

## ABSTRACT

TRAN, ELIZABETH JANE. Structure and Function of the Archaeal Box C/D Ribonucleoprotein Complex. (Under the direction of E. Stuart Maxwell.)

Box C/D ribonucleoprotein complexes (RNPs) are evolutionarily ancient nucleotide modification machines found in both Eukarya and Archaea. The box C/D RNAs are essential for ribosome biogenesis and primarily function by guiding 2'-O-methylation of ribosomal RNA (rRNA). The site of modification is determined by base-pairing between the target RNA and the box C/D RNA through a region of complementarity. The box C/D RNAs possess terminal box C/D and internal C'/D' motifs that fold to form K-turn RNA elements. In eukaryotes, the box C/D RNAs associate with a common set of four core proteins to form an RNP. The core proteins, 15.5kD, Nop56p, Nop58p and Fibrillarin, are differentially distributed on eukaryotic box C/D RNAs to form an asymmetric RNPs.

We have characterized the structure and function of the archaeal box C/D RNP using *Methanocaldococcus jannaschii* sR8 RNP as a model box C/D complex. Archaeal genomes contain genes for a Fibrillarin homolog and a single homolog for both Nop56p and Nop58p termed Nop56/58p. Our initial investigations identified ribosomal protein L7 as the archaeal homolog of the eukaryotic 15.5kD protein. Strikingly, L7 has a dual role as a component of both the ribosome and the box C/D RNP. A methylation-competent sR8 RNP was assembled *in vitro* using the three recombinant *M. jannaschii* box C/D RNA core proteins. This reconstituted complex is symmetric with respect to core protein binding and guides nucleotide modification from both the box C/D and C'/D' RNPs. Additionally, efficient RNA 2'-O-methylation requires juxtaposed box C/D and C'/D' motifs on the same box C/D RNP complex. Finally, the identification of box C/D RNPs in both Archaea and Eukarya led us to question the evolutionary origins of these ancient modification complexes. Based on the demonstration of a common RNP element (L7: K-turn motif) in both the archaeal large ribosomal subunit and the box C/D RNP

complex, we propose that the *trans*-acting nucleotide modification machines evolved elements in the primitive translational apparatus.

**Structure and Function of the Archaeal Box C/D Ribonucleoprotein Complex**

By

**Elizabeth Jane Tran**

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

**Molecular and Structural Biochemistry**

Raleigh, North Carolina

2004

**APPROVED BY:**

\_\_\_\_\_

(Advisory Committee Chair)

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## **DEDICATION**

To my husband, Daniel, for your constant love and support.

And for my mother, Mona, who taught me to persevere.

## **BIOGRAPHY**

Elizabeth (Beth) Jane Tran was raised in Terrell, Texas by her mother Mona Givens. She graduated with honors from Terrell High School in 1994 where she excelled in both the arts and academics.

Beth attended Texas A&M University in College Station, Texas where she graduated Magna cum Laude with a Bachelor's Degree in Genetics. After initially pursuing a career as a veterinarian, Beth realized her passion for basic science. She worked for one year as a research technician in the laboratory of Dr. Jim Golden at Texas A&M University where she met her husband, Daniel Tran.

Beth and Daniel moved to Raleigh, North Carolina in 1999 so Beth could attend graduate school at North Carolina State University under the direction of Dr. Stu Maxwell. Following completion of her doctoral degree, Beth will pursue a career in academia where she hopes to both teach and do research.

## ACKNOWLEDGEMENTS

First, I would like to thank Dr. E. Stuart Maxwell for his direction, support, and advice throughout my graduate career. I would also like to thank members of my committee and the biochemistry faculty for providing critical direction and guidance.

Next, I would like to thank my friends for their support especially Chele DeRider for the many discussions about life, graduate school and science. Also, thanks to current and past members of the Maxwell laboratory for help with various experiments. Thanks, above all, to Xinxin Zhang who maximized and my efficiency as a researcher.

Lastly, I would also like to acknowledge my family. Thank you, Mom, for teaching me the value of learning and for your encouragement throughout my educational process. Thanks, Gramma, for your warm support and listening to my stories about the "ups and downs" of graduate school. Also, thank you Uncle Buddy and Aunt Susan, for always taking caring of me and my family, even in times when we didn't know we needed assistance. Finally, I would also like to express my appreciation to my husband, Daniel, for all of the sacrifices he made so I could attend graduate school. Thanks also to my beautiful daughter, Kendall, just for being you.

## TABLE OF CONTENTS

LIST OF TABLES .....	ix
LIST OF FIGURES .....	x-xii
LITERATURE REVIEW .....	1
I. Eukaryotic Ribosome Biosynthesis .....	1
A. Nucleotide Modifications in rRNA .....	2
B. Ribosome Biosynthesis .....	3
II. Small Nucleolar RNAs .....	3
A. A Brief History of SnoRNAs .....	4
B. SnoRNA Processing in Eukaryotes .....	4
III. Classes of SnoRNAs .....	6
A. Box H/ACA SnoRNAs .....	7
B. Box C/D SnoRNAs .....	8
C. MRP .....	9
IV. SnoRNA-Protein Complexes .....	10
A. Box H/ACA Core Proteins .....	10
B. Box C/D Core Proteins .....	10
V. Archaea: A Separate Domain of Life from Eukaryotes and Bacteria .....	12
VI. Archaeal sRNAs and Core Proteins .....	13
A. Archaeal Box H/ACA sRNPs .....	13
B. Archaeal Box C/D sRNPs .....	14
VII. Identification of the Archaeal Homolog of Eukaryotic 15.5kD .....	15
A. Identification of Archaeal Ribosomal Protein L7 .....	15
B. Identification of the Kink-Turn Motif .....	16

C. Characterization of L7 Association with the Box C/D Core Motif .....	18
VIII. Archaeal Box C/D sRNP Assembly .....	19
A. Analysis of Ribosomal Protein L7 Binding to a Full-length Archaeal Box C/D RNA .....	20
B. Identification of Protein: Protein Interactions Between Nop56/58p And Fibrillarin .....	20
C. Demonstration of Archaeal Box C/D sRNP Symmetry .....	21
IX. RNP-guided Methylation Activity of the In Vitro Reconstituted Archaeal Box C/D RNP .....	21
X. Evolutionary Origins of the RNA-guided Nucleotide Modification Complexes .....	23
A. Origins of the <i>Trans</i> -acting Nucleotide Modification Machines .....	23
B. Model for Evolution of the Asymmetric, Eukaryotic Box C/D snoRNP From a Symmetric, Ancestral Box C/D RNP .....	25
XI. Summary of Dissertation Work .....	26
REFERENCES .....	27

CHAPTER I. Archaeal Ribosomal Protein L7 is a Functional Homolog of the Eukaryotic 15.5kD/Snu13p snoRNP Core Protein .....	34
ABSTRACT .....	35
INTRODUCTION .....	36
MATERIALS AND METHODS .....	38
RESULTS .....	42
Archaeal Ribosomal Protein L7 is a Functional Homolog of Eukaryotic 15.5kD snoRNP Protein and Binds the Box C/D snoRNA Core Motif .....	42

Archaeal L7 Protein Binds the Box C/D Core Motif with the Same Affinity As Eukaryotic 15.5kD Protein .....	43
L7 Requires the Same Box C/D Sequence and Structural Elements for Binding as Does the 15.5kD Protein .....	44
DISCUSSION .....	49
REFERENCES .....	54
CHAPTER II. Efficient RNA 2'-O-methylation Requires Juxtaposed and Symmetrically Assembled Archaeal Box C/D and C'/D' RNPs .....	66
ABSTRACT .....	67
INTRODUCTION .....	68
RESULTS .....	72
SRNP Core Proteins L7, Nop56/58, and Fibrillarin Bind Both the Terminal And Internal C'/D' Motifs to Assemble Symmetric RNP Complexes .....	72
Nop56/58 and Fibrillarin Associate Through Protein: Protein Interactions And Can Bind the C'/D' Motif in the Absence of L7 .....	74
The Terminal Box C/D RNP is the Minimal Methylation Complex but Efficient Methylation Requires that the Box C/D and C'/D' RNPs are Juxtaposed in the Full-length RNA .....	76
Mutations in Conserved Box Elements Impair Methylation from Both Guide Regions .....	78
Archaeal L7 Binds Cooperatively to the Box C/D and C'/D' Motifs .....	79
Archaeal L7 and Eukaryotic 15.5kD Core Protein Homologs Exhibit Different RNA-Recognition Specificities for Box C/D and C'/D' RNA Binding .....	80

DISCUSSION .....	82
MATERIALS AND METHODS .....	88
REFERENCES .....	93
CHAPTER III. Evolutionary Origins of the RNA-Guided Nucleotide Modification	
Complexes: From the Primitive Translational Apparatus? .....	105
ABSTRACT .....	106
Nucleotide Modification Guide-RNAs are Not Limited to Eukarya .....	107
The Guide RNAs Function as Ribonucleoprotein Complexes .....	108
The Kink Turn or "K-Turn" of the Box C/D RNAs is a Widespread and Evolutionarily Conserved RNA Motif .....	109
Archaeal Box C/D and H/ACA sRNAs Share a Common RNP Core .....	110
A Common Ancestral RNP Complex Derived from the Primitive Translation Apparatus .....	111
Continuing Evolution of the RNA-Guided Nucleotide Modification Complexes .....	113
Concluding Remarks .....	115
BIBLIOGRAPHY .....	117
CONCLUDING REMARKS .....	126
Remaining Questions .....	127
REFERENCES .....	129
APPENDIX: Scientific Meeting Abstracts .....	130

## LIST OF TABLES

### CHAPTER I.

Table I. Relative binding affinity of L7 protein for selected box C/D

snoRNA mutants ..... 65

## LIST OF FIGURES

### LITERATURE REVIEW

Figure 1. Schematic representation of the rRNA primary transcript .....	1
Figure 2. Processing pathways (major and minor) for intronic snoRNAs .....	5
Figure 3. Secondary structure of the eukaryotic box H/ACA snoRNA .....	7
Figure 4. Secondary structure of the eukaryotic box C/D snoRNA .....	8
Figure 5. Alignment of <i>Methanocaldococcus jannaschii</i> ribosomal protein L7 with human Nhp2p and 15.5kD .....	16
Figure 6. Secondary structure of the K-turn consensus sequence and the tertiary structure of the 15.5kD-binding site in U4 .....	17
Figure 7. Comparison of the box C/D core motif with the ribosomal binding site of L7 in the ancestral 23S rRNA illustrates that L7 recognizes variants of the K-turn structure .....	18
Figure 8. Crystal structure of Nop56/58p: Fibrillarin heterodimer .....	22
Figure 9. Model for evolution of the methylation guide RNAs from <i>cis</i> -acting elements in the primitive translation apparatus .....	24
Figure 10. Model for the evolution of the symmetric, archaeal box C/D RNP versus the asymmetric, eukaryotic box C/D RNP .....	25

### CHAPTER I

Figure 1. Archaeal ribosomal protein L7 is a homolog of eukaryotic 15.5kD snoRNP protein .....	57
Figure 2. Archaeal ribosomal protein L7 binds the box C/D snoRNA core motif .....	58
Figure 3. Folded structure of the box C/D snoRNA core motif .....	59

Figure 4. The box C/D core motif exhibits a highly ordered RNA structure .....	60-61
Figure 5. L7 requires specific sequence and structural elements of the box C/D core motif for RNA: protein interaction .....	62
Figure 6. Comparison of the U4 snRNA 5' stem loop and the KT-15 K-turn of 23S rRNA .....	63
Figure 7. Comparison of human 15.5kD, <i>H. marismortui</i> L7Ae and <i>M. jannaschii</i> L7 protein .....	64

## CHAPTER II

Figure 1. Archaeal sR8 sRNP assembly requires a defined order of core protein addition and forms symmetric RNP complexes on the terminal box C/D core and internal C'/D' motifs .....	97
Figure 2. Nop56/58 and Fibrillarin interact via protein: protein interactions and can bind the internal C'/D' motif in the absence of core protein L7 .....	98
Figure 3. <i>in vitro</i> assembled sR8 sRNP guides site-specific 2'-O-methylation from both terminal box C/D and internal C'/D' RNPs: Efficient methylation requires juxtapositioning of both RNP complexes in the full-length sRNA .....	99-100
Figure 4. Mutation of either the terminal box C/D core or internal C'/D' motifs in full-length sR8 affects guided 2'-O-methylation of both RNP complexes .....	101
Figure 5. Core protein L7 exhibits cooperative binding to archaeal sR8 sRNA .....	102
Figure 6. Archaeal L7 binds both the C/D and C'/D' motifs: Eukaryotic 15.5kD protein binds only the terminal box C/D core motif .....	103
Figure 7. Evolution of the box C/D RNP core proteins: Archaeal and eukaryotic box C/D RNP complexes exhibit symmetric versus asymmetric distribution of the RNP core proteins bound to sRNAs and snoRNAs, respectively .....	104

### CHAPTER III

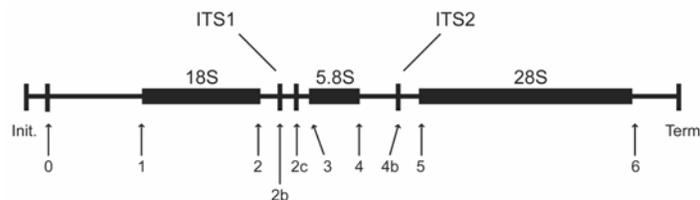
Box I. Eukaryotic and archaeal RNA-guided nucleotide modification complexes .....	121-122
Figure 1. The L7 protein family .....	123
Figure 2. The kink-turn (K-turn): An evolutionarily conserved RNA motif .....	124
Figure 3. Evolution of the RNA-guided nucleotide modification complexes .....	125

## LITERATURE REVIEW

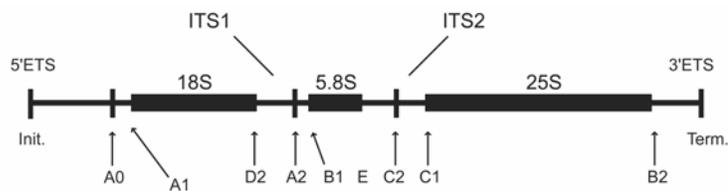
Ribosomes are ubiquitous translational machines that are essential for cell growth and maintenance. Ribosome biogenesis is an active process involving the transcription of precursor RNA molecules, cleavage of the pre-ribosomal RNA (pre-rRNA), modification of nucleotides, and assembly with ribosomal proteins [1-3]. Active cell growth requires maintaining ribosome numbers at an operational level and thus requiring a large input of cellular resources for any given cell. For example, an actively growing human cell may require up to 1 million ribosomes, synthesizing approximately 7500 subunits per minute [4]. Synthesis of such large numbers of ribosomes requires that the steps in ribosome biogenesis be highly orchestrated.

### I. Eukaryotic Ribosome Biosynthesis

A.



B.



**Figure 1. Schematic representation of the rRNA primary transcript (adapted from [3]).** A) Mouse rRNA primary transcript contains the 18S, 5.8S and 28S precursor molecules. Numbers indicate specific pre-rRNA cleavage sites. B) The yeast primary transcript contains the 18S, 5.8S, and 25S precursor molecules. Letters denote cleavage sites for the precursor rRNA.

Eukaryotic ribosome biogenesis begins with transcription by RNA Polymerase I (RNA Pol I) of the 47S precursor in metazoa or 35S in yeast (see Figure 1A or B, respectively, see [3] for review). The precursor transcript contains the 18S, 5.8S and 28S (25S in yeast) rRNAs as well as intergenic (ITS) and extragenic (ETS) transcribed spacer regions (see Figure I). (5S rRNA is also included in the mature ribosome (60S subunit) but is transcribed by RNA polymerase III outside of nucleolus.) The nascent precursor rRNA (pre-rRNA) must undergo a variety of post-transcriptional processing events including cleavage, folding, and nucleotide modification. The vast majority of the nucleotide modifications are of two types, sugar methylation at the 2'-O-ribose position and pseudouridylation of uridine residues. These modifications are numerous with upwards of 200 (100 of each type) present in mature, human rRNAs [5, 6].

#### **A. Nucleotide Modifications in rRNA**

Currently, the function of these rRNA nucleotide modifications is unknown. It is commonly believed, however, that the cumulative effect of these modifications influences ribosome biogenesis and/or function. The fact that these nucleotide modifications are important stems from several lines of evidence. First, modifications do not occur in the ITS or ETS regions of the pre-rRNA [3], regions that are not retained in the mature ribosome RNA. Second, recent positioning of modification sites on modeled yeast ribosomal subunits revealed that these modifications are clustered in functionally important regions, including the peptidyl transferase center and the ribosomal subunit interface [7]. Additionally, rRNA regions involved in ribosomal protein binding are virtually devoid of modifications.

Ribose methylation increases base stacking and pseudouridylation enhances base rotational freedom around the glycosyl bond, which can stabilize secondary and tertiary structures [8, 9]. Curiously, experiments have demonstrated that all but one modification site found in the peptidyl transferase

center are dispensable for function [10]. The collective effect of nucleotide modifications could contribute significantly to the overall structure of the folded rRNA particle.

## **B. Ribosome Biosynthesis**

Ribosome biogenesis in eukaryotes occurs in a dense, fibrillar region within the nucleus called the nucleolus. The nucleolus was first reported in 1898 and was one of the first organelles identified by microscopists [11]. However, the function of the nucleolus remained elusive until the 1960's, when two different research groups identified 18S and 28S ribosomal DNA (rDNA) transcriptional units located within this subcellular compartment [12, 13]. It is now known that the nucleolus is the site of ribosome biogenesis and is often referred to as the "ribosome factory". The nucleolus is organized around 50-1000 actively transcribed rDNA repeats of the primary transcript gene. Since this organelle is not membrane-bound, it is this high level of transcription that is believed to give the nucleolus its dense appearance in the electron microscope. It is also believed that the nucleolus is formed by a high concentration of rRNA processing factors transiently associated with multiple sites of activity [14].

## **II. Small Nucleolar RNAs**

The small nucleolar RNAs comprise a large class of stable cellular RNAs that function in various aspects of ribosome biogenesis. Approximately 100 species of snoRNAs have been identified to date, but estimates indicate that the full cellular complement numbers around 150 [15, 16]. SnoRNAs are typically between 75 and 200 nucleotides in length, although, yeast snoRNAs are often considerably longer due to a lack of 5' end processing (see section IIB). One such example is yeast snR30, which is approximately 600 nucleotides [17]. Some snoRNAs function as pre-rRNA chaperones, base-pairing to the pre-rRNA and either preventing rRNA misfolding or facilitating proper folding of the rRNA molecule. Some snoRNAs also act in pre-rRNA cleavage events and depletion of these specific

snoRNA species results in accumulation of rRNA precursors. However, the vast majority of snoRNAs guide rRNA nucleotide modifications using rRNA-complementary regions within the snoRNA molecule.

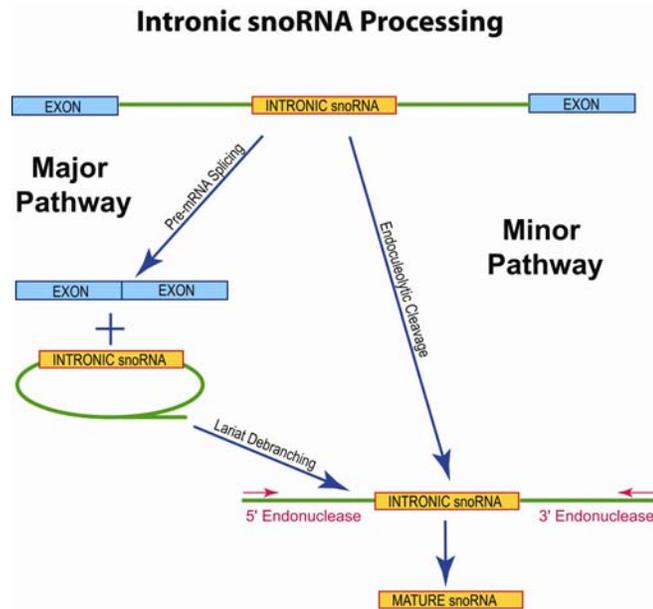
### **A. A Brief History of SnoRNAs**

In 1966, Busch and coworkers published a study of small nuclear and nucleolar RNAs [18]. They observed a fraction of nucleolar RNA sedimenting between the 4 and 6S range in a sucrose density gradient that had not been previously identified. Base composition analysis demonstrated that this low molecular weight RNA had a higher uridine content (U) with respect to the higher molecular weight RNA fraction (i.e. rRNA). Two years later, this group published a report that further analyzed this small nucleolar RNA fraction and revealed that it did not exhibit amino acid acceptor activity (i.e. not transfer (t) RNA) and had low template activity (i.e. not messenger (m) RNA). They concluded that this low molecular weight fraction contained unique small RNAs, which were later termed small nucleolar or snoRNAs [19]. Individual species of snoRNAs are currently named U3, U14, U22, etc. based on the original base composition analysis and the order in which they were discovered.

### **B. SnoRNA Processing in Eukaryotes**

SnoRNAs exhibit several genome organizations depending on the eukaryotic organism. Most snoRNAs in yeast are monocistronic but some (and most in plants) are polycistronic. In metazoa, a few snoRNAs are individually transcribed (U3, U8, U13), but the vast majority are encoded in introns of protein-coding genes ("intronic snoRNAs") and transcribed by RNA polymerase II (Pol II) [20]. U14 was the first snoRNA discovered to be intron-encoded and, in mouse, resides within the coding region of the heat shock cognate 70 (hsc70) chaperone protein [21-23].

It is now evident that Pol II transcripts containing intronic snoRNAs ("snoRNA host transcripts") preferentially code for proteins involved in various aspects of ribosome biosynthesis and translation, including ribosomal proteins and translation factors [20]. This organization may reflect a regulatory mechanism important for maintenance of ribosome numbers at operational levels. There are, however, examples of snoRNA host transcripts whose exonic sequences do not encode protein products and, instead, act only as snoRNA expression vehicles. These include the *gas5* gene, which encodes ten different snoRNAs in ten different introns [24]. Four such non-protein coding snoRNA host genes have been identified to date [24-27]. Interestingly, these genes are members of the 5'-terminal oligopyrimidine (5' TOP) gene family [24, 27]. The 5'-TOP family is associated with gene products that are translationally regulated by changes in growth conditions [28]. Since this type of snoRNA host transcript is not translated, it is possible that expression of the 5'-TOP gene family is regulated at the level of transcription [20].



**Figure 2. Processing pathways (major and minor) for intronic snoRNAs.**

There are two biosynthetic pathways for the intronic snoRNAs (Figure 2). The major pathway

involves exon-intron splicing, lariat debranching, and finally exonucleolytic trimming. This pathway is utilized by the vast majority of intronic snoRNAs. The minor pathway is used when snoRNAs are found in poorly spliced introns (e.g. U16 and U18) [29]. This pathway involves endonucleolytic excision of the intronic snoRNA followed by exonucleolytic trimming of the ends.

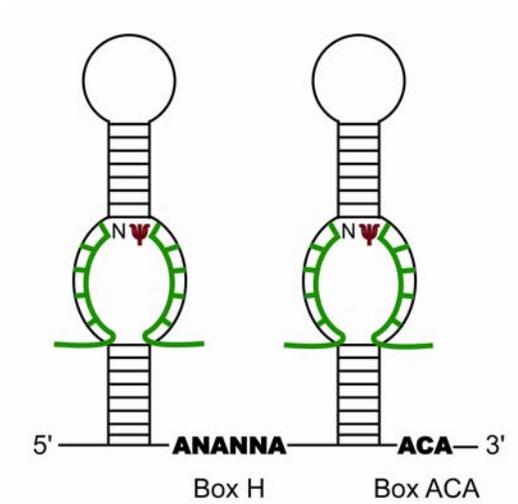
Processing of snoRNAs via either pathway is believed to involve initial packaging of the intronic snoRNA into a ribonucleoprotein complex (see below). Evidence for snoRNP assembly includes the finding that intronic snoRNA processing can be inhibited by addition of excess, unprocessed transcripts indicative of titration of an essential, trans-acting, processing factor(s) [30]. In addition, precursor snoRNAs can be immunoprecipitated with antibodies against snoRNP core proteins [30]. Recent studies have provided evidence that splicing and snoRNA processing are tightly linked. SnoRNAs are preferentially positioned 70 to 80 nucleotides upstream of the 3' splice site in humans and alteration of this spacing is detrimental to snoRNA processing [31, 32]. This conservation of intronic snoRNA spacing may indicate binding of additional, unidentified snoRNA-processing factor(s).

### **III. Classes of SnoRNAs**

There are two major classes of snoRNAs based on conserved sequence elements: the box H/ACA snoRNAs and the box C/D snoRNAs. Both snoRNA classes guide nucleotide modifications through an antisense base-pairing mechanism. Both the box H/ACA and box C/D snoRNAs contain complementary sequences that pair with the target rRNA and designate the specific nucleotide for modification (see [16] for review).

### A. Box H/ACA SnoRNAs

The box H/ACA snoRNAs characteristically fold into a bipartite hairpin-hinge-hairpin tail structure as seen in Figure 3 [33, 34].



**Figure 3. Secondary structure of the eukaryotic box H/ACA snoRNA (reproduced from [35]).** Box elements are indicated in bold while the targeted rRNA molecule (green) is shown base-paired to the snoRNA within the pseudouridylation pocket. The target pseudouridylation nucleotide ( $\Psi$ ) is shown in red.

This snoRNA family is named for conserved box H (sequence ANANNA) and ACA-triplet nucleotide sequences. The majority of box H/ACA snoRNAs guide pseudouridylation and they do this through rRNA antisense regions located within an internal loop in one or both of the hairpin structures [36]. These antisense regions base pair to ribosomal RNA, creating short helices (4-10 base pairs in length) within the loop (see [37] for review). The target uridine lies in an unpaired region at the base of the proximal helix within the hairpin structure. The internal loop region is called the pseudouridylation pocket and modification occurs approximately 15 nucleotides from the box H or ACA sequence within the snoRNA [34]. Only a few members of the box H/ACA family participate in rRNA processing. The best characterized is yeast snR30 (or U17 in humans) which functions in rRNA precursor cleavage events. SnR30 is an essential snoRNA and its depletion results in loss of 18S rRNA accumulation and



nucleotides in length and are located immediately upstream of boxes D and D' [42, 25]. Nucleotide modification occurs at a fixed position within the target RNA paired five nucleotides upstream of box D and/or D' [43]. Unlike box H/ACA snoRNAs, there is no base-specificity for box C/D-guided modification: any ribonucleotide can be modified.

A few box C/D snoRNAs function in pre-rRNA processing. U3, U14, and U22 are required for cleavage events necessary for 18S rRNA production whereas U8 is required for 5.8S and 28S processing [3, 44, 37, 20]. Since no ribonucleases have been identified in association with these snoRNAs, it is anticipated that processing is facilitated by these snoRNAs acting as RNA chaperones to assist pre-rRNA folding. Several lines of evidence support this hypothesis. First, U3 acts by bridging two regions of rRNA that are not adjacent in the precursor transcript, establishing a pre-rRNA structure that is essential for ribosomal RNA processing [45, 46]. Second, U106 is a suspected chaperone because it contains two antisense rRNA elements, both of which are complementary to rRNA regions devoid of modifications [15]. Third, the guide sequence upstream of box D' in U14 base-pairs to a region of 18S rRNA lacking ribose-methylated nucleotides. Experiments have demonstrated that the U14: rRNA helix formed is required for pre-rRNA cleavage and subsequent 18S rRNA production [47].

### **C. MRP**

MRP (mitochondrial RNA-processing) is a unique snoRNA species because it is not a member of either of the two major snoRNA classes. MRP is required for cleavage of pre-rRNA at an ITS I site upstream of the 5.8S sequence [48]. Interestingly, the RNA component of the MRP snoRNP complex resembles that of the ribozyme RNase P, which cleaves the 5' end of tRNA [48]. This suggests that the MRP snoRNA is also a ribozyme, although that has not been demonstrated directly.

#### **IV. SnoRNA-Protein Complexes**

The two major classes of snoRNAs (box H/ACA and box C/D) form ribonucleoprotein complexes (RNPs) by associating with a set of core proteins. Whereas the antisense regions within the snoRNAs select the site for nucleotide modification, the associated proteins provide the necessary enzymatic activities.

##### **A. Box H/ACA Core Proteins**

Box H/ACA snoRNAs associate with four core proteins that are common to all members of this snoRNA class. These core proteins are Gar1p, Nhp2p, Nop10p, and Cbf5p [49, 51]. Cbf5p (Dyskerin in humans or Nap57p in rat) is the pseudouridine synthase. Cbf5p shows significant homology to other known pseudouridine synthases and Cbf5p mutant strains in yeast abolish pseudouridylation of rRNA [6, 52]. The remaining three proteins are believed to perform structural roles in the H/ACA snoRNP. Gar1p is an essential component of the box H/ACA snoRNAs and its depletion in yeast results in pre-rRNA processing defects [53]. Gar1p contains glycine and arginine rich N-terminal domain and has been found to bind the H/ACA snoRNA directly [54, 55]. In fact, all four core proteins may contact the RNA because each protein can be UV-crosslinked to a box H/ACA snoRNA *in vivo* [56].

##### **B. Box C/D Core Proteins**

Box C/D snoRNAs also form ribonucleoprotein complexes by assembling with a set of four core proteins. These core proteins are Fibrillarin, 15.5kD, Nop56p, and Nop58p. Fibrillarin is the methylase enzyme, utilizing S-adenosyl methionine (SAM) as the methyl donor [57-59]. Interestingly, like Gar1p, Fibrillarin is also a member of the RGG-motif family indicating that it may also bind the snoRNA directly [54]. Fibrillarin has been UV-crosslinked to a model box C/D snoRNA *in vivo*,

supporting RNA-binding activity for this core protein [60]. Experiments have demonstrated that the association of Fibrillarin with the box C/D snoRNA is dependent on binding of other core proteins to the snoRNA, indicating that Fibrillarin is only weakly associated with the RNA. However, at least for one snoRNA species, Fibrillarin associates with the RNA in the absence of other core proteins [61].

The snoRNA core protein, 15.5kD, binds directly to the box C/D motif and this binding is believed to function as the nucleation event for snoRNP assembly. 15.5kD was originally identified as a protein that associates with the U4 spliceosomal small nuclear RNA (snRNA) [62]. Subsequent experiments identified 15.5kD as a dual function protein because it also associates with the box C/D snoRNAs [40].

The co-crystal structure of 15.5kD bound to its binding site in U4 snRNA revealed a unique RNA fold resulting from an asymmetric bulge flanked on either side by helices [63]. This secondary structure has been applied to the box C/D motif by analogy [40]. Strikingly, analysis of RNA-folding motifs in the *H. marismortui* 50S ribosomal subunit identified a new RNA motif, termed the kink-turn (K-turn) [64]. Both 15.5kD-binding sites in U4 snRNA and the box C/D snoRNAs exhibit classical features of this motif and are, therefore, members of this newly identified K-turn family (see section VIIB).

Interestingly, the box H/ACA snoRNA core protein, Nhp2p, is similar to 15.5kD in sequence, exhibiting 31% identity or 61% similarity in yeast. However, no kink-turn secondary structure has been identified in the eukaryotic box H/ACA snoRNAs.

The Nop56p and Nop58p core proteins constitute a pair of related proteins, exhibiting 45% sequence identity in yeast [65]. Like the box H/ACA core proteins Nhp2p and Nop10p, Nop56p and Nop58p are predicted to be structural components of the box C/D snoRNP. These proteins were originally identified in synthetic lethal screens using temperature-sensitive Fibrillarin mutants [65]. Interestingly, whereas Nop58p and Fibrillarin associate independently with box C/D snoRNAs, Nop56p is bound

only in the presence of Fibrillarin [66, 67]. Additionally, depletion of either Nop58p or Fibrillarin (but not Nop56p) leads to co-depletion of box C/D snoRNAs [67]. In 2002, Steitz and coworkers utilized *in vivo* crosslinking to reveal a differential distribution of Nop56p and Nop58p on a box C/D snoRNA [60]. This biochemical data supported prior genetic analyses indicating different roles for these two proteins. More specifically, this study revealed that Nop56p binds to the internal box C'/D' motif whereas Nop58p binds the terminal box C/D core motif [60]. In another report from the same laboratory, nucleotide interference analog mapping demonstrated that 15.5kD binds only to the terminal C/D motif and not the internal C'/D' motif [68]. Thus, the methylase Fibrillarin is the only core protein bound to both RNA motifs.

