A TALE OF TWO METHYLATION MODIFICATIONS IN ARCHAEAL RNAs

Kunal Chatterjee
Southern Illinois University Carbondale, kunal@siu.edu

Follow this and additional works at: http://opensiuc.lib.siu.edu/dissertations

Recommended Citation
A TALE OF TWO METHYLATION MODIFICATIONS IN ARCHAEAL RNAs

By

Kunal Chatterjee

B.S., University of Calcutta, 2004
M.S., University of Calcutta, 2006

A Dissertation
Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Molecular Biology, Microbiology and Biochemistry
in the Graduate School
Southern Illinois University Carbondale
May, 2014.
DISSERTATION APPROVAL

A TALE OF TWO METHYLLATION MODIFICATIONS IN ARCHAEAL RNAs

By

Kunal Chatterjee

A Dissertation Submitted in Partial
Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

In the field of Molecular Biology, Microbiology and Biochemistry

Approved by:

Dr. Ramesh Gupta, Chair
Dr. Judith Davie
Dr. Rodney Weilbaecher
Dr. Douglas Fix
Dr. Matthew Geisler

Graduate School
Southern Illinois University Carbondale
7th. February 2014
AN ABSTRACT OF THE DISSERTATION OF

KUNAL CHATTERJEE, for the Doctor of Philosophy degree in Molecular biology, Microbiology and Biochemistry presented on February 7\textsuperscript{th}. 2014, at Southern Illinois University, Carbondale.

TITLE: A TALE OF TWO METHYLATION MODIFICATIONS IN ARCHAEAAL RNAs.

MAJOR PROFESSOR: Dr. Ramesh Gupta

In all the three domains of life, most RNAs undergo post transcriptional modifications both on the bases as well as the ribose sugars of the individual ribonucleotides. 2'-O-methylation of ribose sugars and isomerization of Uridines to Pseudouridines are two most predominant modifications in rRNAs and tRNAs across all domains of life. Besides 2'-O-methylation of ribose sugars, methylation of pseudouridine (Ψ) at position 54 of tRNA, producing m\textsuperscript{1}Ψ, is a hallmark of many archaeal species but the specific methylase involved in the formation of this modification had yet to be characterized. A comparative genomics analysis had previously identified COG1901 (DUF358), part of the SPOUT superfamily, as a candidate for this missing methylase family. To test this prediction, the COG1901 encoding gene, \textit{HVO}_1989, was deleted from the \textit{Haloferax volcanii} genome. Analyses of modified base contents indicated that while m\textsuperscript{1}Ψ was present in tRNA extracted from the wild-type strain, it was absent from tRNA extracted from the mutant strain. Expression of the gene encoding COG1901 from \textit{Halobacterium} sp. NRC-1, VNG1980C, complemented the m\textsuperscript{1}Ψ minus phenotype of the Δ\textit{HVO}_1989 strain. This in vivo validation was extended with in vitro tests. Using the COG1901 recombinant enzyme from \textit{Methanocaldococcus jannaschii} (Mj1640), purified
enzyme Pus10 from *M. jannaschii* and full-size tRNA transcripts or TΨ-arm (17-mer) fragments as substrates, the sequential pathway of m$^1$Ψ54 formation in Archaea was reconstituted. The methylation reaction is AdoMet-dependent. The efficiency of the methylase reaction depended on the identity of the residue at position 55 of the TΨ-loop. The presence of Ψ55 allowed the efficient conversion of Ψ54 to m$^1$Ψ54, whereas in the presence of C55 the reaction was rather inefficient and no methylation reaction occurred if a purine was present at this position. These results led to renaming the Archaeal COG1901 members as TrmY proteins.

Another aim of this study was to investigate the mechanism of target RNA recruitment to a box C/D sRNP. From data obtained, we have made the following hypothesis- aNop5p, either alone or as a heterodimer with Fibrillarin, binds to single stranded bulges and loops of target RNA. This aNop5p bound target is then hybridized to an assembling guide sRNP complex containing the guide RNA and L7Ae or guide RNA, L7Ae and aNop5p. If the guide:target sequences are complementary to each other, they hybridize and the target nucleotide gets modified. We also think that post modification, the guide and target strands separate, the core proteins rearrange themselves on the guide RNA and then prime the guide RNA for next round of modification.

Compared to the general archaeal populations, haloarchaea contain significantly fewer number of box C/D guide RNAs. In archaea, previous studies have underscored the importance of a symmetric assembly of the core proteins on the sRNA. This meant that if the core proteins were unable to bind to either the terminal box C/D or the internal box C'/D' motifs, the sRNP was not efficient to carry out the modification of the target
RNA. Essentially the only two haloarchaeal box C/D sRNPs known before had a symmetric architecture. In this study we discovered the first naturally occurring asymmetric box C/D sRNP called sR-41 in *Haloferax volcanii*. The architecture of *Haloferax volcanii* sR-41 box C/D sRNP seems to be closer in conformation to eukaryal snoRNPs (eukaryal counterparts of archaeal sRNPs) in which the core proteins assemble asymmetrically on the RNA. Till date, no information regarding the catalytic mechanism of an asymmetrically arranged eukaryal box C/D snoRNPs are available, because of unavailability of any assembly systems or crystal structures. Hence, this archaeal sR-41 guide sRNP provides a unique opportunity to study mechanism of modification in an asymmetrically arranged box C/D sRNP molecule.
DEDICATION

Dedicated to my parents.
ACKNOWLEDGEMENTS

First and foremost, I want to thank Dr. Ramesh Gupta for his invaluable guidance and support during my stay in his laboratory. I thank him for sharing his scientific knowledge and his expertise to make me the scientist that I am today. Other than science, personally too, I have learned a lot from him; I am a changed person for the better because of him. I thank him from the bottom of my heart. In this regard, I would also like to thank his family especially Shashi for making my stay in Carbondale more memorable.

I would like to thank my committee members, Dr. Rodney Weilbaecher, Dr. Douglas Fix and Dr. Matthew Geisler for their incredible feedbacks, comments, discussions and support during the course of my PhD. Very special thanks to Dr. Judith Davie for agreeing to be on my committee on such short notice. I would also like to thank Dr. Blaine Bartholomew for his time.

Thank you Stefani Hall and Cindy Filla for being the trouble shooters of the department. A very special thanks to Artemus Holguin for the “extreme lab-makeover” and helping Stefani and Cindy arrange all the potlucks and cake-cutting ceremonies that we love so much.

I would also like to thank Dr. Peter Hardwicke and Dr. Piroska Huvos for their constant words of encouragement. Also, thanks to Dr. Farid Kadyrov and Dr. Lyudmilla Kadyrova for their consistent help.

I really feel privileged to have worked with two of the best seniors in the lab one can hope for, Dr. Archi Joardar and Dr. Priyatansh Gurha. Thanks to the past members (Jiajun, Geena, Parinati, Mrinmoyee) and the present members (Sujata, Manisha,
Shaoni, Saakshi and Mike) for making the lab a fun place to work. They suffered me and my idiosyncrasies. Also, thanks to all the friends I had in the department during my stay here for their help.

Words cannot express my gratitude for my “adopted” family in Carbondale -Dr. Naushad Ali, Donna and Aiden.

Very special thanks to Gregory Sean Hummel for all his help, support and for preventing homelessness.

My stay in Carbondale has been incredible because of all my friends here. Avinash Prakash and Amrita Biswas-cannot thank them enough for their constant support. Big thanks to Julie Cosenza, Dr. Danielle Soto, Robert Carlsen, Christopher Higgerson II and Zack Phillips for always having my back.

I want to thank my sister Indrani Das and my brother-in-law Saurav Das for being there whenever I needed them.

And above all, thanks to Sunanda and Kamal Chatterjee, for literally being the best parents in this world.
PREFACE

This dissertation is divided into four Chapters. Chapter 1 provides a detailed background on the topics discussed in this dissertation. Figure 1.6 and Figure 1.7 are taken from the article titled “Natural history of S-adenosylmethionine-binding proteins” by P.Z Kozbial and A.R.Mushegian, published in BMC structural biology, volume 5, Page 19 (See Bibliography for further details). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chapter 2 discusses the discovery of a novel methyltransferase enzyme TrmY, which modifies a highly conserved $\Psi^{54}$ position of haloarchaeal tRNAs to $\text{m}_1\Psi^{54}$. This work has already been published (Chatterjee et al., 2012). This was a collaborative project. Dr. Ian K. Blaby from the laboratory of Dr. Valérie de Crécy-Lagard at the Department of Microbiology and Cell Science, University of Florida have created the $\Delta$TrmY and the $\Delta$TrmY + pHTrmY strains and verified both the strains with PCR (Figure 2.2 A). Dr. Blaby and Patrick Thiaville from the same laboratory performed the LC-MS on the Wild type *Haloferax volcanii* H26 and the $\Delta$TrmY strains (Figure 2.2 B). The bioinformatics work (Figure 2.5) was also performed at the University of Florida. Dr. Y Adam Yuan from the Department of Biological Sciences and Temasek Life Sciences Laboratory, National University of Singapore, Singapore, provided the MJ1640 (TrmY protein) plasmid for purification. Dr. Henry Grosjean from Université Paris, Orsay, France contributed to the critical writing of manuscript. Dr. Mrinmoyee Majumder
from Dr. Ramesh Gupta’s laboratory performed the Primer Extension reactions on CMCT-treated RNAs (Figure 2.3B). Purified Pus10 proteins were kindly provided by Archi Joardar from Dr. Ramesh Gupta’s laboratory.

Chapter 3 extensively describes the dynamics of GuideRNA mediated modification of archaeal tRNAs. We employed limited Lead (II)-mediated cleavage reactions to probe the conformational changes of both guide and target RNAs during box C/D mediated catalysis. pHVMΔI36 plasmid was made in Dr. Ramesh Gupta’s laboratory by Dr. Ignatius Gomes (Gomes and Gupta, 1997). Figure 3.1 has been taken from the article titled “Dynamic guide-target interactions contribute to sequential 2’-O-methylations by a unique archaeal dual guide box C/D sRNP” by Dr. Sanjay Singh, Dr. Priyatansh Gurha and Dr. Ramesh Gupta, published in RNA, 2008 (Singh et al., 2008) (See Bibliography for more details). This article is also distributed under the terms of the Creative Commons Attribution License.

Chapter 4 describes the process by which we identified the third box C/D sRNA in Haloferax volcanii. ΔFibrillarin strains used in Figure 4.1 and Figure 4.2B was provided by Parinati Kharel from Dr. Ramesh Gupta’s laboratory.

The rest of the work (other than the ones mentioned above) was performed by the author of this dissertation in Dr. Ramesh Gupta’s laboratory under his guidance and supervision.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>PREFACE</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xix</td>
</tr>
</tbody>
</table>

## CHAPTERS

### CHAPTER 1 – Literature Review

1.1 Most RNAs undergo post-transcriptional modifications ......................... 1
1.2 Importance of methylation modifications in RNAs ........................................ 2
1.3 RNA dependent modifications of RNAs ......................................................... 4

#### 1.3.1 Box C/D RNA

1.3.1.1 K-Turn and K-Loop motifs .................................................................. 8

1.3.2 Box H/ACA RNA ................................................................................... 10

1.4 Box C/D sRNP/snoRNPs ........................................................................... 11

1.5 Proteins associated with box C/D sRNP/snoRNPs ...................................... 16

1.5.1 L7Ae/15.5kD/Snu13p ............................................................................ 17
1.5.2 aNop5p/Nop56/Nop58 ........................................................................... 20
1.5.3 Fibrillarin ............................................................................................... 26

1.6 Guide RNA independent methylation modifications in archaea ................. 29
2.3 Results

2.3.1 *trmY* is not essential in *H. volcanii* but deletion of *trmY* leads to absence of m$^1$ψ$^5$4 in *H. volcanii* tRNA

2.3.2 *M. jannaschii* TrmY catalyzes the formation of m$^1$ψ$^5$4 in tRNA

*in vitro*

2.3.3 *M. jannaschii* TrmY can produce m$^1$ψ in just the Tψ-arm

2.3.4 *M. jannaschii* TrmY requires ψ55 in tRNA to produce m$^1$ψ$^5$4

2.4 Discussion.

CHAPTER 3 – Dynamics of box C/D mediated modification

3.1 Introduction

3.2 Materials and Methods

3.2.1 Generation of DNA templates for *in vitro* RNA synthesis

3.2.2 Site directed mutagenesis

3.3.3 *In vitro* RNA synthesis

3.2.4 3' end labeling of RNA

3.2.5 5' end labeling of RNA

3.2.6 *In vitro* RNP-directed nucleotide 2'-O-Methylation and TLC

3.2.7 Purification of His-Tagged *M. jannaschii* L7Ae and aFibrillarin

3.2.8 Purification of *M. jannaschii* aNop5p
3.2.9 Lead (II) mediated footprinting of RNP complexes............................... 84

3.2.10 Iodine induced cleavage of 5'end labeled CTPαS incorporated RNA in denatured condition................................................................. 86

3.2.11 Electrophoretic Mobility shift assay (EMSA)...................................... 86

3.3 Results and Discussions.............................................................................. 87

3.3.1 In vitro functional analyses of box C/D sRNP ................................. 87

3.3.2 sR-tMet 93mer forms stable complexes with Methanocaldococcus jannaschii recombinant box C/D core proteins in solution ................. 91

3.3.3 Structural analyses of an in vitro assembled box C/D sRNP

by lead (II) induced cleavage RNA footprinting technique..................... 95

3.3.4 Mechanism of target recruitment by box C/D sRNPs..................... 101

3.3.5 aNop5p recruits target RNA by recognizing single stranded

loops and bulges in the RNA................................................................. 108

3.3.6 Importance of GAEK and ALFA motifs in aNop5p recognition...... 120

3.3.7 Box C/D sRNPs exhibit multiple turnovers................................. 122

3.4 Conclusions and Perspectives................................................................. 126

3.4.1 Box C/D core protein aNop5p recruits target to an assembling

box sRNP .............................................................................................. 127

3.4.2 aNop5p recognizes single stranded bulges and loops in RNA..... 128
3.4.3 The conserved ALFA but not the GAEK motif is essential in Box C/D mediated catalysis................................................................. 130

3.4.4 Other two box C/D core proteins L7Ae and aFibrillarin have differential effects of aNop5p mediated target recruitment. ........ 130

3.4.5 Guide-Target unpairing is essential for box C/D sRNP mediated multiple turnover reactions................................................. 132

CHAPTER 4 – Discovery of the RNA that mediates 2’-O-methylation of G1934 residue of Haloferax volcanii 23s rRNA ........................................ 134

4.1 Introduction................................................................................................................................. 134

4.2 Materials and Methods ............................................................................................................ 138

4.2.1 Growth of Haloferax volcanii cells.......................................................... 138

4.2.2 Isolation of total RNA from Haloferax volcanii cells......................... 138

4.2.3 Primer extension using limited dNTP ...................................................... 138

4.2.4 Northern Hybridization ................................................................................. 138

4.2.5 Generation of DNA templates for in vitro RNA synthesis............. 141

4.2.6 In vitro RNA synthesis.............................................................................. 142

4.2.7 In vitro RNP-directed nucleotide 2’-O-methylation and TLC .......... 143

4.2.8 Electrophoretic Mobility shift assay (EMSA) ....................................... 143

4.2.9 Lead (II) mediated footprinting of RNP complexes......................... 143

4.3 Results............................................................................................................................... 144
4.3.1 G1934 residue of 23S rRNA in Haloferax volcanii is 2’-O-methylated

