alters the primary sequence and modulates

the function of target RNAs that include rRNAs, snRNAs, tRNAs and perhaps mRNAs [6-10]. In humans it is currently estimated that over 200 2'-*O*-methylations and pseudouridylations are introduced into rRNA and other RNAs by this system [11-14].

There are two large families of modification guide RNAs found in both eukaryotes and archaea: C/D RNAs that guide 2'-*O*-ribose methylation [15,16] and H/ACA RNAs that guide pseudouridylation [17-20]. Both families of guide RNAs function in the context of RNA-protein complexes (RNPs) that include the enzyme responsible for modification [21,22]. The functional organization of modification guide RNPs, including the mechanism by which the enzyme associates with a guide RNA and the roles of the other essential proteins in the complex, is a subject of great interest. In C/D RNPs the 2'-*O*-methyltransferase, fibrillarin, associates with a guide RNA primarily via a bridge formed by the other proteins in the complex, Nop56/58 and L7Ae (or Nop56, Nop58 and p15.5 in eukaryotes). L7Ae binds directly to box C/D RNAs via Kink (K)-turn motifs [23] formed by conserved box C and box D sequences [24-26], and thereby nucleates assembly of the RNP. Binding of L7Ae mediates binding of Nop56/58, which in turn allows association of fibrillarin with the guide RNA [27-30]. Base-pairing of the guide RNA with the target RNA positions the substrate nucleotide for 2'-*O*-methylation by fibrillarin [15,31].

Much less is known about the structure and function of the pseudouridylation guide RNPs or H/ACA RNPs. Four proteins have been identified as components of H/ACA RNPs: Cbf5, Gar1, L7Ae (Nhp2 in eukaryotes), and Nop10 [32-38]. In yeast, these proteins are found associated with H/ACA guide RNAs, and disruption of the corresponding genes affects pseudouridylation [32-34,37]. The sequence and structure of Cbf5 suggest that it is a pseudouridine synthase [33,39-42]. The precise roles of the other proteins are not known. It is

not known whether these four proteins comprise the full set of proteins required for RNA-guided pseudouridylation. It is also not known whether the essential roles of the proteins *in vivo* reflect direct involvement in modification or critical upstream functions (e.g. stabilization or trafficking of guide RNAs). The pseudouridylation guide RNAs are comprised of one to three hairpins, each of which contains a bipartite guide sequence within an internal loop (pseudouridylation pocket) and is followed by a conserved sequence element, either box H or box ACA [17,19,20]. Due largely to the technical difficulties that have been encountered with the protein components of H/ACA RNPs from eukaryotes, there is very little information about the organization and composition of functional complexes.

In this work we report the first reconstitution of RNA-guided RNA pseudouridylation from recombinant components. We have reconstituted functional H/ACA RNPs using components from the hyperthermophilic archaeon *Pyrococcus furiosus*. Our results demonstrate that each of four proteins, Cbf5, Gar1, L7Ae and Nop10, and the guide RNA are essential, and that this set of five components is sufficient for function *in vitro*. The reaction depends upon the pseudouridylation pocket, K-turn and box ACA sequence within the guide RNA. We have also mapped the direct RNA-protein and protein-protein interactions between the components of the archaeal pseudouridylation guide complex. Surprisingly, we have found that Cbf5, the presumptive pseudouridine synthase, interacts directly and specifically with the H/ACA guide RNA. The interaction of Cbf5 with the guide RNA depends on the signature motif, box ACA, and the pseudouridylation pocket (and also to some extent on sequences in the terminal loop of the hairpin), but does not depend on the K-turn. In addition, the archaeal Cbf5 protein can specifically recognize eukaryotic H/ACA RNAs. Our results suggest that the number of molecules of Cbf5 bound to an H/ACA RNA correlates with the number of RNA hairpin units.

As has been reported previously [36], L7Ae also interacts directly with the H/ACA RNA via the K-turn. Our work indicates that L7Ae does not interact independently with the other protein components of the RNP and also is not required for the interaction of the other proteins with the guide RNA. The other two essential proteins, Gar1 and Nop10, do not interact with the guide RNA in the absence of other proteins. We have found that Gar1 and Nop10 each interact independently with Cbf5, which mediates the association of these two proteins with the H/ACA guide RNA.

Results

Requirements for RNA-guided RNA pseudouridylation

Proteins with sequence homology to the four proteins associated with eukaryotic pseudouridylation guide RNPs are encoded in archaeal genomes, but with the exception of L7Ae, these proteins have not been characterized [36,38]. In order to assess the potential role of the four archaeal proteins in RNA-guided pseudouridylation, we investigated whether a functional RNP complex could be reconstituted *in vitro* using proteins and RNAs from *P. furiosus*.

We used Pf9, a single hairpin H/ACA RNA from *P. furiosus*, as the guide RNA for the majority of our work. Pf9 was identified as a potential non-coding RNA by Klein et al. in a computational screen for GC-rich regions in the AT-rich genomes of hyperthermophilic archaea [43]. We have determined that this RNA is an H/ACA RNA (see Figure 2.1A) and verified the corresponding modification at the predicted target site (U910) in *P. furiosus* 16S rRNA (S. Marshburn, R. Terns and M. Terns, unpublished data). The four predicted *P. furiosus* H/ACA RNP proteins (Cbf5, L7Ae, Gar1 and Nop10) were expressed with histidine tags and purified by

affinity chromatography (Figure 2.1B). The substrate for the pseudouridylation assay consisted of the target region of *P. furiosus* 16S rRNA (nts 905-917) flanked by three nucleotide extensions at each end. In addition, to facilitate unequivocal interpretation of results we substituted three of the uridines that base-pair with Pf9 (nts 915-917) with adenines (and made compensatory changes in the sequence of the pseudouridylation pocket of Pf9) to eliminate uridines other than the target uridine from the substrate for this assay. (When transcribed in the presence of radiolabeled UTP, the substrate RNA will be labeled only at the target uridine.)

We incubated the radiolabeled substrate RNA with unlabeled Pf9 guide RNA and various combinations of the four purified proteins (Figure 2.1C). To test for pseudouridylation of the substrate, we extracted and nuclease-digested the RNA, separated uridines and pseudouridines by thin layer chromatography, and examined the products by autoradiography. No pseudouridylation was observed in the absence of proteins (Figure 2.1C, lane 1) or in the absence of Pf9 RNA (Figure 2.1D, lane 4). In addition, no single protein, including the pseudouridine synthase Cbf5, was found to catalyze pseudouridylation of the rRNA substrate (data not shown). However, pseudouridylation was observed upon addition of all four proteins and the guide RNA (Figure 2.1C, lane 2). Importantly, the absence of any one protein from the reaction resulted in substantial loss or elimination of pseudouridylation activity (Figure 2.1C, lanes 3-6). The results indicate that these four proteins, which were implicated in RNA-guided pseudouridylation on the basis of homology to eukaryotic H/ACA RNP proteins, function in this process in *P. furiosus*. Moreover, our results demonstrate for the first time that the activity of an H/ACA guide RNP depends on all four proteins, Cbf5, Gar1, Nop10 and L7Ae, as well as the guide RNA *in vitro*.

We then tested the importance of conserved elements of the guide RNA in function (Figure 2.1D). We incubated the substrate RNA with the four proteins and various Pf9 mutant

RNAs. Disruption of box ACA, the pseudouridylation pocket or the K-turn eliminated or severely reduced function (Figure 2.1D, compare lanes 1-3 with 5). Thus, function of the complex *in vitro* also depends on at least three important elements of the guide RNA: the signature motif (box ACA), the pseudouridylation pocket and the L7Ae binding site (K-turn, see [36].

Mechanism of association of Cbf5 with H/ACA guide RNAs

One key issue is the mechanism by which the enzyme (Cbf5) associates with the guide RNAs. In the case of C/D modification guide RNPs, it is clear that the association of the enzyme (fibrillarin) depends on prior binding of the other protein components of the RNP [27,28,30]. Interestingly, the protein that recognizes C/D RNAs and initiates assembly of the C/D complex is a common component of C/D and H/ACA RNPs in archaea: L7Ae [25,36]. Furthermore, L7Ae has also been shown to bind directly to archaeal H/ACA RNAs via K-turns [36]. Therefore it seemed likely that L7Ae might also be involved in the assembly of the H/ACA proteins on H/ACA RNAs in archaea.

We tested the ability of each of the four H/ACA RNP proteins to interact with the H/ACA guide RNA (Pf9) in the absence of the other proteins by gel mobility shift assay (Figure 2.2A). Consistent with a previous study [36], we found that L7Ae interacts with Pf9 and that the interaction depends on the K-turn motif of the RNA (Figure 2.2A and B). Surprisingly however, we found that Cbf5 also interacts with Pf9 in the absence of the other H/ACA RNP proteins. The apparent K_d of the interaction between Cbf5 and Pf9 (estimated as the concentration of protein resulting in half-maximal binding of the input RNA) was approximately 450 nM (Figure 2.2C and data not shown). Cbf5 failed to interact with *P. furiosus* C/D RNAs sR2 and sR29, and

human $tRNA_{iMet}$ (Figure 2.2D), indicating that the direct interaction of Cbf5 with the guide RNA is specific. Finally, we found that Nop10 and Gar1 do not interact with the guide RNA independently (tested over a range of protein concentrations up to 600 nM and 10 μM, respectively, data not shown). These results indicate that both the pseudouridine synthase Cbf5 and L7Ae interact directly with the guide RNA, but that the interactions of Gar1 and Nop10 with the guide RNA are likely mediated by the other proteins.

To identify the elements of the guide RNA that are important for its recognition by the modifying enzyme, we tested a series of Pf9 mutants and fragments in gel mobility shift assays (Figure 2.3). We found that box ACA is essential, but not sufficient, for recognition by Cbf5. Mutation of box ACA eliminated the interaction observed with wildtype Pf9 (Figure 2.3B). However, an RNA comprised of box ACA and the lower stem of the Pf9 hairpin was not sufficient for Cbf5 binding (Figure 2.3C). The pseudouridylation pocket also plays an important role in the interaction of Cbf5 with Pf9. Addition of the pseudouridylation pocket to the lower stem and box ACA resulted in significant binding by Cbf5 (Figure 2.3D). In addition, the elimination of the pseudouridylation pocket in the context of the full-length Pf9 RNA substantially reduced the ability of Cbf5 to interact with the guide RNA (Figure 2.3E). However, while box ACA and the pseudouridylation pocket are necessary for recognition by Cbf5, it appears that these two are not the only elements that contribute to Cbf5 binding (Figure 2.3D), and that another important element present in the upper region of the hairpin of the RNA is required for full binding activity. The K-turn is important for the interaction of L7Ae with Pf9 (Figure 2.2B), however it is not essential for the interaction of Cbf5 (Figure 2.3F). On the other hand, we found that replacement of the terminal loop of the hairpin with a stable tetra-loop significantly reduced binding (Figure 2.3G). Although neither the sequence nor length of the

terminal loops of H/ACA RNAs are thought to be conserved, we noticed a "GAG" sequence present within the terminal loop of several archaeal H/ACA sRNAs. (This bears some similarity to the CAB box that has been found to be important in the localization of certain guide RNAs to Cajal bodies in eukaryotes [44]). Mutation of the GAG sequence significantly reduced Cbf5 binding (Figure 2.3H), suggesting that this sequence within the terminal loop also plays a role in the interaction of Cbf5 with H/ACA guide RNAs.

We also examined the interaction of Cbf5 with other H/ACA RNAs in gel mobility shift assays. Pf3 is a double hairpin H/ACA RNA from *P. furiosus* [36,43]. Interestingly, we found that Pf3 formed two distinct complexes with Cbf5 (Figure 2.4A). The second complex (Figure 2.4A, marked with **), which co-migrates with a band observed with RNA alone, appears with increasing concentrations of Cbf5. We did not observe the formation of more than one specific complex with the single hairpin RNA Pf9, even at protein concentrations up to 6 μM (Figure 2.2C and data not shown). The results suggest that Cbf5 interacts with each of the two hairpins of a double guide RNA. Mutation of the two ACA elements found in Pf3 disrupted the interaction of Cbf5 with Pf3 (Figure 2.4B). In addition, we tested two types of eukaryotic H/ACA RNAs – a small nucleolar or snoRNA and a small Cajal body or scaRNA. U65 is a typical, human H/ACA snoRNA with two hairpins that guides rRNA modification in the nucleolus [19]. U92 is a double hairpin H/ACA scaRNA that guides pseudouridylation of small nuclear (sn)RNA within Cajal bodies [45]. *P. furiosus* Cbf5 formed two specific complexes with each of these human H/ACA RNAs (Figures 2.4 C and E). As is typical among eukaryotic H/ACA RNAs, a box H sequence (ANANNA) follows the 5' hairpin in U65 and U92, and box ACA follows the 3' hairpin. Mutation of the ACA sequence associated with the 3' hairpin disrupted the interaction of the archaeal Cbf5 with human U65 and U92 (Figure 2.4D and F).

The disruption of binding at both hairpins is consistent with the previous observation that mutation of either the H or ACA sequence element of a eukaryotic RNA eliminates the function of both guide elements [46] and suggests cooperativity in the interaction of proteins with the two hairpins in eukaryotic double guide H/ACA RNAs.

In summary, these results indicate that Cbf5, the pseudouridine synthase, interacts with H/ACA guide RNAs specifically and independently of the other proteins of the pseudouridylation guide complex. Our mutational analysis indicates that Cbf5 depends upon box ACA, the pseudouridylation pocket and sequences within the terminal loop of the hairpin for interaction with the H/ACA RNA. Moreover, it appears that the number of molecules of Cbf5 that binds a guide RNA correlates with the number of hairpins.

Protein-protein interactions within the archaeal H/ACA RNP

Our results indicate that both Cbf5 and L7Ae interact directly and independently with H/ACA guide RNAs, however the means of association of Gar1 and Nop10 with the RNP was still unclear. In addition, we were very interested in identifying protein-protein interactions between components of the complex. We investigated the protein-protein interactions by incubating various combinations of the recombinant proteins (shown in input (I) lanes), one of which was His-tagged (indicated with an asterisk), and identifying the proteins associated with the tagged protein by affinity chromatography (shown in bound (B) lanes). Bovine serum albumin (BSA) was included in all incubations to assess the extent of non-specific interactions, but was not detected in the affinity-purified samples (Figure 2.5). The results indicate that Cbf5 interacts directly with each Gar1 and Nop10 (Figure 2.5, lanes 1-4). Gar1 and Nop10 do not interact with one another (Figure 2.5, lanes 5-6), but Gar1 does co-purify with tagged Nop10 in the presence of Cbf5 (Figure 2.5, lanes 13-14), indicating that these three proteins form a

heterotrimeric complex in which each Gar1 and Nop10 are bound to Cbf5. At the same time, no interaction was observed between L7Ae and either Gar1, Nop10 or Cbf5 (Figure 2.5, lanes 7-12). Moreover, when all four proteins were co-incubated, Cbf5, Gar1 and Nop10 co-purified, but L7Ae did not, suggesting that L7Ae does not interact with the other protein components of the H/ACA RNP in the absence of the guide RNA (Figure 2.5, lanes 15-16).

In vitro assembly of an H/ACA RNP

We next examined the assembly of the H/ACA RNP in gel mobility shift assays (Figure 2.6). As we have shown, Cbf5 and L7Ae (but not Gar1 and Nop10) interact directly with the single hairpin guide RNA Pf9, and the interaction of Cbf5 with Pf9 depends on box ACA (Figures 2.2 and 2.3, and Figure 2.6, lanes 1-5 and 10-14). In protein-protein interaction assays we found that Gar1 and Nop10 interact with Cbf5 in the absence of the guide RNA (Figure 2.5) and thus hypothesized that Cbf5 mediates the interaction of these two proteins with the RNP. Here we show that addition of each Nop10 and Gar1, and both Nop10 and Gar1 to Cbf5 in gel mobility shift assays results in stepwise supershifts of the RNA relative to Cbf5 alone (Figure 2.6, lanes 4, 6, 7 and 8). Like the interaction of Cbf5 alone, these interactions are dependent on box ACA (Figure 2.6, lanes 13, 15, 16, 17). We did not observe a shift in the mobility of Pf9 with the combination of Gar1 and Nop10 in the absence of Cbf5 (data not shown). In addition, Gar1 and Nop10 did not supershift the L7Ae-Pf9 RNA complex (data not shown). These results indicate that Cbf5 mediates the interaction of both Gar1 and Nop10 with the H/ACA RNP. Addition of L7Ae resulted in a further supershift of the complex formed by Cbf5, Gar1 and Nop10 with Pf9 (Figure 2.6, lane 9, asterisk), indicating that L7Ae can interact with Pf9 in the context of the complex formed with the other three proteins. Together, our results indicate that a

functional H/ACA RNP is formed by the independent binding of each Cbf5 and L7Ae to distinct sites on the guide RNA and by independent binding of Gar1 and Nop10 to Cbf5.

Discussion

RNA-guided RNA pseudouridylation

Pseudouridylation is the most common RNA modification and occurs in tRNA, rRNA, snRNA, snoRNA and likely other non-coding RNAs [6,12,41]. There is mounting evidence that pseudouridines occur in functionally important RNA domains and play a vital role in RNA-mediated cellular processes including pre-mRNA splicing and ribosome function [9,10,47,48]. Pseudouridylation of RNA is an evolutionarily ancient process catalyzed by a large family of enzymes known as pseudouridine synthases [39,49].

There are two distinct mechanisms by which pseudouridine synthases select target uridine residues for isomerization. In all known instances in eubacteria, pseudouridylations are carried out by dedicated pseudouridine synthases that each recognize one or a small set of similar RNA substrates [39,41]. Most known pseudouridine synthases are of this type. However, in archaea and eukaryotes, many pseudouridylations are introduced by RNA-guided pseudouridine synthases [6]. The RNA-guided system is versatile and employs armies of H/ACA guide RNAs to direct a common pseudouridine synthase to many different sites. The RNA-guided pseudouridine synthases are members of the TruB subfamily and are called Cbf5(p) in yeast [50] and archaea [38], dyskerin in humans [51], and NAP57 in rat [52]. Three additional proteins are associated with the RNA-guided pseudouridine synthases: Gar1, Nop10 and L7Ae (Nhp2 in eukaryotes). However, the roles of these additional proteins in pseudouridylation are not known. The results presented here provide a substantial amount of new information on the structure and

function of the RNP that catalyzes RNA-guided RNA pseudouridylation in archaea.

The pseudouridine synthase Cbf5 interacts directly with H/ACA guide RNAs via the conserved box ACA element

Box ACA is the signature sequence element of H/ACA RNAs. In eukaryotes, mutational analysis has demonstrated that box ACA is essential for multiple aspects of H/ACA RNA biogenesis and function, including RNP assembly [4,21,22]. It was therefore thought that box ACA served as an important protein binding site, but the identity of the box ACA binding factor remained elusive. The work presented here demonstrates that the pseudouridine synthase itself, Cbf5 is the RNA binding protein that specifically recognizes box ACA in archaea. We show that mutation of box ACA abolishes Cbf5 binding (Figures 2.3 and 2.6). Like other TruB-family pseudouridine synthases, Cbf5 contains a domain that is involved in interaction with substrate RNA [42,53], but an additional RNA binding motif that might have predicted the ability of Cbf5 to interact selectively with H/ACA guide RNAs was not recognized and should now be a focus of further investigation.

Analysis of hundreds of eukaryotic and archaeal pseudouridylation guide RNAs has revealed a conserved (~14 nucleotide) distance between box ACA of the guide RNA and the unpaired target uridine of the substrate RNA positioned within the pseudouridylation pocket [18]. Our finding that Cbf5 interacts with box ACA may provide an explanation: the fixed distance may simply reflect the physical spacing between the domains of Cbf5 that interact with box ACA and catalyze pseudouridylation of the target uridine.

Organization of functional pseudouridylation guide RNPs

Our findings provide a clear model for the basic organization of the archaeal pseudouridylation guide RNP, in which Cbf5 and L7Ae bind independently to distinct sites on the guide RNA, and Gar1 and Nop10 interact with Cbf5 (Figure 2.7).

L7Ae interacts directly with the K-turn of the guide RNA [36], but we did not find evidence of interaction of L7Ae with the other proteins in the absence of the guide RNA (Figure 2.5). Moreover, the interaction of the other proteins with the RNA did not depend on the presence of L7Ae (Figure 2.6), indicating that L7Ae does not nucleate the assembly of the H/ACA RNP as it does the C/D RNP [27,29,30].

Cbf5 also interacts directly with the guide RNA and we found that box ACA, the pseudouridylation pocket and the terminal loop of the hairpin appear to be important for this interaction (Figures 2.2 and 2.3), suggesting extensive contact between Cbf5 and the guide RNA. Our data indicate that Gar1 and Nop10 each interact directly with Cbf5, but not with the other proteins or with the guide RNA in the absence of Cbf5 (Figures 2.2 and 2.5). The interaction of Gar1 and Nop10 with Cbf5 mediates the interaction of these proteins with the complex (Figure 2.6). Further, our results indicate that these three proteins can form a heterotrimeric Cbf5/Gar1/Nop10 complex that can interact with the guide RNA (Figures 2.5 and 2.6 and unpublished data). Based on our results it seems equally possible that these three proteins assemble on the guide RNA sequentially or as a pre-formed complex.

Eukaryotic pseudouridylation guide RNPs

The components of eukaryotic and archaeal pseudouridylation guide RNPs are generally

well conserved, suggesting that the organization and function of the components will be fundamentally similar in the two systems. Unfortunately, detailed analysis of functional eukaryotic H/ACA RNPs has not been reported to date. However, two recent studies describe interactions between various components of eukaryotic H/ACA RNPs – one in a mammalian system and one in yeast [54,55]. The interactions observed in the yeast study [54] are in agreement with those reported here, while there are significant differences in the interactions observed in the mammalian system [56].

Using mammalian proteins expressed in rabbit reticulocyte lysates, Wang et al. found a heterotrimeric protein complex with a different composition – comprised of the mammalian homologs of Cbf5, Nop10 and L7Ae [55], rather than Cbf5, Nop10 and Gar1 (Figure 2.5). In addition, they found that Nop10 is essential for interaction between the mammalian Cbf5 and L7Ae homologs, and thus appears to play the central role in this complex [55], while Cbf5 is at the core of the archaeal complex, interacting independently with each Gar1 and Nop10 (Figure 2.5). In the mammalian system, specific recognition of H/ACA RNAs required all three components of the trimeric complex [55]. On the other hand, we have found that archaeal Cbf5 interacts specifically with guide RNAs in the absence of the other proteins (Figures 2.2-2.4), and that Gar1 and Nop10 do not observably increase the affinity of the interaction (data not shown).

On the other hand, the data from yeast suggest that the organization of the yeast H/ACA RNP resembles the archaeal complex. In studies with complexes expressed and assembled *in vivo*, and purified from *S. cerevisiae*, Henras et al. also found that Cbf5(p), Gar1(p) and Nop10(p) can form a complex independent of both L7Ae (Nhp2p) and guide RNA [54].

At present it is not clear whether the observed discrepancies between the mammalian system, and the archaeal and yeast systems reflect fundamental differences in the RNPs or the

limitations of experimental approaches. The authors of the mammalian study note that no significant pseudouridylase activity could be detected with the complexes assembled in the mammalian system [55]. The functionality of the purified yeast complexes was not reported. The eukaryotic H/ACA RNP proteins, and especially Cbf5, are challenging to express and purify [54,56]. It is possible that both the lack of functionality of the mammalian proteins and the observed differences result from production of defective (perhaps misfolded) mammalian proteins *in vitro*. A better understanding of the extent of differences between the eukaryotic and archaeal RNPs awaits more detailed structural studies of functional eukaryotic complexes.

Roles of the H/ACA RNP proteins in RNA-guided pseudouridylation

All evidence indicates that Cbf5 is the pseudouridine synthase (i.e. catalyzes breakage of the N1-C1' glycosidic bond and reattachment of the free uridine base to the ribose via a C5-C1' glycosidic bond). The sequence and structure of the protein is very similar to other known pseudouridine synthases [39,41,42], and in yeast, mutation of the predicted catalytic aspartate (universally conserved in all pseudouridine synthases) prevents RNA-guided pseudouridylation *in vivo* [33,40]. Our work establishes two additional key roles for Cbf5: direct recognition of the guide RNA and recruitment of both Gar1 and Nop10 (Figures 2.2 and 2.5).

What are the roles of the other proteins? It is clear that L7Ae interacts directly with the guide RNA [36]. In addition, our results indicate that L7Ae does not interact independently with the other proteins and is not responsible for the recruitment of the other proteins to the complex (Figures 2.5 and 2.6 and data not shown). One conceivable role of L7Ae is alteration of the structure of the guide RNA (e.g. introduction of a kink in the upper stem) to induce a conformation in the RNA or RNP that is important for pseudouridylation.

Our finding that Cbf5 interacts directly with the guide RNA indicates that the other proteins do not bridge the interaction of the modifying enzyme with the guide RNA. Gar1 and Nop10 both interact with Cbf5 (Figure 2.5), but this interaction does not apparently increase the affinity of Cbf5 for the guide RNA (data not shown). The association of Gar1 and Nop10 with Cbf5 suggests that they may serve auxiliary roles in H/ACA RNP function. For example, these proteins may promote or stabilize the interaction with the substrate rRNA, ensure proper positioning of the target uridine in the active site, or influence substrate rRNA release following catalysis. Site-specific crosslinking studies support an intimate association of Gar1 (as well as Cbf5) with the target uridine in the mammalian system [55]. In addition, genetic depletion of Gar1(p) in yeast results in partially assembled RNP complexes that are unable to interact with substrate rRNA [37]. Analysis of the sequence of archaeal Nop10 proteins suggests the potential existence of a zinc-finger motif (D. Baker, J. Omichinski, R. Terns and M. Terns, unpublished observation) and the ability to interact directly with nucleic acids – perhaps substrate RNA.

Importantly, our results establish for the first time that each of the four proteins is essential for RNA-guided pseudouridylation *in vitro* (Figure 2.1). Previous studies in eukaryotes established that these proteins are associated with H/ACA guide RNAs and that RNA pseudouridylation is reduced in cells lacking these proteins. Our results indicate that the proteins are not solely required for the stability of the guide RNAs or other upstream functions *in vivo*, but are also necessary for efficient catalysis of the modification.

Materials and Methods

Expression and purification of recombinant proteins

The genes encoding *P. furiosus* Cbf5 (PF1785), Gar1 (PF1791), Nop10 (PF1141), and L7Ae (PF1367) were amplified by PCR from genomic DNA and cloned into modified versions

of pET21d and pET24d. The primers used in the PCR reactions are specified in Tables I and II (supplemental data). The recombinant proteins were expressed in *E. coli* BL21 codon+ cells (DE3, Invitrogen). The cells were grown to a culture OD_{600} of 0.7, and expression of the proteins was induced with 1 mM isopropylthio-b-D-galactoside (IPTG) for 4 h at 37°C. The cells were pelleted, resuspended in Buffer A (20 mM sodium phosphate buffer (pH 7.0), 1M NaCl and 0.1 mM phenylmethylsulfonyl fluoride (PMSF)), and disrupted by sonication (10 second pulse, 20 second rest, repeated 5 cycles using a Branson Sonifier Cell Disruptor 200 and microtip, intensity level 6, duty cycle 60%). The sonicated sample was centrifuged at 45,000 x g for 30 min at 4°C. The supernatant was heated at 75-78°C for 20 min and centrifuged at 45,000 x g for 20 min at 4°C. The supernatant was filtered (0.8 μm pore size Millex Filter Unit, Millipore) and applied to a Ni–NTA agarose (Qiagen) column equilibrated with Buffer A. Proteins were eluted with Buffer A containing 350-500 mM imidazole. The protein samples were dialyzed at room temperature against 40 mM HEPES (pH 7.0), 100-500 mM KCl. Some samples were concentrated using a PL-10 Microcon filter device (Millipore). The purity of the protein samples was assessed by SDS-PAGE and Coomassie blue staining. The concentration of the proteins was determined via BCA protein assay (Pierce).

Synthesis of DNA templates for *in vitro* transcription of RNAs

DNA templates used for *in vitro* transcription of Pf3 and Pf9 RNAs (and mutants) were generated by PCR using *P. furiosus* genomic DNA and oligonucleotides as described in Tables I and II (supplemental data). The oligonucleotides incorporate an SP6 polymerase promoter for *in vitro* transcription. The PCR products were cloned into the pCR2.1-TOPO vector (Invitrogen) and were confirmed by DNA sequence analysis. DNA templates encoding sR2 and sR29 [57]

and human $tRNA^{i}Met$ [58] were generated as described. The template for *in vitro* transcription of the substrate RNA (corresponding to nts 905-917 of *P. furiosus* 16S rRNA with uridines 915-917 replaced by adenosines, and flanked by three nucleotide extensions at each end) was generated by direct annealing of two oligonucleotides (see Tables I and II, supplemental data).

In vitro transcription of guide and target RNAs

PCR product (30-100 ngs), linearized plasmid (1μg), or annealed oligonucleotides were used as templates for *in vitro* transcription, which was performed as described previously [58] using α^{32}P-GTP to uniformly radiolabel guide RNAs or α^{32}P-UTP to label the uridine in the substrate rRNA.

Pseudouridylation assay

0.2-5 pmols of guide RNA, and 0.05 pmols of ^{32}P-labeled rRNA substrate were incubated with purified his-tagged proteins in 40 mM HEPES (pH 7.0), 500 mM KCl, 1.5 mM $MgCl_2$, 10% v/v glycerol, 5 μg *E. coli* tRNA, 1 U/10 μL of RNasin (Promega, Madison, WI) for 1 hour at 70^{o}C. The reaction was terminated by extraction with phenol/chloroform/isoamyl alcohol at 4^{o}C, and the RNA was ethanol-precipitated and digested with nuclease P1 (200 ng, United States Biological, Swampscott, MA). The nucleotide 5' monophosphate mixture was separated via thin layer chromatography on cellulose polyethyleneimine plates (EMD Chemicals, Gibbstown, NJ) with isopropanol-HCl-water (70:15:15) as the solvent [59]. Under these conditions, pseudouridine migrates more slowly than uridine [59].