#### **V. Archaea: A Separate Domain of Life from Eukaryotes and Bacteria**

In 2000, homologs of eukaryotic snoRNAs (termed snoRNA-like or sRNAs) were discovered in Archaea [69, 70]. This finding indicated that RNA-guided nucleotide modification is an evolutionarily ancient mechanism. Archaea are prokaryotic organisms, which constitute a domain of life distinct from both Eukarya and Bacteria [71]. Archaeal metabolism and gene organization resembles Bacteria, but archaeal replication, transcription and translation machinery more closely resembles eukaryotes [72].

Many archaeal species live in extreme environments including low pH and high temperature. In 1998, Noon and coworkers demonstrated that the level of rRNA 2'-O-methylation increases proportionally with increased growth temperature [73]. This suggests that nucleotide modification in the ribosome may help stabilize thermolabile RNA at high temperatures. Therefore, rRNA modification in Archaea may be an inherent necessity for habitat adaptation of these unicellular organisms.

Archaeal sRNAs exhibit a genomic organization distinct from eukaryotes. DNA sequences encoding sRNAs are found on both strands and are dispersed around the entire circular chromosome [74]. Most are found in intergenic regions between protein-coding genes whereas some overlap the adjacent open reading frame. The mechanism for transcription and processing of sRNAs in Archaea remains unknown. Currently, there is only one example of an sRNA encoded within an intron. SR3 (*Archeoglobus fulgidus*) or sR50 (*Pyrococcus abyssi*) is found in the intron of its target substrate, precursor-tRNA<sup>Trp</sup> (pre-tRNA<sup>Trp</sup>) [75]. The mechanism by which this sRNA guides modification was predicted to occur in *cis* via an intramolecular modification reaction [75, 76]. In this model, intron-encoded sR3 guides two, tRNA nucleotide modifications before its removal from the pre-tRNA by splicing. New results, however, now indicate that these modifications occur in *trans* via an intermolecular reaction (R. Gupta, personal comm.). The details of this mechanism are still under investigation.

## **VI. Archaeal sRNAs and Core Proteins**

The nucleotide-modification guide RNAs of Archaea exhibit similar consensus box elements and secondary structures as their eukaryotic counterparts. Not surprisingly, homologs to the eukaryotic snoRNA core proteins are also present in Archaea. Currently, the only characterized function of Box C/D RNPs in Archaea is 2'-O-methylation. Possible roles in pre-rRNA folding and/or cleavage remain to be demonstrated.

### **A. Archaeal Box H/ACA sRNPs**

Although both major classes of modification-guide RNAs are present in Archaea, the H/ACA sRNAs have only recently been discovered and this sRNA class is the least characterized [77]. The presence of H/ACA RNAs in Archaea was contested before their identification because the level of rRNA

pseudouridylation in Archaea is low. Pseudouridylation levels in Archaea more closely resemble those in *Escherichia coli*, where this type of nucleotide modification is catalyzed by protein-only enzymes [74]; [9]. However, homologs of the core proteins Cbf5p, Gar1p, and Nop10p, had been detected in archaeal genomes [78, 79]. In 2002, this debate ended with the discovery of four H/ACA sRNAs in *Archeoglobus fulgidus* [77]. The box H/ACA sRNAs of Archaea exhibit slightly different structures and sequence elements than those of eukaryotes. Archaeal H/ACA sRNAs have either one or three hairpin structures and the box H sequence is sometimes absent. Notably, this is also the case for the box H/ACA snoRNAs of two early branching eukaryotes, *Trypanosoma brucei* and *Euglena gracialis* [80, 81]. This suggests that the single hairpin H/ACA RNA represents an early evolutionary form of the box H/ACA snoRNAs found in vertebrates and yeast.

## **B. Archaeal Box C/D sRNPs**

The box C/D sRNAs were the first class of guide RNAs discovered in Archaea [69]; [70]. Unlike the H/ACA sRNAs, the presence of box C/D snoRNA homologs was anticipated due, in part, to the high level of 2'-O-methylation in archaeal rRNA sequences. Additionally, homologs to core proteins Fibrillarin and Nop56p/Nop58p had been detected [82, 83]. Interestingly, Archaea contain a single homolog to both eukaryotic Nop56p and Nop58p (termed Nop56/58p or Nop5p). This is consistent with the prediction that the two related proteins in eukaryotes arose from a gene duplication event [74]. With knowledge of these core protein homologs, Omer and coworkers identified the first archaeal box C/D sRNAs from *Sulfolobus solfataricus* by co-immunoprecipitating these RNAs from cell extracts using antibodies against archaeal Fibrillarin and Nop56/58p [59]. Subsequent development of a computer algorithm to detect additional archaeal box C/D sRNAs has resulted in the detection of many species. Interestingly, box C/D sRNAs are found in at least two branches of the archaeal domain, Euryarcheota and Crenarcheota, supporting the hypothesis that the nucleotide modification complexes arose before the evolutionary split between eukaryotes and Archaea (see section X).

The box C/D sRNAs exhibit similar sequences and structures as their eukaryotic snoRNA counterparts with a few exceptions. First, archaeal box C/D sRNAs are slightly smaller with an average size of 75 nucleotides (as compared to 150 nucleotides in eukaryotes). Additionally, the internal C'/D' motif is typically a perfect match to the consensus box C and D sequences; Eukaryotic C'/D' sequences are frequently degenerate. Finally, most box C/D sRNAs in Archaea exhibit "dual guide" function: both the D and D' guide regions contain sequences antisense to targeted RNA molecules [74]. In contrast, only 20% of box C/D snoRNAs have been identified as "dual guides".

## **VII. Identification of the Archaeal Homolog of Eukaryotic 15.5kD**

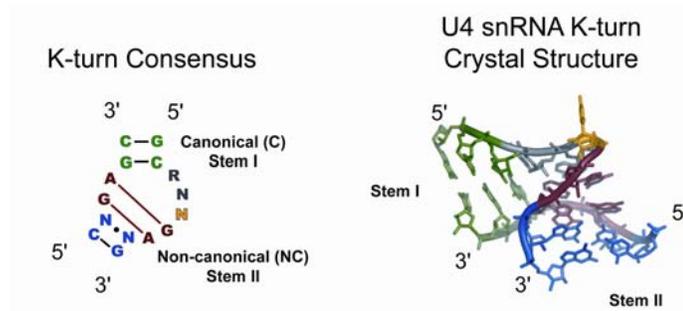
To characterize the assembly and function of the archaeal box C/D sRNPs, it was necessary to identify the full complement of core proteins. Whereas homologs to both Fibrillarin and Nop56p and Nop58p were evident, no clear homolog to 15.5kD had been identified. Because 15.5kD is the first protein to bind and recognize the C/D core motif, it was unlikely that archaeal box C/D sRNP assembly did not require a comparable core protein homolog.

### **A. Identification of Archaeal Ribosomal Protein L7**

In 2002, we began our investigation of the archaeal box C/D sRNP by identifying ribosomal protein L7 as a putative homolog to 15.5kD ([84]; see Chapter I). Sequence comparison of *Methanocaldococcus jannaschii* (*M. jannaschii*) L7 with human 15.5kD revealed that these protein sequences are 33% identical and 60% similar (Figure 5).



into a family termed the kink-turn (or K-turn) motif [64]. This motif is characterized by a helix-bulge-helix structure with a sharp bend between the two helical axes (Figure 6).

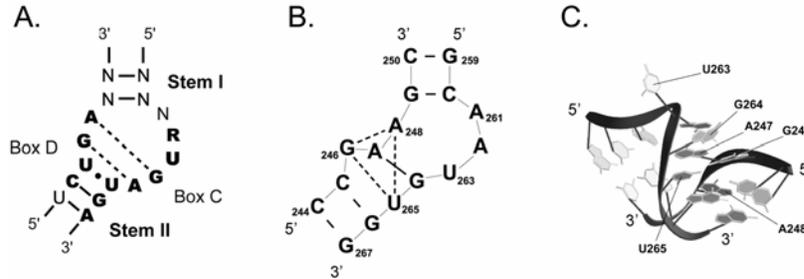


**Figure 6. Secondary structure of the K-turn consensus sequence (left; [64]) and the tertiary structure of the 15.5kD-binding site in U4 (right; [63]).** Stem structures I and II are coordinately colored in green and blue, respectively. Tandem, sheared G•A base pairs appear in red, and the protruded nucleotide of the asymmetric bulge is shown in yellow.

This bend is facilitated by two adjacent, sheared G•A base pairs, which reside at the base of the terminal bulge. This structure was first identified in the L30-binding site of the L30 pre-mRNA [85]. Ribosomal protein L30 in eukaryotes regulates its own expression at the level of splicing by associating with its pre-mRNA. The K-turn structure is also evident in the 15.5kD-binding site in U4 snRNA [63]. In the U4 kink-turn, the RNA is bent at approximately 120 degrees and this bend results in protrusion of a uridine nucleotide that lies in the RNA-binding pocket of 15.5kD. Logic predicted that the 15.5kD-binding site in the box C/D terminal core motif should also be K-turn, although this had not yet been demonstrated directly by detailed structural analysis [40]. Consistently, L7, the archaeal homolog of eukaryotic 15.5kD, binds a K-turn (see below).

### C. Characterization of L7 Association with the Box C/D Core Motif

The archaeal box C/D RNP core protein, L7, is a ribosomal protein that binds a K-turn (KT-15) in the archaeal large ribosomal subunit (Figure 7B and C; [86, 64]).



**Figure 7. Comparison of the box C/D core motif with the ribosomal binding site of L7 in the archaeal 23S rRNA illustrates that L7 recognizes variants of the K-turn structure.** The box C/D consensus secondary structure (A) and the secondary (B) and tertiary (C) structure of the L7-binding site in the 23S rRNA of the 50S subunit, KT-15 [64].

KT-15 is a unique structural variant of the K-turn because it incorporates a base-triple in place of one of the G•A pairs. This indicated that structure may play a larger role in L7 binding than sequence.

To further characterize L7 binding to the box C/D core motif, point mutations in the minimal motif were utilized in a series of binding studies (see Chapter I). Our results confirmed the hypothesis that structure was a critical factor for high affinity binding [84]. Mutations predicted to disrupt the helix-bulge-helix secondary structure, e.g. mutation of the critical GA nucleotides, either reduced or abolished L7 binding. Conversely, mutations that maintained secondary structure had minimal effect on L7 binding. The exception is the protruded uridine whose identity should have no bearing on structure since this residue makes no interactions with other portions of the RNA. Mutation of this nucleotide to a purine greatly impaired L7 association with the box C/D core motif, whereas a guanine nucleotide at this position was

tolerated. Coincidentally, Klein and coworkers noted that the protruded uridine in KT-15 is found deep within the binding pocket of L7 and a purine base ring might be too large to be accommodated in the RNA binding site of this protein [64].

Comparison of the crystal structures of archaeal ribosomal protein L7 and eukaryotic 15.5kD has revealed that the tertiary structures of these two proteins are virtually superimposable [84]. This conservation of structure is striking given that the L7 and 15.5kD proteins share only 30% sequence identity (see Figure 5). Our results are consistent with the identification of L7 and 15.5kD as functional homologs ([84]; see Chapter I).

The co-crystal structure of L7 bound to the C/D motif has confirmed that this RNA motif is indeed a K-turn ([87]; Appendix II). In fact, the L7:C/D RNA complex exhibits high structural similarity to that of the U4:15.5kD co-crystal structure [63]. This finding was anticipated due to the structural similarities of the two proteins. Interestingly, our UV-melting profiles revealed conformational changes in the RNA following L7 binding, indicative of an "induced-fit" mechanism. Interestingly, an "induced-fit" binding mechanism was also predicted for ribosomal protein L30 binding to its pre-mRNA and may be the common mode of protein association with K-turn motifs [88]. Consistent with the idea that the K-turn is not a rigid, preformed structure, Lilley and colleagues demonstrated that the K-turn motif is highly dynamic and requires other factors, such as proteins, to "lock" the K-turn motif into its characteristically "bent" conformation [89].

### **VIII. Archaeal Box C/D sRNP Assembly**

In 2002, Omer and colleagues established an *in vitro* assembly system to study archaeal box C/D sRNP-guided methylation [59]. They demonstrated that addition of the three core proteins from

*Sulfolobus acidocaldarius*, L7, Fibrillarin, and Nop56/58p (Nop56a), to a box C/D RNA from *S. solfataricus* resulted in stable formation of a box C/D RNP *in vitro*.

#### **A. Analysis of Ribosomal Protein L7 Binding to a Full-length Archaeal Box C/D RNA**

Following the work Omer and coworkers, we developed an *in vitro* assembly system using *Methanocaldococcus jannaschii* core proteins and a box C/D RNA to analyze the structure and function of the archaeal box C/D sRNP. Although Omer and coworkers demonstrated that ribosomal protein L7 is assembled in the box C/D RNP, it was not clear which motifs this protein bound. Using full-length, archaeal sR8 box C/D sRNA as well as minimal box C/D or C'/D' RNAs, we determined that L7 associates with both motifs ([90]; see Chapter II). Additionally, we demonstrated that L7 binds with positive cooperativity to the two binding sites in the full-length sRNA. RNA footprinting analysis has now revealed that L7 initially binds to either the box C/D or the C'/D' motif, indicating that the binding is not sequential (Tran et al., unpublished results). These results were surprising in light of recent work reporting that the eukaryotic homolog, 15.5kD, binds the eukaryotic terminal box C/D core but not the internal C'/D' motif [68]. Our comparison of L7 and 15.5kD binding to both a full-length eukaryotic and archaeal box C/D RNA revealed that the difference in distribution of the two core protein homologs on the box C/D and C'/D' motifs is due to divergence in the RNA-recognition capabilities of L7 and 15.5kD ([90]; see Chapter II).

#### **B. Identification of Protein: Protein Interactions Between Nop56/58p and Fibrillarin**

To characterize further assembly of the sRNP complex, we utilized a series of "pull-down" assays to study protein: protein interactions between the archaeal box C/D core proteins. These experiments revealed that Nop56/58p and Fibrillarin associate through protein: protein interactions ([90]; see Chapter II), indicating that these two proteins may associate with the RNA as a protein complex. This

suggestion is consistent with our observation that Nop56/58p binding to an L7: RNA complex is greatly enhanced with the addition of Fibrillarin. Additionally, the Nop56/58p-Fibrillarin protein complex can associate specifically with the C'/D' motif in the absence of L7, albeit inefficiently. This result indicated that the Nop56/58p-Fibrillarin "dimer" contacts the RNA directly and that the C/D and C'/D' motifs are structurally distinct.

### **C. Demonstration of Archaeal Box C/D sRNP Symmetry**

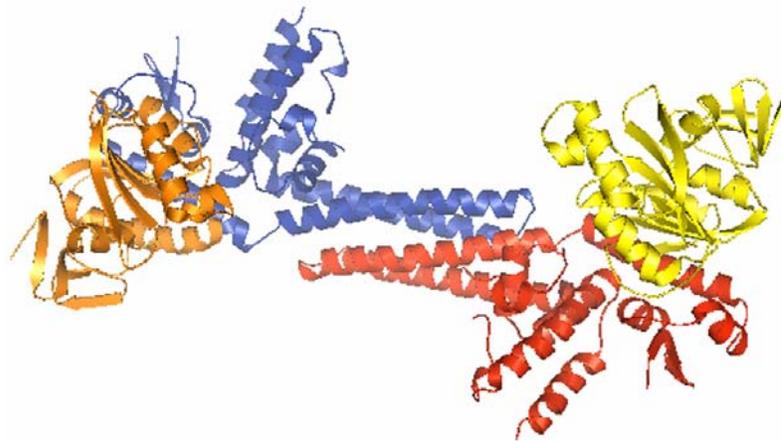
Both electrophoretic mobility-shift and co-purification analyses revealed that all three core proteins (L7, Nop56/58p, and Fibrillarin) bind both the terminal box C/D and internal C'/D' motifs. This work, therefore, established the archaeal box C/D sRNP as a "symmetric" particle with respect to core protein distribution ([90]; see Chapter II). This sharply contrasts with the eukaryotic box C/D snoRNP, where differential binding of the core proteins to the box C/D and C'/D' motifs is "asymmetric". Eukaryotic 15.5kD protein is uniquely bound to the terminal box C/D core motif [68]. Nop58p and Nop56p bind to the terminal C/D and internal C'/D' motifs, respectively [60]. Only Fibrillarin is found associated with both motifs in the eukaryotic snoRNP. From these contrasting RNP structures, we have proposed a pathway for the evolution of the box C/D RNP complexes ([90]; see section X).

### **IX. RNP-guided Methylation Activity of the *In Vitro* Reconstituted Archaeal Box C/D sRNP**

Our reconstituted box C/D sRNP guides site-specific methylation from both the box C/D and C'/D' motif, using both antisense regions to select their respective target RNAs for modification ([90]; see Chapter II). Surprisingly, although both the C/D and C'/D' motifs bind all three of the core proteins, both "RNPs" must be juxtaposed in the same sRNP complex for efficient guided-methylation activity. This requirement was revealed using two different experimental approaches. First, methylation activity of the box C/D "halfmer" RNP was only 30% of the wild-type, full-length sR8 RNP activity,

whereas, the box C'/D' "halfmer" RNP was inactive. This requirement for juxtaposed RNPs in the same sRNP particle was confirmed using full-length sR8 molecules mutated in either the box C/D or C'/D' motifs. Strikingly, mutations in either motif affected methylation not only of the mutated motif, but of the non-mutated motif as well. These results clearly demonstrated that efficient guided methylation requires box C/D and C'/D' RNPs juxtaposed on the same box C/D sRNP particle. From these observations, we have concluded that "crosstalk" interactions between the box C/D and C'/D' motifs are important for guided-nucleotide modification activity ([90]; see Chapter II).

Although the nature of this "crosstalk" is presently unknown, the recent Nop56/58p-Fibrillarin crystal structure from *Archeoglobus fulgidus* has suggested an interesting possibility [91].



**Figure 8. Crystal structure of Nop56/58p-Fibrillarin heterodimer [91].** Nop56/58p molecules are shown in blue or red whereas Fibrillarin molecules are shown in yellow or orange.

This crystal structure confirms the previously reported interaction between Nop56/58p-Fibrillarin [90], but also reveals that this heterodimer potentially homodimerizes through the coiled-coil domain of Nop56/58p (see Figure 8). Aittleb and coworkers proposed that Nop56/58p may dimerize across the

box C/D RNA between the C/D and C'/D' motifs [91]. The importance of this putative protein: protein interaction for "crosstalk" between the box C/D and C'/D' motif in guided methylation remains to be demonstrated.

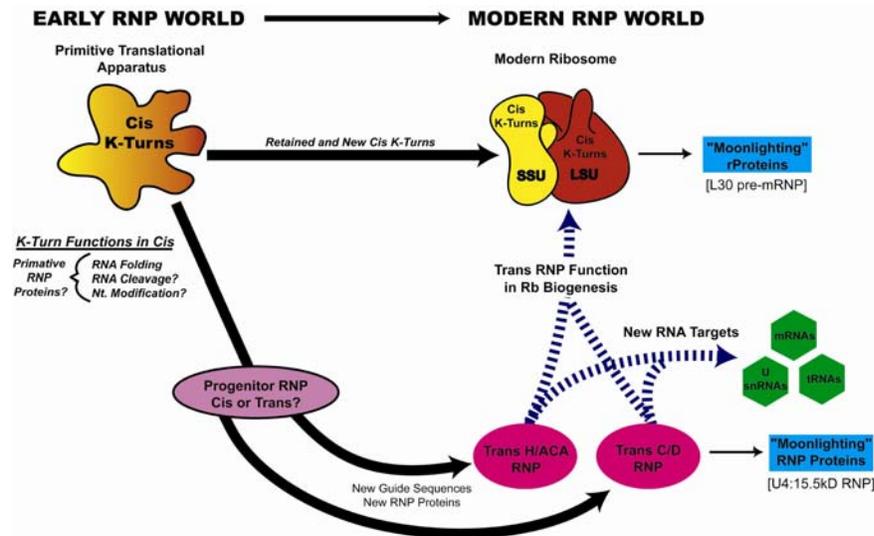
## **X. Evolutionary Origins of the RNA-guided Nucleotide Modification Complexes**

The archaeal and eukaryal domains are estimated to have diverged approximately 2 billion years ago [92], making the nucleotide modification complexes evolutionarily ancient ribonucleoprotein machines. Recently, ribosomal protein L7 was shown to be the archaeal homolog of eukaryotic box H/ACA core protein, Nhp2p ([93]; see Figure 5). Additionally, L7 binds a K-turn motif within the archaeal box H/ACA sRNA and is a core component of the pseudouridylation guide complexes in Archaea. Therefore, archaeal L7 bound to a K-turn motif represents an RNP that is common to both classes of modification guide complexes as well as the ribosome. These observations led us to propose that these modification complexes may have their origins in the primitive translational apparatus.

### **A. Origins of the *Trans*-acting Nucleotide Modification Machines**

Currently, there is no evidence for RNA-guided nucleotide modification in Bacteria. There are no bacterial proteins with significant sequence similarity to archaeal ribosomal protein L7, although there are reported K-turn motifs in Bacteria [94]. While we cannot rule out the presence of undiscovered box C/D and/or H/ACA RNA homologs in Bacteria, it is likely that the modification guide RNAs originated after the branching of Bacteria and Archaea/Eukarya.

Modern day, nucleotide modification-guide RNPs probably originated as K-turn RNP components of the primitive translational apparatus before the division of Archaea and Eukarya domains (Figure 9).

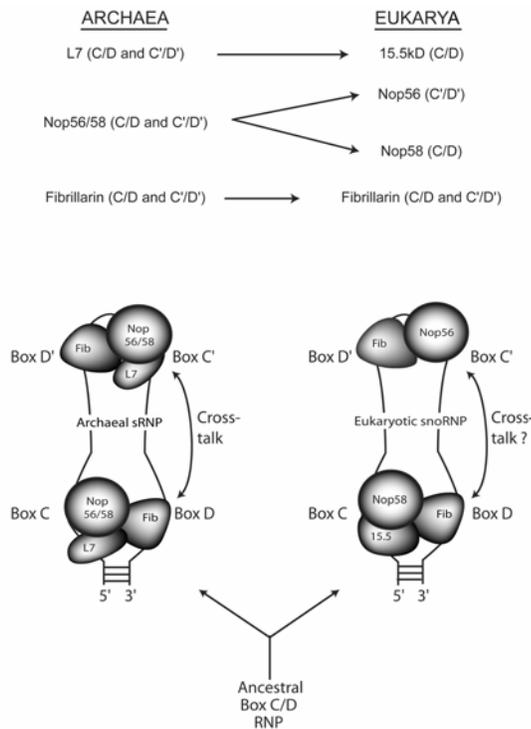


**Figure 9. Model for evolution of the modification guide RNAs from *cis*-acting elements in the primitive translation apparatus (Adapted from [35]).**

K-turn RNPs present within the primordial apparatus may have been retained as necessary folding elements. This is supported by the presence of K-turn RNPs in the modern ribosome. Additionally, some of these *cis*-acting elements may have evolved to guide rRNA nucleotide modification in the ribosome. Evolution of the *trans*-acting modification guide-RNPs probably occurred through a progenitor RNP (common to both the box C/D and H/ACA RNP class) that guided modification as a *cis* or *trans* complex. The box C/D and H/ACA RNPs then evolved from this progenitor complex. It has been suggested that the expansion of the number of guide RNPs and the multitude of RNA targets may have occurred from duplication events (or retrotransposition events) of the box C/D and H/ACA RNA genes followed by variation of the antisense elements [6, 15]. Continued variation of the guide regions has enabled additional guide RNPs to now recognize a variety of non-rRNA targets including tRNAs, mRNAs, and snRNAs.

**B. Model for Evolution of the Asymmetric, Eukaryotic Box C/D SnoRNP from a Symmetric, Ancestral Box C/D RNP**

The demonstration that the archaeal box C/D RNP is symmetric with respect to protein binding contrasts sharply the asymmetric eukaryotic complex [90, 60, 68]. It is likely that the archaeal box C/D RNP more closely resembles that of the last ancestor common to both Archaea and eukaryotes. We have suggested that the evolution of the asymmetric, eukaryotic box C/D RNP is a result of gene duplication and altered RNA-binding capabilities of the core proteins (Figure 10; [90]).



**Figure 10. Model for the evolution of the symmetric, archaeal box C/D RNP (left) versus the asymmetric, eukaryotic box C/D RNP (adapted from [90]).**

Gene duplication of Nop56/58p has resulted two highly similar core proteins in eukaryotes, Nop56p and Nop58p [83, 74]. Nop56p associates with the internal box C'/D' motif, whereas Nop58p binds the

terminal box C/D motif due to differential RNA-binding capabilities [60]. It is possible that Nop56p, in the presence of Fibrillarin, no longer requires an L7-like protein at the C'/D' motif for binding, therefore, eukaryotic 15.5kD protein is found only at the box C/D motif. It will be interesting to determine the core protein distribution on the box C/D RNPs in early branching eukaryotes such as Trypanosomes and Euglena [80, 81]. These box C/D RNPs may represent evolutionary intermediates, more closely resembling the symmetric, archaeal RNP complex. Characterization of snoRNPs from these two species may provide additional information necessary to support this model.

## **XI. Summary of Dissertation Work**

This thesis examines the structure and function of the box C/D sRNP in Archaea. Chapter 1 describes the identification of ribosomal protein L7 as the archaeal homolog of eukaryotic 15.5kD. L7 serves a dual role in Archaea as both a ribosomal protein and a box C/D core protein. In both the ribosome and the box C/D sRNP complex, L7 binds a K-turn motif. Chapter 2 reports development of an *in vitro* assembly and methylation assay for the archaeal box C/D sRNP. The archaeal box C/D sRNP is symmetric with respect to core protein binding which contrasts the asymmetric eukaryotic box C/D snoRNP. Additionally, we reported that efficient methylation requires that both the box C/D and C'/D' motifs be juxtaposed in the full-length box C/D sRNP. Finally, the demonstration of L7 as a component of the ribosome as well as the RNA-modification guide complexes suggested that the modern day sRNPs and snoRNPs may have had their origins as *cis*-acting elements in the primitive translational apparatus. Chapter 3 details a proposed pathway for evolution of the *trans*-acting nucleotide modification machines.

## REFERENCES

1. Gerbi, S., A. Borovjagin, and T. Lange, *The nucleolus: a site of ribonucleoprotein maturation*. *Curr. Opin. Cell Biol.*, 2003. **15**: p. 318-325.
2. Sollner-Webb, B., K.T. Tycowski, and J.A. Steitz, *Ribosomal RNA Processing in Eukaryotes*, in *Ribosomal RNA: Structure, Evolution, Gene Expression and Function in Protein Synthesis*, R.A. Zimmermann and A.E. Dahlberg, Editors. 1996, CRC Press: New York.
3. Eichler, D.C. and N. Craig, *Processing of Eukaryotic Ribosomal RNA*, in *Progress in Nucleic Acids Research and Molecular Biology*, K. Moldave and W. Cohen, Editors. 1994, Academic Press: New York.
4. Lewis, J.D. and D. Tollervey, *Like Attracts Like: Getting RNA Processing Together in the Nucleus*. *Science*, 2000. **288**: p. 1385-1389.
5. Maden, B.E., *The numerous modified nucleotides in eukaryotic ribosomal RNA*. *Prog. Nucleic Acids Res.*, 1990. **39**: p. 241-303.
6. Lafontaine, D. and D. Tollervey, *Birth of the snoRNPs: The evolution of the modification-guide snoRNAs*. *Trends Biochem. Sci.*, 1998. **23**: p. 383-388.
7. Decatur, W. and M.J. Fournier, *rRNA modifications and ribosome function*. *Trends Biochem. Sci.*, 2002. **27**: p. 344-351.
8. Davis, D., *Biophysical and conformational properties of modified nucleotides in RNA (nuclear magnetic resonance studies)*, in *Modification and editing of RNA*, H. Grosjean and R. Benne, Editors. 1998, American society for microbiology: Washington, D.C. p. 85-102.
9. Ofengand, J., *Ribosomal RNA pseudouridines and pseudouridine synthases*. *FEBS Letters*, 2002. **514**: p. 17-25.
10. King, T., et al., *Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center*. *Mol. Cell*, 2003. **11**: p. 425-435.
11. Montgomery, T., *Comparative cytological studies, with especial regard to the morphology of the nucleolus*. *J. Morphol.*, 1898. **15**: p. 265-565.
12. Ritossa, F. and S. Spiegelman, *Localization of DNA complementary to ribosomal RNA in the nucleolus organizer region of Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*, 1965. **53**: p. 737-745.
13. Birnstiel, M. and M. Chipchase, *The nucleolus: pacemaker of the cell*. *Sci. J.*, 1970. **6**: p. 41-48.
14. Lewis, J. and D. Tollervey, *Like attracts like: getting RNA processing together in the nucleus*. *Science*, 2000. **288**: p. 1385-1389.

15. Vitali, P., et al., *Identification of 13 novel human modification guide RNAs*. Nucleic Acids Res., 2003. **31**: p. 6543-6551.
16. Terns, M. and R. Terns, *Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin*. Gen. Express., 2002. **10**: p. 17-39.
17. Atzorn, V., P. Fragapane, and T. Kiss, *UI7/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production*. EMBO J., 2004. **24**: p. 1769-1778.
18. Muramatsu, M., J. Hodnett, and H. Busch, *Base composition of fractions of nuclear and nucleolar ribonucleic acid obtained by sedimentation and chromatography*. J. Biol. Chem., 1966. **241**: p. 1544-1550.
19. Nakamura, T., A. Prestayko, and H. Busch, *Studies on nucleolar 4 to 6S ribonucleic acid of Novikoff hepatoma cells*. J. Biol. Chem., 1968. **243**: p. 1368-1375.
20. Filipowicz, W., et al., *Biogenesis, structure and function of small nucleolar RNAs*. Acta Biochim Pol., 1999. **46**: p. 291-302.
21. Liu, J. and E.S. Maxwell, *Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene*. Nucleic Acids Res., 1990. **18**: p. 6565-6571.
22. Leverette, R., M. Andrews, and E.S. Maxwell, *Mouse U14 snRNA is a processed intron of the cognate hsc70 heat shock pre-messenger RNA*. Cell, 1992. **71**: p. 1215-1221.
23. Xia, L., et al., *Intronic U14 snoRNAs of Xenopus laevis Are Located in Two Different Parent Genes and Can Be Processed From Their Introns During Early Oogenesis*. Nucleic Acids Research, 1995. **23**: p. 4844-4849.
24. Smith, C.M. and J.A. Steitz, *Classification of gas5 as a Multi-Small-Nucleolar-RNA (snoRNA) Host Gene and a Member of the 5'-Terminal Oligopyrimidine Gene Family Reveals Common Features of snoRNA Host Genes*. Molecular and Cellular Biology, 1998. **18**: p. 6897-6909.
25. Tycowski, K., et al., *A Small Nucleolar RNA Requirement for Site-Specific Ribose Methylation of rRNA in Xenopus*. Proc.of the Natl.Acad. Sci. USA, 1996. **93**: p. 1480-1485.
26. Tycowski, K., M. Shu, and J.A. Steitz, *A mammalian gene with introns instead of exons generating stable RNA products*. Nature, 1996. **379**: p. 464-466.
27. Pelczar, P. and W. Filipowicz, *The Host Gene for Intronic UI7 Small Nucleolar RNAs in Mammals Has No Protein-Coding Potential and Is a Member of the 5'-Terminal Oligopyrimidine Gene Family*. Mol. Cell. Biol., 1998. **18**: p. 4509-4518.
28. Amaldi, F. and P. Pierandrei-Amaldi, *TOP genes: a translationally controlled class of genes including those coding for ribosomal proteins*. Prog. Mol. Subcell. Biol., 1997. **18**: p. 1-17.

