ABSTRACT

GAGNON JR., KEITH THOMAS. Archaeal Box C/D sRNP Assembly, Structure and Function. (Under the direction of E. Stuart Maxwell.)

Archaeal box C/D sRNAs function by guiding the nucleotide-specific 2’-O ribose methylation of rRNA and tRNA. These small RNAs contain terminal box C/D and internal C’/D’ motifs which each bind three core sRNP proteins and fold into RNA K-turn motifs. The guide sequences of box C/D sRNAs, located upstream of both box D and D’ sequence elements, base pair with target RNAs to direct methylation. The target nucleotide five base pairs upstream of the box D or D’ is then 2’-O-methylated by bound sRNP core proteins.

Functional archaeal box C/D sRNP s can be assembled using \textit{in vitro} transcribed sRNA and three recombinant core proteins, L7, Nop56/58 and fibrillarin. These complexes can then be affinity purified for use in structural or enzymatic studies. Here methods for the preparation of recombinant core proteins and \textit{in vitro} transcribed sRNA are presented. \textit{In vitro} sRNP assembly, 2’-O-methylation activity assays, and techniques for the affinity purification of active box C/D sRNP complexes are also described.

\textit{In vitro} assemblies of box C/D sRNPs from the thermophilic archaean \textit{Methanocaldococcus jannaschii} have now been used to study core protein-mediated sRNA remodeling that occurs during the assembly of these sRNP complexes. L7 and Nop56/58 core proteins required elevated temperatures to efficiently bind the sRNA and both CD spectroscopy and thermal denaturation revealed an increase in order and stability of the sRNA after core protein binding, thus demonstrating that RNA structural changes take place during RNP assembly. Probing of sRNA structure and guide region accessibility during sRNP assembly of \textit{M. jannaschii} sR8, sR6 and mutant sRNAs further revealed variations in the assembly pathways taken by different sRNA species. RNA remodeling appears to be a common and requisite feature during box C/D sRNP assembly and variations in the remodeling pathway(s) are largely dependent upon the particular guide sequences of each sRNA.
Archaeal Box C/D sRNP Assembly, Structure and Function

by

Keith Thomas Gagnon, Jr.

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

MOLECULAR AND STRUCTURAL BIOCHEMISTRY

Raleigh, North Carolina

2007

APPROVED BY:

__________________________________                    ______________________________
Paul F. Agris                                                          A. Clay Clark
Committee Member                                                Committee Member

__________________________________
James W. Brown
Committee Member

__________________________________                    _____________________________
E. Stuart Maxwell                                                    Paul Wollenzien
Chair of Advisory Committee                                         Committee Member
DEDICATION

To my patient and loving wife Darci for her faith and sacrifice. Also to my mother and father for instilling in me the fortitude and values necessary for success. And finally, to the big guy upstairs for guiding me through the difficult seasons of life.
BIOGRAPHY

Keith Thomas Gagnon, Jr. was born in Fort Kent, Maine to Keith and Debra Gagnon on a cold November morning in 1980. He grew up in a variety of small towns in the New England states of Maine, New Hampshire and Vermont. His father’s work as a carpenter and tradesman also carried their family to Colorado and Kentucky during his youth. His family finally settled in North Carolina in the middle of his freshman year of High School. It was here that he began courting his future wife, Darci Campbell, who had also recently moved from upstate New York. Throughout high school and after graduation Keith held several jobs in the restaurant and construction businesses and nearly joined the Army. However, despite wild teenage years and a rebellious spirit, Keith maintained a high GPA in High School, which helped him to gain acceptance into the Biochemistry program at North Carolina State University in the spring of 2000. He began working in the lab of E. Stuart Maxwell as a junior and continued until graduating summa cum laude with honors in his major. Upon obtaining his Bachelors degree in the spring of 2003 he joined the Maxwell lab as a Masters student and later that year switched to the doctorate program. Now in the fall of 2007 Keith looks forward to completing graduate school with a Ph.D. in Biochemistry, starting a family with his wife and soon-to-be-born first child, and pursuing a career of teaching and research in academia.
ACKNOWLEDGEMENTS

I first thank my advisor and friend, E. Stuart Maxwell, for teaching me more than just how to do good research. He has given me insight into the potential difficulties and joys of being a research professor. I also thank my committee members, as well as all the faculty, staff and graduate students of the Department of Molecular and Structural Biochemistry, for their guidance, sharing of wisdom and knowledge, and often free-spirited humor and opinions.

I thank past and present members of the Maxwell lab for helping me to grow into a better research scientist with their insightful discussions and experimental know-how. Special thanks to Xinxin Zhang, who has been exceptionally helpful in tackling experimental difficulties and contributing to several research projects.

I also thank my family and close personal friends. They have encouraged me, made me laugh, and been a comfort to me during the less-palatable moments of graduate school and life in general. Especially supportive is my best friend and loving wife Darci. She is unconditionally there when the experiments work and there when they don’t.

Above all I thank God. Whether science can ever prove or disprove the existence of such, I acknowledge that a power greater than myself has guided me and a purpose higher than my own ambitions exists.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... ix
LIST OF FIGURES ......................................................................................................... x
LITERATURE REVIEW .................................................................................................. 1

I. A Brief History of the Box C/D RNAs and RNPs ...................................................... 2
II. Diversity of Box C/D RNA Populations ................................................................. 3
   A. Box C/D RNA Nomenclature ............................................................................ 3
   B. Box C/D RNA Structure .................................................................................. 4
   C. Diversity of Box C/D RNA Populations ......................................................... 6
   D. Box C/D RNA Identification .......................................................................... 6

III. Box C/D RNA Functions and Target RNAs ........................................................ 7
   A. Folding and Cleavage of Pre-rRNA ................................................................ 7
   B. 2'-O-Methylation of Diverse RNA Targets ..................................................... 9
   C. Additional Roles and Targets for Box C/D RNAs ............................................ 10

IV. Box C/D RNP Structure and Nucleotide Methylation Function ......................... 11
   A. Eukaryotic Box C/D Core Proteins and snoRNP Structure .......................... 11
   B. Archaeal Box C/D Core Proteins and In Vitro sRNP Assembly .................. 12
   C. Core Protein and RNP Crystal Structures ..................................................... 13
   D. Investigating Methylation Using In Vitro Assembled Archaeal Box C/D sRNPs ................................................................. 15

V. Box C/D RNP Biogenesis ....................................................................................... 16
   A. Genomic Organization of Eukaryotic Box C/D snoRNA Genes ............... 16
   B. Independently-Transcribed and Intronic Eukaryotic Box C/D snoRNA Genes ......................................................................................... 16
   C. Archaeal Box C/D sRNA Genes ................................................................... 18
   D. Transcription and Processing of Independently Transcribed Box C/D snoRNAs ......................................................................................... 19
   E. Transcription and Processing of Intronic Box C/D snoRNAs .................... 20
   F. Box C/D snoRNP Transport .......................................................................... 21
VI. Summary of Relevant Dissertation Research .......................................................... 23
   A. *In Vitro* Box C/D sRNP Assembly and Affinity Purification ..................... 23
   B. RNA Remodeling During Archaeal Box C/D sRNP Assembly ...................... 24
REFERENCES .................................................................................................................. 26

CHAPTER I. *In Vitro* Reconstitution and Affinity Purification of Catalytically
Active Archaeal Box C/D sRNP Complexes ............................................................... 42
   ABSTRACT .................................................................................................................. 43
   INTRODUCTION ........................................................................................................ 44
   RESULTS .................................................................................................................... 46
   I. Cloning, Expression, and Preparation of *M. jannaschii* Box C/D sRNP
      Core Proteins ........................................................................................................ 46
         A. Cloning of *M. jannaschii* L7, Nop56/58, and Fibrillarin Genes ............ 46
         B. Recombinant Core Protein Expression in Bacterial Cells ....................... 47
         C. Affinity Chromatographic Isolation of L7 and Fibrillarin Core
            Proteins ........................................................................................................... 48
         D. Isolation of Nop56/58 Core Protein by Cation-Exchange
            Chromatography ............................................................................................. 49
   II. Cloning and *In Vitro* Transcription of Archaeal Box C/D sRNAs ............. 50
   III. *In vitro* Assembly of the *M. jannaschii* sR8 Box C/D sRNP Complex ....... 51
   IV. Assessment of *In vitro* Assembled *M. jannaschii* Box C/D sRNP
      Methylation ........................................................................................................... 52
   V. Sequential Affinity Purification of *In Vitro* Assembled Box C/D sRNPs ...... 53
      A. An Overview ..................................................................................................... 53
      B. Affinity Chromatography Buffers ................................................................... 55
      C. Preparation of the Affinity Chromatography Resins .................................... 55
      D. Assembly of the sR8 sRNP Complex .............................................................. 56
      E. Tandem Affinity Purification of *In Vitro* Assembled Box C/D sRNP ......... 56
   VI. Concluding Remarks ............................................................................................ 60
ACKNOWLEDGEMENTS ................................................................. 62
REFERENCES ............................................................................. 63

CHAPTER II. Assembly of the Archaeal Box C/D sRNP Can Occur Via Alternative Pathways and Requires Temperature-Facilitated sRNA Remodeling .............................. 69
ABSTRACT ................................................................................... 70
INTRODUCTION ............................................................................ 71
RESULTS ........................................................................................ 74
   A. Efficient L7 Core Protein Binding is Temperature-Dependent and Requires sRNA Remodeling .................................................. 74
   B. L7 Binding Remodels and Stabilizes sR8 sRNA Structure .................. 75
   C. Ordered Assembly of the sR8 sRNP is Temperature-Dependent and Requires sRNA Structural Changes ................................. 76
   D. sR8 sRNA Remodeling Modulates D and D’ Guide Sequence Accessibility for Target RNA Binding ................................................. 78
   E. Temperature-Dependent Assembly of M. jannaschii sR6 sRNP Requires sRNA Remodeling But Uses an Alternative Assembly Pathway ........................................................................ 79
   F. sR6 Guide Regions Are Exposed Throughout sRNP Assembly .......... 81
   G. sRNA Guide Sequences Can Influence the sRNP Assembly and Remodeling Pathway .................................................................. 82
DISCUSSION .................................................................................. 84
MATERIALS AND METHODS ............................................................ 89
ACKNOWLEDGEMENTS ............................................................... 94
REFERENCES ............................................................................. 95

CONCLUDING REMARKS ............................................................ 112
REFERENCES ............................................................................. 115
APPENDICES ........................................................................................................ 116
APPENDIX A.  Scientific Meeting Abstracts  ....................................................... 117
APPENDIX B.  Purification of Recombinant Box C/D snoRNP Core Proteins
from Mouse and In Vitro Assembly of a Box C/D snoRNP  ......................... 131
APPENDIX C.  Protocols for Large-Scale Production of Recombinant
_Methanocaldococcus jannaschii_ Box C/D sRNP Core Proteins ............... 153
LIST OF TABLES

LITERATURE REVIEW
Table I. Box C/D RNA Databases and Search Engines ........................................... 8

CHAPTER I
Table I. In Vitro sRNP Assembly .......................................................... 65

CHAPTER II
Table I. 2’-O-Methylation Activity of Dual Guide Box C/D sRNPs ......................... 102

APPENDIX B
Table I. U15 snoRNP Assembly Reactions for EMSA ................................. 151
# LIST OF FIGURES

## LITERATURE REVIEW

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Archaeal and eukaryotic box C/D RNAs and their target RNAs</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Archaeal and eukaryotic box C/D ribonucleoprotein complexes</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Crystal structures of Archaeal box C/D sRNP core proteins</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Genomic organization of the box C/D RNAs</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>Schematic of eukaryotic box C/D snoRNP biogenesis and transport</td>
<td>22</td>
</tr>
</tbody>
</table>

## CHAPTER I

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Archaeal box C/D recombinant core protein isolation, <em>in vitro</em> RNP assembly, and sRNP-guided nucleotide 2'-O-methylation</td>
<td>66</td>
</tr>
<tr>
<td>2</td>
<td>Sequential affinity purification of <em>in vitro</em> assembled Archaeal box C/D sRNP</td>
<td>67</td>
</tr>
<tr>
<td>3</td>
<td>SDS-polyacrylamide gel electrophoretic analysis, methyltransferase activity, and electron microscopy of <em>in vitro</em> assembled <em>M. jannaschii</em> sR8 box C/D sRNPs</td>
<td>68</td>
</tr>
</tbody>
</table>

## CHAPTER II

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L7 core protein binding to both sR8 box C/D and C'/D' motifs requires elevated assembly temperatures</td>
<td>102</td>
</tr>
<tr>
<td>2</td>
<td>L7 binding to sR8 remodels and stabilizes sRNA structure</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>Complete assembly of the sR8 sRNP complex requires elevated assembly temperatures and sRNA remodeling</td>
<td>104</td>
</tr>
<tr>
<td>4</td>
<td>sR8 D and D’ guide region accessibility to target RNA binding changes during sRNP assembly</td>
<td>105</td>
</tr>
<tr>
<td>5</td>
<td><em>M. jannaschii</em> sR6 requires elevated temperature for efficient L7 binding</td>
<td>106</td>
</tr>
<tr>
<td>6</td>
<td>sR6 sRNP assembly requires sRNA remodeling</td>
<td>107</td>
</tr>
<tr>
<td>7</td>
<td>sR6 D and D’ guide regions are accessible throughout sRNP assembly</td>
<td>108</td>
</tr>
</tbody>
</table>
Figure 8. sRNA guide regions affect sRNA folding and L7 binding .................. 109
Figure 9. Altered sRNA guide regions affect sRNA structure and sRNP assembly ................................................................. 110
Figure 10. Archaeal box C/D sRNPs can follow different remodeling and assembly pathways ........................................................................................................... 111

APPENDIX B
Figure 1. Assembly of a eukaryotic box C/D snoRNP ................................. 152

APPENDIX C
Figure 1. Purification of *M. jannaschii* His-L7 ............................................ 158
Figure 2. Purification of the *M. jannaschii* ΔK Nop56/58:His-fibrillarin dimer ...... 164
LITERATURE REVIEW

Derived From:

The Box C/D RNPs: Evolutionarily Ancient Nucleotide Modification Complexes

Keith T. Gagnon, Xinxin Zhang and E. Stuart Maxwell

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

Published in RNA and DNA Editing: Molecular Mechanisms and their Integration into
Biological Systems, H. Smith editor, John Wiley and Sons, in press.
I. A Brief History of the Box C/D RNAs and RNP

Small RNAs were first identified in eukaryotic nuclei in the 1960's when a handful of abundant species were phenol-extracted from isolated nuclei and characterized on sucrose gradients and polyacrylamide gels (reviewed in (1)). Subsequent nucleotide analysis revealed a base composition distinct from GC-rich ribosomal RNA leading to their designation as U-rich small nuclear RNAs or the U snRNAs (2,3). It quickly became apparent that the U snRNAs are partitioned in the nucleoplasmic and nucleolar compartments (4-8). Eventual sequencing of the predominant snRNAs in the 1970s defined U3, the first nucleolar box C/D RNA (9-12). A few additional nucleolar RNA species were subsequently identified and this RNA population was designated the small nucleolar RNAs or snoRNAs in 1981 (13). The snoRNA population remained rather small until the early 1990's with the discovery of numerous intronic snoRNAs, encoded within pre-messenger RNA introns and excised from pre-mRNA transcripts during splicing (reviewed in (14)). Categorization of the expanding snoRNA population based upon conserved sequence elements defined the two major snoRNA families, the box C/D and the box H/ACA RNAs (15,16). Discovery of box C/D RNAs in archaeal organisms in the mid 1990s suggested an evolutionarily ancient origin for these small non-coding RNAs before the divergence of Archaea (17-20). The number of both eukaryotic and archaeal box C/D RNAs has continued to grow and now includes hundreds of species.

Nucleolar localization of the snoRNAs implied a role in ribosome biogenesis and experiments in *Saccharomyces cerevisiae* as well as *Xenopus laevis* demonstrated the essentiality of the major U3 box C/D snoRNA for rRNA processing (7,21-23). Several additional snoRNAs were identified as crucial for pre-rRNA processing, although many tested snoRNA species revealed no growth phenotype when deleted in yeast (24-26). Critical to understanding their primary function as guide RNAs was the observation that snoRNAs possessed regions of complementarity with rRNA (27). Strikingly, the snoRNA-complementary regions in the rRNAs corresponded to sites of nucleotide modification and it was soon demonstrated that the box C/D snoRNAs direct the nucleotide-specific 2'-O-
methylation of rRNA (28-30). Additional roles for the box C/D snoRNAs have since been demonstrated. U3 and U8 snoRNAs function as RNA chaperones, folding the pre-rRNA transcript which is required for precursor processing (31-34). The list of target RNAs modified by the box C/D RNAs has expanded to include the eukaryotic splicing snRNAs, pre-messenger RNAs, and archaeal tRNAs (18,35-37). The identification of tissue-specific box C/D snoRNAs may well suggest additional box C/D snoRNA populations with new functions guided via the RNA’s guide sequence in specific tissues (38-40).

In the last decade, investigations have focused upon identifying box C/D RNA-binding proteins. Highly homologous eukaryotic and archaean core proteins have been defined and the identification of eukaryotic accessory proteins suggests that snoRNP assembly may be a complex and regulated process (41-47). In vitro assembly systems for the archaean box C/D sRNP complex have been established and used to define box C/D RNP assembly pathways as well as reveal RNP structure and examine the mechanisms of box C/D RNA-guided nucleotide methylation function (48-50). These studies, coupled with emerging crystal structures of the core proteins and RNA:protein complexes, are now leading to a more detailed understanding of the structure and function of this RNA:protein enzyme. This review discusses current knowledge and ongoing studies examining the structure and function of both eukaryotic and archaean box C/D RNP complexes.

II. Diversity of Box C/D RNA Populations

A. Box C/D RNA Nomenclature

Box C/D RNAs are abundant in both Eukarya and Archaea, numbering in the hundreds of species. Following the original nomenclature, most animal and plant homologs are designated U(n) with the species number reflecting the approximate order of discovery. Homologs identified in other organisms usually adopt the same designation. However, there are organisms which use a non-conventional nomenclature. Yeast snoRNAs are designated by a number preceded by the prefix snR(n) (51). In Archaea, the snoRNA-like small RNAs or sRNAs are numbered and preceded by sR(n) (17). Other examples of non-standard
nomenclature include the protozoa *Trypanosoma brucei* (TBxCsyCz, x = chromosome #, y = cluster #, z = RNA species number) and *Euglena gracilis* (Eg-mx, x = RNA species number) (52,53). Different naming schemes have also been employed for tissue-specific or disease related box C/D snoRNAs, such as the mouse brain-specific box C/D snoRNA II-52 (MBII-52) or the box C/D snoRNA linked to neonatal lethality in Prader-Willi syndrome (Pwcr1/MBII-85) (38,54).

B. Box C/D RNA Structure

Box C/D RNAs are defined by their highly conserved nucleotide box C (RUGAUGA) and box D (CUGA) sequences positioned near the 5’ and 3’ termini, respectively (Figure 1), and frequently possess internal C’ and D’ boxes (15,29,55). The C’ and D’ nucleotide boxes are typically present and well conserved in the archaeal sRNAs but often degenerate and difficult to identify, or even missing, in the eukaryotic snoRNAs (17,18,56,57). Individual box C/D RNA species are defined by their unique guide sequences located immediately adjacent and upstream of boxes D and D’. These guide sequences base pair to the different target RNAs and determine the site of box C/D RNA action on the target RNA (28-30,58). Archaeal box C/D sRNAs are typically smaller (50-70 nucleotides) than the eukaryotic box C/D snoRNAs (75 to 150 nucleotides) (18,51,57,59-61). The highly conserved archaeal guide regions of 12 nucleotides most often constitute the entire spacer regions between the box C/D and C’/D’ motifs and account for the smaller size of the archaeal sRNAs (62). The well defined box C/D and C’/D’ motifs fold to establish kink-turn or “K-turn” and “K-loop” elements, respectively. The K-turn consists of canonical stem I and internal stem II separated by an asymmetric bulge which possesses tandem sheared G-A pairs essential for K-turn stability (44,63,64). The internal K-loop lacks stem I and is typically replaced with a small loop. Ultimately, both the K-turn and K-loop serve as core protein binding sites to assemble the box C/D and C’/D’ RNP complexes (48-50). The recent discovery of circularized box C/D sRNAs in *Pyrococcus furiosus* suggests a novel processing pathway that may augment box C/D sRNA stability in hyperthermophiles (65).
FIGURE 1: Archaeal and eukaryotic box C/D RNAs and their target RNAs. The secondary structures of archaeal and eukaryotic box C/D RNAs are shown with box C and D nucleotide sequences designated in blue and C’ and D’ sequences in red. Representative target RNAs (shown in grey) are base paired with the D and D’ guide sequences and the nucleotides targeted for methylation indicated. Box C/D and C’/D’ motifs are designated as well as those structural differences characteristic of the archaeal and eukaryotic box C/D RNAs. Known target RNAs of the archaeal and eukaryotic box C/D RNAs are shown below.
C. Diversity of Box C/D RNA Populations

Appreciation of box C/D RNA diversity in Eukarya and Archaea has come with the characterization of RNA populations in a variety of organisms in both kingdoms. For eukaryotes, approximately two thirds of the yeast snoRNAs involved in rRNA maturation are conserved and found in plants and humans (66). The yeast box C/D snoRNA population has been well defined and currently numbers 46 species (67). Compare this with more than double that number of species in human involved in rRNA maturation, which does not include additional species modifying other cellular target RNAs (68). Thus, the box C/D snoRNA populations of metazoan organisms are characteristically more complex with many of these additional RNAs unique to specific organisms. Box C/D homologs in different organisms may possess unique guide sequences reflecting organism-specific rRNA sequences yet modify corresponding nucleotides in the respective ribosomal RNAs. However, many box C/D snoRNA species modify rRNA nucleotides or nucleotides of other target RNAs that are unique to that organism (52). Thus, different eukaryotes typically possess a unique set of snoRNAs characteristic of that organism. While the majority of box C/D snoRNAs are involved in nucleotide methylation of rRNA nucleotides, conserved box C/D snoRNAs such as U3, U8, and U14 function in animals, plants, and fungi in non-modification roles such as facilitating pre-rRNA folding or cleavage during precursor processing (14). To date, there appears to be no shared box C/D RNA homologs between Archaea and Eukarya. Furthermore, box C/D sRNA populations among Crenarchaeota and Euryarchaeota are typically distinct with respect to the specific organism examined (18,57,69).

D. Box C/D RNA Identification

Hundreds of box C/D RNAs have been defined in numerous eukaryotic and archaeal organisms. Techniques to identify and define box C/D RNAs include both biochemical and computational approaches. Biochemical approaches include isolation of small RNAs fractionated on gradients or immunoprecipitated with antibodies against box C/D RNP proteins, followed by cDNA cloning and sequencing (70-72). Immunoprecipitation of box
C/D RNPs recently identified 66 previously unknown box C/D snoRNAs in the protist *Euglena gracilis* (53). Computational approaches have utilized a variety of search engines and algorithms to analyze available databases using box C and D consensus sequences (Table 1), as well as appropriately positioned guide sequences that exhibit complementarity to cellular target RNAs (73,74). Other sequence analysis-based searches have used snoRNA or intronic sequence databases to search for homologs in particular organisms (60,75). Computational approaches can identify potential box C/D RNAs that are expressed at low levels or in a tissue-specific manner whose detection by biochemical approaches might prove difficult. The utility of computational approaches was recently demonstrated by the identification of 62 novel box C/D RNAs from the trypanosome *Leishmania major* and 50 novel snoRNAs in *Caenorhabditis elegans* (76,77). Both candidate and experimentally confirmed RNAs are found in an array of databases which report ever increasing populations of archaeal and eukaryotic box C/D RNAs (Table 1). Biochemical and computational approaches have identified “orphan” box C/D RNAs where the associated guide sequence is not complementary to any known cellular RNA sequence, thus raising the question as to their possible function.

III. Box C/D RNA Functions and Target RNAs

A. Folding and Cleavage of Pre-rRNA

The base pairing of box C/D RNA guide sequences to their respective target RNAs determines the site of RNA function. The first identified eukaryotic snoRNAs were localized in the nucleolus, thus implying a role in pre-rRNA processing and/or ribosome biogenesis (7,22,78,79). Early experiments indeed demonstrated the importance of U3 (snR17), U8, U14, snR10, and snR30 box C/D snoRNAs for pre-rRNA cleavage (24,26,80-83). Those snoRNAs required for rRNA processing are typically essential RNAs, whereas those snoRNAs guiding nucleotide modification are not. The box C/D snoRNAs required for pre-rRNA cleavage at specific sites in yeast is based upon snoRNA gene disruption experiments. Similar observations have been made for *Xenopus laevis* rRNA processing when these same
TABLE 1: Box C/D RNA Databases and Search Engines

<table>
<thead>
<tr>
<th>DATABASE</th>
<th>WEBSITE</th>
<th>COMMENTS / UTILITY</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>snoRNA Database</td>
<td><a href="http://lowelab.ucsc.edu/snoRNAdb/">http://lowelab.ucsc.edu/snoRNAdb/</a></td>
<td>Box C/D snoRNAs from S. cerevisiae, A. thaliana, and currently 8 Archaea are sorted by name, target and genome locus.</td>
<td>Lowe and Eddy, 1999; Omer et al., 2000; Brown et al., 2001</td>
</tr>
<tr>
<td>snoRNA-LBME-db</td>
<td><a href="http://www.snoana.biotoul.fr/">http://www.snoana.biotoul.fr/</a></td>
<td>Human box C/D and H/ACA snoRNA database. Find snoRNAs by name, target sequence or modification, or genomic location.</td>
<td>Lestrade and Weber, 2006</td>
</tr>
<tr>
<td>Plant snoRNA Database</td>
<td><a href="http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/home">http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/home</a></td>
<td>Database of currently 18 plant species in a tabulated and downloadable sequence format.</td>
<td>Brown et al., 2003</td>
</tr>
<tr>
<td>Human snoRNA Database</td>
<td><a href="http://www.treex.ucsb.edu/snoRNA/">http://www.treex.ucsb.edu/snoRNA/</a></td>
<td>Contains the sequences of currently 463 human snoRNAs. Provides 2-D structure and energy information imported from mFold.</td>
<td></td>
</tr>
<tr>
<td>sno/scaRNAbase</td>
<td><a href="http://gene.fudan.sh.cn/snoRNAbase.html">http://gene.fudan.sh.cn/snoRNAbase.html</a></td>
<td>Database consisting of 1979 sno/scaRNAs from 85 organisms. Contains sequence, target site, accession number, references and allows users to perform BLAST searches.</td>
<td>Xie et al., 2007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALGORITHM / SEARCH ENGINE</th>
<th>WEBSITE</th>
<th>FUNCTION / UTILITY</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>snoSeeker</td>
<td><a href="http://snoSeeker.zsu.edu.cn/snoSeeker/">http://snoSeeker.zsu.edu.cn/snoSeeker/</a></td>
<td>Searches for box C/D snoRNAs in the genomic alignment of two or more organisms.</td>
<td>Yang et al., 2006</td>
</tr>
<tr>
<td>snoScan</td>
<td><a href="http://lowelab.ucsc.edu/snoScan/">http://lowelab.ucsc.edu/snoScan/</a></td>
<td>Searches mammalian, yeast, or archael genomic sequences for box C/D snoRNA genes with a probabilistic model starting from a query or target RNA sequence.</td>
<td>Lowe and Eddy, 1999; Schattner et al., 2005</td>
</tr>
<tr>
<td>SNO.pl</td>
<td><a href="http://ksc.utoledo.edu/bioinfo/cdn/sno/">http://ksc.utoledo.edu/bioinfo/cdn/sno/</a></td>
<td>Searches for conserved structures characteristic of C/D box snoRNAs within the Mammalian Orthologous Intron Database (MOID).</td>
<td>Fedorov et al., 2005</td>
</tr>
</tbody>
</table>
essential snoRNA species are depleted in oligonucleotide knockout experiments (14,21,34,84).