Gel mobility shift assays

0.05 pmol of ^{32}P-labeled RNA was mock treated or mixed with indicated amounts of recombinant Cbf5, Nop10, L7Ae or Gar1 proteins. Reactions were carried out in a final volume of 20 μl containing 20 mM HEPES (pH 7.0), 250 mM KCl, 1.5 mM $MgCl_2$, 0.25μg/μL *E. coli* tRNA, 0.75 mM DTT and 10% glycerol. After incubation at 37^{o}C for 1 hr, samples were loaded on nondenaturing 6% or 8% polyacrylamide gels containing 0.5x TBE. Electrophoresis was performed at 4^{o}C in 0.5x TBE for 12 hours at 125V. The RNA distribution was visualized by autoradiography after gel drying.

***In vitro* protein/protein interaction assay**

Protein samples were dialyzed against Buffer B (20 mM HEPES (pH 7.0), 500 mM KCl, 1.5 mM $MgCl_2$). Approximately equimolar amounts of proteins were incubated for 30 min at 37^{o}C. Bovine serum albumin (Promega, Madison, Wis.) was included as a negative control. Half of the protein mixture was reserved as input sample and concentrated 10-fold using a YM-3 Microcon filter device (Millipore). The other half of the sample was incubated for 10 min at room temperature with 15 μL of Ni-NTA resin (Qiagen) equilibrated in Buffer B. The resin was washed 4 times with Buffer B plus 20 mM imidazole and 0.1% Triton X-100. Bound proteins were eluted with SDS gel loading buffer and heating. Input and bound protein samples were analyzed by 15% Tris-tricine gel electrophoresis and Coomassie blue protein staining.

Acknowledgements

We are grateful to Mike Adams, Gerti Schut, and Frank Jenney (University of Georgia) for generously sharing *P. furiosus* resources, Yi-Tao Yu (University of Rochester) and Tom

Meier (Albert Einstein College of Medicine) for technical advice regarding the pseudouridylation assay, Natalia Starostina for technical insights on protein expression, and Sarah Marshburn for characterization of Pf9 and critical reading of this manuscript. This work was supported by grants from the National Institutes of Health and National Science Foundation to M.P.T. and R.M.T, a fellowship from the Egyptian government to O.A.Y., and a CURO Summer Research Fellowship to M.I.R.C.

References

1. Ambros V: **The functions of animal microRNAs**. *Nature* 2004, **431**:350-355.

2. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**:281-297.

3. He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation**. *Nat Rev Genet* 2004, **5**:522-531.

4. Kiss T: **Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions**. *Cell* 2002, **109**:145-148.

5. Decatur WA, Fournier MJ: **RNA-guided nucleotide modification of ribosomal and other RNAs**. *J Biol Chem* 2003, **278**:695-698.

6. Yu YT, Terns RM, Terns MP: **Mechanisms and functions of RNA-guided RNA modification**. In *Fine-tuninfg of RNA functions by modification and editing*. Edited by Grosjean H: Topics in Current Genetics; 2005. vol In Press.

7. Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A: **Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization**. *Proc Natl Acad Sci U S A* 2000, **97**:14311-14316.

8. Omer AD, Ziesche S, Decatur WA, Fournier MJ, Dennis PP: **RNA-modifying machines in archaea**. *Mol Microbiol* 2003, **48**:617-629.

9. King TH, Liu B, McCully RR, Fournier MJ: **Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center**. *Mol Cell* 2003, **11**:425-435.

10. Yu YT, Shu MD, Steitz JA: **Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing**. *EMBO J* 1998, **17**:5783-5795.

11. Maden BE: **The numerous modified nucleotides in eukaryotic ribosomal RNA**. *Prog Nucleic Acid Res Mol Biol* 1990, **39**:241-303.

12. Ofengand J, Fournier MJ: **The pseudouridine residues of rRNA: Number, location, biosynthesis, and function**. In *Modification and Editing of RNA*. Edited by Grosjean H, Benne B: ASM Press; 1998:229-253.

13. Bachellerie J-P, Cavaille J: **Small nucleolar RNAs guide the ribose methylations of eukaryotic rRNAs**. In *Modification and Editing of RNA*. Edited by Grosjean H, Benne B: ASM Press; 1998:255-272.

14. Vitali P, Royo H, Seitz H, Bachellerie JP, Huttenhofer A, Cavaille J: **Identification of 13 novel human modification guide RNAs**. *Nucleic Acids Res* 2003, **31**:6543-6551.

15. Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T: **Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs**. *Cell* 1996, **85**:1077-1088.

16. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea**. *Science* 2000, **288**:517-522.

17. Balakin AG, Smith L, Fournier MJ: **The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions**. *Cell* 1996, **86**:823-834.

18. Ganot P, Bortolin ML, Kiss T: **Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs**. *Cell* 1997, **89**:799-809.

19. Ganot P, Caizergues-Ferrer M, Kiss T: **The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation**. *Genes Dev* 1997, **11**:941-956.

20. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A: **Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus**. *Proc Natl Acad Sci U S A* 2002, **99**:7536-7541.

21. Terns MP, Terns RM: **Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin**. *Gene Expr* 2002, **10**:17-39.

22. Filipowicz W, Pogacic V: **Biogenesis of small nucleolar ribonucleoproteins**. *Curr Opin Cell Biol* 2002, **14**:319-327.

23. Klein DJ, Schmeing TM, Moore PB, Steitz TA: **The kink-turn: a new RNA secondary structure motif**. *EMBO J* 2001, **20**:4214-4221.

24. Charron C, Manival X, Clery A, Senty-Segault V, Charpentier B, Marmier-Gourrier N, Branlant C, Aubry A: **The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity**. *J Mol Biol* 2004, **342**:757-773.

25. Kuhn JF, Tran EJ, Maxwell ES: **Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein**. *Nucleic Acids Res* 2002, **30**:931-941.

26. Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R: **A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP**. *Cell* 2000, **103**:457-466.

27. Omer AD, Ziesche S, Ebhardt H, Dennis PP: **In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex**. *Proc Natl Acad Sci U S A* 2002, **99**:5289-5294.

28. Rashid R, Aittaleb M, Chen Q, Spiegel K, Demeler B, Li H: **Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles**. *J Mol Biol* 2003, **333**:295-306.

29. Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, Li H: **Structure and function of archaeal box C/D sRNP core proteins**. *Nat Struct Biol* 2003, **10**:256-263.

30. Tran EJ, Zhang X, Maxwell ES: **Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs**. *EMBO J.* 2003, **22**:3930-3940.

31. Cavaille J, Nicoloso M, Bachellerie JP: **Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides**. *Nature* 1996, **383**:732-735.

32. Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gelugne JP, Caizergues-Ferrer M: **Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs**. *EMBO J* 1998, **17**:7078-7090.

33. Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D: **The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase**. *Genes Dev* 1998, **12**:527-537.

34. Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Luhrmann R: **Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA- binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure**. *RNA* 1998, **4**:1549-1568.

35. Dragon F, Pogacic V, Filipowicz W: **In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs**. *Mol Cell Biol* 2000, **20**:3037-3048.

36. Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A: **Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea**. *Nucleic Acids Res* 2003, **31**:869-877.

37. Bousquet-Antonelli C, Henry Y, G'Elugne J P, Caizergues-Ferrer M, Kiss T: **A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs**. *EMBO J* 1997, **16**:4770-4776.

38. Watanabe Y, Gray MW: **Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria**. *Nucleic Acids Res* 2000, **28**:2342-2352.

39. Koonin EV: **Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases**. *Nucleic Acids Res* 1996, **24**:2411-2415.

40. Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J: **Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA**. *Mol Cell Biol* 1999, **19**:7461-7472.

41. Charette M, Gray MW: **Pseudouridine in RNA: what, where, how, and why**. *IUBMB Life* 2000, **49**:341-351.

42. Hoang C, Ferre-D'Amare AR: **Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme**. *Cell* 2001, **107**:929-939.

43. Klein RJ, Misulovin Z, Eddy SR: **Noncoding RNA genes identified in AT-rich hyperthermophiles**. *Proc Natl Acad Sci U S A* 2002, **99**:7542-7547.

44. Richard P, Darzacq X, Bertrand E, Jady BE, Verheggen C, Kiss T: **A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs**. *EMBO J* 2003, **22**:4283-4293.

45. Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T: **Cajal body-specific small nuclear RNAs: a novel class of 2'-O- methylation and pseudouridylation guide RNAs**. *EMBO J* 2002, **21**:2746-2756.

46. Bortolin ML, Ganot P, Kiss T: **Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs**. *EMBO J* 1999, **18**:457-469.

47. Donmez G, Hartmuth K, Luhrmann R: **Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation**. *RNA* 2004, **10**:1925-1933.

48. Zhao X, Yu YT: **Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in Xenopus oocytes**. *RNA* 2004, **10**:681-690.

49. Ofengand J, Malhotra A, Remme J, Gutgsell NS, Del Campo M, Jean-Charles S, Peil L, Kaya Y: **Pseudouridines and pseudouridine synthases of the ribosome**. *Cold Spring Harb Symp Quant Biol* 2001, **66**:147-159.

50. Jiang W, Middleton K, Yoon HJ, Fouquet C, Carbon J: **An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules**. *Mol Cell Biol* 1993, **13**:4884-4893.

51. Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I: **X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions**. *Nat Genet* 1998, **19**:32-38.

52. Meier UT, Blobel G: **NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria**. *J Cell Biol* 1994, **127**:1505-1514.

53. Aravind L, Koonin EV: **THUMP--a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases**. *Trends Biochem Sci* 2001, **26**:215-217.

54. Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M: **Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs**. *RNA* 2004, **10**:1704-1712.

55. Wang C, Meier UT: **Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins**. *EMBO J* 2004, **23**:1857-1867.

56. Wang H, Boisvert D, Kim KK, Kim R, Kim SH: **Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution**. *EMBO J* 2000, **19**:317-323.

57. Speckmann WA, Li ZH, Lowe TM, Eddy SR, Terns RM, Terns MP: **Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus**. *Curr Biol* 2002, **12**:199-203.

58. Narayanan A, Speckmann W, Terns R, Terns MP: **Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli**. *Mol Biol Cell* 1999, **10**:2131-2147.

59. Yu YT, Shu MD, Narayanan A, Terns RM, Terns MP, Steitz JA: **Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of Xenopus oocytes**. *J Cell Biol* 2001, **152**:1279-1288.

Figure 2.1: Reconstitution of functional pseudouridylation guide RNPs from recombinant RNA and protein components. A) Sequence and secondary structure of Pf9 H/ACA guide RNA with important elements indicated. Box ACA is located at the base of the hairpin structure near the 3' end of the RNA (nts 68-70). The pseudouridylation pocket is an internal loop bounded by the upper and lower stems of the hairpin. The nucleotides within the pocket base-pair with the rRNA substrate (represented as solid bold line), positioning the unpaired uridine to be modified (Ψ) at the top of the loop. A kink-turn motif is located in the upper stem, near the terminal loop of the hairpin and consists of an asymmetric loop containing two G-A basepairs and flanked by two short stems [23]. A GAG sequence present in the terminal loop of Pf9 and other archaeal H/ACA RNAs is indicated (nts 30-32). B) Purified samples of H/ACA RNP proteins Cbf5, Gar1, Nop10, and L7Ae analyzed by SDS PAGE and Coomassie protein staining are shown. C) Pseudouridylation activity of various combinations of the four recombinant H/ACA RNP proteins. Pf9 guide RNA and substrate RNA (containing a single, ^{32}P-labeled target uridine) were incubated with the indicated combinations of proteins. Pseudouridylation was assessed by TLC separation of nucleotides (obtained by nuclease P1 digestion of RNA) under established conditions where pseudouridine (Ψp) migrates more slowly than uridine (Up) [59]. Autoradiographs of TLC plates are shown. D) Effect of mutations in Pf9 guide RNA on pseudouridylation activity. Box ACA was mutated to UGU (ΔACA). Pseudouridylation pocket was eliminated by replacement of sequence on one side of the loop with sequence complementary to other side of the loop (ΔΨ pocket). The K-turn was mutated by disruption of critical GA basepairs (substitution of GA with CC; ΔK-turn). The indicated mutant or wildtype Pf9 guide RNA was incubated with the four recombinant proteins and substrate RNA, and pseudouridylation activity was assessed as in C.

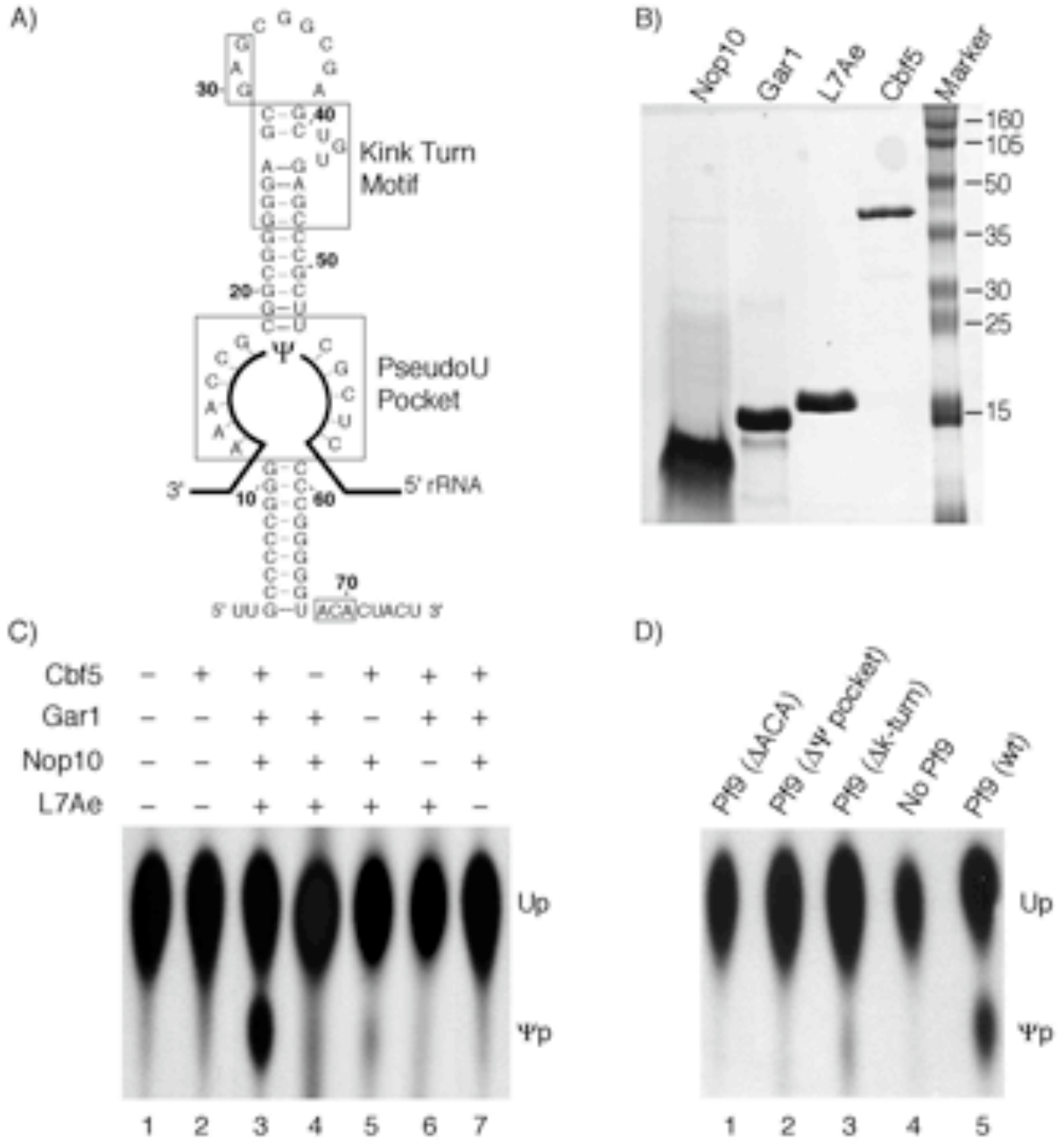

A)
Kink Turn Motif
PseudoU Pocket
5' rRNA
3'
5' UU G—U ACA CUACU 3'
B)
Nop10
Gar1
L7Ae
Cbf5
Marker
160
105
50
35
30
25
15
C)
Cbf5 – + + – + + +
Gar1 – – + + – + +
Nop10 – – + + + – +
L7Ae – – + + + + –
Up
Ψp
1 2 3 4 5 6 7
D)
Pf9 (ΔACA)
Pf9 (ΔΨ pocket)
Pf9 (Δk-turn)
No Pf9
Pf9 (wt)
Up
Ψp
1 2 3 4 5

Figure 2.2: Cbf5 interacts directly and specifically with Pf9 H/ACA guide RNA. Direct interactions of proteins with ^{32}P-labeled RNAs were investigated by native gel mobility shift analysis and autoradiography. A) Pf9 RNA was incubated with each of the four recombinant H/ACA RNP proteins or no protein (-). B) The K-turn of Pf9 was disrupted and the mutant RNA was incubated with L7Ae. C) Wildtype Pf9 was incubated with increasing concentrations of Cbf5 (0 to 2000 nM) to assess the apparent Kd of the observed interaction. D) Cbf5 was incubated with non-H/ACA RNAs including *P. furiosus* C/D RNAs sR2 and sR29, and a human tRNA to assess the specificity of the observed interaction.

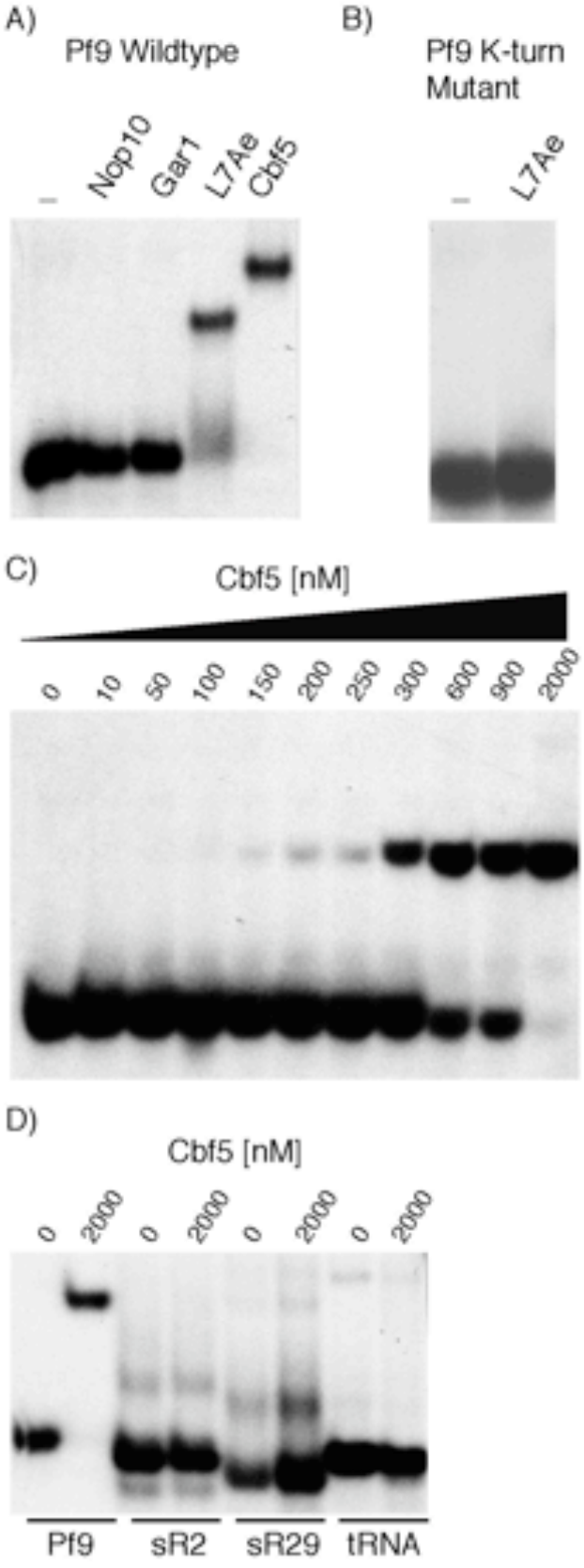

A)
Pf9 Wildtype
–
Nop10
Gar1
L7Ae
Cbf5
B)
Pf9 K-turn Mutant
–
L7Ae
C)
Cbf5 [nM]
0
10
50
100
150
200
250
300
600
900
2000
D)
Cbf5 [nM]
0
2000
0
2000
0
2000
0
2000
Pf9
sR2
sR29
tRNA

Figure 2.3: Elements of the H/ACA guide RNA important for Cbf5 interaction. The ability of Cbf5 to interact with mutants and fragments of Pf9 was assessed by native gel mobility shift analysis with a range of concentrations of Cbf5. Each panel shows a diagram of the RNA tested (location of mutations indicated with X), autoradiograph of gel shift analysis and scaled estimate of the extent of interaction relative to wildtype Pf9 (- to +++). A) Wildtype Pf9; B) mutation of box ACA (to UGU); C) deletion of the terminal loop, K-turn, upper stem and pseudouridylation pocket; D) deletion of the terminal loop, K-turn and upper stem; E) closure of pseudouridylation pocket by replacement of sequence on one side of the loop with sequence complementary to other side of the loop; F) disruption of critical GA basepairs in K-turn by substitution of GA with CC; G) replacement of terminal loop with tetra-loop; and H) mutation of GAG in terminal loop (to CUC).

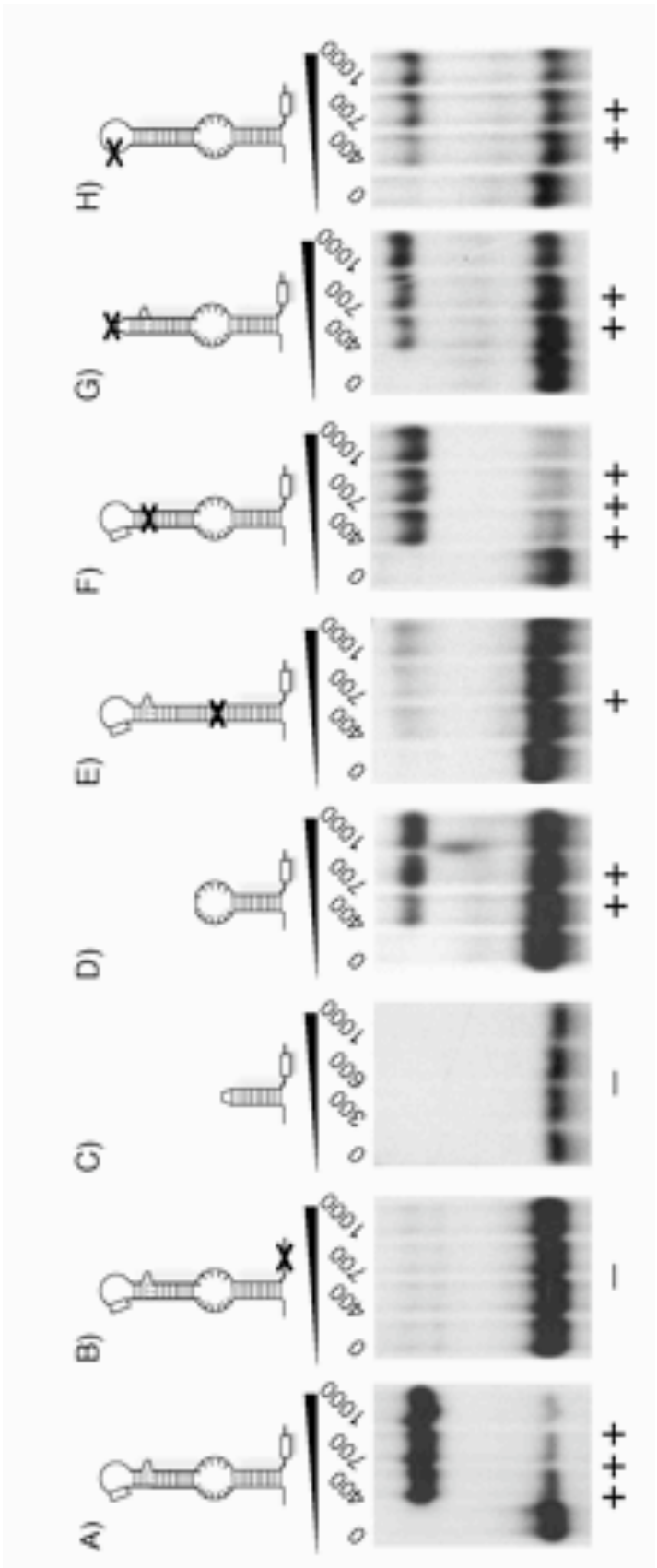

A)
B)
C)
D)
E)
F)
G)
H)
0
400
700
1000
300
600
+++
+
++
+++
++
++
+
++
++
-
-

Figure 2.4: Cbf5 also interacts with archaeal and eukaryotic double hairpin H/ACA RNAs. A, C, E) The ability of Cbf5 to interact with double hairpin H/ACA RNAs Pf3 (a *P. furiosus* guide RNA) U92 (a eukaryotic scaRNA) and U65 (a eukaryotic snoRNA) was assessed by native gel mobility shift analysis with a range of concentrations of Cbf5. Distinct RNP complexes are indicated with single and double asterisks. A diagram of the RNA tested is shown to the left of each panel. B, D, F) In order to assess the specificity of the observed interactions and importance of box ACA, native gel mobility shift analysis was performed with RNAs in which the box ACA elements were mutated (ACA to UGU, or AAA to UUU in the case of the 3' element of Pf3).

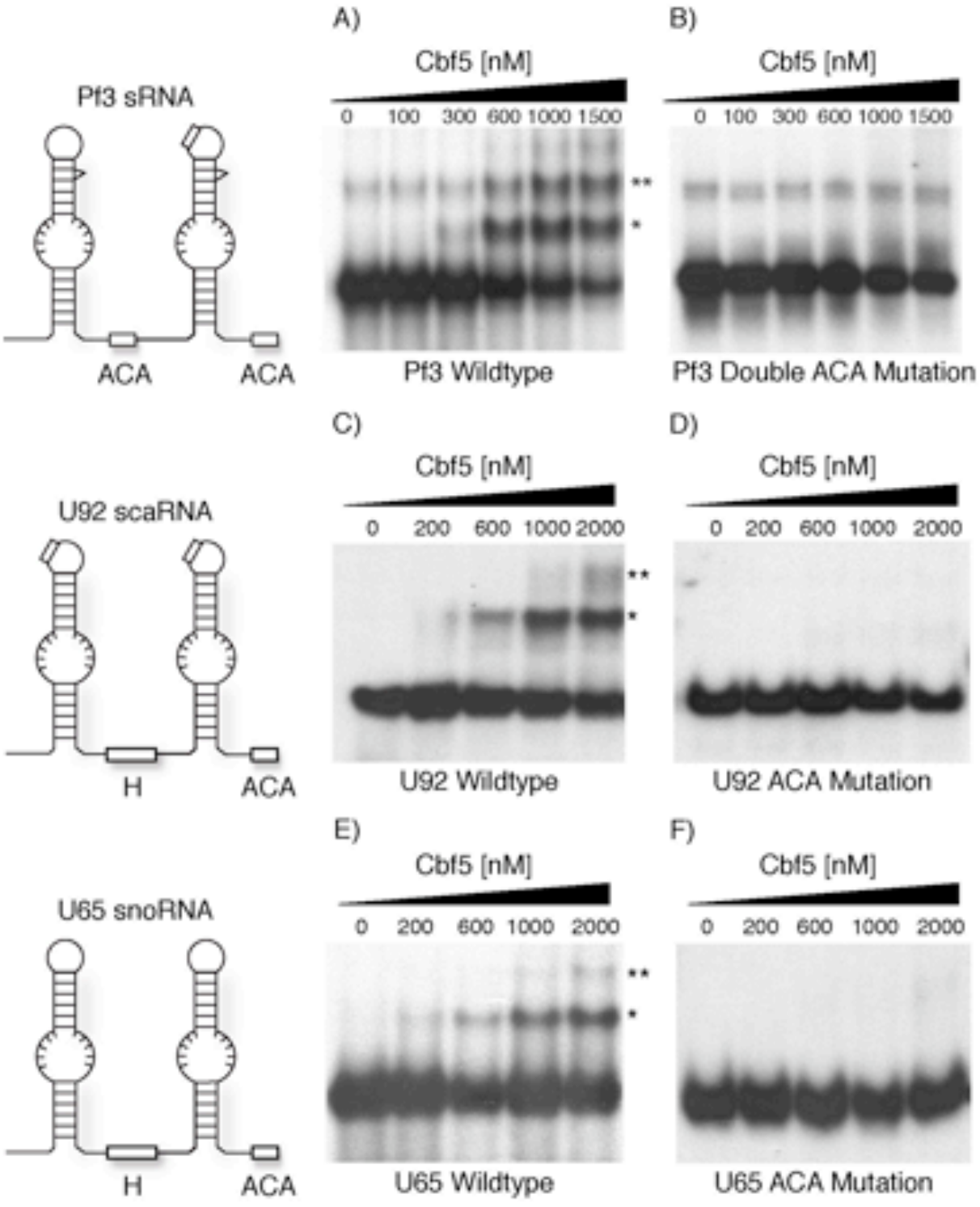

Pf3 sRNA
ACA
ACA
A)
Cbf5 [nM]
0 100 300 600 1000 1500
**
*
Pf3 Wildtype
B)
Cbf5 [nM]
0 100 300 600 1000 1500
Pf3 Double ACA Mutation
U92 scaRNA
H
ACA
C)
Cbf5 [nM]
0 200 600 1000 2000
**
*
U92 Wildtype
D)
Cbf5 [nM]
0 200 600 1000 2000
U92 ACA Mutation
U65 snoRNA
H
ACA
E)
Cbf5 [nM]
0 200 600 1000 2000
**
*
U65 Wildtype
F)
Cbf5 [nM]
0 200 600 1000 2000
U65 ACA Mutation

Figure 2.5: Cbf5 interacts with Gar1 and Nop10 to form a heterotrimeric protein complex. Combinations of the four H/ACA RNP proteins (indicated as C (Cbf5), L (L7Ae), G (Gar1), and N (Nop10)) were incubated in approximately equimolar amounts (I (input) lanes). In each panel the his-tagged protein is designated with an asterisk. Bovine serum albumin (BSA) was also added to the protein mixtures. The his-tagged proteins were purified using nickel agarose resin. Input (I lanes) and bound (B lanes) samples were compared following 15% Tris-tricine gel electrophoresis and Coomassie blue staining.