29. Caffarelli, E., et al., *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**: p. 1121-1131.
30. Watkins, N., et al., *Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D*. RNA, 1996. **2**: p. 118-133.
31. Hirose, T. and J.A. Steitz, *Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells*. Proc. Nat. Acad. Sci. USA, 2001. **98**: p. 12914-12919.
32. Hirose, T., M.-D. Shu, and J. Steitz, *Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells*. Mol. Cell, 2003. **12**: p. 113-123.
33. Balakin, A., L. Smith, and M. Fournier, *The RNA World of the nucleolus: two major families of small RNAs defined by different box elements with related functions*. Cell, 1996. **86**: p. 823-834.
34. Ganot, P., M. Caizergues-Ferrer, and T. Kiss, *The family of box H/ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation*. Genes Dev., 1997. **11**: p. 941-956.
35. Tran, E., J. Brown, and E.S. Maxwell, *Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus?* Trends Biochem. Sci., 2004. **29**: p. 343-350.
36. Ni, J., A. Tien, and M.J. Fournier, *Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA*. Cell, 1997. **89**: p. 565-573.
37. Peculis, B., *RNA processing: pocket guides to ribosomal RNA*. Curr. Biol., 1997. **7**: p. R480-R482.
38. Bally, M., J. Hughes, and G. Cesareni, *SnR30: A New, Essential Small Nuclear RNA from Saccharomyces cerevisiae*. Nucleic Acids Res., 1988. **16**: p. 5291-5303.
39. Tollervey, D., *A Yeast Small Nuclear RNA Is Required for Normal Processing of Pre-Ribosomal RNA*. EMBO J., 1987. **6**: p. 4169-75.
40. Watkins, N., et al., *A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP*. Cell, 2000. **103**: p. 457-466.
41. Kiss-Laszlo, Z., Y. Henry, and T. Kiss, *Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA*. EMBO J., 1998. **17**: p. 797-807.
42. Bachellerie, J.P., et al., *Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA*. Trends Biochem. Sci., 1995. **20**: p. 261-264.

43. Kiss-Laszlo, Z., et al., *Site-specific ribose methylation of pre-ribosomal RNA: a novel function for small nucleolar RNAs*. Cell, 1996. **85**: p. 1077-1088.
44. Maxwell, E.S. and M. Fournier, *The small nucleolar RNAs*. Ann. Rev. Biochem., 1995. **35**: p. 897-934.
45. Borovjagin, A. and S. Gerbi, *Xenopus U3 snoRNA GAC-box A' and box A sequences play distinct functional roles in rRNA processing*. Mol. Cell. Biol., 2001. **21**: p. 6210-6220.
46. Borovjagin, A. and S. Gerbi, *Xenopus U3 snoRNA docks on pre-rRNA through a novel base-pairing interaction*. RNA, 2004. **10**: p. 942-953.
47. Liang, W. and M.J. Fournier, *U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing*. Genes Dev., 1995. **9**: p. 2433-2443.
48. Schmitt, M.E. and D.A. Clayton, *Nuclear RNase MRP Is Required for Correct Processing of Pre-5.8S rRNA in Saccharomyces cerevisiae*. Mol. Cell. Biol., 1993. **13**: p. 7935-41.
49. Bousquet-Antonelli, C., et al., *A Small Nucleolar RNP Protein Is Required for Pseudouridylation of Eukaryotic Ribosomal RNAs*. EMBO J., 1997. **16**: p. 4770-4776.
50. Watkins, N., et al., *Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all box H/ACA-motif snoRNPs and constitute a common bipartite structure*. RNA, 1998. **4**: p. 1549-1568.
51. Henras, A., et al., *Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs*. EMBO J., 1998. **17**: p. 7078-7090.
52. Zebarjadian, Y., et al., *Point Mutations in Yeast CBF5 Can Abolish In Vivo Pseudouridylation of rRNA*. Mol. Cell. Biol., 1999. **19**: p. 7461-7472.
53. Girard, J.P., et al., *GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast*. EMBO J., 1992. **11**: p. 673-682.
54. Shaw, P. and E. Jordan, *The nucleolus*. Annu Rev Cell Dev Biol., 1995. **11**: p. 93-121.
55. Bagni, C. and B. Lapeyre, *Gar1p Binds to the Small Nucleolar RNAs snR10 and snR30 in Vitro through a Nontypical RNA Binding Element*. J. Biol. Chem., 1998. **273**: p. 10868-10873.
56. Dragon, F., F. Pogacic, and W. Filipowicz, *In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs*. Mol. Cell. Biol., 2000. **20**: p. 3037-3048.
57. Tollervey, D., et al., *Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly*. Cell, 1993. **72**: p. 443-457.

58. Wang, H., et al., *Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 Å resolution*. EMBO J., 2000. **19**: p. 7-17.
59. Omer, A., et al., *In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex*. Proc. Nat. Acad. Sci. USA, 2002. **99**: p. 5289-5294.
60. Cahill, N., et al., *Site-specific cross-linking analyses reveal an asymmetric distribution for a box C/D snoRNP*. EMBO J., 2002. **21**: p. 3816-3828.
61. Fatica, A., et al., *Fibrillarin binds directly and specifically to U16 box C/D snoRNA*. RNA, 2000. **6**: p. 88-95.
62. Nottrott, S., et al., *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**: p. 6119-6133.
63. Vidovic, I., et al., *Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment*. Molecular Cell, 2000. **6**: p. 1331-1342.
64. Klein, D., et al., *The kink-turn: a new RNA secondary structure motif*. EMBO J., 2001. **20**: p. 4214-4221.
65. Gautier, T., et al., *Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis*. Mol. Cell. Biol., 1997. **17**: p. 7088-7098.
66. Lafontaine, D. and D. Tollervey, *Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability*. RNA, 1999. **5**: p. 455-467.
67. Lafontaine, D. and D. Tollervey, *Synthesis and assembly of the box C+D small nucleolar RNPs*. Mol. Cell. Biol., 2000. **20**: p. 2650-2659.
68. Szewczak, L.W., et al., *Exclusive interaction of the 15.5kD protein with the terminal box C/D motif of a methylation guide snoRNP*. Chem. Biol., 2002. **9**: p. 1095-1107.
69. Gaspin, C., et al., *Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes*. J. Mol. Biol., 2000. **297**: p. 895-906.
70. Omer, A., et al., *Homologs of small nucleolar RNAs in Archaea*. Science, 2000. **288**: p. 517-522.
71. Woese, C., O. Kandler, and M. Wheelis, *Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya*. Proc. Natl. Acad. Sci. USA, 1990. **87**: p. 4576-4579.
72. Dennis, P., *Ancient ciphers: translation in archaea*. Cell, 1997. **89**: p. 1007-1010.
73. Noon, K.R., E. Bruenger, and J.A. McCloskey, *Posttranscriptional Modifications in 16S and 23S rRNAs of the Archaeal Hyperthermophile Sulfolobus solfataricus*. J. of Bacteriol., 1998. **180**: p. 2883-2888.

74. Dennis, P., A. Omer, and T. Lowe, *A guided tour: small RNA function in Archaea*. Mol. Microbiol., 2001. **40**: p. 509-519.
75. d'Orval, B.C., et al., *Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA<sup>Trp</sup> intron guides the formation of two ribose-methylated nucleosides in the mature tRNA<sup>Trp</sup>*. Nucleic Acids Res., 2001. **29**: p. 4518-4529.
76. Bortolin, M.-L., J.-P. Bachellerie, and B. Clouet-d'Orval, *In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA<sup>Trp</sup>*. Nucleic Acids Res., 2003. **31**: p. 6524-6535.
77. Tang, T., et al., *Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus**. Proc. Nat. Acad. Sci. USA, 2002. **99**: p. 7536-7541.
78. Bult, C., et al., *Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii**. Science, 1996. **273**: p. 1058-1072.
79. Watababe, Y. and M. Gray, *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 archaeobacteria*. Nucleic Acids Res., 2000. **28**: p. 2342-2352.
80. Liang, X., L. Liu, and S. Michaeli, *Identification of the first Trypanosome H/ACA RNA that guides pseudouridine formation on rRNA*. J. Biol. Chem., 2001. **276**: p. 40313-40318.
81. Russell, A., M. Schnare, and M. Gray, *Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis**. RNA, 2004. **10**: p. 1034-1046.
82. Amiri, K., *Fibrillar-like proteins occur in the domain Archaea*. J. Bacteriol., 1994. **176**: p. 2124-2127.
83. Wu, P., et al., *Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast*. J. Biol. Chem., 1998. **273**: p. 16453-16463.
84. Kuhn, J., E. Tran, and E.S. Maxwell, *Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein*. Nucleic Acids Res., 2002. **30**: p. 931-941.
85. Mao, H., S.A. White, and J.R. Williamson, *A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex*. Nat. Struct. Biol., 1999. **6**: p. 1139-1147.
86. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution*. Science, 2000. **289**: p. 905-920.
87. Moore, T., et al., *Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA*. Structure, 2004. **12**: p. 807-818.

88. Mao, H. and J.R. Williamson, *Local Folding Coupled to RNA Binding in the Yeast Ribosomal Protein L30*. J. Mol. Biol., 1999. **292**: p. 345-359.
89. Goody, T., et al., *The kink-turn motif in RNA is dimorphic, and metal ion dependent*. RNA, 2004. **10**: p. 254-264.
90. Tran, E., X. Zhang, and E.S. Maxwell, *Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs*. EMBO J., 2003. **22**: p. 3930-3940.
91. Aittaleb, M., et al., *Structure and function of archaeal box C/D sRNP core proteins*. Nat. Struct. Biol., 2003. **10**: p. 256-263.
92. Runnegar, B. *Proterozoic eukaryotes: evidence from biology and geology*. in *Nobel Symposium*. 1994. New York: Columbia U. P.: p.287-297.
93. Rozhdestvensky, T.S., et al., *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**: p. 869-877.
94. Winkler, W., et al., *The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs*. RNA, 2001. **7**: p. 1165-1172.

**CHAPTER I. Archaeal Ribosomal Protein L7 is a Functional Homolog of the Eukaryotic  
15.5kD/Snu13p snoRNP Core Protein**

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

Department of Molecular and Structural Biochemistry, North Carolina State University  
Raleigh, North Carolina

Published in Nucleic Acids Research, 2002, Vol. 30, No. 4, p. 931-941.

## ABSTRACT

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

Keywords: ribosomal protein L7; archaeal sRNA; sRNP; box C/D snoRNP

## INTRODUCTION

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

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

Most recently, a fourth core protein designated the 15.5kD protein (Snu13p in yeast) has been identified [18]. Eukaryotic 15.5kD snoRNP protein binds directly to the box C/D core motif and is a

structural protein, initiating formation of the box C/D snoRNP core complex. Interestingly, 15.5kD protein also recognizes spliceosomal U4 snRNA and the crystal structure of the U4 snRNA:15.5kD protein complex has been solved [19, 20]. The 5 + 2 nucleotide asymmetric bulge formed between flanking stems I and II is stabilized by tandem sheared G:A pairs and a protruded pyrimidine nucleotide is important for 15.5kD recognition. This RNA motif adopts a highly ordered tertiary structure including a sharp bend of approximately 65° between the helical axes.

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

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

## MATERIALS & METHODS

### ***In Vitro* Transcription and Labeling of RNA**

The U14 box C/D core motif and mutant RNAs were transcribed from linearized plasmid templates pSP64T7U14 $\Delta$ AV and pSP64T7U14 $\Delta$ AV $\Delta$ CD, respectively [27]. RNA transcription was carried out using the RiboMAX system (Promega) according to the manufacturers protocol, at template concentrations of 0.1 g/L. RNA transcripts were purified by polyacrylamide gel electrophoresis. Wild type and mutant box C/D core motif RNAs were also transcribed from short DNA templates generated from PCR-amplification of the plasmid pSP64T7U14 $\Delta$ AV using the DNA oligo pairs listed below. Primer pairs containing the desired mutations were used to produce DNA templates positioned downstream of a T7 promoter site. RNA synthesis from these DNA templates was carried out using the RiboMAX system at template concentrations of 0.2-0.4 $\mu$ g/L and RNA transcripts were purified by polyacrylamide gel electrophoresis. Internally radiolabeled RNAs were transcribed from linearized plasmids pSP64T7U14 $\Delta$ AV and pSP64T7U14 $\Delta$ AV $\Delta$ CD in the presence of  $\alpha$  <sup>32</sup>P-CTP using T7 RNA polymerase [28]. Gel-purified RNA transcripts were radiolabeled at the 5' end and purified from free nucleotides by Sephadex G-25 spin column chromatography.

### **Primer Pairs for PCR-Amplification of DNA Templates**

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

### DNA Oligonucleotide Sequences

- A. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGATGATGGATTCC-3'
- B. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTCATGATGGATTCC-3'
- C. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGCTGATGGATTCC-3'
- D. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGATGATGGATTCC-3'
- E. 5'-ATTCGCTCAGACATCCAAGGAAGGAATTTTGG-3'
- F. 5'-ATTCGCTGAGACATCCAAGGAAGGAATTTTGG-3'
- G. 5'-ATTCGCGCAGACATCCAAGGAAGGAATTTTGG-3'
- H. 5'-ATTCGCTCAGTACTCCAAGGAAGGAATTTTGG-3'
- I. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGATGTGTGATTCC-3'
- J. 5'-CTAATACGACTCACTATAGGCCATTCGCTGGGATGATGGATTCC-3'
- K. 5'-CTAATACGACTCACTATAGGCCATTCGCTGCGATGATGGATTCC-3'
- L. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGAAGATGGATTCC-3'
- M. 5'-ATTCGCTCAGACATCCAAGGAAGGAATTTTGG-3'
- N. 5'-CTAATACGACTCACTATAGGCCATTCGCTGATGATGGATTCC-3'
- O. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGATCATGGATTCC-3'
- P. 5'-ATTCGCTCACACATCCAAGGAAGGAATTTTGG-3'
- Q. 5'-CTAATACGACTCACTATAGGCCATTCGCTGTGACGATGGATTCC-3'

### Cloning of *M. jannaschii* L7 & Mouse 15.5kD Protein Genes: Expression and Purification of Recombinant Proteins

The mouse 15.5kD protein coding sequence was PCR-amplified from I.M.A.G.E. Consortium clone #809008 using gene-specific primers. The *M. jannaschii* L7 protein coding sequence was amplified from genomic DNA. Primers possessed terminal BamHI restriction sites and amplified DNA fragments were digested with BamHI restriction endonuclease and inserted into the BamHI restriction

site of the pGEX-4T-1 plasmid (Amersham Pharmacia Biotech). *E. coli* DH5 cells were transformed with plasmid constructs and selected clones sequenced to confirm that the correct coding sequences were in frame with the N-terminal glutathione-S-transferase (GST) tag. Protein expression was accomplished using standard techniques. At three hours post-induction, harvested cells were disrupted by sonication in buffer D (20 mM HEPES pH 7.0, 100 mM KCl, 3 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5 mM DTT, 20% w/v glycerol) and insoluble protein removed from the sonicate by ultracentrifugation at 39,000 x g for 25 minutes at 4C. Soluble sonicate was then applied to glutathione Sepharose 4B resin (Amersham Pharmacia Biotech) equilibrated in buffer D and bound GST-tagged protein was eluted in buffer D supplemented with 20 mM glutathione. Protein concentrations were determined by the Bradford Assay. N-terminal GST tags were removed by thrombin (Sigma) cleavage overnight. GST cleavage reactions for the 15.5kD recombinant protein were carried out at 4 degrees C at a thrombin concentration of 1 U/50ug protein. GST cleavage reactions for the L7 recombinant protein were carried out at 25C at a thrombin concentration of 2 U/50ug protein.

### **Thermal Denaturation Analysis**

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

### **Protein:RNA Interaction Analysis**

Recombinant 15.5kD and L7 protein binding to the U14 core motif RNA was assessed by electrophoretic mobility-shift analysis as detailed previously [27] with the following modifications: binding reactions contained 0.1 nM <sup>32</sup>P labeled RNA, 2.0 M recombinant protein (GST-15.5kD, GST-L7, or L7), and assembly was carried out for 1 h at 4C. For competition studies, non-radiolabeled competitor RNA was added at 4000-fold molar excess of radiolabeled RNA. RNP complexes were resolved on a 4% non-denaturing polyacrylamide gel in TBE buffer containing 8% glycerol and visualized by autoradiography. Equilibrium RNA:protein binding was assessed using a nitrocellulose filter binding assay to measure affinities of recombinant L7 protein for the U14 box C/D core motif RNAs. Binding reactions were carried out as previously described [27] with the following modifications: 100 uL reactions were assembled at 4C using 5'-labeled RNAs at a concentration of 2.5 nM with the recombinant L7 protein being titrated from 15 pM to 750 nM. Binding reactions were carried out for 1h at 4 C and then the reaction mix was applied to a nitrocellulose membrane. Prior to loading, membranes were equilibrated in binding buffer (30 mM HEPES pH 7.0, 250 mM KCl, 4.5 mM MgCl<sub>2</sub>, 1.0 mM DTT, 0.1 mM EDTA, 10% w/v glycerol) and were washed twice after sample application with binding buffer. Membranes were then dried and bound radioactivity was visualized using a Molecular Dynamics Model 425F PhosphorImager and quantified using the ImageQuant software v3.3.

## RESULTS

### **Archaeal Ribosomal Protein L7 is a Functional Homolog of Eukaryotic 15.5kD snoRNP Protein and Binds the Box C/D snoRNA Core Motif**

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

Based upon the sequence similarity of the eukaryotic 15.5kD and archaeal L7 proteins, we reasoned that L7 could function as a homolog of the 15.5kD protein and bind the box C/D sRNAs. To test this hypothesis, we assessed the ability of recombinant *M. jannaschii* L7 protein to recognize the box C/D core motif and form an RNP complex. The RNA used for these electrophoretic mobility-shift analyses was the box C/D core motif derived from mouse U14 snoRNA consisting of base paired 5' and 3' termini with flanking boxes C and D, respectively. This RNA substrate also possesses the rRNA-complementary sequence immediately upstream of box D but lacks the internal box C/D motif with associated rRNA-complementary sequence as well as the adjacent U14 variable region.

Recombinant mouse 15.5kD protein (+/- GST tag) binds this box C/D RNA, and mutagenesis of boxes C and D demonstrates the essentiality of these conserved nucleotide sequences for protein recognition (Figure 2A). Similarly, the *M. jannaschii* L7 protein (+/- GST tag) binds the U14 box C/D core motif

with the same specificity (Figure 2B). Competition experiments with non-radiolabeled U14 snoRNA also demonstrate the requirement of boxes C and D for L7 binding. Addition of excess non-radiolabeled wild type box C/D RNA disrupts RNP formation while added RNA possessing point mutations in boxes C and D is not an effective competitor.

### **Archaeal L7 Protein Binds the Box C/D Core Motif with the Same Affinity as Eukaryotic**

#### **15.5kD Protein**

The snoRNA box C/D core motif can be folded into a secondary structure consisting of helical stems I and II flanking an asymmetric 5 + 2 nucleotide bulge (Figure 3A). This structure is based upon the crystal structure of the 15.5kD protein-binding site observed in the U4 snRNA:15.5kD RNP complex [20] which has an essentially identical box C/D-like motif [18]. Eukaryotic 15.5kD protein binds the snoRNA box C/D core motif with high affinity and a  $K_d$  of 8 nM has been estimated from electrophoretic mobility-shift analysis of the 15.5kD:box C/D core motif complex. To compare the binding affinity of ribosomal protein L7 with the 15.5kD protein, the  $K_d$  of the L7:box C/D core motif was determined in equilibrium binding analysis using a filter-binding assay. L7 protein binds the box C/D core motif with the same affinity as the 15.5kD protein with a  $K_d$  of 5 nM (Table 1). The mouse 15.5kD protein exhibits a similar  $K_d$  value of 19 nM when its binding to the box C/D core motif is assessed using the filter binding assay (data not shown). However, we believe the determined  $K_d$  values for both 15.5kD and L7 binding are underestimates of their affinity for the box C/D core motif. Both the electrophoretic mobility-shift and filter-binding assays are limited in determining an actual binding constant when analyzing RNA-binding proteins that exhibit such high affinity for the RNA substrate. The concentrations of radiolabeled RNA in our equilibrium RNA:protein binding assay are nearly equivalent to the L7 protein concentrations at the determined  $K_d$  value of 5 nM. Therefore, the filter-binding assay is being carried out under stoichiometric conditions rather than under conditions

of limiting RNA concentrations. Consistent with stoichiometric conditions, the binding curve for RNP formation exhibits a steep shape (data not shown). Unfortunately, the exceptionally high affinity of L7 for the box C/D core motif precludes the possibility of measuring radioactively-labeled RNP formation at sufficiently low RNA concentrations where small amounts of RNA are bound to protein. With sufficiently dilute RNA concentrations, it is impossible to detect RNP formation since the L7:box C/D core motif complex contains too little radioactive RNA to detect. Therefore, the nM  $K_d$  values determined for both 15.5kD (electrophoretic mobility-shift) and L7 (filter-binding assay) binding are most likely upper limits of binding affinity and the actual  $K_d$  values are probably an order of magnitude or more lower.

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

### **L7 Requires the Same Box C/D Sequence and Structural Elements for Binding As Does the 15.5kD Protein**

X-ray crystallographic analysis of the 15.5kD:U4 snRNP complex has revealed that the 15.5kD protein requires specific sequence and structural elements of the folded RNA for protein binding [20]. In particular, the juxtaposition of box C and D-like sequences by flanking stems I and/or II to form the asymmetric bulge is critical. This 5 + 2 nucleotide asymmetric bulge establishes the bulk of the protein binding site where tandem, sheared G:A pairs and a protruded pyrimidine nucleotide contact the bound protein (Figure 3B). These same structural features are presumed to be present in the snoRNA box

C/D core motif and the observed requirements for box C/D snoRNP assembly *in vivo* are consistent with this RNA folded structure [4]. We therefore assessed the importance of specific RNA elements of the box C/D core motif for L7 protein binding to determine if this archaeal ribosomal protein interacts with the RNA in the same manner as the eukaryotic 15.5kD protein.

The folded structure of box C/D core motif RNA and various mutants was monitored in thermal denaturation/renaturation analysis (Figure 4) while RNP assembly (Figure 5) and L7 affinity (Table 1) were determined in electrophoretic mobility-shift and equilibrium binding analyses, respectively. Mutations were created in stem I, stem II, and the asymmetric bulge (Figure 3C) to assess the contribution of each region for both RNA folding and L7 binding. In contrast to the box C/D core motif RNA, the mutant RNAs exhibited lower binding affinities for L7 allowing radiolabeled RNP formation to be measured over a range of protein concentrations sufficient to determine equilibrium binding constants. Calculated  $K_d$  values provide relative L7 binding strengths with respect to the upper limit of 5 nM estimated for the box C/D core motif RNA and the baseline, non-specific protein binding to ribosomal RNA at 400 nM.

***Stem I or External Stem.*** Thermal denaturation of the folded box C/D snoRNA core motif results in a sharp melting profile (Figure 4) that is typical of a highly structured RNA. The large increase and decrease in hyperchromicity over a narrow temperature range around the  $T_m$  is characteristic of an RNA that folds in a highly cooperative manner. Such a melting profile is consistent with the proposed folded structure of the box C/D motif as illustrated for U4 snoRNA in Figure 3B. Mutating the 5' terminal sequence of U14 to disrupt stem I base pairing (stem I mut) significantly alters the folded structure of the box C/D core motif as evidenced by the reduction in cooperativity during thermal denaturation (Figure 4). As a consequence, L7 binding is disrupted (Figure 5) and affinity for the core

motif is reduced to a level equivalent to non-specific association (Table 1). Compensatory changes in the 3'-terminal sequence that reestablish stem I base pairing (stem I comp) restore the wild type RNA melting profile and protein binding, albeit at a lower affinity. Thus, folding of the box C/D core motif for L7 binding is dependent on the formation of this RNA helix.

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

More pronounced effects on both the melting profile and L7 affinity are observed when nucleotides of boxes C and D that establish stem II structure are altered. Mutation of either G11 in box C or C41 in box D alters the structure of the folded core motif as evidenced by changes in the RNA melting profiles. L7 binding to the core motif is also reduced. Restoration of base pairing via compensatory mutations restores the melting profile and L7 affinity indicating that base pairing of these two bases is important for L7 interaction. Thus, hydrogen-bonding of this nucleotide pair within the core motif contributes to the stability of the protein:RNA interaction, but specific nucleotides at these particular positions appear less important for L7 binding. Interestingly, nucleotides G11 and C41 are conserved in boxes C and D, respectively, at greater than 97% of all box C/D RNAs in eukaryotes and Archaea [21, 23].

Covariation of this G:C base pair is not evident suggesting that these nucleotides positioned specifically within boxes C and D may be important for the binding of other proteins. Recent work has indeed indicated that this region of the snoRNA core motif is important for fibrillarin binding [30].

Equally interesting is the U10:U42 base pair of stem II that is adjacent to the asymmetric bulge.

Mutant U10a replaces uridine with adenosine and preserves base pairing. This nucleotide substitution has no effect upon the melting profile nor L7 affinity for the core motif RNA (Figure 5 and Table 1).

In contrast, substitution of the uridine with a cytidine alters core motif folded structure as evidenced by changes in the melting profile and L7 binds with reduced affinity. These results suggest that hydrogen-bonding between these two nucleotide positions is important. U10 is greater than 90% conserved in eukaryotic snoRNAs but is less than 80% conserved in archaeal sRNAs [4, 21].

Covariation analysis of archaeal sRNAs indicates that hydrogen-bonding between this nucleotide pair is most often maintained with alternative nucleotide pairs (data not shown).

***Asymmetric bulge.*** Formation of the 5 + 2 asymmetric bulge with tandem, sheared G:A pairs and protruded pyrimidine nucleotide (Figure 3B) is critical for 15.5kD binding to U4 snRNA [20] and presumably the snoRNA box C/D core motif. If L7 is a functional homolog of 15.5kD, then the sheared G:A pairs and protruded nucleotide should also be critical for L7 recognition. Indeed, mutations within the asymmetric bulge of the core motif have a profound impact on L7 binding. Point mutations in each of the 4 G/A residues that constitute the two sheared base pairs (mutants G8, A9, G43 and A44) substantially alter RNA folded structure as seen in the melting profiles with alteration of G43 and A44 in box D having a particularly severe effect. All four nucleotides are critical for protein binding as each mutation completely disrupts L7 binding. Analysis of the G8c point mutation demonstrates that the L7 binding affinity is reduced to non-specific levels.

The protruded nucleotide which makes important base-specific contacts with the 15.5kD protein is also critical. Mutations made at this critical nucleotide include substitution of both pyrimidine (mutant U7c) and purine (mutant U7g) nucleotides. These substitutions affect the melting profile of the folded RNA similarly but have significantly different effects upon L7 binding. Replacement of uridine with another pyrimidine affects RNP formation and L7 affinity only slightly. In contrast, replacement of the wild type pyrimidine nucleotide with the purine guanosine eliminates L7 binding and reduces L7 affinity to non-specific levels. These observations are consistent with the central role of U7 in sequence-specific protein recognition (see Discussion).

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

## DISCUSSION

While genes for box C/D sRNAs and homologs for the sRNP core proteins fibrillarin and Nop56/58 have been identified in archaeal genomes, a defined homolog for the 15.5kD protein has been notably absent. Database analysis of available archaeal sequences reveals the strong sequence similarity of ribosomal protein L7 with mouse and human 15.5kD protein as well as yeast Snu13p. Other eukaryotic proteins exhibiting homology to archaeal L7 include eukaryotic ribosomal proteins L7, L30, S12, and the H/ACA snoRNP protein NHP2p [19]. Our studies now demonstrate that L7 binds the snoRNA box C/D core motif with the same specificity and affinity as the eukaryotic 15.5kD snoRNP core protein. Because of the high affinity of this RNA-binding protein for the box C/D core motif, it is difficult to determine an accurate  $K_d$  value. Our estimate in the nM range is consistent with that made for eukaryotic 15.5kD protein and both values most likely represent lower estimates of the true strength of these RNA:protein interactions.

The highly conserved L7 gene was first identified from the sequencing of various archaeal genomes and designated L7 because of its relatedness to ribosomal proteins [29]. Recent crystallization of the *H. marismortui* ribosome has demonstrated the presence of this protein (L7Ae) in the 50S large subunit, thus verifying its function as an archaeal ribosomal protein [26]. An experimental demonstration of ribosomal protein L7's presence in the archaeal sRNP complex awaits further characterization of the sRNP core proteins. However, the ability of L7 to bind the box C/D core motif indicates that L7 serves a dual role in Archaea as both ribosomal and sRNP core proteins. The binding of L7 to both 23S rRNA and the sRNAs could potentially play important regulatory roles, coordinating ribosome assembly with sRNP assembly and sRNA stabilization. Strikingly, 15.5kD protein also serves a dual role in eukaryotes, binding to two different RNA substrates to function as both an snoRNP core protein

and a U4-specific snRNP protein [18, 19]. Again, a regulatory role to coordinate snoRNP biogenesis with pre-mRNA processing has been suggested [18]. The fact that this protein homolog binds both identical (snoRNA/sRNA) and unique (rRNA vs. U4 snRNA) substrates in Archaea versus Eukaryotes is noteworthy, suggesting a possible evolution of protein function (see below).

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

Individual K-turn motifs often exhibit unique structural features including variations in asymmetric loop size and protruded nucleotide identity while still exhibiting the overall features of the motif. The L7Ae binding site in 23S rRNA (KT-15) and the U4 (presumably box C/D core motif) share similarities in folded structure (Figures 6A and 6B). Both exhibit helical regions flanking an

asymmetric bulge and the protruded pyrimidine nucleotide is critical for L7 binding. In particular, Klein and coworkers noted that the bulged pyrimidine nucleotide of KT-15 fit the hydrophobic binding pocket of L7 and a uridine nucleotide afforded better interactions with critical amino acids than a cytidine residue. In contrast, a purine ring was too large for insertion into this hydrophobic pocket. Our mutational studies of base substitution at this protruded nucleotide within the box C/D core motif are consistent with their structural predictions of L7 binding at this nucleotide position in the KT-15 turn.

While similar in overall structure, the folded K-turns of U4 snRNA and KT-15 exhibit distinct differences when compared. Most notably, the U4 and box C/D core motif present two tandem, sheared G:A pairs characteristic of most K-turns (Figure 6A). In contrast, the KT-15 k-turn determined from the *H. marismortui* 50S ribosome subunit crystal structure reveals a distinctly different nucleotide configuration in this region (Figure 6B). Specifically, a single sheared G:A pair is followed by a G:U:A base triple where the adenine residue or third nucleotide is derived from the short side of the asymmetric loop. This nucleotide arrangement causes a more pronounced bend in the K-turn motif and results in a sharp kink in the phosphate backbone of both RNA strands. These differences exhibited by individual K-turn motifs typically account for the recognition of the structurally similar K-turns by different proteins. For example, the 6 K-turns of the *H. marismortui* 50S ribosome subunit bind nine different ribosomal proteins. However, despite these structural differences between the KT-15 and the U4 (box C/D core motif) K-turns, each is recognized by L7. Strikingly, L7 binds both motifs utilizing identical regions of the K-turn, including the NC or stem II region which exhibits differences in sheared G:A pair configuration.