Some box C/D snoRNAs such as U3, U8, and U14 have been shown to be important for pre-rRNA folding and cleavage. These snoRNAs are considered to have “chaperone” function as their base pairing to target rRNAs facilitates correct pre-rRNA folding required for precursor maturation. U3 hydrogen-bonds with the 5’ ETS of the rRNA precursor and U14 base pairs with 18S rRNA (31,85,86). Both snoRNAs are essential for pre-RNA cleavage and production of mature 18S rRNA. U8 is required for proper folding of 5.8S and 28S rRNAs within the precursor transcript and is essential for 28S rRNA production (32-34). Interestingly, U14 not only guides 18S rRNA processing using its C’/D’ guide sequence, but also guides nucleotide methylation using its box C/D guide sequence (84). Recent work utilizing computer simulations has suggested that snoRNAs may play important roles in long range rRNA folding (87). To date, eukaryotic box C/D RNAs have not been implicated in chaperone folding or precursor processing functions for any RNA other than pre-ribosomal RNA.

B. 2’-O-Methylation of Diverse RNA Targets

The primary role of most box C/D RNAs is to guide the modification of a targeted nucleotide’s ribose sugar, catalyzing addition of a methyl group at the 2’ hydroxyl position (Figure 1). The nucleotide targeted for methylation is positioned within the RNA duplex formed by the hydrogen bonded target:guide RNA sequences and base paired to the fifth nucleotide upstream from the snoRNA’s D or D’ box (28-30). The large number of nucleotides modified by box C/D RNAs in eukaryotes and archaea explains the extensive box C/D populations found in each. Approximately 100 ribose methylations are found in human rRNA, about 120 in plant Arabidopsis thaliana rRNA, and 55 in yeast rRNA (88-90). The occurrence of methylated nucleotides in archaeal rRNA is also considerable with 67 identified in Sulfolobus solfataricus and at least 26 and 93 predicted in Sulfolobus acidocaldarius and Pyrococcus abyssi, respectively (18,57,91). The specific rRNA
nucleotides methylated vary from organism to organism, but these modifications tend to be clustered in conserved core and functional regions of the rRNA (90). Disruption of one or a few modifications at specific nucleotides often does not result in an obvious phenotype (90), although complete disruption of nucleotide methylation in yeast has severe effects upon rRNA maturation and ribosome function (92).

The types of RNA targeted for methylation by box C/D RNAs extends beyond the ribosomal RNAs to include archaeal tRNAs and the eukaryotic splicing snRNAs. Many of the modifications in archaeal rRNA and tRNAs are predicted by newly defined box C/D sRNAs and their associated D and D’ guide sequences (18). Also found in Archaea and Eukarya are “orphan” RNAs where the guide sequence(s) of an identified C/D RNA is not complementary to any sequence in a known cellular RNA (58,74). In eukaryotes, specific box C/D RNAs guide nucleotide methylation of U1, U2, U4, U5, U6 and U12 snRNAs (35,93,94). Specific species of box C/D RNA family are also found localized to the nuclear Cajal bodies and designated small Cajal RNAs or scaRNAs. These scaRNAs are unique in that they possess both box C/D and box H/ACA RNA motifs and direct the 2’-O-methylation and pseudouridylation of the RNA Pol II-transcribed snRNAs U1, U2, U4, and U5 within the Cajal bodies (35).

C. Additional Roles and Targets for Box C/D RNAs

Investigation of specific tissues has suggested a larger population of box C/D RNAs than originally anticipated. Analysis of small RNAs in mouse, rat and human brain has identified many brain-specific box C/D snoRNAs (38-40). Only the mouse/human brain-specific box C/D snoRNA MBII-52/HBII-52 has been mapped to a putative target. This box C/D snoRNA appears to target an A-to-I editing site in the serotonin receptor 2C (5-HT2C) pre-mRNA. Methylation of this site potentially modulates pre-mRNA editing and alternative splicing, thereby resulting in changes to serotonin signaling capacity (36). Alternatively, it has been suggested that the HBII-52 snoRNA uses its guide region sequences to affect alternative splicing of the 5-HT2C pre-mRNA by masking splicing silencers which then leads
to the inclusion of an exon in the final spliced serotonin receptor mRNA (37). The loss of MBII-85/HBII-85 is implicated in Prader-Willi syndrome (95). Furthermore, regulation of MBII-52 and MBII-48 snoRNA levels has been noted in the early phase of memory consolidation during learning, thus suggesting a possible role for this snoRNA in higher brain function (96).

IV. Box C/D RNP Structure and Nucleotide Methylation Function

A. Eukaryotic Box C/D Core Proteins and snoRNP Structure

Box C/D snoRNAs bind a small set of core proteins to establish ribonucleoprotein (RNP) complexes (Figure 2). Terminal box C/D and internal C’/D’ snoRNA motifs serve as protein binding sites for RNP assembly and it is the bound core proteins that carry out the nucleotide 2’-O-methytransferase reaction. Four well conserved core proteins bind eukaryotic box C/D snoRNAs. They are the 15.5kD protein (Snu13p in yeast), nucleolar proteins 56 (Nop56) and 58 (Nop58), and the methyltransferase fibrillarin. The 15.5kD protein directly binds the box C/D K-turn, but not box C’/D’ motifs, initiating assembly of box C/D snoRNPs (44,97-99). The use of nucleotide analog interference mapping (NAIM) has suggested that 15.5kD’s inability to bind C’/D’ motifs may lie in slight structural differences that arise when a K-turn possesses a terminal loop instead of a stem II structure (100).

The Nop56 and Nop58 core proteins are highly homologous and essential for nucleotide modification and ribosome biogenesis (41,42). Cross-linking experiments have shown that Nop58 interacts with box C whereas Nop56 interacts with box C’ (98). This suggests that Nop58 and Nop56 core proteins bind the K-turn and K-loop, respectively, although their asymmetric distribution in the snoRNP complex has not yet been confirmed by additional experiments. Interestingly though, binding of both Nop56 and Nop58 requires stem II of the box C/D motif (101). These core proteins are likely to be RNP structural proteins, although their role in the methylation reaction or its regulation cannot be ruled out at this time.
The methyltransferase fibrillarin interacts with both Nop56 and Nop58 and can be cross-linked to both box C/D and C’/D’ motifs (41,98). Evidence establishing fibrillarin as the methyltransferase enzyme includes the presence of an S-adenosylmethionine (AdoMet)-dependent methyltransferase-like domain that is required for methylation activity (102,103) and the observation that disruption of the yeast fibrillarin gene results in the loss of prerRNA methylation (92).

B. Archaeal Box C/D Core Proteins and In Vitro sRNP Assembly

Archaeal box C/D sRNAs bind three core proteins which are highly homologous to the eukaryotic core proteins (43,48,49). Ribosomal protein L7, homologous to eukaryotic 15.5kD, recognizes both box C/D and C’/D’ motifs to initiate box C/D and C’/D’ RNP assembly. A single Nop56/58 core protein, homologous to the Nop56 and Nop58 pair, binds both box C/D and C’/D’ motifs. The third core protein, fibrillarin, also binds both motifs to

FIGURE 2: Archaeal and eukaryotic box C/D ribonucleoprotein complexes. Archaeal and eukaryotic box C/D RNAs are shown with the RNP core proteins bound to the box C/D and C’/D’ motifs. The binding of all three archaeal core proteins to both the box C/D and C’/D’ motifs assembles a symmetric sRNP. The differential distribution of the eukaryotic core proteins 15.5kD, Nop56, and Nop58 binding to the box C/D and C’/D’ motifs establishes an asymmetric snoRNP. (Crosslinking experiments have implied the differential distribution of Nop58 and Nop56 on the eukaryotic box C/D and C’/D’ motifs, respectively (94)).
complete box C/D sRNP assembly. Much of what is presently known about box C/D RNP structure and assembly comes from examination of in vitro assembly of the archaeal sRNP using in vitro synthesized box C/D sRNAs and recombinant sRNP core proteins. These in vitro assembly studies revealed an order of assembly, with L7 initiating RNP formation followed by Nop56/58 and then fibrillarin binding (48-50). Binding of the L7 core protein induces structural changes in both box C/D and C’/D’ motifs (104-106). These structural changes are required for subsequent Nop56/58, but not fibrillarin, binding. Nop56/58 and fibrillarin form a stable dimer in the absence of the sRNA and/or L7 and evidence suggests that it is the dimer that binds the assembling box C/D and C’/D’ RNP complexes (49,106-108). In contrast to the “asymmetric” eukaryotic snoRNP complex, the binding of all three core proteins to the box C/D and C’/D’ motifs of archaearal sRNAs assembles a “symmetric” sRNP (Figure 2).

C. Core Protein and RNP Crystal Structures

Emerging crystal structures of core proteins, the K-turn, and sRNA:core protein sub-complexes are beginning to reveal the detailed architecture of the box C/D RNP. The ease of expressing soluble, recombinant archaearal core proteins has greatly facilitated these studies and all three archaearal core proteins have yielded crystal structures at detailed resolution (Figure 3). Crystal structures of free L7 core protein, the K-turn, and the L7:K-turn RNP have shown that L7 binding induces an RNA conformational change (64,99,104,105,109). This induced fit has been confirmed in fluorescence resonance energy transfer (FRET) experiments (105). Both the L7 and 15.5kD core proteins lock the RNA into a tightly kinked structure characterized by stacked stem I and stem II helices bent at approximately 60°, hence the designation of kink-turn for this RNA element (63,110). Particularly important for binding is the requirement of a pyrimidine nucleotide from box C (UGAUGA) which is extended and inserted into a cleft of the L7 protein. A co-crystal structure of Archaeoglobus fulgidus Nop56/58 and fibrillarin has provided the molecular details of this core protein dimer (107). S-adenosylmethionine (AdoMet) is bound to the methyltransferase domain of fibrillarin and biophysical studies have indicated that Nop56/58 binding to fibrillarin helps to
FIGURE 3: Crystal structures of Archaeal box C/D sRNP core proteins. (A) Superimposed crystal structures of ribosomal protein L7 from *M. jannaschii* both free (PDB accession number, 1RA4) and bound to a K-turn RNA (1SDS). (B) Crystal structure of *A. fulgidus* fibrillarin-Nop56/58 core protein dimer complexed with S-adenosylmethionine (1NT2).
stabilize it in the binding pocket (111). A recent crystal structure of the human Nop56/58 homolog hPrp31 bound to the 15.5kD:U4 K-turn RNP has revealed that Nop56/58 and hPrp31 bind the K-turn-L7/15.5kD protein complex through their Nop domains and require interaction with both the K-turn and the L7/15.5kD protein (112). Despite crystallization of the individual components of the complex, a crystal structure of the fully assembled box C/D sRNP has thus far proved elusive.

D. Investigating Methylation Using In Vitro Assembled Archaeal Box C/D sRNPs

The archaeal in vitro assembly system has enabled investigations of box C/D sRNP methylation function with respect to sRNP structure and the role of the sRNA and individual core proteins. The assembled box C/D sRNP is catalytically active and guides site-specific methylation of target RNAs using both D and D’ guide sequences. Target RNA methylation is nucleotide-specific and dependent upon fibrillarin binding and the presence of S-adenosylmethionine (48). Further studies have shown that efficient methylation of D and D’ guide RNAs requires that both box C/D and C’/D’ RNPs be juxtaposed in a fully assembled sRNP (49). Moreover, the juxtaposed box C/D and C’/D’ RNPs are dependent upon the highly conserved spatial positioning of each RNP separated by 12 nucleotides (62). This suggests molecular interactions between the two complexes and/or that concerted sRNA remodeling is important for methylation function. This appears to be unlikely for the eukaryotic box C/D snoRNPs exhibiting dual guide function since the spatial positioning of box C/D and C’/D’ RNPs is not conserved and is often quite distant. Finally, both in vivo and in vitro studies have examined the base pairing interactions of the target RNA with the sRNA guide sequence (113,114). A minimum RNA:RNA duplex of 9-11 nucleotides is required and Watson-Crick pairing is essential. Interestingly, the ability of the in vitro assembled complex to methylate target nucleotides positioned within thermally stable double-stranded secondary structures may suggest an ability of the core complex to facilitate target RNA melting necessary for base pairing with the sRNA guide sequence.
V. Box C/D RNP Biogenesis

A. Genomic Organization of Eukaryotic Box C/D snoRNA Genes

The genomic organization of eukaryotic box C/D RNAs falls into two major categories. Some box C/D snoRNA genes are independently transcribed from snoRNA-specific promoters with the primary transcript possessing single or multiple snoRNA species. However, the majority of box C/D snoRNAs are encoded within introns of host genes and are transcribed by RNA polymerase II as part of the host precursor transcript (Figure 4). Fungi, plants, trypanosomes and unicellular organisms possess primarily independently-transcribed snoRNA genes, although snoRNAs encoded within host gene introns are present. In metazoan organisms, snoRNA coding sequences are overwhelmingly positioned within host gene introns (reviewed in (14) and (58)). For both snoRNA categories, the primary transcripts undergo post-transcriptional processing to produce mature snoRNA species.

B. Independently-Transcribed and Intronic Eukaryotic Box C/D snoRNA Genes

Box C/D snoRNA genes that are independently-transcribed using a snoRNA-specific promoter may be found as single or clustered species. U3 is the best characterized box C/D snoRNA that is independently-transcribed as a single snoRNA. U3 genes have been identified and characterized in fungi, protists, vertebrates, insects and plants (reviewed in (14)). They are most often found in multiple copies and dispersed throughout the genome, although occurrence of pseudogenes in vertebrates can make determination of gene copy number difficult (115-117). Box C/D snoRNA genes in yeast, plants, and trypanosomes are often organized into clusters and are transcribed as a polycistronic precursor using a snoRNA-specific promoter (89,118-121). These polycistronic transcripts are subsequently processed to produce the mature individual box C/D snoRNAs.

The predominant box C/D snoRNA gene organization in metazoans, and particularly vertebrates, is that of encoding within introns of host RNA Pol II transcripts. Typically, the intronic snoRNAs are located in protein-coding pre-mRNA introns and are frequently found
in multiple introns of common host genes with each intron limited to a single encoded snoRNA. Intronic box C/D snoRNAs are also often found in the same host gene in different organisms but this organization is not universal, and frequently a snoRNA species is found in
the same host gene but in different introns (reviewed in (14)). Host genes often possess isomers of the same intronic snoRNA species (122-124). Host genes also encode different snoRNA species in different introns ((125,126) and reviewed in (14)). The intronic snoRNAs are characteristically found within host genes that encode proteins involved in ribosome biogenesis or protein synthesis and ribosomal protein genes very often contain intronic snoRNAs (14,58). The positioning of intronic snoRNAs within protein coding genes important for ribosome biogenesis and protein synthesis suggests a possible coordination of the snoRNA and ribosome biogenesis pathways. Of particular interest are the intronic box C/D snoRNAs encoded within pre-mRNAs that do not encode a protein, such as the UHG genes of human and *Drosophila melanogaster* (127-133). In these cases, the host precursor transcript simply serves as a carrier for the encoded intronic snoRNAs.

C. Archaeal Box C/D sRNA Genes

Most of the limited information concerning archaeal box C/D sRNA genes comes from an analysis of several hyperthermophile *Pyrococcus* genomes (57). Box C/D sRNA genes are primarily intergenic, although some do overlap upstream and/or downstream open reading frames on the same DNA strand (Figure 4). They typically exhibit very little clustering although in *S. solfataricus* and *P. fulgidus* two sRNA genes (sR10/sR11 and sR26/sR60, respectively) are separated by only a few nucleotides and appear to be co-transcribed (18). In *P. abyssi*, the box C/D sRNA genes are preferentially located in non-coding regions of the genome, again with little clustering. Interestingly, the box C/D RNA sR40 of *Pyrococcus* and *A. fulgidus* is found within the intron of the tRNA<sup>Trp</sup> gene and this sRNA is responsible for 2’-O-methylation of two nucleotides in the tRNA<sup>Trp</sup> itself (132-134). Nothing is presently known about the RNA polymerase promoters responsible for sRNA transcription as obvious promoter elements are not observed and detailed analysis of possible sRNA precursor transcription has not been examined.
D. Transcription and Processing of Independently Transcribed Box C/D snoRNAs

Independently transcribed box C/D snoRNA genes possess their own promoter and terminator regions as well as enhancer elements flanking the coding sequence. They typically possess promoters similar to protein coding genes and are typically transcribed by RNA Polymerase II (14,132,135). Transcription of U3 is driven by a Pol II promoter in vertebrates and possesses several conserved sequence elements. These include TATA-like boxes, proximal and distal sequence elements (PSE and DSE), and a “U3 box” specific for U3 at the DSE (116,136-138). Plant U3 genes also possess TATA-like boxes but are transcribed by RNA Pol III due to the shorter spacing between the TATA box and upstream sequence elements (USE) (139,140). The U3 transcripts of vertebrates are capped at the 5’ terminus with trimethylguanosine (TMG) whereas those of plants are O-methyl capped (141). Other Pol III transcribed box C/D snoRNA genes include snR52 in yeast (142) and several plant box C/D snoRNAs clustered with tRNA genes (Figure 4). The plant snoRNA genes are co-transcribed with the tRNA as a tRNA-snoRNA precursor using the tRNA gene’s RNA Pol III promoter (143). Yeast promoters of box C/D snoRNA genes that are independently transcribed by Pol II often contain A/T rich stretches and a TATA box. Upstream Rap1p or Abf1p binding sites that are typical for yeast ribosomal protein genes are also sometimes present (118). Box C/D snoRNA promoters for independent transcripts in plants are not well characterized but do contain putative TATA boxes (119). Independently-transcribed box C/D snoRNAs are often polycistronic. The nascent transcripts are capped with trimethylguanosine (TMG) but undergo 5’ end processing involving endonucleases to produce processing intermediates that are matured by trimming exonucleases. In yeast, the ortholog of bacterial endonuclease RNase III, Rnt1p, and the exonucleases Rat1p and Xrn1p are responsible for snoRNA processing and trimming (118,144-146).

Termination of Pol II-transcribed genes can result in either polyadenylated or non-polyadenylated transcripts. The box C/D snoRNAs are not polyadenylated but current understanding of 3’ end formation is complicated by the overlapping machinery involved in these two pathways and shared components which include the core cleavage and
polyadenylation factor (CPF) complex and the exosome (147). 3’ End formation of independently transcribed box C/D snoRNA precursors involves factors Nrd1p, Nab3p, and Sen1p which appear to be specific for non-poly(A) termination and 3’ end pre-snoRNA processing (148). Nrd1p and Nab3p recognize and bind specific sequence elements upstream of the termination signal of box C/D snoRNA transcripts, although their recognition of these sequences alone is not sufficient to prevent polyadenylation and direct proper 3’ end formation (149). Nrd1p interacts with the exosome, cap-binding complex, and Pol II and is therefore implicated in coordinating transcription and 3’ end formation with exosome processing (150). 3’ End formation is also dependent upon the co-transcriptional assembly of the box C/D snoRNP (151,152). Thus, termination and 3’ end formation may rely on a monitoring mechanism whereby Nrd1p, Nab3p, and Sen1p are involved in detecting the assembled snoRNP upstream of the termination signal and directing a bypass of transcript polyadenylation while activating exosome-mediated 3’ end processing.

E. Transcription and Processing of Intronic Box C/D snoRNAs

Box C/D snoRNA genes, particularly those of metazoan organisms, are positioned within introns of host genes that may be protein- or non-protein coding. The intronic snoRNA host genes are primarily driven by promoters containing terminal oligopyrimidine tracts, or TOP promoters, which are characteristic of a broad family of protein-coding genes involved in ribosome biogenesis and protein synthesis (131,153-155). Intronic snoRNAs are excised from the intron during host precursor splicing (reviewed in (14)). Processing of intronic box C/D snoRNAs requires specific positioning of the snoRNA within the host intron, approximately 50 nucleotides upstream of the branchpoint (156,157). Recently, a splicing factor designated intron binding protein 160 (IBP160) was identified that defined the snoRNA distance from the branchpoint (158). IBP160 binds the intron approximately 35-40 nucleotides upstream of the branch point in a sequence-independent manner. Intronic snoRNA processing also requires snoRNP assembly. Immunoprecipitation experiments with fibrillarin antibodies have shown that the snoRNP complex is assembled while the snoRNA is still part of the unspliced, host pre-mRNA transcript. Accordingly, the box C/D snoRNP
was found to assemble in the C1 splicing complex where IBP160 exerts its function (159). Thus, it appears that IBP160 may be a key factor linking intronic box C/D snoRNP assembly and intronic snoRNA processing with host pre-mRNA splicing events.

Upon pre-mRNA splicing, box C/D snoRNAs are excised from the intron and trimmed to a mature species using exonucleases. In yeast, an RNase III enzyme Rnt1p debranches the lariat intron. Rnt1p cleavage requires interaction with Nop1p (fibrillarin) of the assembled box C/D snoRNP (160). This cleavage provides entry sites for exonucleolytic trimming by the 5’-3’ exonucleases Rat1p and Xrn1p and the 3’-5’ trimming activities of Rrp6p and the exosome (161,162). Using a minor pathway, intron-encoded snoRNAs can be directly excised from unspliced pre-mRNA using an endonuclease followed by exonucleolytic trimming (163,164). Box C/D snoRNAs that are processed directly from unspliced pre-mRNAs appear to be released by a mechanism similar to that observed for processin of independently-transcribed snoRNAs (160).

F. Box C/D snoRNP Transport

Box C/D snoRNA transcription, processing, and snoRNP assembly occurs in the nucleoplasm before snoRNP transport into the nucleolus (Figure 5). Transport to the nucleolus requires the box C/D core motif and all four core proteins for nucleolar localization (165-169). Experiments indicate that in higher eukaryotes, box C/D snoRNPs undergo transit through Cajal bodies for maturation before entering the nucleolus (168). In Hela cells, independently-transcribed U3 snoRNA precursors are found in the nucleoplasm as large multiprotein processing complexes containing factors linked to RNA processing (hRrp46, LSm 2-8, La protein), snoRNP assembly (Tip48, Tip49, Nopp140), and RNA transport (PHAX, CRM1) (170). The snRNA export factor PHAX is required for U3 transport to the Cajal body where it is m^3G-capped and undergoes CRM1-dependent transport to the nucleolus (171,172). Intronic and uncapped box C/D snoRNAs may use a variation of this nucleolar transport pathway. Accordingly, U14 nucleolar localization is impaired when
PHAX is depleted and some intronic and uncapped box C/D snoRNAs can bind PHAX in the absence of the cap binding complex (170,172).

FIGURE 5: Schematic of eukaryotic box C/D snoRNP biogenesis and transport.
VI. Summary of Relevant Dissertation Research

A. In Vitro Box C/D sRNP Assembly and Affinity Purification

The ability to assemble functional archaeal box C/D sRNPs in vitro has greatly facilitated in-depth investigations of assembly, structure and function of these complexes. Importantly, it has complemented high-resolution crystal structure analyses and in vivo studies. In Chapter I of this dissertation, detailed protocols for the production of recombinant core proteins and in vitro transcribed sRNA are supplied. In vitro sRNP assembly and activity assays, as well as effective methods for purification of fully assembled box C/D sRNPs, are also described.

Cloning, overexpression in E. coli, and chromatographic isolation of recombinant core proteins L7, Nop56/58 and fibrillarin from the thermophilic organism Methanocaldococcus jannaschii is discussed in a “how-to” format. Also, the cloning and synthesis of box C/D sRNAs using in vitro T7 transcription kits is provided. Procedures for in vitro assembly of these RNP complexes from recombinant core proteins and synthesized RNAs, as well as methylation assays to test the enzymatic activity of these RNP complexes, are also described. Most importantly, a step-by-step protocol for the purification of fully assembled box C/D sRNPs is put forth.

Full length box C/D sRNAs contain two protein binding sites, the terminal box C/D K-turn and the internal box C’/D’ K-loop, which each bind a copy of L7, Nop56/58 and fibrillarin (48,49). This makes selection of fully assembled sRNP complexes from partially assembled complexes based on protein composition alone difficult. The purification technique presented here exploits the fact that fibrillarin is the last protein to bind and is required for methylation activity (48,49). Box C/D sRNP complexes are assembled with sRNA, L7, Nop56/58 and two differentially tagged fibrillarins, His(6x)-tagged fibrillarin and FLAG-tagged fibrillarin. Complexes are purified by three tandem affinity steps: 1) an RNA specific selection step then 2) selection with ANTI-Flag affinity resin and 3) Ni^{2+}-NTA resin. This purification procedure may be scaled up accordingly to produce pure and fully assembled sRNP suitable for structural and enzymatic studies requiring near homogeneous particles.
B. RNA Remodeling During Archaeal Box C/D sRNP Assembly

The relatively recent advent of in vitro assembly systems for archaeal box C/D sRNPs has resulted in the characterization of many important features of box C/D RNAs and their core protein components. These include the order of core protein binding, protein-protein and protein-RNA interactions necessary for assembly, and requirements for a functional and minimal box C/D RNP (48,49,62,108,113,173,174). Chapter II of this dissertation contributes to these studies by offering new insight to the mechanisms of box C/D sRNP assembly. We found that efficient assembly of the archaeal box C/D sRNP required core protein-mediated RNA remodeling facilitated by elevated temperatures. Moreover, the pathway of RNA remodeling varied among different sRNAs tested and was largely dependent upon the guide region sequence of each sRNA species.