4.3.2 Identification of the sRNA predicted to modify Gm1934 of 23S rRNA of *Haloferax volcanii*

4.3.3 Only the D’guide of sR-41 is functional

4.3.4 Haloferax sR-41 box C’/D’ does not fold into either a K-Turn or K-Loop

4.4 Discussion

4.4.1 sR-41, the third guide box C/D RNA in *Haloferax volcanii*

4.4.2 Only the D’ guide of sR-41 is functional

4.4.3 Asymmetrically assembled sR-41 is still functional

BIBLIOGRAPHY

VITA
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1: Box C/D and box H/ACA s(no) RNAs and their core proteins</td>
<td>13</td>
</tr>
<tr>
<td>Table 2.1: Identification of $m^1\Psi$ and $s^2T$ ($s^3m^5U$) in archaeal tRNAs</td>
<td>42</td>
</tr>
<tr>
<td>Table 2.2: Strains used in Chapter 2</td>
<td>50</td>
</tr>
<tr>
<td>Table 2.3: Plasmids used in Chapter 2</td>
<td>50</td>
</tr>
<tr>
<td>Table 2.4: Oligonucleotides used in Chapter 2</td>
<td>51</td>
</tr>
<tr>
<td>Table 3.1: List of RNA substrates and Primers used in Chapter 3</td>
<td>77</td>
</tr>
<tr>
<td>Table 3.2: Protein Mutants, DNA templates and Primers used in Chapter 3</td>
<td>79</td>
</tr>
<tr>
<td>Table 4.1: RNA Substrates, DNA templates and Primers used in Chapter 4</td>
<td>141</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1.1: The two most predominant methylation modifications in archaeal RNAs</td>
<td>2</td>
</tr>
<tr>
<td>Figure 1.2: The two guide RNA modification families</td>
<td>6</td>
</tr>
<tr>
<td>Figure 1.3: Comparison between archaeal and eukaryal box</td>
<td>14</td>
</tr>
<tr>
<td>C/D s(no)RNP architecture</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.4: Proposed architecture of di-sRNP model</td>
<td>16</td>
</tr>
<tr>
<td>Figure 1.5: Cartoon representation of aNop5p structure</td>
<td>22</td>
</tr>
<tr>
<td>Figure 1.6: Topology and Cartoon representation of Rossman fold</td>
<td>31</td>
</tr>
<tr>
<td>Methyltransferases</td>
<td>31</td>
</tr>
<tr>
<td>Figure 1.7: Topology and Cartoon representation of SPOUT Methyltransferases</td>
<td>33</td>
</tr>
<tr>
<td>Figure 1.8: Chemical structure of SAM</td>
<td>34</td>
</tr>
<tr>
<td>Figure 2.1: Enzymatic posttranscriptional modifications of uridines in RNA</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.2: Construction and phenotype of the trmY-deleted strain of <em>H. volcanii</em></td>
<td>52</td>
</tr>
<tr>
<td>Figure 2.3: Formation of m¹Ψ of <em>H. volcanii</em> tRNA is mediated by TrmY</td>
<td>56</td>
</tr>
<tr>
<td>Figure 2.4: <em>M. jannaschii</em> TrmY can convert Ψ54 of tRNA to m¹Ψ54</td>
<td>58</td>
</tr>
<tr>
<td>Figure 2.5: Phylogeny of the COG1901 family and structural comparison of Mj1640 dimer with VV_1434 dimer</td>
<td>60</td>
</tr>
<tr>
<td>Figure 3.1: A proposed mechanism for the assembly of a functional archaeal box</td>
<td>69</td>
</tr>
<tr>
<td>C/D sRNP complex</td>
<td>69</td>
</tr>
<tr>
<td>Figure 3.2: Ribbon diagram of the C-Terminal domain of aNop5p superimposed on guide RNA</td>
<td>70</td>
</tr>
<tr>
<td>Figure 3.3: Partial multiple Sequence Alignment of Nop5p with representative archaeal and eukaryal homologs</td>
<td>72</td>
</tr>
</tbody>
</table>
Figure 3.4: sR-tMet 93mer and HVM\textsubscript{i36} can act as a model guide-target system to study box C/D mediated catalysis .............................................. 89

Figure 3.5: EMSA of sR-tMet .................................................................................................................. 92

Figure 3.6: Architecture of a single guide sRNP sR-tMet 73mer ......................................................... 97

Figure 3.7: Lane profile analyses of Lead (II)-induced cleavage of assembly reactions of sR-tMet 93 mer ......................................................................................................................... 101

Figure 3.8: Lead (II)-induced cleavage RNA footprinting of pre-tRNA\textsubscript{Met}\textsubscript{i36}i .............. 106

Figure 3.9: HVM\textsubscript{i36} (Target) RNA interactions with the core proteins in presence of guide RNA sR-tMet ..................................................................................................................... 110

Figure 3.10: Titration of aNop5p with 5’end labeled pre-tRNA\textsubscript{Trp}\textsubscript{67} RNA ................... 112

Figure 3.11: A) Sequence and secondary structure of HVM\textsubscript{i36} showing regions of RNA bound by aNop5p ................................................................. 114

B) Sequence and predicted secondary structures of pre-tRNA\textsubscript{Met}, the 3’ stem loop of sR-h45 and mature tRNA\textsubscript{Met} ................................................................. 114

Figure 3.12: Sequence and predicted secondary structure of substrates derived from HVM\textsubscript{i36} .......................................................................................................................... 115

Figure 3.13: A) Modification assays of substrates derived from HVM\textsubscript{i36} .......................... 118

B) Single Turnover kinetics of aNop5p substrates ............................................................................ 118

Figure 3.14: Mutating GAEK and ALFA motifs in aNop5p have differential effects on box C/D sRNP mediated catalysis ................................................................. 120

Figure 3.15: Box C/D sRNPs exhibit multiple turnover kinetics ......................................................... 123

Figure 3.16: Conformational changes in the single guide sRNA sR-tMet in
real time during methylation reactions .................................................. 125

Figure 4.1: Autoradiograms showing primer extension analysis to detect
2-O-methyl modification at position G1934 of Haloferax volcanii

23S rRNA ........................................................................................................... 146

Figure 4.2: sR-41 is present in Haloferax volcanii ............................................ 149

Figure 4.3: D’ guide region of sR-41 can modify target RNA.............................. 150

Figure 4.4: Electrophoretic Gel Mobility Shift Assays of sR-41 with increasing
concentrations of L7Ae ..................................................................................... 152

Figure 4.5: L7Ae has affinity only for the sole K-Turn present in sR-41 box C/D

sRNA ................................................................................................................ 154
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Ammonium Persulfate</td>
</tr>
<tr>
<td>AdoMet</td>
<td>S-Adenosyl Methionine (Also referred to as SAM in this work)</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>BME</td>
<td>β-Mercaptoethanol</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>CMCT</td>
<td>1-Cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethylpyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine Tetraacetic Acid</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic Acid</td>
</tr>
<tr>
<td>h</td>
<td>Hour</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilo Dalton</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>min</td>
<td>Minute</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>nM</td>
<td>nanomolar</td>
</tr>
</tbody>
</table>

**Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenylmethyl Sulfonyl Fluoride</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SAM</td>
<td>S-Adenosyl Methionine (Also referred to as AdoMet in this work)</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>sec</td>
<td>Second</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-Acetate-EDTA</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-Borate-EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N’,N’-Tetramethylethylenediamine</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin Layer Chromatography</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris[hydroxymethyl]aminomethane</td>
</tr>
</tbody>
</table>
tRNA     Transfer RNA
CHAPTER 1
LITERATURE REVIEW

1.1 Most RNAs undergo post-transcriptional modifications

The central dogma of molecular biology suggests that DNA makes up the genetic information, while RNA is just a passive carrier whose classical function seems only to decode this information into polypeptides. However, in the last quarter century knowledge about various kinds of RNA have challenged this view and have shed light on the vast repertoire of cellular functions that are RNA dependent or influenced. The presence of stable “non (protein)-coding” RNAs across all three domains of life underscores the importance of RNA in processes other than protein anabolism (Bachellerie et al., 2002; Cavaille et al., 2000; Dieci et al., 2009; Eddy, 2001; Terns and Terns, 2002; Williams and Farzaneh, ; Williams and Farzaneh, 2012). The abundance and variety of these non-coding RNAs and their involvement in an array of regulatory functions ranging from gene silencing to genomic imprinting have led to the hypothesis of an “RNA world” that may have preceded the evolution of DNA and proteins.

In all the three domains of life, most RNAs undergo post transcriptional modifications both on the bases as well as the ribose sugars of the individual ribonucleotides (Agris, 1996; Limbach et al., 1994). These RNA modifications can be broadly classified into four groups:

i) Isomerization of Uridine to Pseudouridine.

ii) Methylation of 2’-OH group of Ribose sugars.

iii) Chemical alterations to bases, such as addition of methyl groups into base ring Nitrogen or Carbon
iv) Hypermodifications or multiple modifications of bases.

In the present study the ramifications of methylation modifications of both the ribose sugars as well as the base Nitrogen of ribonucleotides (Figure 1.1) have been explored.

1.2 Importance of Methylation modifications in RNA

2′-O-methylation of ribose sugars is one of the most predominant modifications in rRNAs and tRNAs across all kingdoms of life. Extensive studies using nuclear magnetic resonance spectroscopy on nucleoside and oligonucleotide models have demonstrated that 2′-O-methylation results in thermodynamic stabilization of the C3′-endo sugar conformation in both pyrimidines and purines (Clouet-d'Orval et al., 2005). This ribose modification stabilizes the A-type helical conformation found in RNA, resulting in enhanced regional rigidity and higher Tm. As such, ribose methylation is thought to
prevent phosphodiester bond hydrolysis (which requires a free 2′-hydroxyl) at high temperatures as well as prevent adventitious nuclease cleavages associated with an increased population of unfolded RNA structures at higher temperatures (Noon et al., 1998).

Methyl modification of bases dramatically alters the hydrogen bonding capabilities of the parent ribonucleotides (Basti et al., 1996). It is generally acknowledged that methylated pyrimidines confer stability to RNA helices by increasing base-stacking interactions. Studies have also proven that this stability observed by the so called “methyl effect” is independent of either any hydrophobic effect conferred by the introduction of the methyl group or any effect of the methyl group on the ribose sugars (Benne, 1998).

Even though some of the positions of methylation modifications in bacteria, eukarya and archaea are conserved, the pathways by which these methylation modifications are carried out are quite different for bacteria compared to that of eukarya and archaea (Bachellerie et al., 2002; Decatur and Fournier, 2003; Gaspin et al., 2000; Kiss, 2002; Omer et al., 2000). Methylation modifications in bacteria are carried out by standalone enzymes which recognize specific RNA sequences or RNA tertiary structures. In contrast, methylation modifications in archaea and eukarya are carried out by a combination of site or sequence specific proteins as well as RNA-Protein complexes. These RNA-Protein complexes display remarkable division of labor, where the RNA partner “guides” the catalytic protein partner to its cognate target by forming a complementary base pairing with the target site.
Other than these “guided” modification of Ribose sugars, extensive Modomics studies, which are analyses on patterns of RNA modifications across organisms, have unraveled some stand alone protein enzymes which catalyze base modifications in archaea and their homologs in eukarya. Many of these predicted methylations have been verified by biochemical procedures like Primer extensions in presence of limited dNTPs (Maden et al., 1995), which stall primer progression at methylated sites on ribose sugars and presence of some these predicted methylated sites await further experimental validations. In this section, we will limit our discussion to RNA guided methylation modifications of 2'-OH of RNA first followed by RNA independent methylation of ψ ribonucleotides in Archaea.

1.3 RNA dependent modifications of RNAs

Various non-coding RNAs are involved in chemical modifications of bases and ribose sugars in RNA, processing of RNA, synthesis of telomeric repeats and are conserved in archaea and eukaryotes. In eukaryotes, these RNAs are called small nucleolar RNAs (snoRNAs) as they reside in the nucleolus, while their prokaryotic counterparts in archaea are called small nucleolar RNA like RNA (sRNAs). A family of proteins homologous in eukarya and archaea associate with these RNA to form a functional small RiboNucleoProtein complex (sRNPs). Besides their location in the nucleoli, the main criteria by which snoRNAs are identified are their salt resistant, nucleoli like extraction properties, association with rRNAs or nucleolar proteins and most importantly the presence of conserved sequence elements in the snoRNAs (Maxwell and Fournier, 1995).
Based on the sequence of these short consensus motifs, the sRNA component of the sRNPs can be classified into two major groups: box C/D and box H/ACA snoRNAs (Figure 1.2). Members of these two snoRNAs families guide the 2'-O- ribose methylations and pseudouridylations respectively, of RNAs. Some of them are also involved in pre-rRNA and mRNA processing events. Both ribose sugar methylations and pseudouridylations are very widespread in Eukarya and Archaea, numbering over 50-100 of each modification in their mature rRNA populations (Bachellerie et al., 2002). However, in bacteria, the numbers of modified nucleotides are far fewer in comparison to eukarya and archaea. Moreover, these modifications in bacteria are carried out by protein only enzymes, as opposed to sRNP complexes in their Eukaryal and Archaeal counterparts. Since, these classes of nucleoside-modification complexes in archaea share the common concept of protein catalyst being guided by RNA to its precise target location, it is hypothesized that the core elements of these complexes originated from a progenitor RNP (Tran et al., 2004). Also, since some of the members of these core sRNP complexes are located in the ribosomes, it was hypothesized that the origin of the progenitor RNP lies in the primitive translation apparatus (Rozhdestvensky et al., 2003).

In addition to their recognized roles in RNA modification and splicing, recent studies have unraveled a plethora of functions that the snoRNPs perform. For example, small nucleolar box C/D sRNAs U32a, U33, and U35a are found to be critical mediators of lipotoxic metabolic stress in mammalian cells (Michel et al., 2011). A number of human box C/D and box H/ACA snoRNAs with microRNA (miRNA)-like processing signatures
Figure 1.2: The two guide RNA modification families. Left: Schematics of a typical dual guide sno/sRNA with two target RNAs hybridizing to the guide sequences upstream of box D and D’. Note that it is nucleotide on the target RNA that is base paired to the complementary nucleotide five bases upstream of box D and box D’ in the guide that gets methylated. Box C/D motif is located at the terminus, whereas box C’/D’ motif is located internally in the guide RNA. Both the box C/D and C’/D’ motifs are represented as orange boxes, with the consensus sequences written in the boxes. Right: Schematics of a typical dual guide box H/ACA sno/sRNA hybridized to one target RNA, the later interacting with the former in the bulged out pseudouridylation pocket. The two helix-loop-helix RNA hairpins are joined by the Hinge box or Box H, represented as the Yellow box, with its consensus sequence. The ACA box, also represented as a yellow box, is typically present at the 3’end. Notice the bipartite pairing between guide-target RNAs around the target Uridine which gets modified to Pseudouridine.

are involved in the down-regulation of gene expression (Brameier et al.; Scott et al., 2009). SnoRNAs like box C/D HBII-52 (Kishore and Stamm, 2006) and the brain specific
brain-specific box C/D small RNA, MBII-52, are also shown to regulate alternate splicing and RNA editing respectively (Doe et al., 2009).

1.3.1 Box C/D RNA

Both the Eukaryal box C/D snoRNAs and Archaeal box C/D sRNAs are defined by the presence of a pair of conserved sequence motifs called box C and box D (Maxwell, 1995). Box C sequence (RUGAUGA) and box D sequences (RCUGA) are conserved across species and are located at the 5’end and the 3’end of the RNA. Some of the snoRNAs/sRNAs have an additional pair of internally located, but less conserved sequences called box C’ and box D’ (Reichow et al., 2007). A stretch of 10-21 nucleotide sequences upstream of box D and box D’ are complementary to the target and it is these sequences which determine the specific nucleotides that the particular guide RNA targets. As a rule, it is always the nucleotide on the target RNA that is base paired with the complementary nucleotide five bases upstream of box D and box D’ of the guide that gets modified by box C/D snoRNPs (N+5 Rule) (Kiss-Laszlo et al., 1996; Kiss, 2002; Reichow et al., 2007).

Comparisons of the sequences of eukaryal snoRNAs with the archaeal sRNAs reveal that snoRNAs are in general bigger than sRNAs (Omer et al., 2000). Majority of sRNAs are actually “dual-guide” sRNAs; they are able to direct modifications from both their guide sequences (Terns and Terns, 2002). Only a fifth of all snoRNAs known are able to guide ribose methylation from both the guide sequences (Speckmann et al., 2002). Also for the eukaryal snoRNAs, the targets are often present on two different molecules in contrast to archaeal sRNAs, where both the targets are present on the same target molecules (Dennis and Omer, 2005). The spacing between the two motifs
in sRNAs is highly conserved and initial analyses of archaeal sRNA sequences have shown these distance should be at least 12 bases (Tran et al., 2005), although spacer lengths of less than 12 bases are methylation competent but with lower efficiencies. This requirement for this intra-motif spacing is less stringent in eukaryotes. snoRNAs with bigger spacer lengths often fold into evolutionary conserved secondary structures which reduce the inter-RNP distance to the optimal length, a kind of “spatio-functional” coupling (Qu et al., 2010).