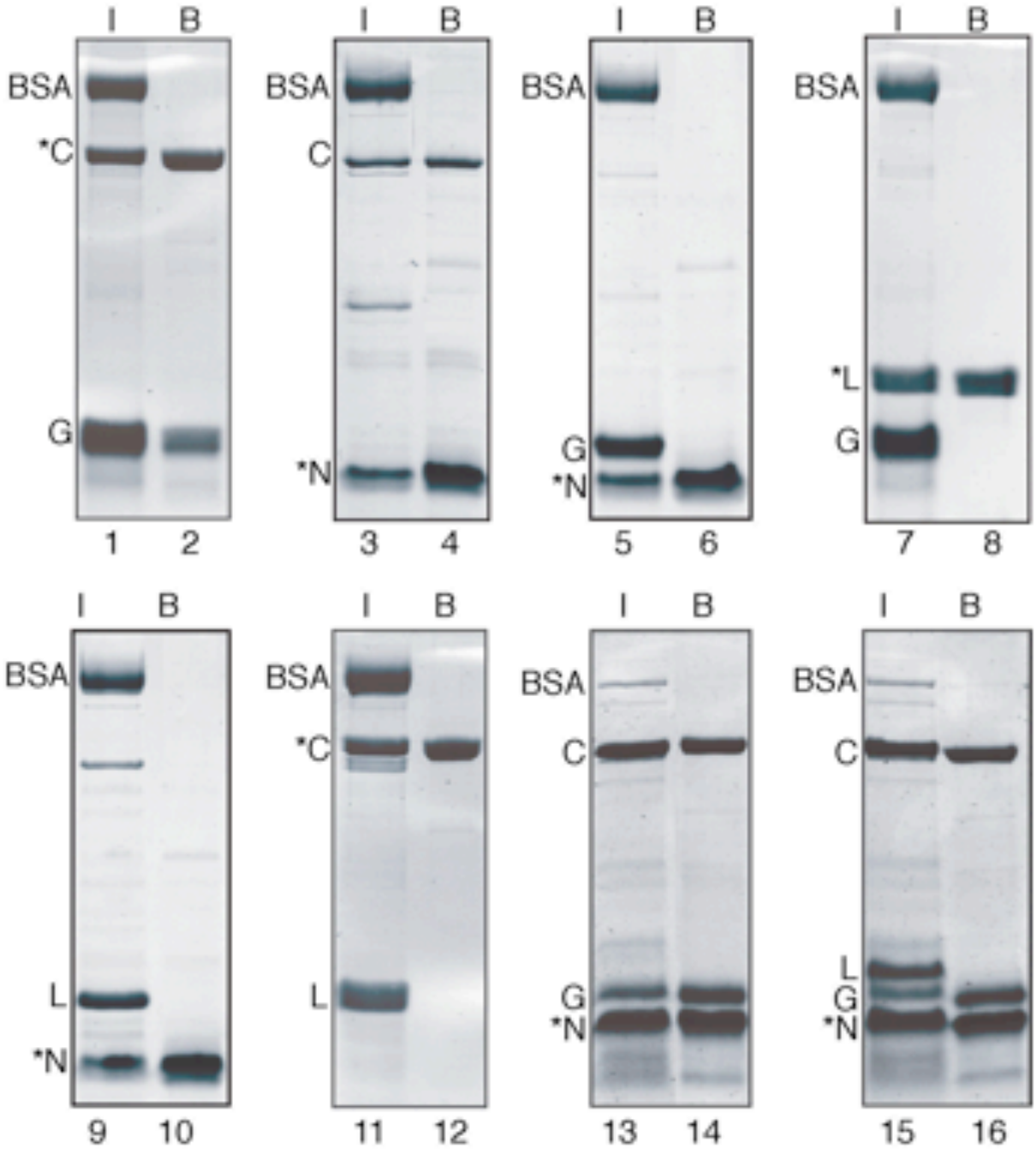

I B
BSA
*C
G
1 2
I B
BSA
C
*N
3 4
I B
BSA
G
*N
5 6
I B
BSA
*L
G
7 8
I B
BSA
L
*N
9 10
I B
BSA
*C
L
11 12
I B
BSA
C
G
*N
13 14
I B
BSA
C
L
G
*N
15 16

Figure 2.6: Assembly of H/ACA RNP proteins with an H/ACA guide RNA. ^{32}P-labeled wildtype (wt) or ACA mutant (ΔACA) Pf9 RNAs were incubated with one or more of the four proteins as indicated. The resultant RNP complexes were detected by native gel shift analysis followed by autoradiography. The distinct complex formed in the presence of all four proteins is indicated with an asterisk.

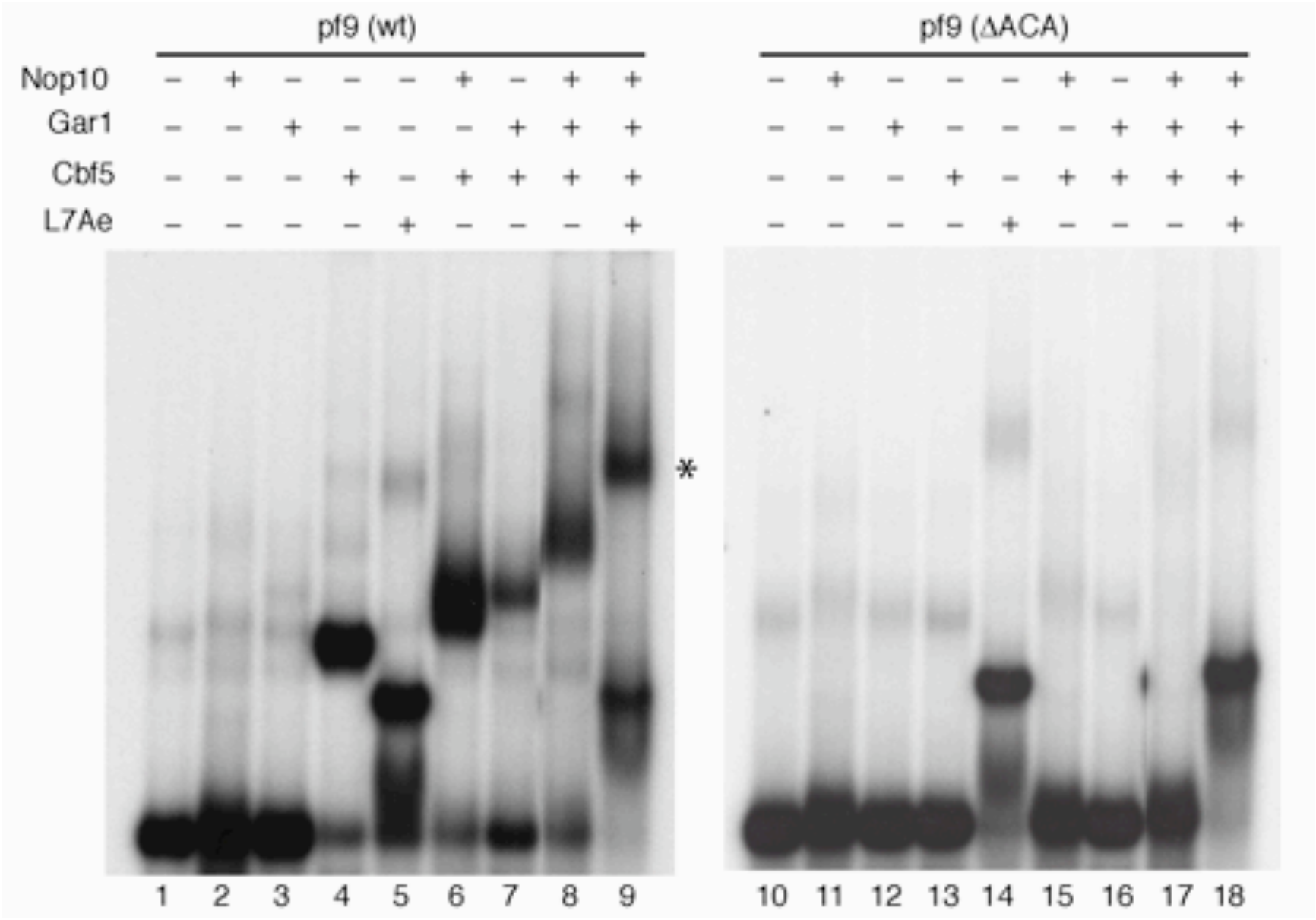
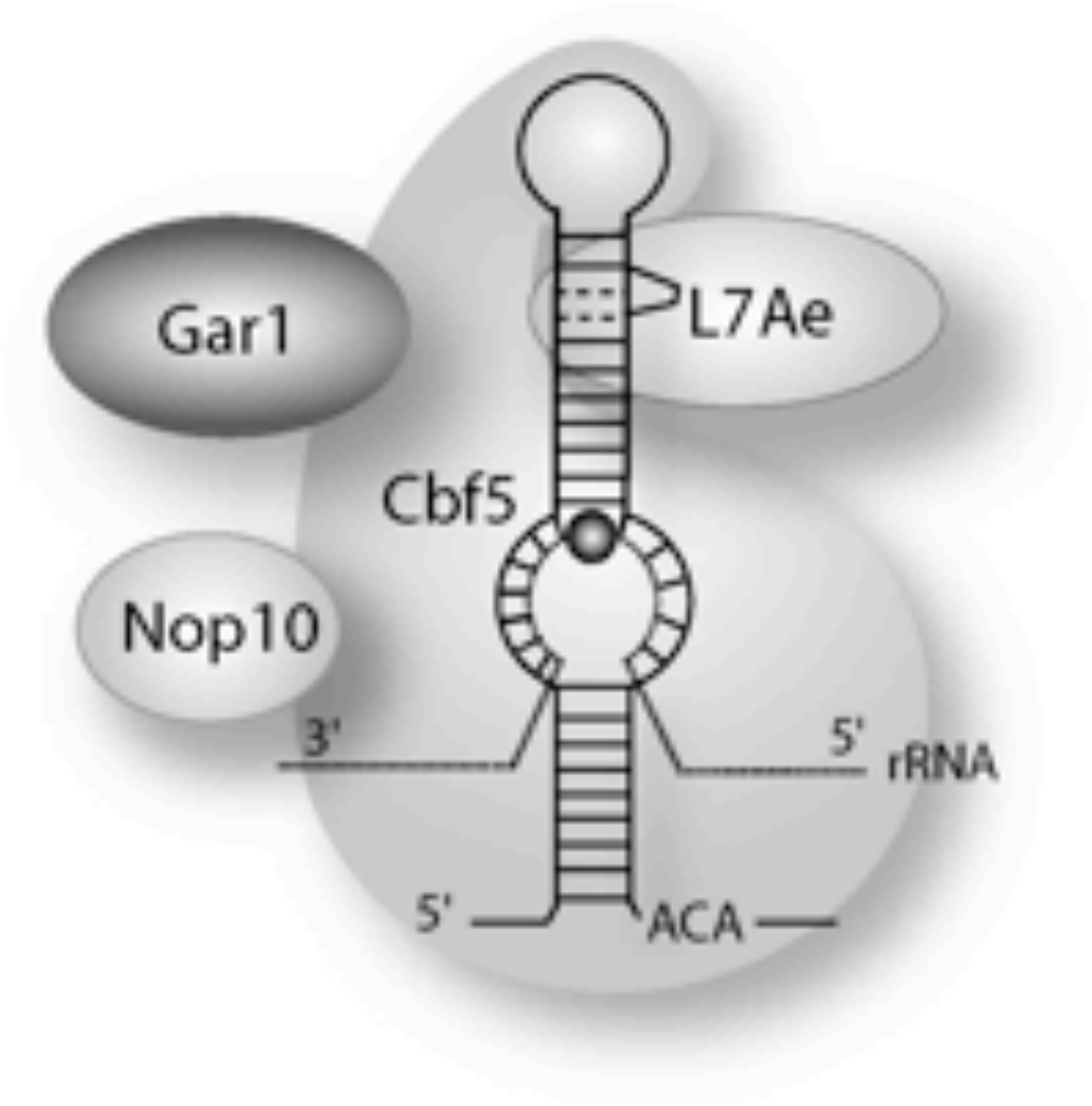

pf9 (wt)
pf9 (ΔACA)
Nop10 – + – – – + – + +
Gar1 – – + – – – + + +
Cbf5 – – – + – + + + +
L7Ae – – – – + – – – +
*
1 2 3 4 5 6 7 8 9
10 11 12 13 14 15 16 17 18

Figure 2.7: Organization of an archaeal pseudouridylation guide RNP complex. The results of this study suggest the model that is shown. L7Ae interacts directly with the K-turn of the guide RNA. Cbf5 also interacts directly and independently with the guide RNA, making extensive contacts that may include box ACA, the pseudouridylation pocket and the terminal loop. Association of Gar1 and Nop10 with the complex is mediated by their individual interactions with Cbf5. Close contacts between the various components may occur in the context of the assembled RNP, but no evidence of additional independent interactions was obtained in this study.

RNA/Protein Interaction Summary



CHAPTER 3

DYNAMIC INTERACTIONS WITHIN SUB-COMPLEXES OF H/ACA PSEUDOURIDYLATION GUIDE RNP[1]

[1] Osama A. Youssef, Rebecca M. Terns, and Michael P. Terns, Nucleic Acid Res. (2007), 35(18):6196-6206
Reprinted here with permission of publisher

Abstract

H/ACA RNP complexes change uridines to pseudouridines in target non-coding RNAs in eukaryotes and archaea. H/ACA RNPs are comprised of a guide RNA and four essential proteins: Cbf5 (pseudouridine synthase), L7Ae, Gar1 and Nop10 in archaea. The guide RNA captures the target RNA via two anti-sense elements brought together to form a contiguous binding site within the pseudouridylation pocket (internal loop) of the guide RNA. Cbf5 and L7Ae interact independently with the guide RNA, and here we have examined the impacts of these proteins on the RNA in nucleotide protection assays. The results indicate that the interactions observed in a fully assembled H/ACA RNP are established in the sub-complexes, but also reveal a unique Cbf5-guide RNA interaction that is displaced by L7Ae. In addition, the results indicate that L7Ae binding at the kink (k)-turn of the guide RNA induces the formation of the upper stem, and thus also the pseudouridylation pocket. Our findings indicate that L7Ae is essential for formation of the substrate RNA binding site in the archaeal H/ACA RNP, and suggest that k-turn binding proteins may remodel partner RNAs with important effects distant from the protein binding site.

Introduction

In all organisms, post-transcriptional modifications play an important role in the maturation and function of cellular RNAs, especially stable non-coding RNAs [1-4]. The human ribosome is estimated to contain over 200 modified nucleotides and these fall primarily in functionally important regions of the rRNAs [1,5,6]. In eukaryotes and archaea, rRNAs and other non-coding RNAs are modified by two classes of RNA-guided modification enzymes: C/D and H/ACA RNPs [7-9]. C/D RNPs methylate the 2'-*O*-hydroxyl group of ribose rings in target

nucleotides [10,11]. H/ACA RNPs isomerize target uridine residues to pseudouridines by base rotation [12,13]. These modification enzymes are comprised of a set of three or four core proteins and a cognate guide RNA that determines the target nucleotide by base-pairing with the substrate RNA [7-9].

Some key aspects of the mechanism of H/ACA RNP function have been well defined [3,14]. Seminal studies revealed that the substrate recognition site is formed by juxtaposing two antisense sequences within an internal loop of the conserved hairpin structure of the guide RNA (see Figure 3.1B) [13,15]. This loop that comprises the substrate recognition site is termed the pseudouridylation pocket. The antisense elements recognize substrate sequences flanking the target uridine, resulting in placement of the uridine to be modified at the apex of the pseudouridylation pocket. It is quite clear based on sequence and structure homology, and mutational analysis that Cbf5 is the pseudouridine synthase [16,17]. The functions of the other three proteins, Gar1, Nop10 and L7Ae (or Nh2p in eukaryotes) are not established, but are known to be essential for the function of the complex [18,19].

Our laboratory and the Branlant laboratory successfully reconstituted and characterized functional H/ACA RNPs using components from *Pyrococcus furiosus* and *Pyrococcus abyssi*, respectively [18,19]. These studies established that the four core proteins and a guide RNA are necessary and sufficient for full activity *in vitro*. We found that both Cbf5 and L7Ae interact directly with the guide RNA in the absence of other proteins. The remaining proteins, Gar1 and Nop10, bind to independent sites on Cbf5.

L7Ae belongs to a family of proteins that interact with RNA kink (k)-turns [20]. The k-turn binding proteins also include components of the ribosome, proteins involved in the assembly of spliceosomes, mRNA binding proteins and components of the RNase P and MRP complexes

that function in tRNA and rRNA processing [21-24]. L7Ae appears to be important for the kinetics of pseudouridylation by the H/ACA RNP [19]. The primary interaction of L7Ae within the H/ACA RNP is with the k-turn of the guide RNA; no substantial interaction with the other proteins is observed in the absence of the RNA [18]. Moreover, L7Ae binding is not required for association of the other three proteins with the guide RNA, though it may enhance their binding [18,19]. L7Ae binding sites are located either nearby (canonical k-turn) or overlapping (non-canonical k-turn) the apical loop of archaeal H/ACA RNAs [16,25]. The essential role of L7Ae in H/ACA RNP function was not apparent, but seemed likely to be accomplished through its interaction with the RNA component.

Mutational analysis in combination with RNA-protein binding assays indicated that Cbf5 requires several important elements of the guide RNA for its interaction, including sequences in the apical loop, pseudouridylation pocket and box ACA (Figure 3.1B), suggesting that Cbf5 may interact with these regions of the RNA [18]. A subsequent crystal structure of the *P. furiosus* H/ACA RNP (including the four proteins and a guide RNA) indicates that in the context of the complete complex, Cbf5 interacts with box ACA and nucleotides in the lower stem, and to a lesser extent with the apex of the pseudouridylation pocket [26]. Similar interactions were mapped in RNA footprinting studies with yeast Cbf5 [27]. No interaction of Cbf5 with the apical loop was observed in the holoenzyme [26].

In this work, we have examined the impacts of Cbf5 and L7Ae, both individually and in combination, on a guide RNA by enzymatic and chemical footprinting. The influences of the proteins on the RNA footprinting patterns substantiate and clarify the RNA-protein interactions predicted by the previous mutational analysis and observed in the crystal structure of the full complex [18,26]. In addition, the results indicate that L7Ae plays an important role in formation

of the pseudouridylation pocket (i.e. substrate recognition site). Finally, we observed an interaction of Cbf5 with the apical loop of the RNA that is disrupted by the binding of L7Ae. Our results indicate that RNA remodeling events triggered by binding of specific components of the H/ACA RNP govern the ability of the RNP to function in target RNA recognition and nucleotide modification.

Results

RNA-Protein Interactions in Cbf5-Guide RNA and L7Ae-Guide RNA Complexes

To assess the interactions of Cbf5 and L7Ae with *P. furiosus* H/ACA guide RNA Pf9, we analyzed the two RNA-protein sub-complexes (Cbf5-Pf9 and L7Ae-Pf9) by hydroxyl radical nucleotide protection assays (Figure 3.1). Hydroxyl radicals cleave the RNA backbone independent of RNA sequence or secondary structure [28,29]. Thus, in the absence of the proteins cleavages were observed at all ribose moieties (Figure 3.1A, lanes 4, 11, 16). Protection of a ribose from hydroxyl radical cleavage upon addition of a protein generally indicates a direct association with the protein [30]. RNA-protein complex formation (with 5' end-labeled Pf9 RNA and purified recombinant proteins) was verified by gel shift analysis (Figure 3.2). The majority of the Pf9 is shifted into RNA-protein complexes at 2 μM Cbf5 and 1 μM L7Ae. Both proteins provided some global protection of the RNA, however, as can be seen in Figure 3.1A, distinct RNA protections were observed with increasing Cbf5 or L7Ae concentrations (see regions indicated with blue and green bars).

Figure 3.1B shows the protection results in the context of a secondary structure model of Pf9 RNA that is based on the well-defined, functional features of the H/ACA RNA family [3]. The pseudouridylation pocket of the H/ACA RNA is the bipartite target recognition site,

established and bounded by the upper and lower stems. The predicted pseudouridylation pocket of Pf9 is complementary to sequences that flank 16S rRNA U910 (i.e. nts. 905-917), and consistent with this model, we have confirmed that U910 is modified in rRNA extracted from *P. furiosus* (Marshburn, S., Terns, R. and Terns, M., unpublished data). Box ACA, the signature sequence element, is located 3' of the lower stem. The k-turn of Pf9 is found within the upper stem, adjacent to the apical loop. Canonical k-turns are helix-bulge-helix structures that produce an ~120^{o} bend between the axes of the two adjacent RNA helices [20]. The bulge of a k-turn is bounded by two G-A base pairs that terminate the first helix and a G-C base pair that initiates the second helix. The motif generally includes several flanking base pairs (Figure 3.1B).

In multiple studies, L7Ae and its close homologs have been found to interact directly with sequences in the k-turn of partner RNAs [18,25,31-33], and as expected, L7Ae provides strong protection on both strands of the k-turn (green shading, Figure 3.1). In particular, the k-turn- binding proteins are consistently found to contact nucleotides in the bulge of the k-turn [18,25,31-33] and L7Ae's protection of Pf9 includes the bulge (Figure 3.1).

The interaction of Cbf5 with H/ACA RNAs is less well studied. Previous gel shift analysis suggested that box ACA, the pseudouridylation pocket and sequences in the apical loop of the RNA may be involved in the interaction; alterations in these elements affect the stability of the Cbf5-guide RNA complex [18]. In the crystal structure of a complex that includes Nop10, Gar1 and L7Ae as well as Cbf5 and a guide RNA, contacts were observed between nucleotides in box ACA, the lower stem and the pseudouridylation pocket of the RNA and Cbf5 [26]. As can be seen in Figure 3.1 (blue shading), Cbf5 provides extensive protection of Pf9 from hydroxyl radical cleavage. Cbf5 significantly reduces cleavage of the 3' strand of the lower stem, the 5' strand of the pseudouridylation pocket, and the 5' half of the k-turn and the apical loop.

Additional weak protection was observed along the 3' strand of pseudouridylation pocket and the upper stem. The 5' strand of the lower stem and the 3' single-stranded region that contains box ACA were not assessed in this experiment due to resolution limitations. The results suggest that Cbf5 interacts directly with the lower stem, pseudouridylation pocket, apical loop and 5' strand of the k-turn in the absence of other proteins.

The observed protections were specific to the individual proteins with the notable exception of the 5' strand of the k-turn (Figure 3.1). Interestingly, the results reveal strong protection of the 5' strand of the k-turn by both Cbf5 and L7Ae, suggesting that both proteins interact with this region of the RNA in the sub-complexes. Direct interactions of Cbf5 and L7Ae with the 5' side of the k-turn would be expected to be mutually exclusive.

RNA-Protein Interactions in Cbf5-L7Ae-Guide RNA Complexes

Cbf5 and L7Ae do not interact directly, either independently [18] or in the context of the fully assembled complex [26], and therefore should not directly influence the interaction of the other protein with the guide RNA; however we were interested in the possibility of effects translated through the RNA (for example, via changes in RNA structure). Moreover, we were interested in examining the footprint in the presence of both proteins on the 5' side of the k-turn (where both proteins were observed to bind, Figure 3.1). To test for any impact of one protein on the interaction of the other protein with the RNA, we examined the hydroxyl radical footprints of combinations of Cbf5 and L7Ae on Pf9 (Figure 3.3). In these experiments, the RNA was mixed with increasing concentrations of one protein in the presence of a constant concentration of the other protein in final amounts that promote nearly complete incorporation of the RNA into complexes (Figure 3.2, lanes 10-14).

For the most part, the guide RNA protection pattern in the presence of both proteins (Figure 3.3, lanes 6 and 12) appears to be the simple sum of the patterns obtained with the individual proteins (Figure 3.3, lanes 2 and 8), however there are several interesting exceptions. Examination of multiple experiments and exposures revealed two regions where greater protection is observed than would be expected from the individual protections (Figure 3.3, turquoise shading). In the 3' strand of the k-turn, several nucleotides (nts 39, 40, 44 and 45) are partially protected by L7Ae, and not significantly protected by Cbf5, but are nearly completely protected in the presence of both proteins (Figure 3.3A, lanes 6 and 12, and indicated in Figure 3.3B). Similarly, little protection of sequences in the 5' strand of the upper stem was observed with either L7Ae or Cbf5 binding, however cleavage of most nucleotides in this region (19, 20, 22-24) is reduced by the combination of the two proteins (Figure 3.3). These increased protections likely reflect enhanced interaction of the proteins with these regions when the other protein is bound to the RNA.

In contrast, protection of nucleotides in the apical loop by Cbf5 (nts 30, 31) is lost upon introduction of L7Ae (Figure 3.3A, lane 8 versus lane 9, purple shading). The de-protected nucleotides are part of a larger, contiguous Cbf5 binding site that also includes the 5'strand of the k-turn, the region where both Cbf5 and L7Ae interact with the RNA (Figure 3.1). The loss of the Cbf5 protection pattern suggests that L7Ae disrupts or prevents the interaction of Cbf5 with the 5' k-turn/apical loop region. This is consistent with the absence of an interaction between the k-turn/apical loop of the RNA and Cbf5 in the crystal structure of the full complex [26].

Interaction of Cbf5 with the Conserved ACA Sequence

Box ACA is the signature sequence of the H/ACA guide RNAs, and RNA-protein binding studies and crystal structure data indicate that Cbf5 specifically binds and recognizes this

family feature [18,19,26]. We examined the individual impacts of Cbf5 and L7Ae on the 3' half of the RNA including box ACA by lead (II) acetate cleavage. Lead (II) acetate induces cleavage preferentially at single-stranded and dynamic regions of RNAs such as bulges and loops [34,35]. As expected from the predicted secondary structure, the 3' strand of the lower stem of Pf9 is inaccessible to lead-induced cleavage in the absence of proteins (see lack of cleavage between 3' pseudouridylation pocket and ACA in Figure 3.4A, lanes 5, 12, 16). However, the 3' strand of the upper stem was unexpectedly sensitive to cleavage (indicated with red arrowheads in Figure 3.4A, lane 12).

Binding of Cbf5 to Pf9 RNA results in substantial protection of the conserved ACA sequence as well as the nucleotide immediately upstream (Figure 3.4, blue shading). Protection of most of the lower stem could not be assessed in this experiment (because this region is already insensitive to lead-induced cleavage), however we observed that Cbf5 also provides some protection to the 3' strand of the pseudouridylation pocket. In contrast, L7Ae does not protect these regions (Figure 3.4, lanes 13-15). The addition of L7Ae to the RNA results in reduced cleavage of the 3' strand of the k-turn and also of the upper stem outside of the k-turn motif (Figure 3.4, green shading). L7Ae also produced increased sensitivity to lead-induced cleavage in nucleotides in the 3' half of the pseudouridylation pocket (Figure 3.4, yellow shading).

Guide RNA Secondary Structure

The unexpected sensitivity of the 3' side of the upper stem of Pf9 to lead-induced cleavage (Figure 3.4) led us to further probe the secondary structure of the RNA in the absence of proteins. We performed partial enzymatic digestions of 5' end labeled Pf9 RNA using RNase T1 and RNase A (Figure 3.5). RNases T1 and A cleave accessible phosphodiester backbones

following un-base paired guanines (Gs) and pyrimidines (Cs and Us), respectively. Adenines (As) are not subject to analysis. Base paired regions are resistant to both enzymes. The results are summarized schematically in the context of the predicted secondary structure in Figure 3.5B. Under the experimental conditions analyzed, in the absence of proteins, accessible regions of the RNA include the 3' tail, the pseudouridylation pocket, the apical loop and the bulge of the k-turn motif as expected (Figure 3.5B, orange shading). Consistent with the predicted secondary structure, nucleotides in the lower stem are inaccessible to the single-stranded nucleases. In addition, nucleotides in this region are susceptible to RNase V1, a nuclease specific for double-stranded regions (data not shown). However, we found that the upper stem is sensitive to single-stranded nuclease digestion [see strong cleavages in red boxed region (Figure 3.5B) and indicated by red arrowheads (Figure 3.5A)]. The results suggest that the upper stem of Pf9 RNA is not stably structured under these experimental conditions.

Effects of L7Ae on the Guide RNA Outside the K-Turn Motif

Although the upper stem of the RNA does not appear to be firmly established in the absence of proteins (Figures 3.4 and 3.5), the interaction of L7Ae with sequences on both sides of the k-turn motif (Figure 3.1) strongly implies the existence of a helix within the upper stem of Pf9 RNA in the presence of the protein. Moreover, the L7Ae-induced resistance of the upper stem beyond the k-turn motif (i.e. outside the region where L7Ae has been shown to directly contact RNA) to lead-induced cleavage (Figure 3.4) suggests the formation of the stem in the presence of the protein. To further investigate the impact of L7Ae on the secondary structure of the RNA, we also analyzed partial enzymatic digestions in the presence of the protein (Figure 3.6).

The strong ribonuclease T1 and A cleavages observed in the "upper stem" both within and outside the k-turn motif in the absence of the protein (red arrowheads, Figure 3.6A, lanes 3 and 5) are significantly reduced upon L7Ae binding (Figure 3.6A, lanes 4, 6). At the same time, L7Ae increases cleavage of the apical loop (Figure 3.6B, yellow shading). The interaction of L7Ae with the k-turn is well documented [18,25,31-33]. No extensive interactions outside the k-turn have been described. The strong protection that we observe in the upper stem, as well as the increased sensitivity in the apical loop, is consistent with formation of the upper stem upon L7Ae binding.

Discussion

In this work we examined the arrangements of the RNAs and proteins in a series of sub-complexes of the H/ACA RNP by chemical and enzymatic footprinting. The combined approach has the potential to detect both physical protein interaction sites and effects on RNA configuration, and we found evidence for both types of impacts in this study. The results provide detailed insight on steps in the assembly of the complex and essential roles of the proteins.