Initial investigation of box C/D sRNA structural changes were based on comparison of crystal structures of L7 free and bound to a K-turn RNA and fluorescence resonance energy transfer (FRET) experiments (64,104,105). These studies concluded that L7 binding to K-turn RNAs was an induced-fit interaction, with the L7 protein undergoing little or no change and the RNA itself being locked into a tight kink-turn structure. The effect on RNA structure of other core proteins binding to box C/D motifs, however, had not been addressed. Using a number of techniques including temperature-dependent binding assays, RNA thermal denaturation, circular dichroism, and nuclease mapping, we were able to verify RNA structural changes upon L7 binding to the M. jannaschii sRNA sR8. However, these same experimental approaches demonstrated a previously unappreciated role for Nop56/58 in RNA remodeling during sRNP assembly. Like L7, efficient Nop56/58 binding required elevated temperatures and had significant effects on the accessibility of guide regions to complementary RNA probes. Interestingly, fibrillarin’s binding was not temperature dependent and had no effect on RNA structure. This result is in agreement with previous work suggesting that fibrillarin binds via protein-protein interactions only (49,107,108). To determine how general RNA remodeling was to box C/D sRNP assembly, other M. jannaschii box C/D sRNAs and mutants thereof were analyzed. Surprisingly, we found that
different RNA remodeling pathways were utilized by different sRNA species in a manner dependent upon the sRNA’s guide region sequences. These conclusions were based upon core protein binding, the degree of temperature-dependent assembly, and the accessibility of guide regions for target RNA binding. Regardless of the different RNA remodeling pathway followed, assembled box C/D sRNPs were all enzymatically active. Investigation of box C/D sRNA remodeling during \textit{in vitro} sRNP assembly has led to a proposed general mechanism of assembly and provided insight into the requirements of assembly and function for these RNP complexes.
REFERENCES


CHAPTER I

*In Vitro* Reconstitution and Affinity Purification of Catalytically Active Archaeal Box C/D sRNP Complexes

Keith T. Gagnon, Xinxin Zhang and E. Stuart Maxwell

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

ABSTRACT

Archaeal box C/D RNAs guide the site-specific 2’-O-methylation of target nucleotides in ribosomal RNAs and tRNAs. *In vitro* reconstitution of catalytically active box C/D RNPs using *in vitro* transcribed box C/D RNAs and recombinant core proteins provides model complexes for the study of box C/D RNP assembly, structure, and function. Described here are protocols for assembly of the archaeal box C/D RNP and assessment of its nucleotide modification activity. Also presented is a novel affinity purification scheme employing differentially tagged core proteins and a sequential three step affinity selection protocol that yields fully assembled and catalytically-active box C/D RNPs. This affinity selection protocol can provide highly purified complex in sufficient quantities for not only for biochemical analyses but also biophysical approaches such as cryo-electron microscopy and x-ray crystallography.
INTRODUCTION

The box C/D RNAs constitute large populations of small non-coding RNAs found in both eukaryotic and archaeal organisms where their primary function is to guide the site-specific 2’-O-methylation of nucleotides located in various target RNAs. Guide sequences within each box C/D RNA base pair to complementary sequences in the target RNA, thereby designating specific nucleotides for post-transcriptional modification. (1-4). Archaeal box C/D sRNAs are defined by highly conserved boxes C and D located near their 5’ and 3’ termini and internally located C’ and D’ boxes (5,6). Both the external boxes C and D and the internal C’ and D’ boxes fold to establish kink-turn or K-turn motifs. These highly structured K-turn (box C/D) and K-loop (C’/D’) motifs serve as binding platforms for the box C/D RNP core proteins (7-11). The archael core proteins ribosomal protein L7, Nop56/58, and fibrillarin bind both box C/D and C’/D’ motifs to establish individual RNP complexes. It is the core proteins, working in concert with the guide regions located immediately upstream of the D and D’ boxes, that direct the 2’-O-methylation of targeted nucleotides (7,8).

Investigation of box C/D RNP structure and function has been greatly facilitated in recent years with the establishment of in vitro systems that assemble catalytically active archaeal sRNP complexes using in vitro transcribed sRNAs and recombinant sRNP core proteins purified from bacterial expression systems. Several laboratories, including our own, have used these in vitro assembled complexes to investigate the assembly, structure, and methylation function of this RNA-protein enzyme (for review see (12)). Presented here are detailed protocols for the assembly of a Methanocaldococcus jannaschii box C/D sRNP and the assessment of the complex’s methylation capabilities. Also described here is a novel sRNP isolation protocol involving three sequential affinity selection steps utilizing differentially tagged fibrillarin core proteins and an oligonucleotide complementary to the sRNA. This purification scheme yields highly purified and fully assembled archaeal box C/D sRNPs in sufficient quantities for not only biochemical analyses, but also biophysical
approaches such as cryo-electron microscopy and x-ray crystallography. While specifically designed for the *M. jannaschii* sR8 box C/D sRNP, this approach can be easily modified for the isolation of other sRNP complexes assembled either *in vitro* or possibly *in vivo* in the cell.
RESULTS

I. Cloning, Expression, and Preparation of *M. jannaschii* Box C/D sRNP Core Proteins

Genes encoding the *Methanocaldococcus jannaschii* core proteins L7, Nop56/58 and fibrillarin were PCR-amplified from isolated genomic DNA, inserted into bacterial expression vectors, and recombinant proteins expressed and then purified using affinity and cation-exchange chromatography as previously outlined (8). The cloning, expression, and purification of each core protein are presented here in greater detail.

A. Cloning of *M. jannaschii* L7, Nop56/58, and Fibrillarin Genes

Ribosomal protein L7, fibrillarin, and Nop56/58 gene coding sequences are PCR-amplified from *M. jannaschii* genomic DNA. DNA oligonucleotide primers are synthesized for each core protein gene and used for PCR amplification. The L7 and fibrillarin upstream and downstream primers contain 5’ terminal NdeI and BamHI restriction sites, respectively, whereas the Nop56/58 primers contain NcoI and BamHI restriction sites. Following PCR amplification, resulting DNA fragments are digested with the appropriate restriction endonucleases and ligated into similarly digested pET28a (Novagen) protein expression vectors using standard methods. Plasmid constructs are transformed into *E. coli* cells and colonies screened for plasmids containing core protein sequences by PCR amplification using L7-, Nop56/58-, or fibrillarin-specific primers. Selected colonies are cultured in liquid LB broth with antibiotics and plasmid DNA harvested using Wizard Midiprep Kits (Promega). Expression of recombinant L7 and fibrillarin proteins produces N-terminal and thrombin-cleavable 6X-histidine tags. The expressed Nop56/58 recombinant protein is untagged. To generate FLAG-tagged fibrillarin, the fibrillarin coding sequence is PCR amplified from genomic *M. jannaschii* DNA and ligated into the pET28a vector engineered to encode a FLAG peptide (DYKDDDDK) in place of the His tag. The sequences of all core protein plasmid constructs are verified by DNA sequencing.
B. Recombinant Core Protein Expression in Bacterial Cells

Core proteins are expressed in the Rosetta (DE3) strain of \textit{E. coli} (Novagen), a protein expression host carrying a plasmid encoding rare tRNA codons to help enhance levels of recombinant protein synthesis. Competent cells are transformed with individual plasmids and spread onto LB agar plates containing kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). A single cell colony is selected and grown in 1 liter of LB broth containing antibiotics with shaking at 37°C until the cell culture reaches an optical density at 600 nm (OD$_{600}$) of 0.8. Expression of His-L7, His-fibrillarin, and FLAG-fibrillarin is initiated by the addition of isopropyl-\(\beta\)-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and shaking is continued for 4 hours at 37°C. For expression of Nop56/58, the transformed cell culture is grown at 37°C with shaking to an OD$_{600}$ of 0.5 and then moved to 15°C and shaken for an additional 30 minutes. Nop56/58 expression is induced with the addition of IPTG (0.4 mM) and shaking continued at 15°C for an additional 24 hours. IPTG induces the expression of a genomically encoded bacteriophage T7 RNA polymerase gene under control of the \textit{lacUV5} promoter. The expressed T7 RNA polymerase transcribes the plasmid-encoded core protein genes under control of the pET28a vector’s T7 promoter.

Following recombinant core protein expression, cells are pelleted by centrifugation at 10,000xg for 10 minutes at 4°C and then resuspended in 5 ml of Buffer D (20 mM HEPES, pH 7.0, 100 mM NaCl, 3 mM MgCl$_2$, 20% glycerol (w/v)) per gram of cell paste. A protease inhibitor cocktail (Cocktail Set VII, Calbiochem) at a final concentration of 1% (w/v) and 10 U/ml of Benzonase Nuclease (Novagen) are added to prevent protein degradation and promote nucleic acid degradation, respectively. To lyse cells, the cell suspension is sonicated (Fisher Sonic Dismembrator, Model 150) at maximum power for 30 seconds (x 3) on ice, with 1 minute cooling intervals. The lysate is then mixed by rocking for 1 hour at room temperature to facilitate nucleic acid degradation. Degradation of nucleic acids at this step is crucial for preparation of L7 and Nop56/58 core proteins as contaminating nucleic acids bind these core proteins tightly, which can result in nucleic acid contamination and protein precipitation. Cell lysates are then separated into soluble and insoluble protein fractions by
ultracentrifugation at 38,000xg for 30 minutes at 4°C. The majority of recombinant protein remains in the soluble fraction. Nop56/58 core protein is expressed at lower levels than either L7 or fibrillarin and only 50% of the total Nop56/58 protein is soluble. However, without low temperature expression, nearly all the Nop56/58 protein is found in the insoluble fraction. These soluble fractions can be stored at -80°C for several weeks before continuing with chromatographic enrichment of the individual core proteins.

C. Affinity Chromatographic Isolation of L7 and Fibrillarin Core Proteins

His-tagged L7 (14.7 kDa, 13 kDa without the His tag) and His-tagged fibrillarin (28 kDa) are isolated from soluble protein fractions to greater than 90% purity using a single nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography step (Fig. 1A, lanes 1 and 2). The soluble lysate fraction containing recombinant, His-tagged L7 or fibrillarin protein is applied at room temperature to 4 ml of Buffer D-equilibrated Ni²⁺-charged Ni-NTA His-bind resin (Novagen) suspended in a 20 ml (1.5 x 10 cm) chromatography column. This resin has an estimated binding capacity of 8 mg of protein per ml of resin. The column is then washed with 25 bed volumes of Buffer D 300 (Buffer D with 300 mM NaCl) containing 40 mM imidazole. Glycerol may be omitted from the wash buffer to increase column flow rates. Bound core protein is eluted with 2 bed volumes of Buffer D containing 250 mM imidazole and collected in 1-2 ml fractions. Elution fractions containing the core protein are pooled and dialyzed against 100 volumes of Buffer D overnight at 4°C to remove imidazole. The His-tag may be removed at this point by digesting the isolated core protein with thrombin (~5 units/ml) either prior to or during dialysis. Protein concentrations are estimated by UV absorbance at 280 nm using Beer’s Law and extinction coefficients of 5,240 cm⁻¹ M⁻¹ (calculated) and 29,900 cm⁻¹ M⁻¹ (calculated) for L7 and fibrillarin, respectively. Utilization of dye-based assays such as the Bradford assay to determine protein concentrations are suitable for fibrillarin. However, L7 is not proportionately stained with respect to protein concentration.
FLAG-tagged fibrillarin (28 kDa) is affinity selected from the soluble cell sonicate using ANTI-FLAG M2 affinity agarose (Sigma). Fibrillarin lysate is applied twice at room temperature to 5 ml of ANTI-FLAG M2 agarose resin previously equilibrated in Buffer D and packed in a 20 ml chromatography column (estimated resin binding capacity of 0.5-1.0 mg of fibrillarin per ml of resin). The affinity column with bound fibrillarin is then washed with 25 bed volumes of Buffer D$_{300}$ (300 mM NaCl). FLAG-fibrillarin can be eluted with 3 bed volumes of Buffer D containing 100 µg/ml of FLAG peptide (Sigma). However, an alternative and more economical approach is to elute bound FLAG-fibrillarin with 0.1 M glycine, pH 3.5 containing 100 mM NaCl. The eluted fractions are collected into 1.5 ml microfuge tubes containing 1/10 the elution volume of 1 M Tris, pH 8.0 (the Tris pH 8.0 buffer reestablishes the eluate pH to approximately 7.0). Both elution protocols yield FLAG-tagged fibrillarin that efficiently assembles box C/D sRNPs which are active for sRNP-guided nucleotide methylation. ANTI-FLAG M2 resin has a lower binding capacity than the Ni-NTA resin. Therefore, this chromatography step is typically repeated several times using the same fibrillarin lysate and the eluted fibrillarin fractions are pooled. Pooled fibrillarin fractions are then concentrated 10-15 fold using sequentially a 50 ml Amicon Centricon Concentrator (10,000 MWCO, Millipore) and a 15 ml spin concentrator (Vivaspin 15R, Vivascience) before dialysis against 100 volumes of Buffer D for 16 hours at 4°C. This single affinity selection step yields FLAG-tagged fibrillarin at >85% homogeneity (Fig. 1A, lane 3). FLAG-tagged fibrillarin concentrations are determined either by a Bradford assay or UV absorbance at 280 nm using a calculated extinction coefficient of 31,400 cm$^{-1}$ M$^{-1}$. Isolated L7 and fibrillarin protein preparations can be stored at -80°C for up to one year.

D. Isolation of Nop56/58 Core Protein by Cation-Exchange Chromatography

Purification of recombinant Nop56/58 (48 kDa) via cation-exchange chromatography takes advantage of this core protein’s strong positively charged character. Initial chromatographic isolation of Nop56/58 used SP Sepharose Fast Flow cation-exchange resin (Amersham Biosciences) (8). We now routinely use heparin agarose as this resin more efficiently binds Nop56/58, has a higher binding capacity, and yields a more homogeneous protein
preparation. Affinity-tagged versions of Nop56/58 previously tested in our lab did not bind their affinity resins. Notably however, an N-terminally His-tagged version of Nop56/58 has been successfully purified using Ni-NTA metal affinity chromatography (developed by the Brown laboratory, Wake Forest University) and this affinity-purified Nop56/58 is comparable to protein isolated using cation exchange chromatography (13).

The Nop56/58 lysate is applied twice at room temperature to 10 ml of Buffer D-equilibrated heparin agarose (MP Biomedicals) packed in a 40 ml chromatography column (1.5 x 20 cm). The resin with bound Nop56/58 is then washed with 30 bed volumes of Buffer D$_{800}$ (800 mM NaCl). Bound Nop56/58 is eluted with 3 bed volumes of high salt Buffer D$_{1300}$ (1.3 M NaCl). The Nop56/58 eluate is diluted with Buffer D to a final NaCl concentration of 500 mM and then concentrated to approximately 2 ml using the concentration techniques outlined above for FLAG-tagged fibrillarin. Concentrated Nop56/58 is then centrifuged at 14,000xg at room temperature for 15 minutes to pellet insoluble protein. Nop56/58 is stored in Buffer D$_{500}$ to avoid aggregation and precipitation, although buffers with lower salt concentrations are suitable for lower Nop56/58 concentrations. This single purification step typically yields Nop56/58 protein to approximately 60-70% homogeneity (Fig. 1A, lane 4). A Bradford assay estimates the concentration of isolated Nop56/58 and this core protein can be stored at -80°C for up to one year.

II. Cloning and In Vitro Transcription of Archaeal Box C/D sRNAs

The Methanocaldococcus jannaschii sR8 box C/D sRNA (Fig. 1B) gene was originally PCR-amplified from genomic DNA and then cloned into a pUC19 plasmid (8). In vitro transcription of sR8 sRNA is accomplished by first generating DNA templates from this plasmid via PCR amplification. The upstream PCR primer possesses a T7 promoter sequence (22 nucleotides) at the 5’ terminus followed by the first 22 nucleotides of sR8 coding sequence whereas the downstream primer is complementary to the last 23 nucleotides of sR8 coding sequence. PCR amplification of the sR8 pUC19 plasmid using these primers generates a DNA template for in vitro T7 RNA polymerase transcription. A standard 100 µl
PCR amplification reaction of 35 cycles produces approximately 8-10 µg of DNA template from 50 ng of plasmid DNA. Ampliscribe Flash T7 Transcription Kits (Epicentre) are used to synthesize sRNA transcripts following the manufacturer’s protocol, except that the 37°C incubation is extended to 5 hours. Transcribed RNA is phenol-chloroform extracted, ethanol precipitated, and then resuspended in Tris-Borate-EDTA (TBE) buffer containing 80% formamide. The RNA is resolved on denaturing 6% polyacrylamide-TBE gels containing 7 M urea. RNA bands are visualized by UV shadowing for excision from the gel. sRNA is eluted from the gel slice (x 3) at room temperature for 45-60 minutes with 2 ml of elution buffer (10 mM Tris, pH 7.4, 0.3 M sodium acetate, 5 mM EDTA, 0.1% SDS) per gram of gel using the crush-and-soak method. Eluted RNA is ethanol precipitated, resuspended in water, and the RNA concentration determined by absorbance at 260 nm. RNA is then aliquoted, dried, and stored at -80°C. This in vitro transcription protocol can produce about 50-80 µg of gel-purified sR8 sRNA per µg of DNA template for a standard 20µL transcription reaction.

III. In vitro Assembly of the M. jannaschii sR8 Box C/D sRNP Complex

Core protein binding capabilities and sRNP assembly are assessed using electrophoretic mobility-shift analysis (EMSA) (8). In vitro assembly of the M. jannaschii sR8 box C/D sRNP is accomplished by incubating 5’-radiolabeled sR8 sRNA (0.2 pmol and 1x10⁴ cpm) with 20 pmol of L7, 32 pmol Nop56/58, and 32 pmol fibrillarin in Assembly Buffer (20 mM HEPES, pH 7.0, 150 mM NaCl, 1.5 mM MgCl₂, 10% glycerol) containing tRNA (1.5 mg/mL). Binding capabilities of individual core proteins are determined by sequentially adding L7, Nop56/58, and then fibrillarin as assembly of the complex requires ordered binding of the core proteins. Assembly of the sRNP complex is accomplished by incubating sRNA and core proteins at 75°C for 10 minutes. In vitro assembly of the sRNP requires elevated temperatures to facilitate sRNA remodeling required for core protein binding (14). Partial or completely assembled complexes are resolved on native 4% polyacrylamide gels containing 25 mM potassium phosphate buffer, pH 7.0 and 2% glycerol. After electrophoresis, gels are dried and assembled RNPs visualized by autoradiography or
phosphor-imager analysis. Figure 1C shows a representative EMSA analysis of the sequential binding of sRNP core proteins to radiolabeled sR8 sRNA.

IV. Assessment of In vitro Assembled M. jannaschii Box C/D sRNP Methylation

Methyltransferase activity of the in vitro assembled M. jannaschii sR8 box C/D sRNP is assessed using an in vitro methylation assay. Assembled sRNP complexes are incubated in the presence of the methyl donor S-adenosyl-L-[methyl-³H] methionine (SAM) (Amersham Pharmacia) and synthetic target RNA oligonucleotide substrates (Dharmacon) that are complementary to the D or D’ guide regions. Methylation activity is assessed by measuring the incorporation of [³H]-CH₃ into these target RNAs. Assembly reactions of 80 µl and approximately 0.5 µM assembled sRNP are incubated on ice and mixed with 30 µl of Assembly Buffer containing 12 µM target RNA substrate(s) and 15 µM SAM (5 µCi of [³H]-SAM at a 1:50 ratio with non-radioactive SAM). This SAM concentration is sufficient for this methylation assay although recent work has indicated that higher concentrations of SAM can drive the reaction to yield higher levels of target RNA methylation (15). We have established that the length of target RNA oligonucleotide substrates affects the level of RNA methylation. Extending the target RNA at both 5’ and 3’ termini by 4-5 nucleotides beyond that region which base pairs with the sRNA guide sequence significantly increases [³H]-CH₃ incorporation (16). Negative controls are target RNAs already possessing a 2’-O-CH₃ at the target nucleotide or target RNAs with a deoxynucleotide replacing the target ribonucleotide.

Target RNA methylation is initiated by incubating the assembled reactions at 68°C. Aliquots of 20 µl are removed at the desired time points and spotted onto 2 cm filter discs (3M Whatman paper). After drying, the filters are washed in 10% trichloroacetic acid (TCA) and then 3 times in 5% TCA. Washed and dried filters are suspended in scintillation fluid and counted in a liquid scintillation counter. Results of a typical in vitro methylation activity assay are shown in Figure 1D. Methylation activity of the sRNP is reported as moles of methylated target RNA per mole of sRNP. Conversion of cpm to moles of incorporated [³H]-CH₃ is accomplished by spotting 1 µCi of [³H]-SAM onto control filters, determining
the cpm/µCi, and then calculating the moles of incorporated CH₃ using the specific activity of the [³H]-SAM (Ci/mole) provided by the manufacturer and the molar ratio of radioactive and non-radioactive SAM in the reaction. This value is then reported with respect to the moles of assembled sRNP in the reaction.

V. Sequential Affinity Purification of In Vitro Assembled Box C/D sRNPs

A. An Overview

Affinity purification using tagged proteins or RNAs has proved to be a powerful approach for isolating multi-component protein and RNA-protein complexes from isolated cellular extracts (17-20). In comparison to more traditional fractionation techniques such as gradient sedimentation centrifugation, gel filtration, and ion exchange chromatography, affinity chromatography typically yields highly purified complexes in only one or two isolation steps. Therefore, we have developed a tandem affinity purification protocol for the rapid isolation of in vitro assembled archaeal box C/D sRNPs. This protocol uses three affinity selection steps which are designed to isolate fully assembled and catalytically active sRNP.

Archaeal box C/D sRNPs are assembled using large preparations of in vitro transcribed box C/D sRNA and the three recombinant sRNP core proteins. Assembled sRNPs are first selected using an oligonucleotide complementary to the sRNA. Two approaches can be used in this step. An sR8 sRNA engineered with a poly-A tail (14 adenines) at the 3’ end is used in sRNP assembly and selection is carried out using oligo-dT cellulose resin. Alternatively, a biotinylated DNA oligonucleotide complementary to the sR8 D guide region is hybridized to the assembled sRNP and then selected using streptavidin resin. For both approaches, the sRNA and bound core proteins are efficiently eluted from the respective resins at elevated temperature. We have noted, however, that sRNP affinity selected with the biotinylated oligonucleotide exhibits a 20-30% reduction in methylation activity guided by the terminal box C/D RNP. Subsequently, assembled complexes are sequentially selected via fibrillarin’s FLAG tag and His tag using ANTI-FLAG M2 and Ni-NTA affinity resins, respectively. Fibrillarin is the third and final core protein to bind the box C/D and C’/D’ motifs.
Sequential affinity selection of the *in vitro* assembled sRNP possessing the two tags assures that each isolated complex contains two fibrillarin proteins and is thus a fully assembled sRNP with catalytically active box C/D and C’/D’ RNPs. This is particularly important for biophysical analyses such as cryo-electron microscopy or X-ray crystallography where a homogeneous population of complexes is crucial for analysis.

Selection of the sRNP using complementary oligonucleotides as the first selection step effectively eliminates free core proteins not bound to the sRNA. Removal of free proteins is particularly advantageous at this point in the isolation protocol. The highly charged character of free Nop56/58 can cause aggregation problems. Protein aggregation may be minimized by using higher salt buffers and working in the presence of very low concentrations of SDS (~0.007%), non-ionic detergents (0.1% Triton X-100), or non-specific RNA (1-2 mg/ml tRNA). These added components can increase sRNP yields although small amounts of detergent may be carried through the purification process and be present in the final purified sRNP fraction. Nop56/58 and fibrillarin also efficiently dimerize in the absence of sRNA and sequential isolation of the sRNP complexes solely via the tagged fibrillarin core proteins results in isolation not only of the sRNP but also free Nop56/58-fibrillarin dimers. ANTI-FLAG M2 affinity chromatography is the second step in sRNP purification as this resin exhibits low binding capacity and low elution efficiency. By placing this affinity step second in the purification protocol, we are able to use larger amounts of resin and elute the sRNP in larger elution volumes for more efficient recovery. As the final selection step, the Ni-NTA resin binds the His-tagged fibrillarin with high affinity. The large capacity of this resin facilitates elution of more concentrated sRNP in smaller final volumes. We have found that this order of affinity selection steps is most efficient for purification of fully assembled and methylation-competent sRNP.

A flowchart for the tandem affinity selection and purification of *in vitro* assembled archaeal box C/D sRNP complexes is presented in Figure 2. The starting amounts of assembled sRNP may be reduced or scaled up depending upon the quantity of purified sRNP desired.
For the particular sRNP purification experiment shown here, additional components such as detergents were omitted. sRNP obtained from the final Ni-NTA affinity column can be exchanged with any buffer of choice during concentration depending upon the requirements of the planned experiments, although buffers of higher ionic strength help to reduce sRNP aggregation. sRNP fractions applied to each resin, collected flow through fractions, and subsequently eluted sRNP fractions are analyzed on an SDS-polyacrylamide gel and shown in Figure 3A.

B. Affinity Chromatography Buffers

10X Binding Buffer (BB): 100 mM HEPES, pH 7.0, 1.0 M NaCl
Buffer R: 20 mM HEPES, pH 7.0, 0.1 M NaCl, 1 mM EDTA
Buffer D: 20 mM HEPES, pH 7.0, 3 mM MgCl2, 100 mM NaCl, 20% glycerol
Buffer E: 20 mM HEPES, pH 7.0, 1.5 mM MgCl2, 500 mM NaCl
Buffer G: 0.1 M glycine, pH 3.5, 100 mM NaCl

C. Preparation of the Affinity Chromatography Resins

1. Oligo-dT cellulose (Ambion): Equilibrate 4 ml of resin packed in a 20 ml (1.5 cm x 10 cm) chromatography column with Buffer E.

1A. Streptavidin agarose (Novagen): Equilibrate 3 ml of resin packed in a 20 ml (1.5 cm x 10 cm) chromatography column with Buffer D. Bind 3 mg of biotinylated DNA oligonucleotide (5’-(biotin)-ACAGTCATCGCTTGCTCATACGGTTCCTC-3’) (Integrated DNA Technologies) complementary to the D guide sequence (underlined) by incubating the oligonucleotide and resin for 10 minutes at room temperature. Then equilibrate the resin with Buffer E.

2. ANTI-FLAG M2 agarose (Sigma): Equilibrate 3.0 ml of resin packed in a 20 ml (1.5 cm x 10 cm) chromatography column with Buffer E.

3. Ni-NTA His-Bind Resin (Novagen): Charge 1.5 ml of resin packed in a 20 mL chromatography column (1.5 cm x 10 cm) with 5 ml of 50 mM NiSO4. Then equilibrate the resin with Buffer E containing 25 mM imidazole.
D. Assembly of the sR8 sRNP Complex

sRNP are assembled \textit{in vitro} using \textit{in vitro} transcribed sRNA and recombinant core proteins with the amount of assembled sRNP dependant upon the desired amount of purified complex. Table 1 lists the components required for assembling 12 nmol (~2.5 mg) of sRNP complex. Approximately 12 nmol (320 ug) of sR8 sRNA, 24 nmol of L7 (325 ug), 24 nmol of Nop56/58 (1200 ug), and 12 nmol (350 ug) each of FLAG-tagged and His-tagged fibrillarin are incubated in a final assembly volume of 2.5 ml. Core proteins are stored in Buffer D except for Nop56/58 which is stored in Buffer D containing 0.5 M NaCl to maintain protein solubility. The final assembly buffer (20 mM HEPES, pH 7.0, 0.5 M NaCl, 1.5 mM MgCl₂, and 10% glycerol (w/v)) contains 0.5 M NaCl to minimize protein and sRNP aggregation. After mixing components, the assembly reaction is heated to 75°C for 5 minutes and then cooled to room temperature. Centrifugation of the assembly reaction at 14,000xg for 5 minutes at room temperature removes insoluble materials from the supernatant fraction containing assembled sRNP. Shown in Figure 3A are the sRNP fractions obtained from the sequential affinity selection steps.

