ABSTRACT

APPEL, CATHRYN DENISE. Structural Features of the Guide:Target RNA Duplex Required for Archaeal C/D sRNA Guided Nucleotide 2'-*O*-methylation. (Under the direction of E. Stuart Maxwell.)

Archaeal box C/D sRNAs guide the 2'-O-methylation of target nucleotides in both ribosomal and tRNAs. These small non-coding RNAs are characterized by conserved terminal box C/D and internal C'/D' RNA motifs. Each RNA motif binds three core proteins to establish individual RNP complexes that catalyze the site-specific 2'-O-methylation of target nucleotides. Specificity of nucleotide modification is determined by target RNA base pairing with complementary sRNA D or D' guide sequences. The fifth target nucleotide upstream from the D or D' box within the guide:target RNA is then methylated by the core proteins. *In vitro* assembly of *Methanocaldococcus jannaschii* sR8 box C/D RNA with recombinant core proteins, L7, Nop56/58, and fibrillarin produces a methylation-competent sRNP complex. This model box C/D sRNP has now been used to determine the structural features of the guide:target RNA duplex that are important for sRNA-guided nucleotide methylation. Watson-Crick pairing of guide and target nucleotides was essential for nucleotide methylation. Mismatched bases within the guide:target RNA duplex also disrupted target nucleotide methylation. Nucleotide methylation required that the guide:target duplex consist of an RNA:RNA helix as target deoxy-oligonucleotides possessing a target ribonucleotide were not methylated. Methylation specificity at the base paired guide:target nucleotide was compromised by elevated Mg^{2+} concentrations. In high divalent cation concentrations, target nucleotides not hydrogen bonded to the guide nucleotide were nevertheless methylated. Interestingly, D and D' target RNAs were methylated to different levels when deoxynucleotides were inserted within the target RNA or when target methylation was carried out in elevated Mg^{2+} concentrations. These results suggested that

structural features unique to the box C/D and C'/D' RNPs affect their nucleotide methylation capabilities. Finally, the ability of the sR8 box C/D sRNP to methylate target nucleotides positioned within highly structured RNA hairpins suggested an intrinsic ability of this archaeal RNA:protein enzyme to unwind double-stranded target RNAs prior to nucleotide modification.

STRUCTURAL FEATURES OF THE GUIDE:TARGET RNA DUPLEX REQUIRED FOR ARCHAEAL C/D sRNA GUIDED NUCLEOTIDE 2'-*O*-METHYLATION

by

CATHRYN DENISE APPEL

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

STRUCTURAL AND MOLECULAR BIOCHEMISTRY

Raleigh, North Carolina

2006

APPROVED BY:

Linda Hanley-Bowdoin	A. Clay Clark

E. Stuart Maxwell
Chair of Advisory Committee

DEDICATION

To my loving husband, Wyat, for his patience, love and support.

And in honor of my mother, Cathy, whose love and guidance shaped my life.

BIOGRAPHY

Denise Appel was born in Robeson County, NC and is one of two daughters reared almost exclusively by her mother Cathy Atkinson. She was raised in Dublin, NC and attended Tar Heel High School where she graduated as Class of 1999 Valedictorian. During high school, Denise discovered a passion for both academics and athletics, excelling in softball and volleyball. During her senior year of high school, Denise met her future husband, Wyat, at a NC State AMS picnic. Six years later, during graduate school, they were married in Raleigh, NC.

Denise began her college career at North Carolina State Univeristy in biology but soon changed her major to biochemistry, as her interests focused more on molecular biology. She also persued her passion for volleyball and joined the NC State Club Volleyball team. This allowed her to travel across the country, playing in many exciting and memorable tournaments. During her college career, Denise participated in a total of six U. S. National Volleyball Championships. Upon graduating, Denise enrolled in graduate school at North Carolina State University in Biochemistry under the direction of Dr. E. S. Maxwell. Following completion of her Masters, Denise plans to enter the foreign world of industry in hopes of making some money.

ACKNOWLEDGEMETNS

I would like to thank Dr. E. S. Maxwell for his patience and guidance throughout my graduate career and for the opportunities he provided to expand my intellectual abilities. Special thanks to the Maxwell lab members for their support and advice, particularly to Keith whose ideas and examples were always a source of inspiration and to Xinxin and Roxanna whose company and support were most appreciated. I would also like to recognize my committee members, Dr. Linda Hanley-Bowdoin, Dr. Clay Clark, Dr. Paul Wollenzien, and Dr. James Brown, and thank them for their assistance and guidance.

In particular I would like to acknowledge my friends and family for unceasing love and encouragement. Thank you, Wyat, for supporting my decisions and helping me find the courage to follow through with those decisions. Your unconditional love and absolute confidence has made such an impact on my life, you are the source of my happiness. Special thanks to my mother, Cathy, and sister, Jessica, just for being my family. Thank you, mom, for your love, your example, and your unyielding support. My abilities and accomplishments are a direct result of you, I would not be who I am today without you. To Jessica, thanks for your love and companionship. Finally, thanks to Mom and Dad Appel, who have loved me like a daughter from the moment we met, your love, support and kindness means so much. Also, thanks to Bret, Brad, Scot and Val, my extended family, whose company, friendship, and advice are always valued.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

LIST OF FIGURES

LITERATURE REVIEW

CHAPTER I

LITERATURE REVIEW

I. A Brief History of the Eukaryotic Small Nucleolar RNAs

In the late 1960's, a novel low molecular weight RNA was isolated from a fraction of nucleolar RNA that had been resolved on a sucrose density gradient. This RNA species was U-rich and distinctly different in base composition from the GC-rich ribosomal RNA of the nucleolus.[1,2] This new RNA species lacked amino acid acceptor activity, generally attributed to transfer RNAs (tRNA), and exhibited low template activity, indicative of messenger RNA (mRNA). Based on these observations, this novel species was reported to be a small, non-messenger RNA of unknown function.[2] Further investigation ultimately demonstrated its specific localization to the nucleolus. Subsequently, this RNA was designated a small nucleolar RNA or snoRNA, indicative of its small size and cellular location.[2] Eventual sequence analysis defined this snoRNA as U3, the first of what would be many eukaryotic snoRNAs. The nomenclature for the snoRNAs follows the designation system originally established with the splicing snRNAs where U is indicative of "uridine-rich" RNAs and the assigned number indicates the order of snoRNA identification.[3,4]

After U3 identification in humans, other organisms were examined for U3 homologs. Investigations demonstrated U3 snoRNAs in diverse vertebrate and plant species as well as Drosophila and yeast. Notably however, no U3 snoRNA was detected in bacterial organisms. The widespread conservation of U3 in eukaryotes indicated a conserved and essential function. Over the last 20 years, a number of eukaryotic snoRNAs has been added to a growing list that now includes more than 100 species.[4,5,6]

All are localized to the nucleus and, by implication, are involved in ribosome biogenesis.[7] Of the known species, U3 is the most abundant at about 200,000 copies per nucleus while other snoRNAs are more modestly represented at between 10,000-20,000 copies per nucleus.[4,8] Yeast generally contain fewer snoRNA copies, ranging from a few hundred to slightly more than 1,000.[9,10] Vertebrate snoRNAs are typically 75-200 nucleotides in length while yeast guide RNAs can reach lengths of up to 600 nucleotides.[7]

II. Eukaryotic snoRNA Function

Investigations into snoRNA function have demonstrated several roles for the snoRNAs in ribosomal RNA processing. Certain snoRNA species are required for pre-rRNA cleavage events essential for rRNA biogenesis while others function as rRNA folding chaperones.[4,11-14] However, the primary role of most snoRNAs is to direct the site-specific nucleotide modification of cellular RNA targets, primarily rRNA.[15-17] The site of nucleotide modification is determined via base pairing between a guide region within the snoRNA and a complimentary sequence in the target RNA.[16,18] Because base pairing is essential to snoRNA function, snoRNAs are often referred to as "guide RNAs".

Eukaryotic snoRNAs or guide RNAs are divided into two major families based upon conserved sequence elements. The box C/D family of guide RNAs is characterized by conserved boxes C (RUGAUGA) and D (CUGA) sequence elements located at the 5' and 3' termini, respectively. The box C/D guide RNAs direct 2'-O- ribose methylation of targeted nucleotides. The site of 2'-0 methylation is determined by the spatial positioning of the designate nucleotide to the upstream box D sequence, where the fifth base paired

nucleotide in the snoRNA:target RNA duplex is targeted for methylation.[16,18] Box H/ACA guide RNAs are defined by a conserved box H element (ANANNA) and an ACA nucleotide triplet, located within a hinge region of its bipartite structure and at the 3' terminus, respectively. This family of guide RNAs is responsible for the site-specific isomerization of uridines to pseudouridines.[19-21]

III. SnoRNA Homologs Are Found in Archaea

In the three domains of life, Archaea are thought to represent an evolutionary link between Bacteria and Eukarya as Archaea possess characteristics of both. Identification of both 2'-0 methylated nucleotides and pseudouridines in archaeal rRNA led to the search for snoRNA homologs.[22] Genome wide searches of several archaeal species and subsequent sequence comparisons led to the discovery of numerous snoRNA-like molecules called sRNAs.[23,24] Like the eukaryotic guide RNAs, archaeal sRNAs are divided into two classes based on conserved sequences, structural features, and functions. Archaeal box H/ACA sRNAs convert uridines to pseudouridines while the box C/D RNAs are responsible for 2'-0 ribose methylation.[24,25] Ribosomal RNA is the primary target for sRNA- guided nucleotide modification, although examples of tRNA targets are documented, such as $tRNA^{Trp}$ in *Haloferax volcanii* and four targeted sites in tRNA Leu of *Pyrococcus*.[25-29]

As 2'-0 methylated bases are more abundant than pseudouridines in Archaea[30,31], there are only a few box H/ACA RNAs currently identified.[23] Overall, there are fewer base modifications in archaeal ribosomal RNA than in eukaryotic rRNA. Thus, the total

number of discovered and predicted sRNAs is less than half that of the eukaryotic snoRNAs. To date, the greatest number of archaeal sRNA species identified in a single organism is approximately 50 in the archaeon *Pyrococcus abyssi*.[24] Though Archaea lack membrane-bound organelles, the sRNAs are localized to a region equivalent to the eukaryotic nucleus. In *P. abyssi*, all sRNAs are approximately equal in cellular abundance except for two which display notably decreased levels.[24] The length of archaeal guide RNAs is noticeably shorter than eukaryotic snoRNAs, averaging between 50 and 70 nucleotides.[24] As there is no evidence for snoRNA homologs in bacteria, guide RNAs are thus confined to Eukarya and Archaea, providing further evidence for an evolutionary link between these two kingdoms.[32]

IV. Box H/ACA Guide RNA Structure and Nucleotide Pseudouridylation.

The box H/ACA RNAs guide the nucleotide-specific isomerization of uridine to pseudoruidine (ψ). Eukaryotic box H/ACA RNAs fold into a phylogenetically conserved hairpin-hinge-hairpin bipartite structure, where the box H element is positioned within the hinge region (Figure 1). The ACA triplet nucleotide is located downstream of the second hairpin, three nucleotides upstream from the 3' terminus.[19,20] The guide regions, complementary to target RNA sequences, comprise an internal loop within the hair pins. The loop formed by the unpaired guide regions constitutes the "pseudouridylation pocket" where uridine isomerization occurs.[33]

Unlike the eukaryotic H/ACA snoRNAs, the archaeal H/ACA sRNAs are structurally distinct in that they may consist of 1, 2, or 3 hairpins with each hairpin possessing a

pseudouridylation pocket and accompanying guide regions. Also, the box H sequence is frequently absent in archaeal H/ACA RNAs.[23] Interestingly, the archaeal H/ACA sRNAs have an additional structural feature, the kink-turn (K-turn), located in the upper region of the stem-loop(s). This RNA element is absent in the eukaryotic H/ACA snoRNAs. The K-turn is a structural motif commonly found in rRNA and, importantly, in both eukaryotic and archaeal box C/D guide RNAs (discussed below in *Box C/D Guide RNAs*).

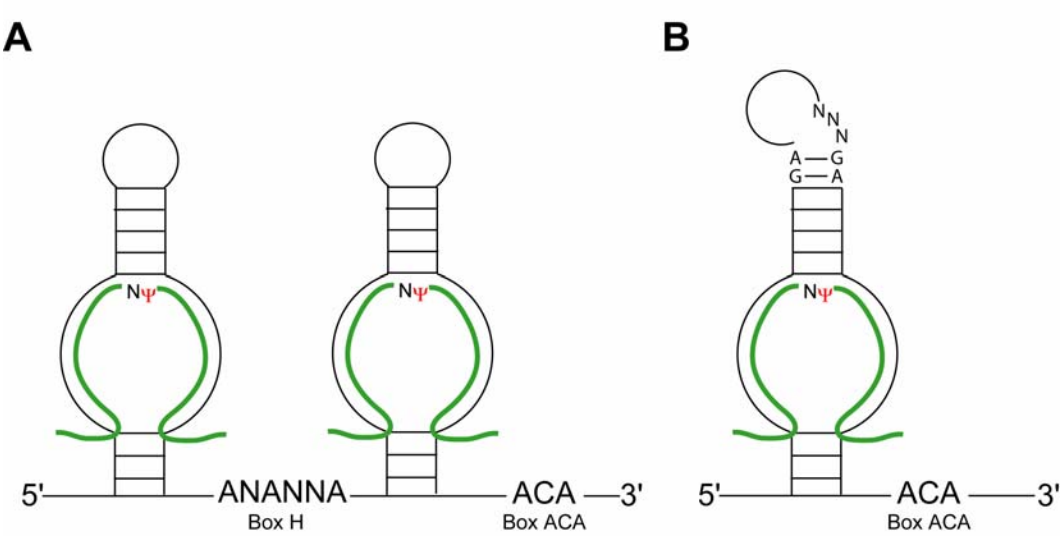

Figure 1. Secondary structures of eukaryotic and archaeal box H/ACA guide RNAs. Box elements are indicted in bold while target RNA (green) is shown base paired to guide regions within the ψ pocket. The targeted nucleotide (N ψ) is shown in red.

The mechanism of nucleotide-specific pseudouridylation is conserved between Eukarya and Archaea and achieved through base pairing between the guide region and the target RNA. The designate uridine to be isomerized, is located within the target RNA sequence and positioned by base pairing between the guide sequences and target RNA so that it is unpaired and accessible for isomerization (Figure 1).[20] This uridine is typically between 14 and 16 nucleotides upstream of the box H or ACA elements depending on

which hairpin is guiding psuedouridylation. Pseudouridylation may be guided from any hairpin containing guide regions complementary to rRNA sequences.[19,20]

V. Box C/D Guide RNA Structure and Nucleotide 2'-O-Methylation

The primary function of the box C/D guide RNAs is to direct the nucleotide-specific 2'-0 ribose methylation of designated nucleotides within a target RNA.[19] The characteristic box C and D sequence elements, RUGAUGA and CUGA, are located at the 5' and 3' ends of the guide RNA, respectively. The 5' and 3' ends typically base pair and establish a terminal stem-asymmetric bulge-internal stem structure (Figure 2). The asymmetric 5+2 bulge is created by partial base pairing of the box elements which fold into a highly conserved kink-turn or "K-turn" motif.[34] This K-turn element was first described in the crystal structure of the *Haloarcula marismortui* 50S ribosomal subunit and subsequently in eukaryotic U4 snRNA.[35] Some eukaryotic box C/D RNAs contain a second, internal pair of box C and D sequences which are homologous to the terminal box C and D sequences. These internal boxes are designated C' and D' boxes.[15,36,37] The C' and D' sequences are often not well conserved and difficult to discern, although the C' and D' boxes are believed to form a K-turn motif. Because of the unique structure of this internal K-turn fold, it has been designated the K-loop.[38] Guide regions, 10-21 nucleotides long, are located immediately upstream of boxes D and D' (Figure 2). The spatial positioning of the terminal box C/D core motif and internal box C'/D' motif is variable and can be separated by short or very long inter-motif sequences. Most often, but not always, they possess identifiable guide sequences upstream of the D' box and complementary to cellular target RNAs.[16,18,39]

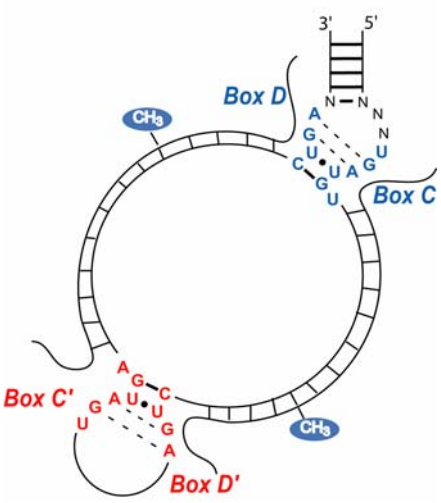

Figure 2. A representative box C/D guide RNA. Conserved box elements C and D (in blue) and C' and D'(in red). Two separate RNA molecules are shown base paired to the antisense regions and the target nucleotide (blue) is 5 nucleotides upstream of boxes D and D'. Eukaryotic box C/D guide RNAs may or may not have the internal C'and D' boxes.