A significant amount is known about the sites of RNA-protein interaction within the fully assembled H/ACA guide RNP from the crystal structure of the *P. furiosus* holoenzyme (using a modified Afu 46 guide RNA) [26]. In the Cbf5-guide RNA and L7Ae-guide RNA sub-complexes that we examined here, we observed footprints consistent with the well-established interaction of L7Ae with the k-turn [18,25,31-33] and with the contacts observed between Cbf5 and the guide RNA in the holoenzyme crystal structure [26]. Our results support the extensive interaction of Cbf5 with the guide RNA from box ACA through the lower stem to the pseudouridylation pocket (observed in RNA-protein binding assays [18,19], holoenzyme crystal

structure [26] and RNA footprinting of the eukaryotic complex [27]). The results indicate that the interactions of both Cbf5 and L7Ae with the guide RNA in the fully assembled enzyme are established in the Cbf5-guide RNA and L7Ae-guide RNA sub-complexes.

In addition however, we found evidence of an interaction between Cbf5 and the guide RNA that is unique to the sub-complex. In the absence of other proteins, Cbf5 protects the 5' strand of the k-turn and apical loop from hydroxyl radical cleavage, an effect that generally reflects a physical interaction (Figure 3.1). Moreover, previous studies showed that mutation of this region of the RNA weakens the binding of Cbf5 [18], supporting the existence of the interaction and suggesting that the interaction is important in formation and stability of the sub-complex. In the crystal structure of the holoenzyme, Cbf5 is not found in proximity with the apical loop and the 5' strand of the k-turn [26]. Our results indicate that L7Ae successfully competes for the site and displaces Cbf5 (Figures 3.3 and 3.7). Accordingly, L7Ae is found in close proximity with this region of the RNA in the crystal structure of the holoenzyme [26]. [The specific equivalent L7Ae-RNA interactions could not be compared as the guide RNA used in the crystal structure differs significantly from Pf9 in this region (non-canonical k-turn) and its structure is also incomplete in this region [26].]

Because the intermediates in the assembly and function of the H/ACA RNP have not been precisely defined, it is not yet clear what role the newly identified Cbf5-guide RNA interaction may play in H/ACA RNP assembly or function. In eukaryotes, evidence indicates that three of the four core proteins, including Cbf5, assemble on the H/ACA RNA at the site of transcription (and that association of Gar1 occurs at a later point in the temporal and spatial assembly pathway) [3,7,8]. Among the core proteins, Cbf5 shows the strongest association with the H/ACA RNA genes, suggesting that Cbf5 could be the first of the H/ACA RNP proteins to

associate with the newly made guide RNA in yeast [36]. Thus, the Cbf5 interactions defined here may provide for the initial recognition of the guide RNA and subsequent complex assembly. It is also possible that a sub-complex lacking L7Ae is involved in the function of the H/ACA RNP (for example, as a step in substrate release).

Our studies also revealed a substantial effect of L7Ae on the guide RNA configuration beyond the k-turn with significant implications for proper establishment of the target recognition site. Previous studies had shown that the structure of the k-turn motif itself is dynamic in the absence of protein and that formation of the kink (i.e. 120° bend from linear) is induced by the binding of L7Ae and related proteins [37-40]. Our results indicate previously undescribed effects on the RNA beyond this region (Figure 3.7). Our data from both partial enzymatic hydrolysis (Figure 3.5) and lead-induced cleavage (Figure 3.4) indicate that the upper stem of the guide RNA is not stably formed in the absence of proteins under the solution conditions used in our work. However, upon addition of L7Ae the upper stem nucleotides become resistant to single-stranded nucleases (Figure 3.6) and lead-induced cleavage (Figure 3.4), strongly suggesting L7Ae-induced formation of the upper stem. The observed increases in the sensitivity of nucleotides in the apical loop and pseudouridylation pocket to single stranded nucleases (Figure 3.6) and lead-induced cleavage (Figure 3.4) upon L7Ae binding are also consistent with formation of the upper stem, which defines these loops. While L7Ae provided strong protection of the upper stem against enzymatic cleavage (Figure 3.6), this region was not significantly protected from hydroxyl radical cleavage outside of the k-turn motif (Figure 3.1), providing further evidence that the observed protection of this region from enzymatic cleavage reflects induction of basepairing (rather than steric interference). Importantly, the upper stem establishes the pseudouridylation pocket - the H/ACA RNP target RNA binding site.

Our results reveal that L7Ae plays a significant role in substrate binding and placement in the archaeal H/ACA RNP via formation of the pseudouridylation pocket. These findings may explain the positive impact that L7Ae appears to have on formation of substrate-containing H/ACA RNP complexes and on activity of the complex [18,19]. In addition, analysis of a crystal structure of a sub-complex of the H/ACA RNP with a substrate RNA recently obtained by Hong Li's laboratory suggests that both pseudouridylation pocket formation and positioning of the substrate uridine in the Cbf5 active site are defective in the absence of L7Ae (Liang, B., Xue, S., Terns, R., Terns, M. and Li. H., submitted for publication). At the same time, this and several other recent studies describe guide-substrate RNA interactions in the absence of L7Ae [41,42]. It is most likely that the difference reflects the high concentrations of molecules used in these structural studies [41,42]. In the cell, substrate capture likely depends on a well-formed pseudouridylation pocket established by L7Ae.

Given that the pseudouridine synthase Cbf5 can interact directly with the guide RNA, which has the capacity to capture and present the substrate, it was previously not clear why L7Ae should be needed. Our findings indicate that the importance of L7Ae in the function of the H/ACA RNP is in remodeling the guide RNA to form the substrate binding site. In eukaryotes, the H/ACA RNP protein homologous to L7Ae is Nhp2, a protein with less well-defined RNA binding properties [43], and it remains to be determined whether Nhp2 will also play a role in definition of the substrate binding site in the eukaryotic H/ACA RNPs. However, L7Ae is also a component of C/D RNPs and the ribosome in archaea [44], and our findings suggest that L7Ae and other k-turn binding proteins could play a similar role in important alterations of RNA structure beyond the k-turn in other complexes as well.

Materials and Methods

Protein expression and purification

Cbf5 and L7Ae genes were amplified by PCR from *P. furiosus* genomic DNA and subcloned into a modified version of pET21D expression vector as previously described [18]. The resultant recombinant proteins containing N-terminal 6x histidine tags, were purified by affinity chromatography on Ni-NTA agarose (Qiagen), eluted with buffer A (20 mM sodium phosphate, pH 7.0, 1 M NaCl, 350 mM imidazole) and quantified using BCA protein assay (Pierce). Prior to use in RNA binding assays, the proteins were dialyzed against 40 mM HEPES-KOH, pH 7.0, 1 M KCl (or K-acetate).

End labeling of H/ACA RNA

The single hairpin, *P. furiosus* H/ACA RNA Pf9 was transcribed *in vitro* from PCR-amplified DNA product containing a SP6 promoter using SP6 RNA polymerase (Epicentre Biotechnologies) as previously described [18]. RNA was gel-purified by electrophoresis through a 15% polyacrylamide/7 M urea gel. Purified RNA was ethanol precipitated and washed with 70% ethanol. Purified RNA was dephosphorylated with calf intestinal alkaline phosphatase according to the manufacturer's protocol (Ambion). The dephosphorylated RNA was ^{32}P labeled with T4 polynucleotide kinase (Ambion) and [γ-^{32}P]ATP (7000 Ci/mmol, MP Biomedicals). 5'-end labeled RNA was then gel purified as described above.

Gel mobility shift assay

Reconstitution of RNP complexes was performed as described previously [18]. Briefly, 5'-end radiolabeled RNA (0.05 pmol) was incubated in buffer B (20 mM HEPES-KOH, pH 7.0, 500 mM KCl, 1.5 mM $MgCl_2$, 5 μg *E. coli* tRNA) alone or with various concentrations of

protein in a final volume of 20 μL for 1 hr at 65°C. RNP complexes were analyzed on an 8% non-denaturing polyacrylamide gel and visualized by autoradiography.

Enzymatic and chemical probing

^{32}P-end labeled Pf9 RNA (0.05 - 0.1 pmoles) was incubated in the absence (free RNA) or presence (RNPs) of increasing concentrations of purified Cbf5 or L7Ae proteins for 1 hr at 65°C in buffer B (described above) in a final volume of either 20 or 50 μl. For ribonuclease cleavage, the reactions were initiated by addition of 0.1 U or 0.2 U RNase T1 (Sigma), or 1 ng or 2 ng RNase A (Sigma) and incubated for 15 min at 37°C. The enzymatic reactions were stopped by extraction with phenol/chloroform/isoamyl alcohol. Hydroxyl radical footprinting experiments were performed essentially as described [45]. Briefly, the cleavage reactions were initiated by adding freshly prepared 18 μM ethylenediaminetetraacetic acid iron (III) sodium salt dihydrate (Aldrich), 2 mM sodium ascorbate (Sigma) and 0.14% (v/v) H_2O_2 (Sigma). The reactions were carried out at 65°C for 30 seconds and stopped by addition of 1 mM thiourea (Aldrich) followed by phenol/chloroform/isoamyl alcohol extraction. For lead (II) footprinting, the reactions were carried out in a modified buffer B where the KCl was substituted with 200 mM K acetate. Lead cleavage was performed essentially as previously described [34] with 15 mM Pb(II) acetate (Merck) freshly prepared in sterile water. The reactions were performed at room temperature for 10 minutes and were stopped by adding EDTA to final concentration of 20 mM before ethanol precipitation. As sequence markers, RNA alkaline hydrolysis ladders (cleavage after each nucleotide) were generated by incubating RNA with 5 μg *E. coli* tRNA in 50 mM sodium carbonate at pH 9.5, 1 mM EDTA for 5 min at 90°C. RNase T1 ladders (ΔT1) (cleavage after each guanosine) were generated by incubating the RNA in 20 mM sodium citrate at pH 4.5, 1

mM EDTA, 7 M urea for 10 min at 50°C. For both enzymatic and chemical probing reactions, the treated RNA samples were then ethanol precipitated in the presence of 0.3 M sodium acetate at pH 5.2 followed by washing with 70% ethanol. The dried RNA pellets were resuspended in RNA loading dye (10 M urea, 2 mM EDTA, 0.5% (w/v) SDS, 0.02% (w/v) each bromophenol blue and xylene cyanol). The cleavage products were separated on 15% or 20% polyacrylamide (acrylamide:bis ratio 19:1) 7M urea-containing gel and visualized by autoradiography.

Acknowledgements

We thank Hong Li for sharing results prior to publication and for stimulating collaborations. We thank Alex Huttenhofer and Pascale Romby for sharing detailed RNA footprinting protocols. This work was supported by NIH grant RO1 GM54682 to M.T. and R.T. and by a fellowship from the Egyptian government to O.Y. Funding to pay the Open Access publication charges for this article was provided by NIH grant RO1 GM54682 to M.T. and R.T.

References

1. Decatur WA, Fournier MJ: **rRNA modifications and ribosome function**. *Trends Biochem Sci* 2002, **27**:344-351.

2. Lane BG, Ofengand J, Gray MW: **Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins**. *Biochimie* 1995, **77**:7-15.

3. Terns M, Terns R: **Noncoding RNAs of the H/ACA family**. *Cold Spring Harb Symp Quant Biol* 2006, **71**:395-405.

4. Yu YT, Terns RM, Terns MP: **Mechanisms and functions of RNA-guided RNA modification**. In *Fine-tuning of RNA functions by modification and editing*. Edited by Grosjean H: Topics in Current Genetics; 2005:223-262. vol 12.

5. Maden BE: **The numerous modified nucleotides in eukaryotic ribosomal RNA**. *Prog Nucleic Acid Res Mol Biol* 1990, **39**:241-303.

6. Ofengand J: **Ribosomal RNA pseudouridines and pseudouridine synthases**. *FEBS Lett* 2002, **514**:17-25.

7. Kiss T, Fayet E, Jady BE, Richard P, Weber M: **Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs**. *Cold Spring Harb Symp Quant Biol* 2006, **71**:407-417.

8. Matera AG, Terns RM, Terns MP: **Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs**. *Nat Rev Mol Cell Biol* 2007, **8**:209-220.

9. Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G: **The structure and function of small nucleolar ribonucleoproteins**. *Nucleic Acids Res* 2007, **35**:1452-1464.

10. Kiss-Laszlo Z, Henry, Y., Bachellerie, J.-P., Caizergues-, Ferrer M, and Kiss, T.: **Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs**. *Cell* 1996, **85**:1077-1088.

11. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea**. *Science* 2000, **288**:517-522.

12. Balakin AG, Smith L, Fournier MJ: **The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions**. *Cell* 1996, **86**:823-834.

13. Ganot P, Bortolin ML, Kiss T: **Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs**. *Cell* 1997, **89**:799-809.

14. Meier UT: **How a single protein complex accommodates many different H/ACA RNAs**. *Trends Biochem Sci* 2006, **31**:311-315.

15. Ni J, Tien AL, Fournier MJ: **Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA**. *Cell* 1997, **89**:565-573.

16. Hamma T, Ferre-D'Amare AR: **Pseudouridine synthases**. *Chem Biol* 2006, **13**:1125-1135.

17. Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J: **Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA**. *Mol Cell Biol* 1999, **19**:7461-7472.

18. Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP: **RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP**. *Genes Dev* 2005, **19**:1238-1248.

19. Charpentier B, Muller S, Branlant C: **Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation**. *Nucleic Acids Res* 2005, **33**:3133-3144.

20. Klein DJ, Schmeing TM, Moore PB, Steitz TA: **The kink-turn: a new RNA secondary structure motif**. *EMBO J* 2001, **20**:4214-4221.

21. Allmang C, Carbon P, Krol A: **The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding**. *RNA* 2002, **8**:1308-1318.

22. Muslimov IA, Iacoangeli A, Brosius J, Tiedge H: **Spatial codes in dendritic BC1 RNA**. *J Cell Biol* 2006, **175**:427-439.

23. Rosenblad MA, Lopez MD, Piccinelli P, Samuelsson T: **Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes**. *Nucleic Acids Res* 2006, **34**:5145-5156.

24. Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, Ficner R: **Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment**. *Mol Cell* 2000, **6**:1331-1342.

25. Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A: **Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea**. *Nucleic Acids Res* 2003, **31**:869-877.

26. Li L, Ye K: **Crystal structure of an H/ACA box ribonucleoprotein particle**. *Nature* 2006, **443**:302-307.

27. Normand C, Capeyrou R, Quevillon-Cheruel S, Mougin A, Henry Y, Caizergues-Ferrer M: **Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA**. *RNA* 2006.

28. Celander DW, Cech TR: **Iron(II)-ethylenediaminetetraacetic acid catalyzed cleavage of RNA and DNA oligonucleotides: similar reactivity toward single- and double-stranded forms**. *Biochemistry* 1990, **29**:1355-1361.

29. Huttenhofer A, Noller HF: **Hydroxyl radical cleavage of tRNA in the ribosomal P site**. *Proc Natl Acad Sci U S A* 1992, **89**:7851-7855.

30. Latham JA, Cech TR: **Defining the inside and outside of a catalytic RNA molecule**. *Science* 1989, **245**:276-282.

31. Hamma T, Ferre-D'Amare AR: **Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution**. *Structure* 2004, **12**:893-903.

32. Moore T, Zhang Y, Fenley MO, Li H: **Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA**. *Structure* 2004, **12**:807-818.

33. Nolivos S, Carpousis AJ, Clouet-d'Orval B: **The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn**. *Nucleic Acids Res* 2005, **33**:6507-6514.

34. Brunel C, Romby P: **Probing RNA structure and RNA-ligand complexes with chemical probes**. *Methods Enzymol* 2000, **318**:3-21.

35. Lindell M, Romby P, Wagner EG: **Lead(II) as a probe for investigating RNA structure in vivo**. *RNA* 2002, **8**:534-541.

36. Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y, Chanfreau G: **Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly**. *Mol Cell Biol* 2005, **25**:3295-3304.

37. Goody TA, Melcher SE, Norman DG, Lilley DM: **The kink-turn motif in RNA is dimorphic, and metal ion-dependent**. *RNA* 2004, **10**:254-264.

38. Marmier-Gourrier N, Clery A, Senty-Segault V, Charpentier B, Schlotter F, Leclerc F, Fournier R, Branlant C: **A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA**. *RNA* 2003, **9**:821-838.

39. Mougin A, Gottschalk A, Fabrizio P, Luhrmann R, Branlant C: **Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles**. *J Mol Biol* 2002, **317**:631-649.

40. Turner B, Melcher SE, Wilson TJ, Norman DG, Lilley DM: **Induced fit of RNA on binding the L7Ae protein to the kink-turn motif**. *RNA* 2005, **11**:1192-1200.

41. Jin H, Loria JP, Moore PB: **Solution structure of an rRNA substrate bound to the seudouridylation pocket of a box H/ACA snoRNA**. *Mol Cell* 2007, **26**:205-215.

42. Wu H, Feigon J: **H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification**. *Proc Natl Acad Sci U S A* 2007, **104**:6655-6660.

43. Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M: **Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p**. *Nucleic Acids Res* 2001, **29**:2733-2746.

44. Terns MP, Terns RM: **Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin**. *Gene Expr* 2002, **10**:17-39.

45. Tullius TD, Dombroski BA, Churchill ME, Kam L: **Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts**. *Methods Enzymol* 1987, **155**:537-558.

Figure 3.1: Hydroxyl radical footprinting of Cbf5-Pf9 and L7Ae-Pf9 complexes. (A) 5'-end labeled Pf9 was incubated in the absence (lanes 4, 11, 16) or presence of increasing concentrations of Cbf5 (lanes 5-10) or L7Ae (lanes 12-15) and subjected to hydroxyl radical cleavage. Lane 1 is undigested RNA and lanes 2 and 3 are size markers generated by alkaline hydrolysis (OH) and RNase T1 digestion (T1) of the free RNA, respectively. Nucleotides corresponding to secondary structure landmarks are indicated to the right. Blue and green bars indicate regions of strong Cbf5 and L7Ae protection, respectively. (B) Summary of protections in the context of a functional secondary structure model of Pf9 RNA. Box ACA, the pseudouridylation pocket and k-turn are boxed. Apical loop, upper and lower stems are labeled. The rRNA target of Pf9 is shown in gray lowercase letters. Cbf5 and L7Ae protections observed in A are shown as indicated in the legend. The regions shaded gray were not assessed due to the resolution limits of the gel.

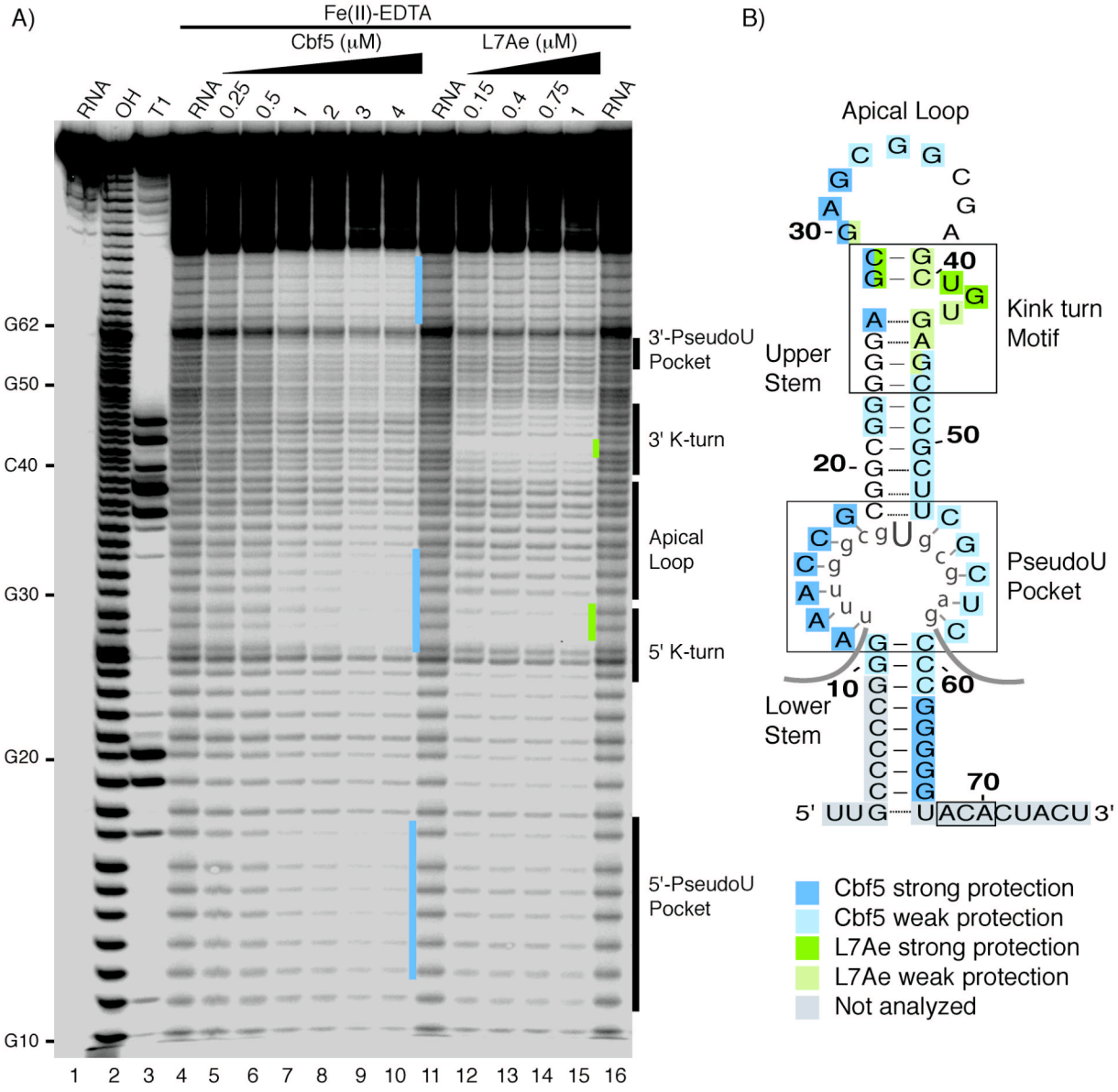

A)
Fe(II)-EDTA
Cbf5 (μM)
L7Ae (μM)
RNA
OH
T1
RNA
0.25
0.5
1
2
3
4
RNA
0.15
0.4
0.75
1
RNA
G62
G50
C40
G30
G20
G10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
3'-PseudoU Pocket
3' K-turn
Apical Loop
5' K-turn
5'-PseudoU Pocket
B)
Apical Loop
Kink turn Motif
Upper Stem
PseudoU Pocket
Lower Stem
30
40
50
20
10
60
70
5' UU G U ACACUACU 3'
Cbf5 strong protection
Cbf5 weak protection
L7Ae strong protection
L7Ae weak protection
Not analyzed

Figure 3.2: Reconstitution of Cbf5-Pf9, L7Ae-Pf9, and Cbf5-L7Ae-Pf9 sub-complexes. (A) Coomassie blue staining of purified recombinant Cbf5 and L7Ae following PAGE. M lane contains protein standards. (B) Gel mobility shift analysis of 5'-end labeled Pf9 RNA, alone (lane 1) or with increasing amounts of Cbf5 and/or L7Ae as indicated. Complexes were separated on an 8% native gel and visualized by autoradiography.

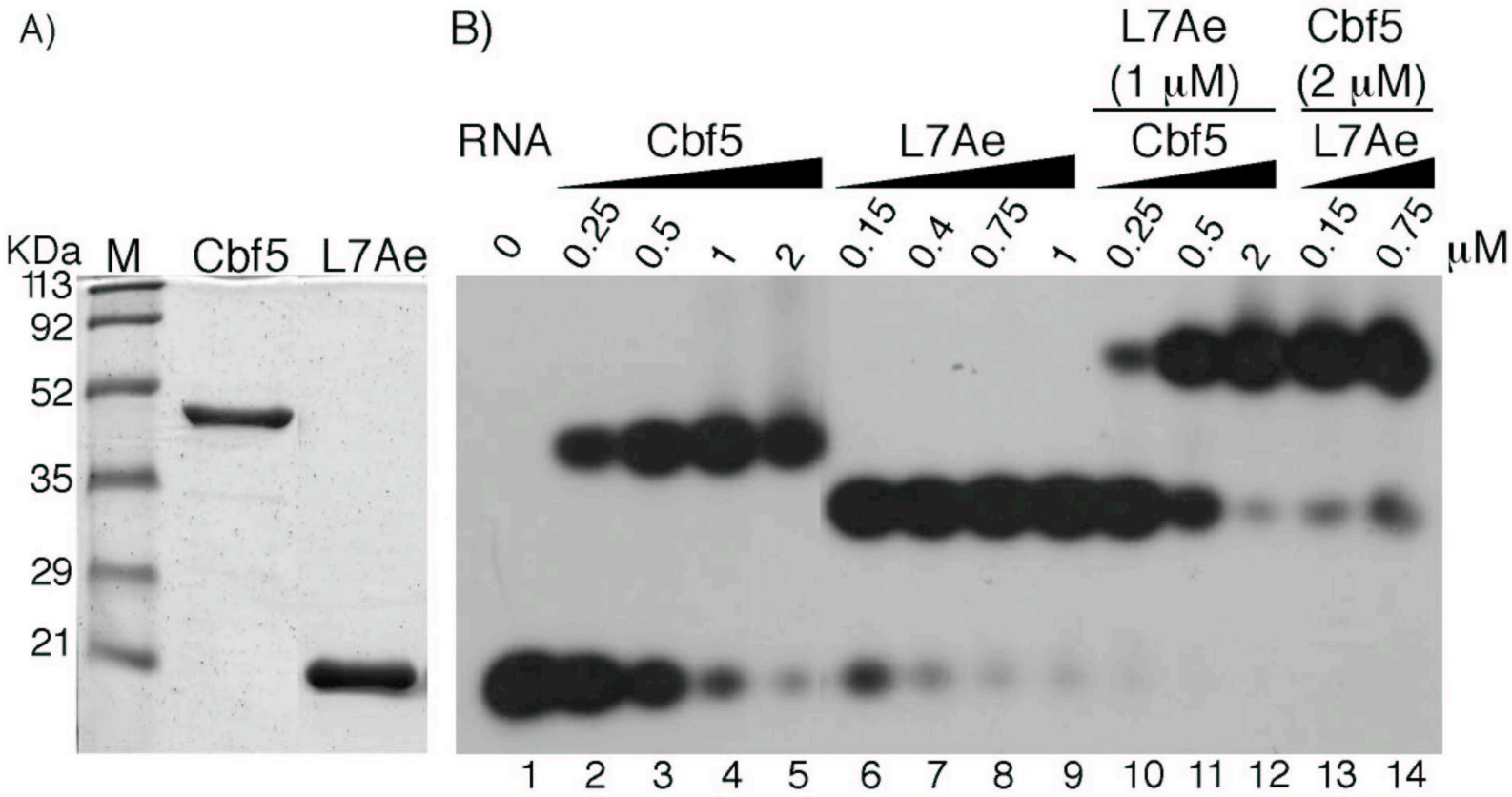

A)
KDa
M
Cbf5
L7Ae
113
92
52
35
29
21
B)
RNA
Cbf5
L7Ae
L7Ae (1 μM) Cbf5
Cbf5 (2 μM) L7Ae
0
0.25
0.5
1
2
0.15
0.4
0.75
1
0.25
0.5
2
0.15
0.75
μM
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 3.3: Hydroxyl radical footprinting of Cbf5-L7Ae-Pf9 complexes. (A) 5'-end labeled Pf9 was incubated in the absence (lanes 1, 7, 13) or presence of either 1 μM L7Ae (lane 2) with increasing concentrations of Cbf5 (lanes 3-6), or 2 μM Cbf5 (lane 8) with increasing concentrations of L7Ae (lanes 9-12), and subjected to hydroxyl radical cleavage. Lanes 14 and 15 are size markers generated by alkaline hydrolysis (OH) and RNase T1 digestion (T1). Turquoise and purple bars indicate new sites of protection observed in the presence of both proteins and of Cbf5 protections lost upon addition of L7Ae, respectively. (B) Summary of changes in protection in the context of Pf9 RNA secondary structure model (as in Figure 3.1).