E. Tandem Affinity Purification of \textit{In Vitro} Assembled Box C/D sRNP

\textit{Affinity Selection Step 1: Oligo dT Cellulose Chromatography}

1. Mix the assembled sRNP (~12 nmol in 2.5 ml; Fig. 3A, lane 2) with 4 ml of oligo dT cellulose resin equilibrated with Buffer E in a 20 ml chromatography column.

2. Bind the sRNP to the oligo-dT resin by rocking the column for five minutes at room temperature and then continue rocking for 30 minutes at 4°C.

3. Begin chromatography by collecting the flow through fraction (Fig. 3A, lane 3) and then washing the oligo dT resin with 20 bed volumes of cold (4°C) Buffer E. Maintain the column at 4°C during washing to stabilize hydrogen bonding of the poly (A) tail to the oligo dT cellulose.

4. Transfer the column with bound sRNP to a 50°C incubator and heat the column and resin for 20 minutes.

5. Elute the sRNP with 4 bed volumes of Buffer E heated to 50°C (Fig. 3A, lane 4).
6. Cool the sRNP eluate to room temperature for subsequent affinity selection on ANTI-FLAG M2 agarose.

7. Regenerate the oligo dT cellulose resin by washing sequentially with 5 bed volumes each of Buffer R, 0.1 M NaOH, and H₂O.

8. Equilibrate the resin with Buffer R containing 0.05% sodium azide and store at 4°C. (For long term storage, after the water wash rinse the resin with ethanol, dry the resin, and store at -20°C.)

*Alternative Affinity Selection Step 1A: Streptavidin Affinity Chromatography*

1. Mix the assembled sRNP with 3 ml of Buffer E-equilibrated streptavidin resin bound with the biotinylated DNA oligonucleotide and suspend in a 20 ml chromatography column.

2. Bind the sRNP to the streptavidin resin by rocking the column for 5 minutes at room temperature and then continue rocking for 30 minutes at 4°C.

3. Begin chromatography by collecting the flow through fraction and then washing the resin with 20 bed volumes of cold Buffer E. Maintain the column with resin at 4°C during washing to stabilize hybridization of the biotinylated DNA oligonucleotide to the sRNA.

4. Transfer the affinity column with resin-bound sRNP to a 60°C incubator and equilibrate the column for 20 minutes at this elevated temperature.

5. Elute bound sRNP with 4 bed volumes of Buffer E heated to 60°C.

6. Cool the eluted sRNP fraction to room temperature for subsequent affinity selection on ANTI-FLAG M2 agarose.

7. Regenerate the DNA oligonucleotide-streptavidin resin by washing with 6 bed volumes of Buffer R heated to 60°C.

8. Equilibrate the resin in Buffer R containing 0.05% sodium azide and store at 4°C.
Affinity Selection Step 2: ANTI-FLAG M2 Affinity Chromatography

1. Mix the cooled sRNP eluate (12-16 ml) with 3 ml of Buffer E-equilibrated ANTI-FLAG M2 agarose suspended in a 20 ml chromatography column.
2. Bind the sRNP to the agarose resin by incubating for 15 minutes at room temperature.
3. Collect the flow through (Fig. 3A, lane 5) and reapply this eluate to the affinity column twice more, each time incubating the sRNP eluate with the resin for 15 minutes at room temperature.
4. Wash the agarose resin with 20 bed volumes of Buffer E at room temperature.
5. Release the bound sRNP by resuspending the agarose resin in one bed volume of Buffer E containing 110 µg/ml of FLAG peptide (Sigma) and incubate for 10 minutes at room temperature.
6. Collect the column eluate containing the sRNP.
7. Resuspend the agarose resin twice more in one bed volume of Buffer E with FLAG peptide, each time incubating the resin for 10 minutes at room temperature before collecting the eluate.
8. Pool all three sRNP eluate fractions (Fig. 3A, lane 6) for subsequent Ni-NTA affinity chromatography.
9. Regenerate the agarose resin by washing twice with 4 bed volumes of Buffer G. Do not allow this resin to be suspended in Buffer G for more than 20 minutes. Wash the resin with Buffer R.
10. Equilibrate the agarose resin in Buffer R containing 0.05% sodium azide and store at 4°C.

Affinity Selection Step 3: Ni-NTA His Bind Resin Affinity Chromatography

1. Mix the pooled eluates from ANTI-FLAG M2 affinity chromatography with 1.5 ml of Buffer E-equilibrated Ni-NTA His-Bind resin suspended in a 20 ml chromatography column and incubate for 10 minutes at room temperature.
2. Collect the column flow through fraction (Fig. 3A, lane 7) and reapply the eluate to the resin, incubating another 10 minutes at room temperature.
3. Wash the affinity resin with 20 bed volumes of Buffer E containing 25 mM imidazole.
4. Release the bound sRNP by resuspending the resin in 2 bed volumes of Buffer E containing 200 mM imidazole and incubating for 5 minutes at room temperature.
5. Collect the eluate fraction.
6. Resuspend the resin in an additional 2 bed volumes of Buffer E containing 200 mM imidazole and incubate for an additional 5 minutes at room temperature.
7. Collect the second eluate fraction.
8. Pool the two eluates (this pooled eluate may be stored overnight at 4°C).
9. Concentrate the pooled eluates to between 200 and 500 µl using a 15 ml spin concentrator (Vivaspin 15R, Vivasience) and a microspin concentrator (Microcon YM-3, Millipore). During concentration, an exchange of buffers to the desired final buffer may be accomplished. Dialysis is not recommended at this step.
10. Regenerate the Ni-NTA His-Bind resin by washing with 5 bed volumes of 100 mM EDTA, pH 7.0 followed by washing with water.
11. Equilibrate the Ni-NTA His Bind resin in 20% ethanol and store at 4°C.

The yield of purified sRNP is estimated by assessing the amount of RNA and/or protein contained in a small aliquot of complex resolved on polyacrylamide gels. Alternatively, the amount of RNA in a phenol-extracted aliquot of sRNP is determined by absorbance at 260 nm. Typically 200-300 pmol (40-60 ug) of purified sRNP is obtained from approximately 2.5 mg of assembled complex. Typical losses are observed for each affinity step with major losses being at the second and third selection step for the FLAG-tagged and His-tagged fibrillarin proteins. At these steps, significant amounts of sRNP are lost as unselected complexes as these sRNP are assembled with similarly tagged but unselected fibrillarin proteins at both box C/D and C’/D’ RNPs. Assessment of box C/D RNP-guided nucleotide methylation activity of the purified complexes reveals methylation of target RNAs for both box C/D and C’/D’ RNPs comparable to that of unpurified sRNPs (Fig. 3B). Determination of sRNP methylation capabilities for complexes suspended in Buffer E with elevated NaCl
concentration has no discernable effect upon the complex’s enzymatic activities. Electron microscopy of isolated sRNP reveals sRNP of homogeneous size with some larger complexes (Fig. 3C). The larger complexes are aggregated sRNP and their presence can be diminished with the addition of ionic (SDS, heparin) compounds, although elevated concentrations can destabilize the sRNP complex.

VI. Concluding Remarks

*In vitro* assembled archaeal sRNPs provide a model complex for the investigation of box C/D RNP assembly, structure, and function. Study of these minimal yet catalytically active RNA:protein enzymes will help define the fundamental principles behind RNA-guided nucleotide modification. The archaeal sRNP also serves as a prototype box C/D RNA-guided nucleotide modification enzyme for understanding the more structurally and functionally complex eukaryotic snoRNPs. Affinity purification of completely assembled and catalytically active sRNPs can now provide not only complexes for more detailed biochemical and functional studies but also for various biophysical approaches requiring larger amounts of material. Of particular advantage is the fact that the isolated complexes are homogeneous in composition, a prerequisite for approaches such as cryo-electron microscopy and X-ray crystallography.

The sequential affinity approach described here may well have broader applications in the study of RNA-guided nucleotide modification complexes. This same isolation protocol should be easily adapted to isolate other box C/D complexes and, in principle, *in vitro* assembled H/ACA sRNPs. Perhaps more exciting is the possibility to use this affinity protocol for the isolation of *in vivo* assembled RNP complexes. Expression of tagged core proteins in various cell lines should make affinity selection of the corresponding RNP complexes from cell lysates or various cellular fractions possible. This could ultimately lead to the identification of additional or “accessory” proteins associated with a family of complexes, thus defining novel proteins important for RNP biogenesis and/or function. Also possible may be the selection of specific RNP complexes from a homogenous RNP
population using oligonucleotides complementary to a given sRNA or snoRNA species. Again, such an approach could lead to the identification of accessory proteins unique to a specific RNP complex. While these suggested approaches have yet to be tested, they have the potential to greatly facilitate more detailed examinations of the diverse and highly conserved populations of RNA-guided nucleotide modification complexes.
ACKNOWLEDGEMENTS

We thank James Brown for providing *Methanocaldococcus jannaschii* genomic DNA, and Franziska Bleichert for critical reading of the manuscript. We are grateful to Franziska Bleichert, Vinzenz Unger and Susan Baserga (Yale University) for providing electron micrographs of affinity purified *M. jannaschii* sR8 box C/D sRNPs. This work was supported by NSF grant MCB 0543741 (ESM).
REFERENCES


### TABLE I

*In Vitro sRNP Assembly*

<table>
<thead>
<tr>
<th>Reaction Component</th>
<th>Concentration</th>
<th>Volume (μL)</th>
<th>Approx. Mass (μg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L7</td>
<td>100 μM</td>
<td>250</td>
<td>325</td>
</tr>
<tr>
<td>Nop56/58</td>
<td>100 μM</td>
<td>250</td>
<td>1200</td>
</tr>
<tr>
<td>His-fibrillarin</td>
<td>100 μM</td>
<td>125</td>
<td>350</td>
</tr>
<tr>
<td>FLAG-fibrillarin</td>
<td>100 μM</td>
<td>125</td>
<td>350</td>
</tr>
<tr>
<td>Buffer D*</td>
<td>1X</td>
<td>500</td>
<td>---</td>
</tr>
<tr>
<td>poly-(A) sRNA</td>
<td>38 μM</td>
<td>320</td>
<td>320</td>
</tr>
<tr>
<td>NaCl</td>
<td>5 M</td>
<td>155</td>
<td>---</td>
</tr>
<tr>
<td>Binding Buffer**</td>
<td>10X</td>
<td>250</td>
<td>---</td>
</tr>
<tr>
<td>H2O</td>
<td>---</td>
<td>525</td>
<td>---</td>
</tr>
<tr>
<td><strong>Final</strong></td>
<td><strong>2500</strong></td>
<td><strong>2545</strong></td>
<td></td>
</tr>
</tbody>
</table>

*20 mM HEPES, pH 7.0, 0.1 M NaCl, 3 mM MgCl2, 20% (w/v) glycerol.

**100 mM HEPES, pH 7.0, 1.0 M NaCl.*
FIGURE 1: Archaeal box C/D recombinant core protein isolation, in vitro RNP assembly, and sRNP-guided nucleotide 2’-O-methylation. (A) Coomassie Brilliant Blue stained SDS-polyacrylamide gel of isolated M. jannaschii box C/D sRNP core proteins used for in vitro sRNP assembly. (B) Sequence and folded secondary structure of the M. jannaschii sR8 box C/D sRNA used for in vitro sRNP assembly. (C) Electrophoretic mobility-shift analysis revealing the hierarchical binding of the sRNP core proteins to 5’-radiolabeled sR8 sRNA in assembly of the box C/D sRNP. (D) In vitro methylation of D and D’ target RNAs by the in vitro assembled sR8 box C/D sRNP. The lack of methylation at the target nucleotide when this nucleotide is already methylated at the ribose 2’ position (D-CH$_3$ and D’-CH$_3$ targets) demonstrates nucleotide-specific modification of each target RNA by the in vitro assembled sRNP.
FIGURE 2: Sequential affinity purification of in vitro assembled Archaeal box C/D sRNP.
FIGURE 3: SDS-polyacrylamide gel electrophoretic analysis, methyltransferase activity, and electron microscopy of in vitro assembled M. jannaschii sR8 box C/D sRNPs. (A) SDS-polyacrylamide gel electrophoretic analysis of sRNP fractions. Silver-stained sRNP fractions analyzed from the individual affinity columns include: the sRNP sample applied (A); the flow through fraction of unbound material (FT); affinity column-bound and eluted sRNP fraction (E). Percentage of the total sample volume used for each electrophoretic analysis is indicated above the lane. (B) Assessment of the methyltransferase activity of tandem affinity purified box C/D sRNP (TAP) compared with in vitro assembled but not affinity selected complexes. Both affinity selected and non-purified sR8 box C/D sRNP methylate D and D' targets indicating fully assembled and catalytically active box C/D and C'/D' RNPs. D and D' targets possessing a 2'-O-CH$_3$ at the target nucleotide (negative controls) are not methylated demonstrating nucleotide-specific modification for both affinity purified and non-purified complexes. (C) Electron micrograph of in vitro assembled and tandem affinity purified M. jannaschii sR8 box C/D sRNP. Electron micrograph courtesy of Franziska Bleichert, Vinzenz Unger and Susan Baserga (Yale University).
CHAPTER II

Assembly of the Archaeal Box C/D sRNP Can Occur Via Alternative Pathways and Requires Temperature-Facilitated sRNA Remodeling

Keith T. Gagnon, Xinxin Zhang, Paul F. Agris and E. Stuart Maxwell

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

Archaeal dual-guide box C/D sRNAs bind three core proteins in sequential order at both terminal box C/D and internal C’/D’ motifs to assemble two RNP complexes active in guiding nucleotide methylation. Experiments have investigated the process of box C/D sRNP assembly and the resultant changes in sRNA structure or ‘remodeling’ as a consequence of sRNP core protein binding. Hierarchical assembly of the *Methanocaldococcus jannaschii* sR8 box C/D sRNP is a temperature-dependent process with binding of L7 and Nop56/58 core proteins to the sRNA requiring elevated temperature to facilitate necessary RNA structural dynamics. Circular dichroism (CD) spectroscopy and RNA thermal denaturation revealed an increased order and stability of sRNA folded structure as a result of L7 binding. Subsequent binding of the Nop56/58 and fibrillarin core proteins to the L7-sRNA complex further remodeled sRNA structure. Assessment of sR8 guide region accessibility using complementary RNA oligonucleotide probes revealed significant changes in guide region structure during sRNP assembly. A second dual-guide box C/D sRNA from *M. jannaschii*, sR6, also exhibited RNA remodeling during temperature-dependent sRNP assembly, although core protein binding was affected by sR6’s distinct folded structure. Interestingly, the sR6 sRNP followed an alternative assembly pathway, with both guide regions being continuously exposed during sRNP assembly. Further experiments using sR8 mutants possessing alternative guide regions demonstrated that sRNA folded structure induced by specific guide sequences impacted the sRNP assembly pathway. Nevertheless, assembled sRNPs were active for sRNA-guided methylation independent of the pathway followed. Thus, RNA remodeling appears to be a common and requisite feature of archaeal dual-guide box C/D sRNP assembly and formation of the mature sRNP can follow different assembly pathways in generating catalytically active complexes.
INTRODUCTION

Ribonucleotide 2’-O-methylation of both eukaryotic and archaeal RNAs is accomplished using box C/D guide RNAs (1-8). Box C/D RNAs are small species of between 50 and 150 nucleotides defined by highly conserved boxes C and D. These terminally-located box C and D sequences of 6 and 4 nucleotides, respectively, define all box C/D RNAs and are essential for folding the box C/D core motif into characteristic kink-turn, or K-turn, structures (5,9-12). The K-turn is a helix-asymmetric bulge-helix RNA motif where two G-A base pairs hydrogen bond across the asymmetric bulge to stabilize the characteristic kinked structure of this RNA motif. Box C/D RNAs often possess internally positioned and highly conserved C’ and D’ boxes that are particularly evident in archaeal box C/D RNAs (10,13-15). The C’/D’ motif also folds into a K-turn structure. Nucleotide methylation is carried out by base-pairing of guide sequences located upstream of box D and D’ to complementary regions of the target RNA (1,2,5,16). Core proteins bound to the box C/D and C’/D’ motifs then catalyze 2’-O-methylation of the fifth base-paired nucleotide positioned upstream from either box D or D’ (N+5 rule) (1,17-20).

In eukaryotes, four core proteins bind the box C/D small nucleolar RNAs (snoRNAs) and are critical for methylation activity. The 15.5kDa core protein initiates snoRNP assembly by recognizing the terminal box C/D core motif and stabilizing this RNA element into a K-turn structure (12,21). Two highly related nucleolar proteins, Nop58 and Nop56, differentially bind the box C/D and C’/D’ motifs, respectively, while the methyltransferase enzyme fibrillarin interacts with both motifs (22-26). The eukaryotic snoRNP is said to be asymmetric with 15.5kDa, Nop58, and fibrillarin binding to the terminal box C/D motif and core proteins Nop56 and fibrillarin binding the internal C’/D’ motif (26,27).

Box C/D snoRNA-like RNAs, or sRNAs, are abundant in archaeal organisms and guide nucleotide methylation of rRNAs and tRNAs (14,28). In vitro assembly of catalytically active archaeal box C/D sRNPs has recently been accomplished and used to structurally and
functionally characterize sRNP complexes (29-33). Archaeal box C/D sRNAs possess characteristic terminal box C/D and internal C'/D’ motifs, each bound by three core proteins which are highly homologous to the eukaryotic box C/D snoRNP core proteins. Ribosomal protein L7, a eukaryotic 15.5kDa homolog, initiates sRNP assembly by binding both box C/D and C'/D’ motifs (21,29,30,34). Subsequently, a single homolog for eukaryotic Nop56 and Nop58, Nop56/58 (Nop5), binds both motifs and is followed by fibrillarin binding (29,30). The three archaeal core proteins exhibit a sequential binding order at both terminal C/D and internal C'/D’ K-turns to establish ‘symmetric’ RNPs of identical protein composition. Optimal methylation activity requires that both box C/D and C'/D’ RNPs be juxtaposed and spatially positioned within the same sRNP complex (30,31,35). Incomplete assembly caused by mutation, loss of protein components at either the C/D or C'/D’ motif, or alterations in inter-RNP spacing, results in lost or reduced methylation activity from both RNPs. Unfortunately, a complete understanding of box C/D RNP assembly is presently lacking and the nature of inter-RNP interactions is still undefined (15,36).

Mounting evidence demonstrating that the L7 core protein strongly stabilizes the box C/D K-turn by inducing RNA structural changes has suggested that sRNP assembly requires sRNA remodeling(37-41). Crystal structures of both free and RNA-bound L7 from several archaeal organisms, as well as FRET analyses, have revealed a mechanism of induced-fit interaction for the RNA component with only minor changes in L7 protein structure upon RNA binding. To further examine this RNA remodeling phenomenon, we have investigated RNA and RNP structural changes induced during assembly of functional archaeal dual-guide box C/D sRNPs. sRNP assembly and concurrent sRNA structural changes were found to be temperature-dependent, indicating that RNA flexibility and remodeling of RNA folded structure are critical for efficient and ordered sRNP assembly. Observed alterations in guide region structure and accessibility during sRNP assembly further reflected RNA remodeling events. Comparison of box C/D sRNAs with different guide regions demonstrated that RNA remodeling can follow alternative pathways during assembly of functional sRNPs. These
observations provide another example of RNA remodeling upon protein binding, which is likely a central theme in the assembly of most, if not all, RNP complexes.
RESULTS

A. Efficient L7 Core Protein Binding is Temperature-Dependent and Requires sRNA Remodeling

Box C/D sRNA structural changes induced by core protein binding during sRNP assembly were investigated using our established in vitro sRNP assembly system (30). Recombinant Methanocaldococcus jannaschii core proteins L7, Nop56/58, and fibrillarin bind M. jannaschii sR8 sRNA in vitro to assemble a dual-guide box C/D sRNP complex that is catalytically active in guiding nucleotide 2’-O-methylation from both terminal box C/D core and internal C’/D’ RNPs. Elevated temperatures of approximately 70°C are required for efficient box C/D sRNP assembly and RNA-guided nucleotide modification, consistent with M. jannaschii being a thermophilic organism. M. jannaschii box C/D sRNA sR8 (Fig. 1A) was chosen as a model dual-guide box C/D sRNA, having perfect consensus C, C’, D, and D’ boxes with D and D’ guide regions of 12 nucleotides each. Box C/D and C’/D’ motifs fold into K-turn elements and their inter-motif spacing of 12 nucleotides is highly conserved in Archaea (13,35).

L7 initiates sRNP assembly at both box C/D core and C’/D’ motifs with an L7 core protein binding each motif in a cooperative manner (30). Titration of L7 with radiolabeled sR8 at the optimal binding temperature of 70°C revealed the formation of two RNP complexes as L7 concentration increased, consistent with L7 binding to one and then the second RNA motif (Fig. 1B). Strikingly, elevated temperature was required for simultaneous L7 binding to both RNA motifs. At low temperature of 4°C, only a single L7 protein was bound (Fig. 1B). This phenomenon was dependent upon having both RNA motifs contained in a single RNA molecule. sR8 “halfmer” RNAs possessing only box C/D or C’/D’ motifs (Figs. 1C and 1E) bound a single L7 protein at both low and elevated temperatures (Figs. 1D and 1F). Since the L7 protein does not undergo significant structural changes during K-turn binding (40), we concluded that elevated temperature was necessary to enhance the flexibility and
conformational freedom of the sRNA, thus facilitating L7 binding through RNA “remodeling” of both motifs juxtaposed in the same sRNA molecule.

B. L7 Binding Remodels and Stabilizes sR8 sRNA Structure

The effect of L7 core protein binding upon sR8 sRNA structure was assessed using circular dichroism (CD) spectroscopy, thermal denaturation analysis, and nuclease mapping. CD analysis revealed that L7 binding at low temperature (4°C) increased the amplitude of the positive ellipticity at 260 nm (Δε_{260}) indicating a more ordered or structured sRNA (Fig. 2A). Protein spectra at this wavelength was negligible and was baseline subtracted so as not to interfere with data collection and interpretation. Thus, increased ellipticity reflected enhanced sRNA base-base interactions and base stacking. This increase in ordered structure was further enhanced when L7 binding was carried out at elevated temperature (70°C) where L7 is bound to both the box C/D and C’/D’ motifs. The two different CD curves obtained at different assembly temperatures suggest that different sRNA structures exist when sR8 is bound at 4°C or 70°C with one or two L7 molecules, respectively. No difference in CD spectra was observed at these two temperatures for sRNA alone, clearly demonstrating that the observed increase was dependent upon L7 binding.

Thermal denaturation analysis of the sR8 sRNA in the presence of increasing amounts of L7 protein revealed a stabilization of RNA structure upon protein binding (Fig. 2B). An increase in T_m of 17.3 ± 0.4°C was observed, with a corresponding decrease in free energy calculated at 4.5 ± 0.6 kcal/mol (Meltwin 3.5), when sR8 was bound by two L7 molecules. Plotting the change in T_m as a function of L7 concentration yielded a sigmoidal curve, which, when plotted using the Hill equation, gave a Hill coefficient of 1.7 indicating strong cooperativity (Fig. 2C, inset). This is consistent with previous observations (30). These results indicate that sR8 conformational changes contribute to the observed cooperativity of L7 core protein binding.
To further investigate sRNA remodeling induced by L7 binding, nuclease mapping with a single strand-sensitive nuclease, mung bean nuclease, was performed at 4°C after binding L7 at either 4°C or 70°C. The low nuclease digestion temperature was necessary to ensure that the RNP assembled at low temperature bound only one L7 protein. Limited digestion of unbound 5’-radiolabeled sR8 at 4°C yielded a pattern of nuclease cleavage sites throughout the sRNA (Fig. 2C, lane 4). Binding of L7 at 4°C significantly altered this digestion pattern in boxes D, D’ and C’ (lane 5). Reduction in nuclease cleavage could be due to protection by protein binding or induced RNA structural changes. However, enhancement of cleavage, such as observed in the D guide region and in box D’ (U30), clearly demonstrated induced RNA structural changes. Unfortunately, these results did not clearly indicate whether L7 was uniquely binding the box C/D or C’/D’ motif. This may suggest that either motif can be bound at this temperature but not both. Nuclease mapping of L7 titration onto sR8 at elevated temperature has also been inconclusive in determining an order of L7 binding (data not shown). Comparison of nuclease digestion when L7 was bound at 4°C versus 70°C revealed unique cleavage sites for each L7 binding temperature. Most distinct was a large reduction in cleavage of the D guide region when L7 was bound at 70°C (lane 6). Thus, the final sR8 folded structure is dependent upon L7 binding temperature and full occupancy of both RNA motifs.

C. Ordered Assembly of the sR8 sRNP Is Temperature-Dependent and Requires sRNA Structural Changes

Assembly of box C/D sRNPs was characterized by sequential binding of the core proteins, with L7 binding followed by Nop56/58 and fibrillarin binding. Analysis of the sR8 assembly pathway revealed that, as with L7, binding of Nop56/58 was also a temperature-dependent process that required changes in sR8 sRNA structure and/or dynamics. As shown previously, L7 binding to both motifs to initiate sRNP assembly required elevated temperature (Fig. 3A, lane 2 vs. lane 6). When only one L7 was bound to sR8 (low temperature), neither Nop56/58 nor fibrillarin were able to bind (Fig. 3A, lanes 3 and 4). Notably, when L7 was bound to box C/D and C’/D’ motifs, Nop56/58 still required elevated temperature for binding (Fig 3A,
lanes 7 vs. lane 11). Prior Nop56/58 dimerization with fibrillarin did not affect Nop56/58's requirement for elevated temperature to bind (data not shown). These results indicated that Nop56/58 binding required RNA conformational flexibility or structural remodeling. Prior heating of Nop56/58 also did not facilitate its binding, implying that necessary structural changes promoted by elevated temperature occur in the sRNA and not the protein (data not shown). Once Nop56/58 was bound, fibrillarin was able to bind independent of temperature (Fig. 3A, lanes 12 and 16). This is consistent with previous work demonstrating that Nop56/58 and fibrillarin dimerize in the absence of sRNA, suggesting that fibrillarin binding occurs primarily through protein-protein contacts (29,30,42).

To assess alterations in sRNA structure during full sRNP assembly, both CD and thermal denaturation experiments were carried out. In these experiments, Nop56/58 tended to precipitate at elevated temperature unless complexed with fibrillarin. Therefore, an assessment of their individual contributions for altering sRNA structure was not possible. CD experiments revealed a pronounced increase in peak amplitude at 260 nm (Δε260) upon Nop56/58-fibrillarin binding, most likely reflecting increased order and base stacking of individual sR8 nucleotides in an already well folded sRNA (Fig. 3B) (43-45). Analysis of sR8’s thermal melting profile revealed that after strong stabilization of sR8 structure by L7 binding, only slight alterations in the profile were observed with Nop56/58-fibrillarin binding (Fig. 3C). From these experiments, we concluded that the initial binding of L7 dramatically stabilizes sRNA structure while subsequent binding of the Nop56/58-fibrillarin dimer primarily affects nucleotide order and base-stacking.

sRNA and sRNP structure were also investigated during full sRNP (70°C) assembly with nuclease mapping (Fig. 3D). The conserved G and A nucleotides of boxes C (G12, A13) and D (G58, A59), as well as the protruded uridine and adjacent adenine of box C (U11, A10), displayed a cleavage pattern consistent with their exposed locations in the asymmetric bulge region of the box C/D core motif (41,46). Cleavage sites were also observed in boxes C’ and D’ as well as both guide regions (Fig. 3D, lane 3). Consistent with previous observations,
binding of L7 to sR8 resulted in the protection of boxes C, D, and C’, enhancement of box D’ cleavage, and small changes in the D guide region (Fig. 3D, lane 4). Subsequent binding of Nop56/58 protected box D’, slightly altered D guide region structure, and increased protection at boxes C’ and C (lane 5). Final binding of fibrillarin generated only minor changes in the digestion pattern (lane 6). Similar box C/D RNA structural changes have also been reported during assembly of a box C/D RNA embedded in the pre-tRNA \textsuperscript{Trp} RNA from \textit{Haloferax volcanii} and \textit{Pyrococcus abyssi} using Pb\textsuperscript{2+}-induced RNA cleavage (32). These cleavage patterns reflect not only protection from digestion as core proteins bind, but also changes in sRNA structure or remodeling as sRNP assembly proceeds.