Archaeal box C/D sRNAs are structurally quite similar to the eukaryotic C/D RNAs, though somewhat simplified, suggesting they are more ancient than their evolved eukaryotic counterparts. Like the eukaryotic box C/D snoRNAs, the archaeal sRNAs contain highly conserved terminal boxes C and D which establish a K-turn fold. Archaeal box C/D sRNAs also possess internal box C' and D' sequences that are generally very well conserved and often identical in sequence to the terminal box C and D sequences.[28] This is in contrast to the highly variable eukaryotic C' and D' sequences.[37] As all archaeal box C/D sRNAs have C' and D' sequences, internal K-loops are a common feature among all species.[19,38] Furthermore, the archaeal box C/D sRNAs exhibit a more constrained inter-motif spacing between internal and terminal box elements. The typical length of the spacer regions is 12 nucleotides and often constitutes the entire guide region for the box C/D and C'/D' motifs.[40]

Box C/D-guided 2'-O-methylation is directed by the base pairing of the guide region with the complementary target RNA sequence. The site of 2'-O-methylation is determined by the spatial relationship of the target nucleotide to box D or D', where the fifth base-paired nucleotide upstream of these box sequences is targeted for modification.[16,18] Box C/D RNAs may direct 2'-O-ribose methylation of target RNAs from one or both guide regions. Sometimes, each guide region in a given sRNA is guiding nucleotide modification in different target RNAs.[25,28] Notably, most archaeal sRNAs guide methylation from both guide regions, in contrast to eukaryotes where only about 20% of the box C/D snoRNAs exhibit a dual guide function.[25]

VI. *MRP RNA*: The Other Eukaryotic snoRNA.

MRP (mitochondrial RNA-processing) RNA is a small nucleolar RNA and a component of a ribonucleoprotein endoribonuclease. It is a unique snoRNA species in that it lacks any conserved or shared sequence or structural element that would classify it as a box C/D or H/ACA RNA.[41] Though MRP was originally identified by its ability to cleave mitochondrial RNA, its primary activity is in the nucleus where it is required for the B1 site-specific cleavage of 5.8S pre-rRNA. Depletion of MRP in yeast has confirmed its role in 5.8S rRNA processing, as all pre-rRNA cleavage sites occur normally except for B1.[41] MRP RNA structurally resembles RNaseP, a ribozyme that cleaves the 5' end of tRNAs. Due to this observation and MRP's activity in rRNA processing, it has been suggested that MRP RNA may function as a ribozyme. However, evidence for this is lacking at present.[41,42]

VII. SnoRNA Core Proteins

Both families of guide RNAs are found in the nucleolus as RNA-protein complexes. Sets of core proteins, absolutely necessary for function, have been defined for both the H/ACA and C/D guide RNAs. *In vivo* analysis has characterized the conserved proteins for both types of guide RNAs while knockout studies have demonstrated their importance in function. *In vitro* assembly of guide RNPs using recombinant core proteins has now been accomplished for both the H/ACA and C/D RNPs, although reconstitution of nucleotide modification activity has only been successful in the archaeal sRNP assembly systems.

Box H/ACA RNPs

In both Eukarya and Archaea, four conserved core proteins specifically bind the H/ACA guide RNAs to assemble an RNP complex. These core proteins are ultimately essential for both biogenesis and function. The eukaryotic core proteins are Nhp2p, Nop10p, GAR1p, and Cbf5p.[43-46] With the exception of GAR1p, all the proteins are required for H/ACA snoRNA accumulation *in vivo*. GAR1, containing N- and C-terminal glycine and arginine rich domains, were originally identified by co-purification with yeast H/ACA snoRNAs snR10 and snR30.[43] Nhp2p and Nop10p, identified by co-immunopreciptation with GAR1, are essential for pre-RNA processing and site-specific pseudouridine synthase (ψ synthase).[44] Cbf5p, the ψ synthase, was identified based on significant sequence homology to the E. coli tRNA ψ synthase, Tru B, a member of an evolutionarily conserved family of ψ synthases.[45,47,48] However, Cbf5p is the only member of the family that requires a RNA cofactor to catalyze the isomerization of

uridines to pseudouridines.[49] Depletion of plasmid-encoded Cbf5p under control of a repressible GAL promoter produced pre-rRNA lacking conserved pseudouridines and resulting in a dramatic decrease in pre-rRNA accumulation.[45]

The archaeal H/ACA sRNP differs from the eukaryotic complex only by the absence of Nhp2 which is replaced by L7, a homolog of the box C/D protein 15.5 kDa and known to specifically bind K-turn motifs.[50,51] The archaeal homologs of the remaining eukaryotic core proteins, Nop10p, Cbf5p, and GAR1, were predicted to exist based on DNA sequence comparisons and they were eventually verified experimentally in several archaeal species. [52,53]

Structural Organization of the H/ACA sRNP Complex.

Recent *in vitro* assembly of eukaryotic proteins suggests that a core trimer of Cbf5p, Nop10 and Nhp2p is formed in the cytoplasm, while GAR1 associates independently with Cbf5 either during or soon after core trimer formation. The entire complex is then transported to the nucleus presumably through Cbf5, the only protein of the core complex to contain a classic, functional nuclear localization signal.[54] Nhp2p, alone, can bind multiple RNA molecules promiscuously, but in concert with other H/ACA core proteins, binds specifically to the H/ACA RNAs.[55,56] Nhp2p binding to the core complex is apparently mediated by Nop10.[56] Nhp2p and Nop10p may play necessary roles in anchoring the core complex to the RNA via the binding of Nhp2p, thus restructuring the guide RNA to facilitate target RNA binding.[55,56]

Biochemical evidence indicates that CBf5p makes specific contacts with both the ACA triplet and the ψ pocket[57], supporting a model that describes the ψ pocket containing the target uridine positioned in the catalytic center of Cbf5p.[49,57] This is in fact confirmed by the recent crystal structure of the core trimer by Singh et al. and subsequent modeling and retro-fitting the core trimer onto model guide and target RNA duplex. The location of GAR1 within the RNP is not well understood, although there is evidence that GAR1 may interact with both guide RNA and target RNA. Crosslinking data indicates a specific contact with the target uridine while crystal structure analysis demonstrates a GAR1 interaction with the target RNA and Cbf5. Interestingly, the region of protein-protein contact between GAR1 and Cbf5 is a domain attributed to Cbf5-target RNA binding.[49,56,58] Possibly, GAR1 may act as a clamp, stabilizing the Cbf5p – target RNA interface and positioning the target uridine correctly in the Cbf5p active site.[49,56]

Further analysis of the core trimer crystal structure suggests that the final arrangement of core proteins upon the H/ACA RNA is conserved between Archaea and Eukarya.[49] The two complexes differ only by the presence of L7, which binds directly and specifically to RNA via the K-turn of archaeal H/ACA sRNAs. This is in contrast to the eukaryotic homolog Nhp2p which associates with RNA only through protein-protein interactions.[49] Despite fundamental differences between Nhp2p and L7, both proteins appear to interact with Nop10 and may play similar roles in the remodeling of their respective H/ACA RNAs.[49,56]

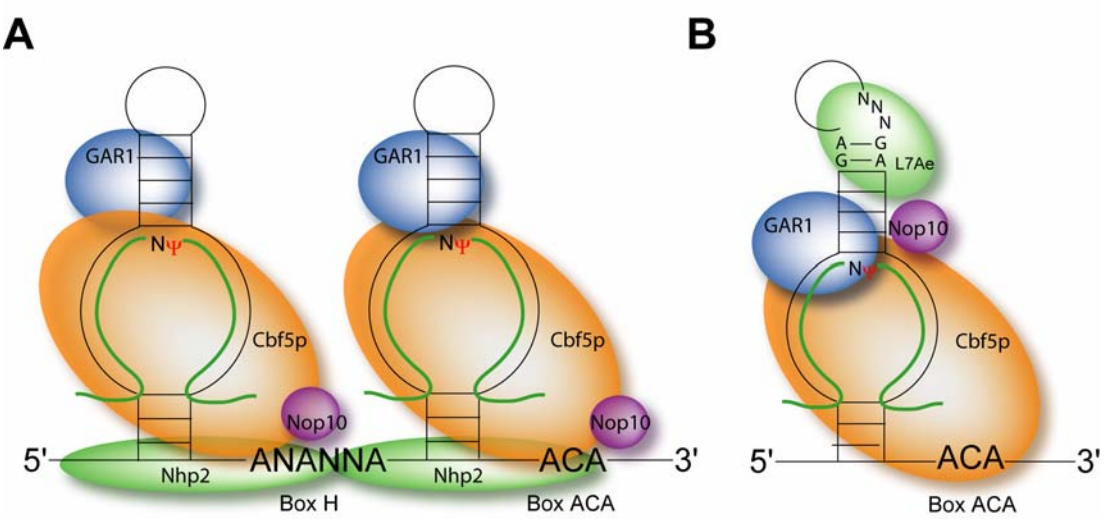

Figure 3. Eukaryotic and Archaeal H/ACA RNPs. A) Eukaryotic H/ACA RNPs consists of four core proteins. B) Archaeal H/ACA RNPs are also composed of four core proteins but Nhp2p is replaced by L7Ae. (Adapted from [49]).

Box C/D RNPs

Box C/D guide RNAs are also tightly associated with a set of core proteins. In eukaryotes, the core snoRNP complex possesses four proteins which are the 15.5kD protein, Nop56p, Nop58p, and fibrillarin (Nop1p in yeast).[19] The archaeal C/D sRNP is comprised of only three core proteins; L7, the archaeal homolog of 15.5 kD, a fibrillarin homolog, and a single homolog of Nop56 and Nop58, denoted Nop56/58 (Nop5p).[45,59-62] Both the 15.5 kDa protein and L7 belong to a family of RNA binding proteins including ribosomal proteins S12, L30, L7a, and the box H/ACA-associated Nhp2p.[50,63,64] 15.5 kDa binds specifically to the terminal K-turn motif created by base pairing of the C and D box elements.[35,65] Interestingly, L7 binds both the terminal and internal K-turn motifs, in contrast to15.5 kDa which binds only the terminal K-turn.[35,62,65] Initial binding of 15.5 kDa and L7 is considered the nucleation event required for subsequent binding of the remaining core proteins.[49,50,62,66]

Fibrillarin (Nop1p) is responsible for the methyl transferase activity of both archaeal and eukaryotic box C/D guide RNAs and uses S-adenosyl methionine as the methyl donor. Fibrillarin was identified as the putative methylase based on its consistent association with box C/D snoRNAs[8] and structural homology to other known methyl transferases.[67-69] Fibrillarin is a member of the RGG-motif family, like the H/ACA component GAR1p, and possesses an N-terminal glycine/arginine rich (GAR) domain as well as an internal RNA binding domain. Fibrillarin knockouts in yeast are lethal while temperature-sensitive mutants in both yeast and mice result in dramatic decreases in conserved pre-rRNA methylation sites coupled with aberrant pre-RNA processing.[69] Point mutations in the active site of yeast fibrillarin inhibit overall ribose methylation in pre-rRNA while similar mutations in the archaeal fibrillarin homolog (*Archaeoglobus fulgidus*) abolish methylation of a synthetic target RNA in an *in vitro* assay.[70]

Nop56 and Nop 58 are homologous, highly conserved eukaryotic core proteins which display 37% and 45% sequence identity in human and yeast, respectively.[71,72] The fact that Nop56 and Nop58 are closely related and found in Archaea as a single homolog strongly suggests a gene duplication event early in eukaryotic evolution.[28,72] Nop56 and Nop58 are essential proteins in yeast and were initially identified as a box C/D snoRNP core proteins through synthetic lethal Nop1 (fibrillarin) screens. The single archaeal Nop56/58 homolog was subsequently identified based upon sequence analysis and comparison to eukaryotic Nop56 and Nop58.[60] Though the function of Nop56 and Nop58 are not yet known, these proteins are essential for box C/D RNA-guided

methylation and are believed to be structural proteins important for stabilizing snoRNA conformation and/or providing a scaffold for RNP assembly and subsequent fibrillarin binding.[49,60,72]

Structural Organization of the Eukaryotic and Archaeal Box C/D RNP Complexes

While archaeal sRNP core protein L7 binds both the terminal K-turn and internal K-loop,[49,50] the eukaryotic homolog 15.5 kDa strikingly binds only the terminal K-turn.[35,62,65,73] The dissociation constant for these two proteins is quite similar and determined to be approximately 10 nM indicating their very tight affinity for the K-turn motif.[50] The binding of this core protein occurs first as demonstrated with *in vitro* RNP assembly systems.[74]

Archaeal Nop56/58 is the next core protein to bind box C/D sRNAs and its binding distribution in archaeal complexes is distinctly different from Nop56 and Nop58 distribution on the eukaryotic box C/D snoRNAs.[49] The eukaryotic Nop56 and Nop58 core proteins are replaced by a single homolog in Archaea, and this single Nop56/58 core protein binds both the terminal box C/D core and internal C'/D' motif. In contrast, the eukaryotic Nop58 binds uniquely with the box C/D motif while Nop56 associates with the C'/D' K-turn. The difference in binding characteristics of the archaeal Nop56/58 and the eukaryotic Nop56 and Nop58 core proteins are believed to reflect the evolution of these core protein binding capabilities. Their binding is dependent upon L7/15.5kD binding but not that of fibrillarin.[72,75,76] Fibrillarin requires Nop56/Nop58 or Nop56 and Nop58 for binding but appears to interact with the guide RNA specifically making

contact with both the C/D and C'/D' motifs. This is the only eukaryotic box C/D snoRNP core protein to do so.[75] In Archaea, fibrillarin appears to dimerize with Nop56/58 and assembles on the guide RNA as a heterodimer. Thus, this heterodimer of fibrillarin and Nop56/58, present on the C/D and C'/D' motif, has been described as a protein "platform" essential for target RNA binding and nucleotide methylation.[77]

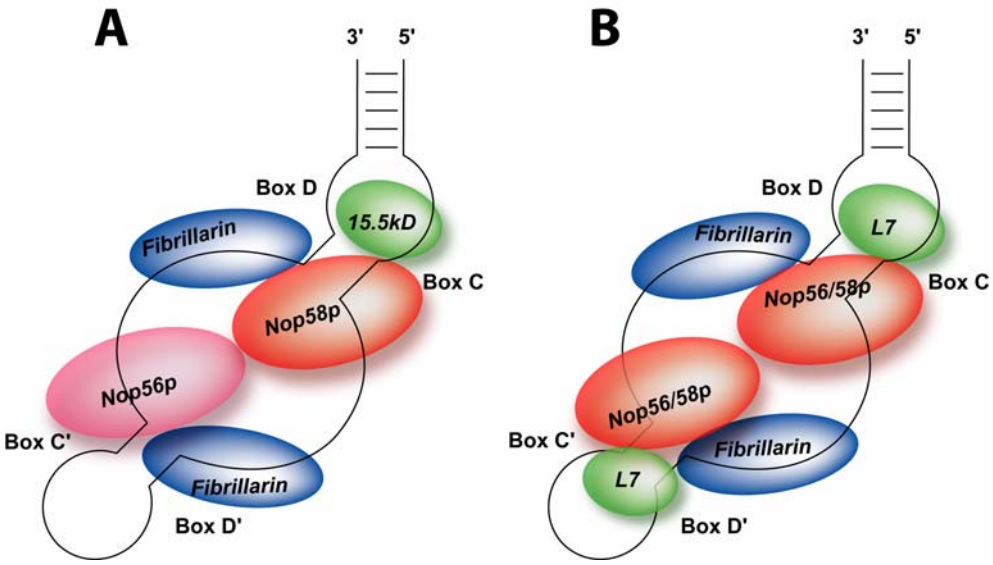

Figure 4. Eukaryotic and Archaeal Box C/D RNPs. A) Eukaryotic C/D RNPs assemble in an asymmetrical arrangement on the guide RNA with a single 15.5kD protein bound at the C/D motif. B) Archaeal C/D RNPs assemble symmetrically with one copy of each protein represented at each motif.

VIII. snoRNA Genomic Organization and Biosynthesis

The eukaryotic snoRNAs exhibit several different genomic organizations with specific organisms using a combination of gene organizations and different gene organizations predominating in different organisms. The majority of metazoan box C/D and H/ACA snoRNAs are encoded within introns of protein-coding genes. These snoRNAs are transcribed as part of the pre-mRNA transcript by RNA Polymerase II (pol II) and released during pre-mRNA splicing events.[4,78-81] Originally, intron-encoded

snoRNAs were thought to exist only in protein-coding pre-mRNAs, but later studies discovered intronic snoRNAs within pre-mRNAs that do not code for proteins.[15,19,82,83] As processed introns, these snoRNAs do not possess methylated 5' caps nor poly A tails. In contrast, the non-intronic snoRNAs, such as U3, U8, and U13, are independently transcribed by RNA Pol II and do possess 5' methyl caps and 3' poly-adenylated tail modifications.[4,9,10]

Host genes that harbor snoRNAs within their introns often encode proteins with related functions. The proteins of most host genes play roles in nucleolar function, ribosome structure, or protein synthesis.[4,84] This organization may be significant in terms of co-regulating snoRNA biosynthesis with the synthesis of specific proteins important in ribosome biogenesis or function. However, since not all intronic snoRNAs are located within such host genes, this hypothesis may not be applicable for all snoRNAs.[4,19,85-87]

Maturation of intron-encoded snoRNAs requires the excision of the host intron from the pre-mRNA primary transcript, lariat debranching of the host intron and subsequent exonucleolytic trimming of the intron sequences flanking the snoRNA.[88,89] A second, less utilized pathway for snoRNA maturation involves direct excision from the host intron using intron-specific endonucleases followed by exonucleolytic trimming.[88,90,91] Correct processing of snoRNAs from host pre-mRNAs requires the initial packaging of the snoRNAs into snoRNP complexes. Immunoprecipitation studies have demonstrated the presence of core proteins bound to the intronic snoRNA precursor prior to final snoRNA maturation.[45,92] Furthermore, mutation of the terminal C and D

boxes of the box C/D snoRNAs and the H box of the H/ACA snoRNAs , sequence elements all critical for core protein binding, are detrimental to intronic snoRNA biosynthesis.[93-94] Also, depletion of fibrillarin and Nop58, which are known to bind the C and D boxes, reduces the accumulation of snoRNAs.[76] These findings have led to the processing model where the core proteins are assembled on the precursor snoRNA while still contained in the intron and the binding of the core proteins defines the stop sites for the trimming exonucleases during intronic snoRNA processing.[68]

Unlike metazoan snoRNAs, most yeast snoRNAs are organized in monocistronic, independently transcribed units, although some are transcribed as polycistronic RNAs under the control of a single promoter. Only a few yeast snoRNAs are processed from pre-mRNA introns.[17,95-98] Interestingly, plant snoRNA biosynthesis and processing is more closely related to that of yeast, as the vast majority of plant snoRNAs are independent Pol II transcripts arranged as polycistronic snoRNAs.[98] Strikingly, the sequence elements crucial for processing in animals, plants, and fungi are identical and conserved for intronic, non-intronic, and polycistronic snoRNAs.