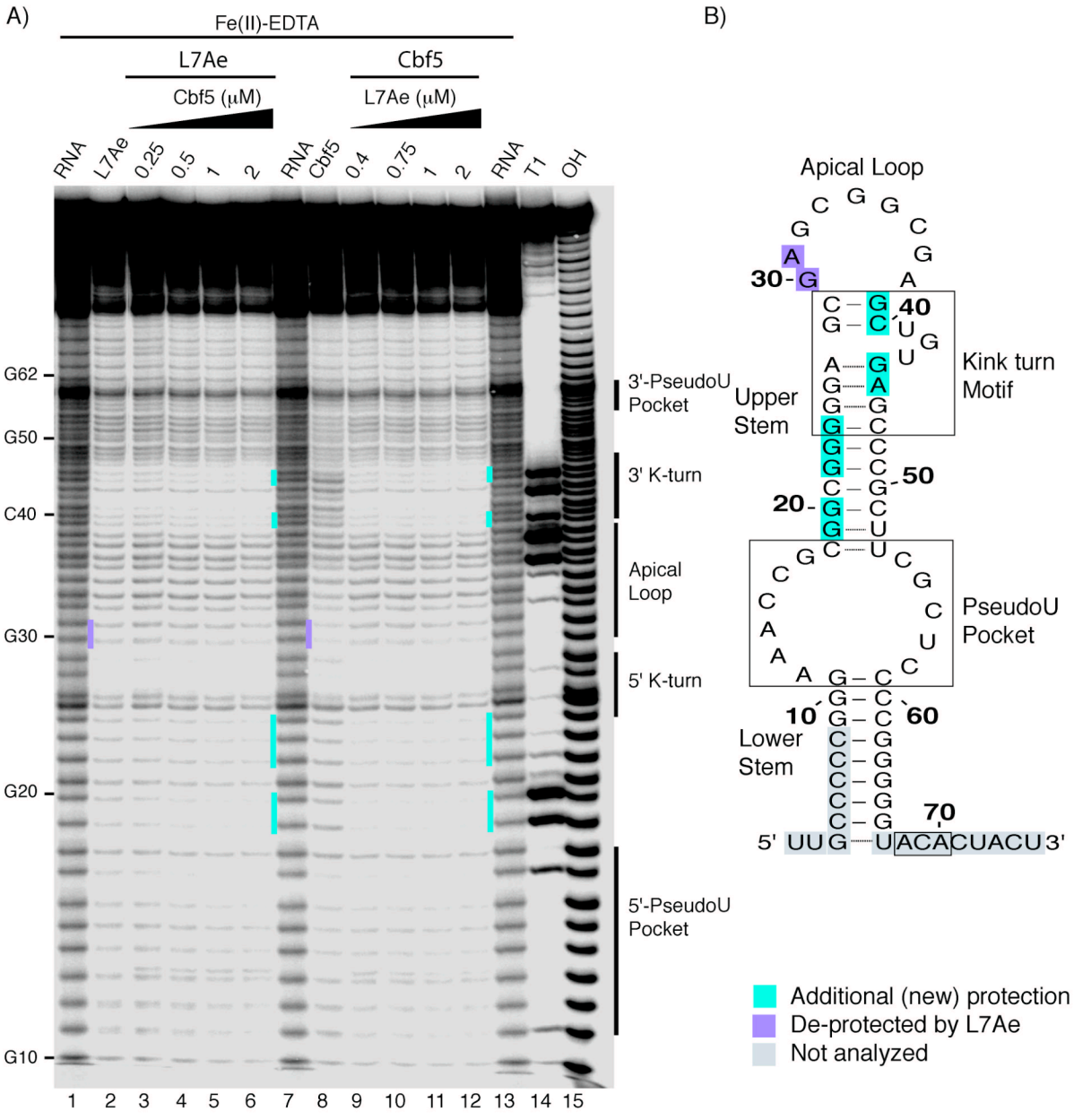

A)
Fe(II)-EDTA
L7Ae
Cbf5 (μM)
Cbf5
L7Ae (μM)
RNA
L7Ae
0.25
0.5
1
2
RNA
Cbf5
0.4
0.75
1
2
RNA
T1
OH
G62
G50
C40
G30
G20
G10
3'-PseudoU Pocket
3' K-turn
Apical Loop
5' K-turn
5'-PseudoU Pocket
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
B)
Apical Loop
Kink turn Motif
Upper Stem
PseudoU Pocket
Lower Stem
30
40
50
20
60
10
70
5' UUG U ACA CUACU 3'
Additional (new) protection
De-protected by L7Ae
Not analyzed

Figure 3.4: Lead-induced cleavage footprinting of Pf9 RNA, and Cbf5-Pf9 and L7Ae-Pf9 sub-complexes. (A) 5'-end labeled Pf9 was incubated in the absence (lanes 5, 12, 16) or presence of increasing concentrations of Cbf5 (lanes 6-11) or L7Ae (lanes 13-15) and subjected to lead (II)-induced cleavage. Lane 1 is undigested RNA and lanes 3, 2 and 4 are size markers generated by alkaline hydrolysis (OH), and RNase T1 digestion under non-denaturing (T1) and denaturing conditions (ΔT1), respectively. Blue and green bars indicate regions of strong Cbf5 and L7Ae protection. Yellow bars indicate cleavage enhancements observed with L7Ae. Red arrowheads indicate unexpected cleavages in the upper stem of the guide RNA in the absence of protein. (B) Summary of cleavage protections and enhancements in the context of Pf9 RNA secondary structure model (as in Figure 3.1).

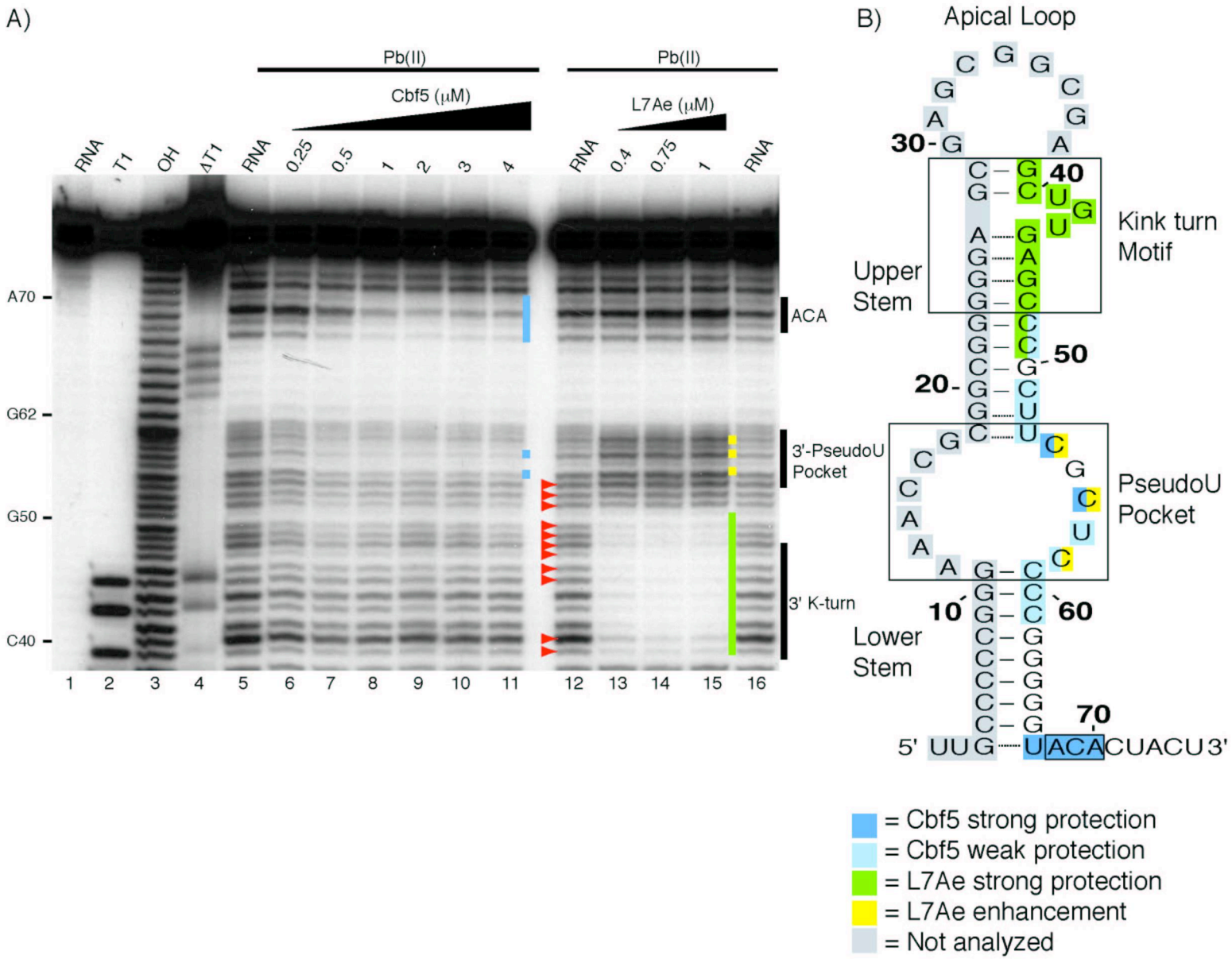

A)
Pb(II)
Pb(II)
Cbf5 (μM)
L7Ae (μM)
RNA
T1
OH
ΔT1
RNA
0.25
0.5
1
2
3
4
RNA
0.4
0.75
1
RNA
A70
G62
G50
C40
ACA
3'-PseudoU Pocket
3' K-turn
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
B)
Apical Loop
Kink turn Motif
Upper Stem
PseudoU Pocket
Lower Stem
30
40
50
20
10
60
70
5' UUG UACACUACU 3'
= Cbf5 strong protection
= Cbf5 weak protection
= L7Ae strong protection
= L7Ae enhancement
= Not analyzed

Figure 3.5: Single-stranded nuclease footprinting of Pf9. (A) 5'-end labeled Pf9 was digested with indicated concentrations of RNase A (lanes 3, 4) or RNase T1 (lanes 5, 6). The cleavage products were separated on a denaturing 20% acrylamide gel. Lane 1 is undigested RNA and lane 2 is a size marker generated by alkaline hydrolysis (OH). Strong cleavages at nucleotides in the upper stem region are indicated with red arrowheads. (B) Summary of Pf9 RNA cleavages by single-stranded nucleases in the context of the predicted secondary structure of Pf9 RNA (as in Figure 3.1). Upper stem region is boxed in red.

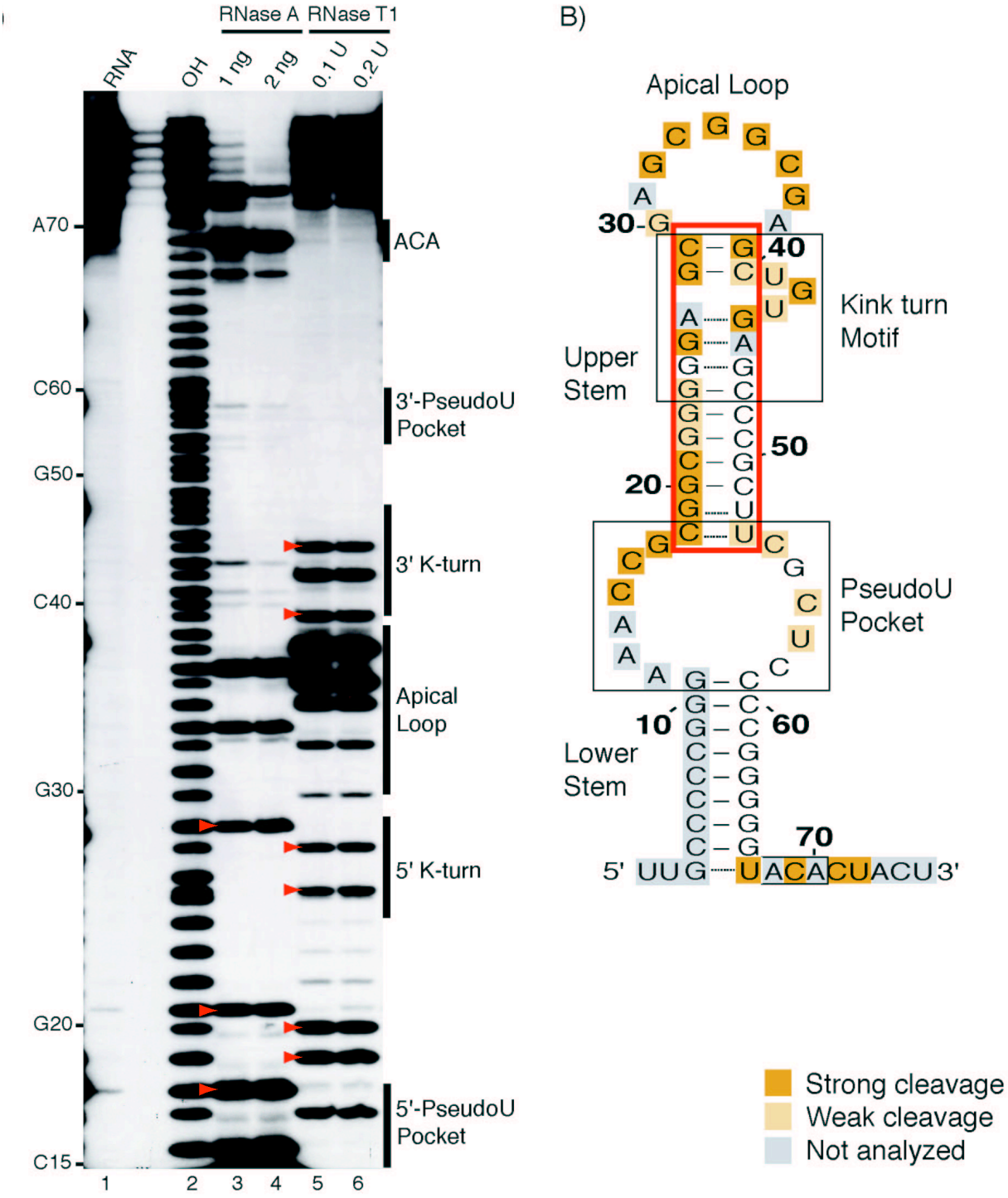

RNase A
RNase T1
RNA
OH
1 ng
2 ng
0.1 U
0.2 U
A70
C60
G50
C40
G30
G20
C15
ACA
3'-PseudoU Pocket
3' K-turn
Apical Loop
5' K-turn
5'-PseudoU Pocket
1
2
3
4
5
6
B)
Apical Loop
Kink turn Motif
Upper Stem
PseudoU Pocket
Lower Stem
30
40
50
20
10
60
70
5' UUG UACACUACU 3'
Strong cleavage
Weak cleavage
Not analyzed

Figure 3.6: Single-stranded nuclease footprinting of L7Ae-Pf9. (A) 5'-end labeled Pf9 was incubated alone (lanes 3, 5) or with 1 μM L7Ae (lanes 4, 6) and digested with RNase T1 (lanes 3, 4) or RNase A (lanes 5, 6). The cleavage products were separated on a denaturing 15% acrylamide gel. Lane 1 is undigested RNA and lane 2 is a size marker generated by alkaline hydrolysis (OH). Red arrowheads indicate cleavages in the upper stem of the guide RNA in the absence of protein. Green and yellow bars indicate strong L7Ae protections and cleavage enhancements, respectively. (B) Summary of L7Ae cleavage protections and enhancements in the context of the predicted secondary structure of Pf9 (as in Figure 3.1).

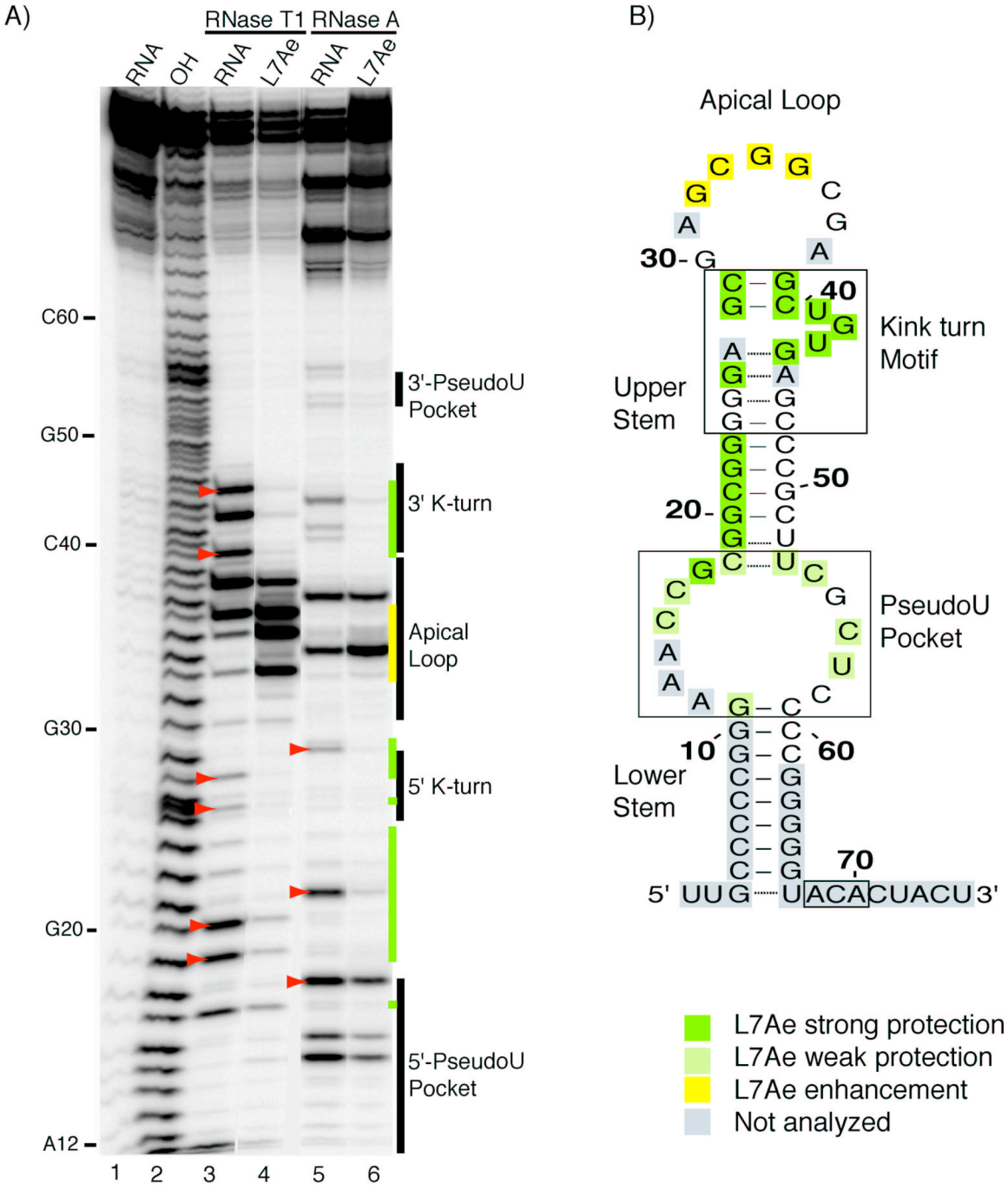
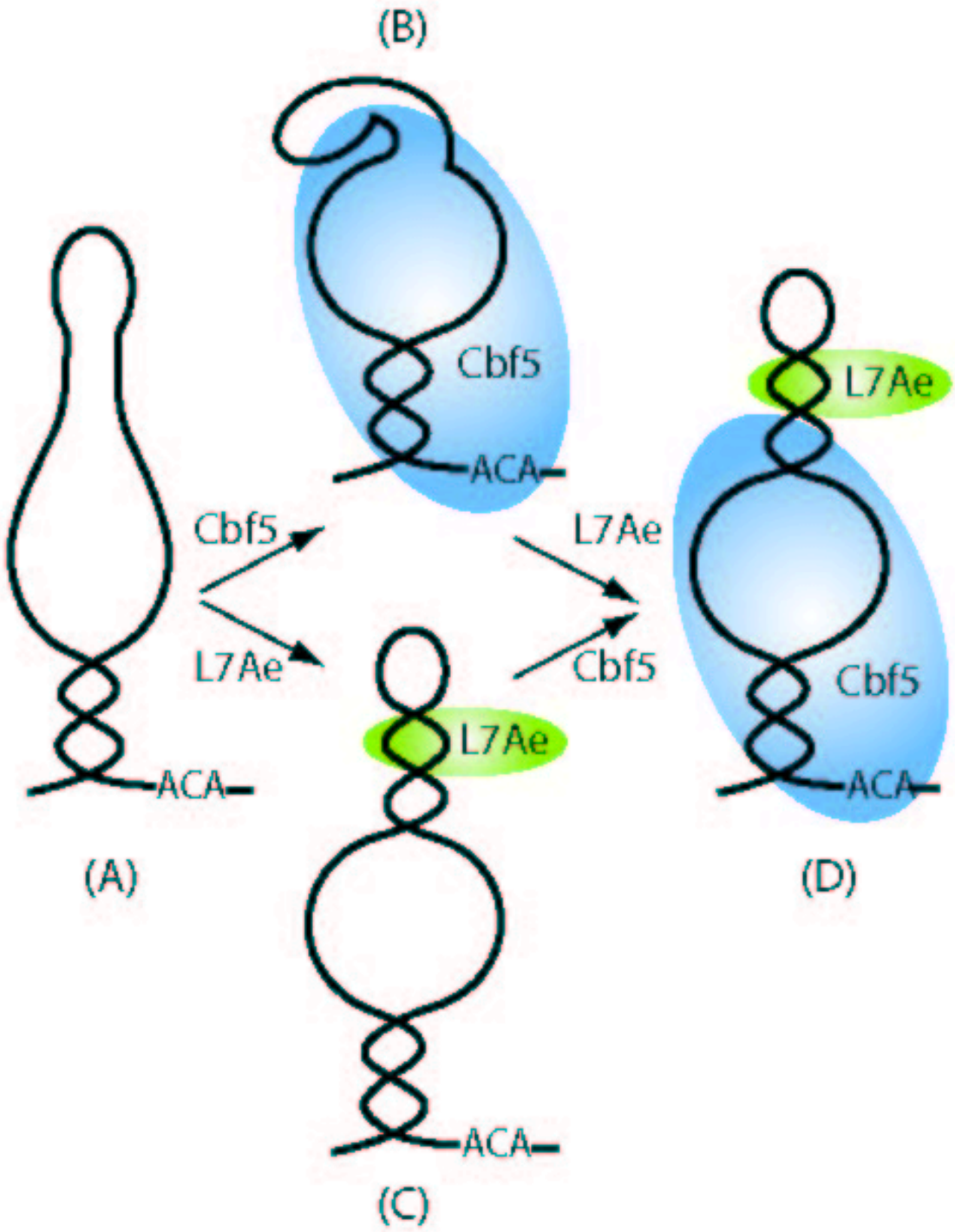

A)
RNase T1
RNase A
RNA
OH
RNA
L7Ae
RNA
L7Ae
C60
G50
C40
G30
G20
A12
1
2
3
4
5
6
3'-PseudoU Pocket
3' K-turn
Apical Loop
5' K-turn
5'-PseudoU Pocket
B)
Apical Loop
Kink turn Motif
Upper Stem
Lower Stem
PseudoU Pocket
30
40
50
20
10
60
70
5' UUG U ACA CUACU 3'
L7Ae strong protection
L7Ae weak protection
L7Ae enhancement
Not analyzed

Figure 3.7: Simplified model depicting the observed conformational changes in the guide RNA and RNA-protein interactions in sub-complexes of the H/ACA RNP. Our footprinting data suggest that the upper stem of the guide RNA is not stably formed in the absence of L7Ae (A and B). Binding of L7Ae at the k-turn induces formation of the upper stem and establishes the pseudouridylation pocket (C). Cbf5 interacts with box ACA, the lower stem, the pseudouridylation pocket, and a unique site that includes the apical loop and the 5' strand of the k-turn in the absence of L7Ae (B). L7Ae successfully competes with Cbf5 for binding to the 5' strand of the k-turn and disrupts Cbf5's interaction with this region of the guide RNA (D).


(B)
Cbf5
ACA
Cbf5
L7Ae
L7Ae
Cbf5
L7Ae
ACA
ACA
(A)
(C)
(D)

CHAPTER 4

A NOVEL ARCHAEAL SM-LIKE PROTEIN FORMS HOMOOCTAMERIC RING-LIKE STRUCTURES AND ASSOCIATES WITH H/ACA AND C/D MODIFICATION GUIDE RNAS[1]

[1] Osama A. Youssef, Lei Huang, Zhi-Jie Liu, Wolfram Tempel, John P. Rose, Bi-Cheng Wang, Rebecca M. Terns and Michael P. Terns, (2008), To be submitted to EMBO J.

Abstract

Sm and Lsm (Like-Sm) proteins comprise a large family of evolutionarily conserved RNA binding proteins. In eukaryotes, bacteria, and archaea, members of this protein superfamily assemble into ring-like complexes that interact with distinct RNA species to influence a wide range of RNA-mediated cellular pathways. Here we report the crystal structure of a novel archaeal Lsm protein (Sm4) from *Pyrococcus furiosus* at 2.8 Å. The Sm4 protein forms homomeric rings of eight molecules and thereby differs from most other Lsm proteins that form six or seven member ring complexes. Coimmunoprecipitation and Northern analysis revealed that Sm4 specifically associates primarily with ribose methylation (C/D) and pseudouridylation (H/ACA) guide RNAs *in vivo* but also with two uncharacterized non-coding RNAs. *In vitro* binding analysis with recombinant Sm4 protein and *in vitro* transcribed RNAs demonstrated that Sm4 binds directly and independently to H/ACA and C/D guide RNAs. Our data suggest a potential role of Sm4 in the biogenesis and/or function of H/ACA and C/D RNPs in archaea.

Introduction

The Sm and Lsm (Like-Sm) proteins comprise an evolutionarily ancient family of RNA binding proteins. Members of this protein superfamily interact with distinct partner and target RNAs to effect a variety of RNA-mediated functions in organisms from all three domains of life [1,2]. The initially-discovered members are the Sm proteins that bind small nuclear RNAs and are required for pre-mRNA splicing in eukaryotes [3,4]. Subsequently, at least 18 distinct eukaryotic Lsm proteins have been identified and found to be involved in the metabolism of a variety of cellular RNAs and to impact diverse cellular processes including pre-mRNA splicing

[5,6], telomere synthesis [7], histone pre-mRNA processing [8-10], mRNA degradation [11,12], and rRNA and tRNA maturation [13-17]. An Sm-like protein, called HfQ (*E. coli* host factor required for Q Beta RNA bacteriophage replication), is present in numerous eubacterial species and functions to regulate gene expression by modulating the stability or translatability of mRNAs [18]. HfQ employs small non-coding (nc)RNA partners and appears to facilitate the base-pairing of these ncRNAs with the target mRNAs [19-23]. Finally, three distinct Sm-like proteins (called Sm1, Sm2, and Sm3) have been identified in various archaeal organisms [2,24,25] and at least one archaeal species appears to have acquired a bacterial HfQ gene through lateral gene transfer [26]. While the precise functions of the identified archaeal Lsm proteins are not known, the association of *Archaeoglobus fulgidus* Sm1 and Sm2 with RNase P RNA (via co-immunoprecipitation) suggests a role for these proteins in tRNA maturation [27].

The Sm/Lsm/HfQ proteins each contain a conserved domain called the Sm-fold that is composed of a closed barrel containing an α helix followed by 5 anti-parallel β strands [18,28]. The individual subunits combine to form hexameric (six) or heptameric (seven), but occasionally other-sized ring structures [28-30]. In prokaryotes (bacteria and archaea), the known complexes are homomeric. However, in eukaryotes, the complexes are heteromeric and various combinations of proteins can form distinct complexes. Interestingly, the composition of the complex seems to determine the partner RNA and the function of the complex. For example, the 18 known eukaryotic Lsm proteins appear to combine into at least 6 different heterohepatmeric rings [28,31]. Complexes formed from Lsm proteins 2-8 bind nuclear RNAs including U6 snRNA, pre-tRNAs, and pre-rRNAs and are required for splicing and translation [32-34]. On the other hand, Lsm1-7 complexes function in cytoplasmic mRNA degradation [5,12]. Some of the characterized ring-like protein structures form only upon binding to an RNA, while others form

in the absence of RNA [35,36]. Moreover, some protein ring structures transiently interact with target RNAs in the cell while others remain stably associated [36]. It is thought that all members of the larger Sm/Lsm/HfQ protein family function by modulating RNA/RNA or RNA/protein interactions [19,20,28,36].

Crystal structures of archaeal Lsm ring complexes, both with and without RNA, and of sub-complexes of eukaryotic Sm RNPs have provided atomic-level resolution of the interaction of the proteins with one another as well as with target RNA molecules [24,35,37-42]. The ring-like architecture is intimately involved in RNA binding; each individual subunit of the ring interacts with a single nucleotide of single-stranded RNA target sequence [27,39,42]. In different complexes, the RNA interaction can take place either in the central cavity of the ring (i.e., the hole of the doughnut) or at the outer surface of the ring [39,42].

Here, we report the 2.8 Å crystal structure of a previously unrecognized Lsm complex (Sm4) from the hyperthermophilic archaeon *Pyrococcus furiosus*. The complex exhibits the hallmark ring-like form of Sm/Lsm/HfQ protein complexes and assembles into a stable homo-octameric complex in the absence of the RNA. We demonstrate that Sm4 is an RNA binding protein that selectively interacts with a specific subset of small non-coding RNAs *in vivo* and *in vitro*. These RNAs are primarily H/ACA and C/D guide RNAs that function in pseudouridylation and 2'-*O*-methylation, respectively, of rRNAs and tRNAs in archaea. In addition, two uncharacterized non-coding RNAs (Pf8 and Pf10) specifically associate with Sm4 *in vivo*. Our results indicate that Sm4 is a novel Sm-like protein important for the biogenesis and/or function of modification guide RNPs and perhaps ribosome biogenesis or activity.

Results

Overview of Sm4 structure

The crystal structure of *P. furiosus* Sm4 (Figure 1) was determined by the single wavelength anomalous diffraction method with selenomethionine-containing crystals. The refined Sm4 model (PDB entry 1YCY) consists of four identical polypeptide chains in the asymmetric unit, with each chain being composed of residues 5-24 and 29-70. Residues 1-4, 25-28 and 71 of each chain were not observed in the electron density maps and are therefore presumed to be disordered. Refinement of the structure against the 2.8 Å resolution data converged to give an R value of 0.281 (R_{free} = 0.311) with good stereochemistry (Table 1).

Sm4 Monomer: The Sm4 monomer (Figure 1A) can be described as a wedge shaped structure with approximate dimensions of 24 X 25 X 38 Å (solvent accessible surface of 4045 $Å^2$) [43] that is formed by an N-terminal helix (αA, residues 5-15) and a five-stranded twisted anti-parallel beta sheet (β1 - residues 17-24, β2 - residues 29-37, β3 - residues 39-48, β4 - residues 51-61, and β5 - residues 63-70). The Sm4 structure strongly resembles that of previously reported Lsm protein structures [24,27,29,35,39-42,44-46], which are characterized by two structural motifs: Sm1 (β strands 1, 2 and 3) and Sm2 (β strands 4 and 5). A pair-wise search of the Protein Data Bank using DALI (distance-matrix alignment) [47] and SSM (secondary structure matching) [48] for structures similar to Sm4 is shown in Table 2. Although the sequence similarity for the identified structural homologs presented is very low, the structures of Sm4 and other Sm/Lsm/HfQ proteins can be superimposed with root mean square deviations (RMSD) ranging from 1 to 2.1 Å. In addition, a PISA search (www.ebi.ac.uk/msd-srv/prot_int/pistart.html) [43,48] suggested that the four Sm4 monomers described in the crystallographic asymmetric unit could associate through crystallographic symmetry operations

to form both octamers and hexadecamers as described below.

Sm4 Octamer: The crystallographic asymmetric unit contains four identical polypeptide chains (chains A, B, C, and D in PDB entry 1YCY) that are linked via main-chain hydrogen bonds between residues Lys55, Gln56, and Leu58 of strand β4 of one monomer and residues Ile67' and Leu69' of strand β5' in an adjacent monomer (' indicates an adjacent molecule). There is also a hydrogen bond observed between Trp14 (NE1) and Gln56' (OE1) (see Table 3).