D. sR8 sRNA Remodeling Modulates D and D’ Guide Sequence Accessibility for Target RNA Binding

Small changes in sR8 guide sequence structures seen in nuclease mapping experiments prompted us to explore their accessibility for target RNA binding during sRNP assembly. Short RNA oligonucleotides complementary to the sR8 D and D’ guide sequences (target RNA probes) were tested for their ability to base-pair to the guide sequences during sRNP assembly. RNA duplex formation was monitored by Pb\textsuperscript{2+} probing, where the sRNA is preferentially cleaved in flexible and/or single stranded regions (Fig. 4A). The utilization of Pb\textsuperscript{2+} as an RNA structural probe produced somewhat different digestion patterns than mung bean nuclease. This is due to the small size of the metal ion and its ability to examine in more detail the fine structure of the folded sRNA and its guide regions. Cleavage results demonstrate that the D, but not D’, guide region was accessible for base pairing with its complementary RNA probe as evidenced by D guide region protection from Pb\textsuperscript{2+} cleavage when the D target RNA was present (Fig. 4A, lanes 4-7). L7 binding to sR8 induced a characteristic cleavage pattern (both enhanced and protected nucleotides) in both D and D’ guide regions, indicating significant sRNA structural changes. Particularly prominent was a strong cleavage at nucleotide C54 of the D guide region (Fig. 4A, lane 8). Addition of D and D’ target RNA probes did not significantly alter this digestion pattern, indicating that the binding of the L7 core protein inhibited target RNA binding (Fig. 4A, lanes 9-11).
Electrophoretic mobility-shift analysis (EMSA) of radiolabeled target RNA binding to sR8 in the absence and presence of L7 at low temperature (4°C) further suggested that L7 causes structural changes during binding that are not conducive to target RNA annealing since L7 was unable to bind sR8 when the D target RNA was basepaired first (Fig. 4B). The converse was also true. Radiolabeled D target RNA incubated with sR8 already bound by L7 was unable to bind its sRNA guide sequence (data not shown). This data indicates that L7 binding to sR8 requires RNA structural dynamics and guide region remodeling. The D’ target did not bind sR8 independent of whether the sRNA was or was not bound with L7, suggesting that the D’ guide sequence is locally folded to prevent target RNA probe base pairing.

Binding of Nop56/58 and fibrillarin core proteins further altered Pb$^{2+}$ cleavage patterns of the D and D’ guide sequences (Fig. 4A, lane 12). Addition of D and D’ target RNA probes protected the D and D’ guide sequences, respectively, showing that both guide regions were accessible for base pairing upon full sRNP assembly (Fig. 4A, lanes 13-15). EMSA of radiolabeled target RNA binding to sR8 at each sRNP assembly step has supported the conclusions of these mapping experiments (data not shown). Also, the use of target RNA probes methylated at the target nucleotide showed no difference in target RNA binding (data not shown). Collectively, these analyses demonstrated that the D and D’ guide sequences undergo distinct structural changes during sR8 sRNP assembly. Notably, both guide regions become fully accessible to target RNA binding only upon complete sRNP assembly with all three core proteins.

E. Temperature-Dependent Assembly of *M. jannaschii* sR6 sRNP Requires sRNA Remodeling But Uses an Alternative Assembly Pathway

To determine if temperature-dependent sRNP assembly and sRNA remodeling are general features of archaeal box C/D sRNP biosynthesis, a second box C/D sRNA from *M. jannaschii*, sR6, was selected for analysis. sR6 is also a dual-guide sRNA possessing box C/D and C’/D’ motifs separated by inter-motif spacing distances of 12 nucleotides (Fig. 5A).
The distinguishing structural difference between sR6 and sR8 is the D and D’ guide sequences, which are complementary to different target RNAs (Fig. 8A). Examination of L7 binding to sR6 in titration experiments revealed more efficient binding at an elevated temperature (Fig. 5B). Interestingly, even at elevated temperature L7 appeared to bind only once. The gradual decrease in RNP mobility suggested that alternative sRNA structures or remodeling may be occurring during L7 binding. A similar effect has been previously observed for the binding of the 15.5kDa protein, the eukaryotic homolog of L7, to the human U15 snoRNA (30). To confirm that L7 was binding only once, an affinity isolation protocol was developed. Both N-terminally His-tagged and untagged L7 were incubated with sR6 at elevated temperature. Assembled RNPs were then affinity selected using the His tag. Low levels of untagged L7 were co-purified with the selected His-L7:sR6 complex (Fig. 5C). In contrast, similar affinity selection of assembled His-L7:sR8 complexes co-purified much higher amounts of untagged L7 protein. These results indicated only very weak binding of a second L7 to sR6.

Temperature-dependent assembly of the complete sR6 sRNP was examined next (Fig. 6A). Similar to sR8 sRNP assembly, elevated temperatures were required for efficient Nop56/58 binding. Some Nop56/58 binding was observed at 4°C, but complete conversion of the L7:sR6 RNP to a larger complex containing Nop56/58 clearly required elevated temperature (Fig. 6A, lane 7 versus lane 11). Interestingly, subsequent binding of fibrillarin at either 4°C or 70°C did not result in the presence of a prominent, slower migrating (larger) complex (Fig 6A, lanes 12 and 16). Rather, the appearance of a faster migrating complex, particularly evident at 70°C, indicated that full complex assembly may result in sRNP compaction. A similar observation has been made for fibrillarin binding in the assembly of the eukaryotic box C/D snoRNP (Gagnon, Zhang, and Maxwell, unpublished results). Assembly of fully functional sR6 sRNP complexes was confirmed in methylation assays (Table 1).

CD and thermal denaturation analyses, along with nuclease mapping of the assembling complex, were used to further investigate sR6 sRNP assembly. As observed with the sR8
sRNP, CD analysis revealed a stepwise ordering of the sR6 sRNA structure upon sequential core protein binding (Fig. 6B). Thermal denaturation experiments demonstrated an increase in the $T_m$ (~20.8°C) upon L7 binding. Binding of the Nop56/58-fibrillarin dimer only slightly altered the melting curve profile (Fig. 6C). Of particular note, however, was a difference in the melting curve profile of free sR6 sRNA compared with that of free sR8 (Fig. 3C). Melting of sR8 yielded a biphasic melting curve whereas that of sR6 was monophasic. Since the major difference between these two sRNAs is their respective guide sequences (Fig. 8A), we concluded that the difference in melting profiles was due to the distinctly folded structures of the guide sequences of each sRNA (see below).

Nuclease mapping experiments assessed changes in sR6 RNA structure during sRNP assembly (Fig. 6D). Prior to core protein binding, boxes C, D, C’, and D’ are clearly accessible, similar to sR8 sRNA. L7 binding protected boxes C, D, and C’ sequences from cleavage with smaller effects for box D’. Since L7 appeared to bind to both motifs and previous experiments indicated a single L7-bound sR6, we concluded that initial binding of a single L7 protein could be at either motif. Only when Nop56/58 was added did box D’ become significantly protected. As with the sR8 sRNP, the binding of fibrillarin had modest effects on the digestion pattern, consistent with its binding primarily through protein-protein interactions with Nop56/58. Cleavage patterns of the sR6 box C/D and C’/D’ motifs upon core protein binding were very similar to those observed during sR8 sRNP assembly. However, different cleavage patterns were noted in the sR6 and sR8 guide regions, again suggesting distinctly folded structures for the respective sRNA guide sequences.

F. sR6 Guide Regions Are Exposed Throughout sRNP Assembly

Accessibility of sR6 guide regions for target RNA probe base pairing during sRNP assembly was assessed by probing with Pb$^2+$ (Fig. 7A). In surprising contrast to sR8, both D and D’ guide regions were available to bind their respective target RNAs throughout sRNP assembly. Both target RNAs bound free sR6 (lanes 4-7) as well as when sR6 was bound with core proteins (lanes 8-15). Consistent with these observations, and in contrast to sR8,
L7 binding to sR6 at low temperatures was largely unaffected by bound D and D’ target RNA probes (Fig 7B).

G. sRNA Guide Sequences Can Influence the sRNP Assembly and Remodeling Pathway

Analysis of sR6 and sR8 sRNP assembly suggested that the sRNA-specific guide sequences could affect the assembly and remodeling pathway. To investigate further the influence of D and D’ guide sequences upon box C/D sRNP assembly, two sR8 sRNA guide region mutants were engineered to possess alternative guide sequences (Fig. 8A). The first mutant switched the D and D’ guide regions within sR8 (sR8 GRrev), whereas the second mutant replaced D and D’ guide regions with the corresponding sR6 guide sequences (sR8 GRsr6). Initiation of sRNP assembly with L7 binding onto sR8 GRsr6 revealed L7 binding properties similar to that observed for sR6, with one L7 bound and only a weak second L7 interaction at both low and elevated temperatures (Fig. 8B and 8D). As the major difference between sR6 and sR8 is their guide sequences, this result was anticipated and reinforced the idea that guide region sequence/structure influences initiation of sRNP assembly. Also anticipated was the binding of two L7 molecules to sR8 GRrev. However, the unexpected ability of sR8 GRrev to now bind L7 twice at low temperature, albeit less efficiently than at elevated temperature, indicated that individual guide regions/sequences can work in concert with flanking box C/D and C’/D’ motifs to influence overall sRNA structure and dynamics (Figs. 8C and 8D). We have also analyzed the binding kinetics of L7 to sR8 GRrev and observed a hyperbolic binding curve at 4°C consistent with one or two site binding but exhibiting no cooperativity. In contrast, L7 binding at 70°C reveals sigmoidal binding and strong cooperatively with a Hill plot slope of nearly 2 (data not shown). These observations are consistent with the importance of sRNA remodeling for sRNP assembly and the influence of guide region sequence upon the remodeling pathway.

Analysis of full sRNP assembly for sR8 guide region mutants was carried out next using thermal denaturation analysis and temperature-dependent assembly. sR8 possessing sR6 or
reversed sR8 D and D’ guide sequences exhibited the expected thermal stabilization of sRNA structure upon core protein binding (Figs. 9A and B). Notable however, was the observation that reversing D and D’ guide regions had significant impact upon the melting profile of the free sR8 GRrev sRNA (Fig. 9B). UV melt analysis revealed the replacement of sR8’s characteristic biphasic melting curve (Fig. 2B) with a monophasic curve. This indicated that reversing the positioning of the D and D’ guide sequences significantly altered sRNA folded structure. Replacement of sR8 guide regions with the respective sR6 sequences resulted, as might be expected, in a monophasic curve that was characteristic of sR6 sRNA. Furthermore, replacement of sR6 guide regions with the corresponding sR8 guide sequences produced the expected biphasic melting curve (data not shown). Thus, individual guide sequences have the ability to fold the complete sRNA molecule into different conformations prior to core protein binding and sRNP assembly.

Core protein binding to the sR8 GRsr6 mutant demonstrated that substitution with sR6 guide sequences generated an sR8 molecule that exhibited the core protein binding characteristic of sR6 (Fig. 9C). Interestingly, sR8 possessing switched D and D’ guide regions (sR8 GRrev) bound two L7 core proteins even at low temperature and required only elevated temperature for Nop56/58 binding (Fig. 9D). Nuclease and Pb$^{2+}$ probing of mutant sRNP assemblies has revealed unique cleavage patterns for these mutants distinct from either sR8 or sR6 during assembly, although the fully assembled complexes exhibited similar patterns (data not show). Assessment of the methylation capabilities of both mutants revealed activities similar to the wild type sRNPs (Table 1). Together, these results support the idea that sRNA guide region sequences can fold into conformations which not only affect global sRNA structure, but also the initiation of sRNP assembly and the remodeling pathway followed to assemble functional sRNP complexes.
DISCUSSION

Archaeal box C/D sRNP assembly requires elevated temperatures to enhance the dynamic nature of the sRNA and permit RNA remodeling essential for core protein binding. The binding of the core proteins then restricts flexibility as sRNA structure becomes more ordered. For box C/D sRNAs, core protein binding at the box C/D and C’/D’ motifs stabilizes sRNA structure primarily through the establishment of K-turns. Bound core proteins may further provide a scaffold for the sRNA, exposing the guide sequences for target RNA base-pairing while establishing the necessary inter-RNP spacing required for nucleotide methylation. Comparison of different box C/D sRNAs has indicated that sRNP assembly can occur via different assembly pathways, both in terms of core protein binding and sRNA structural changes. Nevertheless, RNA structural remodeling upon core protein binding appears to be a common theme for each sRNA species in assembling an enzymatically active sRNA-guided nucleotide modification complex.

Box C/D sRNP assembly is initiated by L7 binding to the box C/D and/or C’/D’ motifs. The free and L7-bound K-turn fold of the box C/D motif has been well characterized by X-ray crystallography, computer simulations, and with FRET and biochemical analyses (21,30,34,37-41,46-50). FRET analysis has revealed an inherent structural flexibility that becomes more ordered with increasing concentrations of Mg$^{2+}$ and Na$^+$ ions. Similarly, the binding of L7 stabilizes the K-turn motif, restricting its conformation and inducing a sharp kink in the RNA backbone. At both low and elevated temperatures, independent box C/D and C’/D’ K-turns bind L7 efficiently. However, when box C/D and C’/D’ motifs are juxtaposed in the same molecule, elevated temperature is critical for increasing the conformational dynamics of the sRNA to facilitate L7 binding at both motifs. We have previously reported the cooperative nature of L7 binding to sR8 and the crosstalk interactions important for efficient methylation activities guided by both box C/D and C’/D’ RNPs (30,36). Therefore, it is reasonable to suggest that formation of inter-RNP interactions
critical for methylation activity may well be dependent upon the inherent structural flexibility of the sRNA itself.

Nop56/58 binding required prior L7 binding as well as elevated temperatures indicating the importance of a defined K-turn structure for binding. The small changes in sRNA conformation upon Nop56/58 binding indicated that structural alterations, such as base stacking and nucleotide positioning, of the K-turns or adjacent guide regions are more subtle and specifically required for Nop56/58 binding. A crystal structure of this protein has suggested that the protein's extended coiled-coil domain may facilitate self-dimerization in the fully assembled sRNP, perhaps establishing inter-RNP interactions between the box C/D and C'/D' complexes (42). However, we have recently shown that this interaction is not required for sRNP assembly and is not likely to occur in the complete sRNP complex (36). The observation that loss of the coiled-coil domain affects overall sRNP structure is, however, consistent with a structural role for Nop56/58 in establishing sRNA and sRNP conformations ultimately important for RNA-guided methylation.

In contrast to L7 and Nop56/58, fibrillarin does not require elevated temperature for binding and has only minor effects upon sRNA structure. The demonstrated dimerization of Nop56/58 and fibrillarin core proteins in solution is consistent with the idea that these two core proteins function in the sRNP as a dimer. We have previously proposed that Nop56/58 and fibrillarin form a protein scaffold or “platform” (36). It is upon this platform that the sRNA-target RNA duplex is formed and the base-paired target RNA for modification positioned. Recent work has suggested that in specific archaean organisms, such as *Archaeoglobus fulgidus*, L7 may help position this platform on the target RNA via L7 protein-protein interactions with Nop56/58 (36). This core protein platform could establish an architectural bridge between both terminal and internal C/D motifs, providing inter-RNP spacing constraints necessary to position the fibrillarin active site for specific 2'-O-methylation of target RNA nucleotides according to the N+5 rule (1,30,35). The need for
sRNA structure to conform to this protein platform would also help explain the nature of Nop56/58- and fibrillarin-induced RNA remodeling.

The methylase fibrillarin-Nop56/58 platform is anticipated to interact with D and D’ guide regions, ultimately positioning fibrillarin’s catalytic domain at the target nucleotide. Protection of specific sR8 and sR6 guide region nucleotides upon Nop56/58 binding is consistent with this positioning, as observed in chemical/nuclease mapping experiments. However, equally notable is the enhanced digestion of specific guide nucleotides upon core protein binding demonstrating induced, and possibly distinct, structural changes in the guide regions. This is apparent in comparing sR8 and sR6 sRNP assemblies where guide regions of sR8 are not available for target RNA binding during sRNP assembly while those of sR6 are continuously exposed. Guide sequences vary with each specific sRNA species, and these differences in functional folding are clearly caused by the individual guide regions, as demonstrated by the sR8 mutant constructs possessing replaced or simply reversed guide sequences. Therefore, it is not surprising that individual sRNA species would exhibit distinct RNA remodeling pathways reflecting their unique guide sequences.

A model summarizing alternative folding and assembly pathways to a functional sRNP is presented in Figure 10. The guide sRNA is conformationally dynamic when free in solution. Transient formation of box C/D and/or C’/D’ K-turn structures allow L7 core proteins to bind and initiate sRNP assembly. Considering the diversity of guide sequences present in the various box C/D sRNAs, it is likely that numerous folding intermediates will be exhibited by the individual RNAs as they assemble RNP complexes. Unfavorable sRNA folding intermediates may become energetically trapped at these early assembly stages depending on the particular sRNA species, thereby generating alternative remodeling and assembly pathways. For example, L7 may bind once or twice to one or both K-turn motifs. However, early assembly pathways begin to converge when Nop56/58 binds at elevated temperature, followed by addition of the methylase fibrillarin which may either bind with or after Nop56/58. Binding of these core proteins further remolds the sRNA/sRNP structure to
assemble complete box C/D and C’/D’ RNPs, expose the guide regions for target RNA binding, and establish the necessary inter-RNP spacing required for nucleotide methylation.

We have concluded that alternative sRNA folding schemes and sRNP assembly pathways may be followed by the many archaeal box C/D sRNAs. Thermophilic organisms such as *M. jannaschii* live at elevated growth temperatures (51). Thus, the ambient environment could facilitate sRNA remodeling during the sRNP assembly process. However, for most organisms RNA helicases, such as DExH/D box family proteins, are likely to guide important RNA structural changes during RNP assembly. For example, the eukaryotic nucleolus possesses a large number of RNA helicases, underscoring their important roles in various steps of ribosome biogenesis. Two DExD/H box proteins in yeast, Prp43p and Dbp4p, were recently found to function in ribosome biogenesis and implicated in chaperoning snoRNA activity (52-54). Box C/D snoRNP biogenesis in yeast requires Tip48 and Tip49 helicases whose presumed function is the rearrangement of RNA folded structures during RNP assembly (22,55-57). Similarly, the human U3 snoRNA is associated with these same helicases during RNP biogenesis in the nucleoplasm (57). RNA helicases are abundant in all kingdoms of life, suggesting that this class of proteins likely plays important structural roles in the biogenesis and function of numerous RNP complexes.

We have shown that archaeal box C/D sRNP assembly is dependent upon elevated temperature to facilitate the dynamic flexibility of sRNA structure necessary for core protein binding. This requirement for elevated temperature is reminiscent of in vitro assembly of *E. coli* 30S ribosomal subunits where heating steps during sequential ribosomal protein binding are necessary to facilitate assembly of a functional ribosome subunit (58). However, protein-induced changes in RNA structure are not limited to elevated temperatures as recently demonstrated with the binding of SRP19 at ambient temperature to alter SRP RNA structure during RNP assembly (59). Thus, remodeling of RNA structure facilitated by bound proteins appears to be a common and perhaps unifying theme in RNP assembly pathways. As with
archaeal box C/D sRNPs, these RNA structural changes are also likely to be critical for RNP function.
MATERIALS AND METHODS

RNA Synthesis

Full length sR8, sR6, and the sR8 C/D and C’/D’ half-molecule RNAs were synthesized as previously described (30). DNA oligonucleotides (Integrated DNA Technologies) were used for PCR-amplification of sR8 and sR8 sRNA genes. PCR amplification introduced a T7 promoter sequence at the 5’-end of the sRNA sequence and allowed introduction of sequence alterations in the sRNA coding region. sR8 (oligo 1), sR6 (oligo 4) sR8 GRrev (oligo 7), and sR8 GRsr6 (oligo 10) template DNA oligonucleotides were PCR-amplified with the following DNA primer pairs: sR8 (primers 2 and 3), sR6 (primers 5 and 6), sR8 GRrev (primers 8 and 9) and sR8 GRsr6 (11 and 12). The sequence of all templates and PCR primers are presented below (5’ to 3’).

1) AAATCGCGAATGATGACAATTTGCCTATCTGATTCTGTGATGACTACTCCCGC
   AGCTGAGGCCGATTT
2) AAATCGCCAATGATGAACCGTATGAGCACTGAGTCTGTGATGACGATTGGCTT
   TGCTGAGGCGATTT
3) AAATCGCCAATGATGACAATTTGCCTATCTGAGTCTGTGATGACTACTCCCGC
   AGCTGAGGCGATTT
4) CTAATACGACTCACTATAGGCCAAAACGTGCGATGACACATTT
5) AAACGTGGCTCAGCTGCGGGAGTA
6) CTAATACGACTCACTATAGGCCAAAACGTGCGATGACACATTT
7) AAATCGCCCTCAGCAAGCACAAGTGCG
8) CTAATACGACTCACTATAGGCCAAAACGTGCGATGACACATTT
9) AAATCGCCCTCAGCTGCGGGAGTAG

sRNAs were in vitro transcribed using Ampliscribe Flash T7 Transcription Kits (Epicentre) and gel-purified by denaturing polyacrylamide gel electrophoresis (PAGE). Purified RNAs were radiolabeled at the 5’ terminus by first digesting with calf intestine alkaline phosphatase (CIAP, Promega) then radiolabeling using T4 polynucleotide kinase (PNK, Promega) and γ-
$^{32}$P] ATP (MP Biomedicals). Radiolabeled sRNAs were re-purified using spin columns (Sephadex G-25 resin) or by denaturing PAGE. Target RNA oligonucleotides were chemically synthesized by Dharmajcon RNA Technologies, Inc. or Integrated DNA Technologies. Target RNA sequences are listed below (5’ to 3’). Nucleotides methylated at the 2’-$O$ position are designated with an ‘m’ preceding the nucleotide. sR8 D target: AUGCUCAUACGGUC; sR8 D-CH$_3$ target: UGAUGCUmCAUACGGUCUGCU; sR8 D’ target: GCUCAAAAGCCAAUCGC; sR8 D’-CH$_3$ target: GCUCAAAAmCCAAUCGC; sR6 D target: ACUGCGGGAGUAGC; sR6 D-CH$_3$ target: ACUGCmGGAGUAGC; sR6 D’ target: GCUAUAGCGAAAUUGC; sR6 D’-CH$_3$ target: GCUAUAGmCGAAAUUGC.

**Electrophoretic Analysis of RNP Assembly**

*M. jannaschii* box C/D sRNP core proteins L7, Nop56/58, and fibrillarin were expressed in *E. coli* and purified as previously described (30). For electrophoretic mobility-shift analysis (EMSA) of L7 binding to sRNAs, 5’-radiolabeled (0.2 pmol and 1x10$^4$ cpn) sRNA was incubated with increasing concentrations of L7 protein in binding buffer (20 mM HEPES, pH 7.0, 150 mM NaCl, 1.5 mM MgCl$_2$, 10% (w/v) glycerol) supplemented with 0.5 mg/mL tRNA (final volume of 20 µL). Reactions were incubated at 70°C for 10 minutes and then cooled to 23°C. Higher-order RNP assembly contained 20 pmol of L7, 32 pmol Nop56/58, and 32 pmol fibrillarin and a final concentration of 1.5 mg/mL tRNA. Complexes were resolved on native 4% or 6% polyacrylamide gels buffered with 25 mM potassium phosphate, pH 7.0, and 2% glycerol. Gels were vacuum-dried and visualized using a phosphorimager or by autoradiography. For temperature-dependent analysis of RNP assembly, reactions were either assembled on ice or heated to 70°C, then equilibrated on ice before resolving on native polyacrylamide gels at 4°C. Radiolabeled target RNA and L7 binding to sRNAs at 4°C was carried out by binding target RNA to the guide sRNA, then cooling and incubating on ice for 15 minutes with increasing concentrations of L7. Assembled complexes were resolved on native polyacrylamide gels at 4°C.
Radiolabeled sRNA (≥ 1x10^5 cpms) was incubated with 20 pmol L7, 32 pmol Nop56/58, 32 pmol fibrillarin, and/or 20 pmol of target RNA(s) in assembly buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 3 mM MgCl₂, 0.2 mM EDTA) containing 0.5 mg/mL tRNA. Reactions were incubated at 70°C then at 50°C before cooling to 23°C. Mung bean nuclease (Epicentre) was added to a final of 0.63 units/µL and reactions incubated at 23°C for 8-10 minutes. Alternatively, lead acetate was added to a final of 2 mM. For mapping of low temperature L7 binding, protein-RNA complexes were assembled at either 70°C or on ice and then probed with mung bean nuclease on ice. T1 digestion ladders were generated by incubating 5'-radiolabeled sRNA with 0.3 mg/mL tRNA and 0.05 units/µL RNase T1 (Ambion) in assembly buffer at 23°C. Alkaline hydrolysis ladders were generated by incubating 5'-radiolabeled sRNA with 0.3 mg/mL tRNA in 10 mM NaHCO₃, pH 10.0, 1 mM EDTA at 95°C. No reaction (NR) samples were handled identically except without added nuclease or Pb²⁺. All reactions were stopped by addition of 10 volumes of 2% LiClO₄ in acetone. Pellets were collected by centrifugation, washed with acetone, and then air dried. RNA samples were resuspended in loading buffer (1X Tris-borate-EDTA, 4 M urea, 0.25 mg/mL bromophenol blue, 0.25 mg/mL xylene cyanol, 12% (w/v) sucrose), boiled for 5 minutes, and resolved on 14% polyacrylamide sequencing gels containing Tris-borate-EDTA buffer, pH 8.3, and 7 M urea. Vacuum-dried gels were visualized using a phosphorimager.

**Circular Dichroism Spectroscopy**

Circular dichroism spectroscopic analysis was carried out using a Jasco Model J-600 spectropolarimeter with a water-cooled 1 cm cuvette. Each experiment consisted of a minimum of five wavelength scans from 300-220 nm in 1 nm steps. Data presented are the average of three or more separate experiments. For L7 binding to sR8, sR8 (1 µM) was incubated with and without L7 (2 µM) in buffer C (20 mM cacodylate, pH 7.0, 100 mM NaCl, 1 mM MgCl₂) at 4°C or 70°C then equilibrated to 4°C before analysis. For CD analysis of higher-order sRNP assembly, L7, Nop56/58, and fibrillarin proteins at final concentrations of 2 µM each were incubated at 70°C in different combinations with and
without sRNA (1 µM) in buffer C then analyzed at 23°C. All presented CD spectra are subtracted for buffer and protein spectra contributions in the 220-300 nm range. CD spectra of free sRNA in solution when previously incubated at either high or low temperature were identical, except for a small loss of total signal (~2%) due to RNA degradation from sample heating and cooling that was observed and accounted for in control experiments.