Transcription and processing of the archaeal sRNAs is not well understood at this time. Most of the identified sRNAs are located within spacer regions between protein open reading frames suggesting their transcription as independent RNAs.[28] To date, there is only one example of an archaeal sRNA being encoded within the intron of another gene. The coding sequences for sR40 sRNA of *Pyrococcus*, sR3 of *A. fulgidus* and *H. volcanii* are encoded within the single intron of the $tRNA^{Trp}$ gene.[27,29,82] The box C/D

and C'/D' guide sequences are complimentary to two sites in the mature $tRNA^{Trp}$ and these intronic elements are responsible for these two $tRNA^{Trp}$ methylation events.[25-28] Despite original suggestions that the tRNA methylations were guided *in cis* by the intron-encoded box C/D and C'/D' RNPs of the precursor tRNA, subsequent work has demonstrated that these two nucleotide modifications are carried out *in trans* using the excised $tRNA^{Trp}$ intron.[29,82]

IX. Alternative Roles for snoRNAs in pre-rRNA Processing

The primary role of the eukaryotic snoRNAs and archaeal sRNAs is to guide the site-specific modification of target RNA nucleotides. Key in the nucleotide modification mechanism is the base pairing of the guide RNA with the target RNA. Intermolecular hydrogen bonding of guide RNAs with other RNA species is a feature shared by all guide RNAs, but for certain species this intermolecular interaction serves additional functions apart from directing nucleotide modification.

Directing rRNA Cleavage or Chaperoning rRNA Folding.

The first snoRNAs described in both humans and yeast were shown to be important for pre-rRNA processing and the site-specific cleavage of the precursor transcript. These cleavage events are directed by base pairing between the pre-rRNA and antisense region of the snoRNA. U3, U14, U22, snR10, and snR30 are all snoRNAs associated with early cleavage events.[4,11-14] Other snoRNAs, such as U8 and MRP, are involved in later processing events. MRP RNA is specifically involved in 5.8S rRNA maturation.[41,99]

Some snoRNAs possess guide sequences that are complimentarity to rRNA but are not involved in nucleotide modification and may or may not be involved in pre-rRNA cleavage events. These snoRNAs act as RNA chaperones, base pairing with the pre-rRNA to either prevent rRNA misfolding or even facilitate the correct folding of the precursor. Two such snoRNAs are U3 which assists in 5' end pre-rRNA folding as well as pre-rRNA cleavage, and U8 which chaperones the folding of the 28S/25S rRNA late in pre-rRNA processing events. For each of these snoRNAs, they base pair with multiple sequences on the pre-RNA to assist in folding of the precursor transcript.[12,100-103]

SnoRNAs Involved in Gene Regulation.

Recent work has suggested a potential role for specific snoRNAs in the regulation of specific target gene transcripts. The lone example and basis for this speculation is the brain-specific snoRNA MBII-52. Experiments have shown that this box C/D snoRNA regulates the splicing pathway of its target mRNA, the serotonin receptor gene 5-HT2c. Preliminary investigation of MBII-52 and 5-HT2c indicate a potential target adenosine for methylation. This particular nucleotide is also the target of adenosine-to-inosine mRNA editing[104,105] within a critical region required for alternative pre-mRNA splicing.[105,106] Methylation of the target adenosine may play a role in regulating both the function and accumulation of the serotonin receptor by impeding 5-HT2c editing and/or correct splicing.[105,107]

X. Targets of Eukaryotic snoRNA and Archaeal sRNA Function

The eukaryotic snoRNAs were originally characterized as guide RNAs that function in the post-transcriptional processing and modification of ribosomal RNA. However, further characterization in Eukarya and Archaea has now expanded the list of target RNAs to include the snRNAs, tRNAs, and even some mRNAs.

Ribosomal RNAs:. Primarily, snoRNAs direct the 2'-O methylation or pseudouridylation of pre-ribosomal rRNA. These RNA-guided nucleotide modifications in yeast, human, and multiple archaeal species tend to cluster in an area of the folded rRNA that is located within the peptidyl transferase center of the ribosome. These studies suggest that the conserved location of these modifications may play essential roles in the structure and/or function of the catalytic center of the ribosome.[108-110] Of course, snoRNA function in the nucleolus also includes roles in pre-rRNA cleavage events and chaperone functions in pre-rRNA folding.

Small nuclear RNAs: Apart from pre-rRNA, other snoRNA targets for nucleotide modifications include the small nuclear RNAs. The splicesomal snRNAs U1, U2, U4, U5, and U6 are known to contain several sites that are either 2'-O-methylated or pseudouridylated by specific snoRNA species.[111,112] The substantial number of modified bases in the snRNAs is conspicuously grouped in locations in the snRNA involved in intermolecular RNA-RNA interactions. This observation has led to the hypothesis that snoRNA-guided modifications of snRNAs are particularly important for assembly and catalytic function of the splicesome.[111-112]

Transfer RNAs: RNA-guided modification of tRNAs is specific to archaeal organisms as no yeast or mammalian tRNA targets have yet been described.[19,27] Archaeal sRNAs which target tRNAs were first identified based on a lack of sRNA sequence complimentarity to rRNA. Subsequent sequence analysis led to the identification of various target tRNAs. The first and most studied example of a tRNA modified by guide RNAs is the $tRNA^{Trp}$, from *Haloferax volcanii*, modified by guide sequences located within pre-$tRNA^{Trp}$ sequence.[25-29] Not long after the initial characterization of the *H. volcanii* $tRNA^{Trp}$ sRNA, four other sRNAs that target tRNAs for modification were identified in *Pyrococcus horokoshii*. sR47, sR48, sR49, and sR50 all contain a nucleotide antisense element complimentary to a sequence within tRNA-Leu (CAA), tRNA-Leu(UAA), elongator tRNA-Met, and tRNA-Trp, respectively.[19,27] Recently, twenty box C/D sRNAs have been identified that direct methylation of 19 tRNAs in *Pyrobaculum aerophilum*.[113] Although experimental evidence confirming tRNA modification by sRNAs is ongoing, the conservation of modified nucleotides at the predicted sites within tRNAs is a good indication of sRNA guided modification. Analysis has indeed demonstrated that the single sR17 sRNA of *Aeropyrum pernix* modifies the G10 nucleotide of 19 different tRNAs.[28]

Messenger RNAs: Messenger RNAs are also known targets of snoRNAs although rarer and less well characterized than other target RNAs. Currently, there are only two known mRNAs that are candidates for modification by snoRNAs. The spliced leader of *Trypanosoma* mRNA contains a highly conserved pseudouridine within a sequence that

is complimentary to another RNA displaying characteristics of the box H/ACA snoRNAs. *In vitro* studies indicate that disruption of the canonical duplex between the guide RNA and spliced leader (SL) RNA abolish pseudouridylation of the target nucleotide.[114] Likewise, the serotonin receptor 5-HT2c mRNA, specifically expressed in brain tissue, is a proposed methylation target for snoRNA MBII-52, also specifically expressed in the brain. Preliminary studies suggest that MBII-52 may play a role in the alternative splicing pathways of 5-HT2c as well as in the modification of its target base.[105]

XI. *In Vitro* Reconstitution of a Catalytically Active Box C/D RNP

In 2002, recombinant archaeal box C/D core proteins from *Sulfolobus acidocaldarius* and *in vitro*-transcribed sRNA were successfully assembled into a functional sRNP complex that was capable of site-specific, 2'-O-methylation of a target RNA *in vitro*. Binding of L7, Nop56/58, and fibrillarin core proteins to the sRNA was demonstrated by an electrophoretic mobility-shift analysis to assemble a stable sRNP complex. Activity of this sRNP was demonstrated with the incorporation of a tritiated methyl group donated by the ^{3}H-S-adenosyl-methionine (SAM) substrate into the target RNA. Specificity of the reaction for the target nucleotide was confirmed by 2-dimensional TLC analysis[74] and this reaction required fibrillarin, the methylase enzyme. Omer and colleagues further showed that Watson-Crick pairing between the sRNA guide sequence nucleotide and the target nucleotide was required for methylation.

Work in our lab has also established an *in vitro* sRNP assembly system using recombinant core proteins and *in vitro* transcribed RNA of *Methanocaldococcus jannischii.* The selected box C/D sRNA, sR8, exhibits perfectly conserved box elements C, D, C', and D', folds into the expected terminal box C/D and internal C'/D' K-turn structures[115], and a five nucleotide terminal stem flanks the box C/D core motif.[62] *In vivo*, sR8 directs methylation of both pre-rRNA and tRNA, thereby employing both of its 12 nucleotide antisense elements for nucleotide modification. Initial experiments demonstrated the sequential binding of the sRNP core proteins and the methylation of two different synthetic target RNAs, each of which base paired with the D or D' guide regions. Subsequent work has demonstrated the importance of juxtaposed box C/D and C'/D' RNPs for efficient methylation activities and the importance of conserved spacing between these two complexes for nucleotide modification.[40,62] Most recently, we have demonstrated that self-dimerization of the Nop56/58 coiled-coil domain is not important for sRNP assembly nor nucleotide modification, but this domain itself is critical for nucleotide methylation activities.[77]

One sRNP structural feature shown to be important for archaeal box C/D sRNA-guided nucleotide methylation is the inter-RNP spacing between the box C/D and C'/D' motifs. Through extensive database searches encompassing many archaeal species, Tran et al (2005)[62] demonstrated a highly constrained spacing between the terminal and internal box elements. Approximately 40% of the 245 RNAs investigated exhibited inter-motif spacing of 12 nucleotides for both spacer regions separating the C/D and C'/D' motifs. For those sRNAs where one spacer region significantly deviates from

typical 12 nucleotides, the other spacer almost always exhibits the conserved 12 nucleotide length.[62] This suggests that inter-RNP interactions are likely to play critical roles in methylation activity and perhaps specificity.

Alteration of the spacing between the box C/D and C/D' motifs had no affect upon core protein binding to either RNA motif nor sRNP assembly.[62] However, shortening or lengthening spacer regions severely disrupted nucleotide methylation activity for both RNP complexes. Interestingly, when only one spacer region was altered, the enzyme activity of the corresponding RNP was disrupted but the other complex was unaffected. This observation again suggests that potential protein crosstalk interactions between the motifs are important for sRNP methylation activity.[62]

XII. Structural and Functional Features of the Target-Guide Duplex

In vitro assembly systems now provide model systems to facilitate the systematic study of target RNA methylation. Apart from the core proteins and conserved box elements, specific structural features of the sRNA guide sequence:target RNA duplex may also play important roles in not only the methylation function but also target nucleotide specificity. Such features of the sRNP complex and the guide RNA:target RNA duplex are discussed here.

Guide Nucleotide:Target Nucleotide Interaction Requires Watson-Crick Base Pairing.

Existing data examining guide sequence:target RNA interaction specifically at the site of nucleotide methylation is somewhat contradictory at present. Omer and

colleagues first demonstrated the importance of a Watson-Crick pair at the target nucleotide when they examined guide sequence:target RNA interaction in their *in vitro* box C/D sRNP assembly system.[74] Since then, other labs have examined this feature of guide RNA:target RNA interaction and found that Watson-Crick base pairing may not be a universal requirement for sRNA-guided methylation. *In vivo* studies examining snoRNA-guided nucleotide modification in mouse L929 cells revealed that in addition to Watson-Crick pairs, methylation proceeded when the target nucleotide was paired using the non-Watson-Crick G:U base pair. However, other non-Watson-Crick pairs did not permit nucleotide methylation.[116] In contrast, Singh et al (2005) recently established an *in vitro* sRNP assembly and target RNA methylation system using $tRNA^{Trp}$ from *Haloferax volcanii*. This box C/D guide RNA and target RNA are unique in that they are one in the same, with the intron-located box C/D and C'/D' motifs guiding the 2'-O-methylation of two target $tRNA^{Trp}$ nucleotides contained in the $tRNA^{Trp}$ precursor RNA. Their analysis of the base pairing requirements at the target site revealed that most any base pair combination other that G:G, C:C, and U:U resulted in some degree of methylation of the target nucleotide. In this study, the specificity of nucleotide methylation was verified by two-dimensional TLC analysis, demonstrating the specific methylation of the target nucleotide despite interacting with the guide nucleotide using non-Watson-Crick pairing or base mismatches.[29]

These discrepancies may reflect not only differences in specificity between *in vivo* and *in vitro* systems, but also differences between the various *in vitro* systems and the conditions under which methylation is carried out. Thus, specificity may not be a

global feature of any model system chosen for experimentation. Possible factors affecting sRNP enzyme specificity are more fully examined and discussed in the following chapter using our *M. jannaschii in vitro* sRNP assembly and methylation system.

Structural Features of the Guide Sequence:Target RNA Duplex

Recent studies have explored some structural features of the sRNA guide:target RNA duplex. *In vivo* studies were carried out with an engineered snoRNA guide sequence of 17 nucleotides complementary to a co-transfected, target RNA-containing vector. These investigations demonstrated a minimal guide:target duplex was required for snoRNP-guided nucleotide methylation. A target-guide duplex of 8 base pairs was incapable of promoting methylation activity. As the guide:target duplex length was increased from 9 to 17 base pairs, a corresponding increase in methylation activity was observed.[116]

Additional examination of guide:target RNA duplex structure has indicated a degree of flexibility in this RNA:RNA helix depending upon its base composition. RNA:RNA duplexes composed of more that fifty percent A:U base pairs required a Watson-Crick base pair between the guide and target nucleotides. However, duplexes composed of primarily G:C base pairs are often able to accommodate a mismatch at the target site and still exhibit methylation activity. Similarly, a long G/C-rich duplex can tolerate bulges introduced away from the target site. Nevertheless, one nucleotide insertions immediately upstream or downstream of the target nucleotide disrupt

methylation. Presumably, bulges in close proximity to the target nucleotide sterically hinder appropriate RNA-protein contacts or, alternatively, alter the positioning of the target nucleotides 2' hydroxyl group within the catalytic center of fibrillarin.[116]

XIII. Synopsis of Thesis Research

The longstanding working model for RNA-guided modification dictates that a Watson-Crick, base paired RNA duplex is formed between the snoRNA/sRNA's guide region and the target RNA sequence. With the identification of snoRNAs that guide methylation despite non-canonical base pairs at the target nucleotide or vary in methylation efficiency based upon RNA:RNA duplex base composition, this working model of guide sequence:target nucleotide interaction requires further examination. Our established archaeal *in vitro* box C/D sRNP assembly system can directly explore those sRNA:target RNA interactions needed for nucleotide methylation both easily and directly.

The research described in Chapter 1 examines the structural features of the archaeal guide-target duplex that influence nucleotide 2'-0-methylation. These studies were carried out using the *in vitro* assembly and methylation assay developed in our laboratory. *M. jannaschii* sR8 sRNA and recombinant core proteins were used to assemble sRNP complexes possessing both box C/D and C'/D' RNP complexes. The importance of Watson-Crick base pairing between guide and target nucleotides was assessed. For the *in vitro* assembled *M. jannaschii* box C/D sRNP, methylation absolutely required Watson-Crick base pairing both at the target nucleotide and within

the guide RNA:target RNA duplex. However, methylation of mismatched base pairs could be induced under very elevated concentrations of $MgCl_2$. Other structural features of this RNA:RNA duplex were also examined. Using chimeric DNA/RNA oligos as target RNAs for the assembled sR8 sRNP, we demonstrated that inserted single DNA nucleotides in RNA target had no effect upon methylation unless inserted at the target nucleotide site. However, replacement of the RNA target with a DNA oligonucleotide revealed the importance of an RNA:RNA duplex for nucleotide methylation. Finally, the ability of the guide RNA to efficiently methylate a target nucleotide positioned within a highly structured RNA hairpin or the loop region of an RNA hairpin-loop structure was examined. Interestingly, target nucleotides located within a highly folded RNA target affected methylation efficiency somewhat, but still permitted modification of the target nucleotide. This suggests an ability of the assembled sRNP complex to facilitate target RNA unfolding needed for the methylation reaction. Collectively, these results demonstrated the importance of forming an A form, Watson-Crick RNA:RNA duplex upon the Nop56/58-fibrillarin protein platform for efficient, box C/D sRNP-guided nucleotide 2'-O-methylation.

REFERENCES

1. Muramatsu, M., J. L. Hodnett, and H. Busch (1968) *Base composition of fractions of nuclear and nucleolar ribonucleic acid obtained by sedimentation and chromatography.* J. Biol. Chem. **241**:1544-1550.

2. Nakamura, T., A. Prestayko, and H. Busch (1968) *Studies on nucleolar 4 to 6S ribonucleic acid of Novikoff hepatoma cells.* J. Biol. Chem. **243**:1368-1375.

3. Busch, H., R. Reddy, L. Rothblum, YC. Choi (1982) *SnRNAs, SnRNPs, and RNA processing.* Ann. Rev. Biochem. **51**:617-54.

4. Maxwell, E. S and M. J. Fournier (1995) *The small nucleolar RNAs*. Ann. Rev. Biochem. **35**:897-934.

5. Vitali, P., et al. (2003) *Identification of 13 novel human modification guide RNAs.* Nucleic Acids Res. **31**:6543-6551.

6. Terns, M. and R. Terns (2002) *Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin*. Gen. Express. **10**:17-39.

7. Fournier, M. J., and E. S. Maxwell (1993) *The nucleolar snRNAs: catching up with the splicesomal snRNAs.* Trends Biochem Sci. **18**:131-135.

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

9. Wise, J. A., et al. (1983) *Yeast contains small nuclear RNAs encoded by single copy genes.* Cell **35**:743-751.

10. Reidel, N., et al. (1986) *Small nuclear RNAs from Saccharomyces cerevisiae: unexpected diversity in abundance, size, and molecular complexity*. Proc Natl Acad Sci,. **83**:8097-101.

11. Li, H. V, J. Zagorski, and M. J. Fournier (1990) *Depletion of U14 small nuclear RNA (snRl28) disrupts production of 18S rRNA in* Saccharomyces cerevisiae*.* Mol. Cell. Biol. **10**:1145-1152.

12. Kass S., K. Tycowski, J. A. Steitz, B. Sollner-Webb (1990) *The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing*. Cell. **60**:897-908.

13. Morrisey J. P. and D. Tollervey (1993) *Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis.* Mol Cell Biol. **13**:2469-2477.

14. Tycowski, K. T., M. D. Shu, J. A. Steitz (1994) *Requirement for intron-encoded*

U22 small nucleolar RNA in 18S ribosomal RNA maturation. Science **2**:1558-61.