The Sm4 octamer is generated by a crystallographic 2-fold symmetry operation applied to chains A-D in the asymmetric unit, generating a roughly 34 Å thick doughnut shaped assembly with a diameter of 74 Å, and a large (~25 Å) central cavity similar to that observed for other Lsm protein structures (Figure 1B). The main-chain hydrogen bonds described above between adjacent beta strands (β4 and β5'), link the two tetramers to form the octamer. The Sm4 octamer differs from most other Lsm protein structures reported to date, which associate as either hexamers or heptamers (see Table 2). The Sm4 octamer is similar however, to a structure reported recently for the Lsm3 octamer from *S. cerevisiae* [29]. Although the yeast protein has no significant sequence identity with the *P. furiosus* protein. the two monomeric structures can be superimosed with an RMSD deviation of 1.0 Å (Table 2).

Sm4 Hexadecamer: The Sm4 hexadecamer is generated by again applying a crystallographic 2-fold symmetry operation to the Sm4 octamer. This operation generates a double doughnut-shaped dimer approximately 74 Å in diameter and 60 Å thick consisting of two stacked octameric rings (Figure 1C). The octameric dimer is formed by the interdigitation of the long loop connecting strands β3 and β4, placing the potential RNA binding face of each octamer on the outside of the structure exposed to solvent [39]. The assembly is stabilized in part by

weak hydrogen bonds between the side chain of Arg53 and the mainchain oxygen of Asn52* of chains B-B*, C-C* and D-A* (* indicates an adjacent symmetry related chain). Two salt bridges are also observed between the side chains of Arg53 (NH2) and Asp48* (OD1 and OD2) between chains D-A* and C-C*.

As noted above, the Sm4 hexadecamer is formed by the stacking of two Sm4 octamers such that the potential RNA binding surface of each octamer is exposed to solvent. This arrangement differs significantly from the stacked assemblies observed in other Lsm structures such as *P. abyssi* Sm1 and yeast Lsm3 (PDB entries 1M8V and 3BW1, respectively), which associate with the putative RNA binding surfaces facing each other in the center of the assembly [29,39]. Based on the structural homology, we conclude that Sm4 belongs to the evolutionarily conserved Sm/Lsm/HfQ family of proteins.

Association of Sm4 protein with C/D and H/ACA guide RNAs

Given that Sm4 is a novel member of the Sm/Lsm/HfQ structural family of RNA binding proteins, we sought to understand if the protein interacts with specific cellular RNAs. Toward this goal, we raised polyclonal antibodies against recombinant *P. furiosus* Sm4 protein that specifically recognize the protein in *P. furiosus* total cell extract (Figure 2). The Sm4 protein is specifically precipitated by immune (I) but not pre-immune (PI) Sm4 antibodies (Figure 2A). Anti-Sm4 antibodies also specifically recognize recombinant Sm4 protein (data not shown). A second preparation of Sm4 antibodies, generated in a distinct rabbit, also selectively recognize Sm4 protein in *P. furiosus* cell extract (data not shown).

To test for the presence of RNAs associated with Sm4 *in vivo*, co-immunoprecipitation experiments were performed using either Sm4 immune or pre-immune antibodies. We identified

six distinct RNA bands that were selectively immunoprecipitated by the anti-Sm4 antibodies as evidenced by their enrichment in the immune samples relative to either the preimmune or total RNA samples (Figures 2B and Figure 3A). Similar RNA profiles were observed after immunoprecipitation using the second preparation of Sm4 antibodies (data not shown).

The identity of RNAs coimmunoprecipitated by anti-Sm4 antibodies was determined by cloning the RNAs from each band and sequencing of the corresponding cDNAs (see Materials and Methods) (Figure 3B). Most of the clones are H/ACA and C/D RNAs that guide 2'-*O*-methylation of the ribose ring of the target nucleotide and conversion of uridines to pseudouridines in archaeal cellular RNAs, respectively [49]. Of the seven H/ACA RNAs so far identified in *P. furiosus* [50-52], three (Pf6, Pf7, and Pf9) were detected in the immunopurified samples. The most abundant clones encoded Pf6 and Pf7 H/ACA RNAs (more than 20 clones for each RNA). A single clone of the Pf9 H/ACA RNA was observed. In *P. furiosus,* there are more than 50 known C/D RNAs [53]. Of these, 15 C/D RNAs were detected at least one time in the clones (Figure 3B). Curiously, for two C/D RNAs (Figure 3B, band # 3) we obtained clones that are longer than the predicted full-length species [53] (sR48 at the 5' end, and sR55 at both the 5' and 3' ends, Table 4) and may represent precursor RNAs or biogenic intermediates.

In addition to the H/ACA and C/D RNAs, we found a relatively high number of clones derived from uncharacterized non-coding RNAs. The first called Pf8 was identified by Klein and coworkers as a *P. furiosus* non-coding RNA of unknown function [54]. The other RNA, which we designate Pf10 appears to be a previously undescribed RNA without significant protein coding potential. Sequences with the potential to encode highly conserved homologs are apparent in the *P. abyssi* and *P. horikoshii* genomes. Finally, we cloned several small fragments

derived from *P. furiosus* 16S, 23S, and 5S rRNAs, a single clone of RNAse P RNA, and a few mRNA species (Figure 3B).

To determine the extent of association of the RNAs identified by cloning, we performed Northern blotting analysis of RNAs immunoprecipitated from total cell extracts using either Sm4 immune (I) or pre-immune (PI) antibodies (Figure 4). Previous coimmunoprecipitation studies with *A. fulgidus* Sm1 and Sm2 found that these proteins specifically associated with RNase P RNA [27]. However, while we isolated one clone of RNAse P RNA (figure 3B) Northern blot analysis failed to detect significant amount of RNase P RNA in the immunoprecipited material (Figure 4E) indicating that Sm4 does not associate with RNase P RNA. We also probed for all 3 detected H/ACA RNAs (Pf6, Pf7 and Pf9) and five representative C/D RNAs (sR2, sR12, sR29, sR46, and sR55). All of the tested H/ACA and C/D RNAs were selectively immunoprecipitated with the anti-Sm4 antibodies (Figure 4A and B). The two non-coding RNAs (Pf8 and Pf10) were also specifically coimmunoprecipiated with Sm4 protein (Figure 4C). Interestingly, the novel Pf10 RNA was the most efficiently coimmunoprecipitated RNA that was tested (Figure 4C).

At the same time, the rRNAs (5S, 16S, and 23S) that we detected by cloning (Figure 3B) were not present in significant amount in the immunoprecipitated fractions analyzed by Northern blotting (Figure 4D). Finally, tRNA-Asp which was not predicted to interact with Sm4 and not detected by cloning was also not immunoprecipitated by the Sm4 antibodies. Our results clearly indicate that Sm4 associates specifically with H/ACA and C/D guide RNAs as well as two uncharacterized non-coding RNAs (Pf8 and Pf10).

The association of the C/D and H/ACA RNAs with Sm4 could be direct or indirect (mediated by other molecules such as a protein or RNA). To address this point, we performed

gel mobility shift assays with *in vitro* transcribed RNAs and recombinant Sm4 protein (Figure 5). Specifically, we tested the ability of Pf6, Pf7, and Pf9 H/ACA RNAs and sR2 and sR29 C/D RNAs to directly interact with Sm4 *in vitro*. We included L7Ae, a protein that binds directly to a common motif called the kink(k)-turn in both, as a positive control for a direct RNA/protein interaction and as expected L7Ae interacted with all 5 of these RNAs [55]. The gel shift results also indicate that Sm4 binds directly to Pf6, Pf7, and Pf9 H/ACA RNAs (Figure 5A) and sR2 and sR29 C/D RNAs (Figure 5B). The apparent dissociation constant for the Sm4/RNA interactions is relatively high (Kd $\geq$ 10 μM). However, we also tested 5S rRNA and tRNA-Asp and consistent with the results of Northern analysis (Figure 4E), we did not observe an interaction between Sm4 (or L7Ae) and 5S rRNA or tRNA-Asp (Figure 5C). In summary, the results indicate that Sm4 protein interacts directly and independently with H/ACA and C/D RNAs.

Archaeal H/ACA and C/D RNAs contain an obvious common sequence/structure motif – the k-turn that is bound by the L7Ae protein [52]. We naturally tested whether Sm4 also recognizes this common motif. However, our results indicate that the interaction of Sm4 with the RNAs is not mediated by the k-turn. We investigated the interaction in gel mobility shift assays. As observed in Figure 6A, L7Ae binds and shifts the mobility of H/ACA RNA Pf9, and Sm4 binds Pf9 and reduces the mobility to an even greater extent (consistent with the possibility that Sm4 interacts with the RNA as a multimer) (Figure 6A). However, while mutation of the k-turn of Pf9 H/ACA RNA abolished the interaction of L7Ae, it did not affect the binding of Sm4 (Figure 6A) indicating that Sm4 does not require the k-turn for interaction with Pf9. In addition, interaction of L7Ae with the double hairpin/double k-turn Pf6 H/ACA RNA produces a shift approaching the mobility of the Sm4 RNP (Figure 6B). However, in the presence of both

proteins, a supershift is observed indicating that both proteins can interact with the RNA simultaneously.

Discussion

Here, we report a fourth archaeal Sm-Like protein, Sm4, and the RNAs that it interacts with *in vivo*. Sm4 bears no obvious sequence homology with other known Sm/Lsm proteins. However, the crystal structure presented here revealed that Sm4 shares the Sm fold common to all identified Sm/Lsm/HfQ proteins (Figure 1). Although most well-characterized Sm/Lsm/HfQ complexes contain six or seven subunits, Sm4 complex contains eight subunits. The only other known octameric Lsm protein is the yeast Lsm3 [29]. Structural comparison between archaeal Sm4 and yeast Lsm3 showed that they have a very similar structural organization (1 Å RMSD). All the well-studied Sm/Lsm/HfQ proteins have been shown or are predicted to be RNA-binding proteins involved in various cellular RNA pathways. In the present study, we found that Sm4 is associated with C/D and H/ACA guide RNAs as well as a couple of ncRNAs of undefined function (but not other cellular RNAs) *in vivo* (Figure 4).

Mode of RNA Binding: Crystal structures of other Sm/Lsm/HfQ complexes with short RNA sequences and available electron microscopy data indicate that RNA binds to the center of the ring-like structure (inner RNA binding site) [35,39,45,58], with some complexes also containing a second RNA binding site located at the periphery of the ring (outer RNA binding site) [39,42]. A cationic central pore that includes conserved basic residues located in loops 3 (between β2 and β3) and 5 (between β4 and β5) forms the inner RNA binding sites (Figure 7B and [27,39,42,45]). However, the electrostatic potential of this region of the Sm4 is negatively charged (Figure 7A) making it unlikely that the RNAs interact with the central pore of the Sm4

complex. In contrast, Sm4 does share structural features that constitute the outer RNA binding site in other Sm/Lsm/HfQ proteins indicating that Sm4 may utilize this mode of RNA binding.

Superimposition [48] of the RNA (seven uridines)-bound *P. abyssi* Sm1 structure with Sm4 [39] revealed striking similarities that suggest that the Sm4 octamer may employ the outer RNA binding site associated with the αA helix from each of the monomers (Figure 7). Sm4 chain C and Sm1 chain D align and give an RMSD of 0.96 Å (Figure 7C). Most Sm1 main-chain RNA interactions could be accommodated in Sm4 by a small reorientation of the αA helix. To explore this possibility a model of the Sm4 structure was manually generated [59] with the Sm4 αA helix positioned in a manner similar to that observed in the Sm1 structure. This was done by the superposition [48] of the Sm4 αA helix onto the Sm1 αA helix and then replacing the reoriented αA residues in the Sm4′ coordinate set (′ denotes the computer generated model). A comparison of RNA binding interactions for the modeled Sm4′ structure with those observed in the Sm1 RNA-bound structure is shown in figure 7D. All RNA/main-chain hydrogen bonds are observed in the Sm4′ model. In addition, in the Sm4′ model, Tyr34 from the Sm1 structure is replaced by Phe36′ and preserves both ring stacking and main-chain hydrogen bonding interactions. It should be noted that the position of Phe36′ at the C-terminus of strand β2 was not included in the modeling process. The alignment also places a serine at the Sm1 Arg4 position, thus the Arg4 (NH2 – U4 OP2) hydrogen bond is lost. In addition, the model places Glu8′ at the Sm1 Asp7 position, which in theory (the Glu8′ side chain atoms are not present in the Sm4 structure) could maintain the hydrogen bonding interactions to RNA at this position. Finally a new hydrogen bond is made between the Lys15′ (NZ) and U4 (O4). Based on these observations it is conceivable that the RNA could bind to Sm4 in a manner similar to that proposed previously for the outer RNA binding site of Sm1 [39].

The available crystal structures of RNA-bound/Sm-like proteins have utilized short model RNAs (typically short poly U) and not physiological RNA partners. Given that we have found that Sm4 interacts both *in vivo* and *in vitro* to members of both the H/ACA and C/D family of RNA modification guide RNAs (Figures 3-5), an important question is which sequences or structures within these diverse RNAs are recognized by the Sm4 protein. We have ruled out the simple possibility that Sm4 interacts with the k-turn motif that is common to these two classes of RNAs (Figure 6). Ultimately, high resolution 3D structures of Sm4 bound with the different guide RNAs will be needed to illuminate the detailed RNA and protein binding sites in Sm4/ RNA complexes.

Biological Roles of Sm4:

In general the Sm/LSm/HfQ proteins are thought to function as RNP chaperones [11,20], Based on the RNA interactions identified here, we hypothesize that Sm4 modulates either RNA/protein or RNA/RNA interactions during one or more steps in the biogenesis or activity of the H/ACA and C/D guide RNPs. For example, a role in the biogenesis of the guide RNPs can be envisioned whereby Sm4 protein transiently interacts with newly synthesized guide RNAs to modulate RNA structure and to facilitate interaction of the guide RNAs with the core C/D or H/ACA RNP proteins. Our observation that potential precursor C/D RNAs (sR48 and sR55 having 5' and/or 3' ends longer than predicted full-length mature RNAs) (Figures 3B and Table 4) is consistent with such a role. While we and others have shown that the pseudouridylation and ribose methylation activities of the guide RNPs do not require proteins in addition to the core proteins *in vitro* [50,61,62], this does not exclude an important function for proteins like Sm4 in guide RNP assembly or function *in vivo*. For example, snRNPs form spontaneously *in vitro*

from purified RNP components [63], but the *in vivo* assembly depends on the activity of SMN [64,65]. Assembly of eukaryotic H/ACA RNPs *in vivo* requires the Naf1 protein [66]

It is also possible that Sm4 plays a key role in promoting the activity of the guide RNPs *in vivo*. The function of the H/ACA and C/D RNPs involves interaction between the guide RNA and substrate RNA. Moreover, the resultant guide RNA/substrate RNA duplexes must be disrupted following the modification. A major unanswered question in the field has been whether there are trans-acting factors that promote dissociation of the guide RNA/target RNA duplexes. A specific role for the Sm4 protein in this process is conceivable. Finally, while sequence homologs of Sm4 appear to be restricted to *Pyrococcus* species, a conserved functional relationship between Sm-like proteins and the modification guide RNPs is evidenced by the association of eukaryotic Lsm proteins associate with H/ACA and C/D RNAs [15-17].

Taken together, we have identified a novel Sm-like protein in Pyrococcus on the basis of structural homology in the absence of sequence conservation. Our findings suggest that the already large family of Sm/Lsm/HfQ proteins will be expanded as more structural determinations of proteins are made. Importantly, our work suggests a new function for archaeal Sm-like proteins in the biogenesis and/or function of modification guide RNPs and the generation of ribosomes. Determination of the detailed mode of RNA binding and precise biological roles of the newly discovered Sm4 protein awaits further investigation.

Materials and Methods

Protein expression and purification

Sm4 (PF1955) and L7Ae (PF1367) genes were amplified by PCR using *P. furiosus* genomic DNA and subcloned into pET28 and pET21d expression vectors, respectively. The

gene insertion was confirmed by sequencing. The plasmid with the correct insertion was then transformed into *E. coli* BL21 codon+ (L7Ae) or BL21 PRIL (Sm4) competent cells (Stratagene). Protein expression was carried out as described in detail previously [50]. 125 mg/L seleno-methionine (SeMet) was used for the production of Se-Met-labeled Sm4 protein. The resultant recombinant protein carries a hexa-histidine tag at the N-terminus. Recombinant proteins were affinity-purified on a Ni-agarose column (Qiagen). The histidine tag at the N-terminus of Sm4 was removed by human thrombin digestion (Calbiochem) prior to crystallization.

Crystallization and X- ray data collection

Initial crystallization conditions for the Se-Met-labeled Sm4 protein were determined by screening against a set of eight commercial crystallization screens (384 conditions) using 200 nL sitting drops setup using a Cartesian Honeybee as described previously [69]. Each drop contained equal volumes of protein (~10 mg/mL) and precipitant solution. Crystal optimization was carried out by the micro-batch-under-oil method using 2 μL drops, setup using a Douglas Instruments ORYX6 [70]. The 72-condition optimization screens were centered at conditions that produced crystals from the initial screen. All setups were incubated and imaged using a CrystalFarm system maintained at 18°C.

For data collection, crystals were harvested using a rayon loop of appropriate size and flash-frozen in liquid nitrogen [71]. Cryoprotection was achieved using a 30% (v/v) glycerol-mother liquor mixture. A 2.67 Å resolution data set was collected at cryogenic temperatures using 0.9785 Å X-rays (to enhance the selenium anomalous scattering signal) on beamline 22ID (SER-CAT), Advanced Photon Source Argonne National Laboratory, as described in Table 1.

All data were indexed, integrated, and scaled using the HKL 1.9.1 software suite [72].

Structure determination and refinement

The selenium substructure and initial protein phases were determined using the SECSG SCA2Structure automated structure determination pipeline [73]. The initial model obtained from SCA2Structure was completed and adjusted manually, where necessary, using XFIT [74]. Structure refinement was carried out using REFMAC5 [75] and validated using MOLPROBITY [76] and PROCHECK [77]. Both coordinates and structure factors have been deposited in the Protein Data Bank (PDB entry 1YCY).

Crystallization: The molecular weight of the recombinant Sm4 protein was estimated to be 7,970 Dalton by ESI-MS analysis, which is in good agreement with the calculated value for the protein (accounting for the N-terminal His_6 affinity tag). The best diffracting crystals were hexagonal bipyramids obtained using a precipitant solution containing 0.1M sodium acetate pH 4.6, 0.1M sodium chloride and 12% W/V PEG 6000. Crystals generally appeared in 3-5 days and grew to usable size (0.15 x 0.15 x 0.35 mm) in two weeks. These crystals diffracted to 2.6 Å using synchrotron X-rays.

Data Collection: Analysis of the systematic absences in the processed data indicated that the crystals belong to either space group $P6_222$ or $P6_422$ with cell dimensions of a = 82.96 and c = 189.78 Å. The space group is consistent with the crystal's hexagonal bipyramidal habit. Based on the unit cell volume and space group symmetry the crystallographic asymmetric unit can accommodate four protein chains with a calculated Matthew's coefficient of 2.89 $Å^3$/Da and corresponding solvent content of 57.37% [78].

Structure solution: The structure was solved using the SCA2Structure pipeline [73] using single wavelength anomalous scattering signal from eight of the possible 16 selenium atoms found in the asymmetric unit (assuming the asymmetric unit contained four molecules as predicted). The parameter space screening approach used in the pipeline allowed for automated solution attempts in both possible ($P6_222$ and $P6_422$) space groups. Convincing solutions (high SOLVE Z score and significant number of residues fit from RESOLVE), however, were only obtained for data processed in space group $P6_222$.

Sm4 antibody production

Approximately 500 μg of purified recombinant Sm4 protein was injected into a rabbit for both primary immunization and a booster to produce polyclonal anti-Sm4 antibodies. Pre-immune sample was collected a week prior to the protein injection. Immune serum was collected once every two weeks after the injection.

Immunoprecipitation

Protein A Sepharose (Sigma) was swollen in nuclease-free water and washed two times with IPP-500 buffer (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1% Igepal). A 100 μl aliquot of Sm4 immune or pre-immune antibodies was added to the beads and incubated at room temperature for 2 hours. The antibody-coupled beads were washed four times with IPP-300 buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.05% Igepal). A 1 mL aliquot of *P. furiosus* cell extract (10 mg/mL total protein) was added and incubated at 4°C overnight. Beads were washed four times with IPP-300 buffer. *P. furiosus* cell extract was prepared as described previously [79].

RNAs and proteins were extracted from *P. furiosus* cell extract, supernatant and immunoprecipiated fractions using TRIazol LS (Invirtogen) according to the manufacturer's recommended protocol. RNA pellets were resuspended in 8 M urea loading dye. Protein pellets were resuspended in 3x SDS loading buffer.

Immunoblotting analysis

Proteins immunoprecipitated with either immune or preimmune anti-Sm4 antibodies were separated on 15% SDS-PAGE and transferred to nitrocellulose membrane (Bio-Rad) according to the manufacturer's protocol. The membrane was blocked with 5% milk in TBS-T buffer (100 mM Tris-HCl pH 7.5, 0.9% NaCl, 0.2% Tween-20) for 1 hour at room temperature. The membrane was then incubated with Sm4 immune or pre-immune antibodies (1:100 dilution) in TBST buffer for 1 hour at room temperature and washed 3 times with TBS-T. The membrane was incubated with HRP secondary antibodies (HorseRadish Peroxidase-coupled IgG antibodies, Amersham, 1:2500 dilution) in TBS buffer for 1 hour at room temperature and washed 3 times with TBS-T. The protein was detected using an ECL-Plus detection system (Amersham Pharmacia Biotech) and phosphorimager.

Identification of Sm4-associated RNAs

RNAs coimmunoprecipitated with Sm4 antibodies were extracted from pellets and fractionated on 8% polyacrylamide/7 M urea gel and visualized by SYBR-Green staining (Invitrogen). Cloning was performed as described previously [53]. Briefly, a primer carrying monophosphate at the 5' end and blocked with dideoxycytidine at the 3' end (5'-pCTCGAGATCTGGATCCGGGddC-3') was ligated to the gel-purified RNA for 1 hour at 37°C

with T4 RNA ligase (Promega). The ligated RNA was reverse transcribed with Superscript Reverse Transcriptase II (Invitrogen) at 42°C for 50 minutes with primer (5'-CCCGGATCCAGATCTCGAG-3'). The RNA template was hydrolyzed with ribonuclease H (Invitrogen) at 37°C for 20 minutes. Poly dA tails were added to the cDNAs with deoxy-ATP and terminal deoxynucleotidyl transferase (Roche). The poly dA-tailed cDNAs were amplified by PCR using (5'-CCCGGATCCAGATCTCGAG-3') and (5'-GCGAATTCTGCAG$(T)_{30}$-3') as primers. The DNA products were subcloned into TOPO-PCR II vector and transformed into TOP10 cells (Invitrogen). White colonies were selected, plasmids were isolated, and their sequences were determined (Laboratory for Genomics and Bioinformatics, University of Georgia, Athens, GA).

Northern blot analysis

RNAs extracted from immune and pre-immune samples, supernatant and *P. furiosus* total cell extract were fractionated into 8% polyacrylamide/7 M urea gel and transferred to Zetaprobe GT nylon membranes (Bio-Rad). The membrane was pre-hybridized at 42°C for 4 hours in the hybridization buffer [5x SSC (75 mM sodium citrate pH 7.0, 750 mM NaCl), 7% SDS, 20 mM sodium phosphate, pH 7.0, 1x Denhardt's solution (Invitrogen)]. DNA oligonucleotides complementary to the target RNAs were 5' end-labeled with ^{32}P-ATP (7000 Ci/mmol, MP Biomedicals) and T4 polynucleotide kinase (Ambion). Hybridization with ^{32}P-labeled probes was carried out overnight at 42°C in the same hybridization buffer. The membrane was washed two times with 2x SSC and 0.5% SDS at 42°C. The RNAs were visualized by phosphorimager.

Oligonucleotides for Northern blotting

All DNA oligonucleotides used were purchased from MWG-Biotech Inc.

sR2: GGCTCCTCATCACTAATCAGAGTGAGGG

sR12: CACAACTCAGACCGGTAAACGC

sR29: GACATCATCACCTTTCAGGCTGGGC

sR46: CGCCTTTGCTCAGCATTGGATTCAG

sR55: ATCATCGACCCCGTTTCAGCAGG

Pf6: AGCACACCCCGCTCATCGAACC

Pf7: TGTATGCATTCTCAGGCGGGCTAACC

Pf8: CATCGGGCACGGTCAGAGGC

Pf9: AGCGAAGCGGGCTCACAGCTCGC

Pf10: TTAACCCGCCCAAGTTCATCGATTCC

5S rRNA: CCCGGCTTCCCGCCCCCTCT

16S rRNA: CTCGACCTGACCTCCCGAAGG

23S rRNA: CCTTAGATGCTTTCAGGCCTTATCGGC

tRNA-Asp: CGGGCTACACCACCCGGGC

RNase P RNA: GCACCCCGCGGGACGGCCG

In vitro transcription

All the DNA templates were generated by PCR amplification from *P. furiosus* genomic DNA. ^{32}P-GTP-labeled RNAs were transcribed *in vitro* with T7 (sR2 guide and target, and sR29) or SP6 (Pf6, Pf7, Pf9, 5S rRNA, and tRNA-Asp) RNA polymerase (Promega) as previously described [50].

Gel mobility shift assay

Approximately 0.05 pmol of ^{32}P-RNA was incubated alone or with various amounts of the recombinant proteins in the binding buffer (20 mM HEPES pH 7.0, 500 mM KCl, 1.5 mM Mg_2Cl) for 30 min at 65^{o}C. For sR2, 0.5 pmol of sR2 target was added in figure 7B. The complexes were then analyzed by 6% or 8% non-denaturing polyacrylamide gel and visualized by autoradiography.

Acknowledgement

The work described here was funded in part by the following organizations: the National Institutes of Health R01-6M54682 (M.T. and R.T.), GM62407 (B.W.), the Georgia Research Alliance and the University of Georgia Research Foundation and a scholarship from the Egyptian government (O.Y.). Data were collected at 22ID, Southeast Regional Collaborative Access Team (SER-CAT, www.ser-cat.org), Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

References

1. Achsel T, Stark H, Luhrmann R: **The Sm domain is an ancient RNA-binding motif with oligo(U) specificity**. *Proc Natl Acad Sci U S A* 2001, **98**:3685-3689.

2. Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Seraphin B: **Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin**. *EMBO J* 1999, **18**:3451-3462.

3. Jarmolowski A, Mattaj IW: **The determinants for Sm protein binding to Xenopus U1 and U5 snRNAs are complex and non-identical**. *EMBO J* 1993, **12**:223-232.

4. Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Luhrmann R: **snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions**. *EMBO J* 1995, **14**:2076-2088.

5. He W, Parker R: **Functions of Lsm proteins in mRNA degradation and splicing**. *Curr Opin Cell Biol* 2000, **12**:346-350.

6. Staley JP, Guthrie C: **Mechanical devices of the spliceosome: motors, clocks, springs, and things**. *Cell* 1998, **92**:315-326.

7. Seto AG, Zaug AJ, Sobel SG, Wolin SL, Cech TR: **Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle**. *Nature* 1999, **401**:177-180.

8. Galli G, Hofstetter H, Stunnenberg HG, Birnstiel ML: **Biochemical complementation with RNA in the Xenopus oocyte: a small RNA is required for the generation of 3' histone mRNA termini**. *Cell* 1983, **34**:823-828.

9. Schaufele F, Gilmartin GM, Bannwarth W, Birnstiel ML: **Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA**. *Nature* 1986, **323**:777-781.

10. Strub K, Galli G, Busslinger M, Birnstiel ML: **The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing**. *EMBO J* 1984, **3**:2801-2807.

11. Bouveret E, Rigaut G, Shevchenko A, Wilm M, Seraphin B: **A Sm-like protein complex that participates in mRNA degradation**. *EMBO J* 2000, **19**:1661-1671.

12. Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R: **Yeast Sm-like proteins function in mRNA decapping and decay**. *Nature* 2000, **404**:515-518.

13. Kufel J, Allmang C, Verdone L, Beggs JD, Tollervey D: **Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p**. *Mol Cell Biol* 2002, **22**:5248-5256.

14. Kufel J, Allmang C, Petfalski E, Beggs J, Tollervey D: **Lsm Proteins are required for normal processing and stability of ribosomal RNAs**. *J Biol Chem* 2003, **278**:2147-2156.

15. Kufel J, Allmang C, Verdone L, Beggs J, Tollervey D: **A complex pathway for 3' processing of the yeast U3 snoRNA**. *Nucleic Acids Res* 2003, **31**:6788-6797.

16. Tomasevic N, Peculis BA: **Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence**. *Mol Cell Biol* 2002, **22**:4101-4112.

17. Fernandez CF, Pannone BK, Chen X, Fuchs G, Wolin SL: **An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5**. *Mol Biol Cell* 2004, **15**:2842-2852.

18. Wilusz CJ, Wilusz J: **Eukaryotic Lsm proteins: lessons from bacteria**. *Nat Struct Mol Biol* 2005, **12**:1031-1036.

19. Valentin-Hansen P, Eriksen M, Udesen C: **The bacterial Sm-like protein Hfq: a key player in RNA transactions**. *Mol Microbiol* 2004, **51**:1525-1533.

20. Geissmann TA, Touati D: **Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator**. *Embo J* 2004, **23**:396-405.

21. Moller T, Franch T, Hojrup P, Keene DR, Bachinger HP, Brennan RG, Valentin-Hansen P: **Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction**. *Mol Cell* 2002, **9**:23-30.

22. Gottesman S: **The small RNA regulators of Escherichia coli: roles and mechanisms***. *Annu Rev Microbiol* 2004, **58**:303-328.

23. Storz G, Opdyke JA, Zhang A: **Controlling mRNA stability and translation with small, noncoding RNAs**. *Curr Opin Microbiol* 2004, **7**:140-144.

24. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D: **Structure and assembly of an augmented Sm-like archaeal protein 14-mer**. *Proc Natl Acad Sci U S A* 2003, **100**:4539-4544.

25. Kilic T, Sanglier S, Van Dorsselaer A, Suck D: **Oligomerization behavior of the archaeal Sm2-type protein from Archaeoglobus fulgidus**. *Protein Sci* 2006, **15**:2310-2317.

26. Nielsen JS, Boggild A, Andersen CB, Nielsen G, Boysen A, Brodersen DE, Valentin-Hansen P: **An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii**. *RNA* 2007, **13**:2213-2223.

27. Toro I, Thore S, Mayer C, Basquin J, Seraphin B, Suck D: **RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex**. *EMBO J* 2001, **20**:2293-2303.

28. Khusial P, Plaag R, Zieve GW: **LSm proteins form heptameric rings that bind to RNA via repeating motifs**. *Trends Biochem Sci* 2005, **30**:522-528.

29. Naidoo N, Harrop SJ, Sobti M, Haynes PA, Szymczyna BR, Williamson JR, Curmi PM, Mabbutt BC: **Crystal structure of Lsm3 octamer from Saccharomyces cerevisiae:**

implications for Lsm ring organisation and recruitment. *J Mol Biol* 2008, **377**:1357-1371.

30. Brennan RG, Link TM: **Hfq structure, function and ligand binding**. *Curr Opin Microbiol* 2007, **10**:125-133.

31. Albrecht M, Lengauer T: **Novel Sm-like proteins with long C-terminal tails and associated methyltransferases**. *FEBS Lett* 2004, **569**:18-26.

32. Achsel T, Brahms H, Kastner B, Bachi A, Wilm M, Luhrmann R: **A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro**. *EMBO J* 1999, **18**:5789-5802.

33. Spiller MP, Boon KL, Reijns MA, Beggs JD: **The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA**. *Nucleic Acids Res* 2007, **35**:923-929.

34. Mayes AE, Verdone L, Legrain P, Beggs JD: **Characterization of Sm-like proteins in yeast and their association with U6 snRNA**. *EMBO J* 1999, **18**:4321-4331.

35. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Luhrmann R, Li J, Nagai K: **Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs**. *Cell* 1999, **96**:375-387.

36. Beggs JD: **Lsm proteins and RNA processing**. *Biochem Soc Trans* 2005, **33**:433-438.

37. Collins BM, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC: **Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs**. *J Mol Biol* 2001, **309**:915-923.

38. Kilic T, Thore S, Suck D: **Crystal structure of an archaeal Sm protein from Sulfolobus solfataricus**. *Proteins* 2005, **61**:689-693.

39. Thore S, Mayer C, Sauter C, Weeks S, Suck D: **Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya**. *J Biol Chem* 2003, **278**:1239-1247.

40. Toro I, Basquin J, Teo-Dreher H, Suck D: **Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus**. *J Mol Biol* 2002, **320**:129-142.

41. Mura C, Cascio D, Sawaya MR, Eisenberg DS: **The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core**. *Proc Natl Acad Sci U S A* 2001, **98**:5532-5537.

42. Mura C, Kozhukhovsky A, Gingery M, Phillips M, Eisenberg D: **The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs)**. *Protein Sci* 2003, **12**:832-847.

43. Krissinel E, Henrick K: **Inference of macromolecular assemblies from crystalline state**. *J Mol Biol* 2007, **372**:774-797.

44. Collins BM, Cubeddu L, Naidoo N, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC: **Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA. Crystal structure of an oligomeric complex of yeast SmF**. *J Biol Chem* 2003, **278**:17291-17298.

45. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG: **Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein**. *EMBO J* 2002, **21**:3546-3556.

46. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, et al.: **Structure of Pseudomonas aeruginosa Hfq protein**. *Acta Crystallogr D Biol Crystallogr* 2005, **61**:141-146.

47. Holm L, Sander C: **Protein structure comparison by alignment of distance matrices**. *J Mol Biol* 1993, **233**:123-138.

48. Krissinel E, Henrick K: **Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**:2256-2268.

49. Matera AG, Terns RM, Terns MP: **Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs**. *Nat Rev Mol Cell Biol* 2007, **8**:209-220.

50. Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP: **RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP**. *Genes Dev* 2005, **19**:1238-1248.

51. Muller S, Leclerc F, Behm-Ansmant I, Fourmann JB, Charpentier B, Branlant C: **Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs**. *Nucleic Acids Res* 2008.

52. Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A: **Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea**. *Nucleic Acids Res* 2003, **31**:869-877.

53. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea**. *Science* 2000, **288**:517-522.

54. Klein RJ, Misulovin Z, Eddy SR: **Noncoding RNA genes identified in AT-rich hyperthermophiles**. *Proc Natl Acad Sci U S A* 2002, **99**:7542-7547.

55. Dennis PP, Omer A: **Small non-coding RNAs in Archaea**. *Curr Opin Microbiol* 2005, **8**:685-694.

56. Youssef OA, Terns RM, Terns MP: **Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP**. *Nucleic Acids Res* 2007, **35**:6196-6206.

57. Terns MP, Terns RM: **Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin**. *Gene Expr* 2002, **10**:17-39.

58. Stark H, Dube P, Luhrmann R, Kastner B: **Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle**. *Nature* 2001, **409**:539-542.

59. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera--a visualization system for exploratory research and analysis**. *J Comput Chem* 2004, **25**:1605-1612.

60. Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G: **The structure and function of small nucleolar ribonucleoproteins**. *Nucleic Acids Res* 2007, **35**:1452-1464.

61. Yu YT, Terns RM, Terns MP: **Mechanisms and functions of RNA-guided RNA modification**. In *Fine-tuning of RNA functions by modification and editing*. Edited by Grosjean H: Topics in Current Genetics; 2005:223-262. vol 12.

62. Tran EJ, Zhang X, Maxwell ES: **Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs**. *EMBO J.* 2003, **22**:3930-3940.

63. Raker VA, Hartmuth K, Kastner B, Luhrmann R: **Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner**. *Mol Cell Biol* 1999, **19**:6554-6565.

64. Fischer U, Liu Q, Dreyfuss G: **The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis**. *Cell* 1997, **90**:1023-1029.

65. Pellizzoni L, Kataoka N, Charroux B, Dreyfuss G: **A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing**. *Cell* 1998, **95**:615-624.

66. Fatica A, Dlakic M, Tollervey D: **Naf1 p is a box H/ACA snoRNP assembly factor**. *RNA* 2002, **8**:1502-1514.

67. Wang C, Meier UT: **Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins**. *EMBO J* 2004, **23**:1857-1867.

68. Wang C, Query CC, Meier UT: **Immunopurified small nucleolar ribonucleoprotein particles pseudouridylate rRNA independently of their association with phosphorylated Nopp140**. *Mol Cell Biol* 2002, **22**:8457-8466.

69. Liu ZJ, Tempel W, Ng JD, Lin D, Shah AK, Chen L, Horanyi PS, Habel JE, Kataeva IA, Xu H, et al.: **The high-throughput protein-to-structure pipeline at SECSG**. *Acta Crystallogr D Biol Crystallogr* 2005, **61**:679-684.

70. Chayen NE, Shaw Sewart PD, Maeder DL, Blow DM: **An automated system for micro-batch protein crystallization and screening**. *J Appl Cryst* 1990, **23**:297-302.

71. Teng TY: **Mounting of crystals for macromolecular crystallography in a freestanding thin-film**. *Journal of Applied Crystallography* 1990, **23**:387-391.

72. Otwinowski Z, Minor W: **Processing of X-ray diffraction data collected in oscillation mode**. *Methods in Enzymology* 1997, **A276**:307-326.

73. Liu ZJ, Lin D, Tempel W, Praissman JL, Rose JP, Wang BC: **Parameter-space screening: a powerful tool for high-throughput crystal structure determination**. *Acta Crystallogr D Biol Crystallogr* 2005, **61**:520-527.

74. McRee DE: **XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density**. *J Struct Biol* 1999, **125**:156-165.

75. Murshudov GN, Vagin AA, Dodson EJ: **Refinement of macromolecular structures by the maximum-likelihood method**. *Acta Crystallogr D Biol Crystallogr* 1997, **53**:240-255.

76. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC: **Structure validation by Calpha geometry: phi,psi and Cbeta deviation**. *Proteins* 2003, **50**:437-450.

77. Laskowski BA, MacArther MW, Moss DS, Thornton JM: **PROCHECK: A Programm to Check the Stereochemical Quality of Protein Structures**. *J. Appl. Cryst.* 1993, **26**:283-291.

78. Matthews BW: **Solvent content of protein crystals**. *J Mol Biol* 1968, **33**:491-497.

79. Starostina NG, Marshburn S, Johnson LS, Eddy SR, Terns RM, Terns MP: **Circular box C/D RNAs in Pyrococcus furiosus**. *Proc Natl Acad Sci U S A* 2004, **101**:14097-14101.

80. Gruber J, Zawaira A, Saunders R, Barrett CP, Noble ME: **Computational analyses of the surface properties of protein-protein interfaces**. *Acta Crystallogr D Biol Crystallogr* 2007, **63**:50-57.

Figure 4.1 Crystal structure of archaeal Sm4 protein (PDB entry 1YCY). A) The Sm4 monomer (A chain) colored from blue (N-terminus) to red (C-terminus) showing secondary structural details (note residues 25 to 28 were model based on PDB entry 1M8V) [39]. B) The Sm4 octamer viewed facing the putative RNA binding face. The figure is colored to show the 2-fold relationship among the four chain pairs that make up the octamer. Note the Sm4 A chain is colored as described above. C) The Sm4 hexadecamer viewed perpendicular to its putative RNA binding faces showing the interdigitation of the β3 and β4 loops (yellow) of the two octamers. The figures were generated using UCSF Chimera [59].

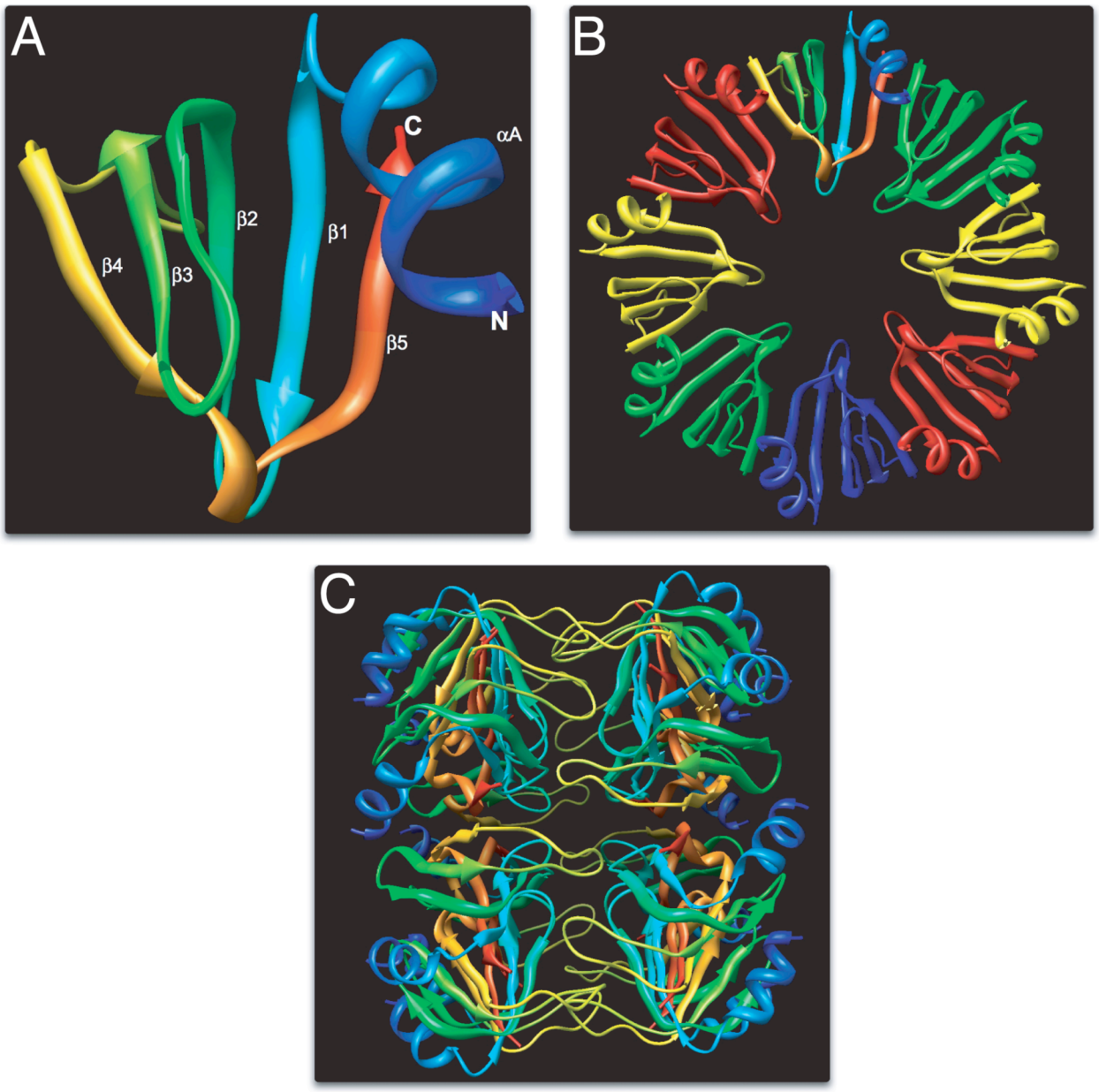

A
C
αA
β2
β1
β4
β3
N
β5
B
C

Figure 4.2 Co-immunoprecipitation of RNAs with Sm4 protein. A) Immunoblots of immunoprecipitated proteins using Sm4 immune (I) or pre-immune (PI) antibodies. B) Co-immunoprecipitated RNAs obtained using Sm4 immune (I) or pre-immune (PI) antibodies, fractionated by electrophoresis on 8% polyacrylamide/7 M urea gels and visualized by SYBR-Green RNA staining. Several RNA species (bands 1-6) from *P. furiosus* are selectively co-immunoprecipitated with anti-Sm4 antibodies (I) relative to preimmune antibodies (PI) and total RNA (T).

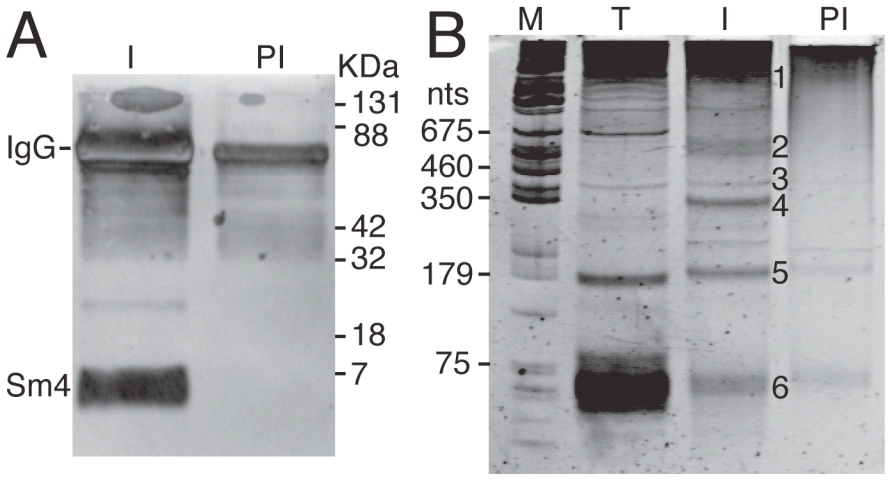
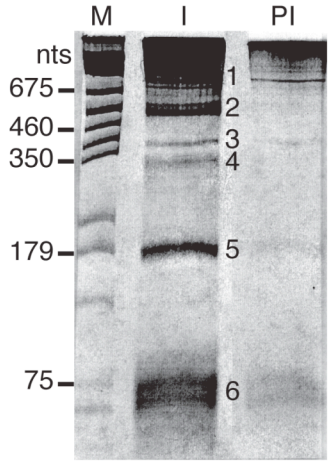

A
I
PI
KDa
131
88
42
32
18
7
IgG
Sm4
B
M
T
I
PI
nts
675
460
350
179
75
1
2
3
4
5
6

Figure 4.3 Identification of Sm4-associated RNAs. A) Six RNA bands from *P. furiosus* selectively co-immunoprecipitated with anti-Sm4 antibodies (I) were cloned (see materials and methods). B) Sequence analysis of Sm4 co-immunoprecipitated RNAs. * indicates that this RNA sequence contains sequence of mature 23S rRNA and the upstream spacer. ** indicates that each sR48 and sR55 detected in band 3 including flanking sequence (see Table 4). The nucleotide coordinates for the novel Pf10 RNA (of band 6) are 128191 to 128132 (also see Table 4).

A



B

Band	RNA	# of Clones
1	23S rRNA	9
	16S rRNA	4
	23S rRNA*	4
	RNase P	1
2	Pf7 (H/ACA)	20
3	23S rRNA	31
	16S rRNA	9
	Pf6 (H/ACA)	4
	Pf7 (H/ACA)	3
	sR48 (C/D)**	1
	sR55 (C/D)**	1
	mRNA (PF1786)	1
	mRNA (PF1558)	1
4	23S rRNA	27
	Pf6 (H/ACA)	21
	16S rRNA	1
	5S RNA	1
5	Pf8	11
	23S rRNA	9
	5S rRNA	2
	16S RNA	1
6	sR29 (C/D)	17
	sR46 (C/D)	9
	Pf10	8
	sR12 (C/D)	6
	sR5 (C/D)	5
	sR16 (C/D)	3
	mRNA (PF2015)	2
	sR55 (C/D)	2
	sR1 (C/D)	1
	sR2 (C/D)	1
	sR11 (C/D)	1
	sR15 (C/D)	1
	sR32 (C/D)	1
	sR34 (C/D)	1
	sR39 (C/D)	1
	sR42 (C/D)	1
	Pf9 (H/ACA)	1

Figure 4.4 Northern blot analysis of Sm4 co-immunoprecipitated RNAs. The specificity of the association of A) C/D RNAs, B) H/ACA RNAs, C) uncharacterized non-coding RNAs, D) ribosomal RNAs, and E) RNase P RNA and tRNA-Asp were tested by Northern blot analysis of RNAs co-precipitated with immune (I) or pre-immune (PI) antibodies. In each panel, 10% of the total (T), 10% of the supernatant (S) and 100% of the pellets (P) recovered from either immune or preimmune antibody reactions were resolved on 8% polyacylamide/7 M urea gels and transferred to membranes. The RNAs were hybridized by ^{32}P-DNA specific probes and visualized by phosphorimager. Note that the probe for sR55 detected an RNA the approximate size of the full-length sR55 [53] and smaller than the fragments obtained by sequencing (Figure 4.4B).

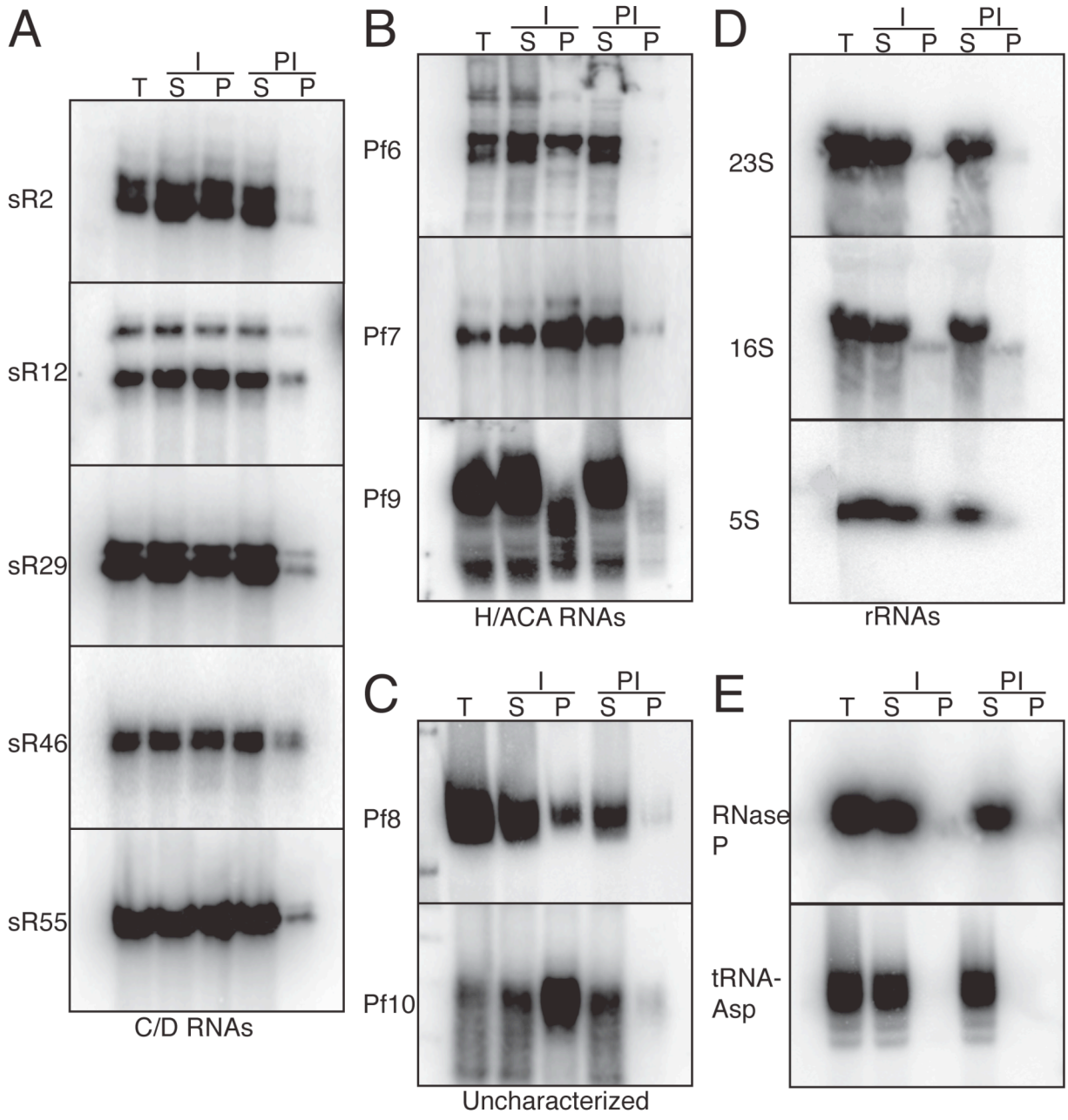

A
T
I
S
P
PI
S
P
sR2
sR12
sR29
sR46
sR55
C/D RNAs
B
Pf6
Pf7
Pf9
H/ACA RNAs
C
Pf8
Pf10
Uncharacterized
D
23S
16S
5S
rRNAs
E
RNase P
tRNA-Asp

Figure 4.5 Gel mobility shift of H/ACA and C/D RNAs. Approximately 0.05 pmol of ^{32}P-radiolabeled H/ACA RNAs (A), C/D RNAs (B) or 5S rRNA or tRNA-Asp (C) were incubated with increasing amounts (μM) of Sm4 recombinant protein or with 100 nM L7Ae protein. The RNA and RNP complexes were separated on 6% or 8% non-denaturing gels and visualized by autoradiography.

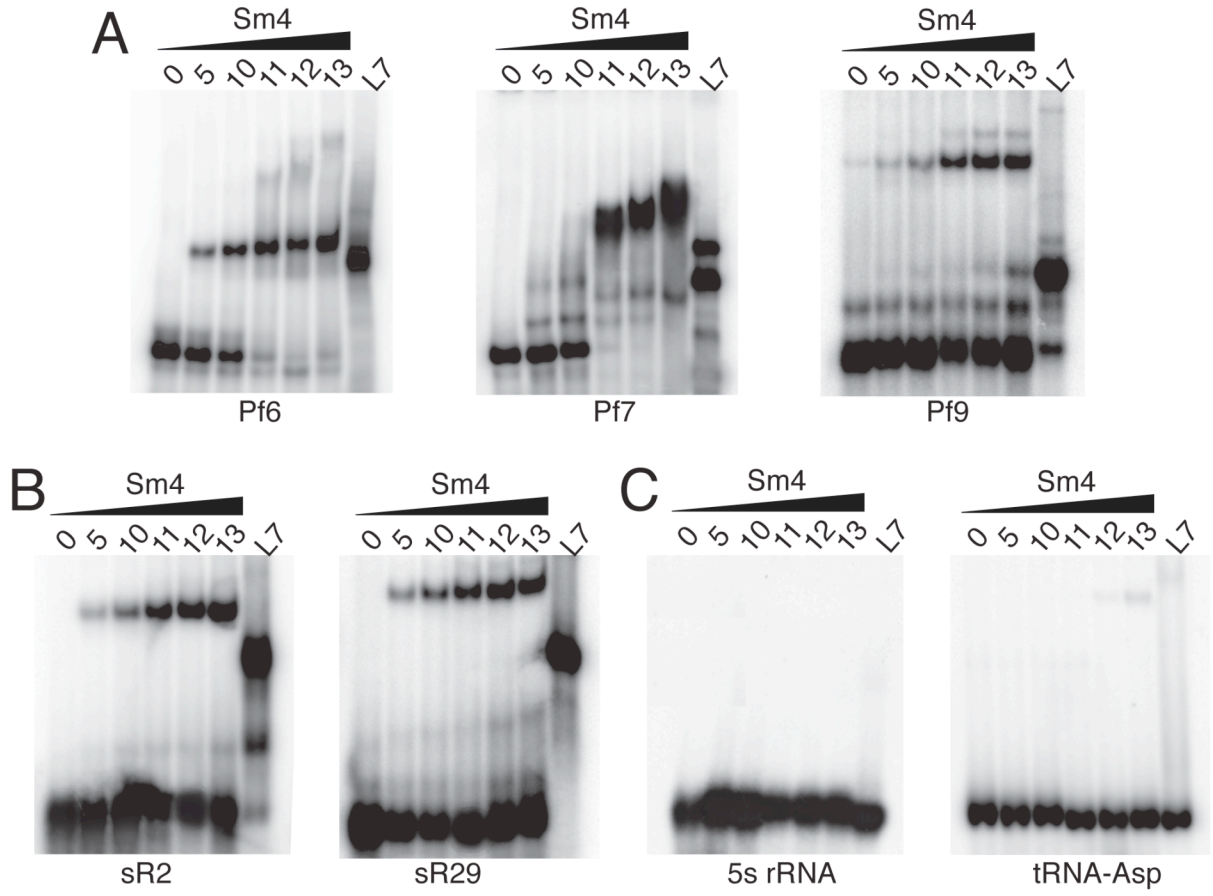

A
Sm4
0 5 10 11 12 13 L7
Pf6
Sm4
0 5 10 11 12 13 L7
Pf7
Sm4
0 5 10 11 12 13 L7
Pf9
B
Sm4
0 5 10 11 12 13 L7
sR2
Sm4
0 5 10 11 12 13 L7
sR29
C
Sm4
0 5 10 11 12 13 L7
5s rRNA
Sm4
0 5 10 11 12 13 L7
tRNA-Asp

Figure 4.6 Native gel mobility assay of H/ACA RNAs with Sm4 and L7Ae proteins. (A) Sm4 protein does not interact with the k-turn motif. ^{32}P-labeled Pf9 wildtype or k-turn mutant was incubated with 13 μM Sm4 and/or 100 nM L7Ae. The RNA and RNP complexes were separated on a native polyacrylamide gel and visualized by autoradiography. (B) Sm4 and L7Ae proteins interact with distinct RNA binding sites. Pf6 wildtype was incubated with Sm4, L7Ae, or both proteins as described in A.

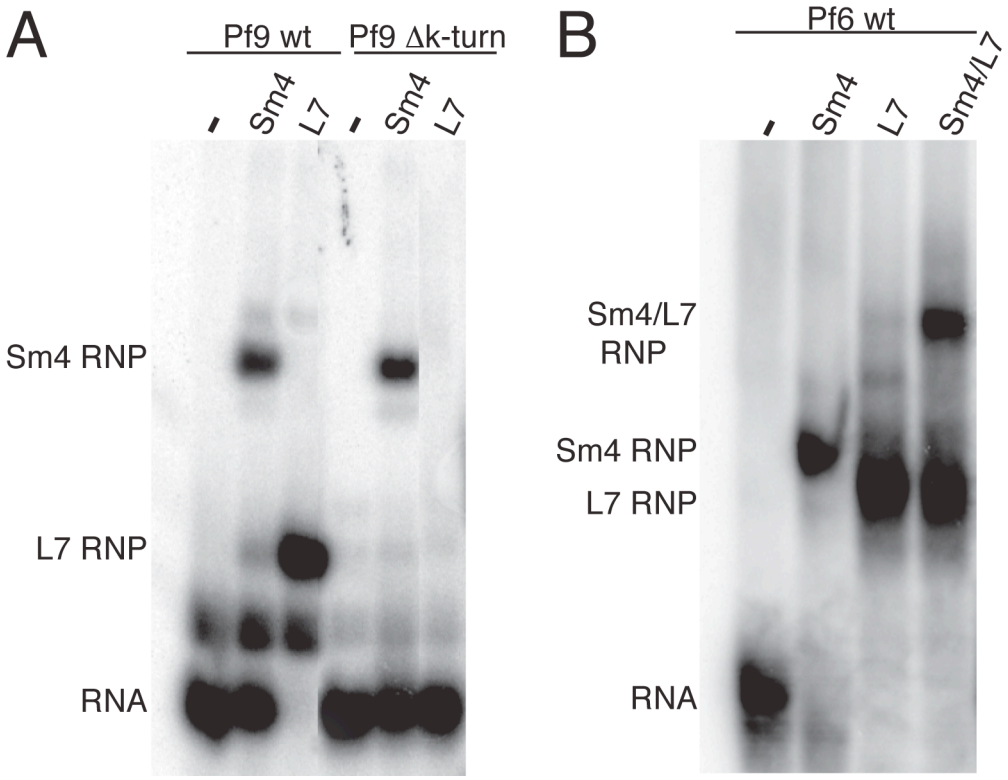

A
Pf9 wt
Pf9 Δk-turn
-
Sm4
L7
-
Sm4
L7
Sm4 RNP
L7 RNP
RNA
B
Pf6 wt
-
Sm4
L7
Sm4/L7
Sm4/L7 RNP
Sm4 RNP
L7 RNP
RNA

Figure 4.7 Electrostatic potential distribution of *P. furiosus* Sm4 and *P. abyssi* Sm1 (PDB entry 1M8V) [39] surfaces. A) A view of the Sm4 octamer (orientation is similar to that of figure 1B) showing the putative RNA binding face. B) A view of the *P. abyssi* Sm1 heptamer with its bound RNA removed for clarity. The electrostatic potential calculated is colored from negative (< 0.5 V, red) to neutral (white) to positive (> 0.5 V, blue) values. Electrostatic potentials were calculated using CCP4 Molecular Graphics [80]. C) Superposition of *P. abyssi* Sm1 monomer (green ribbon) with bound RNA (sticks) onto the Sm4 structure (orange ribbon) showing that with a minor repositioning of αA in the Sm4 structure could accommodate the U_7 RNA. D) A view of the residues involved in RNA binding in the *P. abyssi* structure and corresponding residues in the αA repositioned Sm4 model showing that all of the RNA main-chain interactions can be accommodated. Color scheme: red- oxygen, blue- nitrogen, white- RNA carbons, green- *P. abyssi* carbons and orange- Sm4 carbons.