**Thermal Denaturation Analysis**

Thermal denaturation analysis of sRNAs and sRNPs was carried out using a CARY Varian model 3 UV-Vis spectrophotometer. Absorbance was monitored at 260 nm in a 1 cm quartz cuvette. sRNA (0.83 µM) in buffer UV (20 mM potassium phosphate, pH 7.0, 100 mM NaCl, 1 mM MgCl₂) was annealed and denatured twice from 4°C to 95°C at a ramp rate of 2°C/minute and absorbance collected at a rate of one reading per 0.5°C. Values were averaged, normalized and plotted as a function of temperature. Analysis of thermal stabilization by full sRNP assembly used 2 µM of L7, Nop56/58, and fibrillarin in combinations incubated with or without sR8. Protein and buffer contributions to UV absorbance were subtracted from the absorbance data before normalization and plotting.

**Affinity Isolation of Assembled L7-sRNA Complexes**

L7 occupancy on the sRNA box C/D and C'/D' motifs was evaluated by affinity chromatography. N-terminally His(6X)-tagged L7 was incubated with untagged L7 (5-fold molar excess) in the presence or absence of sRNA (5-fold molar excess) in buffer B (10 mM HEPES, pH 7.0, 150 mM NaCl, 1.5 mM MgCl₂, 10% (w/v) glycerol, 0.5 mg/mL tRNA) at 70°C for 10 minutes. Reactions were cooled to 23°C, passed over Ni-NTA His-bind resin (Novagen), and the resin washed with buffer E (10 mM HEPES, pH 7.0, 150 mM NaCl, 1.5 mM MgCl₂, 10% (w/v) glycerol, 50 mM imidazole, 0.007% SDS, 0.1% Triton X-100). Bound L7-sRNA complexes were eluted in buffer B containing 500 mM imidazole. Eluted proteins were acetone precipitated, resuspended in SDS loading buffer, and resolved on 16% SDS-polyacrylamide gels. Resolved proteins were visualized by Coomassie brilliant blue G-250 staining.
**In vitro Methylation Assays**

*In vitro* 2’-O-methylation assays for assembled box C/D sRNPs were carried out as previously described (30). Aliquots in triplicate were taken at 0 and 60 minute time points, TCA-precipitated, washed, and incorporation of $[^3\text{H}]-\text{CH}_3$ into D or D’ target RNAs determined by scintillation counting. Incorporation of $[^3\text{H}]-\text{CH}_3$ into target RNAs previously methylated at the target nucleotide served as controls and were subtracted from experimental levels.
ACKNOWLEDGEMENTS

We thank the Paul Wollenzien and Kevin Weeks laboratories for their advice on nuclease and Pb$^{2+}$ mapping techniques and Bernard Brown for critical reading of the manuscript. This work was supported by NSF Grant MCB 0543741 (ESM) and NIH Grant 2R01-GM23037 (PFA).
REFERENCES


structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, 103, 457-466.


### Table 1: 2'-O-Methylation activity of dual-guide box C/D sRNPs.

<table>
<thead>
<tr>
<th>Guide RNA</th>
<th>$D\text{ target}$ pmols per 60min minus D-CH3 background (% of WT sR8)</th>
<th>$D'\text{ target}$ pmols per 60min minus D'-CH3 background (% of WT sR8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sR8</td>
<td>12.3 ± 0.2 (100%)</td>
<td>12.9 ± 0.3 (100%)</td>
</tr>
<tr>
<td>sR6</td>
<td>16.2 ± 0.6 (131%)</td>
<td>16.5 ± 1.9 (127%)</td>
</tr>
<tr>
<td>sR8 GRrev</td>
<td>14.9 ± 2.5 (120%)</td>
<td>11.0 ± 0.5 (85%)</td>
</tr>
<tr>
<td>sR8 GRsr6</td>
<td>15.1 ± 0.7 (123%)</td>
<td>15.2 ± 0.4 (117%)</td>
</tr>
</tbody>
</table>
FIGURE 1: L7 core protein binding to both sR8 box C/D and C'/D' motifs requires elevated assembly temperatures. (A) Folded secondary structure of *M. jannaschii* sR8 box C/D sRNA. Conserved boxes C/D and C'/D' sequences are in bold. D and D' guide regions are indicated. (B) Electrophoretic mobility-shift analysis (EMSA) of L7 binding to sR8 sRNA at low and elevated temperatures. Radiolabeled sR8 was incubated with increasing concentrations of L7 at 70°C (left) or 4°C (right) and resultant RNPs resolved on native polyacrylamide gels at 4°C. (C and E) Folded secondary structures of sR8 halfmer RNAs possessing box C/D and C'/D' motifs. (D and F) EMSA analysis of L7 binding to box C/D and C'/D' halfmer RNAs at low and elevated temperatures. Radiolabeled sR8 halfmer RNAs were incubated with increasing concentrations of L7 at 70°C or 4°C and the resultant RNPs resolved on native polyacrylamide gels at 4°C.
FIGURE 2: L7 binding to sR8 remodels and stabilizes sRNA structure. (A) CD spectroscopy of L7 core protein binding to sR8 sRNA at low and elevated RNP assembly temperatures. Protein-only baselines were subtracted. (B) Thermal denaturation analysis of L7 binding to sR8 sRNA at increasing concentrations of L7 core protein. Insets are a plot of T_m as a function of L7 concentration and a Hill plot calculation of the data. Protein-only baselines were subtracted. (C) Mung bean nuclease mapping of L7:sR8 RNP complexes assembled at low and elevated temperatures. Radiolabeled sR8 was bound with L7 at low (4°C) and elevated (70°C) temperature, probed with mung bean nuclease at 4°C, and resulting RNA fragments resolved on a denaturing polyacrylamide sequencing gel. Reaction components and assembly temperatures are indicated above each lane. Nucleotide sequencing ladders were generated by RNase T1 digestion (T1) and alkaline hydrolysis (OH). No reaction (NR) is radiolabeled sR8 not digested with nuclease. Box C, D, C’, and D’ sequences are enclosed and D and D’ guide regions designated by brackets.
FIGURE 3: Complete assembly of the sR8 sRNP complex requires elevated assembly temperatures and sRNA remodeling. (A) Assembly of the sR8 sRNP complex by sequential addition of core proteins at low and elevated assembly temperatures. Radiolabeled sR8 was incubated sequentially with L7, Nop56/58, and fibrillarin core proteins at 4°C or 70°C and resulting RNP complexes resolved on native polyacrylamide gels at 4°C. Indicated above each gel lane are the components for each assembly step with the respective incubation temperature. Arrows indicate the order of protein addition for individual assembly reactions. (B) CD spectroscopy of sR8 sRNP assembly. (C) Thermal denaturation analysis of sR8 sRNP assembly. (D) Mung bean nuclease mapping of sR8 sRNP assembly. 5’-radiolabeled sR8 was sequentially bound with L7, Nop56/58, and fibrillarin core proteins at 70°C and the assembled RNPs then probed with mung bean nuclease at room temperature. Resulting sRNA fragments were resolved on a denaturing polyacrylamide sequencing gel. Reaction components are indicated above each lane. Nucleotide sequencing ladders were generated by RNase T1 digestion (T1) and alkaline hydrolysis (OH). No reaction (NR) is radiolabeled sR8 not digested with nuclease. Box C, D, C’, and D’ sequences are enclosed by boxes and D and D’ guide sequences designated by brackets.
FIGURE 4: sR8 D and D’ guide region accessibility to target RNA binding changes during sRNP assembly. (A) Assessment of sR8 D and D’ guide region accessibility to target RNA base-pairing using lead (Pb$^{2+}$) cleavage probing. Radiolabeled sR8 was sequentially assembled with sRNP core proteins, incubated with D and/or D’ target RNAs, and then cleaved with Pb$^{2+}$. sR8 sRNA cleavage products were resolved on a denaturing polyacrylamide sequencing gel. Components of the individual assembly reactions are indicated above each gel lane. Nucleotide sequencing ladders were generated by RNase T1 digestion (T1) and alkaline hydrolysis (OH). (B) EMSA of target RNA oligonucleotide binding to sR8 D and D’ guide sequences in the presence of L7 core protein. Radiolabeled D (left panel) and D’ (right panel) target RNA probes were incubated with non-radiolabeled sR8 sRNA at 4°C in the presence of increasing concentrations of L7. Free RNA, RNA:RNA hybrids, and assembled RNP complexes were resolved on native polyacrylamide gels at 4°C.
FIGURE 5: *M. jannaschii* sR6 requires elevated temperature for efficient L7 binding. (A) *M. jannaschii* sR6 folded secondary structure. Box C, D, C’, and D’ sequences are presented in bold with D and D’ guide sequences indicated. (B) Efficient L7 binding to sR6 sRNA requires elevated temperatures. Radiolabeled sR6 sRNA was bound with increasing concentrations of L7 core protein at 70°C or 4°C and resultant RNPs resolved on native polyacrylamide gels at 4°C. Concentrations of added L7 protein are indicated above each gel lane. (C) sR6 efficiently binds only one L7 core protein. His-tagged and non-tagged L7 proteins were incubated with sR8 or sR6 sRNAs at 70°C and assembled L7:sRNA complexes isolated by nickel affinity chromatography. Isolated proteins were resolved by SDS-PAGE and stained with Coomassie blue. A: proteins incubated with sRNA. E(-): L7 proteins incubated without sRNA and affinity purified. E(+): L7 proteins incubated with sRNA and affinity purified.
FIGURE 6: sR6 sRNP assembly requires sRNA remodeling. (A) The sR6 sRNP requires elevated temperature for efficient assembly. Radiolabeled sR6 sRNA was incubated sequentially with L7, Nop56/58, and fibrillarin core proteins at 4°C or 70°C and the resulting RNP complexes were resolved on a native polyacrylamide gel at 4°C. Indicated above each gel lane are the components for each assembly step as well as the respective incubation temperatures. Arrows indicate the order of core protein addition for individual assembly reactions. (B) CD spectroscopy of sR6 sRNP assembly. (C) Thermal denaturation analysis of sR6 during sRNP assembly. (D) Mung bean nuclease mapping of sR6 during sRNP assembly. 5’-radiolabeled sR6 was sequentially bound with L7, Nop56/58, and fibrillarin core proteins at 70°C followed by mung bean nuclease probing at room temperature. sRNA fragments were resolved on a denaturing polyacrylamide sequencing gel. Reaction components of each assembly reaction are indicated above each gel lane. Nucleotide sequencing ladders were generated by RNase T1 digestion (T1) and alkaline hydrolysis (OH). Box C, D, C’, and D’ sequences are enclosed by boxes and D and D’ guide sequences designated by brackets.
FIGURE 7: sR6 D and D’ guide regions are accessible throughout sRNP assembly. (A) Assessment of sR6 D and D’ guide region accessibility to target RNA oligonucleotide binding using Pb²⁺ cleavage probing. 5’-radiolabeled sR6 was sequentially assembled with sRNP core proteins, incubated with D and/or D’ target RNAs, and then cleaved with Pb²⁺. sRNA cleavage fragments were resolved on a denaturing polyacrylamide sequencing gel. Components of each assembly reaction are indicated above each gel lane. Nucleotide sequencing ladders were made by RNase T1 digestion (T1) and alkaline hydrolysis (OH). Box C, D, C’, and D’ sequences are enclosed by boxes and D and D’ guide sequences designated by brackets. (B) EMSA analysis of target RNA oligonucleotide binding to sR6 D and D’ guide sequences in the presence of L7 core protein. Radiolabeled D (left panel) and D’ (right panel) target RNA oligonucleotides were incubated with non-radiolabeled sR6 sRNA at 4°C in the presence of increasing concentrations of L7. Free RNA, RNA:RNA hybrids, and assembled RNP were resolved on native polyacrylamide gels at 4°C.
FIGURE 8: sRNA guide regions affect sRNA folding and L7 binding. (A) Primary sequence alignment of *M. jannaschii* sR8, sR6, sR8 GRsr6, and sR8 GRrev sRNAs. (B and C) sR8 GRsr6 and sR8 GRrev sRNAs bind one and two L7 core proteins, respectively, independent of temperature. Radiolabeled sRNA was incubated with increasing L7 concentrations at elevated and low temperatures and assembled RNPs resolved on polyacrylamide gels at 4°C. (D) His-tag affinity purification of L7:sR8 GRsr6 and L7:sR8 GRrev complexes. His-tagged and non-tagged L7 proteins were incubated with indicated sRNAs at 70°C and the assembled His-L7:sRNA complexes isolated by nickel affinity chromatography. Isolated proteins were resolved by SDS-PAGE and stained with Coomassie blue. A; his-tagged and non-tagged L7 proteins incubated with sRNAs. E(−); affinity purified L7 proteins incubated in the absence of added sRNA. E(+) ; affinity purified L7 proteins incubated in the presence of sRNA.
FIGURE 9: Altered sRNA guide regions affect sRNA structure and sRNP assembly. (A and B) Thermal denaturation analyses of sR8 GRsr6 and sR8 GRrev sRNP assembly. (C and D) Assembly of sR8 GRsr6 and sR8 GRrev sRNPs is temperature-dependent. Radiolabeled sRNAs were incubated sequentially with L7, Nop56/58, and fibrillarin core proteins at 4°C or 70°C. The resulting RNP complexes resolved on native polyacrylamide gels at 4°C. Indicated above each gel lane are the components for each assembly step with the respective incubation temperature. Arrows indicate the order of core protein addition for individual assembly reactions.
FIGURE 10: Archaeal Box C/D sRNPs can follow different remodeling and assembly pathways
CONCLUDING REMARKS

Archaeal *in vitro* assembly systems have facilitated detailed study of box C/D sRNP structure and methylation function. However, important questions regarding this minimal RNA-protein enzyme remain. For instance, what are the inter-RNP interactions of juxtaposed box C/D and C’/D’ RNPs within a full-length sRNP? These interactions are required for efficient nucleotide modification but it is not known how they affect RNP structure and, in turn, methylation function (1). Most archaeal box C/D sRNPs possess dual guide function and methylate two different target nucleotides, sometimes positioned within different target RNAs. What is the mechanism of target substrate interaction with the respective guide sequences? Can a box C/D sRNP bind two target RNAs simultaneously to carry out guided methylation from both complexes or is there interplay between the two RNPs such that only one is catalytically active at any given time? Likewise, recent work has indicated the importance of the RNP core proteins in target RNA binding (2) yet the nature of these target RNA-protein interactions is unknown. Do the core proteins assist in unfolding highly structured target RNAs and, moreover, what is the topology of the 12 basepair target RNA-guide RNA duplex which is equivalent to a full turn of an A-form RNA helix?

The archaeal sRNP has served, and will continue to serve, as a minimal model complex for studying the more structurally and functionally diverse eukaryotic snoRNP. However, the diversity of the eukaryotic complex poses interesting questions as well as specific challenges. How similar is the eukaryotic box C/D snoRNP with respect to the archaeal sRNP? Are eukaryotic snoRNPs also constrained by spatial positioning of the box C/D and C’/D’ RNPs for methylation activity and what is the stoichiometry of the snoRNP core proteins on each motif? More specifically, are Nop56 and Nop58 differentially distributed on the box C/D and C’/D’ RNPs as suggested by crosslinking experiments (3)? Also, do snoRNPs which chaperone RNA folding or catalyze target RNA methylation possess specific snoRNP core or accessory proteins? SnoRNA species-specific proteins might include helicases to chaperone RNA folding activity, as was demonstrated for the Dbp4 protein of the yeast U14 snoRNP.
(4). Finally, detailed structural and functional studies of the eukaryotic box C/D snoRNP should begin to rationalize the requirement for a more complex and functionally diverse RNP. Importantly, they should provide new insight into what kinds of evolutionary processes may have occurred during divergence of the archaeal and eukaryotic systems.

While *in vitro* assembly of archaeal box C/D sRNP can provide answers to some of these outstanding questions, the development of new approaches using different techniques will be necessary to address the more challenging problems. For example, crystal structures of fully assembled box C/D sRNP complexes, both free and bound with target RNA, could offer remarkable insight into box C/D RNP structure and function. Likewise, the establishment of an *in vitro* assembly system for eukaryotic box C/D snoRNPs would be a milestone achievement and would finally allow the detailed analysis of box C/D snoRNP structure and function that the archaeal box C/D sRNPs have enjoyed. To this end, the research presented here provides starting points for future structural analyses of box C/D sRNPs and the development of an *in vitro* eukaryotic box C/D snoRNP assembly system. Chapter I describes an affinity purification protocol for the isolation of box C/D sRNPs that may be suitable for x-ray crystallography. In Appendix III working protocols for the large scale production of milligram quantities of L7, Nop56/58 and fibrillarin core proteins are also provided. Chapter II has laid basic groundwork for understanding the fundamental RNA remodeling events that are necessary for a functional box C/D sRNP. These events are very likely common to eukaryotic box C/D snoRNAs as well, considering that they are structurally and functionally homologous in a number of ways. Furthermore, helicases, which have the ability to remodel RNA and RNP structures, have been linked to the processing and function of eukaryotic box C/D snoRNPs (4-8). Appendix II provides working protocols for the expression and purification of recombinant mouse core proteins 15.5kD (L7 homolog), Nop56 and Nop58 (Nop56/58 homologs), and fibrillarin and accessory proteins p50 and p55. Preliminary data demonstrating *in vitro* assembly of a eukaryotic box C/D snoRNP using mouse core proteins and the human snoRNA U15A is also shown. This preliminary work, in combination with further research to identify
accessory proteins and optimal assembly conditions, should in time yield a functional, *in vitro* assembled eukaryotic box C/D snoRNP.
REFERENCES


APPENDICES
APPENDIX A

Scientific Meeting Abstracts
Assembly of Functional Archaeal Dual-Guide Box C/D sRNP Assembly

Keith T. Gagnon, Xinxin Zhang, Paul F. Agris, and E. Stuart Maxwell.

Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA 27695.

Archaeal dual-guide box C/D sRNAs bind three core proteins in a sequential order at both terminal box C/D and internal C’/D’ motifs to assemble two RNP complexes active in guiding nucleotide methylation. Experiments have investigated the process of box C/D sRNP assembly and the resultant changes in sRNA structure, or ‘remodeling’, as a consequence of sRNP core protein binding. Hierarchical assembly of the *M. jannaschii* sR8 box C/D sRNP is a temperature-dependent process, with L7 and Nop56/58 core proteins requiring high temperature to facilitate necessary RNA structural changes. RNA thermal denaturation and circular dichroism (CD) spectroscopy revealed an increased stability of RNA folded structure as a result of L7 binding. Subsequent binding of the Nop56/58 and fibrillarin core proteins to the L7:sRNA complex further stabilized sRNA structure. Probing of sR8 sRNA guide region structure using complementary oligonucleotides uncovered significant alterations in guide region accessibility during hierarchical sRNP assembly.

Another dual-guide box C/D sRNA from *M. jannaschii*, sR6, also exhibited RNA remodeling upon temperature-dependant sRNP assembly, although to a lesser extent. Conversely, sR6 sRNA showed little change in guide region structure during sRNP assembly. Further experiments using sR8 mutants possessing alternative guide sequences suggested that structure introduced by guide regions can impact overall sRNA folded structure, sRNA unfolding/folding pathways, and initiation of sRNP assembly. Nevertheless, sR8 mutants with alternative guide sequences are catalytically active, indicating correct sRNP assembly and accessible guide region structure. Thus, RNA remodeling appears to be a common and requisite feature of archaeal dual-guide box C/D sRNP assembly and formation of mature sRNP complexes appears independent of the sRNA refolding pathway taken.
Assembly of the Archaeal Box C/D sRNP Can Occur Via Alternative Pathways and Requires Temperature-Facilitated sRNA Remodeling. Keith T. Gagnon, Xinxin Zhang, Paul F. Agris, and E. Stuart Maxwell. Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA 27695.

Archaeal dual-guide box C/D sRNAs bind three core proteins in sequential order at both terminal box C/D and internal C’/D’ motifs to assemble two RNP complexes active in guiding nucleotide methylation. Experiments have investigated the process of box C/D sRNP assembly and the resultant changes in sRNA structure or ‘remodeling’ as a consequence of sRNP core protein binding. Hierarchical assembly of the Methanocaldococcus jannaschii sR8 box C/D sRNP is a temperature-dependent process with binding of L7 and Nop56/58 core proteins to the sRNA requiring elevated temperature to facilitate necessary RNA structural dynamics. Circular dichroism (CD) spectroscopy and RNA thermal denaturation revealed an increased order and stability of sRNA folded structure as a result of L7 binding. Subsequent binding of the Nop56/58 and fibrillarin core proteins to the L7-sRNA complex further remodeled sRNA structure. Assessment of sR8 guide region accessibility using complementary RNA oligonucleotide probes revealed significant changes in guide region structure during sRNP assembly. A second dual-guide box C/D sRNA from M. jannaschii, sR6, also exhibited RNA remodeling during temperature-dependent sRNP assembly, although core protein binding was affected by sR6’s distinct folded structure. Interestingly, the sR6 sRNP followed an alternative assembly pathway, with both guide regions being continuously exposed during sRNP assembly. Further experiments using sR8 mutants possessing alternative guide regions demonstrated that sRNA folded structure induced by specific guide sequences impacted the sRNP assembly pathway. Nevertheless, assembled sRNPs were active for sRNA-guided methylation independent of the pathway followed. Thus, RNA remodeling appears to be a common and requisite feature of archaeal dual-guide box C/D sRNP assembly and formation of the mature sRNP can follow different assembly pathways in generating catalytically active complexes.
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’ RNP s. Each of these RNP s 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, predating the divergence of Archaea and Eukarya some two billion years ago.
Assembly of the Archaeal Box C/D sRNP Requires Temperature-Facilitated sRNA Remodeling and Can Occur via Alternative Pathways. Stuart Maxwell, Keith Gagnon, and Xinxin Zhang. Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA 27695.

Archaeal dual-guide box C/D sRNAs bind three core proteins in sequential order at both terminal box C/D and internal C’/D’ motifs to assemble two RNP complexes active in guiding nucleotide methylation. Experiments have investigated the process of box C/D sRNP assembly and the resultant changes in sRNA structure or ‘remodeling’ as a consequence of sRNP core protein binding. Hierarchical assembly of the Methanocaldococcus jannaschii sR8 box C/D sRNP is a temperature-dependent process with binding of L7 and Nop56/58 core proteins to the sRNA requiring elevated temperature to facilitate necessary RNA structural dynamics. Circular dichroism (CD) spectroscopy and RNA thermal denaturation revealed an increased order and stability of sRNA folded structure as a result of L7 binding. Subsequent binding of the Nop56/58 and fibrillarin core proteins to the L7-sRNA complex further remodeled sRNA structure. Assessment of sR8 guide region accessibility using complementary RNA oligonucleotide probes revealed significant changes in guide region structure during sRNP assembly. A second dual-guide box C/D sRNA from M. jannaschii, sR6, also exhibited RNA remodeling during temperature-dependent sRNP assembly, although core protein binding was affected by sR6’s distinct folded structure. Interestingly, the sR6 sRNP followed an alternative assembly pathway, with both guide regions being continuously exposed during sRNP assembly. Further experiments using sR8 mutants possessing alternative guide regions demonstrated that sRNA folded structure induced by specific guide sequences impacted the sRNP assembly pathway. Nevertheless, assembled sRNPs were active for sRNA-guided methylation independent of the pathway followed. Thus, RNA remodeling appears to be a common and requisite feature of archaeal dual-guide
box C/D sRNP assembly and formation of the mature sRNP can follow different assembly pathways in generating catalytically active complexes.
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 of the designated nucleotide. Archaeal box C/D RNP complexes possess terminal box C/D and internal C'/D’ RNPs. Each RNP guides nucleotide methylation and is comprised of a folded RNA K-turn motif bound with three core proteins. We have used the in vitro assembled *M. jannaschii* sR8 sRNP as a model complex to investigate box C/D RNP assembly, structure, and function. Binding of the three core proteins is sequential and efficient methylation activities require juxtaposed box C/D and C'/D’ RNPs. 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. Box C/D sRNP assembly requires sRNA remodeling and assembly of different sRNP complexes can follow different sRNA folding pathways. However, fully assembled sRNPs present exposed D and D’ guide regions for target RNA binding. Examination of sRNA guide sequence-target RNA duplex formation has demonstrated that target nucleotide methylation requires an A helix RNA:RNA duplex formed between the sRNA guide sequence and target RNA. Efficient nucleotide methylation also requires interaction of the target RNA with the flanking sRNP core proteins. Finally, methylation experiments indicate that the archaeal sRNP complex has an intrinsic ability to melt double-stranded target RNAs for subsequent nucleotide methylation.
The Role of Nop56/58 Core Protein in Archaeal Box C/D sRNP Assembly, Target RNA Binding, and sRNP-Guided Nucleotide Methylation. Keith T. Gagnon, Xinxin Zhang, Denise Appel, Paul Wollenzien, and E. Stuart Maxwell. Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA 27695.

Archaeal box C/D RNAs guide the 2’-O-methylation of ribosomal RNA and tRNAs. These small, non-coding guide RNAs are complexed with three core proteins to form the box C/D sRNP. The L7 core protein initiates sRNP assembly by binding terminal box C/D and internal C’/D’ RNA motifs. sRNP assembly is completed with the sequential binding of Nop56/58 followed by fibrillarin to form the box C/D and C’/D’ RNPs. While fibrillarin functions as the methyltransferase enzyme, the role of Nop56/58 is less clear. Nop56/58 and fibrillarin dimerize and evidence suggests that this protein:protein interaction plays a key role in both sRNP assembly and methylation function (EMBO J. (2003) 22:3930-40). Mutational analysis is investigating the importance of Nop56/58 dimerization with fibrillarin for both sRNP assembly and nucleotide methylation. Preliminary results demonstrate that interaction with Nop56/58 is required for fibrillarin binding to the sRNA, indicating that fibrillarin assembly onto the box C/D sRNP complex occurs primarily through protein:protein interactions. Mutational analysis and chemical stability studies have also shown that this protein-protein interaction is exceptionally stable.