The conserved box C/D sequences play important roles in myriad cellular functions. U75 snoRNA synthesis in vitro depends on its box C and D sequences and requires an appropriate spacer length (Hirose and Steitz, 2001). Mutational studies reveal that the first GA bases of the box C sequence UGAUGA are essential for U14 sRNA stability in *Saccharomyces cerevisiae* (Huang et al., 1992). Moreover, these consensus sequences are also important for proper localization of these RNAs as nucleolar localization was severely disrupted by targeting even one or two of these consensus bases (Lange et al., 1998).

### 1.3.1.1 Kink Turn and K-Loop motifs

Another hallmark of the box C/D sRNAs is the presence of the Kink turn and the K-Loop motifs. The K-Turn is a widespread RNA structural motif characterized by two helices flanking a trinucleotide bulge whose nucleic acid chains’ phosphodiester backbone causes a 60° sharp turn in the RNA helix (Klein et al., 2001). This double-stranded, helix–loop–helix motif is comprised of approximately 15 nucleotides. The first helical stem ends at the internal loop with two Watson–Crick base pairs, typically C–Gs
and is called the 'canonical stem' or 'C-stem'. The second helical stem, the 'non-canonical stem' or 'NC-stem', follows the internal loop and starts with two non-Watson–Crick base pairs, typically sheared G–A base pairs. The internal loop between the helical stems is always asymmetrical and usually has three unpaired nucleotides on one strand and none on the other. The 5'-most nucleotide of the loop stacks on the C-stem, the second extends to stack on the NC-stem, and the third, which is mostly a highly conserved Uridine protrudes into solution (Klein et al., 2001). Although substitutions can be tolerated for most of these positions, K-turns that are close to the consensus sequences are the most stable (Daldrop and Lilley, 2013). Because of the kink in the phosphodiester backbone in this strand, the orientation of the axes of the C-stem and the NC-stem differ by 60° (Schroeder et al., 2011; Schroeder et al., 2010). The K-turn RNA exists in a dynamic equilibrium between a tightly kinked conformation and a more open structure similar to a simple bulge. Moreover, crystal structure of Azoarcus group I intron revealed a reverse kink turn that bent in the direction opposite to that of a consensus Kink-turn, showing a remarkable plasticity of this structure (Antonioli et al., 2010). The highly kinked form is stabilized by the non cooperative binding of divalent metal ions like magnesium (Goody et al., 2004). The K-turn is stabilized by multiple contacts between bases and sugars in NC-stem and internal loop, and interactions between the C-stem and the NC-stem (Klein et al., 2001; Vidovic et al., 2000). This bend or kink has huge physiological importance as it forms a platform for protein binding. The wider major groove of the C-Stem, flat minor groove of the NC-Stem, the protruding uridine and the enhanced planes of the bases are the structural cues that most proteins recognize for binding (Winkler et al., 2001).
Other than the K –Turn, box C/D sRNAs also have another distinct RNA fold called K-Loop. Like the K-turn, this RNA motif contains the non-canonical stem, but unlike the K-turn, the canonical stem is replaced by a short terminal loop, hence the name K-loop (Nolivos et al., 2005). However, because of the lack of the canonical stem, K-loops do not bend the RNA or cause any kinks in them (Nolivos et al., 2005). Like the K-turn, the K-loop motifs, which are present in the internal regions of the box C/D sRNA, form an independent RNA binding module and is thought to bring about stability to the RNAs by eliminating free ends (Nolivos et al., 2005). Although most archaeal box C/D sRNAs contain a K-loop at the C'/D' motif, intron of pre-tRNA^{Trp} and the guide RNA for methylating C34 of pre-tRNA^{Met}, called sR-tMet are the only two known haloarchaeal box C/D sRNAs which are devoid of this motif (Joardar et al., 2012).

1.3.2 Box H/ACA RNAs

The other class of RNA guide modification families is made up by box H/ACA RNAs. They together with their cognate set of proteins isomerizes specific uridines to Pseudouridine, which one of the most abundant modifications present in RNA (Dennis and Omer, 2005; Dennis et al., 2001; Henras et al., 2004; Kiss et al., 2010). Box H/ACA RNAs were first discovered in the nuclei of eukaryotes, but subsequently was found in archaea and in the Cajal body of the eukaryotic nucleoli. Box H/ACA RNAs carry out this specific isomeric modification in a variety of RNAs like rRNAs in both eukaryotes and archaea (Kiss et al., 2010) as well as splice leader RNAs in Trypanosomes (Liang et al., 2002). Box H/ACA RNAs are characterized by their unique stem loop structures, which may be single or triple stem loops as in the case of archaea or more
characteristically double stem loop structures as in the case of eukaryotes (Matera et al., 2007). Box H/ACA RNAs derive their name from the consensus H box or Hinge region present between the two stem loops and also the consensus ACA motif present at the 3’end of last stem loop RNA (Omer et al., 2000; Terns and Terns, 2006). Each stem-loop is interrupted by an internal bulge that is complementary in sequence to nucleotides flanking the pseudouridylation site in a substrate RNA (this bulge is also known as the Ψ pocket). This complementarity is sufficient for targeting the box H/ACA RNP to its’ correct cellular substrate.

Some box H/ACA RNPs have functions unrelated to Ψ (Meier, 2005; Terns and Terns, 2006). For example, the box H/ACA guide RNA snR30 (yeast nomenclature) is required for cleavage of the 35 S precursor to 18 S rRNA but is not known to introduce Ψ to any cellular RNA. Vertebrate telomerase RNA contains an H/ACA domain that is important for telomerase RNP assembly and activity. This domain is also not known to introduce Ψ into any cellular RNA.

1.4 Box C/D sRNP/snoRNP

Both the families of the guide RNAs assemble a set of cognate proteins which essentially provide the catalytic functions of the guide sRNP molecule. Due to the difficulties faced in the development of a suitable assembly systems for eukaryotic snoRNP complexes, the majority of knowledge regarding snoRNP architecture have been derived from high resolution structures of archaeal box C/D and box H/ACA complexes. Also, the fact that archaeal proteins can be expressed and purified relatively easily in Escherichia coli led to the widespread use of Archaeal assembly
systems to mimic simpler versions of more complex Eukaryal snoRNP systems (Bortolin et al., 2003; Charpentier et al., 2005; Omer et al., 2000; Rashid et al., 2003; Singh et al., 2008; Tran et al., 2005). The components of box C/D sRNPs are listed in Table 1.1.

An archaeal box C/D sRNA first binds protein L7Ae on both the C/D box as well as C’/D’ boxes in a cooperative manner, which exposes the corresponding guide sequences (Singh et al., 2008). This is followed by binding of the second protein aNop5p alone or aNop5p protein as a dimer with aFibrillarin protein, the last protein being the putative methyltransferase. Thus stoichiometrically, each archaeal box C/D sRNP consists of one copy of the box C/D guide RNA with two copies of the core proteins at its two consensus motifs making the entire sRNP complex very symmetric. However, studies have shown that a “hemi-complex” containing only one of the guide sequences with one set of core proteins are also able to methylate target RNAs, albeit with a much lower competency (Hardin and Batey, 2006).

The molecular architecture of eukaryal snoRNAs have been difficult to elucidate due to non-availability of any functional in vitro reconstituted systems using purified eukaryal sRNAs and proteins (Jady and Kiss, 2001). However, some in vitro modification and cross linking studies using eukaryal cell extracts (Galardi et al., 2002; Wang et al., 2002) immunoprecipitated snoRNPs (Jady and Kiss, 2001; Patton, 1991) and Xenopus oocyte microinjection (Zhao et al., 2002) have been performed. These studies reveal that the architecture of box C/D s(no)RNPs differs between archaeb and eukaryotes. In an eukaryotic box C/D snoRNA, the homolog of L7Ae protein, the 15.5 kD protein recognizes only the K-turn of the terminal box C/D motif and in vitro it is
Table 1.1: box C/D and box H/ACA s(no) RNAs and their core proteins

<table>
<thead>
<tr>
<th>Core Proteins</th>
<th>Box C/D snoRNPs</th>
<th>Box H/ACA snoRNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukarya</td>
<td>Archaea</td>
<td>Eukarya</td>
</tr>
<tr>
<td>Nop1p (Fibrillarin)</td>
<td>Nop56p</td>
<td>Nop58p</td>
</tr>
<tr>
<td>Nop56p</td>
<td>Fibrillarin</td>
<td>Gar1p</td>
</tr>
<tr>
<td>Nop58p</td>
<td>aNop5p</td>
<td>Nop10p</td>
</tr>
<tr>
<td>Snu13p(15.5 kD)</td>
<td>L7Ae</td>
<td>Nhp2</td>
</tr>
<tr>
<td>RNA Targets</td>
<td>rRNA, snoRNA</td>
<td>rRNA, tRNAs</td>
</tr>
</tbody>
</table>

unable to recognize K-loop motifs (Biswas et al., 2011; Charron et al., 2004; Gagnon et al., 2010). Also, instead of a single aNop5p protein, eukaryotes have two orthologs of aNop5p protein called Nop56 and Nop58 which show differential binding specificities to the two motifs of box C/D RNA (Cahill et al., 2002). Cross linking studies have shown that whereas Nop56 protein interacts with the C’ box and Nop58 protein binds the C box. However, the asymmetric binding of the two Nop proteins does not hinder the ability of the two antisense RNAs to guide their respective methylations.
Fibrillarin was the only protein that was shown to interact with both box D and box D’ (Cahill et al., 2002). Thus, in contrast to an archaeal box C/D sRNPs, eukaryotic box C/D sRNPs are asymmetric in structure.

**Figure 1.3. Comparison between archaeal and eukaryal box C/D s(no)RNP architecture.** Left: Archaeal box C/D sRNP has one copy each of the core proteins binding to the consensus box C/D and box C’/D’ motifs, making the structure “symmetric”. The consensus motifs in the RNA are represented as gray boxes and the core proteins have the following color scheme: L7Ae in purple, aNop5p in Fuchsia and aFibrillarin in Yellow. **Right**: In contrast, eukaryal box C/D snoRNP is essentially “asymmetric” in structure with differential binding of the core proteins to box C/D and box C’/D’. The eukaryal sno RNA core proteins are represented by the following colors: 15.5 kD protein in Cyan, Nop56 in Fuchsia, Nop58 in violet and Fibrillarin in Yellow.
As pointed out before, unlike eukaryotic snoRNP, archaeal sRNPs can be reconstituted in vitro and are catalytically active (Omer et al., 2002). Recent studies with such in vitro reconstituted archaeal box C/D sRNP from Methanococcus jannaschii (Mj) has been shown by glycerol gradient sedimentation, gel filtration chromatography, native gel analysis and single-particle electron microscopy (EM) to adopt a di-sRNP architecture, containing four copies of each box C/D core protein and two copies of the sRNA (Bleichert et al., 2009), that are held together by the coiled coil domain of aNop5p proteins. Deletion of this coiled coil domain results in the assembly of a catalytically inactive mono-sRNP (Bleichert et al., 2009; Zhang et al., 2006). This di-sRNP structure is catalytically active and is conserved across the domain archaea (Bower-Phipps et al., 2012). These studies show that the preservation of all the consensus sequences as well as the K-turn/K-Loop architecture of the box C/D sRNAs are absolutely indispensable for the formation of catalytically active di-sRNP which connects the importance of RNA structure with its function (Bleichert and Baserga, 2010). Moreover, internal loops, omnipresent in most physiologically important box C/D sRNAs, predominantly form di-sRNPs, whereas in mutant RNA molecules where the internal loops are deleted leading to the formation of two independent stem loop or “two piece” RNAs, form mono-sRNPs (Bower-Phipps et al., 2012). Further studies using crystal structure of fully assembled archaeal sRNP complexes support the di-sRNP model and propose a novel cross-catalysis mechanism suggesting that the target nucleotide is modified by Fibrillarin from the other half-mer RNP (Xue et al., 2012). A recent study has suggested that a dual guide di-sRNP is capable of binding more than one substrate molecules simultaneously.
The same group also speculates that a fully formed catalytic sRNP is probably recycled by removal of the products post modification.

Figure 1.4. Proposed architecture of di-sRNP model. In this model, two independent sRNPs are joined to each other via the coiled coil domain of aNop5p. This model envisages a functional sRNP as a homodimer of two sRNAs and four sets of box C/D core proteins binding to the consensus motifs of individual sRNAs. L7Ae is shown in Violet, aNop5P is represented by Fuchsia and Fibrillarin in Yellow.

1.5 Proteins associated with box C/D sRNPs/snoRNPs

Methylation competent box C/D sRNPs/snoRNPs employ a division of labor in which the guide RNA directs the proteins to the correct target nucleotide to be modified. As pointed out before, the eukaryal snoRNAs forms complexes with four distinct core proteins, namely 15.5 kD (Snu13p in yeast) protein, Nop56, Nop58 and Fibrillarin (Nop1 Yeast). Fibrillarin is the catalytic subunit of this sRNP complex and
functions as an AdoMet-dependent Methyltransferase. In archaea however, a single protein called aNop5p with homology to both Nop56 and Nop58 replaces these two eukaryotic proteins. By employing an affinity purification technique with a box C/D motif containing RNA derived from mouse U14 snoRNA, separate groups of proteins were identified to be transiently associated with the core box C/D sRNP proteins (King et al., 2001; Newman et al., 2000). For example, a pair of mouse proteins designated as p50/55 was shown to bind with the core proteins via these pull down experiments. The fact that p55 interacts with TATA-binding protein (TBP) and replication A protein as well as both p55 and p50 having DNA helicase activity suggest the coordination of snoRNA processing and snoRNP assembly with replication and/or transcriptional events in the nucleus. Besides the p50 and p55 proteins, another RNA helicase Dbp4 was needed to unwind the pairing of two guide RNAs sR41 and U14 with the pre-ribosomal RNA (Kos and Tollervey, 2005). These experiments actually hinted towards a possibility of involvement of RNA helicases in sRNP turnover.

Box H/ACA sRNAs on the other hand, binds 4 core proteins L7Ae (Nhp2p in eukaryotes), Nop10, Gar1 and aCbf5, the last protein being the catalytic pseudouridine synthase, with the other proteins playing mainly structural roles.

1.5.1 L7Ae/ 15.5 kD protein/ Snu13p

The L7Ae group of proteins (Koonin et al., 1994; Watkins et al., 2000) forms a large diverse family (Nottrott et al., 1999) whose members include the ribosomal proteins L7Ae and L30e, yeast Snu13p, human 15.5-kD proteins and bacterial homologs like YbxF (Baird et al., 2012). As the name suggests, L7Ae is a component of
the large ribosomal subunit. The members of this family of proteins recognize Kink-turn structures described before. L7Ae is the first protein that binds to the sRNA/snoRNA and starts the nucleation process for an effective box C/D sRNPs/snoRNP formation. As described before, L7Ae and its eukaryal homologs, are also a component of box H/ACA sRNPs and U3 snoRNPs (Marmier-Gourrier et al., 2003). Recently, L7Ae has been shown to be a subunit of RNase P (Cho et al., 2010) and is required for optimal tRNA maturation in an in-vitro reconstituted RNase P system. The eukaryal 15.5 kD/Snu13 protein also binds the spliceosomal U4 snoRNA (Vidovic et al., 2000). Comparisons of RNA binding specificities of archaeal L7Ae and yeast Snu13 proteins reveals that protein L7Ae binds terminal loops with at least 5 nucleotides closed by two A:G and G:A pairs and canonical K-turn structures with similar efficiencies, where as Snu13p does not (Cahill et al., 2002). Again, in contrast to Snu13p, binding of protein L7Ae to canonical K-turn structures is not dependent on the identity of the residue at position 2 in the bulge. The human homolog of Snu13p, the 15.5 kD protein, is also unable to bind to the terminal box C'/D' motifs, which donot form a canonical K-turn structure. Archaeal L7Ae can also bind to the atypical KT15 turns in Pyrococcus furiosus RNAs whereas Snu13p protein cannot (Charron et al., 2004). Hence archaeal L7Ae seems to have a broader substrate binding specificity compared to its eukaryal counterparts.