15. Tycowski, K., et al.(1996) *A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus.* Proc. of the Natl. Acad. Sci. **93**:1480-1485.

16. Kiss-Laszlo, Z, et al. (1996) *Site-specific ribose methylation of pre-ribosomal RNA: a novel function for small nucleolar RNAs.* Cell. **85**:1077-1088.

17. Lowe T. M., S. R. Eddy (1999) *A computational screen for methylation guide snoRNAs in yeast.* Science **283**:1168-1171.

18. Cavaille J., M. Nicoloso, J.P. Bachellerie (1996) *Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides*. Nature **383**:732–735.

19. Bachellerie, J. P., J. Cavaille, A. Huttenhofer (2002) *The expanding snoRNA world*. Biochemie **84**:775-790.

20. Ganot , P., M. Caizergues-Ferrer, and T. Kiss (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.

21. Balakin, A., L. Smith, and M. Fourneir (1996), *The RNA World of the nucleolus: two major familes of small RNAs defined by different box elements with related functions.* Cell **86**:823-834.

22. Noon K. R., E. Bruenger, J.A . McCloskey (1998) *Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus.* J Bacteriol. **11**:2883-2888.

23. Tang, T.H., et al. (2002) *Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus*. Proc. Natl Acad. Sci. **99**:7536–7541.

24. Gaspin, C., J. Cavaille, G. Erauso and J. P. Bachellerie (2000) *Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: Lessons from the Pyrococcus genomes.* J. Mol. Biol. **297**:895-906

25. Omer, A. D., et al. (2000) *Homologs of small nucleolar RNAs in Archaea*. Science **288**:517-522.

26. Gupta, R. et al (1984) *Halobacterium volcanii* tRNAs. *Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNA.,* J. Biol. Chem. **259**:9461–9471.

27. d'Orvall, B. C. et al. (2001) *Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guide the formation of two ribose-methylated nucleosides in the mature tRNATrp.* Nucleic Acids Res. **29**:4518-4529.

28. Dennis P. P., A. Omer, and T., Lowe (2000) *A guided tour: small RNA function in Archaea.* Mol Microbiol. **40**:509-19.

29. Singh S. K., et al. (2004) *Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA.* J Biol Chem. **279**:47661-71.

30. Ofengand J., M.Del Campo, Y. Kaya (2001) *Mapping pseudouridines in RNA molecules.* Methods **25**:365-73.

31. Olsen, G. J. and C. R. Woese (1997) *Archaeal genomics: an overview*. Cell **89:**991–994.

32. Tran, E., J. Brown, E.S. Maxwell (2004) *Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus?* Trends in Biochem. Sci. **29**:343-350.

33. Ganot P, et al. (1997) *Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs*. Cell **89**:799-809.

34. Klein, D., et al. (2001) *The kink-turn: a new RNA secondary structure motif*. EMBO J. **20**:4214-4221.

35. Vidovic I., S. Nottrott, K. Hartmuth, R. Luhrmann, R. Ficner (2000) *Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment*. Mol. Cell. **6**:331–1342.

36. Kiss-Laszlo, Z., Y. Henry, and T. Kiss (1998) *Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA.* EMBO J. **17**:797-807.

37. Rashid, R., et al. (2003) *Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles*. J. Mol. Biol. **33**:295-306

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

39. Nicoloso, M., L.H. Qu, B. Michot, J.P. Bachellerie (1996) *Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points*

to their direct role as guides for the 2'-O-ribose methylation of rRNAs. J. Mol. Biol. **260:**178–195.

40. Tran E., et al. (2005) *Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs.* RNA **11**:285-293.

41. Schmitt, M. E. and D. A. Clayton (1993) *Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in Saccharoyces cerevisiae*. Mol. Cell. Biol. **13**:2433-2443.

42. Morrissey, J. P. and D. Tollervey (1995) *Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system*. TIBS **20**:78-82.

43. Girard, Jean-Philippe et al. (1992) *GAR1 is an essential small nucleolar RNP protein required fro pre-rRNA processing in yeast*. EMBO J. **11**:673-682.

44. Henras, A. Y. et al., (1998) *Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs*. EMBO. J. **17:**7078–7090.

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

46. Watkins N.J., A. Gottschalk, G. Neubauer, B. Kastner, P. Fabrizio, M. Mann, R. Luhrmann (1998) C*bf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure*. RNA **4**:1549–1568.

47. Koonin, E.V., et al. (1996) Pseudouridine synthases: *Four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases*. Nucleic Acids Res. **24:**2411–2415.

48. Nurse, K., J. Wrzensinski, A. Bakin, B.G. Lane, and J. Ofengand (1995) *Purification, cloning, and properties of the tRNA C55 synthase from* Escherichia coli. RNA **1:**102–112.

49. Rashid, R. et al. (2006) *Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and Dyskeratosis Congenita*. Mol. Cell **21**:249-260.

50. Kuhn, J., E. Tran, and E. S. Maxwell (2002) *Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5 kDa/Snu13p snoRNP core protein*. Nucleic Acids Res. **30**:931-941.

51. Rozhdestvensky T.S., 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*.Nucleic Acids Res. **31**:869-877.

52. Watanabe, Y. and M. W. 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.* Nucleic Acids Res. **28**:2342-2352.

53. Bult C. J., et al. (1996) *Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii*. Science **73**:1058-1073.

54. Girard JP, et al. (1994) *Identification of a segment of the small nucleolar ribonucleoprotein- associated protein GAR1 that is sufficient for nucleolar accumulation*. J. Biol. Chem. **269**:18499–18506.

55. Henras A., et al. (2001) *Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p*. Nucl. Acids Res. **29**:2733–2746.

56. Weng, C. and U. T. Meier (2004) *Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins*. EMBO J. **23**:1857-1867.

57. Baker, D. L., et al.(2005) *RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP*. Genes and Dev. **19**:1238-1248.

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

59. Amiri, K. (1994) *Fibrillarin-like proteins occur in the domain Archaea*. J. Bacteriol. **176**:2124-2127.

60. Wu, P., et al. (1998) *Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast*. J. Biol. Chem. **273**:16453-16463.

61. Lafontaine, D. L., and D. Tollervey (1999) *Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability*. RNA **3**:455-67.

62. Tran, E., X. Zhang, and E. S. Maxwell (2003) *Efficient 2'-0 methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' sRNP core proteins.* Nat. Struct. Biol. **22**:3930-3940.

63. Koonin, E., P. Bork, and C. Sander (1994) *A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme?*, Nuc Acids Res. **22**:2166-2167.

64. Watkins NJ, A. Dickmanns, R. Luhrmann (2002) *Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP*. Mol. Cell Biol. **23**:8342-52.

65. 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

66. Nottrott S., et al. (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.

67. Wang H., et al. (2000) *Crystal structure of a fibrillarin homologue from* Methanococcus jannaschii*, a hyperthermophile, at 1.6 A resolution*. EMBO. J. **19**:317–323.

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

69. Tollervey D., et al. (1993) *Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly*. Cell **72:**443–457.

70. Aittaleb, M., et al. (2004) *Structural and thermodynamic evidence for a stabilizing role of Nop5p in S-adenosyl-L-methionine binding to fibrillarin*. J Biol Chem. **279**:41822-41829.

71. Lyman S. K., Gerace L., Baserga S. J. (1999) *Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins*. RNA **12**:1597-604.

72. Gautier, T., et al. (1997) *Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis*. Mol. Cell. Biol. **17**:7088-7098.

73. Szewczak, L. B., et al. (2002) *Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP*. Chem Biol. **10**:1095-107.

74. Omer A.D., et al. (2002) *In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex*. Proc Natl Acad Sci. **99**:5289-5294.

75. Cahill, N. M., et al. (2002) *Site specific cross-linking analyses reveal an asymmetric protein distribution for a Box C/D snoRNA*. EMBO J. **21**:3816-3828.

76. Lafontaine, D and D. Tollervery (2000) *Synthesis and assembly of the box C+D snoRNPs*. Mol. Cell. Biol. **20**:2650-2659.

77. Zhang X., et al. (2006) *The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation*. RNA **12**:1092-1103.

78. Tycowski K.T., M.D. Shu, J.A. Steitz (1993) *A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3*. Genes Dev. **7**:1176-90.

79. Fragapane P., et al. (1993) *A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA*. EMBO J. **12**:2921-2928.

80. Kiss, T and W. Filipowicz (1993) *Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1*. EMBO J. **12**:2913-2920.

81. Liu J, and E. S. Maxwell (1990) *Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene*. Nucleic Acids Res. **18**:6565-6571.

82. Bortolin, M. L. and T. Kiss (1998) *Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding*. RNA **4**:445-454.

83. Pelczar, P. and W. Filipowicz (1998) *The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family.* Mol Cell Biol. **18**:4509-4518.

84. Filipowicz, W., P. Pelczar, V. Pogacic, F. Dragon (1999) *Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification*. Acta Biochim Pol. **46**:377-389.

85. Renalier, M. H., et al. (2005) *The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP*. RNA **11**:1051-1063.

86. Xia, L. et al. (1995) *Intronic U14 snoRNAs of Xenopus laevis are located in two different parent genes and can be processed from their introns during early oogenesis*. Nucl. Acids Res. **23**:4844–4849.

87. Cervelli, M. et al. (2002) *Comparative structure analysis of vertebrate U17 small nucleolar RNA (snoRNA)*. J. Mol. Evol. **54:**166–179.

88. Petfalski E., et al. (1998) *Processing of the precursors to small nucleolar RNAs and rRNAs requires common components*. Mol Cell Biol. **18**:1181-1189.

89. Weinstein, L.B. and J.A. Steitz (1999) *Guided tours: from precursor snoRNA to functional snoRNP*. Curr. Opin. Cell. Biol. **11:**378–384.

90. Ooi, S. L., et al. (1998) *Intronic snoRNA biosynthesis in* Saccharomyces cerevisiae *depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA.* RNA **4**:1096-1110.

91. Villa, T., et al. (1998) *Processing of the intron-encoded U18 small nucleolar RNA in the yeast* Saccharomyces cerevisiae *relies on both exo- and endonucleolytic activities*. Mol Cell Biol. **18**:3376-83.

92. Watkins N. J., et al. (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.

93. Huang L., et al. (1998) *A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst*. Biochemistry **37**:344-351.

94. Cafarelli, E., et al. (1998) *In vivo identification of nuclear factors interacting with conserved elements of Box C/D small nucleolar RNAs*. Mol Cell Biol. **18**:1023-1028.

95. Samarsky, D. A., and M. J. Fournier (1999) *A comprehensive database for the small nucleolar RNAs from* Saccharomyces cerevisiae. Nucleic Acids Res. **27**:161-164.

96. Qu, L. H., et al. (1999) *Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast*. Mol Cell Biol. **19**:1144-1158.

97. Eliceiri, G. L. (1999) *Small nucleolar RNAs*. Cell. Mol. Life Sci. **56**:22–31.

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

99. Peculis, B. A., J. A. .Steitz (1993) *Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte.* Cell **73**:1233-1245.

100. Savino, R., S.A. Gerbi (1990) *In vivo disruption of Xenopus U3 snRNA affects ribosomal RNA processing.* EMBO J. **9**:2299-2308.

101. Hughes J. M. X. and M. Ares Jr.(1991) *Depletion of U3 small nucleolar RNA inhibits cleavage in the 5% external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA*. EMBO J. **10:** 4231–4239.

102. Peculis, B. A., J. A. Steitz (1994) *Sequence and structural elements critical for U8 snRNP function in Xenopus oocytes are evolutionarily conserved*. Genes Dev. **8**:2241-2255.

103. Peculis, B. A., J. G. .Gall (1992) *Localization of the nucleolar protein NO38 in amphibian oocytes.* J Cell Biol. **116**:1-14.

104. Burns, C., et al. (1997) *Regulation of serotonin-2C receptor G-protein coupling by RNA editing*. Nature **387**:303–308.

105. Cavaille, J., et al. (2000) *Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization.* PNAS **97**: 14311-14316.

106. Canton H., et al. (1996) *Identification, molecular cloning, and distribution of a short variant of the 5-hydroxytryptamine2C receptor produced by alternative splicing*. Mol Pharmacol. **50**:799-807.

107. Yi-Brunozzi, H. Y., et al. (1999) *Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2.* Nucleic Acids Res. **27**:2912-2917.

108. Decatur, W. A. and M. J. Fournier (2002) *rRNA modifications and ribosome function*. Trends Biochem Sci. **27**:344-351.

109. Hansen, M. A., et al. (2002) *Posttranscriptional modifications in the A-loop of 23S rRNAs from selected archaea and eubacteria*. RNA **8**:202-213.

110. Tollervey, D. (1996) *Trans-acting factors in ribosome synthesis*. Exp Cell Res. **229**:226-232.

111. Massenet, S., et al. (1998) Posttranscriptional modifications in the U snRNAs, in: H. Grosjean, R. Benne (Eds.), Modification and Editing of RNA: The Alteration of RNA Structure and Function, ASM Press, Washington, DC,. 201–228.

112. Gu, J., et al. (1996) *Localization of modified nucleotides in Schizosaccharomyces pombe spliceosomal small nuclear RNAs: Modified nucleotides are clustered in functionally important regions*. RNA **2**:909-918.

113. Renalier, M. H., et al. (2005) *The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP*. RNA **11**:1051-1063.

114. 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.

115. Suryadi, J., et al. (2005) *The crystal structure of* the Methanocaldococcus jannaschii *multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs*. Biochemistry **44**:9657-9672.

116. Cavaille, J. and J. P. Bachellerie (1998) *SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction.* Nucleic Acids Res. **26**:1576-1587.

CHAPTER I.

Structural Features of the Guide:Target RNA Duplex Required for Archaeal C/D sRNA Guided Nucleotide 2'-*O*-methylation.

C. Denise Appel and E. Stuart Maxwell

Department of Molecular and Structural Biochemistry

North Carolina State University

Raleigh, NC

ABSTRACT

Archaeal box C/D sRNAs guide the 2'-O-methylation of target nucleotides in both ribosomal and tRNAs. These small non-coding RNAs are characterized by conserved terminal box C/D and internal C'/D' RNA motifs. Each RNA motif binds three core proteins to establish individual RNP complexes that catalyze the site-specific 2'-O-methylation of target nucleotides. Specificity of nucleotide modification is determined by target RNA base pairing with complementary sRNA D or D' guide sequences. The fifth target nucleotide upstream from the D or D' box within the guide:target RNA is then methylated by the core proteins. *In vitro* assembly of *Methanocaldococcus jannaschii* sR8 box C/D RNA with recombinant core proteins, L7, Nop56/58, and fibrillarin produces a methylation-competent sRNP complex. This model box C/D sRNP has now been used to determine the structural features of the guide:target RNA duplex that are important for sRNA-guided nucleotide methylation. Watson-Crick pairing of guide and target nucleotides were essential for nucleotide methylation. Mismatched bases within the guide:target RNA duplex also disrupted target nucleotide methylation. Nucleotide methylation required that the guide:target duplex consist of an RNA:RNA helix as target deoxy-oligonucleotides possessing a target ribonucleotide are not methylated. Methylation specificity at the base paired guide:target nucleotide was compromised by elevated Mg^{2+} concentrations. In high divalent cation concentrations, target nucleotides not hydrogen bonded to the guide nucleotide were nevertheless methylated. Interestingly, D and D' target RNAs were methylated to different levels when deoxynucleotides were inserted within the target RNA or target methylation was carried out in elevated Mg^{2+} concentrations. These results suggested that structural features unique to the box C/D and C'/D' RNPs affect their nucleotide methylation capabilities. Finally, the ability of the sR8 box C/D sRNP to methylate target nucleotides positioned within highly structured RNA hairpins suggested an intrinsic ability of this archaeal RNA:protein enzyme to unwind double-stranded target RNAs prior to nucleotide modification.