Examination of the L7 crystal structure reveals that this ribosomal protein folds into the same folded

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

Particularly striking is the homology exhibited between snoRNP and ribosomal proteins and the similar ribonucleoprotein structures formed between these proteins and their cognate RNAs. Archaeal protein L7 and eukaryotic 15.5kD protein share sequence similarity with ribosomal protein L30, H/ACA snoRNP core protein Nhp2p, and ribosomal protein S12 [19]. All are members of a family of RNA-binding proteins originally identified by Koonin and coworkers [33]. The homology exhibited by these ribosomal proteins and snoRNP core proteins indicate a common ancestral coding sequence. The 15.5kD:box C/D snoRNA and L7Ae:rRNA complexes are also strikingly similar in RNP structure, perhaps indicating a common ancestral ribonucleoprotein structure. The similarity in RNP organization suggests to us that the snoRNPs and sRNPs could well have their evolutionary origins in primordial ribosomes. Such a progenitor ribosomal RNP structure could have evolved to acquire cis-acting sequences complementary to rRNA that are important for RNA folding. Additional proteins bound to the RNP motif might assist in rRNA precursor folding and/or bring new function to the

complex such as nucleotide modification activities. With time, these RNP motifs would have become independent from the ribosome itself, yet still retained their critical functions in ribosome biogenesis. The occurrence of two box C/D motifs within the intron of the *A. fulgidus* Trp-tRNA that apparently function in cis to guide 2'-O-methylation of specific nucleotides in the unspliced tRNA is consistent with such a scenario [22]. A similar model for the evolution of the snRNP splicing complexes has been previously suggested where folded elements of group II introns evolved to become independent snRNP complexes while maintaining their functional roles in pre-mRNA splicing [34]. Thus, present day RNP complexes critical for post transcriptional RNA processing and modification may have their origins in cis-acting ribonucleoprotein elements of the ancient RNA world.

#### **ACKNOWLEDGMENTS**

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

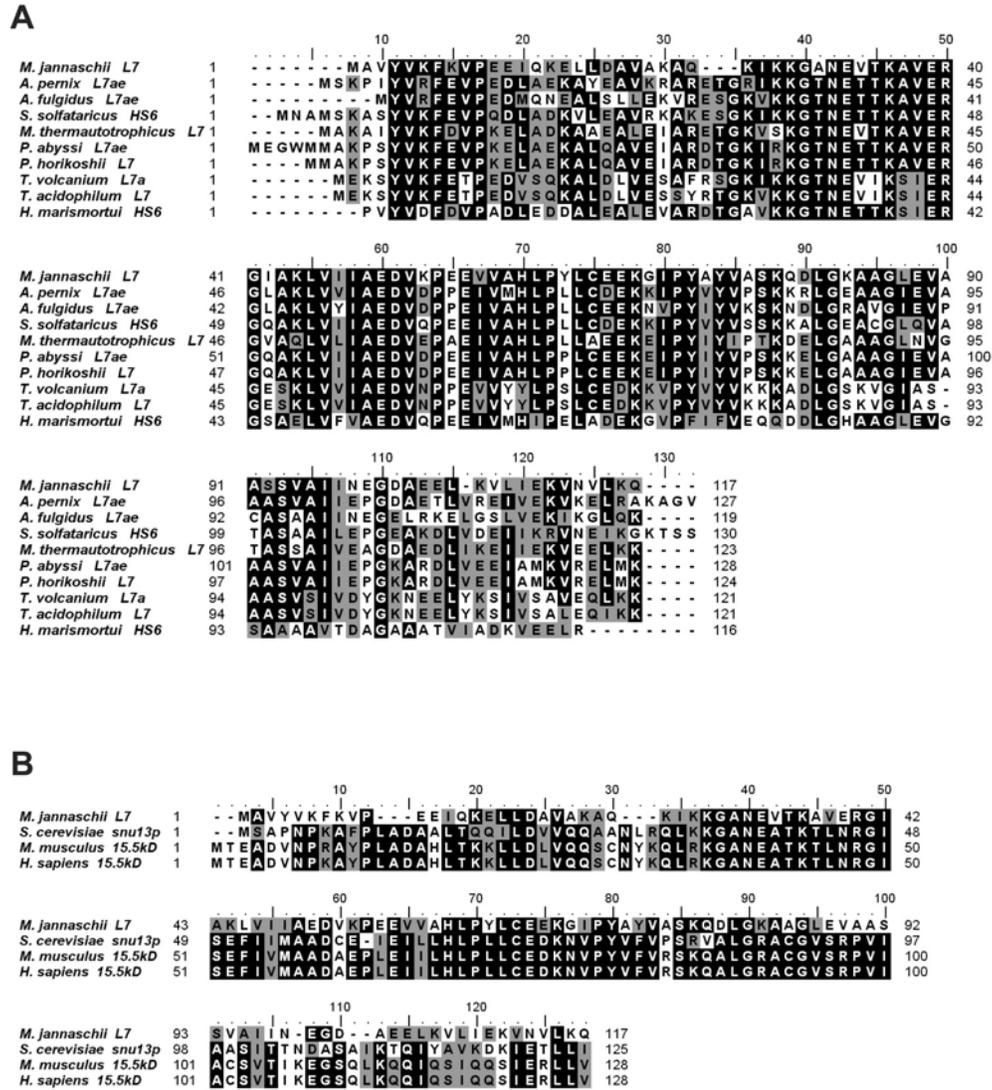
## REFERENCES

1. Maxwell, E. S. and Fournier, M. J. (1995) The small nucleolar RNAs, in *Ann. Rev. Biochem.*, **35**, 897-934.
2. Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617-3622.
3. Smith, C. and Steitz, J. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669-672.
4. Xia, L., Watkins, N. and Maxwell, E. S. (1997) Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA*, **3**, 17-26.
5. Balakin, A., Smith, L. and Fournier, M. J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823-834.
6. Ganot, P., Caizergues-Ferrer, M. and Kiss, T. (1997) The family of box H/ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941-956.
7. Tollervy, D. (1996) Small nucleolar RNAs guide ribosomal RNA methylation. *Science*, **273**, 1056-1057.
8. Bachellerie, J.-P., Cavaille, J. and Qu, L.-H. (2000) Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNA guides revisited, in *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions*. ASM Press, Washington, D. C., pp. 191-203.
9. Watkins, N., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., and Luhrmann, R. (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all box H/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549-1568.
10. Henras, A., Henry, H., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelunge, J.-P. and Caizergues-Ferrer, M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.*, **17**, 7078-7090.
11. Tyc, K., and Steitz, J. (1989) U3, U8, and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J.*, **8**, 3113-3119.
12. Wang, H., Boisvert, D., Kim, K., Kim, R., and Kim, S.-H. (2000) Crystal structure of a fibrillarin homologue from *Metanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J.*, **19**, 7-17.

13. Newman, D., Kuhn, J., Shanab, G. and Maxwell, E. S. (2000) Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA*, **6**, 861-879.
14. Lafontaine, D. and Tollervey, D. (1999) Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA*, **5**, 455-467.
15. Lafontaine, D. and Tollervey, D. (2000) Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol. Cell. Biol.*, **20**, 2650-2659.
16. Caffarelli, E., Losito, M., Giorgi, C., Fatica, A. and Bozzoni, I. (1998) In vivo identification of nuclear factors interacting with the conserved elements of box C/D small nucleolar RNAs. *Mol. Cell. Biol.*, **18**, 1023-1028.
17. Lyman, S., Gerace, L. and Baserga, S. (1999) Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins. *RNA*, **5**, 1597-1604.
18. Watkins, N., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Branlant, C. and Luhrmann, R. (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457-466.
19. Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. and Luhrmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.*, **18**, 6119-6133.
20. Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331-1342.
21. Omer, A., Lowe, T., Russell, A., Ehardt, H., Eddy, S. and Dennis, P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517-522.
22. Dennis, P., Omer, A. and Lowe, T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.*, **40**, 509-519.
23. Gaspin, C., Cavaille, J., Erauso, G. and Bachellerie, J.-P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.*, **297**, 895-906.
24. Amiri, K. (1994) Fibrillarin-like proteins occur in the domain of Archaea. *J. Bacteriol.*, **176**, 2124-2127.
25. Lafontaine, D. and Tollervey, D. (1990) Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *TIBS*, **23**, 383-388.

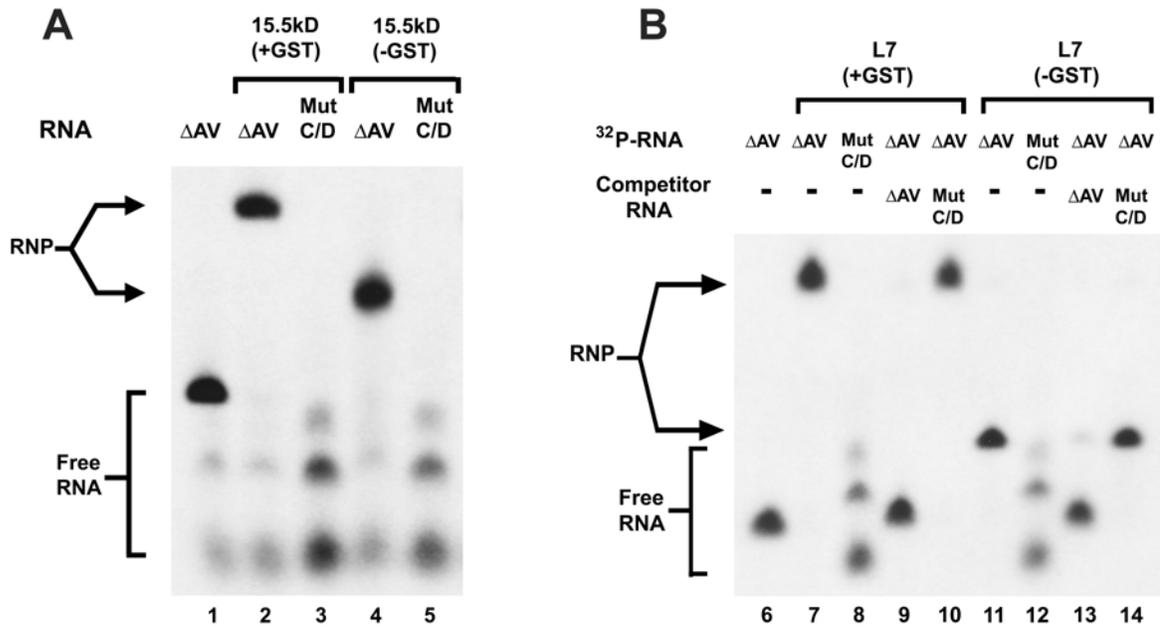
26. Ban, N., Nissan, P., Hansen, J., Moore, P. and Steitz, T. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.
27. Watkins, N., Newman, D., Kuhn, J. and Maxwell, E. S. (1998) In vitro assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA*, **4**, 582-593.
28. Watkins, N., Leverette, R., Xia, L., Andrews, M. and Maxwell, E. S. (1996) Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA*, **2**, 118-133.
29. Bult, C., White, O., Olsen, G., Zhu, L., Fleischmann, R., Sutton, G., Blake, J., Fitzgerald, L., Clayton, R., Gocayne, J., et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058-1072.
30. Fatica, A., Galardi, S., Altieri, F. and Bozzoni, I. (2000) Fibrillarlin binds directly and specifically to U16 box C/D snoRNA. *RNA*, **6**, 88-95.
31. Klein, D., Schmeing, T., Moore, P. and Steitz, T. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214-4221.
32. Winkler, W., Grundy, F., Murphy, B. and Henkin, T. (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA*, **7**, 1165-1172.
33. Koonin, E., Bork, P. and Sander, C. (1994) A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res.*, **22**, 2166-2167.
34. Sharp, P. (1991) Perspective: five easy pieces. *Science*, **254**, 663.
35. Thompson, J., Higgins, D. and Gibson, T. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **11**, 4673-4680.
36. Peitsch, M. (1996) ProMod and Swiss-Model: internet-based tools for automated comparative protein modeling. *Biochem. Soc. Trans.*, **24**, 274-279.

## FIGURES



**Figure 1. Archaeal ribosomal protein L7 is a homolog of the eukaryotic 15.5kD snoRNP protein.**

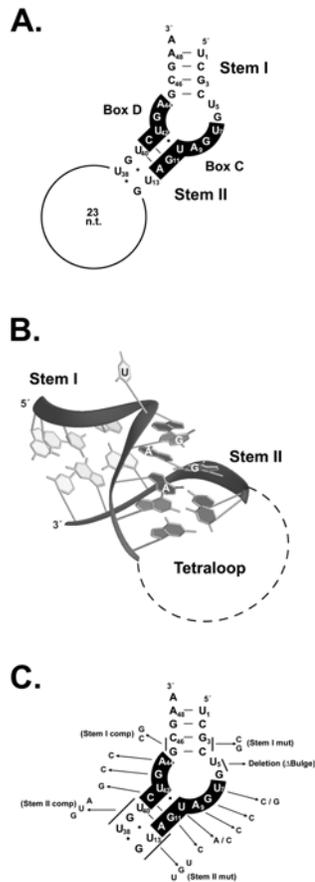
Amino acid sequence alignments of (A) archaeal L7 proteins and (B) the *M. jannaschii* L7 protein with eukaryotic 15.5kD/snu13p box C/D snoRNP core proteins. Primary sequence alignments were accomplished using ClustalW [35]. Identical residues are indicated by white letters highlighted in black and similar residues are indicated by gray shading. Dashes indicate inserted gaps to aid in the alignment.



**Figure 2. Archaeal ribosomal protein L7 binds the box C/D snoRNA core motif.**

(A) Radiolabeled, wild type, mouse U14 box C/D core motif (AV construct) and mutant U14 box C/D RNA (mut C/D with altered box C and D sequences) were incubated with purified mouse 15.5kD protein (+/-GST tag) and RNP formation assessed by electrophoretic mobility-shift analysis (lanes 1-5). The relative positions of free RNA and assembled RNP complex are indicated at the side.

(B) Radiolabeled, wild type, mouse U14 box C/D core motif and mutant U14 box C/D RNA were incubated with purified *M. jannaschii* L7 protein (+/-GST tag) in the presence or absence of excess, non-radiolabeled competitor RNA as designated (lanes 6-14). The relative positions of free RNA and assembled RNP complex are indicated at the side. Note: Wild type U14 RNA used for 15.5kD binding possesses a 12 nucleotide extension of its 3' terminus and thus migrates more slowly on the gel than wild type U14 RNA used for L7 binding. Also, free mut C/D RNA migrates on the polyacrylamide gel in two different RNA conformations due to altered box C and D sequences.



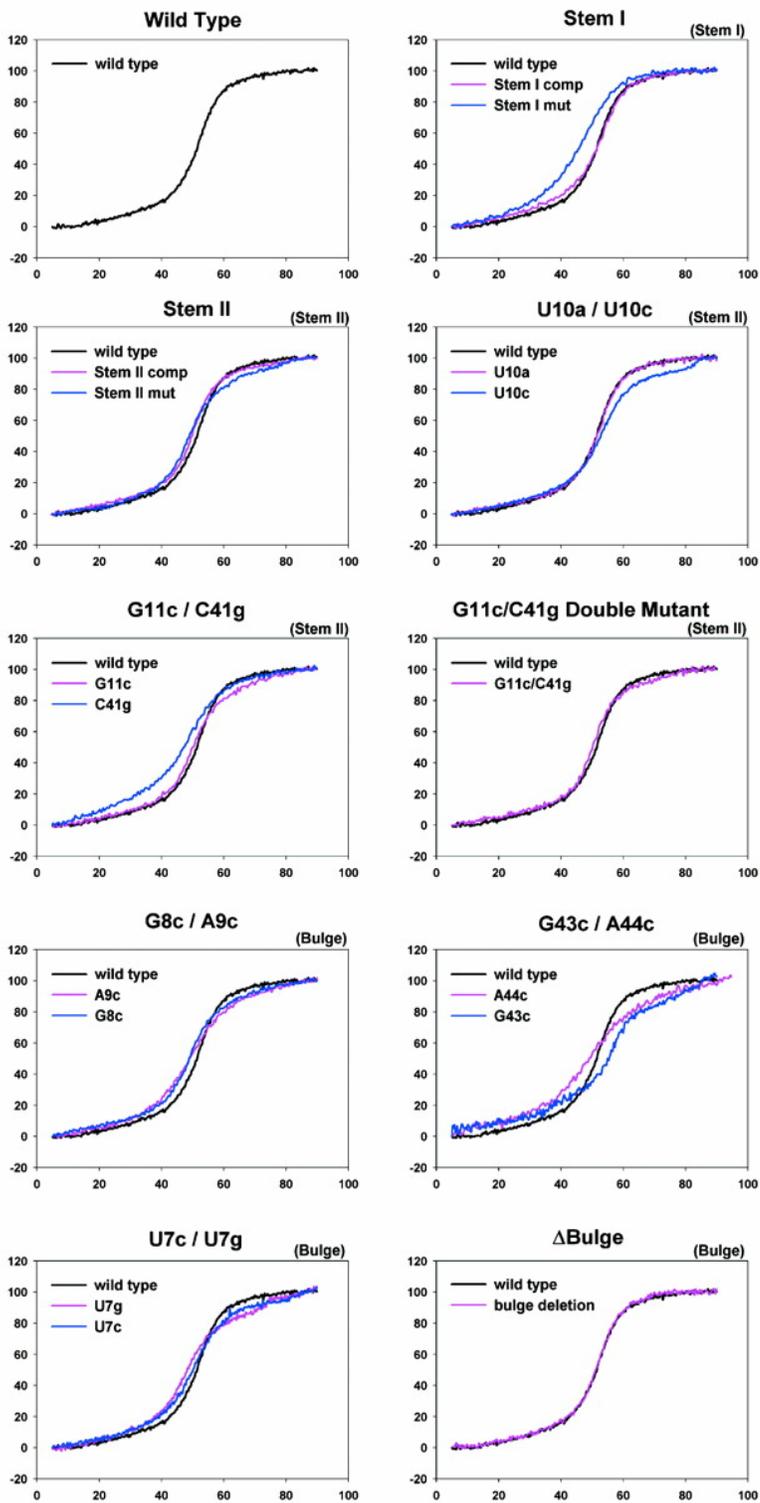
**Figure 3. Folded structure of the box C/D snoRNA core motif.**

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

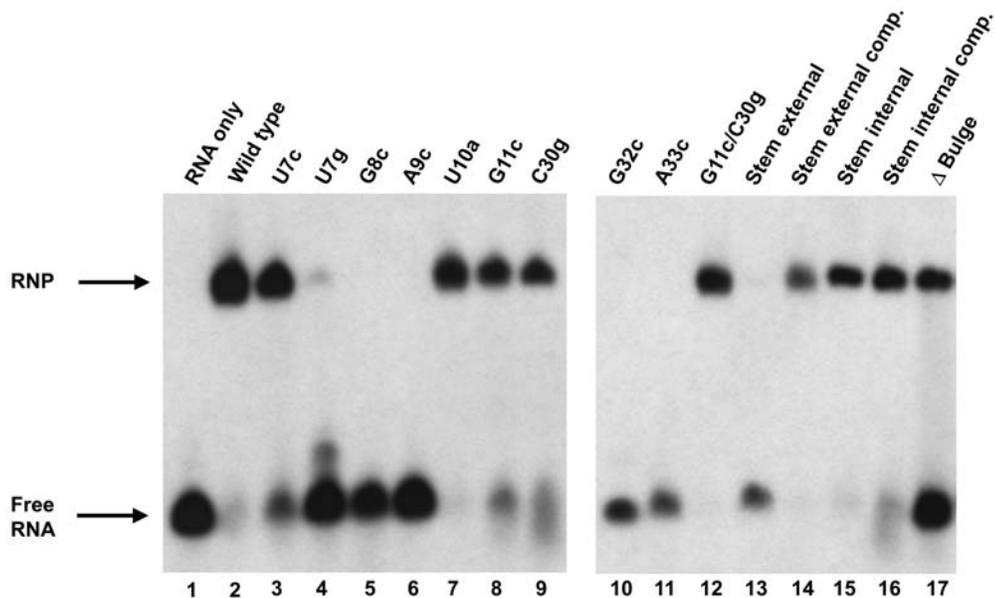
**Figure 4. The box C/D core motif exhibits a highly ordered RNA structure.**

Wild type and mutant box C/D core motif snoRNAs were subjected to thermal denaturation and renaturation and UV absorbance values at 260 nm were collected over a temperature range of 50 to 90°C as described in Materials and Methods. Absorbance values were normalized and plotted vs. temperature. Specific nucleotide alterations for the individual box C/D core motif mutants indicated in each panel are detailed in Figure 3C. All RNAs were diluted and subjected to additional thermal denaturation (data not shown) to insure that a unimolecular event was being measured.

Normalized Absorbance

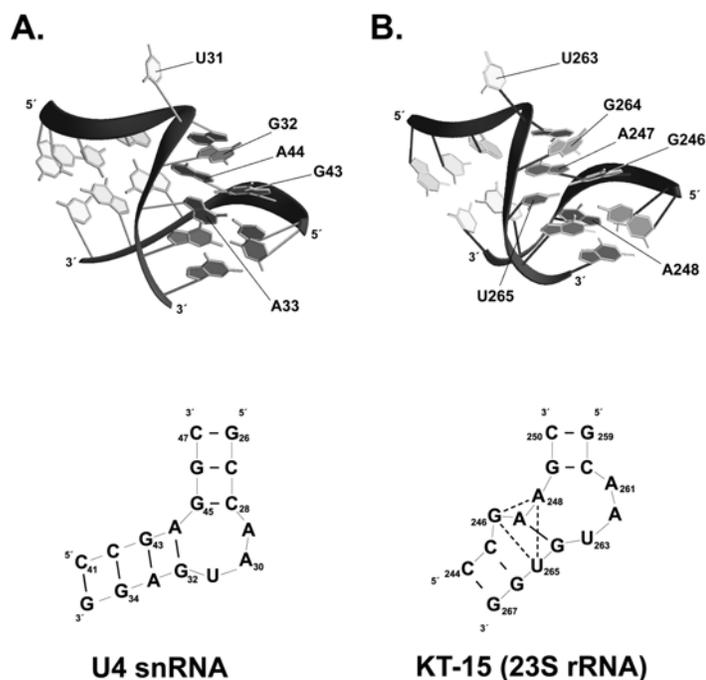


Temperature (°C)

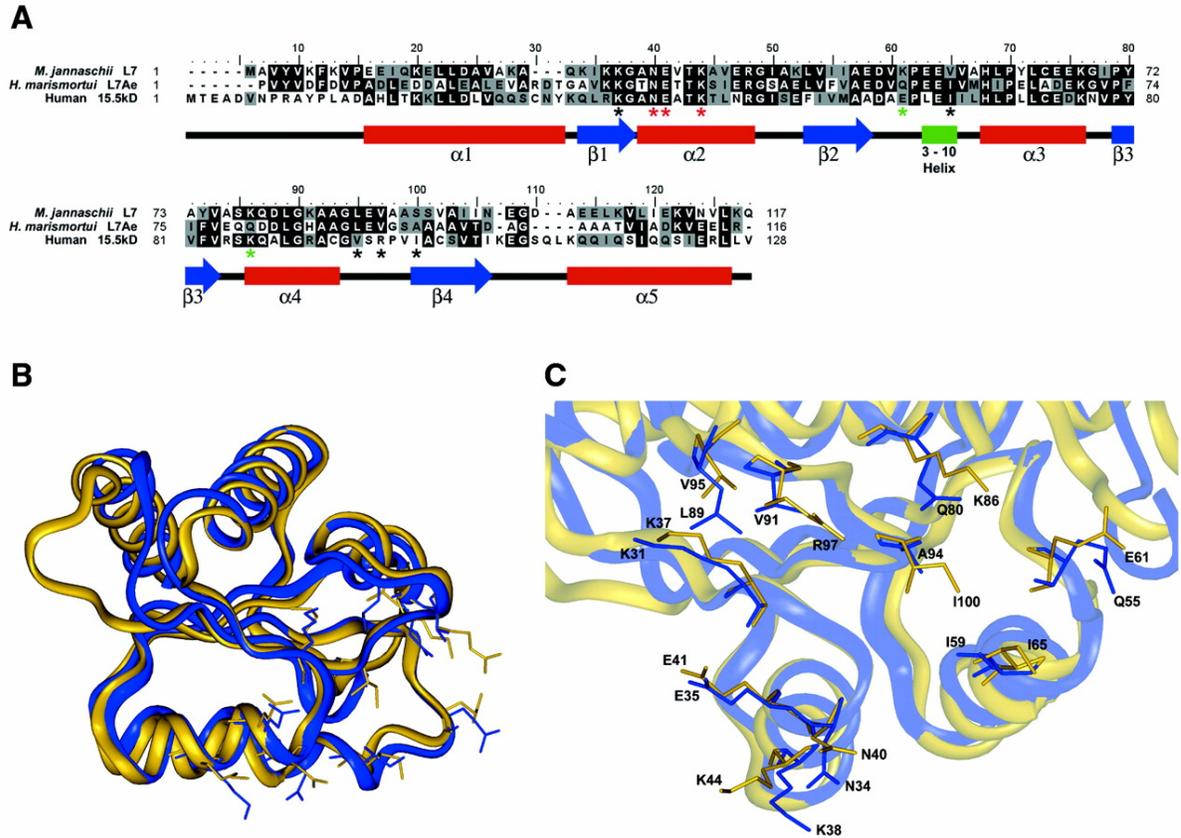


**Figure 5. L7 requires specific sequence and structural elements of the box C/D core motif for RNA:protein interaction.**

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



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



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

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

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

<b>RNA</b>	<b><math>K_d</math> Value (nM)</b>
Wild-Type	$\approx 5$
Non-specific rRNA	$\approx 400$
Stem I Region	
Stem I mutant	$\approx 400$
Stem I compensatory	53
<b>Stem II region</b>	
Stem II mutant	51
Stem II compensatory	67
C41g	137
G11c	122
G11c/C41g	$\approx 5$
U10a	$\approx 5$
U10c	115
<b>Asymmetric Bulge Region</b>	
U7c	73
U7g	$\approx 400$
G8c	$\approx 400$
<b><math>\Delta</math>Bulge</b>	118

**CHAPTER II. Efficient RNA 2'-O-methylation Requires Juxtaposed and Symmetrically  
Assembled Archaeal Box C/D and C'/D' RNPs**

Elizabeth J. Tran, Xinxin Zhang and E. Stuart Maxwell

Department of Molecular and Structural Biochemistry, North Carolina State University  
Raleigh, North Carolina

Published in The EMBO Journal, 2003, Vol. 22, No. 15, p. 3930-3940.

## ABSTRACT

Box C/D RNP complexes direct the nucleotide-specific, 2'-O-methylation of ribonucleotide sugars in target RNAs. *In vitro* assembly of an archaeal box C/D sRNP using recombinant core proteins L7, Nop56/58, and fibrillarin has yielded an RNA:protein enzyme that guides methylation from both the terminal box C/D core and internal C'/D' RNP complexes. Reconstitution of sRNP complexes containing only box C/D or C'/D' motifs has demonstrated that the terminal box C/D RNP is the minimal methylation-competent particle. However, efficient ribonucleotide 2'-O-methylation requires that both the box C/D and C'/D' RNPs function within the full-length sRNA molecule. In contrast to the eukaryotic snoRNP complex where the core proteins are distributed asymmetrically on the box C/D and C'/D' motifs, all three archaeal core proteins bind both motifs symmetrically. This difference in core protein distribution is a result of altered RNA-binding capabilities of the archaeal and eukaryotic core protein homologs. Thus, evolution of the box C/D nucleotide modification complex has resulted in structurally distinct archaeal and eukaryotic ribonucleoprotein particles.

Keywords: Archaea / box C/D RNP / snoRNA / sRNA / ribonucleotide methylation

## INTRODUCTION

The small nucleolar RNAs (snoRNAs) play critical roles in ribosome biogenesis, functioning in the processing and modification of pre-ribosomal RNA (Smith and Steitz, 1997; Kiss, 2001; Bachellerie et al., 2002; Kiss, 2002; Terns and Terns, 2002). The primary role of the vast majority of snoRNAs is to guide the site-specific modification of rRNA nucleotides. Guide regions within the snoRNA base pair with complementary sequences in the rRNA and direct snoRNA-associated enzymes to the designate nucleotide for ribose or base modification. Recent work has also revealed guide RNAs in Archaea (Gaspin et al., 2000; Omer et al., 2000; Dennis et al., 2001; Speckman et al., 2002). While archaeal organisms do not possess a nucleus, they nevertheless utilize snoRNA-like RNAs (sRNAs) for site-specific modification of rRNA nucleotides. The occurrence of guide RNAs in both Eukarya and Archaea indicates that the process of RNA-guided nucleotide modification is an ancient mechanism predating the divergence of Eukarya and Archaea more than 2 billion years ago (Runnegar, 1994).

The box C/D RNAs are one of two major snoRNA families. Box C/D RNAs direct the site-specific, 2'-O-ribose methylation of targeted nucleotides with rRNA and other RNA substrates (Tollervey, 1996; Smith and Steitz, 1997; Tycowski et al., 1998; Jady and Kiss, 2001). Members of this family are defined by the conserved boxes C and D located at the 5' and 3' termini, respectively (Tyc and Steitz, 1989; Caffarelli et al., 1996; Cavaille and Bachellerie, 1996; Watkins et al., 1996; Darzacq and Kiss, 2000; Watkins et al., 2000). These conserved sequences fold into a stem-loop-stem structure which is essential for the binding of box C/D RNP proteins as well as the nucleotide modification reaction itself. Additional internal sequences designated C' and D' boxes can be identified in eukaryotic snoRNAs and archaeal sRNAs (Kiss-Laszlo et al., 1998). Although the C' and D' boxes have been defined based

upon the terminal box C and D sequences, they are not as strictly conserved and are not easily identified in all eukaryotic snoRNAs. Both the terminal box C/D core and internal C'/D' motifs direct 2'-O-methylation of targeted ribonucleotides using rRNA-complementary regions located immediately upstream of boxes D and D'. Antisense sequences of 10-21 nucleotides in length base pair with the target region and direct methylation to the designate nucleotide positioned 5 nucleotides upstream of box D/D' within the snoRNA:rRNA duplex (Kiss-Laszlo et al., 1996; Cavaille and Bachellerie, 1998).

Eukaryotic box C/D snoRNAs associate with four core proteins: 15.5kD protein, nucleolar protein 56 (Nop56p), nucleolar protein 58 (Nop58p), and fibrillarin. Assembly of the terminal box C/D snoRNP complex is required for snoRNA processing and stable accumulation of the box C/D snoRNAs (Caffarelli et al., 1996; Cavaille and Bachellerie, 1996; Watkins et al., 1996; Darzacq and Kiss, 2000).

The 15.5kD protein binds the terminal box C/D core motif in the absence of the other core proteins and initiates snoRNP assembly (Watkins et al., 2000). Nop56p and Nop58p (alternatively Nop5p) are highly related snoRNP core proteins also required for ribosome biogenesis (Wu et al., 1998; Lafontaine and Tollervey, 1999; Lafontaine and Tollervey, 2000; Newman et al., 2000). However, only Nop58p is essential for snoRNA biosynthesis and does not require fibrillarin for RNA binding. In contrast, Nop56p is not necessary for box C/D snoRNA accumulation and its binding to the snoRNA does require fibrillarin. Fibrillarin is the fourth core protein and all available biochemical and genetic evidence indicates that it is the methylase enzyme (Tollervey et al., 1993; Wang et al., 2000; Galardi et al., 2002; Omer et al., 2002). Recent work has revealed the asymmetric distribution of the box C/D snoRNP core proteins upon the terminal box C/D core and internal C'/D' motifs (Cahill et al., 2002; Szewczak et al., 2002). Based upon nucleotide modification experiments, the 15.5kD core protein binds exclusively to the terminal box C/D core motif. *In vivo* crosslinking revealed that core

proteins Nop58 and Nop56 are differentially bound to the box C/D and C'/D' motifs, respectively.

Only fibrillarin, the putative methylase, is a component of both RNP complexes.

The archaeal box C/D sRNP complexes exhibit a distinctly different protein composition from that of eukaryotic box C/D snoRNPs. Only three core proteins are required for sRNP assembly. Ribosomal protein L7 functions as the archaeal homolog of 15.5kD protein and a single archaeal Nop56/58 protein takes the place of the eukaryotic Nop56p/Nop58p protein pair (Kuhn et al., 2002; Omer et al., 2002; Tang et al., 2002). An archaeal homolog of fibrillarin which exhibits high homology with its eukaryotic counterpart is the third archaeal sRNP core protein (Amiri, 1994). The crystal structure of archaeal fibrillarin has revealed an S-adenosyl methionine-binding site, consistent with its role as the methylase enzyme (Wang et al., 2000). Recently, a *Sulfolobus acidocaldarius* box C/D sRNP complex has been assembled *in vitro* using recombinant core proteins (Omer et al., 2002). This complex was shown to direct 2'-O-methylation from the terminal box C/D core motif.