Though archaeal L7Ae and eukaryotic 15.5 kD proteins are very similar in structure, the crystal structures of both proteins bound to K-turn or K-loop RNAs being essentially super- imposable three-dimensional structures (Gagnon et al., 2010), there are certain differences that might account for broader substrate specificity of L7Ae proteins. For example, Pyrococcus L7Ae has a greater number of acidic amino acids on its surface
contacting the RNA surface compared to the human 15.5 kD protein (Charron et al., 2004). A comparative sequence alignment of the human 15.5 kD protein, Pyrococcus L7Ae and its counterpart from Haloarcula marismortuii have identified 27 conserved amino acids that are implicated in RNA recognition (Gagnon et al., 2010). Moreover a signature of five amino acids in Loop 9, which is adjacent to the RNA binding domain of the protein and in close proximity to the bound RNA motifs, are thought to be responsible in this differential recognition of RNA substrates among archaeal and eukaryal L7Ae counterparts. The differences in Loop 9 amino acids L, E, and V in L7Ae and the eukaryotic replacements V, S, and R were thought to account for L7Ae’s recognition of the K-loop and 15.5kD’s inability to do the same. This signature was found to be clearly critical for L7Ae’s recognition of the K-loop motif, though whether this motif alone is sufficient for K-Loop recognition is not sure. The importance of these signature sequences was further highlighted by studies on Giardia lambia 15.5 kD protein which shares closer sequence similarities with higher eukaryotes compared to archaea in residues of loop 9 (Biswas et al., 2011). The Giardia 15.5 kD protein loop 9 amino acids have a VSVP motif which is closer to the eukaryotic homologues (VSPR), and is again unable to bind to K-loop motif (Biswas et al., 2011).

Recent single particle and crystal structure studies have moreover identified two regions of L7Ae protein involved in K-turn recognition. First, a highly basic \( \beta \) strand: \( \alpha \) helix domain which both recognizes and enters the major groove formed in the non-canonical stem of the K-turn, making both non-specific backbone interactions and specific interactions with the conserved Guanosine nucleobases of the conserved G:A pairs of the non-canonical stems. The other region is a short hydrophobic loop in the
protein which makes van der Waals interactions with hydrophobic nucleobases of the K-turn loop and a terminal Glutamate side chain forming hydrogen bonds. Taken together, this model provides a rationale for an induced fit phenomenon where these multiple interactions are highly selective for the structure of the K-turn. Moreover, studies on *Methanocaldococcus jannaschii* L7Ae protein which show that free protein is virtually identical to that of its RNA bound structure as well as circular dichroism experiments which show that box C/D and C'/D' RNA motifs undergo conformational changes in presence of the L7Ae protein, corroborate this induced-fit model for L7Ae-box C/D RNA interactions (Suryadi et al., 2005). Thus, archaeal L7Ae binding to box C/D sRNA leads to a conformational change in the RNA which in turn provides a potential platform for the binding of the next protein of box C/D sRNP complex, aNop5p (Singh et al., 2008).

1.5.2 aNop5p/Nop56/Nop58

The Nop family of proteins is associated with diverse ribonucleoprotein complexes having wide range of functions starting from RNA modification, pre-rRNA and pre-mRNA processing as well as binding to matrix and scaffold attachment regions in chromosomes (Gautier et al., 1997; van Drunen et al., 1999; Weidenhammer et al., 1996). Affinity pull down studies using box U14 snoRNA identified two highly homologous proteins Nop56 and Nop58 from Yeast to be associated with C/D motif containing snoRNAs (Newman et al., 2000). These observations were further confirmed by affinity purification of these proteins from box C/D snoRNPs from Mouse and also from co-immunoprecipitation experiments using Protein A tagged Nop58 (Lafontaine
Both Nop56 and Nop58 reside in the nucleus and are known in Yeast to be essential proteins (Gautier et al., 1997). Over expression of either Nop56 or Nop58 in Yeast does not compensate for the deletion of either Nop56 or Nop58. Before being identified as box C/D snoRNA associated protein, Nop58 was designated as “Sik-similar protein” because of its close similarity to the “suppressor of 1 Kappa B protein” (Sik1p) in mouse (Newman et al., 2000). Nop58 is highly conserved in eukaryotes showing 47% identity and 71% similarity between mouse and Yeast. Nop56 is also highly related to mouse Nop58 being 36% identical and having 58% sequence similarity. This protein is also highly conserved among eukaryotes showing 49% identity and 71% similarity between mouse and Yeast. Though eukaryotes have two distinct members of this nucleolar protein pairs, members of archaea posses only a single homolog aNop5p (Newman et al., 2000). This single archaeal protein has sequence similarity to both Nop56 and Nop58 which suggests a gene duplication event since divergence of eukarya and archaea.

Most of the information pertaining to the structure of aNop5p is obtained from an array of crystal structure studies of archaeal Nop5p protein, either complexed with box C/D sRNAs and/ or another core protein. As pointed out before archaeal Nop5p protein has structural similarities with both eukaryal Nop56 and Nop58, but some archaeal Nop5p lack a KKxD repeat motif at its C terminus, a motif that is found to be dispensable for function in Yeast Nop56p (Gautier et al., 1997). Even in the archaeal species which has this motif, like the Pyrococcus furiosus Nop5p protein, deletion of this motif has very minimal impact on its structure (Oruganti et al., 2007).
Structurally, Nop5p protein can be broadly divided into three broad domains (Figure 1.5) –

i) The N-terminal domain that interacts with Fibrillarin,

ii) The C-terminal domain that binds RNA

iii) A coiled coil domain connecting the N and C termini of this protein (Aittaleb et al., 2003)

Figure 1.5. Cartoon representation of aNop5p structure. Schematics of aNop5p architecture, with the N-terminal domain of the protein represented as a Yellow box with bound Fibrillarin represented by a red box. The bound SAM is shown as a red star. The C-terminal domain is represented as a green box. The C-terminal domain which folds into an independent NOP domain binds the L7Ae-RNA interface. The coiled coil domain with which aNop5p self-dimerizes is shown by a blue box.

Crystal structure studies of Archaeoglobus fulgidus aNop5p revealed that the protein's N terminus was only stabilized in the presence of bound Fibrillarin and was not an autonomously folding unit (Aittaleb et al., 2003). Though the N-terminal domain of aNop5p from various archaeal species shows very less sequence similarity, the domain folds into a much conserved composite surface for Fibrillarin binding (Gagnon et al.,
2012; Oruganti et al., 2007). The Nop5p-Fibrillarin interface is quite large and predominantly non-polar, which minimizes the contribution of one particular amino acid that is absolutely indispensable for Fibrillarin binding. In fact, it is hypothesized that surface complementarity mediated by numerous main chain interactions, plays an important role in Nop5p-Fibrillarin binding (Reichow et al., 2007). This hypothesis is supported by extensive mutagenesis studies which revealed an unusually strong aNop5p-Fibrillarin interaction (Gagnon et al., 2012). aNop5p binding to Fibrillarin is also suggested to play a role in helping Fibrillarin bind SAM. As such, mutations of key amino acids that prevent aNop5p-Fibrillarin binding in A. fulgidus were also shown to be methylation incompetent (Oruganti et al., 2007). Though it is unclear that during the course of box C/D sRNP assembly, whether it is aNop5p alone or a heterodimer of aNop5p-aFib that actually binds to the L7Ae-sRNA platform, overwhelming evidences points to the latter. First, in vitro, aNop5p-Fibrillarin forms a heterodimer in solution (Tran et al., 2003). Moreover, it was demonstrated that Fibrillarin and Nop56 interact in vivo, even in the absence of box C/D sRNA (Lechertier et al., 2009).

Two fibrillarin-Nop5p heterodimers further dimerize through the long anti parallel coiled coil domain of aNop5p forming a four helix bundle (Aittaleb et al., 2003). Coiled-coils typically consist of two or more alpha-helices that wrap around each other with a super helical twist. Sequences with a propensity to assume coiled-coil structures are characterized by the heptad repeat pattern (abcdefg)n, where a and d are hydrophobic, and e and g are charged or polar. Coiled-coils may interact with each other to form homotypic oligomers, or with other coiled-coil domains to form heterotypic oligomers (Liu et al., 2006; Mason and Arndt, 2004). Helices 4 and 5 (residues 85-149) of
Pyrococcus furiosus aNop5p similarly intertwine around each other to form a homotetrameric complex while still associated with Fibrillarin with its N-terminus (Lapinaite et al., 2013). Domain deletion studies prove that the aNop5p dimerization and interaction of aNop5p with Fibrillarin are mutually exclusive protein-protein interactions with aNop5p association with Fibrillarin being the dominant association (Zhang et al., 2006). The importance of self dimerization of aNop5p through the coiled coil domains was further proved with studies of coiled coil domain deleted aNop5p, which though were perfectly able to assemble on a box C/D sRNA but was modification deficient (Rashid et al., 2003; Zhang et al., 2006). Mutations in the coiled coil domains that disrupted aNop5p homodimerization were also methylation incompetent (Zhang et al., 2006). It is observed by electron microscopy and glycerol gradient centrifugation that aNop5p protein lacking the coiled coil domains also assembles into complexes of smaller dimensions compared to the wild type consistent with the dimensions of a methylation incompetent monomeric sRNP (Bleichert et al., 2009). Hence self-dimerization of aNop5p through its coiled coils domains seems to be an important feature for optimal functioning of box C/D sRNP. Comparison of sequences that fold coiled coil domains are conserved from archaean to eukaryotic and hence eukaryotic Nop56 can also potentially dimerize with its coiled coil domains. However, since the length of the spacer regions in eukaryotic snoRNA are highly variable, it is unlikely that eukaryotic Nop56 and Nop58 employ a self-dimerization mechanism for their catalytic activity.

The C terminal domain of aNop5p is implicated as the region of the protein responsible for binding the L7Ae-RNA complex. This highly conserved domain of
approximately 120 residues folds into an independent NOP domain (Pfam PF01798) that is observed in various other RNA binding proteins, one of the more studied being the human Prp31 protein (Gautier et al., 1997; Liu et al., 2007; Makarova et al., 2002; Vithana et al., 2001). Human Prp31 and Nop56/58 are homologs, with the NOP domain in human Prp31 closely resembling that of archaeal aNop5p in absence of any bound RNA. Human Prp31 and archaeal Nop5p binds to very different RNP complexes. Prp31 recognizes complexes of U4 small nuclear RNA and the 15.5 kD protein and the aNop5p binds to the archaeal homolog of 15.5 kD protein L7Ae bound box C/D RNA. However, it seems that the overall schematics of their substrate recognition are the same. The NOP super family of proteins maintains high sRNP binding selectivity despite relaxed RNA sequence requirements. This is because as opposed to a purely RNA binding domain, NOP domain is a "genuine RNA binding" module that binds to a composite RNA-Protein interface as compared to RNA alone (Liu et al., 2007). A combination of structural topology and electrostatic distribution of key amino acid residues in the NOP domain contribute to the unique binding feature of this module. Calculated electrostatic distribution of the aNop5p –Fibrillarin surface has shown a patch of positively charged surface on aNop5p located on the C-terminal domain extending from the SAM binding pocket of Fibrillarin which will enable the negatively charged backbone of RNA molecule to sandwich between the positively charged residues (Aittaleb et al., 2004). To this effect Alanine scanning have identified key amino acids in the NOP domain which have deleterious effects on the binding capability of aNop5p to a L7Ae-box C/D sRNP complex (Hardin et al., 2009). Moreover, by shape complementarity, aNop5p associates with the L7Ae in the L7Ae-C/D RNA complex (Ye
et al., 2009). The importance of various domains of Nop5p in box C/D mediated methylations will be discussed in more details in the later chapters.

1.5.3 Fibrillarin

Fibrillarin is the catalytic subunit of box C/D snoRNPs/sRNPs (Lapinaite et al., 2013; Lin et al., 2011; Singh et al., 2008; Xue et al., 2012; Ye et al., 2009). Fibrillarin gets its name from the fact that it is the most abundant protein in the fibrillar regions of the eukaryotic cell nucleolus (Warner, 1990). In humans, nucleolar auto antigen against Fibrillarin has been implicated in non-hereditary autoimmune disease Scleroderma (Aris and Blobel, 1991). Gene disruption studies have found Fibrillarin to be essential for survival in mouse embryonic stem cells and Yeast (Newton et al., 2003; Tollervey et al., 1993). Fibrillarin homologs have been found in a number of eukaryotes and archaea. Sequence comparisons as well as immunological cross reactivity and functional complementarity studies have shown that these proteins have a high level of conservation across a wide phylogenetic range (Amiri, 1994; Aris and Blobel, 1991; Bult et al., 1996; David et al., 1997; Lapeyre et al., 1990). Eukaryotic Fibrillarin are generally bigger than their archaeal homologs having an extra Glycine-Arginine-Rich (GAR) domain which localizes the eukaryotic Fibrillarin to the nucleolus (Amiri, 1994; Bult et al., 1996). The N-terminal domain is also less conserved in terms of both sequence and structure in proteins isolated from various archaea (Oruganti et al., 2007). Specific β-strand interactions in the N-terminal domain of Fibrillarin from Methanocaldococcus jannaschii (Mj-Fib) were reported to facilitate dimerization of fibrillarin molecules (Deng et al., 2004). However, Fibrillarin from Archaeoglobus fulgidus (Af-Fib) and Pyrococcus furiosus (Pf-Fib) were shown to exist as monomers in solution, proving Fibrillarin homo
dimerization is a not general property of all Fibrillarin of archaea. Other than the N-terminal domain, the rest of the protein has been found to be quite similar, having ~40% sequence identity between eukaryotes and archaea.

Information regarding Fibrillarin protein architecture comes from several crystal structures of archaeal Fibrillarin (Aittaleb et al., 2004; Oruganti et al., 2007; Reichow et al., 2007; Wang et al., 2000). *Methanocaldococcus jannaschii* Fibrillarin has a total of seven α helices and 12 β strands which can be divided into a smaller N terminal domain and a larger C terminal domain. Despite the absence of any significant sequence similarities, the overall fold of the C-terminal domain is similar to the catalytic domain common to many AdoMet dependent Methyltransferases. This prototypical methyltransferase structure consists of a “core fold” consisting of alternating β sheets and α helices (Cheng and Roberts, 2001; Schubert et al., 2003). This characteristic “Rossmann fold” motif is responsible for AdoMet binding and supporting the catalytic reaction. The catalytic mechanism of Fibrillarin remains to be determined, but it is thought to occur, like most Rossmannoid Methyltransferase reactions, via a SN$_2$-type mechanism (Reichow et al., 2007). Studies using chiral AdoMet had shown that the transfer of methyl group from AdoMet to substrates occurred with the inversion of configuration about the methyl groups, strongly suggesting a direct displacement of methyl group from AdoMet to the target substrates (Kealey et al., 1991). This would require initiation by a concerted action of several side chains of the enzymes. As such, the residues involved in AdoMet recognition are conserved in both archaeal and eukaryal Fibrillarin and occupy similar positions in each of the characterized Fibrillarin structure (Reichow et al., 2007). An absolutely conserved Aspartate (D133 of
Archaeoglobus fulgidus Fibrillarin) has been proposed to act as the general base responsible for the deprotonation of the 2'-OH of ribose sugar and stabilizing SAM interaction (Aittaleb et al., 2004). In the co-crystal structure of AF fibrillarin-Nop5p complex bound with AdoMet (holocomplex), a number of conserved fibrillarin residues were observed to interact directly with AdoMet. Glu-88 forms two hydrogen bonds with the ribose hydroxyl groups of AdoMet. Thr-70 also forms a hydrogen bond with the carboxyl group of AdoMet. Asp-133 is situated near the positively charged thiomethyl group and thus may facilitate cofactor binding through favorable electrostatic interactions. Finally, Tyr-89 establishes an aromatic stacking interaction with the adenine ring of the cofactor. Moreover some positively charged residues which are in close proximity to the positively charged patch of aNop5p C terminal domain seem to interact with box C/D sRNA. Hence it is proposed that together these positively charged amino acids of Fibrillarin and aNop5p collectively forms multiple contact points with the negatively charged backbone of RNA and facilitate the accurate transfer of methyl groups to the target RNA molecule.