INTRODUCTION

The primary function of the eukaryotic small nucleolar RNAs (snoRNAs) is to guide the nucleotide modification of ribosomal RNA and other cellular RNAs.[1-5] The snoRNAs are complexed with core proteins to assemble snoRNP complexes with the associated proteins carrying out the nucleotide modification reactions.[5-7] Nucleotide modification is accomplished through snoRNA guide regions that base pair with target RNAs to designate the specific nucleotide for modification.[2,8] The snoRNAs are classified into two large families based on conserved RNA sequence and secondary structural elements. The box H/ACA RNAs direct the isomerization of uridines to pseudouridines whereas the box C/D guide RNAs guide 2'-O-ribose methylation of targeted nucleotides.[9]

Investigations of archaeal genomes have identified numerous snoRNA-like small RNAs designated sRNAs with homologous functions to the eukaryotic snoRNAs.[10,11] Like the snoRNAs, archaeal sRNAs are classified as box C/D and H/ACA RNAs, assemble as sRNP complexes, and direct site-specific nucleotide 2'-O-methylation and pseudouridylation, respectively.[5,10] Similar to the snoRNAs, the sRNAs determine the specific nucleotide for modification also using guide sequences which base pair with the target RNA.[10] The conservation of box C/D and H/ACA guide RNAs in both Eukarya and Archaea strongly suggests that RNA-guided nucleotide modification is an ancient and evolutionarily conserved process.[12]

Archaeal box C/D sRNAs direct the site-specific 2'-O-methylation of designated nucleotides in rRNA and tRNA.[13-15] Box C/D RNAs are defined by conserved boxes C (RUGAUGA) and D (CUGA) located at the 5' and 3' termini of the sRNA, respectively. These sequence elements fold into a unique stem-bulge-stem structure where one stem is established by partial base pairing of boxes C and D. This unique secondary structure is known as a kink-turn (K-turn) and serves as a platform for core protein binding and sRNP assembly.[16] Internal boxes C' and D' are typically found in archaeal sRNAs and are highly conserved.[10,11,17] Boxes C' and D' fold into a K-turn-like structure, now referred to as the "K-loop" to denote the replacement of one of the terminal stems with a loop. The K-loop also serves as a core protein binding platform to assemble a second RNP complex.[18] Both the box C/D and C'/D' RNP motifs direct site-specific nucleotide modification via base pairing of target sequences to the guide regions located adjacent to the D and D' boxes.[19] Base pairing between the sRNA guide region and the target RNA positions that nucleotide to be modified five base pairs upstream from the D or D' box.[4]

Archaeal box C/D RNPs are assembled with three core proteins. Ribosomal protein L7, Nop56/58 (Nop5p), and fibrillarin were first identified as homologs to the eukaryotic box C/D snoRNP core proteins.[20-22] L7 specifically recognizes the K-turn and K-loop motifs and serves as the nucleation event for binding of the remaining core proteins.[23-25] Heterodimers of Nop56/58 and fibrillarin, the methyltransferase, bind the box C/D and C'/D' motifs such that one copy of each protein is present on both RNA motifs.[20,23,26] 2'-O-methylation depends on complete RNP assembly, as the absence of any one of the core proteins results in loss of nucleotide methylation. The assembly of

RNP complexes at both box C/D and C'/D' motifs allows the fully assembled sRNP to function as a dual guide RNA, directing 2'-O-methylation of target RNAs from both RNPs.[23]

Omer and colleagues first successfully assembled an archaeal sRNP *in vitro* using *Sulfolobus solfataricus* sRNA and recombinant core proteins and then demonstrated that this complex was capable of *in vitro* methylation of synthetic target RNAs.[27] Work in our laboratory has also established an *in vitro* sRNP assembly and methylation system using sR8 sRNA and recombinant core proteins of the archaeon *Methanocaldococcus jannaschii* .[23] sR8 sRNA exhibits perfect consensus C, D, C' and D' sequences and contains guide regions of 12 nucleotides in length, the typical and highly conserved inter-motif spacing of archaeal sRNAs.[23] This *in vitro* assembled *M. jannaschii* sRNP guides methylation of synthetic target RNAs from both the C/D and C'/D' RNP motifs. Strikingly, efficient methylation requires that the box C/D and C'/D' RNPS be juxtaposed in the sRNP indicating critical inter-RNP interactions are required for nucleotide modification.[23]

The *in vitro* assembled *M. jannaschii* sR8 sRNP provides a model box C/D RNP to examine in detail the interaction of the target RNA substrate with the guide sequence of the sRNP enzyme. The ability to mutate the target RNA permits the determination of those structural features of the guide:target RNA duplex that are important for sRNA-guided nucleotide methylation. Initial experiments demonstrated that Watson-Crick base

pairing between the guide and target nucleotide is absolutely essential for nucleotide methylation using this *in vitro* assembled sRNP. Subsequent experiments demonstrated that Watson-Crick pairing of an RNA:RNA duplex was also critical for nucleotide modification. The efficient methylation of target nucleotides positioned in highly structured target RNAs also suggested an inherent capability of box C/D sRNP to unfold target RNAs prior to nucleotide modification. Collectively, these studies further define structural features of the target RNA and the guide:target RNA duplex that are important for box C/D sRNP-guided nucleotide 2'-O-methylation.

MATERIALS AND METHODS

Construction and Synthesis of Wild Type sR8 and Mutant sR8 RNAs

Wild type and mutant sR8 DNA oligonucleotides were purchased from Integrated DNA Technologies. The sequences for wild type sR8 and sR8 GR Dtl were as follows:

sR8 WT: AAATCGCCAATGATGACGATTGGCTTTGCTGAGTCTGTGATGAAC
CGTATGAGCACTGAGGCGATTT

sR8 GR Dtl: AAATCGCCAATGATGACGATTGGCTTTGCTGAGTCTGTGATGAG
TATGATGAGCACTGAGGCGATTT

Point mutations in sR8 guide regions were introduced during PCR-amplification using the full length, wild type sR8 template and primers containing the mutated nucleotide. T7 polymerase promoter sequences were added to sR8 and sR8 guide mutants during PCR-amplification. RNA was transcribed *in vitro* using the AmpliScribe T7-Flash transcription kit (EPICENTRE) according to the manufacturer's protocols. Following transcription and DNase I treatment, all RNAs were gel purified.

Primer Pairs for Wild Type sR8 and sR8 Guide Mutants

Wild type sR8 (primers 1+2), sR8 C24G (primers 2+3), sR8 C24A (primers 2+4), sR8 C24U (primers 2+5), sR8 G51C (primers 1+6), sR8 G51A (primers 1+7), sR8 G51U (primers 1+8).

(1) CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTG

(2) AATCGCCTCAGTGCTCATACGG

(3) CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTGGGTTTG

(4) CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTGGATTTG

(5) CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTGGTTTTG

(6) AAATCGCCTCAGTGCTGATACGGTTC

(7) AAATCGCCTCAGTGCTTATACGGTTC

(8) AAATCGCCTCAGTGCTAATACGGTTC

sR8 Targets

Wild type, mutant, chimeric, and hairpin target RNAs were purchased from Integrated DNA Technologies. Wild type and methylated control RNA targets are shown below. Additional RNA oligonucleotides used as mutant, chimeric, and hairpin target RNAs are described in the text and figures.

D target RNA: AUGCUCAUACGGUC

Methylated D target RNA: AUGCUmCAUACGGUC

D' target RNA: GCUCAAA dG CCAAUCGC

Methylated D'target RNA: GCUCAAAmGCCAAUCGC

Cloning, Expression, and Purification of Proteins

Cloning and expression of recombinant, archaeal, box C/D sRNP core proteins, L7Ae, Nop56/58, and fibrillarin was accomplished as previously described (Tran et al 2003). His-tagged L7 and fibrillarin were purified by nickel-affinity chromatography using 'His-bind' Resin (Novagen) according to the manufacturer's protocols. Untagged

Nop56/58 was purified by ion exchange chromatography as previously detailed (Tran et al 2003).

RNP Assembly and In Vitro Methylation

sRNP complexes were assembled by incubating 40 pmol of guide sRNA with 20 pmol of L7, 32 pmol of Nop56/58, and 32 pmol of fibrillarin in assembly buffer containing 20 mM HEPES (pH 7.0), 150 mM NaCl, 0.75 mM dithiothreitol, 1.5 mM MgCl2, 0.1 mM EDTA, and 10% glycerol at 70°C for 10 minutes. Following sRNP assembly, 360 pmol of target RNA(s) was added along with 360 pmol SAM (*S*-adenosyl-l-methionine dihydrogen sulfate; Calbiochem) and 1.7 μCi of [^{3}H] SAM (63 Ci/mmol; Amersham) for a final reaction volume of 55 μl. This reaction was incubated at 68°C and 20 μl aliquots were blotted on 3MM Whatman filters at 0 and 60 minutes. RNA on dried filters was precipitated by soaking the filters in 10% TCA for 15 minutes at 4°C. Filters were then washed three times in 5% TCA for 15 minutes at room temperature before air drying. ^{3}H-SAM incorporation into RNA was determined by scintillation counting. Assays were performed in triplicate, standard errors calculated, and results reported as percent activity of wild type RNA.

RESULTS

M. jannaschii sR8 Box C/D sRNA and In Vitro Assembly of the sRNP Complex

The folded secondary structure of *Methanocaldococcus jannaschii* sR8, a model archaeal box C/D sRNA, is illustrated in Figure 1. This double-guide sRNA contains both box C/D and C'/D' motifs possessing identical box nucleotide sequences which fold into K-turn elements.[10,23] It is these K-turns that serve as binding sites for the three box C/D sRNP core proteins.[16] A single copy of each protein, L7, Nop56/58, and fibrillarin, bind each motif to assemble identical or "symmetric" RNP complexes.[23] Guide regions of 12 nucleotides separate the two RNP complexes and base pair to target RNAs possessing the designate nucleotide to be modified. Using the "n+5 rule", the fifth nucleotide from the D or D' box within the guide sequence:target RNA duplex becomes methylated at the 2' position of the ribose sugar (Figure 1).[23]

Work in our lab has established an *in vitro* archaeal box C/D sRNP assembly system using *in vitro* transcribed sR8 sRNA and recombinant L7, Nop56/58, and fibrillarin core proteins.[23] This reconstituted sRNP is catalytically active and 2'-O-methylates target nucleotides in synthetic target RNAs which are complementary to sR8 D or D' guide regions. S-adenosyl methionine (SAM) is the methyl donor and $^{3}H\text{-}CH_3$ incorporation into the target RNA measures sRNP methyltransferase activity. This *in vitro* assembled sRNP provides a convenient system to study the structure and function of box C/D RNP complexes and the mechanisms of box C/D RNA-guided nucleotide 2'-O-methylation. In the current study, we have taken advantage of this system to examine the

structural features of target RNA interaction with the sRNA guide sequences that are important for efficient, site-specific, nucleotide 2'-O-methylation.

Watson-Crick Pairing of Guide and Target Nucleotides is Essential for Nucleotide Methylation.

Previous *in vivo* and *in vitro* systems have assessed the importance of Watson-Crick pairing between guide and target nucleotides for 2'-O-methylation. Results have indicated the importance of Watson-Crick hydrogen bonding although some data has suggested the methylation of non-Watson-Crick pairs depending upon the specific experimental system used for analysis.[28,29] In light of these conflicting results, we began our analysis of sRNA guide sequence:target RNA interaction defining the hydrogen-bonding features of the guide and target nucleotides important for target nucleotide methylation using our *in vitro* assembled box C/D sRNP.

Synthetic RNA oligonucleotides of 14 and 16 nucleotides in length and complimentary to the sR8 D and D' guide regions, respectively, served as target RNAs. Incorporation into these target RNAs of $^{3}H\text{-}CH_3$ donated from S-adenosyl-methionine measured the methytransferasae activity guided by both box C/D and C'/D' RNPs. Target RNA oligonucleotides possessing 2'-O-methylated target nucleotides served as control target RNAs. To determine the importance of Watson-Crick pairing between the guide and target nucleotides for nucleotide methylation, mutant D and D' target RNAs were synthesized which altered guide:target nucleotide base pairing. For some non-

Watson-Crick pairs between guide and target nucleotides, point mutations were also made in the sR8 D and D' guide sequences. Each sR8 sRNP complex tested contained only one altered guide:target nucleotide pair. Therefore, the unaltered, Watson-Crick guide:target base pair of the partner RNP complex was used as an internal methylation control.

Table 1 summarizes the methylation capabilities of the various guide:target nucleotide base pairs for both box C/D and C'/D' guide:target RNA duplexes. Watson-Crick G:C base pairs for both box C/D and C'/D" guide:target nucleotide pairs exhibited robust target nucleotide methylation. This was also the case for Watson-Crick U:A base pairs although the level of A residue methylation was significantly reduced for the D target U:A pair. This lower level of methylation may be a result of overall guide:target RNA duplex structure (see Discussion). In contrast, all non-Watson-Crick target pairs or mismatches were severely deficient in target nucleotide methylation. This was true for sRNA-guided methylation for both box C/D and C'/D' RNPs. However, compensatory mutations of the D and D' sRNA guide sequences which restored Watson-Crick pairing also restored target nucleotide methylation (data not shown). Particularly interesting was the lack of target nucleotide methylation for the non-Watson-Crick U:G base pair. Despite the ability of this non-Watson-Crick pair to hydrogen bond, this stable pairing was not sufficient to promote target nucleotide methylation. This contrasts the low level of methylation observed for the G:U nucleotide pair *in vivo* (see Discussion).[29] Collectively, these results clearly demonstrated that for this *in vitro* sRNP assembly and

methylation system, Watson-Crick hydrogen bonding between guide and target nucleotide was essential for site-specific, target nucleotide 2'-O-methylation.

Target Nucleotide Methylation Requires a Continuous, Watson-Crick Paired, Guide:Target RNA Duplex

Having demonstrated the importance of Watson-Crick pairing between guide and target nucleotides, we next examined the importance of the base pairing of nucleotides within the guide:target RNA duplex surrounding the guide:target nucleotide pair. Again, target RNAs were mutated to disrupt Watson-Crick pairing of selected RNA duplex nucleotide pairs and the effect of these mismatches upon target nucleotide methylation determined (Figure 2A). Not surprisingly, three continuous mismatches either immediately upstream or downstream of the target site abolished methytransferase activity (Figure 2B). It is likely that mismatches in the guide:target RNA duplex immediately adjacent to the site of methylation also disrupt guide:target nucleotide interactions. Interestingly, single nucleotide mismatches removed from the guide:target nucleotide pair also disrupted target nucleotide methylation. This indicated that even a slight disruption of continuous Watson-Crick base pairing around the target site is detrimental for nucleotide methylation with this *in vitro* system. The single exception for methylation disruption was the single nucleotide mismatch immediately adjacent to box D. Similar, single mismatches at this position have been reported for some wild type box C/D RNAs.[8] This suggested that this position is less critical for guide:target RNA

interaction and may be more important for the spatial positioning of the target nucleotide with respect to the box D sequence and assembled sRNP (see Discussion).

Target Nucleotide Methylation Requires an RNA:RNA Duplex Formed Between the sRNA Guide Sequence and the Target RNA

Utilization of an *in vitro* assembled box C/D sRNP methylation system allowed us to assess the importance of forming an RNA:RNA duplex between the guide sequence and target RNA. Target DNA oligonucleotides complementary to D and D' guide sequences were synthesized and then incubated with *in vitro* assembled sR8 sRNP complexes in the presence of S-adenosyl-methionine. For each of these target DNA oligonucleotides, the 5th or target nucleotide was a ribonucleotide, thus providing a 2'-OH for methyl addition. Target nucleotide methylation activity was tested for both the box C/D and C'/D' RNPs using a target DNA oligonucleotide and chimeric DNA-RNA oligonucleotides that gradually restored the RNA/RNA duplex, in a symmetric (Figure 3A) or asymmetric manner (Figure 3B) progressing outward from the target nucleotide.

Replacement of target RNA oligonucleotides with a DNA oligonucleotide completely disrupted target nucleotide methylation for both box C/D and C'/D' RNPs. Thus, formation of an RNA:RNA duplex between guide and target RNA sequences is essential for target nucleotide methylation. Progressive, symmetric, replacement of deoxynucleotides with ribonucleotides gradually restored methylation activity for both RNP complexes, although the methylation activity of the two RNPs recovered at different

rates (Figure 3A). Only slight methyltransferase activity was observed for the box C/D RNP when nine ribonucleotides replaced the substituted deoxynucleotides and full recovery of methyltransferase activity required a complete ribonucleotide target RNA. In contrast, significant D' target nucleotide methylation was observed when seven nucleotides were substituted for the corresponding deoxynucleotides (~20% activity) and full activity was restored with the replacement of nine ribonucleotides. The differential rate of recovery of target nucleotide methylation between the box C/D and C'/D' guide:target RNA duplexes suggested a significant difference in RNP structure for the two complexes (see Discussion). Despite this difference, these results clearly indicated that a significant length of RNA/RNA duplex surrounding the target nucleotide was required for efficient methylation.

The importance of a guide:target RNA duplex upstream between the target nucleotide and box D/D' as well as downstream of the target nucleotide was examined next. Deoxynucleotides were asymmetrically replaced with ribonucleotides and these chimeric target oligonucleotides tested for methylation capabilities. Complete restoration of the upstream guide RNA:target RNA duplex between the target nucleotide and the D or D' box did not result in target nucleotide methylation (Figure 3B). Strikingly however, methylation of the target nucleotide was partially restored as the downstream region of the guide:target duplex was restored to an RNA/RNA helix. Approximately 20% of wild type methylation levels were observed for the D target nucleotide when the entire downstream guide:target duplex was an RNA:RNA helix. The D' target nucleotide was methylated to approximately 70% of wild type methylation levels when the

downstream guide:target duplex was entirely ribonucleotide base pairs. These results again demonstrated the difference in box C/D and C'/D' RNP methylation capabilities likely reflecting differences in RNP structure. In addition, these results strongly suggested that the upstream and downstream guide:target RNA duplexes play different roles in sRNA-guided nucleotide methylation (see Discussion).

Elevated Mg^{2+} Concentrations Promote 2'-0-Methylation between Mismatched Guide and Target Nucleotides.

Previous studies have shown that increasing the strength of guide sequence and target RNA by increasing the RNA duplex GC content can stimulate methylation, even when this duplex possesses base mismatches.[29] We therefore further investigated this phenomenon by increasing the salt concentration of the methylation reaction to stabilize guide RNA:target RNA interactions. Previous work in our laboratory has demonstrated that the sR8 sRNP complex is stable in salt concentrations approaching 1 M (Tran and Maxwell, unpublished results). The typical NaCl concentration of 150 mM was incrementally increased up to 1000 mM and methylation activity of both box C/D (Figure 4 A) and C'/D' (Figure 4 B) RNPs assessed.

Preliminary experiments utilizing target RNAs with the 2'-OH removed from nucleotides 3, 4, 5 (target nucleotide), 6, or 7 confirmed the site-specific 2'-O-methylation of the fifth or target nucleotide (Figure 4 A and B, upper panels). Site-specific methylation of the target nucleotide is important for some of the results discussed

below. Wild type D and D' target RNAs possessing Watson-Crick guide nucleotide:target nucleotide base pairs as well as target RNAs with non-Watson-Crick guide:nucleotide pairs were assessed for nucleotide methylation under increasing concentrations of NaCl (Figure 4 A and B; middle panels). For wild type Watson-Crick base pairs, increasing NaCl concentrations had no significant effect upon the methylation levels of the D target nucleotide. Interestingly however, target nucleotide methylation levels increased almost two fold for the D' target nucleotide as NaCl concentrations increased to 1 M. This methylation for both D and D' target RNAs was site-specific and control target RNAs possessing a 2'-OH at the target nucleotide were not methylated (data not shown). Little effect was seen under increasing NaCl concentrations for non-Watson-Crick pairs although a low level of methylation was observed at low salt concentrations. Thus, higher salt concentrations that would increase guide:target RNA duplex stability did not affect the site-specificity of the methyltransferase reactions. However, the methyltransferase activities of the two RNP complexes were differentially affected (see Discussion).