Previous work has also revealed a role for Nop56/58 in sRNA remodeling as this core protein facilitates sRNA conformational changes required to expose the D and D’ guide sequences during sRNP assembly (JMB (2006) 362:1025-42). Investigations have now implicated Nop56/58 in target RNA binding to the sRNP complex. Target RNA methylation requires the Watson-Crick pairing of an RNA:RNA duplex to the sRNA guide sequence, thereby determining the specific nucleotide for modification. However, target RNA substrates with sequence that extends beyond the guide-target RNA base-paired region dramatically
increases the level of nucleotide methylation, resulting in enhanced catalytic turnover (RNA (2007) 13:899-911). These results suggest possible target RNA interaction with the sRNP core proteins. Indeed, extended target RNAs with 4-thiouridine incorporated into the 5’ or 3’ termini outside of the base-pairing region were synthesized and found to crosslink to the Nop56/58 core protein when bound to the sRNP complex and UV irradiated.
Box C/D RNP nucleotide methylation complexes are abundant in both Archaea and Eukarya, functioning to direct the 2’-O-methylation of numerous nucleotides found in rRNAs, tRNAs, snRNAs, and some mRNAs. Guide sequences in 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. Archaeal box C/D sRNP provide a model complex for both structure and function studies as the box C/D RNP can be assembled in vitro using in vitro transcribed sRNAs and recombinant core proteins. Three archaeal core proteins bind both the terminal box C/D and the internal C’/D’ motifs to form box C/D and C’/D’ RNPs, each capable of guiding target nucleotide methylation. Ribosomal protein L7 initiates sRNP assembly at each motif followed by Nop56/58 and fibrillarin binding. Efficient methylation requires that both RNPs be fully assembled and juxtaposed in a single RNP complex. Box C/D sRNP assembly requires sRNA remodeling with the sRNA’s guide sequences being fully exposed for target RNA binding upon complete sRNP assembly. Nop56/58 and fibrillarin core proteins dimerize to establish a protein “platform” upon which the target RNA lays and base pairs with the sRNA guide sequence to carry out nucleotide methylation. Nop56/58 mutagenesis and affinity pull-down experiments have demonstrated the interaction between Nop56/58 and fibrillarin is strong, with fibrillarin binding utilizing the N-terminus of Nop56/58 for protein-protein interactions. The in vitro sRNP assembly system has enabled a detailed analysis of target RNA binding to the sRNA guide sequence. Nucleotide methylation requires Watson-Crick pairing of the target RNA with the sRNA guide sequence and mismatches are not tolerated. Methylation also requires formation of an A helix RNA:RNA duplex between target RNA and guide sequence. Extension of the target RNA
beyond those 12 nucleotides hydrogen-bonded to the guide sequence enhances nucleotide methylation activity by increasing target RNA turnover. Crosslinking of lengthened target RNAs possessing 4-thio-uridine at the termini have shown that the target RNA interacts with Nop56/58. This suggests that the Nop56/58 core protein plays an important role(s) in target RNA binding and perhaps the methylation reaction itself. Current work is determining the sites of target RNA crosslinking with the RNP core proteins with the goal of using available core protein crystal structures to model the binding of the target RNA base paired with the sRNA guide sequence upon the Nop56/58-fibrillarin protein platform.
Conserved Sequence Elements of the Archaeal L7 and Eukaryotic 15.5kD Box C/D Core Protein Homologs Determine Their Differential Binding to RNA K-turn and K-loop Motifs. Xinxin Zhang, Keith T. Gagnon, Jimmy Suryadi, Bernard A. Brown, and E. Stuart Maxwell. Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA 27695.

Box C/D RNAs are abundant, small, non-coding RNAs found in both Archaea and Eukarya. Their primary function is to guide the 2'-O-ribose methylation of target RNAs which include rRNAs, tRNAs, and snRNAs. Core proteins recognize the box C/D RNA’s terminal K-turn and internal K-loop motifs to assemble RNP complexes. The guide RNA directs nucleotide methylation of target RNA through its complementary guide regions while bound core proteins catalyze the methyltransfer reaction. The 15.5kD and L7 core proteins initiate assembly of the eukaryotic snoRNP and archaeal sRNP complexes, respectively. The eukaryotic 15.5kD core protein recognizes only the terminal box C/D K-turn motif whereas the homologous archaeal L7 core protein binds both terminal K-turn and internal K-loop motifs. Although available crystal structures of both 15.5kD and L7 proteins reveal remarkably similar structures when binding the K-turn and K-loop motifs, specific sequence elements and structures of the respective homologs are responsible for their differential binding specificities. These high resolution crystal structures have been used to guide the site-directed mutagenesis of eukaryotic 15.5kD and archaeal L7 proteins to determine the specific sequence elements of each protein that are responsible for their distinct RNA binding capabilities. A conserved network of polar contacts and hydrogen bonding between 15.5kD loop 9 amino acids and the RNA are crucial features for its recognition of the terminal K-turn motif. The absence of most of these homolog-specific contacts in archaeal L7 binding to the K-turn and K-loop motifs suggest they are responsible for 15.5kD’s unique binding to only the K-turn motif. Also of significance is archaeal L7 amino acid K26 and eukaryotic amino acid Q34. These amino acids mediate loop 1 and loop 9 interactions in both protein
homologs and are 100% conserved in the respective proteins. Substitution of L7 loop 9 amino acids with their respective 15.5kD counterparts, in combination with substitution of the loop 1 amino acid (K26Q), alters L7 binding to mimic that of 15.5kD with recognition of only the K-turn motif. Ongoing experiments are further characterizing features of the L7/15.5kD loop 9 and K26/Q34 contacts which contribute to the different binding capabilities of these two protein homologs. We propose that during evolution, mutation of eukaryotic 15.5kD’s RNA binding site to recognize only the terminal K-turn motif has facilitated the expansion of box C/D RNA’s size and sequence, thus resulting in a more structurally and functionally diverse box C/D guide RNA and corresponding RNP.
APPENDIX B

Purification of Recombinant Box C/D snoRNP Core Proteins from Mouse and *In Vitro* Assembly of a Box C/D snoRNP
I. GST-Tagged Mouse 15.5kD Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* 15.5kD (m15.5kD) DNA sequence was amplified from I.M.A.G.E Consortium cDNA clone #809008 (Research Genetics, Inc.) with primers introducing BamHI restriction sites both upstream and downstream. Amplified DNA was inserted into the pGEX-4T-1 vector and sequencing gave a positive match to the published gene sequence. Plasmid was subsequently transformed into DH5α *E. Coli* host cells for protein expression.

B. Growth and Expression

1. Inoculate 2 mL LB broth (+ 50 µg/mL carbenicillin (Cb)) with a single colony and grow overnight with shaking at 37°C.
2. Inoculate 300 mL LB + 0.5% glucose (+ Cb) with 2 mL overnight culture and grow with shaking at 37°C to an OD$_{600}$ = 0.8.
3. Induce protein expression by adding IPTG to a final of 0.5 mM and continuing to shake at 37°C for 3 hours.
4. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl$_2$, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.
5. Add PMSF to a final concentration of 1 mM, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals. Rock lysate at room temperature for 2 hours.
6. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.
C. Purification

1. Pack a 15 mL column with 2 mL bed volume of GST resin (Glutathione Sepharose 4B, Amersham Pharmacia). Equilibrate resin with 1 column volume of Buffer D.

2. Apply soluble protein fraction to resin by gravity and wash with 20 bed volumes of wash buffer (20 mM HEPES, pH 7.2, 0.5 M NaCl, 0.1% Triton X-100). Wash with 5 bed volumes of Buffer D.

3. Elute with 2.5 bed volumes (5 mL) of Buffer D + 20 mM glutathione.

4. Regenerate resin with alternative 5 mL washes of high (0.1 M tris, pH 8.5, 0.5 M NaCl) and low (0.1 M sodium acetate, pH 4.5, 0.5 M NaCl) pH buffers. Wash with 20% ethanol and store column at 4°C.

5. Dialyze eluate against 100 volumes (500 mL) of Buffer D at 4°C.

D. Quantification and Purity

1. Resolve fractions on a 14% SDS-PAGE.

2. Quantify GST-m15.5kD protein after dialysis using a Bradford Assay using BSA as a protein standard. Note that GST-m15.5kD molecular weight is ~ 32 kDa.

3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
II. (6X)Histidine-Tagged Mouse 15.5kD Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* 15.5kD (m15.5kD) DNA sequence was amplified from the m15.5kD in pGEX-4T-1 clone 2 (described above) with primers introducing NdeI and HindIII restriction sites upstream and downstream, respectively. Amplified DNA was inserted into pET28a (Novagen) and sequencing gave a positive match to the published gene sequence. Plasmid was subsequently transformed into Rosetta (DE3) *E. Coli* host cells (Novagen) for protein expression.

B. Growth and Expression

1. Inoculate 2 mL LB broth (+ 34 µg/mL chloramphenicol (Cm), + 30 µg/mL kanamycin (Km)) with a single colony and grow overnight with shaking at 37°C.
2. Inoculate 300 mL LB + 0.5% glucose (+ Cm, + Km) with 2 mL overnight culture and grow with shaking at 37°C to an OD_{600} = 0.8.
3. Induce protein expression by adding IPTG to a final of 1 mM and continue shaking at 37°C for 3 hours.
4. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl₂, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1 % Triton X-100 per gram of cells.
5. Add PMSF to a final concentration of 1 mM, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals. Rock lysate at room temperature for 2 hours.
6. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.
C. Purification

1. Pack a 15 mL column with 2 mL bed volume of Ni$_{2+}$-NTA His-Bind resin (Novagen). Charge resin with 50 mM NiSO$_4$ and equilibrate with 1 column volume of Buffer D.

2. Apply soluble protein fraction to resin by gravity and wash with 20 bed volumes of wash buffer (20 mM HEPES, pH 7.2, 0.5 M NaCl, 0.1% Triton X-100, 25 mM imidazole). Wash with 5 bed volumes of Buffer D.

3. Elute with 2.5 bed volumes (5 mL) of Buffer D + 0.2 M imidazole.

4. Regenerate resin by washing with 4 bed volumes of strip buffer (0.1 M EDTA), 4 volumes of H$_2$O, and 6 volumes of 20% ethanol and store column at 4°C.

5. Dialyze eluate against 100 volumes (500 mL) of Buffer D at 4°C.

D. Quantification and Purity

1. Resolve fractions on a 16% SDS-PAGE.

2. Quantify His-m15.5kD protein after dialysis using absorption at 280 nm and Beer’s Law. ($A_{280} = E \times B \times C$) with $B = 1$ cm and $E = 4,320$ M$^{-1}$ cm$^{-1}$ (calculated, http://www.basic.northwestern.edu/biotools/proteincalc.html). Note that His-m15.5kD molecular weight = 16,335 g/mol.

3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
III. Mouse Nop56 Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* Nop56 (nucleolar protein 56) DNA sequence was amplified from mouse cDNA library with primers introducing NdeI and SalI restriction sites upstream and downstream, respectively, and ligated into pET15b cut with NdeI and XhoI. Sequencing of this clone matched the published sequence. Nop56 in pET15b has a 6X-His tag that is not accessible for binding to Ni$^{2+}$-NTA resin. Therefore, Nop56 was subcloned from pET15b into pET21a using NdeI and BamHI restriction sites. Nop56 in pET21a was subsequently transformed into Rosetta-pLysS (DE3) *E. Coli* host cells (Novagen) for protein expression.

B. Growth and Expression

1. Inoculate 4 mL of LB broth (+ 50 µg/mL carbenecillin (Cb), + 34 µg chloramphenicol (Cm)) with a single colony and grow with shaking overnight at 37ºC.
2. Inoculate two flasks containing 450 mL of LB + 0.5% glucose (+ Cb, + Cm) with 2 mL of overnight culture each. Grow with shaking at 37ºC to an OD$_{600}$ = 0.7.
3. Move flasks to coldroom and shake at 12ºC for 45 minutes. Induce expression by adding IPTG to a final concentration of 0.2 mM and continue shaking at 12ºC for 48-72 hours.
4. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl$_2$, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.
5. Add protease inhibitor (Cocktail Set I, Calbiochem) to 1%, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals.
6. Rock lysate at room temperature for 4 hours. Add 5 M NaCl for a final concentration of 0.35 M in Buffer D.
7. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.

C. Purification

1. Set up 50mL (1.5 x 30 cm) chromatography column in coldroom and prepare a 20 mL bed volume of SP Sepharose (strong cation exchanger, Pharmacia or Fisher).
2. Wash resin bed with 30 mL H2O and equilibrate with 40 mL of wash buffer (20 mM HEPES, pH 7.2, 0.35 M NaCl, 0.1% Triton X-100).
3. Apply soluble protein fraction to the resin twice. Wash with 150 mL of wash buffer. Wash with 50 mL Buffer D. Elute with 40 mL of Buffer D500 (Buffer D at 0.5 M NaCl instead of 0.1 M).
4. Clean resin by washing with 80 mL of 2 M NaCl and 80 mL of 20% EtOH. Store in 20% ethanol at 4°C.
5. Concentrate the eluate to 5mL with spin concentrator (5,000 MWCO) at 4°C then dialyze against Buffer D300 (Buffer D at 0.3 M NaCl) at 4°C.
   NOTE: mNop56 tends to precipitate in low ionic strength buffers.
6. Remove any insoluble material from sample after dialysis by centrifugation.
7. Concentrate further if necessary with Amicon Microcon YM-3 concentrators.

D. Quantification and Purity

1. Resolve fractions on a 12% SDS-PAGE.
2. Quantify mNop56 using Bradford Assay or absorption at 280 nm and Beer’s Law. 
   \(A_{280} = E \times B \times C\) with \(B = 1\ cm\) and \(E = 23,290\ M^{-1}\ cm^{-1}\) (calculated, http://www.basic.northwestern.edu/biotools/proteincalc.html). Note that mNop56 molecular weight = 64,462 g/mol.
3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
IV. Mouse Nop58 Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* Nop58 (nucleolar protein 58) DNA sequence was amplified from mouse cDNA library with primers introducing NdeI and XhoI restriction sites upstream and downstream, respectively, and ligated into pET15b cut with NdeI and XhoI. The sequence of this clone matched the published sequence. Nop58 in pET15b has a 6X-His tag that is not accessible for binding to Ni$^{2+}$-NTA resin. Therefore, Nop58 was subcloned from pET15b into pET21a using NdeI and BamHI restriction sites. Nop58 in pET21a was subsequently transformed into Rosetta (DE3) E. Coli host cells (Novagen) for protein expression.

B. Growth and Expression

1. Inoculate 4 mL of LB broth (+ 50 µg/mL carbenecillin (Cb), + 34 µg chloramphenicol (Cm)) with a single colony and grow overnight with shaking at 37ºC.
2. Inoculate two flasks containing 450 mL of LB + 0.5% glucose (+ Cb, + Cm) with 2 mL each of overnight culture. Grow with shaking at 37ºC to an OD$_{600}$ = 0.7.
3. Move flasks to coldroom and shake at 12ºC for 45 minutes. Induce expression by adding IPTG to a final concentration of 0.3 mM and continue shaking at 12ºC for 48-72 hours.
4. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl$_2$, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.
5. Add protease inhibitor (Cocktail Set I, Calbiochem) to 1%, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals.
6. Rock lysate at room temperature for 4 hours. Add 5 M NaCl to give a final [NaCl] in Buffer D of 0.35 M.
7. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.

C. Purification
1. Set up 50mL (1.5 x 30 cm) chromatography column in coldroom and prepare a 20 mL bed volume of SP Sepharose (strong cation exchanger, Pharmacia or Fisher).
2. Wash resin bed with 30 mL H₂O and equilibrate with 40 mL of wash buffer (20 mM HEPES, pH 7.2, 0.35 M NaCl, 0.1% Triton X-100).
3. Apply soluble protein fraction to the resin twice. Wash with 300 mL of wash buffer. Wash with 50 mL Buffer D. Elute with 40 mL of Buffer D₆₀₀ (Buffer D at 0.6 M NaCl instead of 0.1 M) and collect fractions.
4. Clean resin by washing with 80 mL of 2 M NaCl and 80 mL of 20% EtOH. Store in 20% ethanol at 4°C.
5. Combine fractions containing protein and concentrate to 5mL with spin concentrator (5,000 MWCO) at 4°C then dialyze against Buffer D₃₀₀ (Buffer D at 0.3 M NaCl).
   NOTE: mNop58 tends to precipitate in low ionic strength buffers.
6. Remove any insoluble material from sample after dialysis by centrifugation.
7. Concentrate further if necessary with Amicon Microcon YM-3 concentrators.

D. Quantification and Purity
1. Resolve fractions on a 12% SDS-PAGE.
2. Quantify mNop58 using Bradford Assay or absorption at 280 nm and Beer’s Law. \( A_{280} = E \times B \times C \) with \( B = 1 \text{ cm} \) and \( E = 36,070 \text{ M}^{-1} \text{ cm}^{-1} \) (calculated, http://www.basic.northwestern.edu/biotools/proteincalc.html). Note that mNop58 molecular weight = 60,340 g/mol.
3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
V. GST-Tagged Mouse Fibrillarin Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* fibrillarin DNA sequence was initially amplified from ATCC cDNA clone #1245532 and cloned into pET15b (Novagen) using NdeI and XhoI restriction sites. Fibrillarin was then cloned into pGEX-4T-1 from the above vector to give an N-terminal GST-tag. Plasmid was subsequently transformed into DH5α *E. Coli* host cells for protein expression.

B. Growth and Expression

1. Inoculate 8 mL LB broth (+ 50 µg/mL carbenicillin (Cb)) with a single colony and incubate overnight with shaking at 37°C.
2. Inoculate four flasks containing 450 mL of LB + 0.5% glucose (+ Cb) with 2 mL of overnight culture each and grow with shaking at 37°C to an OD600 = 0.8.
3. Move flasks to coldroom and shake at 12°C for 45 minutes. Induce protein expression by adding IPTG to a final of 0.3 mM and continue shaking at 12°C for 16-24 hours. Move to room temperature and shake for an additional 6-8 hours.
4. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl2, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.
5. Add PMSF to a final concentration of 1 mM, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals. Rock lysate at room temperature for 4 hours.
6. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.
C. Purification

1. Pack a 50 mL column with 10 mL bed volume of GST resin (Glutathione Sepharose 4B, Amersham Pharmacia). Equilibrate resin with 20 mL of Buffer D.

2. Add soluble protein fraction to column and resuspend resin with it. Let resin settle then collect flow-through and wash with 150 mL of wash buffer (20 mM HEPES, pH 7.2, 0.5 M NaCl, 0.5 mM DTT, 0.1% Triton X-100). Wash with 30 mL of Buffer D$_{300}$ (Buffer D with 300 mM NaCl instead of 100 mM).

3. Elute with 20 mL of Buffer D$_{300}$ + 20 mM glutathione.

4. Regenerate resin by doing alternative 40 mL washes of high (0.1 M tris, pH 8.5, 0.5 M NaCl) and low (0.1 M sodium acetate, pH 4.5, 0.5 M NaCl) pH buffers. Wash with 20% ethanol and store column at 4°C.

5. Concentrate eluate and dialyze against 100 volumes of Buffer D$_{300}$ at 4°C.

D. Quantification and Purity

1. Resolve fractions on a 12% SDS-PAGE.

2. Quantify GST-mFibrillarin protein after dialysis using a Bradford Assay and BSA as a protein standard. Note the molecular weight of GST-mFibrillarin is ~ 52 kDa.

3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
VI. (6X)Histidine-Tagged Mouse Fibrillarin Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* fibrillarin DNA sequence was initially amplified from ATCC cDNA clone #1245532 and cloned into pET15b (Novagen) using NdeI and XhoI restriction sites. Fibrillarin was then sub-cloned from pET15b into pET33b using NdeI and BamHI restriction sites. This gives an N-terminal His-tag with an extra long intervening amino acid sequence containing a protein kinase A recognition site after the thrombin cleavage site. Plasmid was subsequently transformed into Rosetta (DE3) *E. Coli* host cells (Novagen) for protein expression.

B. Growth and Expression

1. Inoculate 4 mL LB broth (+ 34 µg/mL chloramphenicol (Cm), + 30 µg/mL kanamycin (Km)) with a single colony and incubate overnight with shaking at 37°C.
2. Inoculate two flasks containing 450 mL LB + 0.5% glucose (+ Cm, + Km) with 2 mL overnight culture each and grow with shaking at 37°C to an OD$_{600}$ = 0.6.
3. Move cultures to coldroom and shake at 15°C for 45 minutes.
4. Induce protein expression by adding IPTG to a final of 0.3 mM and continue shaking at 15°C for 48 hours.
5. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl$_2$, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.
6. Add PMSF to a final concentration of 1 mM, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals. Rock lysate at room temperature for 4 hours.
7. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.
C. Purification

Primary Purification with Ni-NTA His-Bind Resin:

1. Pack a 20 mL column with 4 mL bed volume of Ni²⁺-NTA His-Bind resin (Novagen). Charge resin with 50 mM NiSO₄ and equilibrate with 20 mL of Buffer D.
2. Resuspend resin in column with soluble protein fraction. Let resin settle and collect the flow-through.
3. Wash with 20 bed volumes of wash buffer (20 mM HEPES, pH 7.2, 0.5 M NaCl, 0.1% Triton X-100, 25 mM imidazole). Wash with 5 bed volumes of Buffer D₃₀₀ (Buffer D with 300 mM NaCl instead of 100 mM).
4. Elute with 10 mL of Buffer D₃₀₀ + 0.2 M imidazole.
5. Regenerate resin by washing with 4 bed volumes of strip buffer (0.1 M EDTA), 4 volumes of H₂O, and 6 volumes of 20% ethanol and store column at 4°C.
6. Concentrate eluate 2-3 fold and dialyze against Buffer D₃₀₀ at 4°C.

Secondary Purification with Heparin Agarose Resin (Optional):

1. Apply eluate from His-resin (step 4 above) to 3 mL of Heparin Agarose equilibrated with Buffer D₇₀₀ (0.7 M NaCl) in a 15 mL column. Bind by gravity flow.
2. Wash with 20 bed volumes of Buffer D₇₀₀.
3. Elute with 8 mL of Buffer D₁₃₀₀ (1.3 M NaCl).
4. Concentrate eluate 2-3 fold and dialyze against Buffer D₃₀₀ at 4°C.

D. Quantification and Purity

1. Resolve fractions on a 14% SDS-PAGE.
2. Quantify His-mFibrillarin protein after dialysis using absorption at 280 nm and Beer’s Law. \( A_{280} = E \times B \times C \) with \( B = 1 \) cm and \( E = 14,890 \, \text{M}^{-1} \, \text{cm}^{-1} \) (calculated, http://www.basic.northwestern.edu/biotools/proteincalc.html). Note that His-mFibrillarin molecular weight = 37,041 g/mol.
3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
VII. (6X)Histidine-Tagged Mouse p50 Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* p50 DNA sequence was PCR amplified from an I.M.A.G.E. clone with gene specific primers introducing NdeI and SalI restriction sites upstream and downstream, respectively. The PCR product was inserted into the pET15b expression vector at the NdeI and XhoI restriction sites to provide an N-terminal His-tag. Plasmid was subsequently transformed into Rosetta-pLysS (DE3) *E. coli* host cells (Novagen) for expression.

B. Growth and Expression

1. Inoculate 2 mL LB broth (+ 34 µg/mL chloramphenicol (Cm), + 50 µg/mL carbenicillin (Cb)) with a single colony and incubate overnight with shaking at 37°C.
2. Inoculate 450 mL of LB + 0.5% glucose (+ Cm, + Cb) with 2 mL overnight culture and grow with shaking at 37°C to an OD$_{600}$ = 0.8.
3. Induce protein expression by adding IPTG to a final of 0.5 mM and continue shaking at 37°C for 4 hours.
4. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl$_2$, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.
5. Add protease inhibitor (Cocktail Set VII, Calbiochem) to 1%, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals. Rock lysate at room temperature for 2 hours.
6. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.
C. Purification

1. Pack a 15 mL column with 2 mL bed volume of Ni$^{2+}$-NTA His-Bind resin (Novagen). Charge resin with 50 mM NiSO$_4$ and equilibrate with 20 mL of Buffer D.
2. Apply soluble protein fraction to resin by gravity flow. Collect the flow-through.
3. Wash with 30 bed volumes of wash buffer (20 mM HEPES, pH 7.2, 0.5 M NaCl, 0.1% Triton X-100, 50 mM imidazole). Wash with 5 bed volumes of Buffer D.
4. Elute with 2.5 bed volumes of Buffer D + 0.2 M imidazole.
5. Regenerate resin by washing with 4 bed volumes of strip buffer (0.1 M EDTA), 4 volumes of H$_2$O, and 6 volumes of 20% ethanol and store column at 4°C.
6. Dialyze eluate against Buffer D at 4°C.

D. Quantification and Purity

1. Resolve fractions on a 12% SDS-PAGE.
2. Quantify His-mp50 protein after dialysis using Bradford Assay or absorption at 280 nm and Beer’s Law. ($A_{280} = E \times B \times C$) with $B = 1$ cm and $E = 17,570$ M$^{-1}$ cm$^{-1}$ (calculated, http://www.basic.northwestern.edu/biotools/proteincalc.html). Note that His-mFibrillarin molecular weight = 53,136 g/mol.
3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
VIII. (6X)Histidine-Tagged Mouse p55 Protein Expression and Purification

A. Expression Vector and Host

*Mus musculus* p55 DNA sequence was PCR amplified from an I.M.A.G.E. clone with gene specific primers introducing NdeI and Sall restriction sites upstream and downstream, respectively. The PCR product was inserted into the pET15b expression vector at the NdeI and XhoI restriction sites to provide an N-terminal His-tag. Plasmid was subsequently transformed into Rosetta-pLysS (DE3) *E. coli* host cells (Novagen) for expression.

B. Growth and Expression

7. Inoculate 2 mL LB broth (+ 34 µg/mL chloramphenicol (Cm), + 50 µg/mL carbenicillin (Cb)) with a single colony and incubate overnight with shaking at 37°C.

8. Inoculate two flasks containing 450 mL of LB + 0.5% glucose (+ Cm, + Cb) with 2 mL each of overnight culture and grow with shaking at 37°C to an OD$_{600}$ = 0.8.

9. Induce protein expression by adding IPTG to a final of 0.8 mM and continue shaking at 37°C for 4 hours.

10. Harvest cells by centrifugation. Resuspend in 4 mL Buffer D (20 mM HEPES, pH 7.2, 0.1 M NaCl, 3 mM MgCl$_2$, 0.5 mM DTT, 20% (w/v) glycerol) + 0.1% Triton X-100 per gram of cells.

11. Add protease inhibitor (Cocktail Set VII, Calbiochem) to 1%, benzonase nuclease (Novagen) at 10 units/mL, and sonicate 3x for 30 seconds on ice with 1 minute cooling intervals. Rock lysate at room temperature for 2 hours.

12. Clarify lysate by ultracentrifugation at 15,000 RPMs (38,000xG) at 4°C for 30min or 1hr at 10,000 RPMs in a Sorvall RC-5B. If necessary, the soluble protein fraction may be stored at -80°C for several days before purification.
C. Purification

1. Pack a 20 mL column with 4 mL bed volume of Ni\textsuperscript{2+}-NTA His-Bind resin (Novagen). Charge resin with 50 mM NiSO\textsubscript{4} and equilibrate with 5 bed volumes of Buffer D.
2. Apply soluble protein fraction to resin by gravity flow. Collect the flow-through.
3. Wash with 30 bed volumes of wash buffer (20 mM HEPES, pH 7.2, 0.5 M NaCl, 0.1% Triton X-100, 50 mM imidazole). Wash with 5 bed volumes of Buffer D.
4. Elute with 2.5 bed volumes of Buffer D + 0.2 M imidazole.
5. Regenerate resin by washing with 4 bed volumes of strip buffer (0.1 M EDTA), 4 volumes of H\textsubscript{2}O, and 6 volumes of 20% ethanol and store column at 4°C.
6. Dialyze eluate against Buffer D at 4°C.