Comparison between crystal structures of free fibrillarin and fibrillarin-Nop5p-AdoMet tertiary complex revealed large conformational differences at the cofactor-binding site in fibrillarin (Deng et al., 2004). It was shown that the C-terminal domain of aNop5p helps orient a conserved aromatic amino acid residue in Fibrillarin for AdoMet recognition, which adopts an inhibitory conformation when aNop5p is not bound (Oruganti et al., 2007; Reichow et al., 2007). Thus aNop5p plays a dual role in Fibrillarin binding, first helping stabilize AdoMet binding to Fibrillarin molecule and also orienting the catalytic site of box C/D sRNP to its correct target.
1.6 Guide RNA independent methylation modifications in Archaea

Extensive Modomics studies have been performed to predict the plethora of methyltransferase enzymes that are responsible for methylating target ribonucleotides either at the bases or Ribose sugars themselves. Some of these predicted enzymes have been biochemically and genetically verified to be the putative methyltransferases, while some of them await verification (Grosjean et al., 2008). *Haloferax volcanii*, whose genome has been sequenced and in which the modified positions in tRNAs and rRNA have been extensively studied, serves a model organism in these studies (Grosjean et al., 2008; Gupta, 1984; Gupta et al., 1983).

The first discovered archaeal guide independent methyltransferase catalyzes the formation of the universally conserved Cm at position 56 in most sequenced archaeal tRNAs (Gupta, 1984; Renalier et al., 2005). This enzyme is called aTrm56 and is present in all sequenced archaeal genomes except in *Pyrobaculum*. Additionally, Cm56 modification is found only in archaeal tRNAs whereas in bacteria and eukaryotes this position is usually a universal C56. Another modification which is a hallmark of a majority of known archaeal tRNAs is the methylation modification of \( \Psi54 \) to \( \text{m}^1 \Psi54 \), except for *Thermococcales* where \( \text{m}^5 \text{U54} \) is found (Chatterjee et al., 2012; Grosjean et al., 2008). The methyltransferase responsible for this conserved modification was not known till recently. Using comparative genomic approaches, by studying modifications and their responsible genes in bacteria like *E.coli*, studies have predicted the presence of a large number of methyltransferases in archaea carrying out a host of methylation modifications in either the 23S rRNAs and 16S rRNAs as well as tRNAs (Grosjean et al., 2008). These methylations occur on a wide range of substrates –nitrogen as well as
carbon atoms in bases of individual ribonucleotides as well as Ribose sugars as pointed as before.

1.6.1 Classification of Methyltransferase Enzymes

Based on the topology of the methyl donor binding sites, the Methyltransferases might be classified broadly into Methyltransferases containing Folate dependent domains and Methyltransferases containing SAM binding domains.

1.6.1.1. Methyltransferases containing TIM barrels

Methyltransferases of this category, the domain/protein that binds the methyl group donor folds into $\alpha/\beta$ triose phosphate isomerase (TIM) barrel structure as observed in crystal structures of multitude of B$_{12}$ and Folate dependent Methyltransferases (Doukov et al., 2000; Doukov et al., 2007; Hagemeier et al., 2006). In all these proteins, the methyl group binds within a cavity formed by the TIM barrel. Interestingly, the Radical SAM Methyltransferases also have TIM like barrels with additional inserted elements having distant sequence similarity with the corrinoid Methyltransferases (Kozbial and Mushegian, 2005).

1.6.1.2 Methyltransferases containing SAM binding domains

By an iterative comparative strategy, using the known or suspected SAM-binding domains as queries in sequence modeling and database searching, most
Methyltransferases containing SAM binding sites can be divided structurally and phylogenetically into five distinct classes (Kozbial and Mushegian, 2005; Schubert et al., 2003).

1.6.1.2.1 Class I: The Rossmanoids

Majority of SAM-MTs belong to a large class of enzymes with Rossmann like fold, a common arrangement of protein’s spatial structure, observed in many diverse families of enzymes. SAM-MTs are a large group of enzymes within the Rossmannoid class. Fibrillarin is an example of a Rossmannoid methyltransferase. In the most basic form, the Rossmann like fold consists of seven-stranded \( \beta \) pleated sheet, with a centre topological switch point and a characteristic reversed \( \beta \) hairpin at the Carboxyl end of the sheet (Fig. 1.6) (Schubert et al., 2003). This sheet is flanked by helices to form a double wound open \( \alpha \beta \alpha \) sandwich. Another typical feature of the
Rossmanoid enzymes is the location of functionally important conserved residues at the C terminal part of the β strands or in the adjoining loops.

1.6.1.2.2 Class II MTases:

Class II MTases contain a long central anti parallel β sheet flanked by groups of helices at either side. Both in terms of structure and AdoMet binding Class II MTases are quite different from the Rossmanoids. AdoMet is bound to a shallow groove along the edges of the β-strands, forming hydrogen bonds to a conserved RxxxGY motif (Kozbial and Mushegian, 2005).

1.6.1.2.3 Class III MTases

In this structural family, which contains CbiF, a MTase that acts on ring carbons of large, planar percorrin substrates during cobalamin synthesis, the active site is located inside a cleft between two αβα domains each containing five strands and four helices (Kozbial and Mushegian, 2005). In spite of the presence of GxGxGx motif, AdoMet binding is not observed at this site. Instead AdoMet appears to be tightly folded and binds between two domains of this homodimeric protein.

1.6.1.2.4 Class IV MTases:

The SPOUT Class of RNA MTases are the only known members of the Class IV structure, the nomenclature coming from spoU and trmD RNA methylase superfamilies
Figure 1.7. Topology and Cartoon representation of SPOUT methyltransferases. Figure taken from Kozbial et al., 2005

(Anantharaman et al., 2002). These structural family contains six stranded parallel $\beta$-sheet flanked by seven $\alpha$ helices of which the first three strands form half of the Rossmann fold. Though the SPOUT domain is a highly conserved structural fold, the amino acid sequences are not conserved throughout the SPOUT superfamily (Liu et al., 2013). Also, the specificity of substrate recognition cannot be predicted based on sequence or structural homology. They form homodimers, with the catalytic sites located at the interface of the two subunits (Chen and Yuan, 2010). The most characteristic feature of the SPOUT superfamily is however, the presence of a trefoil knot at the C-terminus (Michel et al., 2002). A “trefoil” knot, which is the most common motif among knotted proteins, occurs when a few residues at one end of the chain gets tucked or “threaded” through a loop exposed on the protein surface (Nureki et al., 2002; Taylor, 2000). Crystal studies have revealed that this knotted region which forms a cleft in the protein is SAM binding site (Michel et al., 2002). Recent studies have revealed that some of the RNA methyltransferase belonging to the SPOUT superfamily
have this “trefoil” knots (Elkins et al., 2003; Tkaczuk et al., 2007; Watanabe et al., 2006).

1.6.1.2.5 Class V MTases:

The SET domain proteins form the 5th structural family of AdoMet dependent MTases (Manzur et al., 2003; Xiao et al., 2003). These enzymes contain a series of eight β-strands forming three small sheets, with the C terminus tucked underneath a surface loop forming a knot like structure similar to Class IV MTases, but totally different topology. There are many examples of proteins having SET domains and they have been mostly seen to methylate Lysine residues in the flexible tails of Histones or in Rubisco. Flanking the SET domain are diverse sequences termed as pre and post-SET regions, which is thought to participate in substrate recognition and specificity.

1.6.2 S-Adenosyl Methionine as Methyl Group Donor

The strong preference for SAM (Figure 1.7) over other methyl donors (Martin and McMillan, 2002), such as folate, reflects the highly favorable thermodynamics of SAM.

![Chemical structure of SAM](image)

Figure 1.8. Chemical structure of SAM
dependent methyl transfer reactions. The $\Delta G^0$ (Grove et al., 2011) for methyl transfer from AdoMet is -17 Kcal mol$^{-1}$, over double than that for (ATP $\rightarrow$ ADP + Pi) (Fontecave et al., 2004). The favorable energetics results from the charged methylsulfonium centre of SAM. The methyl carbon is electrophilic because it is bonded to positively charged sulfur which is a powerful electron withdrawing group. The positive charge on the sulfur also makes it an excellent leaving group as the resulting product will be neutral and very stable sulfide. This strong electrophoretic character of the methyl group in SAM has been exploited by SAM-dependent methyl transferases to bring SAM into nucleophilic groups of substrates (Grove et al., 2011).

1.7. RNA modifications and disease

Since snoRNAs are involved in multiple and diverse cellular activities, any defect in their formation or function, which includes RNA modification, may be deleterious for the cell and may lead to several diseases. Absence or defective expression of certain human box C/D RNA gene clusters has been implicated for Prader-Willi syndrome (Kishore and Stamm, 2006; Matera et al., 2007; Sahoo et al., 2008). This syndrome is the leading genetic cause of obesity with a prevalence rate of 1 in 10,000 to 20,000 (Ding et al., 2008). The disease is characterized by neonatal hypotonia leading to feeding difficulties, failure to thrive, short stature, and hyperphagia leading to obesity, mental retardation and hypogonadism. Autoantibodies against core proteins of box C/D and box H/ACA sRNPs have been detected in patients suffering from several kinds of autoimmune diseases like Systemic Lupus Erythematosus, systemic sclerosis, primary Reynaud’s phenomenon, Rheumatoid arthritis and myositis (Van Eenennaam et al., 2002).
The most defining example of lack of modification in RNA leading to disease has been unraveled in recent studies that have found that up to 20 percent of human mRNA is routinely methylated (Meyer et al., 2012). Over 5,000 different mRNA molecules contain m\(^6\)A, which means that this modification is likely to have widespread effects on how genes are expressed. The study demonstrated that the obesity risk gene, FTO (fat mass and obesity-associated), encodes an enzyme capable of reversing this modification, converting m\(^6\)A residues in mRNA back to regular adenosine. Humans with FTO mutations have an overactive FTO enzyme, which results in low levels of m\(^6\)A and causes abnormalities in food intake and metabolism that lead to obesity. m\(^6\)A is present in many mRNAs encoded by genes linked to human diseases, including cancer as well as several brain disorders, such as autism, Alzheimer's disease, and schizophrenia (Meyer et al., 2012).
CHAPTER 2

THE ARCHAEAAL COG1901/DFU358 SPOUT-METHYLTRANSFERASE MEMBERS, TOGETHER WITH PSEUDOURIDINE SYNTHASE PUS10, CATALYZE THE FORMATION OF 1-METHYLPSUEDOURIDINE AT POSITION 54 OF tRNA

2.1 Introduction

Transfer RNAs in all three domains of life contain a large variety of characteristic post-transcriptionally modified nucleosides. Among them pseudouridine, an isomer of uridine (abbreviated Ψ) and base or ribose methylations (mN, or Nm respectively; N being any of the four base A, U, C or G) are by far the most abundant (Cantara et al., 2011; Czerwoniec et al., 2009; Motorin and Helm, 2011). Many of these non-canonical nucleotides are highly conserved at specific sites in functionally important parts of tRNA molecules. For example Ψ, m^5U (riboT) or m^1Ψ are nearly always found at position 54 in the so-called Ψ-Ψ-loop at the elbow of the L-shaped tRNA molecules, while Ψ is found at the neighboring position 55. The conservation of these modifications at these two positions is certainly due to essential structural roles. Indeed, the presence of Ψ55 reinforces tertiary base pairing with the conserved G18, favors intra-loop stacking with the conserved purine at position 57 and with the neighboring m^5U54/m^1Ψ54 engaged in a reverse-Hoogsteen pair with the conserved A58 (Romby et al., 1987). Together with the conserved tertiary pair C56-G19 between the Ψ-Ψ-loop and D-loop, this ‘ensemble’ of conserved interacting nucleotides allow crucial interactions that form a stable ‘tertiary core’ and hence the canonical tRNA L-shape architecture (Kotlova et al., 2007). In
agreement with these ‘locking’ function of both Ψ55 and m$^5$U54 (probably also m$^1$Ψ54 in archaeal tRNAs, see below), these base modifications are among the earliest to appear during the complex tRNA maturation process (Grosjean et al., 1996; Nishikura and De Robertis, 1981).

Even if the modifications at position 54 and 55 are conserved and have critical structural roles, strains lacking these modifications are viable. A mutant of Escherichia coli deleted in the truB gene (encoding for the Ψ55 producing enzyme) grew normally on all media tested (Gutgsell et al., 2000). It did exhibit a competitive disadvantage in extended co-culture with its wild-type progenitor and a defect in surviving rapid transfers from 37°C to 50°C (Gutgsell et al., 2000). Moreover, combining truB mutations with mutations affecting the catalytic activity of TrmA, the enzyme catalyzing the formation of the adjacent m$^5$U54, further increased the temperature sensitivity phenotype (Kinghorn et al., 2002). The high temperature survival of hyperthermophiles such as Thermus thermophilus, is strictly dependent on an additional hypermodification of m$^5$U54 into s$^2$m$^5$U54 (s$^2$T) (Shigi et al., 2006). These modifications protect the tRNA architecture against heat inactivation because of a stronger intra-loop reverse-Hoogsteen interaction with A58 (also modified to m$^1$A58), a better stacking of s$^2$T harboring the bulky highly polarizable 2-thiocarbonyl group with the nearest-neighbor G53-C61 base pair and consequently a reduction of the motional dynamics of the tRNA molecule at high temperature (Davanloo et al., 1979; Shigi et al., 2006; Watanabe et al., 1976). Thus, enzymatic formation of both m$^5$U54 or s$^2$m$^5$U54 and Ψ55 are not essential per se, but clearly contribute at least to thermal stress tolerance.
In the present work, we focus on the biosynthesis of 1-methylpseudouridine in tRNAs ($m^1\Psi54$, Fig. 2.1A). This $\Psi$ derivative was first characterized from the bulk tRNA of archaeon *Halococcus morrhuae* (Pang et al., 1982).