In the work reported here, we have reconstituted a methylation-competent *Methanococcus jannaschii* box C/D sRNP complex which guides methylation from both the terminal box C/D core and internal C'/D' RNA motifs. We demonstrate that the terminal box C/D RNP is the minimal complex still capable of methylation. Unexpectedly, efficient methylation guided by both the terminal box C/D and internal C'/D' RNP complexes requires that they function within the full length sRNA molecule. RNA:protein binding studies demonstrated that a Nop56/58-fibrillarin complex can uniquely bind the internal C'/D' motif indicating a folded RNA structure distinct from that of the terminal box C/D core motif. In contrast to the eukaryotic snoRNP complex, RNP assembly studies demonstrated that the archaeal terminal box C/D core and internal C'/D' RNP complexes are symmetric with respect to protein distribution, requiring all three core proteins for RNP assembly on both RNA motifs. Strikingly, comparative protein binding revealed that while archaeal core protein L7 can bind both the

box C/D and C'/D' motifs of both archaeal and eukaryotic box C/D RNAs, the eukaryotic 15.5kD homolog is limited to binding only the terminal box C/D motif. Thus, the unique core protein distribution of the archaeal and eukaryotic methylation guide ribonucleoprotein particles is a consequence of the evolution in RNA-binding capabilities of the box C/D RNP core proteins.

## RESULTS

### **SRNP Core Proteins L7, Nop56/58, and Fibrillarin Bind Both the Terminal Box C/D Core and Internal C'/D' Motifs to Assemble Symmetric RNP Complexes**

To study the structure and enzymology of the box C/D RNP complex, we have reconstituted *in vitro* an archaeal box C/D sRNP using *in vitro* transcribed or synthetic RNAs and recombinant sRNP core proteins. For these studies, *Methanococcus jannaschii* sR8 was selected as the model sRNA for RNP assembly. sR8 sRNA guides the methylation of both rRNA and tRNA substrates *in vivo*. Figure 1A (upper panel) shows the predicted secondary structure of sR8 with its terminal box C/D core motif and internal box C'/D' motif. This RNA contains each of the conserved elements characteristic of box C/D small nucleolar RNAs: a 5',3' base-paired terminal stem, terminal conserved boxes C (AUGAUGA) and D (CUGA), internal boxes C' (UGAUGA) and D' (CUGA), and sRNA guide sequences positioned immediately upstream of boxes D and D'.

The sR8 sRNP complex was assembled *in vitro* by incubating radiolabeled sR8 sRNA with recombinant *M. jannaschii* sRNP core proteins L7, Nop56/58, and fibrillarin. sRNP complex formation exhibited an ordered assembly of core proteins with L7 binding first followed by Nop56/58 and then fibrillarin (Figure 1A - lower panel). Interestingly, L7 binding occurred over a wide range of temperatures whereas Nop56/58 and fibrillarin binding required elevated temperatures around 70°C (data not shown). This is not surprising since *M. jannaschii* is a thermophile that lives in high temperature environments around 80°C. Both the observed order of core protein binding and the requirement of elevated temperature for assembly are consistent with recent observations for *in vitro* assembly of a *Sulfolobus acidocaldarius* box C/D sRNP complex (Omer et al., 2002). The elevated temperature required for Nop56/58 and fibrillarin binding suggests that these two core proteins, or

perhaps the sRNA, undergo conformational changes at higher temperatures which are necessary for protein binding.

Addition of all three core proteins resulted in several large complexes (designated RNP III, lane 8). At present, we do not know the distinct composition of each of these larger RNPs but believe that incomplete RNP assembly upon one or both RNA motifs or alternative conformations of the assembled sR8 RNP generate the multiple complexes observed. Addition of Nop56/58 to the L7:sR8 RNP complex resulted in the formation of only modest amounts of higher order RNP II (lane 5). In contrast, addition of both Nop56/58 and fibrillarin in combination results in assembly of RNP III complexes in greater amounts. We believe that Nop56/58 and fibrillarin bind the L7:sR8 RNP as a dimer despite the fact that Nop56/58 alone exhibits some binding activity (see below). The highly charged nature of Nop56/58 protein when this protein is not complexed with fibrillarin results in aggregation with the L7:sR8 RNP. Thus, a significant loss of soluble RNP complex(es) is observed when only Nop56/58 is added to the L7:sR8 RNP (lane 5).

To study the structure and function of the two RNP complexes independent of each other, sR8 "halfmers" were constructed possessing either the terminal box C/D core motif or internal C'/D' motif (Figures 1B and C - upper panels). The C/D RNA contains the 5'/3' terminal stem, boxes C and D, and the guide sequence upstream of box D contained in a loop. The C'/D' RNA possesses internal boxes C' and D' as well as both guide sequences. The second, box D-associated guide sequence and terminal stem were added to the C'/D' halfmer to facilitate RNP assembly. Smaller RNAs lacking these elements assembled little RNP complex suggesting a perturbed C'/D' structure (data not shown).

Assembly of the box C/D and C'/D' RNPs was assessed in reconstitution experiments utilizing these

halfmers and various combinations of core proteins. RNP assembly for each RNA motif exhibited the same order of protein binding as the complete sR8 sRNA and the fully assembled RNP III complex contained all three sRNP core proteins (Figures 1B and 1C - lower panels). Analysis of L7 binding to sR8, C/D, and C'/D' RNAs to form RNP I complexes revealed a predominantly slower migrating complex for full-length sR8 when compared with the faster migrating complexes for the C/D and C'/D' halfmers (compare lanes 2 of lower panels 1A, 1B, and 1C). This is consistent with L7 binding both C/D and C'/D' motifs on the full-length sR8 whereas a single L7 protein binds each halfmer. Nuclease mapping of L7 binding upon sR8 has demonstrated L7 binding to both motifs (data not shown). Interestingly, subsequent binding of the Nop56/58 and fibrillarin proteins to the C/D and C'/D' halfmers to assemble RNP complexes II and III is less efficient than that observed for the full-length sR8 RNA, particularly for the terminal box C/D core motif. We suspect the smaller size of the box C/D RNA with a constrained loop structure likely affects the interaction of Nop56/58 and fibrillarin with the RNA, thus assembling less RNP II and III complexes. Collectively, these results demonstrated that both archaeal RNP complexes are symmetric with respect to core protein composition and require all three core proteins for RNP assembly.

#### **Nop56/58 and Fibrillarin Associate Through Protein:Protein Interactions and Can Bind the C'/D' Motif in the Absence of L7**

The binding, albeit limited, of Nop56/58 and fibrillarin to sR8 in the absence of L7 (Figure 1A, lane 7) suggested that these two core proteins may work as a complex capable of independently binding box C/D sRNAs. To explore this possibility further, Nop56/58 and fibrillarin were incubated with the sR8 half-molecules and RNP formation assessed by native gel electrophoresis (Figure 2A).

Despite limited protein binding, these two core proteins interacted with the internal C'/D' motif but not

the terminal box C/D core motif. Competition studies using non-radiolabeled C/D and C'/D' competitor RNAs demonstrated the specificity of this interaction, indicating a distinct structural difference presented by the two folded RNA motifs for protein binding. Although considerably weaker in binding affinity than when assembled in the complete RNP complex, these results clearly demonstrated the ability of Nop56/58 complexed with fibrillarin to bind the C'/D' motif independent of L7. This observation has implications for the evolution of RNA-binding capabilities of these core proteins and the resultant contrasting structures of archaeal and eukaryotic box C/D RNPs (see Discussion).

Protein:protein interactions between the sRNP core proteins were explored using *in vitro* "pull-down" experiments. sRNP core proteins were incubated in pairs with one of the proteins possessing a His-tag and coprecipitation of the untagged core protein with its His-tagged partner assessed by SDS-polyacrylamide gel electrophoresis (Figure 2B). No interaction between L7 and Nop56/58 or L7 and fibrillarin was noted. However, interaction between fibrillarin and Nop56/58 was observed, consistent with our belief that these two proteins likely bind the sRNA as a complex. The possibility that Nop56/58 and fibrillarin were interacting via contaminating RNA in the recombinant protein preparations was ruled out by demonstrating identical results using proteins pre-treated with RNase (data not shown). Recently, the crystal structure of the *Archeoglobus fulgidus* Nop56/58-fibrillarin complex has been reported (Aittaleb et al., 2003). This co-crystal structure of these two proteins is consistent with our pull-down experiments and supports the idea that these proteins function *in vivo* as a dimer. The lack of observed interaction between L7 and Nop56/58 or fibrillarin does not rule out protein:protein interactions between these core proteins within the sRNP complex, but indicates that these interactions are not sufficiently strong to be observed in pulldown experiments in the absence of bound sRNA.

The ability of the Nop56/58-fibrillarin complex to bind the C'/D' motif in the absence of L7 raised the question as to whether L7 is ultimately displaced upon Nop56/58-fibrillarin binding. The small amount of Nop56/68-fibrillarin:sRNA complex formed and the weak binding of these proteins suggested that this is not the case. However, to assess this possibility, RNP “pull-down” experiments were carried out using His-tagged fibrillarin and the RNA halfmers (Figures 2C & 2D). Identical results were obtained for both box C/D and C'/D' RNP assembly. Each complex required all three core proteins for efficient RNP assembly and co-selection of bound C/D or C'/D' RNAs with tagged fibrillarin (Figure 2C). The absence of L7 or Nop56/58 resulted in loss of C/D and C'/D' RNA co-selection with tagged fibrillarin (Figure 2D). Interestingly, incubation of tagged fibrillarin with Nop56/58 resulted in only slight affinity selection of fibrillarin and no Nop56/58 for both RNAs (Figure 2D, lanes 6 & 8). This reflects the loss of soluble Nop56-fibrillarin complex and RNP when incubated in the presence of RNA. These results are consistent with the EMSA analysis of Figure 1A where formation of the Nop56/58-fibrillarin:C'/D' RNP complex can only be observed using radiolabeled RNA. Collectively, these results demonstrate the importance of L7 for efficient and symmetric box C/D and C'/D' RNP assembly (see Discussion).

### **The Terminal Box C/D RNP Is the Minimal Methylation Complex But Efficient Methylation Requires That the Box C/D and C'/D' RNPs Are Juxtaposed in the Full-Length RNA**

The ability of the *in vitro* assembled sR8 sRNP to guide the methylation of RNA substrates was assessed by monitoring the incorporation of  $^3\text{H-CH}_3$  donated from S-adenosyl-L-methionine (SAM) into substrate RNAs. Target substrates of 21 and 16 nucleotides in length contained sequences complementary to the terminal box C/D RNP guide sequence (D target) and the internal C'/D' guide sequence (D' target), respectively (Figure 3A). Determination of TCA-precipitable counts into the target RNAs revealed that both the terminal box C/D core and internal C'/D' motifs of the assembled

sR8 RNP guided methylation of their target RNAs during the 60 minute incubation period (Figure 3B). sR8 sRNP catalyzed the incorporation of between 25 and 30 pmoles of  $^3\text{H-CH}_3$  into each target RNA. The fact that 10 pmoles of L7 were used in the sRNP assembly reaction indicates that multiple rounds of methylation were catalyzed by the sRNP complexes. Similar to *in vitro*-guided methylation by the *Sulfolobus acidocaldarius* box C/D sRNP core complex (Omer et al., 2002), an elevated incubation temperature of  $68^\circ\text{C}$  was required for 2'-O-methylation activity (data not shown). Target RNAs possessing a methyl group at the 2'-O-ribose position of the designated nucleotide prior to incubation with the assembled sRNP (D- $\text{CH}_3$  and D'- $\text{CH}_3$ ; Figure 3A) showed no incorporation of  $^3\text{H-CH}_3$ . Incubation of both target RNAs in the same methylation reaction correspondingly increased the incorporation of  $^3\text{H-CH}_3$  into RNA and electrophoretic analysis of these substrates demonstrated  $^3\text{H-CH}_3$  incorporation into each RNA was blocked when either substrate was previously methylated at the designate nucleotide (Figure 3C). At present, we do not know if both RNAs can be methylated simultaneously or if binding of one substrate precludes binding of the second.

The ability of the terminal box C/D core and internal C'/D' RNP complexes to independently guide methylation of their respective target RNAs was subsequently examined. Strikingly, guided methylation from each assembled half molecule RNP was adversely affected (Figure 3C). The level of 2'-O-methylation for the terminal box C/D RNP complex was reduced to approximately one third of that observed for this complex in full-length sR8. The assembled C'/D' RNP was completely inactive in guiding methylation. Incubation of the two RNP complexes together in *trans* as separate RNPs did not restore the methylation activity of either RNP (see Figure 4B). These results revealed two important catalytic features of the archaeal box C/D sRNP complex. First, the minimal RNP structure capable of directing nucleotide-specific 2'-O-methylation, albeit at reduced levels, is the terminal box C/D RNP. Second, maximal methylation efficiency of both the box C/D and C'/D' RNP

complexes requires that they be juxtapositioned within the full-length sRNA molecule.

### **Mutations in Conserved Box Elements Impair Methylation from Both Guide Regions**

The importance of positioning the two RNP complexes within the full-length sRNA for optimal methylation activity was further examined in sR8 mutagenesis experiments. Mutations were made in the critical GA dinucleotides of each box element (Figure 4A). These nucleotides are important for K-turn formation and L7 binding to the box C/D core motif (Klein et al., 2001; Kuhn et al., 2002). Each mutant sRNA assembled with sRNP core proteins was assayed for methylation activity guided from each motif (Figure 4B). Mutation of box C reduced methylation guided from the terminal box C/D motif as expected, but also severely disrupted methylation guided by the C'/D' motif. Similar results were seen also for the box C' mutation. Mutations in box D and D' sequences also affected methylation of both target RNAs, but resulted in greater inhibition of methylation guided by the mutated C/D or C'/D' motifs. These results confirmed the critical nature of positioning the box C/D and C'/D' RNPs within the full-length sRNA for obtaining efficient methylation activity.

The effects of the box C/D and C'/D' mutations upon sRNP assembly were examined to determine if the lack of methylation activity in the non-mutated motif was due to lack of RNP formation (Figure 4C). Mutant C, D, C', and D' sR8 sRNAs were incubated with the sRNP core proteins and assembled RNP complexes resolved on native polyacrylamide gels. All four mutant sRNAs bound all three sRNP core proteins and assembled higher order RNP III complexes. Interestingly, closer inspection of RNP assembly indicated differences in complex formation when the terminal box C/D core motif was mutated as opposed to the internal C'/D' motif. Binding of L7 to the box C or D mutants resulted in a faster migrating RNP I complex as compared to the C' and D' mutants (Figure 4C). The different migration of these RNP I complexes indicated that L7 is primarily binding the C'/D' motif. However, reduced methylation is still guided from the box C/D core RNP, particularly with the box C mutant,

suggesting some RNP assembly on the C/D motif. In contrast, the C' and D' mutant sRNAs bind two L7 proteins as evidenced by slower migrating RNP complexes. Indeed, this has been confirmed with L7 titration experiments (data not shown). Despite L7 binding at both RNA motifs and subsequent RNP assembly which mirrors formation of the wild type RNP complex (Figure 1), methylation is adversely affected at both the mutated C'/D' and non-mutated C/D motifs. Therefore, loss of methylation activity at the non-mutated motifs is not a consequence of failure to assemble the RNP complex at the non-mutated motif but due to probable alterations in sRNA and/or sRNP structure resulting from crosstalk between the two RNP complexes.

#### **Archaeal L7 Binds Cooperatively to the Box C/D and C'/D' Motifs.**

The importance of juxtaposed RNPs in guided methylation prompted us to examine the possible cooperative nature of box C/D and C'/D' RNP assembly. Since L7 binding initiates sRNP assembly, the nature of its binding to the two RNA motifs was examined. Titration of sR8 with increasing concentrations of L7 revealed the formation of two RNP complexes (Figure 5A). Subsequent footprinting analysis of the sR8 titration demonstrated that both the C/D and C'/D' motifs are bound by L7 (data not shown). The slowest migrating or third apparent RNP in lane 2 is seen with excess concentrations of L7 protein and results from non-specific binding of L7 to RNA. The strength and cooperative nature of L7 binding to sR8 was also assessed using a filter binding assay. L7 association with sR8 revealed a sigmoidal binding curve, indicative of cooperative binding (Figure 5B). From this data, dissociation constants (K<sub>d</sub>) of 9 and 19 nM were calculated. At this time, we are not able to determine which K<sub>d</sub> value corresponds to L7 recognition of a specific RNA motif. Hill plot analysis confirmed the cooperative nature of L7 binding with a determined slope of 1.8 (slopes greater than 1 indicate positive cooperativity).

L7 binding to the individual box C/D and C'/D' motifs was then assessed (Figures 5C and D). L7 bound both sR8 half molecules, with higher affinity noted for the box C/D core motif. Calculation of binding affinities yielded K<sub>d</sub> values of 10 and 54 nM for the box C/D and C'/D' motifs, respectively. Interestingly, the L7 binding affinities to the two motifs relative to each other were affected when the motifs were no longer positioned in the full-length sRNA. A significant increase in one K<sub>d</sub> was noted (9 and 19 nM versus 10 and 54 nM) when the two motifs bound L7 independently. This alteration in relative binding affinities is consistent with the positive cooperativity of L7 association seen with full-length sR8 sRNA.

#### **Archaeal L7 and Eukaryotic 15.5kD Core Protein Homologs Exhibit Different RNA-Recognition Specificities for Box C/D and C'/D' RNA Binding.**

The archaeal box C/D sRNP complex exhibits a symmetric core protein distribution on the box C/D and C'/D' motifs. This is in contrast to that reported for the eukaryotic snoRNP where the eukaryotic core proteins are differentially distributed between the two RNA elements (Cahill et al., 2002; Szewczak et al., 2002). Specifically, the eukaryotic 15.5kD core protein binds only the terminal box C/D core motif of the snoRNAs. To compare RNA-recognition specificities of the archaeal and eukaryotic homologs of this core protein, the binding of archaeal L7 to eukaryotic U15 snoRNA was examined first. Strikingly, like its interaction with archaeal sR8 sRNA, archaeal L7 protein bound both the box C/D and C'/D' motifs of U15 snoRNA (Figure 6A). Binding analysis revealed a sigmoidal curve with calculated K<sub>d</sub> values of 5 and 8 nM and Hill plot analysis was consistent with the cooperative nature of L7 binding yielding a slope of 2 (Figure 6B).

The binding of 15.5kD, the eukaryotic homolog of archaeal L7, was assessed next for its interactions with eukaryotic and archaeal box C/D RNAs. As previously reported (Szewczak et al., 2002), 15.5 kD

protein bound only once to U15 snoRNA (Figure 6C) at the box C/D motif. Consistent with single-site binding, the 15.5kD binding curve was not sigmoidal (Figure 6D). 15.5kD protein was then bound to archaeal sR8 sRNA. This eukaryotic core protein also bound the full-length archaeal RNA only once at the terminal box C/D core motif (Figure 6E-G). Collectively, these binding studies demonstrate the different RNA-recognition specificities of the L7 and 15.5kD core protein homologs. These differences in RNA-binding capabilities provide a biochemical rationale for the different distribution of this protein observed in the archaeal versus eukaryotic box C/D RNPs (see Discussion).

## DISCUSSION

Development of an *in vitro* RNP assembly system provides the opportunity to biochemically dissect the structure and function of this RNA:protein enzyme. The reconstituted archaeal sR8 box C/D sRNP complex guides site-specific nucleotide modification from both the terminal box C/D and internal C'/D' RNP complexes. The terminal box C/D complex is the minimal RNP capable of methylation. However, efficient enzyme activity of the terminal core complex as well as the internal C'/D' RNP requires that both complexes function within the full-length sRNA molecule. Archaeal sRNAs bind all three core proteins at both the terminal box C/D and internal C'/D' motifs, thus establishing a symmetric sRNP. This is in contrast to the eukaryotic snoRNP where four core proteins bind differentially to the two RNA motifs and create an asymmetric snoRNP. Our protein binding studies have demonstrated the different binding capabilities of the eukaryotic 15.5kD and archaeal L7 core protein homologs. Thus, the structural distinctness of the archaeal and eukaryotic box C/D RNP is a consequence of the evolved RNA-recognition specificities of the core proteins.

Assembly of the terminal box C/D core and internal C'/D' RNP complexes requires the ordered addition of the sRNP core proteins as previously observed for a *Sulfolobus acidocaldarius* box C/D sRNP complex (Omer et al., 2002). Omer and coworkers reported the formation of multiple RNPs when L7 was bound to the sRNA. These differently migrating RNPs on native polyacrylamide gels were attributed to conformational isomers of the L7:RNA complex. We now demonstrate that L7 binds to both the box C/D and C'/D' motifs in a cooperative manner, thus providing an explanation for the multiple bands previously observed. The exact nature of this cooperativity is not known but may reflect induced changes in sR8 structure following initial L7 binding, and/or protein:protein interactions between the two L7 molecules.

Strikingly, the differential binding of Nop56/58 and fibrillarin to the C/D and C'/D' halfmers demonstrates the structural distinctness of the two RNA motifs. While the canonical box C and D sequences are well conserved in the C'/D' motifs of Archaea, considerable variation is observed in the C'/D' sequences of eukaryotic snoRNAs (Kiss-Laszlo et al., 1998; Omer et al., 2000). Of particular note is the lack of strict conservation of the G:A nucleotides of the C'/D' motif. These bases and their hydrogen-bonding interactions are crucial for establishing the K-turn structure of the box C/D motif and L7 binding (Klein et al., 2001; Kuhn et al., 2002). The lack of G:A nucleotides in some C' and D' sequences suggests that the C'/D' motif does not adopt a canonical kink-turn structure. Observations made by Cahill and coworkers (2002) have also suggested that G:A nucleotides in box D' are not required, indicating that hydrogen bonding across the asymmetric loop may not be critical for establishing C'/D' motif structure. At the least, available evidence suggests that the C'/D' motif is a variation of the classical K-turn fold.

Order of assembly experiments indicate that Nop56/58 associates with the sRNA following L7 association and is then joined by fibrillarin (Omer et al., 2002; this study). However, it is clear that Nop56/58 binding is enhanced in the presence of fibrillarin and Nop56/58 can bind the C'/D' motif in the absence of L7 when fibrillarin is present. We believe that these two core proteins likely associate prior to binding the RNA. Our pull-down experiments demonstrate interaction between these two core proteins which is consistent with this idea. In addition, the co-crystal structure of this protein complex has recently been solved and clearly shows that Nop56/58 and fibrillarin form a heterodimer in the absence of the sRNA substrate (Aittaleb et al., 2003). These proteins forming a complex for RNA binding is consistent with the core protein composition of the eukaryotic C'/D' RNP (see below).

Establishment of an *in vitro* assembly system for the archaeal box C/D sRNP complex facilitates a

biochemical dissection of the structure and function of the RNA:protein enzyme, overcoming the inability to genetically manipulate the components of this archaeal complex *in vivo*. The *in vitro* assembly system also offers other distinct advantages. It is a simple system allowing the use of the minimal components required for enzyme activity and it also permits quantitative assessment of the methylation reaction. Equally important is the ability to uncouple sRNP biosynthesis events from catalytic functions. *In vivo*, the sRNA/snoRNA requires a terminal box C/D RNP complex to protect and stabilize the RNA termini from trimming exonucleases (Cavaille and Bachellerie, 1996; Watkins et al., 1996). The opportunity for mutagenesis of the box C/D motif *in vivo* is therefore restricted. For *in vitro* analyses, stabilization of the RNA is not a concern.

Unexpectedly, methylation activity of the sR8 half molecules was severely reduced or completely abolished when the two RNP complexes are no longer juxtaposed within the full-length sRNA. It is not yet clear how each RNP complex affects the methylation activity of the other. It is possible that binding of core proteins at one motif affects the folded RNA structure at the second motif and/or involves protein:protein interactions between the two RNP complexes. However, the ability of both motifs to assemble complete RNP complexes with halfmer RNAs demonstrates that complete RNP assembly on each can proceed in the absence of the other. Therefore, it would appear that crosstalk interactions between the two RNPs ultimately affect the overall conformation of each complex, thus maximizing their respective methylation activities.

Interestingly, mutations in boxes C and C' of full-length sR8 affected the activity of the non-mutated motif more significantly than the mutated motif. In contrast, the box D and D' mutants had the most severe effect upon the methylation activity of the mutated motif. These observations suggest that the two nucleotide sequence elements contribute differently to methylation activity, possibly through core

protein binding. This is consistent with the protein:RNA crosslinking results of Cahill and coworkers (2002) which revealed that different core proteins associate with each box sequence element. Our results demonstrating that juxtapositioning of the two RNP complexes for efficient methylation contrasts the *in vivo* analysis of C'/D'-directed methylation activity reported by Kiss-Laszlo and coworkers (1998). Their investigation of guided methylation from a C'/D' RNP complex functioning in the absence of a box C/D core motif indicates that the terminal RNP complex is not required for methylation activity. However, it is difficult to assess methylation efficiency directed from a given RNP complex *in vivo* since reduced enzymatic activity over time could yield measurable methylation at a designate nucleotide. Thus, their observed methylation activity may reflect very low levels of guided methylation from the C'/D' RNP complex. Alternatively, these contrasting observations may reflect a difference between archaeal and eukaryotic box C/D RNA-guided methylation activity.

At present, it is not known if juxtapositioning the terminal box C/D and internal C'/D' motifs in eukaryotic snoRNAs is important for efficient methylation activities. However, two lines of evidence suggest that protein:protein interactions may occur between the asymmetrically distributed core proteins of the eukaryotic box C/D and C'/D' RNPs. First, in characterizing U14 snoRNP proteins, we isolated Nop56 as well as Nop58 using a minimal terminal box C/D core motif for affinity purification of box C/D-binding proteins (Newman et al., 2000). Since the core proteins of the eukaryotic snoRNP complex are asymmetrically distributed with Nop56 bound only to the internal C'/D' motif, this suggests that the co-isolation of Nop56 with Nop58 may be the result of protein:protein interactions among the snoRNP core proteins. Second, Watkins and coworkers (2002) have shown that mutations within stem II of the terminal box C/D affect the binding of both Nop56 and Nop58. The loss of both proteins upon alteration of only one motif again suggests crosstalk between the two RNPs.

Comparison of the archaeal and eukaryotic box C/D RNPs reveals structurally distinct complexes for this RNA:protein enzyme (Figure 8). The eukaryotic snoRNP complex is distinguished by the utilization of four core proteins as compared with three for the archaeal sRNP. A duplication event has resulted in the two Nop56/58 genes found in eukaryotes with each coding sequence evolving to produce proteins with different snoRNA-binding capacities (Dennis et al., 2001). Interestingly, the primary sequence of archaeal Nop56/58 shows similar homology with both eukaryotic Nop56 and Nop58 (data not shown). The ability of archaeal Nop56/58 and fibrillarin to recognize the C'/D' motif in the absence of L7 is strikingly similar to that observed in eukaryotes. In Archaea, L7 contributes to the binding strength of the Nop56/58-fibrillarin and this third core protein has been retained in the C'/D' RNP complex. In contrast, eukaryotic Nop56 has apparently evolved after the gene duplication event such that 15.5kD core protein is no longer necessary for Nop56-fibrillarin stability on the C'/D' motif. The inability of 15.5kD to bind either the eukaryotic or archaeal C'/D' motif indicates that this core protein is distinctly different from its archaeal homolog in RNA recognition specificity. This is in spite of the fact that the crystal structures of 15.5kD and L7 proteins are virtually superimposable (Kuhn et al., 2002). Since the 15.5kD protein is not present at the eukaryotic C'/D' motif, this core protein has apparently adapted its binding site to the terminal box C/D core motif. The structural basis for the different binding capabilities of these core protein homologs await a fine structure analysis of their interaction with the box C/D RNA. In the absence of 15.5kD, the structure of the C'/D' motif may have drifted from the canonical K-turn motif to accommodate the Nop56-fibrillarin complex. This could explain the stricter conservation of C'/D' sequences in Archaea.

Archaeal and eukaryotic box C/D RNAs assemble distinctly different RNPs despite using homologous proteins for RNP construction. The more simple and symmetric archaeal complex has evolved in eukaryotes to a more complex and asymmetric particle. Strikingly, this progression from simple and

symmetric to complex and asymmetric may reflect a general evolutionary progression of RNA processing/modification systems that are shared by both Archaea and Eukarya (Li et al., 1998; Trotta et al., 1997; Kleman-Leyer, et al., 1997). The tRNA splicing machinery of *M. jannaschii* is a tetramer composed of four identical subunits that recognize the two splice sites and carry out both endonuclease and ligase functions. This same processing complex in eukaryotes is again a tetramer but composed of four different proteins. Despite their homology with the archaeal proteins, the eukaryotic subunits have evolved specialized functions and even exhibit a different mechanism for splice site recognition. Analysis of eukaryotic box C/D snoRNAs suggests that at least some species, in addition to their asymmetric RNP organization, have also undergone a similar specialization of RNP function. Such an example is U14 snoRNA. Characteristic of box C/D snoRNAs, U14 guides rRNA methylation from its terminal box C/D RNP. However, the internal C'/D' element functions as an RNA chaperone to direct critical cleavage events in pre-rRNA processing. The recent observation in *Drosophila melanogaster* that the methylation activity and chaperone function of U14 are contained on two different snoRNA species suggests that RNP specialization may have progressed even further in Diptera (Yuan et al., 2003). Determination of possible differences in RNP organization reflecting the observed increased in complexity of snoRNA function awaits further investigation.

## MATERIALS AND METHODS

### DNA template construction and RNA synthesis.

Target RNA substrates and the box C/D sR8 half-molecule were purchased from Dharmacon Research, Inc.. Sequences are reported in the 5' to 3' direction and the 2'-O-methylated nucleotide is designated with an 'm' preceding the modified nucleotide.

Box C/D RNA halfmer: AAAUCGCCAAUGAUGAAACGUAUGAGCACUGAGGCGAUUU

D target: CUGAUGCUCAUACGGUCUGCU

D' target: GCUCAAAGCCAAUCGC

D-CH<sub>3</sub>: CUGAUGCUmCAUACGGUCUGCU

D'-CH<sub>3</sub>: GCUCAAAmGCCAAUCGC

Longer RNAs were synthesized by *in vitro* transcription using DNA templates. Full-length sR8 DNA was amplified from *Methanococcus jannaschii* genomic DNA and mutant sR8 templates were generated by PCR amplification from this full-length DNA template. Human U15A snoRNA template was amplified from plasmid pBS-U15 (Watkins et al., 1996). RNA transcripts were synthesized using the RiboMAX Large Scale RNA Production System- T7 (Promega) according to manufacturer's protocol. RNAs were gel-purified and 5' labeled with T4 polynucleotide kinase and  $\gamma$  <sup>32</sup>P-ATP.

### PCR Primer Pairs for amplification of DNA templates.

Wild-type sR8 DNA template (1+2); C mut (2+3); C' mut (1+4); D mut (1+5); D' mut (1+6); C'/D' RNA halfmer (7+8); U15 (9+10).