D. Quantification and Purity

1. Resolve fractions on a 12% SDS-PAGE.
2. Quantify His-mp50 protein after dialysis using Bradford Assay or absorption at 280 nm and Beer’s Law. \( A_{280} = E \times B \times C \) with \( B = 1 \) cm and \( E = 16,080 \text{ M}^{-1} \text{ cm}^{-1} \) (calculated, http://www.basic.northwestern.edu/biotools/proteincalc.html). Note that His-mFibrillarin molecular weight = 52,237 g/mol.
3. Determine the degree of nucleic acid contamination by performing a wavelength scan on purified protein or checking absorbance at 280 nm vs. 260 nm.
IX. *In Vitro* Assembly of a Eukaryotic Box C/D snoRNP

A. Introduction

*In vitro* assembly of a functional eukaryotic box C/D snoRNP has been a long-term goal for several research labs, including our own. An *in vitro* assembly would allow for systematic and controlled study of the structure and function of this protein-RNA complex, similar to investigations of the *in vitro* assembled archaeal box C/D sRNP. We therefore undertook the task of assembling a eukaryotic box C/D snoRNP from purified components. Presented below is a protocol for assembly of a box C/D snoRNP complex from synthesized RNA and recombinant core proteins.

This work began with the cloning of genes for mouse box C/D core proteins 15.5kD, Nop56, Nop58, fibrillarin, and accessory proteins p50 and p55. They were originally cloned from a mouse cDNA library by Donna Newman in our lab in 1998. Since then, other students, including Jeff Kuhn and Elizabeth Tran, have contributed to expression and purification of these proteins. Following their preliminary work, I refined and developed purification protocols (sections I-VIII) for each protein and successfully assembled a box C/D snoRNP complex *in vitro* (this section). However, a functional complex has not been demonstrated and this continues to frustrate our efforts. Nevertheless, a basic assembly protocol for a eukaryotic box C/D snoRNP is provided here as a starting point for future students and researchers.

B. *In Vitro* Transcription and Radiolabeling of Human U15 Box C/D snoRNA

The human box C/D snoRNA U15 was chosen as a model snoRNA for assembly. It is ~150 nucleotides, approximately twice as large as the minimal box C/D sRNAs from archaeal organisms. However, empirical studies comparing assembly with other smaller snoRNAs, such as human U24 or U14, have shown that the core proteins appear to bind U15 more efficiently *in vitro* (data not shown).
U15A was amplified from HeLa cell genomic DNA and includes 156 nucleotides upstream and 128 nucleotides downstream of the snoRNA coding sequence in intron 1 of ribosomal protein S3. This DNA fragment was inserted between EcoRI and HindIII restriction sites of pBluescript SK+ (RNA (1996), 2:118-133). The pBS+hU15A plasmid is used as a template for subsequent PCR amplification and T7 transcription.

**Generate T7-hU15 DNA by PCR for in vitro T7 Transcription**

1. Setup PCR reaction:
   
   10 µL 10x PCR buffer (+Mg²⁺)
   83 µL dH₂O
   2 µL up primer (0.2 mg/mL)
   2 µL down primer (0.2 mg/mL)
   1 µL dNTP mix (10 mM)
   1 µL pBS+hU15A (50 ng/µL)
   1 µL T4 DNA Taq Polymerase (5 units)

   Up primer: CTAATACGACTCACTATAGGCCCTTCGATGAAGATGATGACG
   Down primer: CCTTCTCAGACAAATGCCTCTAAG

2. Follow the PCR program below:

<table>
<thead>
<tr>
<th>Step</th>
<th>Time (min)</th>
<th>Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>---- cycle to step two 30 times ----</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>72</td>
</tr>
<tr>
<td>7</td>
<td>--</td>
<td>4</td>
</tr>
</tbody>
</table>

3. Clean up PCR product by gel-purification or phenol-chloroform extraction followed by ethanol precipitation.
RNA synthesis by T7 Transcription and Gel-Purification:

1. Follow manufacturer’s protocol using Ampliscribe T7 Flash Transcription Kit (Epicentre) or any in vitro T7 transcription kit.
2. Remove the T7 enzyme from the reaction by phenol-chloroform extraction and resolve RNA product on a 5% denaturing urea-PAGE. Visualize bands by UV shadowing or light staining with methylene blue in dH₂O.
3. Extract RNA bands and elute RNA from gel slices by crush-and-soak method using RNA elution buffer (10 mM Tris, 0.3 M sodium acetate, 5 mM EDTA, 0.1% SDS, pH 7.4).
4. Precipitate RNA with 70% ethanol. Pellet by centrifugation, wash with 70% ethanol, resuspend dry pellet in dH₂O and quantitate by UV absorbance at 260 nm.

5'-end Radiolabel Purified hU15 snoRNA

1. Remove 5’ phosphate from RNA by treatment with calf-intestinal alkaline phosphatase (CIAP) following manufacturer’s protocol. Use 20 µg or more to make a sufficient stock of CIAP-treated RNA. Phenol-chloroform extract and ethanol precipitate RNA. Quantitate by UV absorbance at 260 nm.
2. Add a radioactive phosphate (³²P) to the 5'-end with T4 polynucleotide kinase (PNK):
   1 µg CIP treated RNA
   5 µL 10x PNK buffer
   15 µL [γ]-³²P ATP
   2 µL PNK (20 units)
   Add dH₂O up to 50 µL
3. Incubate at 37°C for 1.5 hours. Phenol-chloroform extract and ethanol precipitate RNA. (Optionally, radiolabeled RNA can be gel-purified for the cleanest RNA bands in EMSA)
4. Check radioactivity of 1 µL in a liquid scintillation counter.

C. In Vitro Assembly and Electrophoretic Mobility Shift Analysis (EMSA)

Although enzymatic activity of this in vitro assembled complex has not been demonstrated, gel-shifts have revealed an apparent order of assembly. Below is a protocol for the assembly
of the U15 box C/D snoRNP with radiolabeled RNA and recombinant core proteins. The assembly reactions are resolved on a native polyacrylamide gel to demonstrate binding of each protein and order of assembly. The core proteins 15.5kD, Nop56, and Nop58 can each bind U15. Nop56 and Nop58 binding appears to be more efficient when 15.5kD is first bound. Addition of fibrillarin has no effect on U15 or the U15-15.5kD RNP. When Nop56 or Nop58 are first bound to U15 or the U15-15.5kD RNP, then addition of fibrillarin results in a faster migrating RNP band. Neither p50 nor p55 appear to stably bind to the assembled complex and their presence has no effect on methylation activity. (Figure 1 and data not shown).

1. Assemble reactions according to Table I below:

<table>
<thead>
<tr>
<th>Rxn #</th>
<th>15.5kD</th>
<th>Nop56</th>
<th>GST-Fibrillarin</th>
<th>Nop58</th>
<th>His-p50</th>
<th>His-p55</th>
<th>Buffer D</th>
<th>U15*</th>
<th>IRNA</th>
<th>Buffer B</th>
<th>NaCl</th>
<th>ATP</th>
<th>H2O</th>
<th>final vol. (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.9</td>
<td>1</td>
<td>0.1</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.4</td>
<td>1</td>
<td>4.1</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.4</td>
<td>1</td>
<td>4.6</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>1</td>
<td>3</td>
<td>0.5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.4</td>
<td>1</td>
<td>4</td>
<td>4.6</td>
<td>20</td>
</tr>
</tbody>
</table>

Buffer D (1x): 20 mM HEPES, pH 7.2, 100 mM NaCl, 0.5 mM DTT, 3 mM MgCl₂, 20% glycerol (w/v)
Buffer B (10x): 1 M NaCl, 0.1 M HEPES, pH 7.2

NOTE: All proteins are in Buffer D. Nop56 and Nop58 are in Buffer D₃₀₀ (300 mM NaCl instead of 100 mM).

2. Assemble at room temperature then move to 30°C for 20 minutes.

3. Add 2.5 μL of 10X phosphate dye and resolve on a 4.5% native (PO₄-buffered) polyacrylamide gel.

4. Dry gel and expose to phosphorimager cassette or film.
FIGURE 1: Assembly of a eukaryotic box C/D snoRNP. 15.5kD binds the radiolabeled U15 box C/D RNA first (lane 2), followed by binding of either Nop56 or Nop58 (lane 3) and fibrillarin (lane 4). p50 and p55 do not appear to bind the assembled complex (lane 6).
APPENDIX C

Protocols for Large-Scale Production of Recombinant Methanocaldococcus jannaschii
Box C/D sRNP Core Proteins
I. Large-Scale Production of His-tagged *M. jannaschii* L7 Core Protein

A. Bio-Reactor (from New Brunswick Scientific) Preparation (day 1)

1. Autoclave Bio-Reactor

   The Bio-Reactor tank (12 L) and all installed components attached to the top casing that will contact the media in the reactor tank must be autoclaved prior to set-up. The entire tank and stainless steel casing can be assembled and laid down in a large plastic autoclave-safe container and autoclaved at normal settings for 20 min. The dissolved oxygen (dO$_2$) sensor and pH probe are autoclave-safe.

2. Make Media

   Remove top of Bio-Reactor casing and add 250 g dry LB media, 26.8 g of K$_2$HPO$_4$ and 13 g of KH$_2$PO$_4$. Add 10 mL of 34 mg/mL chloramphenicol, 10 mL of 30 mg/mL kanamycin, and 20 drops of anti-foam agent. Add ddH$_2$O up to 10 L mark on tank. Remove pH probe and re-attach top casing. Agitate at 300 rpms to dissolve media.

3. Calibrate pH Probe

   Use pH 7.0 and pH 4.0 commercial buffers to calibrate. Place probe in pH 7.0 buffer until the pH reading stops changing. Set this as pH 7.0 in the “zero” option. Rinse well and place in pH 4.0 buffer until it stops changing. The pH should be near 4.0. Now set the “span” as 4.0. Install the calibrated pH probe into the top casing of the Bio-Reactor. When not in use, the pH probe should be stored with its tip in 3M KCl.

4. Calibrate the dO$_2$ Sensor

   Set the agitation of the reactor to as high as it will go during culture growth (320-340 rpms), turn the compressed air on to the highest pressure that will be used (>5 LPM with in-line 0.2 µm filter), and raise the temperature of the reactor to the culture growth temperature (37°C). Also, make sure the pH is set to auto and set to add base (1 M NaOH) if it goes below 6.9 and acid (1 M HCl) if it goes above 7.1. When the dO$_2$ level stops rising, set the “zero” as “100.” Temporarily disconnect the dO$_2$ sensor. When it stably reaches 0%, set the “span” to 0 and re-connect the sensor.

5. Turn off Bio-Flow 110 until ready for use (next morning).
B. Protein Expression (day 1 & 2)

1. Remove ~250 mL of prepared media from Bio-Reactor tank and inoculate with a colony of Rosetta (DE3) cells previously transformed with L7 in pET28a and plated. Shake at 37°C overnight.

2. Next morning turn on Bio-Flow 110, turn on agitation and air flow, and warm up to 37°C. Inoculate with 250 mL overnight culture. Make sure that pH is set to auto, dO₂ is around 100%, and compressed air is flowing. The dO₂ levels will steadily decrease as the culture grows into exponential phase. It will drop to around 0% and then slowly begin to rise again as the culture growth slows. The OD₆₀₀ at this point should be around 1.2-1.5 OD. **NOTE:** If your Bio-Flow allows, set it to cascade the agitation from 300 up to 900 rpms to offset the drop in dO₂ during culture growth. The dO₂ will still decrease then rise, but the increased agitation will allow for longer culture time with greatly increased cell densities before growth slows. If not, maintain at highest/safest setting up to 900 rpms (ours currently only reaches ~330-340 rpms due to a warped agitation shaft).

3. Induce protein expression by addition of 10 mL of 1M IPTG for 6 hours.

4. Siphon out culture into 0.5 L bottles and spin cells in GS-3 rotor (5,000xg, 10 min) to pellet. Cells may be stored at -20°C for overnight.

5. Resuspend cells in 4 mL of Buffer D (20 mM Hepes, pH 7.2, 100 mM NaCl, 3 mM MgCl₂, 0.5 mM β-ME, 20% glycerol) + 0.1% Triton X-100 per gram of cell pellet. Cells may be stored at -20°C for overnight.

6. Add lysozyme at 1 mg/gram of cell paste, PMSF (in 100% EtOH) to a final concentration of 1 mM, and benzonase nuclease (Novagen, >90% purity) at a final of 20 units/mL cell suspension. Rock at room temperature for 16-24 hours. Add 1/10 volume of 5 M NaCl for a final of ~600 mM.

7. Heat lysate to 75°C for 10 minutes then cool to room temperature.

8. Pellet insoluble debris by spinning in Sorvall RC-5B for 1 hour at >10,000xG. Soluble lysate may be stored at 4°C or -80°C for several days before purification.
C. Protein Purification (day 3 & 4)
Two methods can be used for purification, either a simple gravity flow column or our low pressure liquid chromatography (LPLC) setup. Both protocols are shown below.

Gravity-flow Column (day 2 & 3)
1. Charge 50 mL of His-bind resin (Novagen) in a ~150 mL econo-column with 3 bed volumes (150 mL) 50 mM NiSO₄ and equilibrate with wash buffer (20 mM HEPES, pH 7.2, 0.6 M NaCl, 20 mM imidazole, 0.1% Triton X-100).
2. Bind soluble lysate to resin and wash with 600 mL wash buffer. Put manual pump on column during wash step. Wash with 100 mL Buffer D.
3. Elute protein from resin with 150 mL of Buffer D containing 180 mM imidazole. Eluate can be stored at 4°C for a few days or -80°C for longer term. Concentrate ~10 fold and dialyze into 100 volumes of Buffer D at room temperature.

BioRad® BioLogic LPLC (day 2 & 3)
1. Pack ~150 mL econo-column with 50 mL His-bind Resin (Novagen).
2. Make buffers:
   A. wash buffer – same as wash buffer for gravity-flow column (800 mL)
   C. charge buffer - 50 mM NiSO₄ (120 mL)
   D. strip buffer - 100 mM EDTA, pH 7.0 (220 mL)
   E. elution buffer – same as elution buffer for gravity-flow column (150 mL)
3. Place column on HPLC, connect, and manually flush all lines with ddH₂O. Manually pump to fill each tubing line with appropriate buffer, doing the charge buffer last. Run “LARGE L7 HIS” method. Make sure to keep a small volume of buffer on top of the resin at all times during purification so that the resin does not go dry. This method first charges the resin with 100 mL charge buffer then equilibrates with 120 mL wash buffer.
   A. After equilibration you must zero the UV detector.
4. Turn on computer and run LP Data program. This synchronizes with the LPLC and graphs absorbance levels over time or volume. Click “record” in the program to begin.
5. After equilibration, an alarm will sound. Stop the alarm and PAUSE the method. Switch the machine manually from “fill” to “inject” using knob on side. Manually pump soluble lysate using separate pump system (Pharmacia Biotech Pump P-1) into the top inlet of the knob (injection port, only port not being used); this should send the soluble fraction to the column. Collect flow through to save for analysis by moving the “waste” tubing to a clean collection beaker when the UV absorbance begins to rise significantly (above 0.2 OD).

6. After applying the soluble fraction, switch the knob from “inject” back to “fill” and resume method. It will wash with 600 mL of wash buffer.

7. After washing another alarm will sound. Do two things:
   A. Switch buffer A with 200 mL of ddH₂O and buffer B with 200 mL 20% ethanol. This is in preparation for cleaning the resin after stripping it later. The method will pump A first and B second in order to wash the resin and equilibrate it in 20% ethanol for storage after elution step.
   B. Pause the method, take the column offline and reconnect the tubing so it is continuous to the detector. Resume the method and after 15 mL of elution buffer has run through, zero the UV detector and pause the method. Reconnect the column and resume method. It will pump through another 135 mL of elution buffer and begin collecting fractions when the OD exceeds 0.15.

8. After eluate fractions have been collected, store them at 4°C. When method is finished, take column offline and store in coldroom. Flush all tubing lines with ddH₂O. Eluate can be stored at 4°C for several days. Concentrate and dialyze the eluate into Buffer D. L7’s N-terminal 6x-His tag can be removed by addition of thrombin (5-10 units/mL). Keep at 4°C until fully digested (takes a few days, check on gel).

D. Qualitative and Quantitative Analysis of His-L7 Purification (day 4 & 5)
   1. Resolve a few µL of insoluble and soluble lysate fractions, as well as column flow-through and eluate fractions, on a 16% SDS-PAGE (Figure 1a).
2. Check concentration by absorbance at 280 nm. Run a wavelength scan to ensure that no nucleic acids (peak at 260 nm) co-purified with L7 (Figure 1b). If nucleic acid contamination is present, then treatment with fresh benzonase nuclease and repurification may be necessary for accurate quantification. Use absorbance at 280 nm to estimate concentration using Beer’s Law:

\[ A = (E)(C)(b), \quad A = A_{280} \]

\[ E = \text{extinction coefficient} = 5240 \text{ M}^{-1} \text{ cm}^{-1} \]

\[ C = \text{Molar concentration}, \]

\[ b = \text{cell pathlength} = 1 \text{ cm} \]

NOTE: The molecular weight of His-MjL7 is 14,850 g/mol (Daltons).

A.  B.

FIGURE 1: Purification of *M. jannaschii* His-L7. (A) Fractions collected throughout His-Resin purification of L7 using the BioRad Biologic LPLC purification protocol above were resolved on a 16% SDS-PAGE and coomassie stained. (B) Purified, concentrated and dialyzed L7 was diluted 1/10 in Buffer D and a wavelength scan performed in a Beckman DU-64 spectrophotometer.
II. Large-Scale Production of *M. jannaschii* His-fibrillarin:ΔK-tail Nop56/58 Core Protein Dimer

A. Introduction

For production of large amounts of the His-fib:ΔK-tail Nop56/58 dimer this protocol utilizes the Bio-Flow 110 Bio-Reactor from New Brunswick Scientific. Purification is achieved via the N-terminal 6x-Histidine tag on fibrillarin. Nop56/58 stably dimerizes with fibrillarin and thereby co-purifies. For this protocol, a Nop56/58 lacking the lysine-rich C-terminal domain (K-tail) is utilized.

B. Bio-Reactor Preparation (day 1)

1. Autoclave Bio-Reactor

   The Bio-Reactor tank (12 L) and all installed components attached to the top casing that will contact the media in the reactor tank must be autoclaved prior to set-up. The entire tank and stainless steel casing can be assembled and laid down in a large plastic autoclave-safe container and autoclaved at normal settings for 20 min. The dissolved oxygen (dO₂) sensor and pH probe are autoclave-safe.

2. Make Media

   Remove top of Bio-Reactor casing and add 250 g LB media, 26.8 g of K₂HPO₄, 13 g of KH₂PO₄, 0.34 g chloramphenicol, 1 g ampicillin, 0.3 g kanamycin, and 20 drops of anti-foam agent. Add ddH₂O up to 10 L mark on tank. Remove pH probe and re-attach top casing. Agitate at 300 rpms to mix.

3. Calibrate pH Probe

   Use pH 7.0 and pH 4.0 commercial buffers to calibrate. Place probe in pH 7.0 buffer until the pH reading stops changing. Set this as pH 7.0 in the “zero” option. Rinse well and place in pH 4.0 buffer until it stops changing. The pH should be near 4.0. Now set the “span” as 4.0. Install the calibrated pH probe into the top casing of the Bio-Reactor. When not in use, the pH probe should be stored with its tip in 3 M KCl.
4. Calibrate the dO₂ Sensor

Set the agitation of the reactor to as high as it will go during culture growth (330-340 rpms), turn the compressed air on to the highest pressure that will be used (>5 LPM, with in-line 0.2 µM filter), and raise the temperature of the reactor to the culture growth temperature (37°C). Also, make sure the pH is set to auto and set to add base (1 M NaOH) if it goes below 6.9 and acid (1 M HCl) if it goes above 7.1. When the dO₂ level stops rising, set this as 100% in the “zero.” Temporarily disconnect the dO₂ sensor. When it stops going down, set the “span” to 0 and re-connect the sensor.

5. Turn off Bio-Flow 110 until ready for use (next morning).

C. Protein Expression (day 1 & 2)

1. Remove ~250 mL of prepared media from Bio-Reactor tank and inoculate with a plated colony of Rosetta (DE3) cells previously transformed with both fibrillarin in pET28a and AK-tail Nop56/58 in pET21a. Shake at 37°C overnight.

2. Next morning turn on Bio-Flow 110, turn on agitation (320 rpms) and air flow (>5 Lpm), and warm up to 37°C. Inoculate with 250 mL overnight culture. Make sure that pH is set to auto, dO₂ is around 100%, and compressed air is flowing. The dO₂ levels will steadily decrease as the culture grows into exponential phase. It will drop to nearly 0% and then slowly begin to rise again as the culture growth slows. The OD₆₀₀ at this point should be around 1.2-1.5 OD. **NOTE:** If your Bio-Flow allows, set it to cascade the agitation from 300 up to 900 rpms to offset the drop in dO₂ during culture growth. The dO₂ will still decrease then rise, but the increased agitation will allow for longer culture time with greatly increased cell densities before growth slows. If not, maintain at highest/safest setting up to 900 rpms (ours currently maxes out at ~330-340 rpms due to a warped agitation shaft).

3. When the culture’s dO₂ levels reach near zero, attach water-cooling system and bring temperature of culture down to 12°C.

4. Induce protein expression by addition of 5 g glucose (0.05% final), 10 g lactose (0.1% final), and 0.47 g IPTG (0.2 mM final) for ~60 hours at 12°C.
5. Siphon-out culture into 0.5 L bottles and spin cells in GS-3 rotor (8,000xg, 10 min) to pellet.
6. Resuspend cells in 4 mL of Buffer D_M (20 mM Hepes, pH 7.2, 100 mM NaCl, 3 mM MgCl₂, 20% glycerol, 0.5 mM β-mercaptoethanol, 0.1 % Triton X-100) per gram of cell pellet.
7. Add PMSF to a final concentration of 1 mM, Benzonase Nuclease (Novagen, 90% purity) to 15 units/mL, and Lysozyme at ~1 mg per gram of cell paste. Rock lysate at 37°C for two hours.
8. Add 1/10 volume of 5 M NaCl (final [NaCl] = 0.6 M) and heat lysate to 85°C for 10 min. Pellet insoluble debris by spinning in Sorvall RC-5B for 1 hour at 10K rpms. Pour off soluble fraction into new tubes and proceed to purification or store at 4°C for several days until ready to purify.

D. Protein Purification

Two methods can be used for purification, either a simple gravity flow column or our low pressure liquid chromatography (LPLC) setup. Both protocols are shown below.

*Gravity-flow Column (day 5-6)*

1. Charge 50 mL of His-bind resin (Novagen) in a ~150 mL econo-column with 3 bed volumes (150 mL) 50 mM NiSO₄ and equilibrate with wash buffer (20 mM HEPES, pH 7.2, 0.6 M NaCl, 30 mM imidazole, 0.1% Triton X-100).
2. Bind soluble lysate to resin and wash with 600 mL wash buffer. Soluble lysate may need to be filtered first to remove any protein that has precipitated. Wash with 100 mL Buffer D₅₀₀ (Buffer D at 0.5 M NaCl instead of 100 mM).
3. Elute protein from resin with 150 mL of Buffer D₅₀₀ + 0.2 M imidazole. Eluate can be stored at room temperature overnight. Concentrate down to 20-30 mL and buffer exchange into Buffer D₅₀₀ at room temperature until imidazole concentration is reduced to ≤20 mM. Aliquot protein and flash-freeze with liquid nitrogen. Store samples at -80°C.
BioRad® BioLogic LPLC (day 5-6)

1. Pack ~150 mL econo-column with 50 mL His-bind Resin (Novagen).
2. Manually Charge resin with 50mM NiSO₄ (150mL). Manually Equilibrate resin with wash buffer (same as above) (~150 mL).
3. Make buffers:
   A. Wash Buffer – same as wash buffer for gravity-flow column (700 mL)
   B. Elution Buffer – same as elution buffer for gravity flow column (180 mL)
   C. Strip Buffer - 100 mM EDTA, pH 7.0 (180 mL)
   D. RNase free water (180 mL)
   E. 20% ethanol (180 mL)
4. Manually flush all LPLC lines with ddH₂O then place column on-line. Manually pump soluble lysate using separate pump system (Pharmacia Biotech Pump P-1) into the top inlet of the knob (injection port, only port not being used); this should send the soluble fraction to the column. Collect flow-through to save for analysis by moving the “waste” tubing to a clean collection beaker when the UV absorbance begins to rise significantly (above 0.2 OD). Alternatively, bind protein by gravity flow off-line and collect all flow-through.
5. Zero UV detector with elution buffer and run “DIMER WASH” method. Make sure to keep a small volume of buffer on top of the resin at all times during purification so that the resin does not go dry.
6. Turn on computer and run LP Data program. This synchronizes the LPLC and graphs absorbance levels over time or volume. Click “record” in the program to begin.
7. Allow program to run. It will wash column with 600 mL of wash buffer, elute with 150 mL of elution buffer, strip with 150 mL strip buffer, rinse with 150 mL of water and store in 20% ethanol. All these steps will be at a flow rate of 3.5 mL/min. During elution it will collect 8 mL fractions.
8. After eluate fractions have been collected, pool those that contain protein. When method is finished, take the column off-line and store in coldroom. Flush all tubing lines with ddH₂O. Eluate can be stored at room temperature overnight.

9. Concentrate down to 20-30 mL and buffer exchange into Buffer D_{500} at room temperature until imidazole concentration is reduced to \( \leq 20 \text{ mM} \). Aliquot protein and flash-freeze with liquid nitrogen. Store samples at -80°C.

E. Qualitative and Quantitative Analysis of Dimer Purification (day 6-7)

1. Resolve a few \( \mu \text{L} \) of insoluble and soluble lysate fractions, as well as column flow-through and eluate fractions, on a 12% SDS-PAGE (Figure 2a).

2. Check concentration by absorbance at 280 nm. Run a wavelength scan to ensure that no nucleic acids (peak at 260 nm) co-purified with the dimer (Figure 2b). Use \( A_{280} \) OD to estimate concentration using Beer’s Law:

\[
A = (E)(C)(b) = A_{280}
\]

\[
E = \text{extinction coefficient} = 74,630 \text{ M}^{-1} \text{ cm}^{-1}
\]

\[
C = \text{Molar concentration,}
\]

\[
b = \text{cell pathlength} = 1 \text{ cm}
\]

NOTE: The molecular weight of the His-fib:ΔK Nop dimer is \(~70,000\) daltons (g/mol). Its maximum solubility in Buffer D_{500} appears to be around 5 mg/mL.
FIGURE 2: Purification of the *M. jannaschii* ΔK Nop56/58:His-fibrillarin dimer. (A) Fractions collected throughout His-Resin purification of the dimer using the BioRad Biologic LPLC purification protocol above were resolved on a 12% SDS-PAGE and coomassie stained. frxn1-frxn13 represent eluate fractions. (B) Purified, concentrated and buffer-exchanged dimer was diluted 1/10 in Buffer D_{500} and a wavelength scan performed in a Beckman DU-64 spectrophotometer.