Figure 2.1. Enzymatic posttranscriptional modifications of selected uridines in RNA. (A) Schematic consensus of tRNA secondary structures indicating the U54 target within the highly conserved seven-nucleotides T-$\Psi$-loop. The dashed line indicates a reverse Hoogsteen base-pair within the loop. In the majority of Archaea (mostly Euryarchaea and Crenarchaea), U54 is first isomerized into $\Psi54$ by a tRNA...
pseudouridine synthase, aPus10 (symbol ‘a’ preceding the acronym of the enzyme denotes ‘archaeal’, ‘b’ means ‘bacterial’ and ‘e’ means ‘eukaryal’). Next, Ψ54 can be methylated into m^1Ψ54 by a SPOUT-type, SAM-dependent methyltransferase designated TrmY (this work). In certain Archaea this methylase is absent and Ψ54 is no further modified (see Table 2.1). (B) In Archaea belonging to the order Thermococcales, U54 is first methylated to form m^6U54 (riboT) by another SAM-dependent tRNA-U54 methyltransferase aTrmU54. This methylase is of the Rossmann-fold type and belongs to a distinct COG (COG2265) than aTrmY (COG1901). In Bacteria and Eukarya, the same U54 methylation is catalyzed by bTrmA and eTrm2p respectively, both belonging to the same COG2265 as aTrmU54. Depending on the temperature at which the cell is grown, m^6U54 can be further thiolated into s^2m^6U (s^2T). In hyperthermophilic bacteria, this thiolation reaction is catalyzed by the heteromeric enzyme TtuA/TtuB, using IcsS as a cofactor for transfer of the thio-group, while in archaea, the enzymatic system and cofactors remain to be identified. (C) Schematic consensus of a portion of 16S rRNA as part of domain IV, encompassing the highly conserved eight-nucleotide helix 35 Ψ-loop. U914 in 16S rRNA of M. jannaschii, corresponding to U1191 in 18S rRNA of S. cerevisiae, is first isomerized into Ψ by an archaeal enzyme (or enzymatic system) that remains to be identified. In S. cerevisiae this reaction is mediated by the snoRNP complex consisting of the pseudouridine synthase eCbf5 and snR35 guide RNA. Next, both in Archaea and Eukarya, the methylation of N1-atom of the uracil ring is catalyzed by a SPOUT-type and SAM-dependent methyltransferase, Nep1 (also referred as Emg1). Only in Eukarya (such as yeast, Drosophila, HeLa cells), the m^1Ψ is further hypermodified into acp^3m^1Ψ derivative. This last reaction occurs at a very late step of ribosome biogenesis on the cytoplasmic 40S ribosomal subunit. The initial pseudouridylation and N1-methylation are both catalyzed much earlier during rRNA maturation process; in Eukarya, these reactions occur within the nucleolus. On the right part of the figure are the different uridine derivatives. The dashed lines through the structure of m^1Ψ and acp^3m^1Ψ show the axis of base rotation during the isomerization process. The asterisks with small arrows indicate the atoms normally engaged in the reverse Hoogsteen base-pair with A58.
It has now been found in tRNAs isolated from many Archaea, mainly Euryarchaeota (Table 2.1). Its location at position 54 of tRNA was inferred from sequence analysis of tRNAs of *Haloferax volcanii* (Gupta, 1984). In contrast, tRNA sequences from *Thermoplasma acidophilum* (two sequences), *Methanobacterium thermoautotrophicum* (one sequence) and *Sulfolobus acidocaldarius* (one sequence) revealed the presence of a non-methylated Ψ or a ribose-methylated U (Um) in place of m¹Ψ54 (http://www.uni-bayreuth.de/departments/biochemie/trna/ and (Cantara et al., 2011). In tRNAs from Pyrococcales (*Pyrococcus furiosus* and *Pyrococcus abyssi*), m⁵U54 or s²m⁵U54 are found instead of m¹Ψ54 (Fig. 2.1B). (Constantinesco et al., 1999; Kowalak et al., 1994). As previously noted (Gupta and Woese 1980; Pang et al. 1982), methyl groups on the uracil ring at tRNA position 54 in the two cases (N1 in m¹Ψ and C5 in m⁵U) bear similar orientations with respect to the ribose and the polynucleotide chain and both methylated products have similar shapes. This could be an example of evolutionary convergence of structures (and probably of function). Only the first step in m¹Ψ formation in tRNA, the formation of Ψ54, has been elucidated with the recent characterization of the tRNA-Ψ54 synthase, Pus10 (Gurha and Gupta, 2008b). The enzyme responsible for the subsequent N1-methylation of the uracil ring of Ψ54 and the corresponding gene has yet to be identified. However, experiments with whole-cell extract of *H. volcanii* incubated with ³²P-radiolabeled T7-transcript and AdoMet (SAM), showed enzymatic formation of m¹Ψ54, indicating the enzyme of interest was SAM-dependent (Grosjean et al., 1995).
Purified bulk tRNAs were analyzed for their nucleoside content after complete digestion by nuclease, usually nuclease P1, followed by dephosphorylation and LC/MS or 2D-TLC after 5'-32P- post-labeling. An alternative TLC separation method used RNase T2-digests of uniformly labeled tRNAs isolated from 32P-labeled cells. The original LC-MS method is described in details in: (Edmonds et al., 1985) and (Crain,
For \(^{32}\)P-post-labelling technique and TLC analysis, detailed information can be found in: (Grosjean et al., 2007). Analyses of uniformly labeled tRNAs can be found in: (Gupta and Woese 1980) and (Gupta 1984). \(\textit{m}^1\Psi, \textit{m}^5\textit{U}\) or \(\textit{s}^2\textit{T}\) have been found to date only at position 54 in tRNA. Their presence in bulk tRNA hydrolysates are therefore diagnostic of their presence at position 54 exclusively. Presence of \(\Psi\) alone or Um, or any other U derivatives at position 54 can be deduced only from direct tRNA sequencing (indicated in the table by an asterisk (*) next to the symbol of the modified nucleoside). The indication ‘no \(\textit{m}^1\Psi\’ does not exclude the possibility of a non-methylated \(\Psi\) in the \(\Psi\)-loop. Beside Um found in initiator tRNA\(^{\textit{Met}}\) sequenced from \textit{S. acidocaldarius}, \(\textit{s}^2\textit{U}\) and \(\textit{s}^3\textit{Um}\) were also found in bulk tRNA in some Crenarchaeota (Edmonds et al., 1991). Since no sequence analyses are available, this information was omitted from the table even if these modifications could be found at position 54.

The presence or absence of a gene coding for \(\textit{apUs10}, \textit{aTrmY}\) and \(\textit{aTrmU54}\) were derived from the SEED database and from the taxonomic tool of Blink at NCBI. The ‘nc’ symbol means ‘not applicable’ because the genome sequence is not available. The \(^*\) symbol refers to the fact that a gene coding for a \(\textit{Pus10}\) homolog is found in these genomes but the corresponding proteins do not meet structural criteria for a functioning \(\textit{Pus10}\) (Elisha Fitzek, Ramesh Gupta and Matt Geisler, unpublished). Names for \textit{Halobacterium cutirubrum} have changed with time; \textit{Halobacterium salinarium}, \textit{Halobacterium halobium} are the same species. \textit{Halobacterium} sp. NRC1, a different strain from \textit{H. cutirubrum} but with a sequenced genome was used to identify the presence of gene \(\textit{pus10}, \textit{trmY}\) and \(\textit{trmU54}\).

Nep1 (Nucleolar Essential Protein1, formerly named Emg1) from \textit{Saccharomyces cerevisiae} and \textit{Methanocaldococcus jannaschii} was recently identified as a genuine SAM-dependent N1-pseudouridine methyltransferase (Wurm et al., 2010). This enzyme belongs to Cluster of Orthologous Group (COG) 1756 (Tatusov et al., 2003) and converts \(\Psi\) into \(\textit{m}^1\Psi\) in small synthetic fragments of 8, 9 or 11 nucleotides in length (GAUUCAACGCC where second of the two adjacent U is \(\Psi\)) (Wurm et al., 2010). This motif corresponds to the sequence in helix 35 of SSU rRNA of \textit{S. cerevisiae}, as well as of \textit{M. jannaschii}, where Nep1 tightly binds (Buchhaupt et al., 2006). Nep1 is responsible for the methylation step in the formation of the hypermodified 1-methyl-3-(3-amino-3-carboxypropyl) pseudouridine (\(\textit{m}^1\textit{acp}^3\Psi\)) found at position 1191 in helix 35 of yeast 18S.
rRNA and also has a role in ribosome assembly (Meyer et al., 2011; Thomas et al., 2011) (Fig. 2.1C). These functions appear to be conserved in other eukaryotes (Brand et al., 1978; Youvan and Hearst, 1981). Despite the obvious sequence analogy of the Nep1 target with the conserved sequence GUUCAANC/U (underlined, see also Fig. 1A) present in nearly all tRNAs sequenced so far (Grosjean et al., 2008), Nep1 is certainly not the missing tRNA N1-pseudouridine methyltransferase. The reasons are:

(a) the genome of S. acidocaldarius contains a Nep1 homolog (Saci_0034) but its bulk tRNA lacks m^1Ψ and its only sequenced tRNA harbors U_m54 (Gupta and Woese 1980; Kuchino et al. 1982) (Table 2.1); (b) conversely, the genome of H. volcanii lacks the gene coding for Nep1, while nearly all its tRNAs harbor m^1Ψ54 (Gupta 1984, 1986) (Table 1); (c) this absence is consistent with the fact that helix 35 of H. volcanii 16S rRNA harbors an acp^3U and not the ‘hypermodified’ m^1acp^3Ψ as in eukaryotes (Kowalak et al., 1994).

A better candidate for the missing m^1Ψ54 methyltransferase came from a bioinformatics analysis of a large variety of orphan genes coding for putative AdoMet-dependent methyltransferases in genomes of microorganisms belonging to the three domains of life. This analysis identified one of the methyltransferases belonging to COG1901, encompassing an alpha/beta knot fold (also named SPOUT) superfamily of methyltransferases as a valid candidate (Tkaczuk et al., 2007). This prediction fits with the observation that genes of this family usually cluster with pus10 in several archaeal genomes (Grosjean et al., 2008). Lastly, the crystal structure of a COG1901 family member, Mj1640 from M. jannaschii, was solved in complex with AdoMet at 1.4 Å resolution (Chen and Yuan, 2010). Mj1640 protein shares much structural similarity with
the highly conserved eukaryotic nucleolar Emg1/Nep1 protein at its C-terminal half containing the conserved deep trefoil knot fold (Z-score 15.7, r.m.s. 2.5 Å, 163 Cα).

However, because of significant differences in N-terminal extension domain and overall surface charge distribution, it was suggested that the two proteins Nep1 and Mj1460 target different RNA sequences (Chen and Yuan, 2010). This structural data greatly strengthened the prediction that the COG1901 (or DUF358) RNA methylase family was the missing tRNA-\(m^1\Psi\) forming enzyme and we set out to test this experimentally using both in vivo and in vitro assays. Here we show that a \textit{H. volcanii} strain deleted in the COG1901 family gene \textit{HVO_1989} lacks \(m^1\Psi\) in tRNA and that in vitro Mj1640 catalyzes the formation of \(m^1\Psi\) at position 54 on tRNA. These results led to the renaming of the Archaeal COG1901 proteins as TrmY (as \(Y\) is commonly used to represent \(\Psi\)) for tRNA (pseudouridine54-N1)-methyltransferase.

\section*{2.2 MATERIALS AND METHODS}

\subsection*{2.2.1 Bioinformatics}

Sequences and distribution of all genes/proteins analyzed in this work are available through the “COG1901 sub-system” on the public SEED server (http://pubseed.theseed.org/SubsysEditor.cgi?page=SubsystemOverview) (Bult et al., 1996). We also used the BLAST tools and resources at NCBI (Altschul et al., 1997). Multiple protein alignments were performed with the ClustalW tool (Chenna et al., 2003) in the SEED database or the MultiAlign software (http://omics.pnl.gov/). \textit{H. volcanii}
genome sequences were obtained from the UCSC browser (http://archaea.ucsc.edu/cgi-bin/hgGateway?db=haloVolc1). For the phylogenetic analysis, candidate protein sequences were identified using the BLAST tool at NCBI (Altschul et al., 1997), all were COG1901 members. Sequences were aligned using MUSCLE (Edgar, 2004) through the EMBL-EBI website. Phylogenetic analysis was performed using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003) from the command line using the Dayhoff-6 amino acid categories and inferred a tree with the CAT+Γ model to account for evolutionary rate site variations. MrBayes was run for 1,100,000 MCMC iterations with a burnin of 100,000. Trees were sampled every 1000 iterations. A consensus tree was generated and visualized using FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/)

2.2.2 Strains, media and transformation

All strains used in this study are detailed in Supplemental Table S1. *H. volcanii* H26 was used as the wild-type (WT) strain. *E. coli* was routinely grown in LB-Lennox (LB) (Fisher) or LB agar (Fisher) at 37°C, supplemented when required with ampicillin (Amp; 100 µg/mL), isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.2 mM) and bromochloro-indolyl-galactopyranoside (X-gal; 40 µg/mL). When required, novobiocin was added to a final concentration of 0.3 µg/mL. *H. volcanii* cells were routinely grown at 44°C (unless specified) in rich medium (Hv-YPC) or minimal medium (Hv-min). *H. volcanii* media were made according to the recipes provided in the HaloHandbook (http://www.haloarchaea.com/resources/halohandbook/) (Dyall-Smith, 2009). Transformations of chemically-competent *E. coli* were performed as described by the manufacturer’s directions (Invitrogen, CA). Transformation of *H. volcanii* was performed
as described in the HaloHandbook using the “standard PEG-mediated transformation of Haloarchaea” protocol.

2.2.3 Plasmid and deletion strain construction

All plasmids and oligonucleotides used in this study are listed in Supplemental Tables S2 and S3, respectively. plKB227, a pTA131 derivative used to disrupt trmY (HVO_1989) was produced as described previously (Blaby et al., 2010), using oligonucleotide pairs HVO1989_N_IfKO_Fwd, HVO1989_N_IfKO_Rev, HVO1989_C_IfKO_Fwd, and HVO1989_C_IfKO_Rev to amplify the regions upstream and downstream of the target gene. Demethylated plKB227 was prepared by passaging the plasmid through E. coli (INV110, Invitrogen) and was subsequently used in the pop-in/pop-out procedure (Allers et al., 2004) to delete trmY on the H. volcanii H26 chromosome. trmY of Halobacterium sp. NRC1 (VNG_1980C) was amplified by PCR from purified genomic DNA using Phusion Hot Start polymerase (Finnzymes, Espoo, Finland) and Hsal_cog1901_Fwd and Hsal_cog1901_Rev as primers. The amplicon was inserted, after digestion with the compatible enzymes, between the Ndel and Blpl sites of pJAM202 to generate plKB421 (pHTrmY). This plasmid was verified by PCR, restriction digestion and sequencing and then passed through E. coli (INV110, Invitrogen) before transformation into H. volcanii strain VDC2376 (∆trmY) creating complemented strain VDC2604. Transformants were selected for by plating onto Hv-YPC containing novobiocin (0.3 µg/mL).
2.2.4 Detection of m1\(\Psi\) residues in tRNAs

For analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS), *H. volcanii* bulk tRNA was prepared, hydrolyzed and analyzed as described previously (de Crécy-Lagard et al., 2010). tRNA purifications and analyses were performed at least twice independently. For analysis by 2D-TLC, uniformly \(^{32}\)P-labeled total tRNA was prepared as described before (Gupta, 1984; Joardar et al., 2012). Labeled tRNAs were digested with nuclease P1 and digests were resolved on cellulose plates (EM Science) using isobutyric acid/0.5 N NH\(_4\)OH (5:3, v/v) in the first dimension and isopropanol/HCl/H\(_2\)O (70:15:15, v/v/v) in the second dimension (Gupta, 1984). Radioactivity in the plates was revealed by phosphorimaging. Presence of \(\Psi\) at a specific position in tRNA was analyzed by CMCT-RT method as described before (Blaby et al., 2010) using total small RNA. For this, initially total RNA was prepared from the cells using TRI reagent (Molecular Research Center, Inc). Small RNA fraction from the total RNA was extracted by suspending total RNA in 1 M NaCl by vigorous vortexing. The high salt soluble fraction was separated from insoluble material by centrifugation. RNA was precipitated from the supernatant by ethanol.

2.2.5 Expression and purification of TrmY of *M. jannaschii*

Recombinant TrmY protein of *M. jannaschii* (MJ1640) was purified as described before (Chen and Yuan, 2010). These procedures are considered native conditions in this work. The protein was also prepared under renaturing conditions. For this, harvested cells were resuspended in a buffer (10 mM Na\(_2\)HPO\(_4\), 2 mM KH\(_2\)PO\(_4\), 2.7 mM KCl, 500 mM NaCl) containing 8 M urea and lysed by sonication. The supernatant from
a 25 min spin at 13,000 rpm at 4°C was loaded onto Ni\textsuperscript{2+} column at room temperature equilibrated with the same buffer. The protein was renatured on the column in a gradient of 8 M urea to no urea in the same buffer containing 2.5 mM imidazole. After a brief wash with 25 mM imidazole containing buffer the protein was eluted using 250 mM imidazole in the same buffer. Pooled fractions were then dialyzed using 500 mM NaCl, 10 mM DTT and 50 mM Tris-Cl, pH 8.0 and 20% Glycerol.

2.2.6 \textit{In vitro} tRNA methylation assays

Transcripts for \textit{H. volcanii} tRNA\textsubscript{Trp} (lacking its intron), elongator tRNA\textsubscript{Met}, T-arm-Trp and T-arm-Met were generated as described previously (Gurha et al., 2007). Mutants of tRNA\textsuperscript{Trp} were prepared by independently changing U55 to A, C and G, and of tRNA\textsuperscript{Met} by changing U54 to A as described before (Gurha et al., 2007). Sequences of these transcripts are shown in Fig 2.4A. Appropriate labeled transcripts were prepared using relevant [$\alpha$-\textsuperscript{32}P] NTP. Recombinant \textit{M. jannaschii} Pus10 was used first to convert U54 and U55 equivalents of these transcripts into $\Psi$ as described previously (Gurha and Gupta, 2008a). \textit{M. jannaschii} TrmY efficiently works under a range of salt concentrations (100-500 mM NaCl). Our standard 20 $\mu$l methylation reaction contained 100-150 nM pseudouridylated RNA and 1 $\mu$M renatured \textit{M. jannaschii} TrmY in 20 mM Tris-Cl, pH 7.0, 150 mM NaCl, 0.75 mM DTT, 1.5 mM MgCl\textsubscript{2}, 0.1 mM EDTA, 0.1 mM AdoMet and 5% Glycerol. The reaction mixtures were incubated at 68°C for 1 hr. RNA was purified by phenol/chloroform extraction and ethanol precipitation and digested with RNase T2 or nuclease P1. The digests were resolved by TLC as described before for uniformly labeled tRNA. For methylation inhibition
reactions, proteins isolated under native conditions were used and 0.1 mM of S-adenosylhomocysteine, instead of AdoMet, was used in the reactions.