(1.) CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTG

(2.) AAATCGCCTCAGTGCTCATAACGG

(3.) CTAATACGACTCACTATAGGCCAAATCGCCAATCCTGACGATTG

- (4.) AAATCGCCTCAGTGCTCATAACGGTTCAGGACAGAC
- (5.) AAATCGCCAGAGTGCTCATAACGG
- (6.) AAATCGCCTCAGTGCTCATAACGGTTCATCACAGACAGAGC
- (7.) CTAATACGACTCACTATAGGCCTCTGGCGATTGGCTTTGCTGAGTC
- (8.) TCTGGAGTGCTCATAACGGTTCATC
- (9.) CTAATACGACTCACTATAGGCCCTTCGATGAAGAGATGATGACG
- (10.) CCTTCTCAGACAAATGCCTCTAAG

#### **Cloning, expression, and purification of proteins.**

Recombinant, GST-tagged, mouse 15.5kD protein was expressed and purified as previously described (Kuhn et al., 2002). Genes encoding L7, Nop56/58, and fibrillarin proteins were amplified from *M. jannaschii* genomic DNA using primer pairs which placed restriction sites on the 5' and 3' sides of the template (NdeI/BamHI for L7 and fibrillarin and NcoI/BamHI for Nop56/58) for subsequent cloning. Following restriction digestion, protein-coding templates were ligated into a pET28a vector (Novagen) resulting in N-terminally His-tagged L7 and fibrillarin and untagged Nop56/58. Proteins were expressed in Rosetta (DE3) cells (Novagen) at 37°C for 3 hours (L7 and Fibrillarin) or 15°C overnight (Nop56/58). His-tagged proteins were purified by nickel-affinity chromatography using “His-bind” Resin (Novagen) according to the manufacturer's protocol. Affinity tags were removed with thrombin cleavage overnight at 4°C. Nop56/58 was purified (to approximately 80% homogeneity) by cation-exchange chromatography using SP-Sepharose Fast-Flow Resin (Sigma) and bound protein eluted with a NaCl gradient. Isolated archaeal proteins were dialyzed against buffer D (20 mM Hepes pH 7, 100 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 20% glycerol) at 4°C overnight.

### **Protein-RNA interactions analyzed by EMSA and filter-binding.**

RNP complexes were assembled under the following conditions. Radiolabeled RNA (0.2 pmoles) was incubated with increasing concentrations of L7 or GST-tagged 15.5kD for binding analysis. L7 RNP complex formation was assessed by titrating L7 (8 to 600nM) in 20  $\mu$ l reactions supplemented with 10  $\mu$ g tRNA in binding reaction buffer (20mM HEPES pH 7, 150mM NaCl, 0.75mM DTT, 1.5mM MgCl<sub>2</sub>, 0.1mM EDTA, 10% glycerol). Complexes were assembled by incubation at 70°C for 10 min. GST-15.5kD complexes were assembled at 30°C for 30 min under the same binding conditions with increasing protein concentrations from 0.6 to 24 $\mu$ M. Formation of higher order archaeal RNP complexes included 10 pmoles L7, 32 pmoles Nop56/58 and 33 pmoles fibrillarlin added either individually or in different combinations. Complexes were resolved by electrophoretic mobility-shift analysis (EMSA) on 4 or 6% phosphate-buffered polyacrylamide gels (pH 7) containing 2% glycerol and RNPs visualized by autoradiography. RNP complexes for competition analysis were assembled under the same conditions as above but with 64 pmoles of Nop56/58 and no L7. Unlabeled competitor RNAs (C/D or C'/D' RNA halfmers) were added at 1000-fold molar excess and RNP complexes resolved by native gel electrophoresis. L7 and 15.5kD binding affinity with sR8 and U15 was assessed using a nitrocellulose filter-binding assay. Binding conditions were the same as electrophoretic mobility-shift analyses with the following modifications: 50  $\mu$ l reactions were assembled with 0.1nM radiolabeled RNA and L7 (2.5 to 58nM) or 15.5kD (0 to 4  $\mu$ M). Samples were applied to a nylon membrane using a dot blot apparatus, washed, and dried. Bound radioactivity was visualized using a Molecular Dynamics Model 425F PhosphorImager and quantified using the ImageQuant v3.3 software package. Calculations of dissociation constants and cooperativity measurements were performed using the Prism v3.00 software package (GraphPad).

**Protein:protein interactions analyzed by *in vitro* co-selection.**

Box C/D sRNP core proteins were incubated at equimolar concentrations in buffer A (20mM HEPES, 100mM NaCl, 1mM MgCl<sub>2</sub>) for 30 minutes at 23°C. 20% of the protein mixture was removed and designated applied sample while the remainder was incubated with 15 µl of "His-bind" resin (Novagen). Protein-bound resin was washed six times with buffer A supplemented with 60 mM imidazole, 0.1% Triton X-100, and 0.025% SDS. Bound proteins were eluted with buffer A containing 1M imidazole. Applied and eluted protein samples were resolved on 14% SDS-polyacrylamide gels and visualized by Coomassie Blue staining. Selected protein samples were treated for 1 hr at 23°C with .064 and 32 units of RNases A and T1, respectively, prior to affinity selection on His-bind resin. RNP pulldown experiments were accomplished by assembling C/D and C'/D' RNP complexes in binding buffer minus EDTA and at RNA concentrations of 90ng/ul. RNP complex selection was accomplished using His-tagged fibrillarlin and His-bind resin as detailed above. Portions of the RNP eluates were phenol extracted, ethanol precipitated, and RNA analyzed on 10% polyacrylamide-7M urea gels.

***In Vitro* Methylation.**

sRNP complexes were assembled as described for EMSA analysis but in 80 µl reaction volumes containing 52 pmoles of RNA. Assembled RNP complexes were mixed with 720 pmoles anti-sense target RNA, 360 pmoles SAM (S-adenosyl-L-methionine, dihydrogen sulfate, Calbiochem), and 1.6 µCi <sup>3</sup>H-SAM (55 Ci/mmol, ICN Biomedicals) in a final reaction volume of 120 µl in binding reaction buffer. Reactions were incubated at 68°C and 20 µl aliquots removed at 0, 10, 20, 40 and 60 minutes for TCA precipitation. Aliquots were spotted onto Whatman 3MM filters, dried, precipitated onto filters with 10% TCA for 15 minutes at 4°C, and then washed (x3) with 5% TCA at 23°C for 1 hr. Filters were dried and <sup>3</sup>H-CH<sub>3</sub> incorporation determined by scintillation counting. Assays performed

in triplicate consistently revealed small mean deviations of determined counts. Methylated RNA targets were also phenol-extracted, ethanol precipitated, resolved on 12% polyacrylamide-7M urea gels, and visualized by radiography using BioMax intensifying screens and film (Kodak).

## REFERENCES

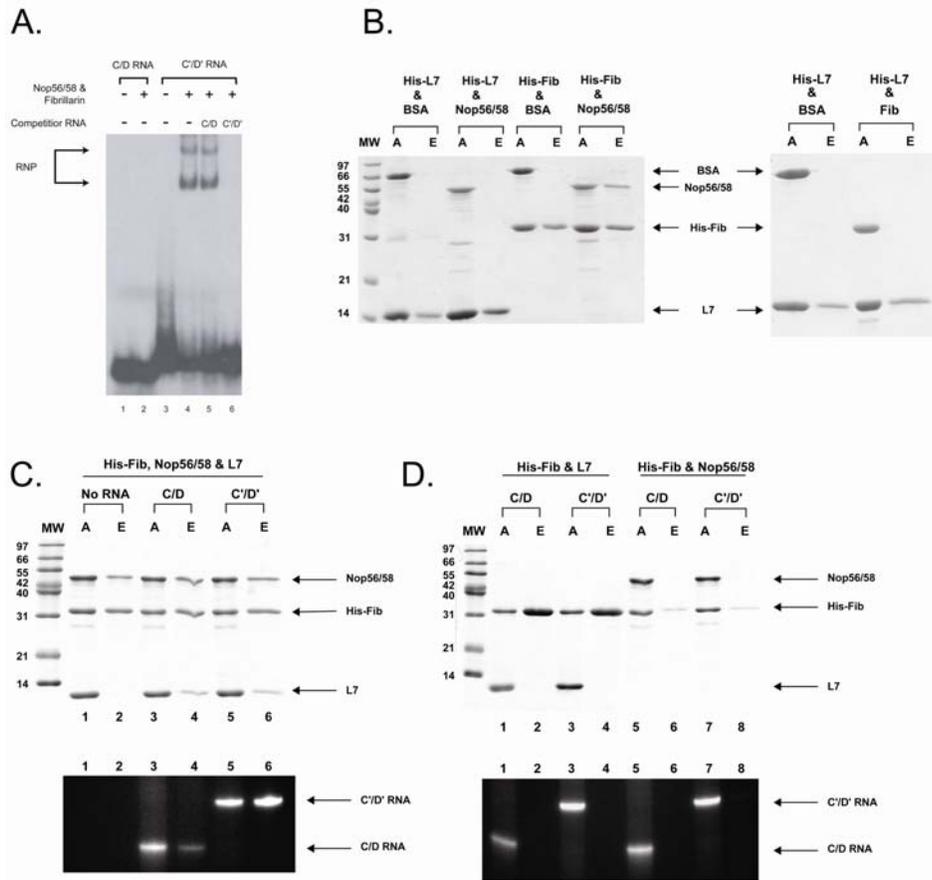
- Aittaleb, M., Rashid, R., Chen, Q., Palmer, J., Daniels, C., and Li, H. (2003) Structure and function of archaeal box C/D sRNP core proteins. *Nat. Struct. Biol.* **10**, 256-263.
- Amiri, K. (1994) Fibrillar-like proteins occur in the domain *Archaea*. *J. Bacteriol.* **176**, 2124-2127.
- Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie* **84**, 775-790.
- Caffarelli, E., Fatica, A., Prislei, S., DeGregorio, E., Fragapane, P., and Bozzoni, I. (1996) 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.* **15**, 1121-1131.
- Cahill, N., Friend, K., Speckman, W., Li, Z., Terns, R., Terns, M. and Steitz, J. (2002) Site-specific cross-linking analyses reveal an asymmetric distribution for a box C/D snoRNP. *EMBO J.* **21**, 3816-3828.
- Cavaille, J. and Bachellerie, J. P. (1996) Processing of fibrillar-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure. *Biochimie* **78**, 443-456.
- Cavaille, J. and Bachellerie, J. P. (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nuc. Acids Res.* **26**, 1576-1587.
- Darzacq, X. and Kiss, T. (2000) Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5',3'-terminal stem structure. *Mol. Cell. Biol.* **20**, 4522-4531.
- Dennis, P., Omer, A. and Lowe, T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.* **40**, 509-519.
- Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C., and Bozzoni, I. (2002) Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA *in vitro*. *Mol. Cell. Biol.* **22**, 6663-6668.
- Gaspin, C., Cavaille, J., Erauso, G. and Bachellerie, J.-P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.* **297**, 895-906.
- Jady, B. and Kiss, T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J.* **20**, 541-551.

- Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20**, 3617-3622.
- Kiss, T. (2002) Small nucleolar RNAs: An abundant group of non-coding RNAs with diverse cellular functions. *Cell* **109**, 145-148.
- Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. and Kiss, T. (1996) Site-specific ribose methylation of pre-ribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077-1088.
- Kiss-Laszlo, Z., Henry, Y. and Kiss, T. (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* **17**, 797-807.
- Klein, D., Schmeing, T., Moore, P. and Steitz, T. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214-4221.
- Kleman-Leyer, K., Armbruster, D. and Daniels, C. (1997) Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell* **89**, 839-847.
- Kuhn, J., Tran, E. and Maxwell, E. S. (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nuc. Acids Res.* **30**, 931-941.
- Lafontaine, D. and Tollervey, D. (1999) Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* **5**, 455-467.
- Lafontaine, D. and Tollervey, D. (2000) Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol. Cell. Biol.* **20**, 2650-2659.
- Li, H., Trotta, C. and Abelson, J. (1998) Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* **280**, 279-284.
- Newman, D., Kuhn, J., Shanab, G. and Maxwell, S. (2000) Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6**, 861-879.
- Omer, A., Lowe, T., Russell, A., Ebhardt, H., Eddy, S., and Dennis, P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science* **288**, 517-522.
- Omer, A., Ziesche, S., Ebhardt, H. and Dennis, P. (2002) *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Nat. Acad. Sci. USA* **99**, 5289-5294.
- Runnegar, B. (1994) Proterozoic eukaryotes: evidence from biology and geology. *In Early Life on Earth*, In Bengtson, S (ed.) Columbia Press, New York, pp. 287-297.

- Smith, C. and Steitz, J. A. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell* **89**, 669-672.
- Speckman, W., Li, Z., Lowe, T., Eddy, S., Terns, R., and Terns, M. (2002) Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus. *Curr. Biol.* **12**, 1-5.
- Szewczak, L. W., DeGregorio, S., Strobel, S. and Steitz, J. (2002) Exclusive interaction of the 15.5kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem. and Biol.* **9**, 1095-1107.
- Tang, T., Rozhdestvensky, T., d'Orval, B., Bortolin, M., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J.P., Brosius, J., and Huttenhofer, A. (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nuc. Acids Res.* **30**, 921-930.
- Terns, M. and Terns, R. (2002) Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gen. Express.* **10**, 17-39.
- Tollervey, D. (1996) Small nucleolar RNAs guide ribosomal RNA methylation. *Science* **273**, 1056-1057.
- Tollervey, D., Lehtonen, H., Jansen, R., Kern, H. and Hurt, E. C. (1993) Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* **72**, 443-457.
- Trotta, C., Miao, F., Arn, E., Stevens, S., Ho, C., Rauhut, R., and Abelson, J. (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active sites subunits homologous to the archaeal tRNA endonucleases. *Cell* **89**, 849-858.
- Tyc, K. and Steitz, J. (1989) U3, U8, and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J.* **8**, 3113-3119.
- Tycowski, K., You, Z., Graham, P. and Steitz, J. (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Cell* **92**, 629-638.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331-1342.
- Wang, H., Boisvert, D., Kim, K., Kim, R. and Kim, S.H. (2000) Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J.* **19**, 7-17.
- Watkins, N., Leverette, R., Xia, L., Andrews, M. and Maxwell, S. (1996) Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA* **2**, 118-133.

- Watkins, N., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Roshbash, M., Branlant, C., and Luhrman, R. (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457-466.
- Wu, P., Brockenbrough, J., Metcalfe, A., Chen, S. and Aris, J. (1998) Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast. *J. Biol. Chem.* **273**, 16453-16463.
- Yuan, G., Klambt, C., Bachellerie, J.P., Brosius, J., and Huttenhofer, A. (2003) RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nuc. Acids Res.* **10**, in press.





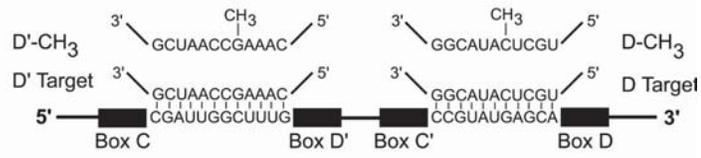
**Figure 2. Nop56/58 and fibrillarin interact via protein:protein interactions and can bind the internal C'/D' motif in the absence of core protein L7.**

**A.)** Nop56/58 and fibrillarin core proteins bind specifically to the C'/D' motif in the absence of bound L7 core protein. Nop56/58 and fibrillarin were incubated with radiolabeled box C/D or C'/D' RNA. Assembled RNP complexes were resolved on a native polyacrylamide gel and visualized by autoradiography. For competition experiments, a 1000-fold molar excess of non-radiolabeled box C/D or C'/D' RNA was added to the RNP assembly mixture. **B.)** *In vitro* co-purification demonstrates that Nop56/58 interacts with fibrillarin. sRNP core proteins were incubated in equimolar concentrations for 30 minutes at room temperature. 20% of the protein mixture was removed as the applied sample (A) and the remainder was applied to affinity resin. Bound proteins were eluted (E), resolved on 14% SDS polyacrylamide gels and visualized by Coomassie Blue staining. Incubated protein combinations are indicated above each gel with protein migration positions and molecular weight markers indicated to the side. **C. & D.)** Pull-down analysis demonstrates the presence of L7 in the C'/D' RNP complex. His-tagged fibrillarin, Nop56/58, and L7 were incubated with the C/D and C'/D' RNAs in various combinations as indicated and fibrillarin affinity-selected via its His-tag. Co-isolated sRNP core proteins (upper panels C and D) and RNAs (lower panels C and D) were resolved on polyacrylamide gels and visualized by Coomassie blue and EtBr staining, respectively.

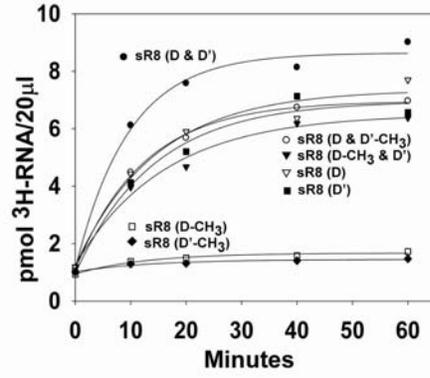
**Figure 3. *In vitro* assembled sR8 sRNP guides site-specific 2'-O-methylation from both terminal box C/D and internal C'/D' RNPs: Efficient methylation requires juxtapositioning of both RNP complexes in the full-length sRNA.**

**A.)** Schematic presentation of *M. jannaschii* sR8 base paired with target RNAs. Boxes C, D, C', and D' are designated by black boxes. Target RNAs complementary to the guide regions for the terminal box C/D RNP (D target) and internal C'/D' RNP (D' target) are shown base paired with their guide sequences. D-CH<sub>3</sub> and D'-CH<sub>3</sub> target RNAs possessing a previously 2'-O-methylated sugar at the designate nucleotide are illustrated above. **B.)** The assembled sR8 sRNP complex guides site-specific methylation of both the D and D' target RNAs. Assembled sR8 sRNP was incubated at 70°C with the indicated RNAs (in parentheses) and <sup>3</sup>H-S-adenosyl-L-methionine. At various times, aliquots of the reaction were collected, TCA-precipitated, and <sup>3</sup>H-methyl incorporation measured by scintillation counting. D-CH<sub>3</sub> and D'-CH<sub>3</sub> target RNAs served as control RNA substrates for site-specific methylation. **C.)** Methylation of D and D' RNA substrates demonstrates site-specific, 2'-O-methylation at designate nucleotides of the target RNAs. Target RNAs in various combinations indicated above the gel were resolved by electrophoresis and 2'-O-methylated RNAs revealed by radiography. Migration positions of the D and D' targets are indicated at the side. **D.)** Efficient methylation requires juxtapositioned RNPs. RNP complexes assembled upon the box C/D and C'/D' halfmer RNAs were incubated with their respective target RNAs and <sup>3</sup>H-S-adenosyl-L-methionine and assayed for methylation activity. sR8 (D) target RNA (solid square) is the control methylation level of the box C/D core RNP determined when positioned in full-length sR8 sRNP.

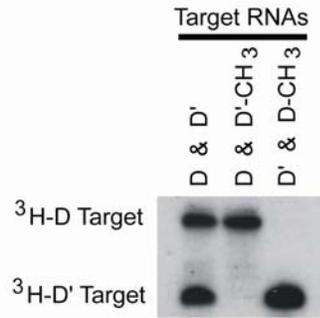
A.



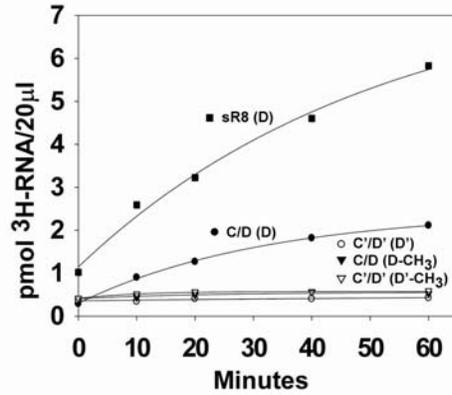
B.

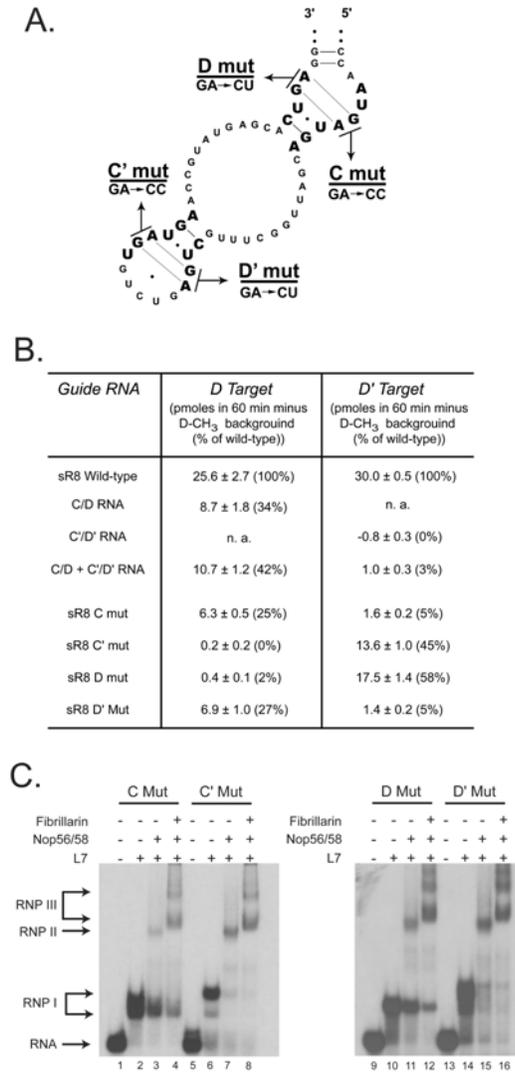


C.



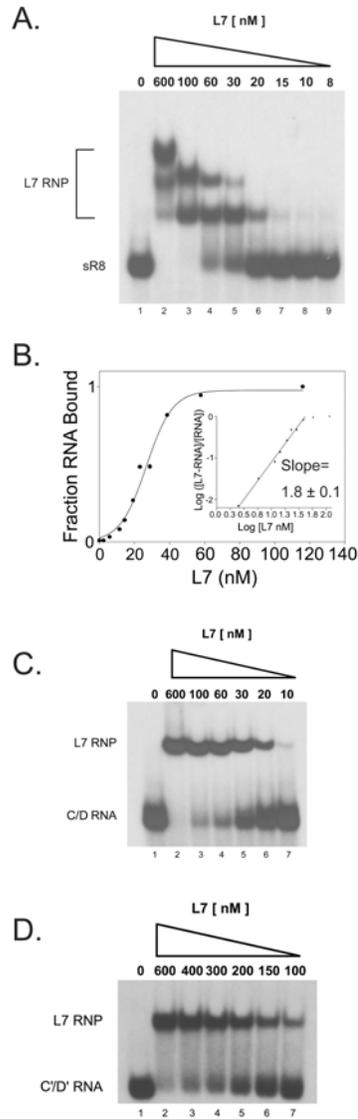
D.





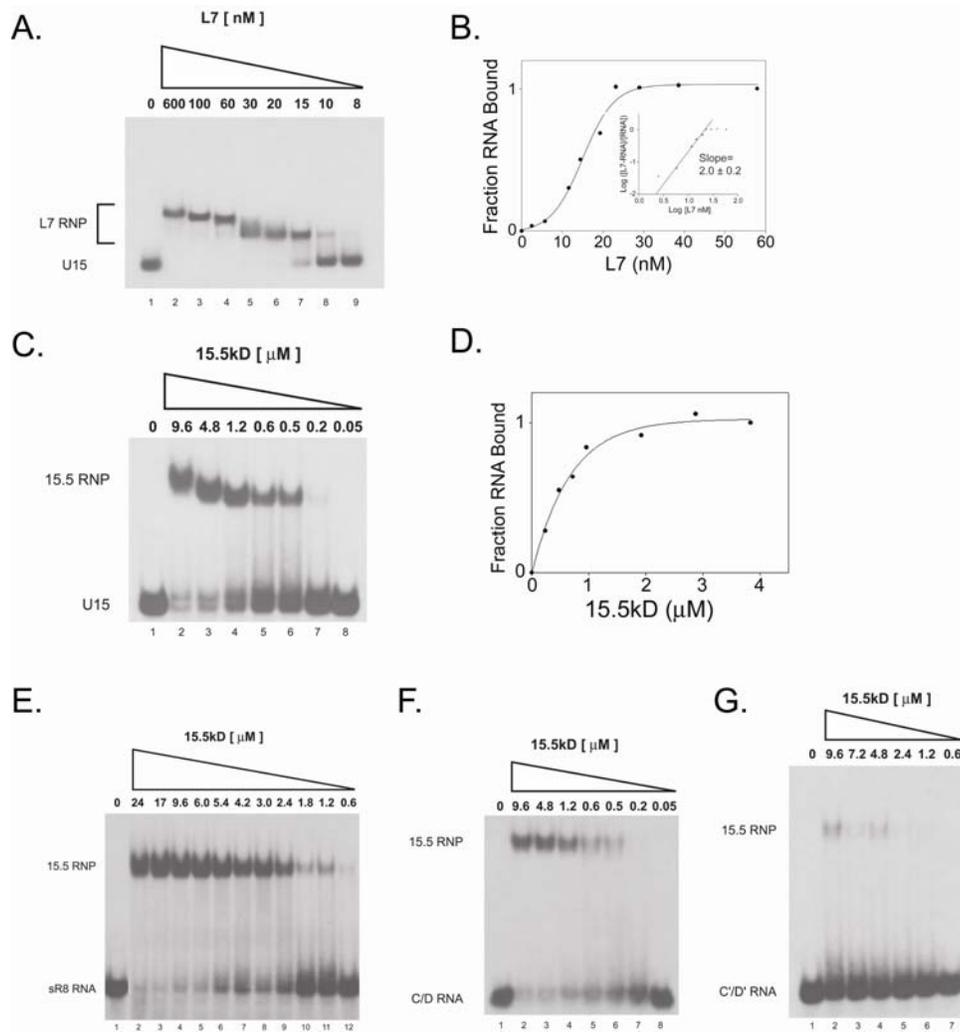
**Figure 4. Mutation of either the terminal box C/D core or internal C'/D' motifs in full-length sR8 affects guided 2'-O-methylation of both RNP complexes.**

**A.)** Schematic presentation of sR8 sRNA with various mutations in the box C, D, C', and D' sequences. The complete terminal helix is not shown for simplicity of illustration but is present in all mutant sRNAs. **B.)** Box C/D or C'/D' mutations affect methylation from both RNP complexes. Methylation efficiencies of wild type sR8, halfmer RNAs, and the full-length mutants are reported as total pmoles of <sup>3</sup>H-methyl incorporation into the target RNAs in one hour. Methylation of the control methylated D and D' target RNAs (D-CH<sub>3</sub> and D'-CH<sub>3</sub>) is subtracted as background. Numbers in parentheses indicate percent of methylation with respect to activity of the respective complexes in the full-length sR8 sRNA. **C.)** RNP complexes are assembled on sR8 box C/D and C'/D' RNA mutants. sR8 RNAs containing the individual mutants illustrated in panel A were radiolabeled and incubated with the sRNP core proteins as indicated above the gel. Assembled complexes were resolved on native polyacrylamide gels and visualized by autoradiography.



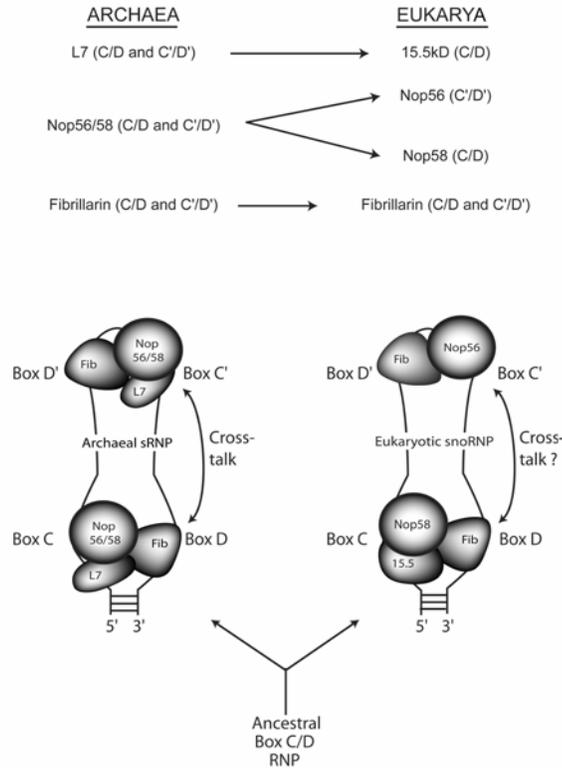
**Figure 5. Core protein L7 exhibits cooperative binding to archaeal sR8 sRNA.**

**A.)** L7 protein binds the box C/D and C'/D' motifs of sR8 sRNA. Increasing concentrations of L7 (indicated above the gel) were incubated with radiolabeled sR8. Assembled RNP complexes were resolved on native polyacrylamide gels and visualized by autoradiography. The slowest migrating L7 RNP (lane 2) is observed only at excess L7 concentrations and represents non-specific L7 binding. **B.)** L7 association with the sR8 box C/D and C'/D' motifs is cooperative. Increasing concentrations of L7 (indicated above the gel) were incubated with radiolabeled sR8, assembled RNP complex blotted to nitrocellulose membranes and bound RNP quantitated using a phosphorimager. The fraction of RNA bound in an RNP complex is plotted as a function of L7 concentration. The L7 binding data is also presented as a Hill plot (inset). **C + D.)** L7 binds to both the terminal box C/D core and internal C'/D' halfmer RNAs. Increasing concentrations of L7 were incubated with radiolabeled terminal box C/D (**C**) or internal C'/D' halfmers (**D**). Assembled RNP complexes were resolved on native polyacrylamide gels and visualized by autoradiography.



**Figure 6. Archaeal L7 binds both the C/D and C'/D' Motifs: Eukaryotic 15.5 kD protein binds only the terminal box C/D core motif.**

**A.)** Increasing amounts of *M. jannaschii* sRNP core protein L7 were incubated with radiolabeled human U15 snoRNA. Assembled RNP complexes were resolved on a native polyacrylamide gel and visualized by autoradiography. **B.)** Archaeal sRNP core protein L7 exhibits cooperativity in binding to U15 snoRNA. Binding analyses were performed as detailed in Figure 5. **C.)** Eukaryotic snoRNP core protein 15.5kD binds only the box C/D motif of U15 snoRNA. Increasing amounts of mouse 15.5kD protein were incubated with radiolabeled human U15 snoRNA. Assembled RNP complexes were resolved on a native polyacrylamide gel and visualized by radiography. **D.)** The fraction of U15 snoRNA bound in an RNP complex is plotted as a function of 15.5kD concentration. Increasing amounts of mouse 15.5kD core protein were incubated with radiolabeled, full-length archaeal sR8 sRNA **(E)** and halfmer RNAs possessing either the terminal box C/D **(F)** or internal C'/D' **(G)** motifs. Assembled RNP complexes were resolved on native polyacrylamide gels and visualized by radiography.



**Figure 7. Evolution of the box C/D RNP core proteins: Archaeal and eukaryotic box C/D RNP complexes exhibit symmetric vs. asymmetric distribution of the RNP core proteins bound to sRNAs and snoRNAs, respectively.**

Evolution of the box C/D RNP core protein's RNA-binding capabilities result in the differential distribution of core protein homologs for the archaeal and eukaryotic box C/D RNP complexes. The illustrated structure of the eukaryotic snoRNP is based upon the work of Cahill et al., (2002) and Szewczak et al., (2002).

**CHAPTER III. Evolutionary Origins of the RNA-Guided Nucleotide Modification Complexes:  
From The Primitive Translational Apparatus?**

Elizabeth Tran<sup>1</sup>, James Brown<sup>2</sup>, and E. Stuart Maxwell<sup>1\*</sup>

Department of Molecular and Structural Biochemistry<sup>1</sup>, Department of Microbiology<sup>2</sup>,  
North Carolina State University, Raleigh, North Carolina, 27695 USA

Published in Trends in Biochemical Sciences, 2004, Vol. 29, No. 7, p. 343-350.

## ABSTRACT

Eukarya and Archaea possess scores of RNA-guided nucleotide modification complexes which target specific ribonucleotides for 2'-O-methylation or pseudouridylation. Recent characterization of these RNA modification machines has yielded striking results with implications for their evolutionary origins. Demonstration that the two major classes of nucleotide modification complexes in Archaea share a common ribonucleoprotein (RNP) core element indicates that both have evolved from a progenitor RNP. The fact that this common RNP element is also found in ribosomes suggests that the origin of the progenitor RNP lies in the primitive translation apparatus. Thus, the trans-acting, RNA-guided nucleotide modification complexes of the modern RNP world appear to have evolved from cis-acting RNA/RNP elements contained in the primitive translation apparatus during transition from the ancient RNA world to the modern RNP world.