Initial examination of *in vitro* box C/D RNP assembly as well as in *vitro* sRNP methyltransferase activities have demonstrated the essential nature of $MgCl_2$.[23] The divalent Mg^{2+} cation likely plays important roles in establishing sRNA structure.[30] We therefore tested the effect of increasing Mg^{2+} concentrations upon possible guide:target RNA interactions and altered methyltransferase activities. Again, both Watson-Crick and non-Watson-Crick guide:target nucleotide base pairs were assessed under increasing concentrations of $MgCl_2$ (Figure 4 A and B, bottom panels). Increasing $MgCl_2$

concentrations from the standard assay conditions of 1.5 mM to 50 mM had little effect upon methyltransferase activity for both the box C/D and C'/D' RNPs and Watson-Crick guide:target nucleotide base pairs. However, particularly striking was the increase in methylation of target nucleotides that were not hydrogen bonded to the guide nucleotide via Watson-Crick base pairing. The methylation of non-Watson-Crick paired target nucleotides was particularly significant for the C'/D' RNP and approached 40% that of wild type target RNA at 50 mM $MgCl_2$ concentrations. This was not a specific effect for a particular non-Watson-Crick pair but a general phenomenon for all non-Watson-Crick pairs tested. Importantly, this methylation was specific for the fifth target nucleotide as mismatched target nucleotides possessing a 2'-O-CH_3 exhibited no nucleotide methylation. Additional analysis of the simultaneous incorporation of ^{3}H-CH_3 into D and D' target RNAs via gel electrophoretic analysis has confirmed that the D' target RNAs possessing mismatched guide:target nucleotides are indeed being methylated (data not shown). Again, these analyses demonstrated a differential response of box C/D and C'/D' RNP-guided nucleotide modification under non-standard assay conditions and suggested that structural differences between the two different RNPs have a significant affect upon the catalytic differences of the two complexes (see Discussion).

Target Nucleotides Positioned Within Highly Folded Target RNAs Are Efficiently Methylated.

Nucleotides targeted by box C/D RNPs for site-specific nucleotide methylation are often positioned within highly folded or structured regions of the target RNA (i.e.

highly folded tRNA or rRNA). As the sequence surrounding the target nucleotide must be single stranded to base pair with the guide region, a mechanism for unfolding this structured region of the target RNA must be utilized. To assess the possible intrinsic ability of the archaeal core proteins to unfold a target RNA, two target RNAs were constructed that positioned the target nucleotide to be 2'-O-methylated in highly structured regions of the target RNA (Figure 5A). The first target RNA positioned the target nucleotide in the center of a double-stranded RNA hairpin whereas the second target RNA positioned the target nucleotide in a 5 nucleotide loop of a hairpin-loop structure. The Tm of both RNA targets was determined by UV melting analysis and shown to be 84°C and 74°C for the target-in-stem and target-in-loop RNAs, respectively. Both melting temperatures were above the standard methylation assay conditions of 70°C. The target-in-stem RNA was constructed to be complimentary to the wild type D guide region. The standard, unstructured D target RNA served as the wild type control target for methylation. The target-in-loop target RNA possessed a slightly altered target RNA sequence for base pairing to the sRNA guide sequence. For this target RNA, a new guide RNA, sR8 GRDtl, was designed with an altered D guide region that maintained complimentarity to the target-in-loop RNA. The target RNA control for this structure RNA was an essentially unstructured target RNA but still complimentary to sR8 GRDtl D guide region (see Materials and Methods).

At 70°C, both structured target constructs exhibited high levels of methylation comparable to their unstructured target RNA controls (Figure 5B). To address the possibility that the assay conditions were sufficiently close to the determined melting

temperatures of the target RNAs, such that transient melting of the folded target RNA could facilitate target nucleotide methylation, the temperature of the reaction was lowered to 50°C. This reduction in temperature reduced the overall levels of nucleotide methylation from optimal levels observed at 70°C. However, this incubation temperature was 25°C and 35°C degrees below the melting temperature of these two RNAs and should therefore greatly stabilize their target RNA secondary structure and reduce transient unfolding. At this lower incubation temperature, both target RNAs were strongly modified, although methylation levels were reduced to approximately half of their unstructured RNA controls (Figure 5B). These results suggested that the archaeal sRNP complex possesses some inherent ability to unwind or denature RNA hairpin structures to promote guide:target base pairing and facilitate nucleotide methylation (see Discussion).

DISCUSSION

The archaeal sR8 sRNP provides a model complex for the study of box C/D RNP-guided nucleotide modification. The ability to assemble *in vitro* a catalytically active RNP facilitates a molecular analysis of both the RNP structure and function. In addition, the *in vitro* system also permits a detailed examination of particular features of the complex not easily addressed *in vivo*. The work presented here has examined those structural features of the guide:target RNA duplex that are important for sRNA-guided nucleotide 2'-O-methylation. These results demonstrated the importance of a continuously Watson-Crick paired, guide:target RNA:RNA duplex for nucleotide methylation. Results also imply that unique structural/functional features of the box C/D and C'/D' RNPs affect the nucleotide modification capabilities of the individual complexes. Finally, the ability to methylate target nucleotides within highly folded target RNAs suggests an intrinsic ability of this RNP to unfold double stranded RNA.

Efficient 2'-O-methylation of target RNAs stringently requires a Watson-Crick base pair at the target nucleotide. The level of methylation for an A:U, guide:target pair although substantial, was consistently lower in comparison to a G:C base pair. This may reflect the lower hydrogen bonding strength of the A:U base pair itself or simply be a consequence of reduced G:C content of the guide:target RNA duplex. Previous work *in vivo* has suggested that increased G:C content of the guide:target RNA duplex can significantly affect the methylation levels of the target nucleotide.[29] Particularly notable was the fact that a hydrogen bonded, non-Watson-Crick, G:U base pair was not sufficient for target nucleotide methylation. Previous *in vivo* analysis of guide sequence-target

RNA interaction has indicated low levels of target nucleotide 2'-O-methylation for such a G:U pair (~1%) when positioned within an elongated, G:C-rich, guide:target RNA duplex.[29] These observations have suggested that increased RNA:RNA duplex stability could promote, albeit at low levels, methylation of a non-Watson-Crick paired target nucleotide. Interestingly, no G:U pair has been identified at a guide:target nucleotide position for any eukaryotic or archaeal box C/D RNA, although several G:U pairs have been found within the guide:target RNA duplex itself.[29]

This stringent requirement for Watson-Crick pairing between guide and target nucleotide exhibited by both *in vivo* eukaryotic and *in vitro* archaeal systems contrasts that recently observed for the *in vitro* assembled archaeal box C/D RNP of *H. volcanii* pre-tRNA^{Trp}.[28] This box C/D RNP is unusual in that the sRNA is encoded within the pre-tRNA^{Trp} intron and guides the nucleotide methylation of two nucleotides in the tRNA^{Trp} itself. Originally thought to occur in *cis*, these two methylation reactions have recently been shown *in vitro* to occur in *trans* and likely to occur *in vivo* in *trans* as well, guided by the excised and circular intron. *In vitro* assembly of the box C/D RNP on the pre-tRNA^{Trp} precursor intron or on the excised circular intron can guide the methylation of the two tRNA^{Trp} target nucleotides.[28] Interestingly, a number of non-paired guide:target nucleotides are substantially methylated by this *in vitro* assembled sRNP. A G:U pair was methylated to 50% that of the wild type Watson-Crick pair and other mismatched target nucleotides ranged from 13-40% of wild type levels.[28] This may be the result of suboptimal guide:target RNA duplex lengths (D guide of 11 nucleotides and D' guide of 8 nucleotides) and the low G:C content of the D' guide:target duplex (25%).

Alternatively, structural constraints imposed by the highly folded pre-tRNA^{Trp} precursor or the excised circular intron could cause this atypical methylation of mismatched target nucleotides.

The stringent requirement for a Watson-Crick pair at the target site can be compromised with increased concentrations of $MgCl_2$. Methylation of mismatched base pairs approached 40% of the wild type levels for some of the guide:target mismatches examined. This methylation indeed occurred at the mismatched target nucleotide and not at an adjacent, Watson-Crick paired nucleotide. The importance of $MgCl_2$ for stabilizing RNA structure, particularly secondary structure, is well known and the observed effect is likely a result of altered RNA:RNA duplex structure at the target nucleotide.[30] Increasing the NaCl concentrations did not have the same effect upon mismatched guide:target nucleotides, except perhaps with some small effect at low salt concentrations. Interestingly, an increase in overall methylation of Watson-Crick paired guide:target nucleotides was observed for the D' target RNA as NaCl concentration increased. Thus, the box C/D and C'/D' RNPs responded differently to altered $MgCl_2$ and NaCl concentrations suggesting distinct RNP structures and accompanying methylation activities for each complex (see below).

Mismatched pairs in the guide:target RNA duplex away from the guide:target nucleotide pair also affected nucleotide methylation. With the exception of the first guide:target nucleotide pair proximal to the D box, single mismatches at nucleotide positions 2 through 8 disrupted methylation of the Watson-Crick paired target nucleotide.

These mismatches apparently cause a disturbance in local helical structure at the target nucleotide sufficient to disrupt methylation. These results are similar to those originally reported by Cavaille and coworkers for eukaryotic box C/D snoRNA-guided methylation *in vivo*.[8,29] It is interesting that non-Watson-Crick pairs and mismatches have been described in native guide:target RNA duplexes.[29] However, almost all examples are from eukaryotic snoRNAs with the majority non-Watson-Crick pairs consisting of G:U base pairs, though never at the guide:target nucleotide pair. Only a few examples of truly mismatched, non-hydrogen bonded pairs have been reported, and again none are at the target nucleotide.[29]

Our minimal, *in vitro* assembled archaeal box C/D sRNP complex may be particularly sensitive to guide:target RNA duplex mismatches in the absence of additional cellular or auxiliary proteins that may be involved in the methylation reaction *in vivo*. Such proteins may play important roles for the eukaryotic snoRNP complex which is more diverse both in terms of structure and its function. It is interesting to speculate that increased flexibility in pairing between the guide and target RNA may allow recognition of more than one target RNA for methylation. A similar scenario is seen for the microRNAs where imperfect pairing of a miRNA with its target RNA allows recognition of more that one mRNA sequence for regulatory control of translation by a single miRNA.[31]

Replacement of single ribonucleotides with deoxynucleotides around the target nucleotide had no effect upon methylation. However, replacement of the target RNA

with a deoxy-oligonucleotide target possessing only a target ribonucleotide severely affected nucleotide methylation. These results clearly demonstrated the importance of an RNA:RNA duplex interacting with the protein methylation machinery for nucleotide modification. While double stranded DNA forms a B helix, a DNA oligonucleotide (target) paired with an RNA oligonucleotide (guide sequence) would establish an A helix typical of an RNA:RNA duplex. This A helix would possess a smaller major groove and incorporate more bases per helix turn. The chimeric DNA-RNA target oligonucleotides utilized in some of these studies should also form an A helix with the guide sequence. It is likely, however, that these small but significant differences in double stranded duplex structure affect how the guide:target duplex interacts with the Nop56/58-fibrillarin dimer or "protein platform" and the proper positioning of the target nucleotide at fibrillarin's catalytic site. Such perturbations, however small, are likely to have detrimental effects upon the methylation capabilities of the assembled RNP.

Interestingly, restoration of the upstream or downstream deoxynucleotides with ribonucleotides produced a differential effect upon the methylation reaction. Restoration of ribonucleotides upstream of the target nucleotide (nucleotides 1-4) did not result in any observable methylation whereas restoration of the downstream ribonucleotides did gradually restore activity. The differential rate of methylation recovery in response to upstream versus downstream ribonucleotide restoration suggests that correct duplex formation downstream of the target site (nucleotides 6-12) may well be more important for stabilizing the guide:target duplex and establishing important contacts with fibrillarin (substrate:enzyme interactions). It is likely that this downstream region of the

guide:target RNA duplex requires an RNA:RNA duplex presenting proper helix structure and conformation necessary for stable methylase interaction. The rigid requirement for nucleotide spacing between the box D or D' sequence and the target nucleotide suggests that the upstream region of the guide:target RNA duplex is more critical for spacing or "measuring" the target nucleotide with respect to the RNP complex assembled upon the box D/D' sequence, and less critical for stable binding of the target RNA to the guide sequence. The ability to disrupt Watson-Crick pairing at the nucleotide pair proximal to the D box without affecting methylation activity is consistent with this suggestion.

In vivo, nucleotides targeted for box C/D RNA-guided methylation are frequently positioned within highly structured regions of rRNA or tRNA. Therefore it is necessary to unfold target RNAs so that they may base pair with the guide sequences. The demonstration that target nucleotides positioned within highly folded target RNAs are still methylated by our *in vitro* assembled sRNP suggests an intrinsic ability of this complex to unfold double stranded RNAs. The box C/D core proteins themselves may well serve such an unfolding function. Neither L7, Nop56/58, nor fibrillarin proteins possess Walker A and Walker B motifs characteristic of protein helicases. However, a few proteins which lack these ATP-dependent helicase motifs but are nonetheless capable of melting double-stranded RNA structures have been described. One such protein is the eukaryotic polypyrimidine tract binding protein (PTB) which utilizes RNA recognition motifs (RRM) to bind an IRES and thereby mediate its correct folding.[32] The Y RNA-associated proteins La, hnRNP1 and hnRNP K are yet other examples of protein chaperones which, while lacking identifiable helicase motifs, nevertheless possess bona

fide chaperone activity in vitro.[33] Interestingly, both Nop56/58 and fibrillarin interact with guide and the target RNAs.[34] Fibrillarin exhibits non-sequence-specific RNA binding via electrostatic interactions between N-terminal, positively charged amino acid residues and the phosphate backbone.[35] Thus, the fibrillarin-Nop56/58 core protein may function to assist in restructuring target RNAs as they bind the box C/D RNA guide sequences. Consistent with this suggestion are recent computer simulations which suggest that sRNPs are indeed capable of such RNA remodeling activities.[36]

Finally, the differential methylation response of the box C/D and C/'D' RNPs to chimeric DNA-RNA targets as well as elevated concentrations of NaCl and $MgCl_2$ is striking. A different response for nucleotide methylation was previously observed when these different K-turn motifs were assembled as individual RNP complexes and functioned independently as sRNP halfmer complexes.[23] We suspect that the differential response reflects an underlying structural difference between the two RNP complexes. Both the box C/D and C'/D' motifs are assembled with the same three core proteins.[37] However, the terminal box C/D K-turn is distinctly different in folded structure from the internal C'/D' K-turn or "K-loop" motif. While the terminal K-turn possesses two RNA stems flanking the asymmetric bulge, the internal K-loop is missing stem I or the canonical stem.[16,18] While L7 initiates assembly of both RNP complexes by binding this K-turn/loop element[23], the established RNPs must be sufficiently different to result in unique methylation capabilities when presented with chimeric DNA-RNA targets or challenged with altered reaction conditions. Recent experiments have indeed demonstrated different L7 binding characteristics for the box C/D and C'/D' motifs

(Pourdehymi, Gagnon, and Maxwell, unpublished results). Ultimately, such questions concerning the unique character of the box C/D and C'/D' RNPs and their individual methylation characteristics await detailed characterization of this sRNP's fine structure most likely to be revealed by X-ray crystallography.

ACKNOWLEDGEMENTS

I would like to thank Keith Gagnon and Xin Xin Zhang for assistance with protein purification and providing sR8 templates. This work was funded by NSF Grant MCB 0543741 (ESM).

REFERENCES

1. Maxwell, E. S and M. J. Fournier (1995) *The small nucleolar RNAs*. Ann. Rev. Biochem. **35**:897-934.

2. Tycowski, K. T., M. D. Shu, J. A. Steitz (1994) *Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation*. Science **2**:1558-61.

3. Tycowski, K., et al. (1996) *A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus*. Proc. Natl. Acad. Sci. **93**:1480-1485.

4. Kiss-Laszlo, Z, et al. (1996) *Site-specific ribose methylation of pre-ribosomal RNA: a novel function for small nucleolar RNAs*. Cell **85**:1077-1088.

5. Bachellerie, J. P., J. Cavaille, and A. Huttenhofer (2002) *The expanding snoRNA world*. Biochemie **84**:775-790.

6. 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

7. Filipowicz, W., P. Pelczar, V. Pogacic, F. Dragon (1999) *Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification*. Acta Biochim Pol. **46**:377-389.

8. Cavaille J., M. Nicoloso, J.P. Bachellerie (1996) *Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides*. Nature **383**:732–735.

9. Balakin, A.G., L. Smith, and M.J. Fournier (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.

10. Gaspin, C., J. Cavaille, G. Erauso and J. P. Bachellerie (2000) *Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: Lessons from the Pyrococcus genomes*. J. Mol. Biol. **297**:895-906.

11. Omer, A. D., et al. (2000) *Homologs of small nucleolar RNAs in Archaea*. Science **288**:517-522.

12. Tran E, J. Brown, and E.S Maxwell (2004) *Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus?* Trends Biochem Sci. **29**:343-350.

13. Tollervey, D. (1998) *Small nucleolar RNAs guide ribosomal RNA methylation*. Science **273**:1056-1057.

14. Tycowski, K. et al. (1998) *Modification of U6 spliceosomal RNA is guided by other small RNAs*. Mol Cell. **2**:629-638.

15. Clouet d'Orval, B., et al. (2001) *Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp*. Nucl. Acids Res. **29:**4518–4529.

16. Klein D.J., et al. (2001) *The kink-turn: a new RNA secondary structure motif.* EMBO J. **20**:4214–4221.

17. Kiss-Laszlo, Z., Y. Henry, and T. Kiss (1998) *Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA*. EMBO J. **17**:797-807.

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

19. Tran, E.J., Zhang, X., Lackey, L., and E.S. Maxwell (2005) *Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-Omethylation of target RNAs*. RNA **11**:285-293.

20. Amiri, K. (1994) *Fibrillarin-like proteins occur in the domain Archaea*. J. Bacteriol. **176**:2124-2127.

21. Wu, P., et al. (1998) *Nop5pis a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast*. J. Biol. Chem. **273**:16453-16463.

22. Lafontaine, D. L., and D. Tollervey (1999) Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. RNA **3**:455-67.

23. Tran, E., X. Zhang, and E. S. Maxwell (2003) *Efficient 2'-0 methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' sRNP core proteins*. Nat. Struct. Biol. **22**:3930-3940.

24. 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.

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

26. Wang H., D. Boisvert, K.K. Kim, R. Kim, S.H. Kim (2000) *Crystal structure of a fibrillarin homologue from* Methanococcus jannaschii*, a hyperthermophile, at 1.6 A resolution*. EMBO. J. **19**:317–323.