**Table 2.2 - Strains used in Chapter 2**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> Top10 DH5α</td>
<td>F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG λ-</td>
<td>Invitrogen, CA</td>
</tr>
<tr>
<td><em>E. coli</em> inv110</td>
<td>F⁻ {traΔ36 proAB lacI₄ lacZΔM15} rpsL (StrR) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) Δ(mcrC-mrr)102::Tn10 (TetR)</td>
<td>Invitrogen, CA</td>
</tr>
<tr>
<td><em>E. coli</em> BL21 (DE3)</td>
<td>F⁻ ompT gal dcm lon hsdS₈(rB⁻ mB⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])</td>
<td>(Studier and Moffatt, 1986)</td>
</tr>
<tr>
<td><em>H. volcanii</em> DS70</td>
<td></td>
<td>(Wendoloski et al., 2001)</td>
</tr>
<tr>
<td><em>H. volcanii</em> H26</td>
<td>DS70 ΔpyrE2</td>
<td>(Allers et al., 2004)</td>
</tr>
<tr>
<td>VDC2376</td>
<td>H26 ΔHVO_1989</td>
<td>This work</td>
</tr>
<tr>
<td>VDC2604</td>
<td>H26 ΔHVO_1989 pIKB421</td>
<td>This work</td>
</tr>
</tbody>
</table>

**Table 2.3 – Plasmids used in Chapter 2**

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTA131</td>
<td>AmpR, ColE1</td>
<td>(Allers et al., 2004)</td>
</tr>
<tr>
<td>pJAM202</td>
<td><em>H. volcanii</em> shuttle vector, NovR</td>
<td>(Kaczowka and Maupin-Furlow, 2003)</td>
</tr>
</tbody>
</table>
### Table 2.3 – Plasmids used in Chapter 2 (continued)

<table>
<thead>
<tr>
<th>Plasmid (Name)</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pIKB227</td>
<td>649 bp upstream and 610 bp downstream of \textit{HVO_1989} recombed between the \textit{EcoRI} and \textit{XhoI} sites of pTA131</td>
<td>This work</td>
</tr>
<tr>
<td>pIKB421</td>
<td>591 bp fragment encoding \textit{VNG_1980C} inserted between the \textit{BspI} and \textit{NdeI} sites of pJAM202</td>
<td>This work</td>
</tr>
<tr>
<td>pET28Mj1640</td>
<td>Mj1640 expressing clone Km\textsuperscript{R}, ColE1</td>
<td>(Chen and Yuan, 2010)</td>
</tr>
</tbody>
</table>

### Table 2.4 – Oligonucleotides used in Chapter 2

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Sequence</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{HVO_1989_N_IfKO_Fwd}</td>
<td>cgggccccccccctcgagcgaggtcgtgttcc</td>
<td>Amplification of region upstream of \textit{HVO_1989}</td>
</tr>
<tr>
<td>\textit{HVO_1989_N_IfKO_Rev}</td>
<td>gacgcgttcatatgcatgtcgtcgagcgagaagtc</td>
<td></td>
</tr>
<tr>
<td>\textit{HVO_1989_C_IfKO_Fwd}</td>
<td>gcatatgaacgcgtcattacggtcgcgcacaacta</td>
<td>Amplification of region downstream of \textit{HVO_1989}</td>
</tr>
<tr>
<td>\textit{HVO_1989_C_IfKO_Rev}</td>
<td>cgggctgcaggaattcttgcaacgactgtccaaaac</td>
<td></td>
</tr>
<tr>
<td>\textit{HvO_1989_Int_F}</td>
<td>tctgtcgctgtgcaactcc</td>
<td>Internal checks for \textit{HVO_1989} deletion</td>
</tr>
<tr>
<td>\textit{HvO_1989_Int_R}</td>
<td>gtcagcgcgctggaagatg</td>
<td></td>
</tr>
<tr>
<td>\textit{HvO_1989_Ext_F}</td>
<td>aacggcccatcaggtgaagta</td>
<td>External checks for \textit{HVO_1989} deletion</td>
</tr>
<tr>
<td>\textit{HvO_1989_Ext_R}</td>
<td>gtcagcgcgctggaagatg</td>
<td></td>
</tr>
<tr>
<td>\textit{Hsal_cog1901_Fwd}</td>
<td>atgct catatg atgagacagttgctgtcct</td>
<td>Amplification of VNG1980C from \textit{Halobacterium} sp. NRC-1 for insertion into pJAM202.</td>
</tr>
<tr>
<td>\textit{Hsal_cog1901_Rev}</td>
<td>atgct gtcagc tcagtcgctggaacccggg</td>
<td></td>
</tr>
<tr>
<td>\textit{Met-CCA2}</td>
<td>tgggccccccccccgtggcc</td>
<td>To characterize residue at position 54 of tRNA\textsuperscript{Met} by CMCT-RT reaction</td>
</tr>
</tbody>
</table>
2.3 RESULTS

2.3.1 trmY is not essential in *H. volcanii* but deletion of *trmY* leads to the absence of m$^{1}$Ψ54 in *H. volcanii* tRNA

The *H. volcanii* *trmY* (COG1901 homolog, *HVO_1989*) was deleted using previously published methods (Allers et al., 2004). Deletion of *trmY* in this strain (VDC2376) was confirmed by PCR using two primer pairs, one designed to anneal outside the deleted fragment and the other within (Fig. 2A). *trmY* is therefore dispensable in *H. volcanii*. Deletion of *trmY* did not lead to any obvious phenotype when analyzing growth in rich or minimal liquid medium or on solid medium at 20, 30, 37, 44 and 50°C or with total salt concentrations ranging from 12% (1.812 M NaCl, 65.28 mM MgCl$_2$, 63 mM MgSO$_4$, and 41.64 mM KCl) to 25% (3.775 M NaCl, 136 mM MgCl$_2$, 131.25 mM MgSO$_4$, and 86.75 mM KCl) (compared to the standard YPC media with final concentration of 18% salt; data not shown). To test if TrmY was involved in m$^{1}$Ψ formation, we first extracted bulk tRNA from VDC2376 (Δ*trmY*) and H26 (the isogenic parent strain as wild type - WT). The bulk tRNA preparations were enzymatically hydrolyzed, dephosphorylated, and the ribonucleosides analyzed by LC/MS-MS as described previously (de Crécy-Lagard et al., 2010). The 259 m/z ion that corresponds to the protonated molecular weight (MH$^+$) of m$^{1}$Ψ was detected by MS at 13.36 min in the WT background, while no 259 m/z ion was detected in the Δ*trmY* strain (Fig. 2B). However, in our set-up, the m$^{1}$Ψ peak eluted as a trail of the U peak at 13-13.6 min and these two peaks were difficult to separate. A two-dimensional thin layer chromatography (2D-TLC) method was therefore used to detect the m$^{1}$Ψ modification. Also, to rescue the effects of the *trmY* deletion, a derivative of the Δ*trmY* strain was
Figure 2.2. Construction and phenotype of the trmY-deleted strain of *H. volcanii*. (A) Deletion of the *trmY* gene was confirmed by PCR; (left panel) PCR products using primers designed to anneal outside the gene (HvO_1989_Ext_F and HvO_1989_Ext_R) confirm a genomic rearrangement of correctly predicted sizes in WT and mutant strains; (middle panel) PCR products using primers designed to anneal within the target gene (HvO_1989_Int_F and HvO_1989_Int_R) show the absence of the *trmY* internal segment in the mutant and its presence in the WT strain; (right panel). To confirm the presence of VNG_1980c in trans, primers were designed to anneal to the complementing gene (Hsal_COG1901_Fwd and Hsal_COG1901_Rev). The predicted size of the fragment is observed in the rescued strain. (B) LC-MS/MS analysis of tRNA extracted from H26 (wild type) and VDC2376 (∆trmY) showing the UV trace at 254nm (top panels). The 259 m/z ion that corresponds to the protonated molecular weight (MH⁺) of m¹Ψ was detected by MS at 13.36 min in the WT background, while no 259 m/z ion was detected in the ∆trmY strain (bottom panels). The peak observed in the mutant strain and not in the WT strain between 14 and 14.5 min corresponds to background (see the differences in the intensity scales between the right and left panels).

created VDC2604), which contained a plasmid-borne copy of the COG1901 encoding gene from *Halobacterium sp* NRC1, VNG_1980C (Fig. 2.1A). The WT, ∆trmY (VDC2376) and ∆trmY+pHTrmY (VDC2604) strains were in vivo labeled with ³²P. Total tRNA was isolated, digested by nuclease P1 and resolved by 2D-TLC. As shown in Fig. 2.3A, radioactive m¹Ψ spot is present in both WT and complemented strains but is
absent from the ΔtrmY strain (VDC2376). These results show that, at least in halophilic Archaea, COG1901 proteins (TrmY) are involved in the formation of m¹Ψ in tRNA.

The position of the non-methylated Ψ in the tRNA of the ΔtrmY strain was determined by CMCT-primer extension reactions, which allow the identification of Ψ at specific positions in RNAs (Bakin and Ofengand, 1998). As seen in Fig. 2.3B, the Ψ-specific band (marked by an arrow) is present at position 54 of tRNA in the ΔtrmY strain but not in the WT and complemented strains, as expected if m¹Ψ54 and not Ψ54 is present in these last two strains. These results also suggest that the lack of TrmY only affects methylation and not Ψ formation at this position. It can also be noted that the presence of Ψ54 instead of m¹Ψ54 in tRNAs causes subtle structural change in the TΨ-loop, such that reverse transcriptase reaction stops at position 56 become more frequent in the ΔtrmY strain compared to the other two strains. (Note: a dark band, marked by an asterisk, at position 56 in 2 min lane in ΔtrmY, but not in other two strains). This effect was reproducible in three independent reactions. Normally H. volcanii as other archaeal tRNAs, contain ribose-methylated C (Cm) at position 56, instead of the unmodified C56 found in Eukarya and Bacteria (Renalier et al., 2005). Because of the increased hydrophobicity of 2’O-methylated residue, the structural flexibility of the Cm56 containing TΨ-loop and thus the ease of reverse transcription is probably slightly different when Ψ54 instead of m¹Ψ54 is present in the tRNA molecule.

2.3.2 M. jannaschii TrmY catalyzes the formation of m¹Ψ54 in tRNA in vitro

The [α-32P]UTP-labeled tRNA^{Trp} substrate, generated by transcription of a plasmid carrying a synthetic H. volcanii tRNA^{Trp} lacking its intron, was first modified by
purified recombinant *M. jannaschii* Pus10, thus converting U at both positions 54 and 55 to \( \Psi' \), as described before (Gurha and Gupta, 2008a). This in vitro pseudouridylated tRNA\(^{\text{Trp}}\) was used as a substrate for further enzymatic methylation by the recombinant *M. jannaschii* TrmY protein prepared under renaturing conditions. The doubly modified RNA was then phenol extracted, digested with nuclease P1 and the resulting digested products analyzed by 2D-TLC. Autoradiograms of the TLC plates revealed the presence of a radiolabeled spot corresponding to pm\(^1\Psi'\) only when the tRNA\(^{\text{Trp}}\) transcript was first incubated with Pus10 and then with TrmY (Fig. 2.4B). When tRNA\(^{\text{Trp}}\) was treated TrmY without pretreatment with Pus10 it did not show any pm\(^1\Psi'\) (data not shown). Nearest neighbor analyses using RNase T2 digestion on the same radiolabeled tRNA samples confirmed the presence of m\(^1\Psi'p\) only in the tRNA\(^{\text{Trp}}\) incubated with both Pus10 and TrmY (Fig 2.4C), attesting that within tRNA\(^{\text{Trp}}\), the m\(^1\Psi'\) residue was located on 5' side of a \(^{32}\text{P}\)-labeled-uridine residue. Indeed, RNase T2 produces ribonucleotide 3'-monophosphates (Np) and in the process transfers labeled phosphate on the 5' side of a residue in the RNA to the 3' side of the preceding residue. In *H. volcanii* tRNA\(^{\text{Trp}}\) there is only one U residue (position 54) that precedes another U residue (position 55) (see Fig. 2.4A). Therefore, the m\(^1\Psi'p\) observed in Fig. 2.4C is obviously derived from the residue at position 54 of this RNA. Moreover, \( \Psi'p \) is absent in the right most panel of Fig. 2.4C, indicating that the methylation reaction was very efficient, all \( \Psi'54 \) being converted into m\(^1\Psi'54\). This methylation is AdoMet-dependent, as it was not observed when
Figure 2.3. Formation of $m^1\Psi$ of *H. volcanii* tRNA is mediated by TrmY. (A) Nuclease P1 digests of uniformly labeled tRNAs were resolved by 2D-TLC. pA, pC, pG, pU, p$\Psi$ and pm$^1\Psi$ indicate 5'-phosphorylated A, C, G, U, $\Psi$ and $m^1\Psi$, respectively. The radioactive spot corresponding to pm$^1\Psi$ is
present in both wild type (H26) and complemented (∆trmY+pHTrmY) strains but is absent from ∆trmY strain (middle panel). (B) CMCT-primer extension analyses to determine the modification status of residue at position 54 of H. volcanii elongator tRNA^{Met} were done using primer Met-CCA2 and total small RNA of wild type, ∆trmY and ∆trmY+pHTrmY strains. RNAs were treated with (+) or without (-) CMCT for the indicated time (in min), followed by alkali (OH-) treatment (+) or no treatment (-). Positions of tRNA residues 54 and 55 are marked on the side. A dark band in CMCT followed by alkali treatment lanes, with an increased intensity in the 20 min treatment lane, indicates the presence of Ψ at that position. These reactions show that unmethylated Ψs are present at position 55 in all three strains but at position 54 only in ∆trmY strain (the band is marked by an arrow). Asterisk indicates the specific band produced by reverse transcriptase stop in ∆trmY, but not in other strains.

AdoMet was omitted from the reaction mixture (data not shown). The same results were obtained by using molar excess of recombinant TrmY protein prepared under native conditions, except that because this protein contains bound AdoMet (Chen and Yuan, 2010), addition of exogenous AdoMet in the reaction was no longer required (data not shown). Furthermore, methylation by AdoMet-containing TrmY could be inhibited by addition of S-adenosyl-homocysteine, an inhibitor of SAM-dependent methyltransferases (data not shown). Similar methylations were obtained when H. volcanii elongator tRNA^{Met} was used as substrate instead of H. volcanii tRNA^{Trp}. This tRNA^{Met} contains residues G58 and G60 instead of the conserved A58 and semiconserved pyrimidine-60, at these positions of the TΨ-loop (Fig. 2.4A). M. jannaschii TrmY converts only tRNA Ψ54 and not Ψ55 to m^1Ψ. This is shown by TLC analyses of RNase T2 digestion of [α-32P]CTP-labeled tRNA^{Trp} first incubated with Pus10 then with TrmY (Fig. 2.4D). Autoradiography of the TLC plate reveals the presence of labeled Ψp and the absence of m^1Ψp, indicating that Ψ55 is not converted
to $m^1\Psi$ in this tRNA. Also, analyses of a nuclease P1 digest obtained from a [$\alpha$-$^{32}$P]UTP-labeled mutant tRNA$^{\text{Met}}$ harboring A54 instead of U54 (Fig. 2.4A) and incubated as above with Pus10 and TrmY, reveals the absence of $p$m$^1\Psi$ (data not shown), further confirming that TrmY produces $m^1\Psi$ only at position 54.

### 2.3.3 *M. jannaschii* TrmY can produce $m^1\Psi$ in just T-$\Psi$-arm (stem-loop) of tRNA

Previously it was shown that *M. jannaschii* Pus10 can produce $\Psi$ in the T-arm-Trp (Fig. 2.4A), a 17 base fragment of tRNA$^{\text{Trp}}$, at positions that correspond to positions 54 and 55 of tRNA (Gurha and Gupta, 2008a). The same pseudouridylated fragment and the equivalent fragment from tRNA$^{\text{Met}}$ (Fig. 2.4A) are both substrates for the *M. jannaschii* TrmY (Fig. 2.4E and data not shown). Methylation of $\Psi$ in Pus10-treated T-arm-Trp occurs only at positions equivalent to 54 and not 55 of the tRNA (data not shown).
FIGURE 2.5. Phylogeny of the COG1901 family and structural comparison of Mj1640 dimer with VV2_1434 dimer. (A) Unrooted Bayesian tree of 42 bacterial and 72 archaeal proteins identified as COG1901 in Genbank and are listed in Supplemental file 1. For clarity, only taxa from major clades are labeled. The scale bar indicates the average number of substitutions per site. Numbers at branches represent posterior probabilities as inferred by Mr.Bayes; for clarity, only major branches are labeled. The division between Bacteria and Archaea is well supported. (*) A. borkumensis SK2; (+) S. baltica OS155, OS183 and OS185. (B) ribbon representation of Mj1640 dimer (left panel); Electrostatic surface potential presentation of Mj1640 dimer (right panel), with blue and red colors corresponding to positively and negatively charged patches, respectively. (C) Ribbon (left panel) and Electrostatic surface potential representation (right panel) of VV2_1434 dimer.