Although the first small nucleolar RNA (snoRNA) was reported more than 35 years ago [1, 2], it has only been in the last decade that the diversity of this small RNA population and their importance in ribosome biogenesis have been fully appreciated (for reviews, see [3-6]). The eukaryotic nucleolus possesses numerous snoRNAs with predictions of their full complement numbering several hundred. The snoRNAs use complementary sequences to base pair with ribosomal RNA, the site of snoRNA function being determined by snoRNA:rRNA duplex formation. A select few snoRNAs facilitate pre-rRNA folding whereas several others have been shown to play critical roles in endonucleolytic cleavage of the ribosomal RNA precursor. However, the primary role of the vast majority of snoRNAs is to guide the multitude of rRNA nucleotide modifications that occur during rRNA maturation and ribosome subunit assembly. From sequence analysis, investigators have established two major snoRNA families based upon conserved nucleotide “boxes” (Box 1). The box C/D snoRNAs guide the 2'-O-methylation of targeted nucleotides while the box H/ACA snoRNAs guide the conversion of targeted uridine nucleotides to pseudouridine. With the number of 2'-O-methylated and pseudouridylated nucleotides in mature vertebrate rRNAs each numbering approximately 100, it is easy to see how the snoRNAs have become one of the largest classes of small stable RNAs in the eukaryotic cell.

### **Nucleotide Modification Guide-RNAs Are Not Limited to Eukarya**

Archaea also use “snoRNA-like” RNAs or sRNAs to guide 2'-O-methylation and pseudouridylation of targeted rRNA nucleotides [5, 7, 8]. The presence of box C/D and H/ACA RNAs in Archaea, a second domain of life distinct from Eukarya, strongly suggests that utilization of guide RNAs for nucleotide modification is an evolutionarily ancient mechanism. Although modified archaeal rRNA nucleotides are typically fewer in number and differ in position from those in Eukarya, the common structural elements of the guide RNAs are well conserved between the two domains (Box 1). Archaeal and

eukaryotic box C/D RNAs possess terminal box C/D core and internal C'/D' motifs, both of which can guide 2'-O-methylation using adjoining rRNA-complementary sequences [7, 9, 10]. The archaeal and eukaryotic H/ACA RNAs possess pseudouridylation pockets where an unpaired uridine within the H/ACA-rRNA duplex is converted to pseudouridine [11-13].

### **The Guide RNAs Function as Ribonucleoprotein Complexes**

Eukaryotic snoRNAs and archaeal sRNAs function as ribonucleoprotein particles with enzymatic activities for nucleotide modification residing in the protein components of the RNP complex (Box 1).

The eukaryotic box C/D snoRNP includes the 4 core proteins 15.5kD or Snu13p, Nop56p, Nop58p, and the methylase Fibrillarin [14, 15]. Recent *in vitro* binding as well as *in vivo* crosslinking analyses have indicated an “asymmetric” distribution of core proteins between the terminal box C/D core and internal C'/D' RNA motifs [16, 17]. 15.5kD and Nop58p are bound to the box C/D core motif whereas Nop56p binds the box C'/D' motif. Only the methylase Fibrillarin associates with both RNA motifs. Analysis of the archaeal box C/D sRNP complex revealed a simpler protein composition and a “symmetric” core protein distribution. Three core proteins, highly similar to the eukaryotic core proteins, bind both the box C/D core and C'/D' motifs: ribosomal protein L7 (the archaeal homolog of eukaryotic 15.5kD or Snu13p), a single Nop56/58p homolog, and Fibrillarin [18-20].

Eukaryotic H/ACA snoRNPs also contain 4 core proteins: Nhp2p, Gar1p, Nop10p, and the pseudouridyl synthase Cbf5p or dyskerin [21, 22]. Little is presently known about the structure of this RNP complex or the distribution of core proteins on the conserved RNA sequences and structures characteristic of this snoRNA family. Genome analysis has identified candidate archaeal homologs for Cbf5p, Gar1p and Nop10p [21, 23]. Sequence analysis has also noted high similarity between eukaryotic Nhp2p and archaeal ribosomal protein L7 (Figure 1.). Most recently, it has been

demonstrated that L7 is indeed the functional archaeal homolog of eukaryotic H/ACA core protein Nhp2p [13]. Both these guide RNP core proteins belong to a larger family of L7 homologous proteins which include several ribosomal proteins (Figure 1).

### **The Kink Turn or “K-Turn” of the Box C/D RNAs is a Widespread and Evolutionarily Conserved RNA Motif**

Investigators have begun to dissect the structure of the guide RNAs and bound core proteins with the majority of progress coming from analysis of the box C/D RNAs and RNPs.

Initial insight into box C/D RNP structure came indirectly from structural analysis of the 15.5kD core protein bound to U4 snRNA [24]. 15.5kD protein binds both box C/D snoRNAs and spliceosomal U4 snRNA, thus serving a dual function in eukaryotes. The crystal structure of 15.5kD-U4 complex revealed an RNA motif consisting of two stems flanking an asymmetric bulge region (2 + 5 nucleotides) (Figure 2). Critical for folding of this RNA structure are two tandem, sheared G-A base pairs hydrogen-bonding across the asymmetric loop. These non-Watson Crick pairs are stacked upon the adjacent helix and the resulting structure exhibits a sharp bend or kink within the RNA, hence the designation kink-turn or “K-turn” [25] (Figure 2). With knowledge of the 15.5kD-U4 snRNP structure, it was reasoned that the 15.5kD core protein interacts with the terminal box C/D of snoRNAs in an analogous manner [15]. Analysis of core protein binding to both eukaryotic and archaeal box C/D RNAs has confirmed the importance of the K-turn motif for box C/D RNP assembly [26, 27].

Kink-turns are widespread in nature and are found in all three domains of life. In the elegant structural analysis of the archaeal *Haloarcula marismortui* ribosome, the K-turn was first defined with six examples being found in the 50S subunit [25]. These K-turns are structural variants of the consensus structure (Figure 2) but all depend upon the tandem, sheared G-A pairs for RNA folding. As a result

of these variations, the K-turns of the archaeal 50S subunit interact with nine different ribosomal proteins, one of them being the sRNP core protein L7. In eukaryotes, K-turn motifs have been found in not only the box C/D snoRNAs, U4 snRNA and ribosomal RNA, but also in the pre-messenger RNA encoding ribosomal protein L30 [25, 28]. In bacteria, K-turn motifs have been observed or predicted in 23S rRNA and the leader sequences of mRNAs regulated by the T box and S box antitermination mechanisms [29].

### **Archaeal Box C/D and H/ACA sRNAs Share a Common RNP Core**

Although dissection of H/ACA RNP structure has lagged behind that of the box C/D complex, recently light has been shed upon the organization of this RNA:protein enzyme with rather striking implications for the evolution of the nucleotide modification complexes. The first archaeal H/ACA sRNAs have been identified along with the demonstration that ribosomal protein L7 is also a core protein of this sRNA family [12, 13]. The finding of a common core protein shared by both the box C/D and H/ACA RNPs indicated their evolutionary origins in a common ancestral complex or progenitor RNP as previously suggested [11].

The demonstration that L7 is a core protein of both archaeal sRNP families raised the question of the structure of the L7 binding site on the H/ACA sRNAs. Logic dictated that L7 should also recognize a K-turn in the H/ACA sRNAs. Mapping of L7's binding site in several archaeal H/ACA sRNAs has revealed that it does indeed bind a K-turn [13]. Several of the archaeal H/ACA K-turns are quite similar in folded structure to the internal C'/D' RNA motif of the archaeal box C/D sRNAs (Figure 2). Identification of a K-turn in archaeal H/ACA sRNAs implies that eukaryotic Nhp2p should also bind a K-turn in the eukaryotic H/ACA snoRNAs. A K-turn motif within these snoRNAs has not been identified. However, it might be premature to rule out this possibility because the K-turn motif

exhibits great variation in folded structure. For example, in the archaeal 50S ribosome subunit one variant utilizes a base triple for RNA folding whereas another requires three separate RNA strands [25]. Therefore, a definitive answer to this question will probably require biophysical analysis of the folded H/ACA snoRNAs, both unbound and bound with core proteins.

### **A Common Ancestral RNP Complex Derived From the Primitive Translation Apparatus?**

The fact that L7 serves a dual function in Archaea as both an sRNP core protein and a large subunit ribosomal protein has led several research groups to suggest that the common ancestral RNP of the RNA-guided nucleotide modification complexes had its evolutionary origins in the primitive translation apparatus (Figure 3) [3, 13, 27]. One would predict in this evolutionary pathway, that cis K-turns within the primitive translation apparatus first functioned as RNA folding elements. Recent analysis revealing that bound protein greatly stabilizes this RNA motif would suggest the early participation of proteins and formation of an RNP [30]. These RNA elements or RNPs might have played additional roles at this early stage. At some point, an RNA or RNP complex would have evolved that ultimately served as the progenitor complex of the box C/D and H/ACA RNPs. It is reasonable to suggest that this progenitor complex was an RNP because the C/D and H/ACA RNPs now share identical (archaeal L7) and highly homologous (eukaryotic 15.5kD or Snu13p and Nhp2p) core proteins (Figure 1). This progenitor RNP might have first functioned as a cis element within the primitive translation apparatus and evolved within this machinery into distinct C/D and H/ACA RNPs. Alternatively, the progenitor RNP could have first become an independent trans complex before evolving into the two structurally and functionally distinct guide RNP families.

The apparent absence of RNA-guided nucleotide modification complexes in Bacteria indicates that the progenitor RNP arose in the archaeal and eukaryal branch after divergence from Bacteria. However,

the finding of K-turns in bacterial RNAs, including rRNA, makes it plausible to now suggest that this fundamental RNA motif was likely present in the translation apparatus of organisms prior to the last common ancestor [31]. The utilization of specific protein enzymes for nucleotide modification in Bacteria has resulted in a limited number of modified rRNA nucleotides [32]. In contrast, archaeal and eukaryal ribosomal RNA possess greater numbers of 2'-O-methylated sugars and pseudouridine nucleotides. It has been suggested that the many more numerous C/D and H/ACA RNPs of Archaea and Eukarya are likely to have arisen through variation of the guide sequence [32]. Those mutant C/D and H/ACA RNAs proving beneficial to ribosome synthesis and/or function would be retained by the organism. Such a scenario would explain the lack of conservation of modified nucleotides shared between Archaea and Eukarya and even between organisms in the same domain of life. The observation that thermophilic archaeal organisms appear to have higher numbers of modified nucleotides in response to these organism's need to stabilize rRNA in extreme environments is consistent with this scenario [8].

Although the C/D and H/ACA RNPs share common or homologous core proteins, each complex has developed a distinctive protein composition unique to each RNP. The C/D- and H/ACA-specific RNP proteins are highly similar between Archaea and Eukarya indicating their presence in each complex before Archaea and Eukarya divergence. As previously suggested, some of the C/D- and H/ACA-specific core proteins are likely to have evolved from ancestral coding sequences prior to the last common ancestor and so are now found in Bacteria, Archaea, and Eukarya [32]. For example, the pseudouridine synthase Cbf5p of the H/ACA RNPs is related to the bacterial tRNA pseudouridine synthase TruB. Gene duplication in eukaryotes explains two modification enzymes, one for tRNA modification (Pus4p) and the other for rRNA modification (Cbf5p). One would anticipate that enzymatically-competent RNPs were established before variation in their respective guide sequences

produced the large array of modification complexes now observed.

### **Continuing Evolution of the RNA-Guided Nucleotide Modification Complexes**

In the last several years it has become clear that RNA-guided RNP complexes have continued to evolve in both structure and function. Eukaryotic box C/D snoRNP organization is more complex than that of Archaea. Although the core proteins of the archaeal sRNP and eukaryotic snoRNP are homologous, the number of core proteins (three vs. four) and their symmetric vs. asymmetric distribution on the C/D core and C'/D' motifs is thus far characteristic of the sRNP and snoRNPs, respectively [16, 17, 20]. The differing number of core proteins and asymmetric structure of the eukaryal snoRNP is the result of gene duplication (archaeal Nop56/58p to eukaryotic Nop56p and Nop58p), specialization of the individual core proteins, and divergence of the symmetrical halves of the ancestral RNP. Evolution of RNP structure is also seen with the additional, species-specific proteins of the U3 and U8 snoRNPs in eukaryotes [33, 34]. Less is known about evolution of H/ACA RNP structure. The number of H/ACA core proteins for eukaryotic snoRNP and archaeal sRNP is the same (four), but the RNAs themselves exhibit apparently distinctive folded structures (bipartite vs. single hairpin structures). Interestingly, the single RNA hairpin H/ACA snoRNAs of trypanosomes is strikingly similar to the archaeal H/ACA sRNAs [35]. It thus appears that establishment of the bipartite H/ACA snoRNA structure occurred after the divergence of this group of eukaryotes. It will be interesting to see if and how the H/ACA RNP organization of trypanosomes and Archaea compares with those of plants, animals, and fungi.

Eukaryotic snoRNP evolution is also seen in additional RNP functions in ribosome biogenesis.

Several snoRNA species are essential for pre-rRNA processing and ribosome biogenesis [36-38].

Others have apparent “chaperone” activity in which intermolecular base pairing between snoRNA and

rRNA has been suggested to promote specific rRNA folding or prevent the formation of non-productive secondary/tertiary structures [36, 39]. It seems likely that similar or alternative functions will also be guided by the archaeal sRNAs.

SnoRNA and sRNAs also guide modifications of additional target molecules other than rRNA. These RNA targets include archaeal tRNAs, the eukaryotic splicing U snRNAs, and apparently eukaryotic mRNAs [10, 12, 40-43]. The snoRNAs that modify the splicing snRNAs have been shown to possess both box C/D and H/ACA motifs in a single snoRNA species, thus guiding both 2'-O-methylation and pseudouridylation [41]. Signal sequences localize these species to Cajal bodies where they carry out their modification reactions; they have been designated small Cajal body RNAs or scaRNAs [44]. Even more striking and thought provoking are the snoRNA species that are expressed in a tissue-specific manner within the brain [42]. One such species reveals complementarity to the serotonin receptor messenger RNA and has been proposed to potentially regulate mRNA maturation.

Finally, evolution of guide RNP function is not limited to the guide RNPs themselves as certain core proteins have taken on new roles as “moonlighting” proteins [45]. The 15.5kD core protein of eukaryotic box C/D RNP functions as a structural protein of the U4 splicing RNP [15]. It is not clear if the origin of the U4 K-turn is the same as those of the ribosome and the snoRNPs. Interestingly, the 61K protein of the U4 snRNP exhibits significant sequence similarity with the box C/D core proteins Nop56p and Nop58p [46]. Another example of a moonlighting protein is the yeast ribosomal protein L30 that binds to K-turns both in the 26S rRNA and its own pre-messenger RNA, thus regulating pre-mRNA splicing to the mature mRNA [47].

## CONCLUDING REMARKS

The proposal that cis-acting elements important in RNA modification reactions have evolved to become trans-acting RNPs while still maintaining their original functions is not unprecedented. A similar evolutionary scenario has been proposed for the splicing RNP complexes in which present-day splicing snRNPs are derived from the folded RNA elements of self-splicing group II introns [48]. It was also suggested that the mechanism of trans-splicing may be an intermediate step in, or an offshoot of, the evolution of cis- to trans-acting RNPs. A potentially analogous evolutionary intermediate has been found in Archaea where cis-positioned box C/D elements in the pre-tRNA<sup>Trp</sup> intron guide the 2'-O-methylation of two nucleotides in the tRNA molecule [40].

Particularly intriguing is the diversity of gene organizations for the coding sequences of these trans-acting small RNAs. Some snoRNAs are transcribed directly from monocistronic or polycistronic genes. However, the vast majority of snoRNA genes, particularly in metazoan organisms, are encoded within the introns of protein coding genes and these intronic snoRNAs are processed from their pre-mRNA introns [3, 5]. The recent identification of a putative H/ACA pseudogene possessing a poly (A) tail suggests that intronic snoRNAs may be a consequence of retroposon events [49]. What insertion mechanism has resulted in the intronic snoRNAs becoming predominantly positioned within pre-mRNA introns of proteins that function in the ribosome or protein biosynthetic pathway? Does this biased positioning play a role in coordinating mRNA synthesis with ribosome biogenesis? What is the relationship between archaeal rRNA synthesis and intron splicing [50]?

Clearly, many additional questions concerning the evolution of the RNA-guided nucleotide modification complexes are unanswered. What is the total complement of guide RNAs in a given

organism? Why do the numbers of guide RNAs vary widely between Archaea and Eukarya, between different organisms within a domain, and between the different tissues of a given organism? Are these varying population sizes indicative of additional target RNAs and/or alternative functions? What role(s) do some guide RNAs play in regulating gene expression? What other moonlighting roles have the RNP core proteins assumed? Unexpected answers are sure to arise from future dissection of the structure and function of these RNA-guided nucleotide modification machines.

### **Acknowledgments**

We thank M. J. Fournier and R. Hartmann for their helpful suggestions in the critical reading of this manuscript. This work was supported by NSF Grant MCB-0215545 to ESM.

## BIBLIOGRAPHY

1. Nakamura, T. *et al.* (1968) Studies on nucleolar 4 to 6S ribonucleic acid of Novikoff hepatoma cells. *J. Biol. Chem.* **243**, 1368-1375.
2. Weinberg, R. and S. Penman. (1968) Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* **38**, 289-304.
3. Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20**, 3617-3622.
4. Terns, M. and R. Terns. (2002) Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gen. Express.* **10**, 17-39.
5. Bachellerie, J.P. *et al.* (2002) The expanding snoRNA world. *Biochimie* **84**, 775-790.
6. Fatica, A. and D. Tollervey. (2003) Insights into the structure and function of a guide RNP. *Nat. Struct. Biol.* **10**, 237-239.
7. Omer, A., *et al.* (2000) Homologs of small nucleolar RNAs in Archaea. *Science* **288**, 517-522.
8. Dennis, P. *et al.* (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.* **40**, 509-519.
9. Kiss-Laszlo, Z. *et al.* (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* **17**, 797-807.
10. Gaspin, C. *et al.* (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.* **297**, 895-906.
11. Ganot, P. *et al.* (1997) The family of box H/ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Gen. Develop.* **11**, 941-956.
12. Tang, T. *et al.* (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Nat. Acad. Sci.* **99**, 7536-7541.
13. Roshdestvensky, T. *et al.* (2003) 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. *Nuc. Acids Res.* **31**, 869-877.
14. Lafontaine, D. and D. Tollervey. (2000) Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol. Cell. Biol.* **20**, 2650-2659.

15. Watkins, N. *et al.* (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457-466.
16. Cahill, N. *et al.* (2002) Site-specific cross-linking analyses reveal an asymmetric distribution for a box C/D snoRNP. *EMBO J.* **21**, 3816-3828.
17. Szewczak, L.W. *et al.* (2002) Exclusive interaction of the 15.5kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem. and Biol.* **9**, 1095-1107.
18. Omer, A. *et al.* (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Nat. Acad. Sci.* **99**, 5289-5294.
19. Aittaleb, M. *et al.* (2003) Structure and function of archael box C/D sRNP core proteins. *Nat. Struct. Biol.* **10**, 256-263.
20. Tran, E. *et al.* (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archael box C/D and C'/D' RNPs. *EMBO J.* **22**, 3930-3940.
21. Watkins, N. *et al.* (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all box H/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4**, 1549-1568.
22. Henras, A. *et al.* (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.* **17**, 7078-7090.
23. Watanabe, Y. and M. Gray. (2000) 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. *Nuc. Acids Res.* **28**, 2342-2352.
24. Vidovic, I. *et al.* (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331-1342.
25. Klein, D. *et al.* (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214-4221.
26. Watkins, N. *et al.* (2002) Conserved stem II of the box C/D motif is essential for nucleolar localisation and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol. Cell. Biol.* **22**, 8342-8352.
27. Kuhn, J. *et al.* (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nuc. Acids Res.* **30**, 931-941.
28. Mao, H. *et al.* (1999) A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat. Struct. Biol.* **6**, 1139-1147.
29. Winkler, W. *et al.* (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* **7**, 1165-1172.

30. Goody, T. *et al.* (2004) The kink-turn motif in RNA is dimorphic, and metal ion dependent. *RNA* **10**, 254-264.
31. Penny, D. and A. Poole. (1999) The nature of the last universal common ancestor. *Curr. Opin. Genet. Dev.* **9**, 672-677.
32. Lafontaine, D. and D. Tollervey. (1998) Birth of the snoRNPs: The evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.* **23**, 383-388.
33. Dragon, F. *et al.* (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967-970.
34. Tomasevic, N. and B. Peculis. (1999) Identification of a U8 snoRNA-specific binding protein. *J. Biol. Chem.* **274**, 35914-35920.
35. Liang, X. *et al.* (2001) Identification of the first Trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. *J. Biol. Chem.* **276**, 40313-40318.
36. Peculis, B. (1997) The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Mol. Cell. Biol.* **17**, 3702-3713.
37. Tycowski, K. *et al.* (1994) Requirement for intron-encoded U22 small nucleolar RNA for 18S ribosomal RNA maturation. *Science* **266**, 1558-1561.
38. Liang, W. and M.J. Fournier. (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Gen. Develop.* **9**, 2433-2443.
39. Borovjagin, A. and S. Gerbi. (2001) Xenopus U3 snoRNA GAC-box A' and box A sequences play distinct functional roles in rRNA processing. *Mol. Cell. Biol.* **21**, 6210-6220.
40. d'Orval, B.C. *et al.* (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA<sup>Trp</sup> intron guides the formation of two ribose-methylated nucleosides in the mature tRNA<sup>Trp</sup>. *Nuc. Acids Res.* **29**, 4518-4529.
41. Darzacq, X. *et al.* (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746-2756.
42. Cavaille, J. *et al.* (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Nat. Acad. Sci.* **97**, 14311-14316.
43. Liang, X. *et al.* (2002) 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* **8**, 237-246.
44. Darzacq, R. *et al.* (2003) A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* **22**, 4283-4293.

45. Jeffery, C. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* **19**, 415-417.
46. Makarova, O. *et al.* (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6-U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.* **21**, 1148-1157.
47. Vilardell, J. and J. Warner. (1997) Ribosomal protein L32 of *Saccharomyces cerevisiae* influences the splicing of its own transcript and the processing of rRNA. *Mol. Cell Biol.* **17**, 1959-1965.
48. Sharp, P. (1991) Perspective: five easy pieces. *Science* **254**, 663.
49. Vitali, P. *et al.* (2003) Identification of 13 novel human modification guide RNAs. *Nuc. Acids Res.* **31**, 6543-6551.
50. Tang, T. *et al.* (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nuc. Acids. Res.* **30**, 921-930.

## FIGURES

### BOX 1. Eukaryotic and Archaeal RNA-Guided Nucleotide Modification Complexes

RNA-guided nucleotide modification complexes are composed of small RNAs bound with core proteins to form ribonucleoprotein (RNP) complexes. The box C/D RNAs direct the methylation at the 2' OH position on the sugar of the targeted nucleotide whereas the H/ACA RNAs convert targeted uridines to pseudouridine. The specificity for the nucleotide modification reaction lies in the RNA component of these RNP enzymes where complementary sequences within the guide RNA base pair with target RNAs to determine the nucleotide for modification. The nucleotide modification reactions are carried out by the associated core proteins possessing methylase or pseudouridine synthase activities bound to the respective guide RNAs. The box C/D RNAs follow the "N+5 rule" for methylation where the 5<sup>th</sup> nucleotide of the target RNA upstream of box D and within the guide RNA-target RNA duplex is methylated. Many box C/D RNAs contain two RNA motifs which guide 2'-O-methylation. Nucleotide boxes C and D located near the 5' and 3' termini of the RNA form the terminal box C/D core motif whereas similar sequences located internally form the C'/D' motif. Despite the sequence similarity of the C/D and C'/D' boxes, it is not yet clear that the folded structure of each RNA motif is identical. Analysis of core protein binding to the box C/D core and C'/D' motifs has indicated symmetric versus asymmetric protein distributions on the archaeal and eukaryotic complexes, respectively. The box H/ACA RNAs use discontinuous guide sequences to hydrogen bond with the target RNA. The unpaired uridine within the guide RNA-target RNA duplex is converted to pseudouridine. Eukaryotic box H/ACA RNAs characteristically exhibit a bipartite RNA composed of two stem-loop structures, each possessing a bulged loop which contains the guide sequence and forms the pseudouridylation pocket. Analysis of archaeal H/ACA RNAs has far revealed either one or three stem-loop structures with the signature H box sometimes absent.

**Figure 1.** Folded RNA Structures and Core Proteins of the Box C/D and H/ACA RNPs.

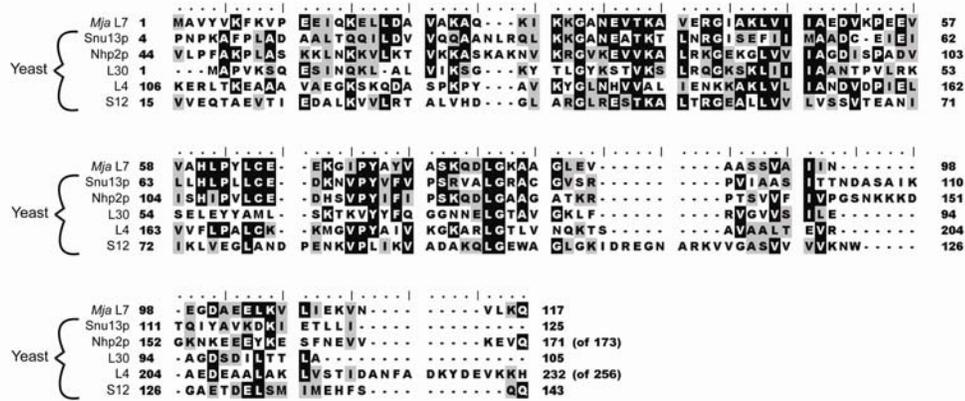
(a) Box C/D methylation guide RNAs and core proteins. The folded secondary structures of the archaeal (i) and eukaryotic (ii) box C/D RNAs are shown. Terminal box C and D and internal C' and D' sequences are in bold. Target RNAs (green) are base paired to the box C/D guide sequences located upstream of boxes D and D'. The designate nucleotide in the target RNA is methylated (purple). Core proteins associated with the terminal box C/D core motif and internal C'/D' motif are listed.

Homologous archaeal and eukaryotic proteins are indicated with arrows and presented in matching colors. (b) Box H/ACA pseudouridylation guide RNAs and core proteins. The folded secondary structures of archaeal (i) and eukaryotic (ii) box H/ACA RNAs are shown. Box H, located in the hinge region, and the 3' terminal ACA triplet sequences are in bold. Target RNAs (green) are based paired in a discontinuous RNA-RNA duplex to the H/ACA guide sequences. Unpaired nucleotides N (any nucleotide) and the modified pseudouridine nucleotide (red) of the target RNA are indicated. Core proteins are listed with homologous archaeal and eukaryotic proteins are indicated with arrows and presented in matching colors.

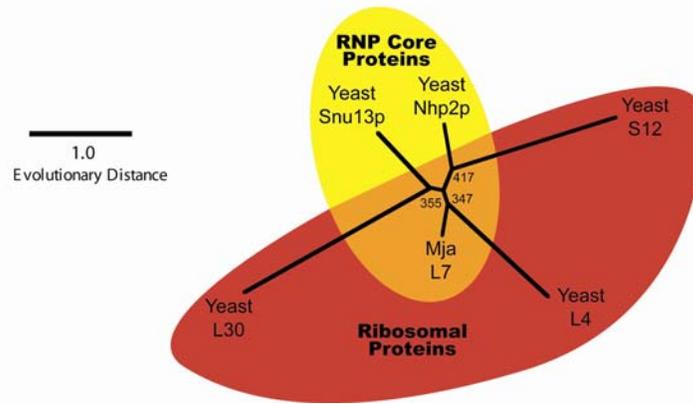


## L7 Protein Family

(a)



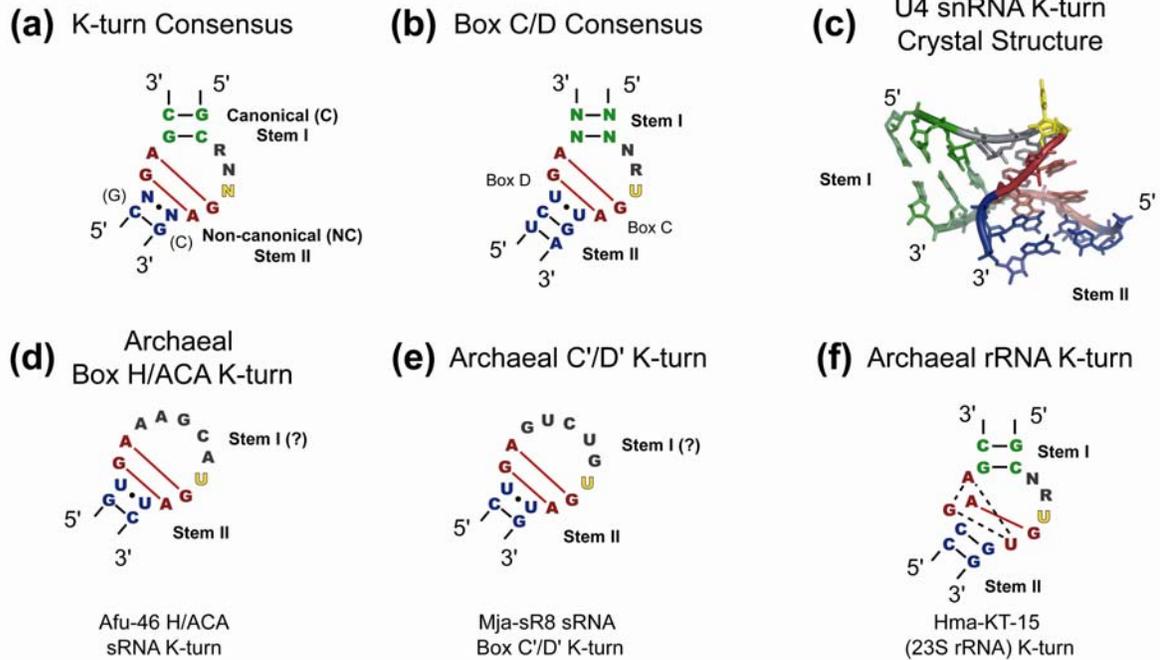
(b)



**FIGURE 1. The L7 Protein Family.**

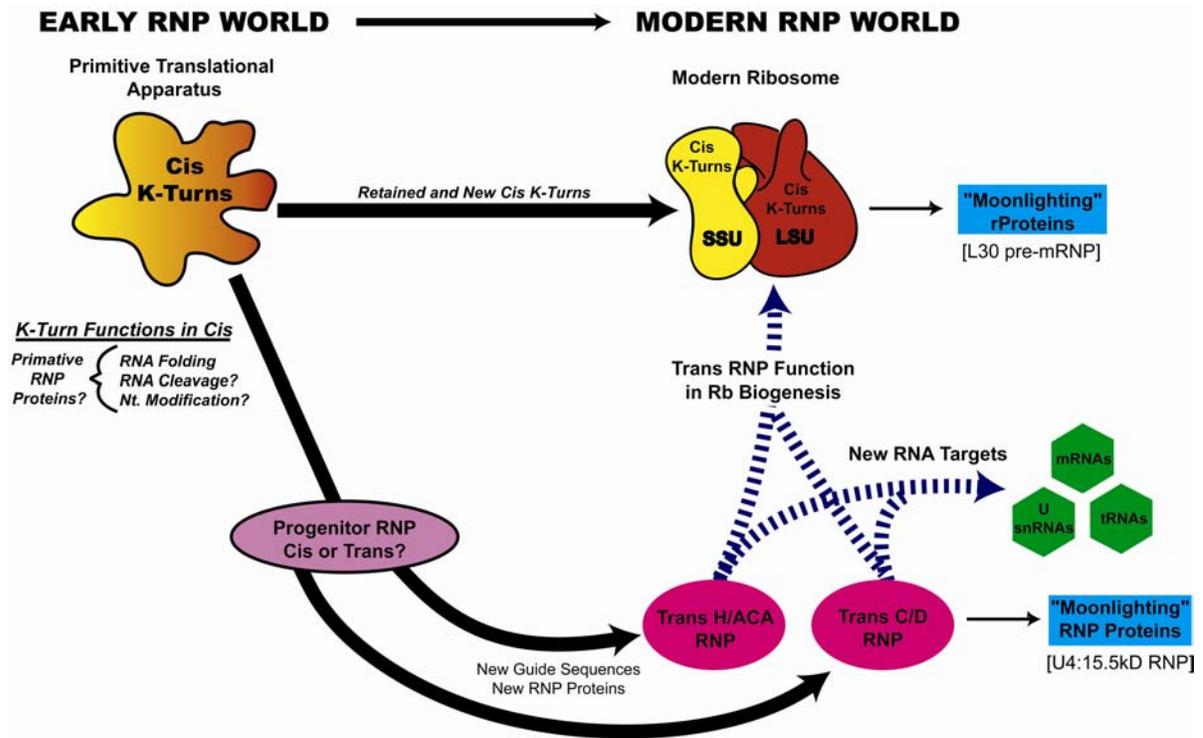
(a) Alignment of the L7 protein family members. The amino acid sequence of archaeal *Methanococcus jannaschii* ribosomal protein L7 (B64450) is compared with homologous yeast small nucleolar RNP (snoRNP) proteins Snu13p (NP\_01088), Nhp2p (NP\_010073) and yeast ribosomal proteins L30 (NP\_011485), L4 (NP\_011830), and S12 (NP\_015014). (GenBank accession numbers are indicated in parentheses following the individual proteins.) L7 protein homologs from a single eukaryotic organism (*Saccharomyces cerevisiae*) have been compared to emphasize sequence similarities. Identical amino acids are designated in white with black backgrounds while similar amino acids are shaded in gray. This alignment displays the central regions of these proteins (indicated by amino acid numbering) which exhibit high sequence similarity. (b) An evolutionary tree of the L7 protein family members. This tree is presented with the snoRNP proteins enclosed in yellow and ribosomal proteins enclosed in red. Ribosomal protein L7 which is common to both groups is enclosed in orange. The scale bar indicates 1.0 change per amino acid position.