27. Omer A.D., et al. (2002) *In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex*. Proc Natl Acad Sci. **99**:5289-94.

28. Singh S. K., et al. (2004) *Sequential 2'-O-methylation of archaeal pre-tRNA^{Trp} nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA*. J Biol Chem. **279**:47661-71.

29. Cavaille, J. and J. P. Bachellerie (1998) *SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction*. Nucleic Acids Res. **26**:1576-1587.

30. Misra, V. K, R. Shiman, D. E. Draper (2003) *A thermodynamic framework for the magnesium-dependent folding of RNA*. Biopolymers **69**:118-136.

31. Lim, L.P., et al. (2005) *Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs*. Nature **433**:769-773

32. Song, Y., et al. (2005) *Evidence for an RNA chaperone function of polypyrimidine tract-binding protein in picornavirus translation*. RNA **11**:1809–1824.

33. Basilov, A., et al. (2005) *RNA chaperone activity of protein components of human Ro RNPs RNA* **11**:1084–1094.

34. Deng, L., et al. (2004) *Structure determination of fibrillarin from the hyperthermophilic archaeaon* Pyrococcus furiosus. Biochem Biophys Res Commun. **315**:726-732.

35. Aittaleb, M., et al. (2004) *Structural and thermodynamic evidence for a stabilizing role of Nop5p in S-adenosyl-L-methionine binding to fibrillarin*. J Biol Chem. **279**:41822-41829.

36. Schoemaker, R. and A. P. Gultyaev (2006) *Computer simulation of chaperone effects of Archaeal C/D box sRNA binding on rRNA folding*. Nucleic Acids Res. **34**:2015–2026.

37. Rashid, R., et al. (2003) *Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles*. J. Mol. Biol. **33**:295-306.

BOX C/D-GUIDED METHYLATION		
sRNA/Mutation	***D (mutants)*** *guide: target (nts)*	***D' (control)***[b] *guide : target (nts)*
sR8/(wt)	**G:C**	**C:G**
^{3}H incorporated[a]	14.6 ± 2.9 pmol	16.6 ± 2.1pmol
(% of control)	*(88%)*	
sR8/G51U[c]	**U:A**	**C:G**
	3.0 ± 0.4	13.5 ± 1.8
	(22.2%)	
sR8/G51A	**A:C**	**C:G**
	0.23 ± 0.4	16.1 ± 4.2
	(1.4%)	
sR8/G51U	**U:C**	**C:G**
	0.04 ± 0.07	16.4 ± 3.9
	(0.2%)	
sR8/G51C	**C:C**	**C:G**
	0.02 ± 0.03	15.9 ± 1.9
	(0.1%)	

BOX C'/D'-GUIDED METHYLATION		
sRNA/Mutation	***D (control)***[b] *guide: target (nts)*	***D' (mutants)*** *guide : target (nts)*
sR8/(wt)	**G:C**	**C:G**
^{3}H incorporated[a]	14.6 ± 2.9 pmol	16.6 ± 2.1 pmol
(% of control)		*(113%)*
sR8/C24U[c]	**G:C**	**U:A**
	10.6 ± 0.3	11.1 ± 2.2
		(104.7%)
sR8/C24A	**G:C**	**A:G**
	15.1 ± 4.4	0.04 ± 0.04
		(0.3%)
sR8/C24U	**G:C**	**U:G**
	9.6 ± 3.1	0.22 ± 0.01
		(2.3%)
sR8/C24G	**G:C**	**G:G**
	12.9 ± 2.2	0.03 ± 0.07
		(0.2%)

TABLE 1. Archaeal Box C/D sRNA-Guided Nucleotide 2'-O-Methylation Requires Watson-Crick Base Paired Guide and Target Nucleotides For Both D and D' Target RNAs

[a]sRNA-guided nucleotide methylation during one hour of incubation is reported in pmols of incorporated of ^{3}H-CH_3 into D and D' target RNAs of each sRNA tested.

[b]Both mutated and non-mutated guide:target nucleotide pairs for each sRNA were assessed. The non-mutated guide:target nucleotide pair served as the control against which the percent (%) methylation of the altered nucleotide pair was compared and calculated.

[c]Alternative base paired guide and target nucleotides were accomplished by mutating the nucleotide of the target RNA, except for examined Watson-Crick A:U pairs where both guide and target nucleotides were mutated.

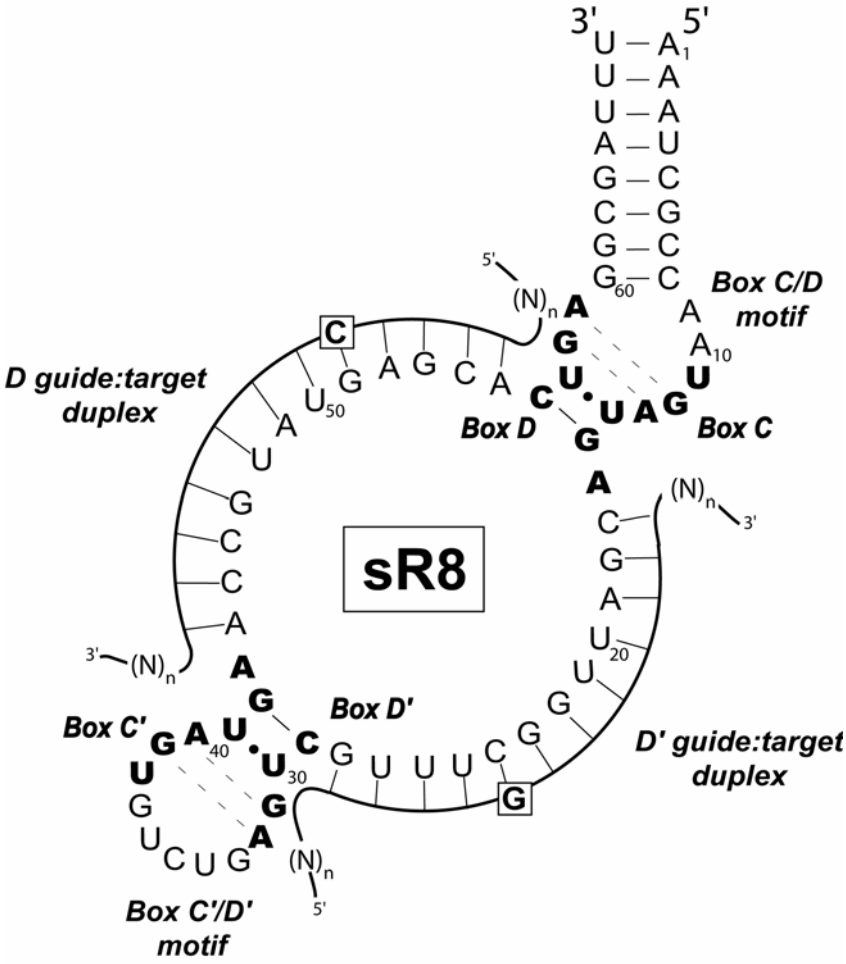

FIGURE 1. Secondary Structure of the *M. jannaschii* sR8 Box C/D sRNA Base Paired With D and D' Target RNAs

M. jannaschii sR8 box C/D sRNA folded into its secondary structure to form the box C/D and C'/D' motifs. Boxes C, D, C', and D' are indicated in bold. The box C/D and C'/D' motifs form K-turn structures critical for the binding of archaeal box C/D sRNP core proteins. D and D' sRNA guide regions are base-paired with their respective target RNA sequences. The boxed nucleotide is that nucleotide which will be 2'-O-methylated during the sRNP-guided nucleotide modification reaction.

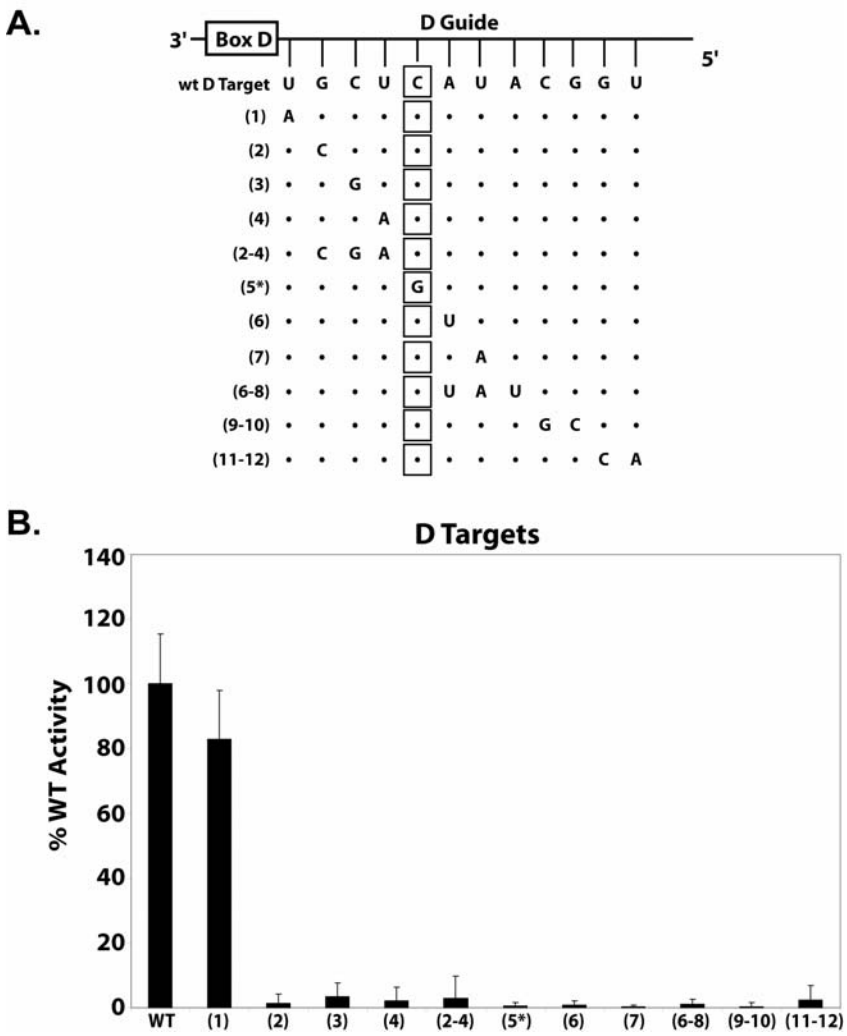

FIGURE 2. Target Nucleotide Methylation Requires A Continuous, Watson-Crick Paired, Guide:Target RNA Duplex

D target RNAs possessing mismatched nucleotides were incubated with *in vitro* assembled sR8 box C/D sRNP and 2'-O-methylation of the target nucleotide assessed. **A)** Schematic presentation of sR8 D guide region with associated box D base paired to D target RNA. Mutated D target RNAs are listed below with the distance of mismatched nucleotides from the D box indicated in parentheses. Target nucleotides are enclosed in squares with dots indicating conserved wild type D target ribonucleotides. Mismatched bases are indicated as G, C, A, U nucleotides for the individual mutated D target RNAs. **B)** Incorporation of $^{3}H\text{-}CH_3$ into D target RNAs. Levels of D target nucleotide 2'-O-methylation are indicated as percent of wild type D target RNA. Individual target RNAs are designated in parentheses.

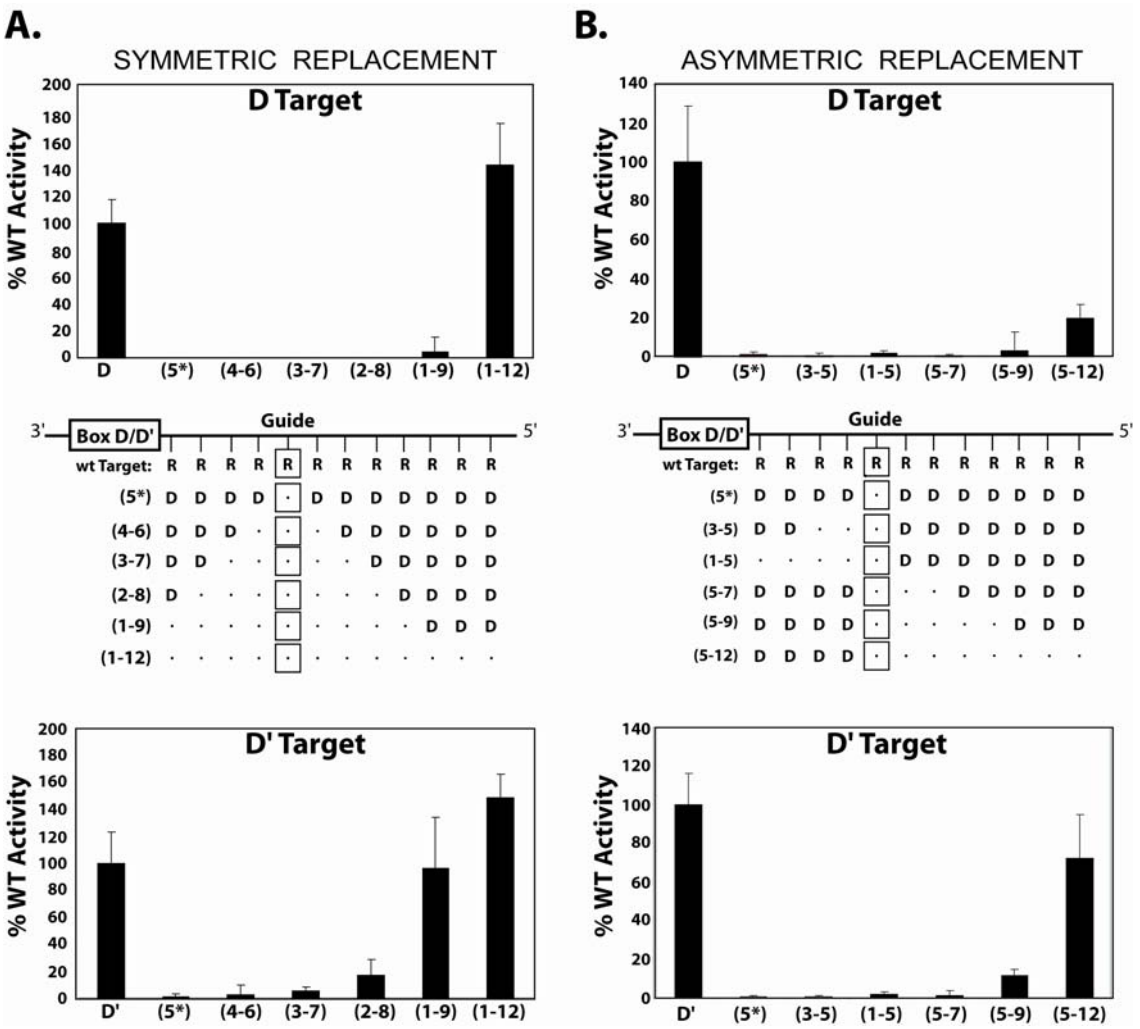

FIGURE 3. Target Nucleotide Methylation Requires An RNA:RNA Duplex Formed Between the Guide and Target RNA Sequences

Target D and D' deoxy-oligonucleotides were incubated with *in vitro* assembled sR8 box C/D sRNP and 2'-O-methylation of target nucleotides assessed. Schematic presentation of sR8 D/D' guide region with associated box D/D' illustrates base pairing of the guide sequence with the target deoxy-oligonucleotides. All target deoxy-oligonucleotides possessed a ribonucleotide with 2'-OH at the 5th or target nucleotide which is enclosed in squares. Deoxynucleotides are indicated by D and wild type ribonucleotides replacing deoxynucleotides indicated with dots. Deoxynucleotides replaced with ribonucleotides are indicated in parentheses. Incorporation of $^{3}H\text{-}CH_3$ into D and D' target deoxy-oligonucleotides is indicated as percent of wild type D target RNA incorporation. **A)** D and D' target deoxy-oligonucleotides symmetrically replaced outward from the target nucleotide with increasing numbers of ribonucleotides. **B)** D and D' target deoxy-oligonucleotides asymmetrically replaced upstream and downstream of the target nucleotide with increasing numbers of ribonucleotides.