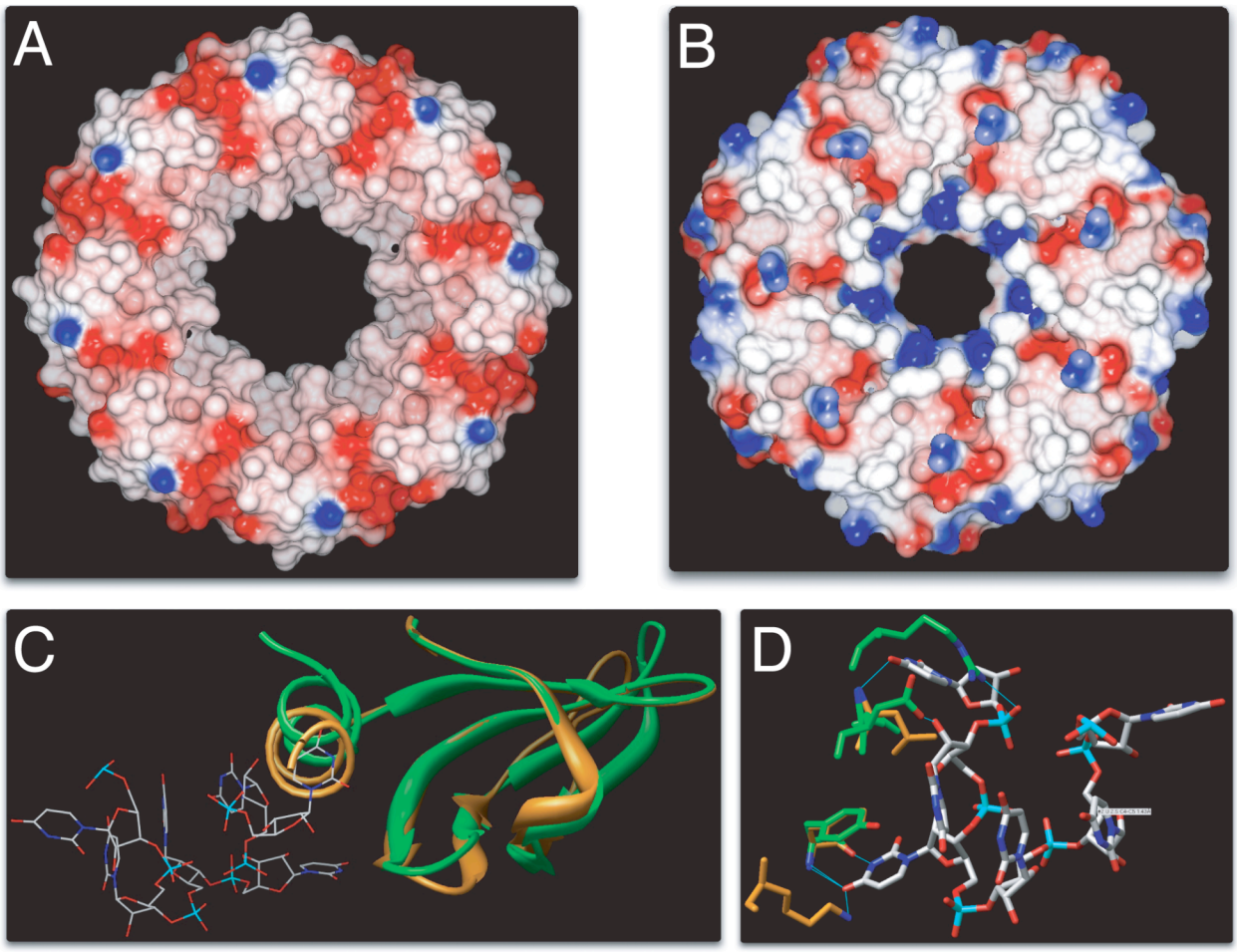

A
B
C
D

Table 1. Statistics of data collection and structure refinement

Crystal	
Space group	$P6_222$
a (Å)	82.96
c (Å)	189.78
Data collection	
Source	22ID APS
Detector	MAR 225 CCD
Wavelength (Å)	0.9785
Distance (mm)	150
2θ	0.0
Phi step (°):	0.5
Total Rotation (°)	360
Data processing:	HKL 1.9.1
Resolution	2.67
Completeness (%)	95.0 (61.6)
R_{sym}*	0.089 (0.294)
*outer shell (2.77 - 2.67 Å) values in parentheses	
Refinement	
Program	REFMAC 5
Resolution (Å)	20.00 - 2.80
Completeness* (%)	100.00
$R_{crystal}$*	0.281 (0.381)
R_{free}*	0.311 (0.469)
*outer shell (2.87 - 2.80 Å) values in parentheses	
R.M.S. Deviations from ideality	
Bond lengths (Å)	0.015
Bond angles (°)	1.343
*Ramachandran analysis***	
Most favored (%)	88.2 (100.0)**
Disallowed (%)	0.0
**PROCHECK percent in all allowed regions shown in parentheses	
Final model	
Residues	5 - 24, 29 - 70 (chains A - D)
Solvent atoms	0
PDB ID	1YCY

Table 2: Sm4 protein structural comparison

DALI (distance-matrix alignment) Search Results						
PDB ID	Z-SCORE	%ID	RMSD	Assembly	Kingdom	Comment
1YCY_A	15.7	100	0.0	2(8-mer)	Archaea	
1M5Q_W	10.6	21	1.4	14-mer	Archaea	Sm-Like RNP
1I5L_C	10.3	19	1.9	7-mer	Archaea	Sm-Like RNP
1I8F_F	10.2	22	1.6	7-mer	Archaea	Sm-Like RNP
1LJO_A	10.1	18	1.5	6-mer	Archaea	Sm-Like RNP
1M8V_A	10.1	19	1.9	2(7-mer)	Archaea	Sm-Like RNP
1D3B_E	9.7	20	2.0	2-mer	Eukaryota	Sm-Like RNP
1JBM_A	9.4	24	1.7	7-mer	Archaea	Sm-Like RNP
1B34_A	8.7	19	1.9	2-mer	Eukaryota	Sm-Like RNP
1N9S_C	8.7	25	2.1	2(7-mer)	Eukaryota	Sm-Like RNP
1KQ1_H	8.7	29	1.4	6-mer	Bacteria	Sm-Like RNP
1U1T_E	8.6	25	1.0	6-mer	Bacteria	Sm-Like RNP
1B34_B	8.5	13	2.1	2-mer	Eukaryota	Sm-Like RNP
SSM (secondary structure matching) Search Results						
1I8F_E	6.2	23	1.4	7-mer	Archaea	Sm-Like RNP
1I5L_M	6.7	20	1.4	7-mer	Archaea	Sm-Like RNP
1I4K_L	6.0	20	1.2	7-mer	Archaea	Sm-Like RNP
1LNX_D	6.6	23	1.5	2(7mer)	Archaea	Sm-Like RNP
1M8V_N	6.2	20	1.3	2(7-mer)	Archaea	Sm-Like RNP
3BW1	5.6	30	1.0	2(8-mer)	Eukaryota	Sm-Like RNP

Table 3: Sm4 interchain hydrogen bonds

Donor				**Acceptor**				
Residue	Number	Chain	Atom	Residue	Number	Chain	Atom	Distance
TRP	14	A	NE1	GLN	56	B	OE1	3.12
LEU	58	A	N	ILE	67	C	O	2.91
LEU	69	A	N	GLN	56	B	O	3.07
GLN	56	B	N	LEU	69	A	O	3.35
LEU	58	B	N	ILE	67	A	O	2.87
TRP	14	C	NE1	GLN	56	A	OE1	3.16
ILE	67	C	N	LEU	58	A	O	3.28
Leu	58	C	N	ILE	67	D	O	2.82
Ile	67	D	N	LEU	58	C	O	3.40
Leu	69	D	N	GLN	56	C	O	3.23

Table 4: Nucleotide coordinates of Pf10, sR48 and sR55 genes

	Start Coordinates	End Coordinates	Length (nts)	Strand
Pf10	128191	128132	60	-
sR48	361376	361318	59	-
sR48**	361371	361219	153	
sR55	1245430	1245493	64	+
sR55**	1245400	1245586	187	

CHAPTER 5

DISCUSSION AND FUTURE DIRECTIONS

The work presented in this thesis discussed two major subjects: (1) the *in vitro* assembly and dynamic interactions of H/ACA RNP components in archaea and (2) the crystal structure of archaeal Sm4 complex and its potential role in RNA modification.

Pseudouridine is the most abundant modified nucleotide found in tRNAs, rRNAs, snRNAs, snoRNAs, and other non-coding RNAs in three domains of life [1,2]. Experimental studies showed the importance of pseudouridylation for cell viability in eukarya and eubacteria [3]. In archaea, the function of pseudouridylation is still greatly unknown. However, the observation that the number of modified nucleotides correlates well with the archaeon's optimum growth temperature suggests that the modifications may be required for the stability of the modified RNAs at elevated temperatures [1].

In eukaryotes and archaea, many pseudouridylations are introduced by RNA-guided pseudouridine synthases [2]. The RNA-guided mechanism is versatile and uses H/ACA guide RNAs to direct a common pseudouridine synthase to many different sites. The H/ACA RNAs, like most cellular RNAs, associate with proteins to form RNPs. H/ACA guide RNA associates with Cbf5 (pseudouridine synthases) and three accessory proteins, Gar1, Nop10, and L7Ae (Nhp2 in eukaryotes) [3]. The roles of these accessory proteins in pseudouridylation were greatly unknown. The overall specific aim of the first project was to determine in greater detail the architecture of the archaeal H/ACA RNPs in the hopes of gaining some insight into the mechanism of RNA-guided RNA pseudouridylation. The production of stable and soluble active archaeal proteins was an advantage in this project. The results presented in Chapters 2 and 3 provide a substantial amount of new information on the structure and function of the RNP that catalyzes RNA-guided RNA pseudouridylation in archaea.

We used a variety of *in vitro* approaches to determine the minimal components required for the pseudouridylation activity and to investigate the interactions (RNA/protein and protein/protein) between H/ACA RNP components [4]. We first showed that *in vitro* reconstituted H/ACA RNP complexes are enzymatically active under the experimental conditions. We then investigated the RNA/protein and protein/protein interactions under the same reaction conditions. Our findings in Chapter 2 provided a basic model for the organization of the archaeal H/ACA RNP complexes, in which Cbf5 and L7Ae bind directly and independently to different sites on the guide RNA, and Gar1 and Nop10 interact independently with Cbf5. The subsequent Cbf5/Gar1/Nop10 crystal structure confirmed the formation of the protein complex and the independent binding of Gar1 and Nop10 to Cbf5 [7]. Since no one has tried to address the order of addition of the core proteins to the guide RNA, it is equally possible that either Cbf5 (alone or in complex with Gar1 and Nop10) or L7Ae binds first to the guide RNA. However, it was shown that Cbf5 interacts with the newly made H/ACA RNA in yeast, suggesting that Cbf5 could be the first core protein to bind to the guide RNA [8].

It was to our surprise that L7Ae (but not eukaryotic Nhp2) is not a part of the protein complex, in the absence of the RNA. This difference between the eukaryotes and archaea in the assembly of the complexes may be attributed to the inability of Nhp2 to directly bind to the guide RNA [5]. The RNA footprinting results presented in Chapter 3 showed the dynamic interactions within H/ACA RNP sub-complexes (RNA/Cbf5 and RNA/Cbf5/L7Ae) [6]. The results showed how binding of L7Ae to the guide RNA causes the remodeling of the Cbf5 binding site. Mapping of the Cbf5 binding site in the Cbf5/guide RNA sub-complex revealed expected binding at the conserved ACA elements, the lower stem, and pseudouridylation pocket (as shown in the crystal structure of H/ACA RNP holoenzyme [7]), but also unexpected binding

at the 5' half of the k-turn and apical loop of the RNA [6]. In addition, protein footprinting of the Cbf5/guide RNA sub-complex found amino acid protections along the same RNA binding site observed in the crystal structure of the holoenzyme [11,12]. However, the protein footprinting results also showed a set of protections in the catalytic domain of Cbf5 that appears to represent an additional interaction of the guide RNA with the protein in the sub-complex [8]. In summary, the results of the two footprinting studies strongly suggest a novel interaction between Cbf5 and the guide RNA in the RNA/Cbf5 sub-complex. The importance of this interaction in the formation of the Cbf5-guide RNA complex is demonstrated by the disruption of the complex that occurs upon mutation of the unexpected binding region of the RNA [4]. Addition of L7Ae to the Cbf5/RNA sub-complex disrupted the additional interaction between Cbf5 and the guide RNA (the 5' half of the k-turn and apical loop). Moreover, the footprinting data unveiled the functional role of L7Ae (see below).

On the basis of the results of this thesis and other recent work, the functions of H/ACA RNP core proteins can be predicted. Available evidence in eukaryotes indicated that Cbf5 is indeed the enzymatic component of the complex. Work from our lab also confirmed that Cbf5 is the modifying enzyme. Mutation of the conserved aspartate residue (which binds to the target uridine to initiate the modification) in the catalytic subunit of archaeal Cbf5 completely abolished the pseudouridylation activity of H/ACA RNP complexes [8]. Moreover, the crystal structure of archaeal Cbf5 showed similarities between Cbf5 and TruB, the *E. coli* pseudouridine synthase, providing more evidence that Cbf5 is the enzymatic component of H/ACA RNP complexes [9].

The footprinting data shown in Chapter 3 unexpectedly supported a model in which L7Ae binding is required for the organization of the secondary structure of the guide RNA [6].

Previous biochemical studies showed that L7Ae is absolutely important for the pseudouridylation activity of H/ACA RNPs [4,10]. Our footprinting assays showed that binding of L7Ae to the guide RNA is required for the formation of the k-turn motif and also the pseudouridine pocket (the target RNA binding site) [6]. This study unmasked the functional role of L7Ae in pseudouridylation activity. The co-crystal structure of archaeal H/ACA RNP/substrate RNA in the absence of L7Ae showed that the target uridine is about 11 Å away from the catalytic site of Cbf5 and subsequently cannot be modified [11]. However, fluorescence experiments showed that addition of L7Ae caused the movement of the target uridine bound to the guide RNPs in to close proximity to the active site [11]. These observations provide more evidence for the importance of L7Ae in the correct positioning of the target uridine.

Although there is no experimental evidence to date about the exact function of either Gar1 or Nop10, data from the biochemical analysis and crystal structures provided a glimpse into the role of both proteins. The crystal structures of the Cbf5/Nop10 dimer [12] and Cbf5/Gar1Nop10 trimer [7] are very similar, suggesting that Gar1 and Nop10 act independently. The crystal structure of H/ACA RNP holoenzyme showed that Nop10 is located between Cbf5 and L7Ae interacting with the catalytic subunit of Cbf5 and with L7Ae in addition to the upper stem of the guide RNA [7]. The interactions between Nop10 and both L7Ae and the guide RNA have not been detected via biochemical analysis. These interactions apparently occur in the context of the RNP complexes. The location of Nop10 in this position in the holoenzyme suggests that the binding of Nop10 at this position is required to keep the catalytic subunit of Cbf5 close to the target uridine.

To data, the function of Gar1 is still mysterious. Genetic depletion of Gar1 in yeast results in partially assembled RNP complexes that are unable to interact with substrate rRNA [13]. Moreover, cross-linking experiments showed that mammalian Gar1 can bind close to the target uridine in the substrate RNA and may be required for the binding and/or release of the target RNA [5]. However, biochemical approaches showed that Gar1 is not required for the binding of the target RNA to the guide RNA *in vitro* [10]. The X-ray structure of H/ACA RNPs holoenzyme showed that Gar1 binds to the thumb loop of Cbf5 (the corresponding thumb loop of *E. coli* pseudouridine synthase TruB binds extensively to the target tRNA [9]). This observation indicates that Gar1 may hold the thumb loop in an open conformation to promote target RNA loading and release. However, the X-ray structure of H/ACA RNPs (minus L7Ae and the k-turn motif)/target RNA showed that Gar1 only binds to the catalytic subunit of Cbf5 and does not contact either the guide or substrate RNAs [11].

Archaeal H/ACA RNA can have one, two, or three hairpins, but the vast majority of eukaryotic H/ACA RNAs contain two hairpin domains. Electron microscopy data have been interpreted to reflect that each hairpin domain associates with one set of the four proteins [14]. Although eukaryotic H/ACA RNP complexes were discovered more than ten years ago, little was known about the assembly pathway of H/ACA RNP complexes. The biochemical analysis and the subsequent crystal structures of H/ACA RNP components from our lab along with other labs greatly helped to understand the archaeal H/ACA RNP complexes [6,10,11,14,15]. The similarities between archaeal and eukaryotic organisms helped us and others to better understand the eukaryotic H/ACA RNPs. The components of eukaryotic and archaeal pseudouridylation guide RNPs are generally well conserved, suggesting that the organization and function of the components will be fundamentally similar in the two systems. No enzymatically active

eukaryotic H/ACA RNPs has been reported to date. However, two studies tried to describe interactions between various components of eukaryotic H/ACA RNPs [9,19]. The interactions observed in the yeast study [15] are in good agreement with those reported in Chapter 2, while there are significant differences in the interactions observed in the mammalian system [5].

Mammalian homologs of Cbf5, Nop10 and L7Ae [5], rather than Cbf5, Nop10 and Gar1 form a complex in the absence of the guide RNA [5,15]. In addition, Nop10 is essential for interaction between the mammalian Cbf5 and L7Ae homologs, and thus appears to play the central role in this complex [5], while Cbf5 is at the core of the archaeal complex, interacting independently with each Gar1 and Nop10. In the mammalian system, specific recognition of H/ACA RNAs required all three components of the trimeric complex [5]. On the other hand, we have found that archaeal Cbf5 interacts specifically with guide RNAs in the absence of the other proteins, and that Gar1 and Nop10 do not observably increase the affinity of the interaction.

On the other hand, the data from yeast suggest that the organization of the yeast H/ACA RNP resembles the archaeal complex. In yeast, Cbf5, Gar1 and Nop10 can form a complex independent of both Nhp2 (archaeal L7Ae) and guide RNA [15].

At present it is not clear whether the observed differences between the mammalian system and the archaeal and yeast systems reflect essential differences in the RNPs or the limitations of experimental approaches. No pseudouridylation activity could be detected with the complexes assembled in the mammalian system [5]. The functionality of the purified yeast complexes was not reported. The eukaryotic H/ACA RNP proteins, and especially Cbf5, are challenging to express and purify [5,15]. It is possible that both the lack of functionality of the mammalian proteins and the observed differences result from production of misfolded mammalian proteins *in vitro*. A better understanding of the extent of differences between the

eukaryotic and archaeal RNPs awaits more detailed structural studies of functional eukaryotic complexes. Taken together, the results in the first part of this thesis formed the backbone of many other studies that greatly contributed in understanding H/ACA RNP complexes.

The work presented in Chapters 2 and 3 in addition to other related works leaves one important question open, why is Gar1 required for the pseudouridylation activity of H/ACA RNPs? The uncertainty function of Gar1 needs to be addressed. The 3D structure of H/ACA RNP (minus L7Ae and k-turn)/target RNA complexes could not explain the function of Gar1 [11]. A high-resolution structure of the fully assembled H/ACA RNP containing the target RNA and complementary biochemical studies are required to answer this question.

The second project presented in Chapter 4 discussed a novel archaeal Lsm protein (Sm4) and the associated RNAs *in vivo* and *in vitro*. In all organisms, Sm/Lsm/HfQ proteins form homo- or hetero- polymeric ring-like structures that are involved in the function of a wide variety of RNA species [16-18]. The available crystal structures of Lsm/HfQ complexes revealed that Lsm proteins can form stable polymeric complexes in the absence of the RNA. Three different Lsm proteins (Sm1, Sm2 and Sm3) were identified in different archaeal species [19-21]. We reported a forth archaeal protein (a second *Pyrococcus* Lsm protein) in Chapter 4. Sm1 is the most abundant Lsm protein among the archaeal species [19], while Sm3 [21] and Sm4 (this work) are considered the less abundant archaeal Lsm proteins. Sequence analysis showed that only *Pyrobaculum aerophilum, Sulfolobus tokodaii*, and *S. solfataricus* contain an Sm3 gene [21]. On the other hand, *Pyrococcus* (*P. furiosus*, *P. abyssi* and *P. horikoshii*) species in addition to *Thermococcus kodakarensis* are the only archaeal species that contain the Sm4 gene.

The crystal structure described in Chapter 4 showed that unlike most of the known Lsm proteins, the archaeal Sm4 protein can form a stable homo-octameric complex. The only known Lsm protein that can form an octamer is the yeast Lsm3 protein [22]. Multiple sequence alignment failed to identify this protein as a member of Sm/Lsm/HfQ proteins family. This is mainly due to the lack of sequence homology between Sm4 and other Sm/Lsm/HfQ proteins in different organisms. However, the crystal structure described in detail here showed that Sm4 has structural homology to all known Sm/Lsm/HfQ proteins. This finding suggests that other Sm/Lsm/HfQ proteins that are structural homologs but not primary sequence homologs may be out there to discover.

Very limited information is available about the functional role of archaeal Lsm proteins. Preliminary studies indicated that Sm1 and Sm2 from *A. fulgidus* may be involved in the processing of tRNAs as they associate with RNase P RNA that is required for tRNA maturation [19]. In the same time, there is no prediction or experimental evidence about the function of archaeal Sm3. To provide an insight into the function of Sm4 in archaea, we used different biochemical approaches. This study indicated that Sm4 has a distinct function from other archaeal Lsm proteins. Our results demonstrated that members of H/ACA and C/D guide RNAs are associated specifically with Sm4. These results suggest that Sm4 protein may be a functional component of H/ACA and C/D guide RNAs. We and others demonstrated that pseudouridylation and ribose methylation activities of the guide RNPs do not require additional factors (e.g., proteins, energy or helicase) *in vitro* [3,4,23]. At the present time, we limit our studies to *in vitro* approaches, as the genetic manipulation of *P. furiosus* is still very limited. However, we cannot rule out that the modification mechanism inside the living cell may require other factors such as Sm4 protein. For example, Sm4 could transiently interact with the guide

RNAs to release the target RNA after the modification. This could explain the competition that we found between Sm4 protein and the target RNA to interact with the antisense region of sR2 C/D guide RNA.

Another possible function is that Sm4 can assist in the maturation of guide RNAs. Information about the archaeal guide RNAs biogenesis and the factor(s) included in this process are still very limited. The role of Sm4 in the guide RNA biogenesis is suggested by our finding that Sm4 associates with a longer sequence of each sR48 and sR55 C/D guide RNAs. Previous studies also found that eukaryotic Lsm proteins associate with members of H/ACA and C/D RNAs suggesting that Lsm proteins may require for the function or/and maturation of these RNAs [24-26].

Pf8 and Pf10 are uncharacterized non-coding RNAs that do not contain the conserved elements characterized of H/ACA or C/D RNAs and hence they are unlikely belong to guide RNAs family. Theses two non-coding RNAs are also associated specifically with Sm4. This indicates that Sm4 probably involves in other processes inside the cell that are distinct from its potential role in RNA-guided RNA modification. What is the function of each of these two RNAs in a complex with Sm4? More bioinformatics and experimental studies are required to answer this question since we do not have any hint about the identity of these RNAs.

We tried to determine the protein and RNA binding sites in Sm4/guide RNA complexes. Previous studies indicated that the RNA can bind to Lsm complexes in the central cationic pore of the ring (inner binding site). This cationic pore is highly conserved in terms of sequence and overall structure. In contrast, Sm4 bears a negatively charged batch in the central pore of the complex. It seems unlikely that the RNA binds to the negatively charged central pore of the complex. However, other crystal structures show additional RNA-binding sites on the top of the

ring (outer binding site) [20,27]. We addressed the possibility that the associated RNAs could bind to Sm4 in the same way. We superimposed Sm4 on the *P. abyssi* Sm1/RNA complex and we found that the RNA can interact with Sm4 on the top of the ring (outer binding site in a similar manner.

Our partial analysis to determine the protein binding sites on the guide RNAs showed that the single-stranded region (antisense) located between the conserved C and D' elements in sR2 C/D guide RNA mediate the interaction. This single stranded region is common in H/ACA and C/D guide RNAs (the binding site of the target RNAs). This finding needs to be tested in other guide RNAs. Ultimately, high-resolution 3D structure of Sm4 with the guide RNAs will provide tremendous information about the RNA and protein binding sites in the RNA/Sm4 complexes.

Taken together, this work sheds light on the existence of a novel Sm protein in archaea and more important, this work gives insights into the potential roles of Sm4 in archaea.

References

1. Dennis PP, Omer A, Lowe T: **A guided tour: small RNA function in Archaea**. *Mol Microbiol* 2001, **40**:509-519.

2. Terns M, Terns R: **Noncoding RNAs of the H/ACA family**. *Cold Spring Harb Symp Quant Biol* 2006, **71**:395-405.

3. Yu YT, Terns RM, Terns MP: **Mechanisms and functions of RNA-guided RNA modification**. In *Fine-tuning of RNA functions by modification and editing*. Edited by Grosjean H: Topics in Current Genetics; 2005:223-262. vol 12.

4. Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP: **RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP**. *Genes Dev* 2005, **19**:1238-1248.

5. Wang C, Meier UT: **Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins**. *Embo J* 2004, **23**:1857-1867.

6. Youssef OA, Terns RM, Terns MP: **Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP**. *Nucleic Acids Res* 2007, **35**:6196-6206.

7. Li L, Ye K: **Crystal structure of an H/ACA box ribonucleoprotein particle**. *Nature* 2006, **443**:302-307.

8. Baker DL, Seyfried NT, Li H, Orlando R, Terns RM, Terns MP: **Determination of protein-RNA interaction sites in the Cbf5-H/ACA guide RNA complex by mass spectrometric protein footprinting**. *Biochemistry* 2008, **47**:1500-1510.

9. Hoang C, Ferre-D'Amare AR: **Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme**. *Cell* 2001, **107**:929-939.

10. Charpentier B, Muller S, Branlant C: **Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation**. *Nucleic Acids Res* 2005, **33**:3133-3144.

11. Liang B, Xue S, Terns RM, Terns MP, Li H: **Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex**. *Nat Struct Mol Biol* 2007.

12. Hamma T, Reichow SL, Varani G, Ferre-D'Amare AR: **The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs**. *Nat Struct Mol Biol* 2005, **12**:1101-1107.

13. Bousquet-Antonelli C, Henry Y, G'Elugne J P, Caizergues-Ferrer M, Kiss T: **A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs**. *EMBO J* 1997, **16**:4770-4776.

14. Lubben B, Fabrizio P, Kastner B, Luhrmann R: **Isolation and characterization of the small nucleolar ribonucleoprotein particle snR30 from Saccharomyces cerevisiae**. *J Biol Chem* 1995, **270**:11549-11554.

15. Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M: **Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs**. *Rna* 2004, **10**:1704-1712.

16. Beggs JD: **Lsm proteins and RNA processing**. *Biochem Soc Trans* 2005, **33**:433-438.

17. Khusial P, Plaag R, Zieve GW: **LSm proteins form heptameric rings that bind to RNA via repeating motifs**. *Trends Biochem Sci* 2005, **30**:522-528.

18. Wilusz CJ, Wilusz J: **Eukaryotic Lsm proteins: lessons from bacteria**. *Nat Struct Mol Biol* 2005, **12**:1031-1036.

19. Toro I, Thore S, Mayer C, Basquin J, Seraphin B, Suck D: **RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex**. *EMBO J* 2001, **20**:2293-2303.

20. Thore S, Mayer C, Sauter C, Weeks S, Suck D: **Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya**. *J Biol Chem* 2003, **278**:1239-1247.

21. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D: **Structure and assembly of an augmented Sm-like archaeal protein 14-mer**. *Proc Natl Acad Sci U S A* 2003, **100**:4539-4544.

22. Naidoo N, Harrop SJ, Sobti M, Haynes PA, Szymczyna BR, Williamson JR, Curmi PM, Mabbutt BC: **Crystal structure of Lsm3 octamer from Saccharomyces cerevisiae: implications for Lsm ring organisation and recruitment**. *J Mol Biol* 2008, **377**:1357-1371.

23. Tran EJ, Zhang X, Maxwell ES: **Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs**. *EMBO J.* 2003, **22**:3930-3940.

24. Tomasevic N, Peculis BA: **Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence**. *Mol Cell Biol* 2002, **22**:4101-4112.

25. Kufel J, Allmang C, Petfalski E, Beggs J, Tollervey D: **Lsm Proteins are required for normal processing and stability of ribosomal RNAs**. *J Biol Chem* 2003, **278**:2147-2156.

26. Fernandez CF, Pannone BK, Chen X, Fuchs G, Wolin SL: **An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5**. *Mol Biol Cell* 2004, **15**:2842-2852.

27. Mura C, Kozhukhovsky A, Gingery M, Phillips M, Eisenberg D: **The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs)**. *Protein Sci* 2003, **12**:832-847.