## K-Turn RNA Family



**FIGURE 2. The Kink-Turn (K-Turn): An Evolutionarily Conserved RNA Motif.**

(a) The consensus secondary structure of the kink-turn [25] and (b) the consensus box C/D terminal core motif folded as a K-turn [15] are presented. This motif consists of stem I or canonical stem (green) and stem II or non-canonical stem (purple) flanking the asymmetric (5 + 2) bulge region. Hydrogen bonding between the tandem-sheared G:A base pairs (red) of the asymmetric bulge are characteristic of K-turns and essential for RNA folding. Stems I and II, the tandem-sheared G:A base pairs, and the protruding nucleotide (yellow) characteristic of the asymmetric bulge are coordinately colored in all the folded RNAs presented. (c) The three dimensional structure of the splicing U4 small nuclear RNA (snRNA) K-turn determined from the crystal structure of the U4 snRNA:15.5kD RNP complex [24] is presented. Also illustrated are three additional K-turns from (d) archaeal box H/ACA sRNAs [13] and (e) archaeal box C'/D' RNA motifs [20] compared with (f) the L7-binding K-turn of archaeal 23S ribosomal RNA [25]. Note the variation in these K-turn motifs and the utilization of a base-triple (broken lines) in the KT-15 turn of 23S rRNA. Each of these three K-turns (d-f) bind archaeal ribosomal protein L7.



**FIGURE 3. Evolution of the RNA-Guided Nucleotide Modification Complexes.**

Schematically presented is the proposed evolutionary pathway of the RNA-guided nucleotide modification ribonucleoprotein (RNP) complexes with origins in the primitive translation apparatus of the early RNP world. Solid thick arrows (black) depict evolution of the cis-positioned K-turns in a primitive translational apparatus proceeding to independent, trans-acting box C/D and H/ACA RNP complexes. This evolution is shown proceeding through a common progenitor RNP which could have been positioned in cis within the primitive translational apparatus or have already evolved as an independent, trans-acting complex. Some cis-positioned K-turns in the primitive RNP translational apparatus have been retained in present day ribosomes of the modern RNP world. Broken arrows (purple) depict functional activities of the trans-acting box C/D and H/ACA RNP upon both modern day ribosomes and other RNA targets such as tRNA, mRNA, and the splicing small nuclear RNAs (snRNAs). Thin arrows (black) indicate K-turn binding proteins of the ribosome and the nucleotide modification RNPs functioning as "moonlighting" proteins, binding the K-turn motifs of other RNA molecules and serving alternative functions.

## CONCLUDING REMARKS

Our investigations have contributed to a better understanding of archaeal box C/D RNP structure and function as well as its evolution as an RNA-guided nucleotide modification complex. First, these studies have detailed the requirements for both box C/D RNP assembly and RNA-guided methylation activity. This work serves as a platform for future characterization of both archaeal sRNPs and eukaryotic snoRNPs. Secondly, our results demonstrated that both the box C/D and C'/D' motifs must be juxtaposed in the same sRNP for efficient guided-methylation activity. This finding is in sharp contrast to our prior assumption that the box C/D and C'/D' motifs in the full-length sRNP complex are independently functioning motifs. It is not clear if the eukaryotic box C/D snoRNP requires the same juxtaposed complexes for activity. The identification of crosstalk between the C/D and C'/D' motifs in archaeal sRNPs is novel and future biochemical studies are necessary to determine the nature of the crosstalk interactions. These interactions may include induced changes in RNA structure and/or inter-RNP protein: protein interactions.

Characterization of the archaeal box C/D RNP as a symmetric particle with respect to protein binding has revealed the differences in RNP structure between Archaea and eukaryotes (see Chapter II). Contrasting symmetric versus asymmetric RNA-processing machinery in Archaea versus eukaryotes is not unprecedented; this has been noted previously in comparison of the archaeal and eukaryotic tRNA splicing endonuclease [1]. The tRNA splicing machinery of *Methanocaldococcus jannaschii* is a homotetramer of four, identical proteins. In yeast, this complex consists of four different proteins that carry out the same splicing and ligation reactions as its archaeal counterpart. For both the box C/D RNPs and the tRNA splicing endonuclease, evolution has resulted in a more complex enzymatic particle in eukaryotes as compared to Archaea.

Dissection of the archaeal sRNP structure revealed that a common RNP element (L7 bound to a K-turn motif) is present in both classes of modification guide RNPs as well as the ribosome. This suggests that the modern day modification guide RNAs may have their origins in a primitive translational complex (see Chapter III). We have proposed a model in which modification guide RNAs have evolved from *cis*-acting elements within the primitive translation apparatus to modern, *trans*-acting RNP complexes.

### **Remaining Questions**

Fundamental questions in the box C/D snoRNP field remain unanswered. First, what structural components are necessary for box C/D snoRNA function in eukaryotes? It is possible that additional proteins, such as accessory proteins p50 and 55 [2], are required for guided methylation activity? Do the box C/D and C'/D' RNPs guide methylation independently in the full-length snoRNP?

Interestingly, Kiss-Laszlo and coworkers have shown that an independent C'/D' RNP guides 2'-O-methylation in eukaryotes *in vivo* [3]. However, it is possible that the level of methylation was inefficient since the method employed did not assess relative efficiency of guided methylation activity compared to wild-type, full-length snoRNPs. Our recent analysis of the distance between the box C/D and C'/D' motifs in archaeal box C/D sRNAs shows that this spacing is highly conserved with an average length of 12 nucleotides (Tran et al., unpublished results). Shortening or lengthening of this conserved distance abolishes guided methylation. This is consistent with crosstalk interactions between the box C/D and C'/D' motifs in archaeal box C/D RNPs. However, this conservation of spacing between box elements is not evident in the eukaryotic snoRNAs analyzed. This may indicate that box C/D snoRNP-guided methylation in eukaryotes does not require crosstalk between the box C/D and C'/D' motifs. The establishment of an *in vitro* assembly and methylation system to study the eukaryotic box C/D snoRNPs will help address this question.

With our recent characterization of archaeal box C/D sRNP structure and function, it is now possible to address the mechanism of RNA-guided modification in more detail. A current issue is the structure of the rRNA: snoRNA complementary region. In eukaryotic box C/D snoRNAs, the complementary region can range from 10 to 21 nucleotides in length [4]. Does the entire complementary region within the snoRNA base pair with the target rRNA molecule? It seems implausible that the full complementary region pairs with the RNA target since this would result in a pseudoknot structure. However, it is possible that different regions of the complementary element may be utilized at different stages of the methylation reaction. The *in vitro* sRNP complex offers opportunities to analyze this aspect of substrate-enzyme interaction as well as examine the kinetics of methylation. For example, what are the on and off rates of the RNA substrate and does the reaction require RNA helicases?

The snoRNA field is rapidly expanding, and with this expansion comes new questions distinct from structure/function issues of the box C/D RNP. First, how are archaeal sRNAs processed? They are rarely found in introns, so any processing pathway will be distinct from that of the eukaryotic snoRNAs. When did eukaryotic snoRNAs become intron-encoded during evolution? How did snoRNAs become preferentially encoded in genes whose protein products are involved in various aspects of ribosome biosynthesis? The recent identification of a putative box H/ACA snoRNA encoded in a pseudogene possessing a poly A<sup>+</sup> tail indicates that they may have been incorporated by a retroposon mechanism. Finally, what is the function of snoRNAs with regions complementary to mRNAs? Do snoRNAs modulate gene expression and if so, is it species specific/ tissue specific? Unanticipated results are sure to follow with the future investigations characterizing snoRNA processing, function and evolution.

## REFERENCES

1. Li, H., C. Trotta, and J. Abelson, *Crystal structure and evolution of a transfer RNA splicing enzyme*. *Science*, 1998. **280**: p. 279-284.
2. Newman, D., et al., *Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription*. *RNA*, 2000. **6**: p. 861-879.
3. Kiss-Laszlo, Z., Y. Henry, and T. Kiss, *Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA*. *EMBO J.*, 1998. **17**(3): p. 797-807.
4. Bachellerie, J.P., et al., *Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA*. *Trends Biochem. Sci.*, 1995. **20**: p. 261-264.

## APPENDIX. Scientific Meeting Abstracts

- A. Kuhn, J., Newman, D., Liu, J., Tran, E., Maxwell, E.S. (2000). The Box C/D snoRNA Proteins Are of Ancient Origin and Provide a Possible Link Between Ribosome Biogenesis and Replication/Transcription. Fifth International Conference on Ribosome Biogenesis and Nucleolar Function, Lake Tahoe, Ca., August 17-21, 2000.
- B. Tran, E., Kuhn, J., Zhang, X., Maxwell, E.S. (2002). Ordered Assembly of the Archaeal Box C/D sRNP Core Complex Involves both RNA: Protein and Protein: Protein interactions. Seventh Annual Meeting of the RNA Society, Madison, WI. May 28-June 2, 2002.
- C. Tran, E. J., Zhang, X., Maxwell, E. S., Brown, B.A. (2003) Structural and Functional Analysis of Archaeal sRNPs. Nucleic Acids Gordon Research Conference, Newport, RI. June 1-6, 2003.
- D. Tran, E. J., Zhang, X., Maxwell, E. S. (2003). Evolution of the Box C/D RNP Complex: Efficient Ribonucleotide Methylation Requires Juxtaposed and Symmetrically Assembled Archaeal Box C/D and C'/D' RNPs. Sixth Ribosome Synthesis 2003, Arcachon, France. June 6-10, 2003.
- E. Maxwell, E. S., Tran, E. J., Zhang, X. (2003). Evolution of the Box C/D RNP Complex: Efficient Methylation Requires Juxtaposed and Symmetrically Assembled Archaeal Box C/D and C'/D' RNPs. Eighth Annual Meeting of the RNA Society, Vienna, Austria. July 1-6, 2003.
- F. Brown, B. A., Tran, E., Zhang, X., Maxwell, E.S. (2003) Structural and Functional Analysis of Archaeal sRNPs. Symposium on RNA Biology V: RNA, Tool and Target, Research Triangle Park, NC. October 17-18, 2003.
- G. Singh, S. K., Gurha, P., Tran, E. J., Maxwell, E. S., Gupta, R. (2004) A Trans Mechanism for Archaeal tRNA<sup>Trp</sup> Nucleotide 2'-O-Methylation Guided by the Pre-tRNA<sup>Trp</sup> Intron-Encoded Box C/D RNPs. Ninth Annual Meeting of the RNA Society, Madison, WI. June 1-6, 2004.
- H. Tran, E., Lackey, L., Zhang, X., Maxwell, E. S. (2004) Conserved Spacing Between the Box C/D and C'/D' RNPs of Archaeal Box C/D RNAs is Required for Efficient 2'-O-methylation of Targeted RNA Nucleotides. Ninth Annual Meeting of the RNA Society, Madison, WI. June 1-6, 2004.
- I. Singh, S. K., Gurha, P., Tran, E. J., Maxwell, E. S., Gupta, R. (2004) A Trans Mechanism for the Sequential 2'-O-methylation of Archaeal Pre-tRNA<sup>Trp</sup> Nucleotides Guided by the Pre-tRNA's Intron-encoded Box C/D RNPs. Extremophiles 2004: 5th International Conference on Extremophiles (An ASM Conference). Cambridge, MD. September 19-24, 2004.

**The Box C/D snoRNA proteins are of ancient origin and provide a possible link between ribosome biogenesis and replication/transcription.** *Jeffrey F. Kuhn, Donna R. Newman, Joyce C. Liu, Elizabeth J. Tran and E. Stuart Maxwell* Department of Biochemistry, North Carolina State University, Raleigh, North Carolina, USA.

Ribosome biogenesis requires a diverse population of small nucleolar RNAs (snoRNAs) for pre-rRNA processing and nucleotide modification. The box C/D snoRNAs possess conserved nucleotide boxes C and D contained within a folded structural element defined as the box C/D core motif that is essential for snoRNA biogenesis and nucleolar transport, as well as indispensable for formation of the snoRNP core complex. Using the box C/D core motif derived from mouse U14 snoRNA, we have affinity purified four proteins that consist of two pairs, each protein pair related in sequence. The first pair corresponds to the yeast nucleolar proteins Nop56 and Nop58, both of which have been shown to be essential for snoRNA biogenesis. The second pair of proteins, designated p50 and p55, have been previously identified as interacting with the TATA-binding protein (TBP) as well as replication protein A. Depletion experiments in yeast (see Decatur et al. abstract) have shown that both p50 and p55 are essential for snoRNA biogenesis and pre-rRNA processing. Evolutionary analysis has revealed that both are highly conserved in eukaryotes and homologs are found in Archaea. Western blot analysis revealed that both p50 and p55 are localized to the nucleoplasm and immunoprecipitation experiments demonstrated that both p50 and p55 are transiently associated with the box C/D snoRNP complex in the nucleoplasm, consistent with a role in snoRNA biogenesis and/or snoRNA transport events. The nucleoplasmic localization of p50/p55 and their reported association with TBP and rep A may suggest a regulatory role for these nucleoplasmic proteins, coordinating ribosome biogenesis with transcription and/or replication events.

**Ordered assembly of the archaeal box C/D sRNP core complex involves both RNA:protein and protein:protein interactions.** *Elizabeth Tran, Jeffrey Kuhn, and E. Stuart Maxwell* Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA.

Archaeal box C/D sRNAs are bound with archaeal proteins L7, Nop56/58 and fibrillarin to establish the sRNP core complex. We have studied the assembly and structure of this ribonucleoprotein particle by cloning these three sRNP protein genes from *Methanococcus jannaschii*, expressing recombinant proteins in *E. coli*, and investigating protein binding to the box C/D core motif RNA. Ribosomal protein L7 binds the RNA in the absence of other core proteins and requires the folded stem-asymmetric bulge-stem structure of the box C/D core motif with tandem sheared G:A base pairs and protruding pyrimidine nucleotide of the asymmetric bulge. Fe-EDTA mapping revealed that L7 interact with box C of the folded core motif. Binding studies have begun to dissect protein:protein interactions between the sRNP core proteins. In vitro pull-down experiments demonstrated a strong interaction between Nop56/58 and fibrillarin. In contrast, no such interaction was seen in vitro between L7 and fibrillarin or L7 and Nop56/58. Co-immunoprecipitation of L7 and fibrillarin from a *M. jannaschii* cell lysate suggests that fibrillarin association with L7 requires sRNP complex assembly. These observations coupled with electrophoretic mobility-shift analysis examining sRNP assembly has determined that the order of core protein binding with the box C/D core motif is L7, then Nop56/58, followed by fibrillarin. Finally, ribosomal protein L7 is a component of both sRNP complexes and the 50S ribosomal subunit, binding “k-turn” motifs in both the box C/D and 23S rRNA substrates. The striking similarity in structure of both RNP complexes suggests to us that the archaeal sRNPs and eukaryotic snoRNPs could well have their evolutionary origins in primordial ribosomes of the ancient RNA world.

**Structural and Functional Analysis of Archaeal sRNPs** Elizabeth Tran<sup>1</sup>, Xinxin Zhang<sup>1</sup>, and E. Stuart Maxwell<sup>1</sup>, and Bernard A. Brown II<sup>2,3</sup> <sup>1</sup>Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695; <sup>2</sup>Department of Chemistry and <sup>3</sup>Center for Structural Biology, Wake Forest University, Winston-Salem, NC 27109

Ribosomes are molecular machines that interpret the genetic message and synthesize proteins in all cells. The small nucleolar RNAs (snoRNAs) play important roles in the synthesis of eukaryotic ribosomes by directing specific covalent modification on ribosomal RNAs. The box C/D family of snoRNAs guides 2'-O-methylation of ribose. These modifications are thought to be important for proper folding and structural stability of ribosomal RNA and affect the efficiency of protein synthesis. Archaeal organisms have small RNAs (sRNAs), which are functionally analogous to the snoRNAs of eukaryotes. At present, the archaeal box C/D small ribonucleoproteins (sRNPs) are the only biochemically-active system that can be reconstituted from purified components. In collaboration with E. Stuart Maxwell's research group (N. C. State Univ.), we have been investigating the biochemical, biophysical, and structural aspects of archaeal sRNP-guided 2'-O-methylation. Recent experiments have demonstrated that the minimal functional archaeal sRNP complex consists of three proteins (L7, Nop56/58, fibrillar) and a 35 nt sRNA containing box C.D/ The entire sRNP complex, consisting of a 70 nt RNA stoichiometry of the active complex is unclear, but thought to contain two copies of each protein. Circular dichroism experiments have demonstrated that the L7 core protein interacts with the box C/D motif in a magnesium-dependent manner, and induces structural changes in the RNA upon binding. The crystal structure of the *M. jannaschii* L7 protein has been solved to 1.6 Å, and has revealed significant structural homology to the eukaryotic 15.5kDa homolog, the *Haloarcula marismortui* L7Ae (HS6) protein, and the yeast ribosomal protein, L30. We have recently obtained crystals of the *M. jannaschii* L7 protein complexed with box C/D and C'/D' RNA motifs. Crystal optimization and data collection are underway.

**Evolution of the box C/D RNP complex: Efficient ribonucleotide methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs.** *Elizabeth Tran, Xinxin Zhang, and E. Stuart Maxwell* Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA.

Box C/D RNP complexes direct the nucleotide-specific, 2'-O-methylation of ribonucleotide sugars in target RNAs. RNA guide sequences associated with the terminal box C/D core and internal C'/D' motifs base pair with complementary sequences in the target RNAs while the box C/D RNA-bound core proteins carry out the nucleotide modification reaction. We have assembled *in vitro* an archaeal box C/D sRNP complex using *Methanococcus jannaschii* sRNA sR8 and recombinant sRNP core proteins L7, Nop56/58, and fibrillarin. The reconstituted sR8 sRNP guides methylation from both the terminal box C/D core and internal C'/D' RNP complexes. Reconstitution of sRNP complexes containing only the box C/D or C'/D' motifs demonstrated that only the box C/D RNP exhibits methylation activity, albeit at reduced levels. Efficient ribonucleotide 2'-O-methylation directed from both motifs requires that the box C/D and C'/D' RNPs be juxtaposed in the full-length sRNA molecule. In contrast to the eukaryotic snoRNP where core proteins are distributed asymmetrically on the box C/D and C'/D' motifs, all three archaeal core proteins bind both motifs creating a symmetric sRNP particle. Analysis of archaeal L7/eukaryotic 15.5kD core protein binding to box C/D RNAs revealed that these homologs exhibit different RNA-recognition capabilities for the box C/D and C'/D' motifs. L7 and 15.5kD protein's unique binding characteristics contribute to the differential distribution of core proteins on the archaeal and eukaryotic box C/D and C'/D' RNP complexes. Thus, the asymmetric vs. symmetric distribution of proteins on the eukaryotic and archaeal box C/D RNP particles reflects an evolution in RNA-binding capabilities of the core proteins.

**Evolution of the box C/D RNP complex: Efficient methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs** *E. Stuart Maxwell, Elizabeth Tran, and Xinxin Zhang.* Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA.

Box C/D RNP complexes direct the nucleotide specific, 2'-O-methylation of ribonucleotide sugars in target RNAs. RNA guide sequences associated with the terminal box C/D core and internal C'/D' motifs base pair with complementary sequences in the target RNAs while the box C/D RNA-bound core proteins carry out the nucleotide modification reaction. We have assembled *in vitro* an archaeal box C/D sRNP complex using *Methanococcus jannaschii* sRNA sR8 and recombinant sRNP core proteins L7, Nop56/58, and fibrillarin. The reconstituted sR8 sRNP guides methylation from both the terminal box C/D core and internal C'/D' RNP complexes. Reconstitution of sRNP complexes containing only the box C/D or C'/D' motifs demonstrated that only the box C/D RNP exhibits activity, albeit at reduced levels. Efficient ribonucleotide 2'-O-methylation directed from both motifs requires that the box C/D and C'/D' RNPs be juxtaposed in the full-length sRNA molecule. In contrast to the eukaryotic snoRNP where core proteins are distributed asymmetrically on the box C/D and C'/D' motifs, all three archaeal core proteins bind both motifs creating a symmetric sRNP particle. Analysis of the Nop56/58-fibrillarin dimer binding to sR8 sRNA revealed that this protein complex can specifically bind the internal C'/D' motif in the absence of L7, albeit at reduced affinity. Interaction of the Nop56/58-fibrillarin complex with the internal C'/D' motif but not the terminal C/D core motif suggests that each RNA element presents a uniquely folded structure for protein binding. Analysis of archaeal L7/eukaryotic 15.5kD binding to archaeal sR8 sRNA and eukaryotic U15 snoRNA revealed that each homolog exhibits differential binding to the box C/D RNAs. Despite essentially identical L7 and 15.5kD folded structures, L7 binds both the C/D and C'/D' motifs whereas 15.5kD recognizes only the terminal C/D core motif. Thus, the differential distribution of proteins on the eukaryotic vs. archaeal box C/D RNP complexes reflects an evolution of RNA-binding capabilities of the core

proteins. This evolution of RNA-binding capabilities ultimately results in the unique RNP structures for the sRNA-guided eukaryotic and archaeal box C/D nucleotide modification complexes.

**Structural and Functional Analysis of Archaeal sRNPs** Bernard A. Brown II<sup>1,2</sup>, Elizabeth Tran<sup>3</sup>, Xinxin Zhang<sup>3</sup>, and E. Stuart Maxwell<sup>3</sup> <sup>1</sup>Department of Chemistry and <sup>2</sup>Center for Structural Biology, Wake Forest University, Winston-Salem, NC 27109; <sup>3</sup>Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695

Ribosomes are molecular machines that interpret the genetic message and synthesize proteins in all cells. The small nucleolar RNAs (snoRNAs) play important roles in the synthesis of eukaryotic ribosomes by directing specific covalent modification on ribosomal RNAs. The box C/D family of snoRNAs guides 2'-O-methylation of ribose. These modifications are thought to be important for proper folding and structural stability of ribosomal RNA and affect the efficiency of protein synthesis. Archaeal organisms have small RNAs (sRNAs), which are functionally analogous to the snoRNAs of eukaryotes. At present, the archaeal box C/D small ribonucleoproteins (sRNPs) are the only biochemically-active system that can be reconstituted from purified components. In collaboration with E. Stuart Maxwell's research group (N. C. State Univ.), we have been investigating the biochemical, biophysical, and structural aspects of archaeal sRNP-guided 2'-O-methylation. Recent experiments have demonstrated that the minimal functional archaeal sRNP complex consists of three proteins (L7, Nop56/58, fibrillarin) and a 35 nt sRNA containing box C.D/ The entire sRNP complex, consisting of a 70 nt RNA stoichiometry of the active complex is unclear, but thought to contain two copies of each protein. Circular dichroism experiments have demonstrated that the L7 core protein interacts with the box C/D motif in a magnesium-dependent manner, and induces structural changes in the RNA upon binding. The crystal structure of the *M. jannaschii* L7 protein has been solved to 1.6 Å, and has revealed significant structural homology to the eukaryotic 15.5kDa homolog, the *Haloarcula marismortui* L7Ae (HS6) protein, and the yeast ribosomal protein, L30. We have recently obtained crystals of the *M. jannaschii* L7 protein complexed with box C/D and C'/D' RNA motifs. Crystal optimization and data collection are underway.

**A Trans Mechanism for Archaeal tRNA<sup>Trp</sup> Nucleotide 2'-O-Methylation Guided by the Pre-tRNA<sup>Trp</sup> Intron-Encoded Box C/D RNPs** Sanjay K. Singh, Piryatansh Gurha, Elizabeth J. Tran<sup>1</sup>, E. Stuart Maxwell<sup>1</sup> and Ramesh Gupta Biochemistry and Molecular Biology, Mailcode 4413, Southern Illinois University, Carbondale, IL 62901-4413, USA and <sup>1</sup>Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-7622, USA

RNA processing events to produce mature tRNA<sup>Trp</sup> of *Haloferax volcanii*, an archaeon, include C/D box-mediated 2'-O-methylation of nucleotides C<sub>34</sub> and U<sub>39</sub> as well as excision of an introns. Modification of these two residues, located in the 5' and 3' exons, respectively is guided by two box C/D motifs positioned within the pre-tRNA introns. Therefore, it is believed that the two C/D RNP complexes are positioned in cis within the introns when they guide the two ribose methylation reactions. Folding of pre-tRNA required for the cis methylations and splicing are mutually exclusive and the two cis methylation reactions would require either formation of a pseudoknot or two different structures. Here we show that recombinant archaeal box C/D core proteins from *Methanocaldococcus jannaschii*, L7Ae, aNop5p, and Fibrillarin, form functional C/D RNP complexes with the *H. volcanii* pre-tRNA<sup>Trp</sup> transcript and methylate both C<sub>34</sub> and U<sub>39</sub> nucleotides in the pre-tRNA. Nucleotide methylation requires complementarity between the guide and target nucleotides. Single mutation of either guide or target residue in the pre-tRNA abolishes corresponding methylation. However, trans methylations of both pre-tRNAs in the reaction are observed when one pre-tRNA has a mutated target residue and the second pre-tRNA contains the complementary guide residue. In addition, we demonstrate that both circular and linear introns formed during the splicing reaction assemble C/D RNPs and can guide methylation of pre-tRNAs as trans complexes. The observation that the circular and linear forms of the pre-tRNA<sup>Trp</sup> introns accumulate *in vivo* may suggest that methylation of C<sub>34</sub> and U<sub>39</sub> is guided by trans RNP complexes in the cell and not via cis-positioned C/D RNPs within the introns as originally assumed. A trans mechanism would thus eliminate the problem of requiring structural changes within the folded introns to accomplish methylation at both modification sites.

**Conserved Spacing Between the Box C/D and C'/D' RNPs of Archaeal Box C/D sRNAs Is Required for Efficient 2'-O-Methylation of Targeted RNA Nucleotides** *Elizabeth Tran, Lela Lackey, Xinxin Zhang, and E. Stuart Maxwell* Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina.

Box C/D RNAs function in ribosome biogenesis by guiding the site-specific 2'-O-methylation of numerous ribosomal RNA nucleotides. The box C/D sRNAs of Archaea possess terminal box C/D and internal C'/D' RNA motifs, each of which directs methylation using guide regions located upstream of boxes D and D'. Each RNA motif binds sRNP core proteins L7, Nop56/58, and fibrillarin (the methylase) to assemble C/D and C'/D' RNPs. Recently, our laboratory has developed an *in vitro* assembly system that reconstitutes enzymatically-active sRNP complexes. We have shown that all three core proteins are bound to both box C/D and C'/D' motifs to establish a symmetric sRNP. Efficient methylation requires that the C/D and C'/D' RNPs to be juxtaposed in the same sRNP particle. Core protein binding studies revealed the cooperative nature of L7 binding to the C/D and C'/D' motifs. From these studies, we have demonstrated the importance of crosstalk interactions between the symmetric C/D and C'/D' RNP for optimal methylation activity. To further characterize these “inter-RNP” interactions, we have examined the importance of spacing between the two juxtaposed complexes. Comparative analysis of archaeal box C/D sRNA folded structure revealed the highly constrained spatial positioning of the C/D and C'/D' motifs. A median distance of 12 nucleotides separating the two motifs was noted for archaeal box C/D sRNAs. Mutagenesis of motif spacing has revealed striking effects upon the enzyme activities of both the C/D and C'/D' RNPs, ranging from enhancement to complete loss of 2'-O-methylation activity. Surprisingly, a similar comparative analysis of eukaryotic box C/D RNAs does not indicate the same strict conservation of motif spacing. This may reflect the distinct structural differences observed between the box C/D RNPs of Archaea and Eukarya.

**A *Trans* Mechanism for the Sequential 2'-O-Methylation of Archaeal Pre-tRNA<sup>Trp</sup> Nucleotides Guided by the Pre-tRNA's Intron-encoded Box C/D RNPs** Sanjay K. Singh, Priyatansh Gurha, Elizabeth J. Tran<sup>1</sup>, E. Stuart Maxwell<sup>1</sup> and Ramesh Gupta Department of Biochemistry and Molecular Biology, Mailcode 4413, Southern Illinois University, Carbondale IL 62901-4413, USA and <sup>1</sup> Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-7622, USA

*Haloferax volcanii* pre-tRNA<sup>Trp</sup> processing requires box C/D RNP-guided 2'-O-methylation of nucleotides C<sub>34</sub> and U<sub>39</sub> followed by excision of the pre-tRNA intron. Location of the target nucleotides in the pre-tRNA<sup>Trp</sup> exons with corresponding guide RNPs positioned within the intron has led to the assumption that methylation occurs via an intramolecular or cis mechanism. We have investigated 2'-O-methylation of this tRNA precursor *in vitro* by assembling methylation-competent box C/D RNPs on both the pre-tRNA<sup>Trp</sup> and excised intron (**both linear and circular forms**) using recombinant *Methanocaldococcus jannaschii* box C/D core proteins. Site-directed mutagenesis of target and guide nucleotides revealed that box C'/D'-guided U<sub>39</sub> methylation first requires box C/D-guided C<sub>34</sub> methylation. Analysis also demonstrated that the intron-encoded box C/D RNP of the pre-tRNA<sup>Trp</sup> could guide the *trans* 2'-O-methylation of another pre-tRNA<sup>Trp</sup> as did the excised intron. Finally, analysis of pre-tRNA<sup>Trp</sup> methylation in a cell extract again revealed that sequential nucleotide modification was carried out in *trans*. Previous observations demonstrating the accumulation of excised pre-tRNA<sup>Trp</sup> introns *in vivo* may therefore suggest that C<sub>34</sub> and U<sub>39</sub> methylation is guided by the intron-encoded box C/D RNP using a *trans* mechanism.