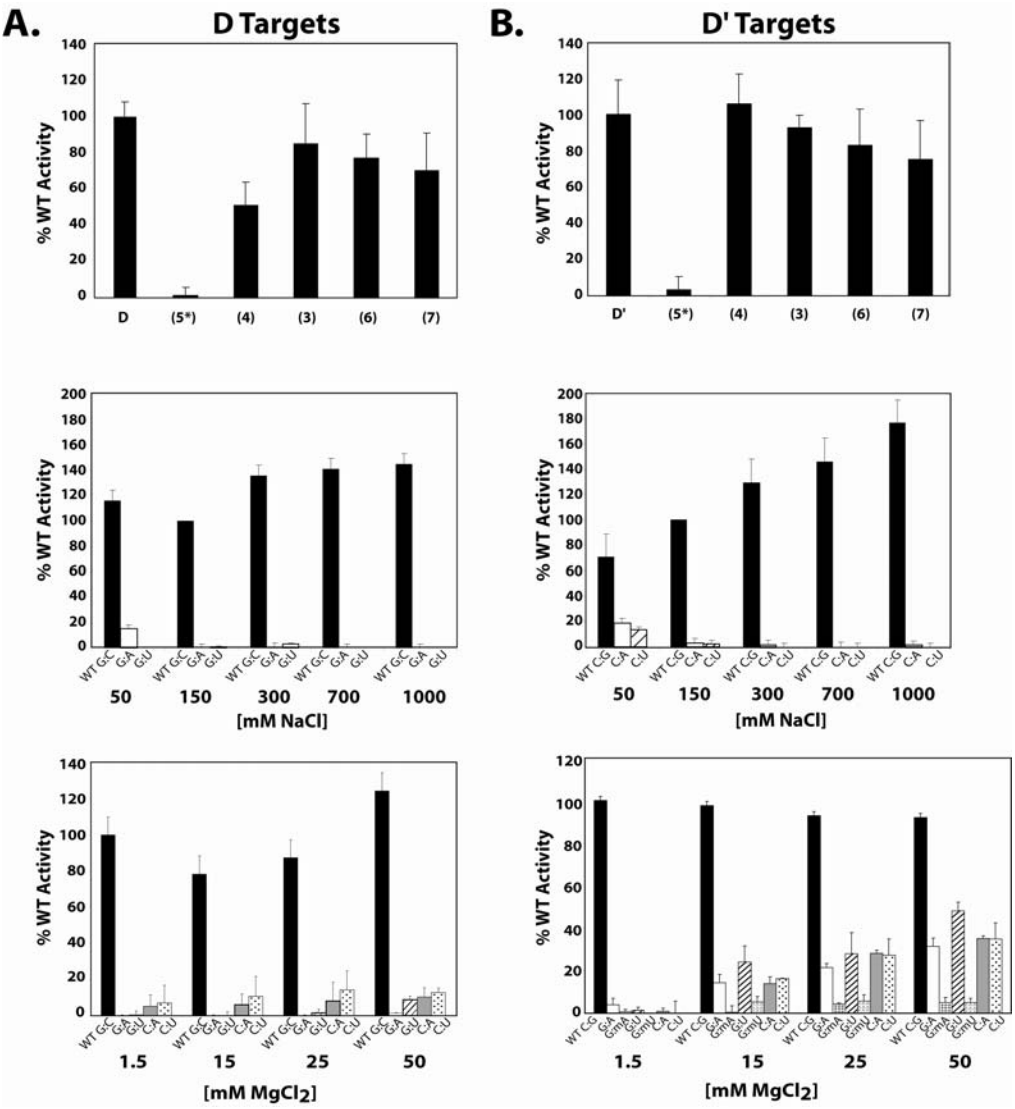

FIGURE 4. Elevated Mg^{+2} Concentrations Promote 2'-O-Methylation of Mismatched Guide and Target Nucleotides

D and D' target RNAs possessing a deoxynucleotide substituted at positions 3, 4, 5, 6, or 7 were incubated with *in vitro* assembled sR8 sRNP and the specificity of nucleotide methylation assessed (top panels). The position of substituted ribonucleotide with deoxynucleotide is indicated in parentheses. D and D' target RNAs with wild type or mismatched guide:target nucleotides were incubated with *in vitro* assembled sR8 sRNP in elevated concentrations of NaCl (middle panels) or $MgCl_2$ (bottom panels) and the 2'-O-methylation of both Watson-Crick and mismatched guide:target nucleotides assessed. Specific mismatched guide:target ribonucleotides are indicated below each panel along with increasing NaCl and $MgCl_2$ concentrations. **A)** D target nucleotide specificity and 2'-O-methylation of Watson-Crick and mismatched guide:target nucleotides for box C/D RNP-guided nucleotide 2'-O-methylation. **B)** D' target nucleotide specificity and 2'-O-methylation of Watson-Crick and mismatched guide:target nucleotides for box C'/D' RNP-guided nucleotide 2'-O-methylation.

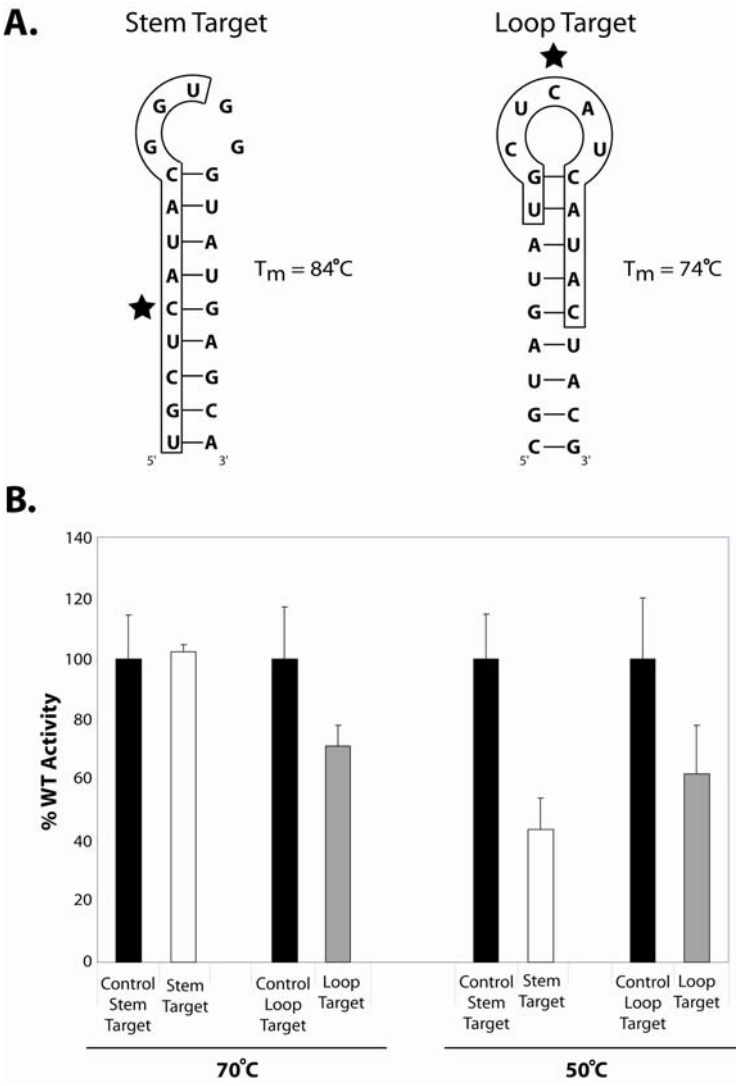

FIGURE 5. Target Nucleotides Positioned Within Highly Folded Target RNAs Are Efficiently Methylated

D target RNAs possessing target nucleotides within highly folded RNA secondary structures were incubated with *in vitro* assembled sR8 box C/D sRNP and 2'-O-methylation of the target nucleotides assessed. **A)** D target RNAs folded into a hairpin-loop structure with the target nucleotide positioned within the RNA hairpin (stem target) or RNA loop region (loop target). Target RNA sequences base pairing with sR8 D guide region are enclosed in boxes and those nucleotides targeted for methylation are indicated by stars. Target RNA melting temperatures (Tm) indicated at the side were determined by UV thermal denaturation analysis. **B)** Incorporation of $^{3}H\text{-}CH_3$ into D target RNAs. Levels of D target nucleotide 2'-O-methylation are indicated as percent of non-folded control D target RNAs. Incubation temperatures of 70°C and 50°C are indicated.

APPENDICES

APPENDIX A

Structure and Function of the Box C/D RNA-Guided Nucleotide Methylation Complexes

2006 FASEB Research Conference on Biological Methylation

June 24-29, 2006

Vermont Academy

Saxtons River, VT

STRUCTURE AND FUNCTION OF THE BOX C/D RNA-GUIDED NUCLEOTIDE METHYLATION COMPLEXES

Denise Appel, Keith Gagnon, Xinxin Zhang, Elizabeth Tran, and Stuart Maxwell.
Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-7622

Box C/D RNP nucleotide methylation complexes are abundant in both archaeal and eukaryotic organisms, functioning to direct the 2'-O-methylation of numerous nucleotides found in ribosomal RNAs, tRNAs, snRNAs, and even some mRNAs. Guide sequences within the box C/D RNA base pair to complementary sequences in the target RNA while associated core proteins carry out ribose methylation at the designated nucleotide. Box C/D RNP complexes typically possess terminal box C/D and internal C'/D' RNPs. Each of these RNPs is comprised of a folded RNA K-turn motif bound with RNP core proteins and each guides nucleotide methylation. We have been investigating both archaeal and eukaryotic box C/D RNP complexes with the goal of elucidating the structure-function relationships of this RNA:protein enzyme. Towards this objective, we have established an *in vitro* assembly system that reconstitutes an enzymatically-active archaeal box C/D sRNP. sRNP assembly proceeds with the ordered addition of three core proteins that bind to both box C/D and C'/D' K-turn RNA motifs. Core protein binding is dependent upon sRNA remodeling during assembly. sRNP assembly may proceed via alternative sRNA folding and sRNP assembly pathways, but fully assembled complexes present exposed sRNA guide regions for target RNA base pairing and the subsequent nucleotide methylation reaction. Emerging crystal structures of RNA K-turn motifs, sRNP core proteins, and sRNA:core protein complexes are beginning to contribute to our understanding of box C/D RNP structure. Assessment of box C/D and C'/D' RNP methylation activities has demonstrated that efficient methylation requires that both complexes be juxtaposed in a single sRNP complex. Consistent with crosstalk interactions between juxtaposed complexes, inter-RNP spacing is highly conserved and critical for box C/D- and C'/D'-guided nucleotide methylation. Examination of sRNA guide sequence-target RNA duplex formation has demonstrated that Watson-Crick pairing between the sRNA guide nucleotide and the target nucleotide to be methylated is essential. Furthermore, establishment and maintenance of an A helix duplex formed between base paired guide and target RNA sequences is also important. Finally, comparison of archaeal sRNPs and eukaryotic box C/D snoRNPs reveal distinct structural features for each complex. These structural differences reflect the evolved RNA-binding capabilities of the core proteins. However, overall folded box C/D RNA structure and RNP organization of the archaeal and eukaryotic complexes are highly conserved and each complex accomplishes nucleotide methylation utilizing an identical mechanism. Thus, shared structural and functional features strongly argue for a common and ancient evolutionary origin for RNA-guided nucleotide modification complexes, pre-dating the divergence of Archaea and Eukarya some two billion years ago.

APPENDIX B

Primers, Template Oligonucleotides, Target Oligonucleotides

sR8 DNA Template

AAAUCGCCAAUGAUGACGAUUGGCUUUGCUGAGUCUGUGAUGAACCGUAU
GAGCACUGAGGCGAUUU

sR8 Guide Mutant Primers

1. T7-sR8 C24A Up:
CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTGGATTTG

2. T7-sR8 C24U Up:
CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTGGTTTTG

3. T7-sR8 C24G Up:
CTAATACGACTCACTATAGGCCAAATCGCCAATGATGACGATTGGGTTTG

4. sR8 G51A Dn: AAATCGCCTCAGTGCTTATACGGTTC

5. sR8 G51U Dn: AAATCGCCTCAGTGCTAATACGGTTC

6. sR8 G51C Dn: AAATCGCCTCAGTGCTGATACGGTTC

sR8 Mutant Targets

D Targets

1. D Tgt C5A: rArUrGrCrUrArArUrArCrGrGrUrC

2. D Tgt C5U: rArUrGrCrUrUrArUrArCrGrGrUrC

3. D Tgt C5G: rArUrGrCrUrGrArUrArCrGrGrUrC

D' Targets

1. D' Tgt G8A: rGrCrUrCrArArArArCrCrArArUrCrGrC

2. D' Tgt G8U: rGrCrUrCrArArArUrCrCrArArUrCrGrC

3. D' Tgt G8C: rGrCrUrCrArArArCrCrCrArArUrCrGrC

4. D' Tgt mG8A: rGrCrUrCrArArAmArCrCrArArUrCrGrC

5. D' Tgt mG8U: rGrCrUrCrArArAmUrCrCrArArUrCrGrC

sR8 DNA Targets

1. D sR8gDNA: ATGCTCATACGGTC
2. D' sR8gDNA: GCTCAAAGCCAATCGC

sR8 D/R Targets

DNA(RNA nt)

1. D sR8g D/R: ATGCTrCATACGGTC
2. D' sR8g D/R: GCTCAAArGCCAATCGC

sR8 R/D Targets

RNA (DNA nt)

1. D sR8g R/D: AUGCUdCAUACGGUC
2. D' sR8g R/D: GCUCAAAdGCCAAUCGC

DNA/RNA Chimeras – symmetric with respect to target nt

1. D sR8 D/R3: ATGCrUrCrATACGGTC

2. D sR8 D/R5: ATGrCrUrCrArUACGGTC

3. D sR8 D/R7: ATrGrCrUrCrArUrACGGTC

4. D(21) sR8 D/R7:
CTGATrGrCrUrCrArUrACGGTCTGCT

5. D sR8 D/R9: ArUrG rCrTrCrArUrArCGGTC

6. D sR8 D/R12: ArUrGrCrUrCrArUrArCrGrGrUC

7. D(21) sR8 D/R12:
CTGA rUrGrCrUrCrArUrArCrGrGrUCTGCT

8. D' sR8 D/R3: GCTCAArArGrCCAATCGC

9. D' sR8 D/R5: GCTCArArArGrCrCAATCGC

10. D' sR8 D/R7: GCTCrArArArGrCrCrAATCG C

11. D' sR8 D/R9: GCTrCrArArArGrCrCrArATCGC

10. D' sR8 D/R12: GCTrCrArArArGrCrCrArArUrCrGC

DNA/RNA Chimeras – asymmetric with respect to target nt
("-" upstream; "+" downstream of target nt)

1. D sR8 D/R-2: ATGrCrUrCATACGGTC

2. D sR8 D/R-4: ArUrGrCrUrCATACGGTC

3. D sR8 D/R+2: ATGCTrCrArUACGGTC

4. D sR8 D/R+4: ATGCTrCrArUrArCGGTC

5. D sR8 D/R +7: ATGCTrCrArUrArCrGrGrUC

6. D' sR8 D/R-2: GCTCArArArGCCAATCGC

7. D' sR8 D/R-4: GCTrCrArArArGCCAATCGC

8. D' sR8 D/R+2: GCTCAAArGrCrCAATCGC

9. D' sR8 D/R+4: GCTCAAArGrCrCrArATCGC

10. D' sR8 D/R+7: GCTCAAArGrCrCrArArUrCrGC

Single DNA nt replacements
("-" upstream; "+" downstream of target nt)

1. D sR8 R/D-1: rArUrGrCdTrCrArUrArCrGrGrUrC

3. D sR8 R/D-2: rArUrGdCrUrCrArUrArCrGrGrUrC

2. D sR8 R/D+1: rArUrGrCrUrCdArUrArCrGrGrUrC

4. D sR8 R/D+2: rArUrGrCrTrCrAdTrArCrGrGrUrC

5. D' sR8 R/D-1: rGrCrUrCrArAdArGrCrCrArArUrCrGrC

7. D' sR8 R/D-2: rGrCrUrCrAdArArGrCrCrArArUrCrGrC

6. D' sR8 R/D+1: rGrCrUrCrArArArGdCrCrArArUrCrGrC

8. D' sR8 R/D+2: rGrCrUrCrArArArGrCdCrArArUrCrGrC

Chimeric, Asymmetric Length Mutants – Length of RNA Duplex (not total target)

1. D sR8 D2/R10: ATGrCrUrCrArUrArCrGrGrUC

2. D sR8 D1/R11: ATrGrCrUrCrArUrArCrGrGrUC

3. D sR8 R7/D5: ArUrGrCrUrCrArUACGGTC

4. D sR8 R9/D3: ArUrGrCrUrCrArUrArCGGTC

5. D sR8 R11/D1: ArUrGrCrUrCrArUrArCrGrGTC

6. D' sR8 D3/R9: GCTCAArArGrCrCrArArUrCrGC

7. D' sR8 D2/R10: GCTCArArArGrCrCrArArUrCrGC

8. D' sR8 D1/R11: GCTCrArArArGrCrCrArArUrCrGC

9. D' sR8 R7/D5: GCTrCrArArArGrCrCAATCGC

10. D' sR8 R9/D3: GCTrCrArArArGrCrCrArATCGC

11. D' sR8 R11/D1: GCTrCrArArArGrCrCrArArUrCGC

Hairpin/Loop Targets – D Only

1. sR8 Dtl: rCrGrUrArGrUrArUrGrCrUrCrArUrCrArUrArCrUrArCrG

2. CTLsR8 Dtl: rCrCrUrUrCrArUrUrGrCrUrCrArUrCrArUrArCrUrArCrG

3. sR8Dts: rUrGrCrUrCrArUrArCrGrGrUrGrGrGrUrArUrGrArGrCrA

Guide sR8 Dtl (DNA template)

AAATCGCCAATGATGACGATTGGCTTTGCTGAGTCTGTGATGAGTATGATGA
GCACTGAGGCGATTT

Helix Mismatches – D Target Only

1. sR8 U1A: rArArGrCrUrCrArUrArCrGrGrUrC

2. sR8 G2C: rArUrCrCrUrCrArUrArCrGrGrUrC

3. sR8 C3G: rArUrGrGrUrCrArUrArCrGrGrUrC

4. sR8 U4A: rArUrGrCrArCrArUrArCrGrGrUrC

5. sR8 2-4CGA: rArUrCrGrArCrArUrArCrGrGrUrC

6. sR8 A6U: rArUrGrCrUrCrUrUrArCrGrGrUrC

7. sR8 U7A: rArUrGrCrUrCrArArArCrGrGrUrC

8. sR8 6-8UAU: rArUrGrCrUrCrUrArUrCrGrGrUrC

9. sR8 9-10GC: rArUrGrCrUrCrArUrArGrCrGrUrC

10. sR8 11-12CA: rArUrGrCrUrCrArUrArCrGrCrArC

Helix Base Pair Segments – D' Target Only

1. sR8 D'bp 1-7: rGrCrUrCrArArArGrCrCrUrCrArUrUrC

2. sR8 D'bp 3-9: rGrCrUrGrUrArArGrCrCrArArCrUrUrC

sR8 Length Mutants – D Target Only

1. D sR8 (5): rCrUrCrArU

2. D sR8 (7): rGrCrUrCrArUrA

3. D sR8 (9): rUrGrCrUrCrArUrArC

4. D sR8 (11): rUrGrCrUrCrArUrArCrGrG

5. D sR8 (18)5': rGrArGrUrCrUrGrCrUrCrArUrArCrGrGrUrC

6. D sR8 (18)3': rCrUrGrCrUrCrArUrArCrGrGrUrCrGrUrArG

sR6 Guide Mutant Primers

1. sR6 C51G Dn: AAACTGGCTCAGCTGCCGGAGTAGTC

2. sR6 C51A Dn: AAACTGGCTCAGCTGCAGGAGTAGTC

3. sR6 C51U Dn: AAACTGGCTCAGCTGCTGGAGTAGTC

4. T7-sR6 G24A:
CTAATACGACTCACTATAGGCCAAACTGGCGATGATGACAATTTCACTAT

5. T7-sR6 G24U:
CTAATACGACTCACTATAGGCCAAACTGGCGATGATGACAATTTCTCTAT

6. T7-sR6 G24C:
CTAATACGACTCACTATAGGCCAAACTGGCGATGATGACAATTTCCCTAT