2.3.4 *M. jannaschii* TrmY requires $\Psi55$ in the tRNA to produce $m^1\Psi54$

In a mutant tRNA$^{Trp}$ where U55 is changed to any other residue (C, A or G), the U54 can still be converted to $\Psi54$ by *M. jannaschii* Pus10 (Gurha and Gupta, 2008a).
However, $\Psi^{54}$ of a mutant tRNA$^{Trp}$, where the sU55 is changed to A55 (Fig. 2.4A), is not a substrate for *M. jannaschii* TrmY. Nuclease P1 digest of [α-$^{32}$P]UTP-labeled transcript of this mutant tRNA does not show any $pm^{1}\Psi$ after treatment with TrmY (Fig.2.4F). Similarly a U-to-G change at position 55 did not lead to any $m^{1}\Psi$ formation, while a trace amount of $pm^{1}\Psi$ was observed when U55 was changed to C55 (data not shown). These experiments suggest that the presence of $\Psi^{55}$ in the tRNA is required for enzymatic methylation of the adjacent $\Psi^{54}$ by TrmY to produce $m^{1}\Psi^{54}$. However, at present we cannot distinguish whether an unmodified U55 can substitute for the requirement of $\Psi^{55}$, because Pus10 treatment of RNA converts both U54 and U55 in our substrates to $\Psi$ (Gurha and Gupta, 2008a).

2.4 DISCUSSION

In the present work, by combining genetic and biochemical experiments, we showed that in at least two Euryarchaeota, *H. volcanii* and *M. jannaschii* COG1901 genes encode the missing tRNA (pseudouridine54-N1)-methyltransferase or TrmY. This newly identified TrmY enzyme belongs to the SAM-dependent SPOUT-super family of dimeric enzymes (Tkaczuk et al., 2007). This superfamily encompasses RNA methyltransferases displaying various base and RNA specificities, such as TrmC56 catalyzing the ribose methylation of C56 (Cm) in T$\Psi$-loop of archael tRNA, TrmD catalyzing the N1-methylation of G34 ($m^{1}\text{G}$) in anticodon loop of bacterial tRNA and RlmH catalyzing N3-methylation of uridine or pseudouridine ($m^{3}\text{U}/m^{3}\Psi$) in helix 69 of bacterial 23S rRNA (Tkaczuk et al., 2007). It was proposed that these different SPOUT-
methyltransferases originated from a common, possibly promiscuous ancestral precursor, which after multiple gene duplications, have later diverged to adapt to more specific cellular functions, the SPOUT domain being itself traceable to the last universal common ancestor (LUCA – (Anantharaman et al., 2002; Tkaczuk et al., 2007)). The closest COG1901 homologs among SPOUT-methyltransferase of known catalytic function is the newly identified dimeric Nep1 (COG1756) catalyzing the N1-methylation of ψ (m^1ψ) in helix 35 of 18S rRNA of Eukarya and Archaea (Fig. 2.1C, (Wurm et al., 2010)). Hence, these two families represent yet another example of two closely related SPOUT subfamilies that target different RNA molecules (an rRNA and a tRNA).

_**M. jannaschii** TrmY can produce m^1ψ54 in full-sized tRNAs as well as at the equivalent position in Tψ-stem-loop substrates (17-mer Tψ-arm of tRNA). Likewise, the tRNA:m^5U54 methyltransferases of _E. coli_ (TrmA or RumT), of _S. cerevisiae_ (Trm2p) and of _P. abyssi_ (PAB_0719, aTrmU54) can produce m^5U54 in similar full-sized or truncated RNAs (Becker et al., 1997; Gu et al., 1996; Urbonavicius et al., 2008), suggesting a possibility of similar substrate recognition modes as described for _E. coli_ TrmA (Alian et al., 2008) and archaeal TrmU54 (Walbott et al., 2008). However, archaeal TrmY can methylate ψ54 in tRNA harboring G58 and G60 (Fig. 2.4A). This is different from the yeast Trm2p enzyme, that requires the presence of both A58 and a of pyrimidine at position 60 to methylate (Becker et al., 1997). This specificity of yeast Trm2p probably depends on the formation of a reverse intra-loop Hoogsteen pair between U54 and A58 in the tRNAs (Becker et al., 1997). The _M. jannaschii_ TrmY requires a ψ at position 55 to efficiently methylate ψ54. This is reminiscent of the requirements of the yeast Trm2p enzyme, where U55 cannot be replaced by any other
residue (Becker et al., 1997). We could not determine whether the residue 3’ adjacent to the target Ψ residue needs to be Ψ or an unmodified U, because of our experimental design, but in the case of the archaeal Nep1, it was shown that the 5’ adjacent residue to the target Ψ residue can be an unmodified U (Wurm et al., 2010).

In addition to *M. jannaschii* TrmY, high resolution crystal structures of other SAM-dependent RNA methyltransferases, such as Nep1 of *M. jannaschii* and *S. cerevisiae*, TrmA and RumA both of *E. coli*, are known (Alian et al., 2008; Kealey et al., 1991; Leulliot et al., 2008; Taylor et al., 2008; Thomas et al., 2011). Nep1 is an rRNA N1-Ψ methyltransferase, TrmA is a tRNA C5-m⁵U methyltransferase, whereas RumA is an rRNA C5-m⁵U tRNA methyltransferase. Structural analysis of these RNA methyltransferases reveals that TrmY and Nep1 belong to SPOUT-family, whereas TrmA and RumA belong to Rossmann (α/β) fold. These two groups share less structural similarities, except a somehow similar α/β fold for AdoMet and RNA substrate binding. Although TrmY and Nep1 share the same characteristic deep trefoil knot core, different structural domains/motifs extended from the deep trefoil knot core are observed. These extra domains/motifs are proposed to facilitate RNA substrate binding and/or selection. Similarly, TrmA and RumA share close structural similarity both in RNA binding and catalytic domains. However, significant structural differences of the extended loop linking the RNA binding domain and the catalytic domain are observed, suggesting the functional role of this loop in substrate selection.

TrmY homologs are found in the majority of Euryarchaeota and in a few Crenoarchaeota. All Archaea that contain a TrmY homolog contain a Pus10 homolog (see the COG1901 Subsystem in the SEED database)
consistent with our finding that the presence of $\Psi$ at position 54 is a prerequisite to the formation of $m^1\Psi$. The presence or absence of $m^1\Psi$ in archaeal tRNAs, has been reported by different research groups over many years (Table 2.1). The presence of $\Psi$ or $m^1\Psi$ correlates well with the presence and absence of Pus10 and TrmY with the exception of the Thermococcales (Table 2.1). Despite the presence of both $pus10$ and $trmY$ genes in their genomes (Table 2.1 and COG1901 Subsystem in the SEED database), $m^5U$ or $s^2T$ and not $m^1\Psi$ was formally identified in bulk tRNAs from Thermococcales (Table 2.1 and Fig. 2.1B). Likewise, when a tRNA transcript specific for Phe, Asp or Ile, lacking modifications, was incubated with AdoMet and cell extracts from $P. furiosus$ or $P. abyssi$, $m^5U$ was almost quantitatively formed at position 54 of the $T\Psi$-loop (Constantinesco et al., 1999; Urbonavicius et al., 2008). Under the same experimental conditions, U54 in the tRNA$^{Ile}$ transcript was shown to be ‘doubly’ modified into $m^1\Psi$ when incubated with cell extracts from $H. volcanii$ (Grosjean et al., 1995). The enzyme responsible for the formation of $m^5U54$ in $P. abyssi$ was recently identified as PAB_0719 (aTrmU54) (Auxilien et al., 2011; Urbonavicius et al., 2008). This protein is a member of the COG2265 SAM-dependent, Rossmann-like RNA $m^5U$ methyltransferase family. Phylogenetic analysis showed that it has been acquired by lateral gene transfer of a bacterial RlmD gene responsible for the site-specific formation of $m^5U$ in 23S rRNA (position 1939 in $E. coli$). Thus, during evolution, the RlmD-like protein in $P. abyssi$ (PAB_0719, aTrmU54) has changed target specificity from rRNA to tRNA. The question of the role of TrmY proteins in Thermococcales remains open. Are these still active as a tRNA $m^1\Psi$ methyltransferase, but in specific conditions that have escaped detection, or
have they acquired another function? Further experimental work is required to address this issue.

Only a few sequenced Crenarchaeota encode both genes *pus10* and *trmY* (COG1901 Subsystem in SEED), these include *Ignococcus hospitalis* KIN4_1, *Aeroyrum pernix* K1, the *Pyrobaculum* species and a few other (but not all) Thermoproteales. Indeed $^1\Psi$54 was detected in bulk tRNAs extracted from two thermoproteales (*Pyrobaculum islandicum* and *Thermoproteus neutrophilus*) fitting with the genomic data (Table 2.1). The only other experimental data available for a Crenarchaeota is for *S. acidocaldarius* where 2'-O-methyluridine at tRNA position 54 (Um54) was identified (Gupta and Woese, 1980; Pang et al., 1982), there again fitting with the absence of both active Pus10 proteins and TrmY in Sulfolobales (Table 2.1).

While COG1901 is primarily found in Archaea, homologs are surprisingly found in most sequenced Vibrio genomes, in two *Shewanella baltica* strains (OS195 and OS185), two *Photobacterium profundum* strains (SS9 and 3TCK) and in *Alcanivorax borkumensis* SK2 (Fig. 5A and COG1901 subsystem in the SEED database). $^1\Psi$54 is typically an archaeal tRNA signature modification (Gupta and Woese, 1980; Pang et al., 1982) and has never been identified in any bacterial tRNA sequenced so far (reviewed in (Grosjean et al., 2008)). Therefore, the presence of TrmY was not expected in bacterial genomes. In addition, the gene coding for the enzyme TrmA responsible for the formation of the canonical $^5$U54 in the T-loop of bacterial tRNA (Edvardsson et al., 2003) is found in all sequenced Vibrio genomes (VV1_1171 in *Vibrio vulnificus*, see COG1901 subsystem in the SEED database). The presence of COG1901 homologs in just a few bacteria could suggest the gene family arose from recent horizontal gene
transfer event from Archaea. However, the phylogenetic analysis of the COG1901 family (Fig 2.5A) shows that bacterial and archaeal COG1901 proteins form two distinct monophyletic groups, which is not easily compatible with the above hypothesis.

Although the overall fold of VV2_1434 (COG1901 homolog in V. vulnificus) shares close structural similarity to that of Mj1640 (Z-score 26.0, r.m.s. 1.6 Å, 181 Cα), significant differences for local conserved motifs and overall charge distributions are observed between these two structures (Fig. 2.5B). In general, VV2_1434 has a more compact structure compared to Mj1640 suggesting that the bacterial and archaeal homologs may fulfill different functions. For example, Mj1640 displays short β-strands with flexible loops at the middle part of the β–sheet compared to those of VV2_1434, whereas VV2_1434 displays an extra α-helix at its “head” of the overall “butterfly” shape (Fig. 2.5B). Further experimental work is required to establish the function of the bacterial COG1901 members.

This work generally reinforces the difficulty of transferring functional annotations of RNA methylases. TrmY are homologs of Nep1 but carry the same reaction (methylation of a Ψ residue) in different substrates, one tRNA the other rRNA. Also, even if we have functionally characterized the function of COG1901 as TrmY in a few Archaea, some of the COG1901 homologs (for example in Thermococcales or Bacteria) might have another function. Experimental validation of judiciously chosen members is the only solution to correctly annotating RNA methylase superfamilies.
CHAPTER 3
DYNAMICS OF ARCHAEAL BOX C/D MODIFICATION

3.1 INTRODUCTION

Methylation of the 2’-OH group of ribose sugars is one of the most predominant post-transcriptional modifications of rRNA and other RNAs in eukarya and archaea. Most of these 2’-O- methylations are carried out by small nucleolar RNA (snoRNAs) in eukaryal and snoRNA like RNAs (sRNAs) in archaena (Bachellerie et al., 2002; Decatur and Fournier, 2003; Henras et al., 2004; Kiss, 2002; Matera et al., 2007; Maxwell and Fournier, 1995; Terns and Terns, 2002; Tran et al., 2004). These s(no)RNAs are called box C/D RNAs because of the presence of consensus box C and box D motif at the termini and imperfect repeats of these consensus sequences called C’ and D’ boxes located internally to the RNA molecules. These box C/D RNAs are functional only as components of ribonucleoprotein (sRNPs), whereby the RNAs associate with an eclectic set of proteins and carry out the relevant methylation reactions. These s(no)RNPs show a remarkable division of labor where the specificity of the guide RNA is determined by a stretch of sequences (about 10-21 nucleotide long) upstream of box D and box D’ binding to complementary positions in the target RNA and then the protein members of the s(no)RNP bringing about the catalysis. It is the precise nucleotide on the target RNA which pairs with the base located 5 bases upstream of box D or box D’ of the guide RNA, which is methylated (Cavaille et al., 1996; Kiss-Laszlo et al., 1996; Kiss, 2002; Reichow et al., 2007; Tycowski et al., 1996).
Eukaryal snoRNPs are formed by the asymmetric assembly of Fibrillarin (the methyltransferase), Nop56, Nop58 and 15.5 kDa protein (Snu13 in Yeast) (Cahill et al., 2002; Szewczak et al., 2002). Archaeal sRNPs on the other hand are more symmetric in structure with all the core proteins binding to both the consensus motifs (Aittaleb et al., 2003; Aittaleb et al., 2004; Dennis and Omer, 2005; Dennis et al., 2001). Archaeal homolog of the 15.5 kDa protein, L7Ae, recognizes a Kink-Turn motif formed by the folding of the box C and box D sequences in the sRNA and through an induced fit mechanism binds to this motif changing the relaxed conformation of the RNA towards a highly kinked form (Suryadi et al., 2005). This binding of L7Ae occurs in a cooperative manner, box C'/D' being occupied first followed by box C/D (Singh et al., 2004) exposing the corresponding guide sequences in the order the box motifs were occupied (Figure 3.1). This exposed guide sequences do not pair with the target sequences in the presence of L7Ae alone. However, this altered RNA conformation of guide RNA with exposed guide sequences is recognized by aNop5p (orthologs of both Nop56 and Nop58 protein) (Omer et al., 2002; Rashid et al., 2003; Singh et al., 2004; Tran et al., 2003) or by a Nop5p-aFib heterodimer (Aittaleb et al., 2003; Aittaleb et al., 2004; Rashid et al., 2003). Studies have pointed that aNop5p-aFib forms an extremely strong heterodimer in vivo even in the absence of RNA (Lechertier et al., 2009). Moreover multiple lines of evidence suggest that aNop5p binding to aFibrillarin has dual purpose. aNop5p not only helps recruit aFib to the guide assembling complex, but it also helps stabilize AdoMet (methyl group donor) binding to aFib (Lapinaite et al., 2013). Hence, it seems highly probable that it is aNop5p-aFib heterodimer that actually binds to an
assembling sRNP. After the assembly is complete, the proteins are arranged towards sequences that hybridizes with the target, upstream of the D box and D' box

**Figure 3.1. A proposed mechanism for the assembly of a functional archaeal box C/D sRNP complex (Singh et al., 2008).** The arrows demonstrate assembly an archaeal box C/D sRNP in a stepwise manner. The proteins are represented with colors as indicated in the figure key. The bifurcation of the pathway after aNop5p binding indicate the target binding to as assembling guide sRNP can occur either before aFib (archaeal Fibrillarin ) binding or after incorporation of aFib. The shaded box indicates unknown changes in the region of the RNA between D’ and C’ boxes which may accompany the modification reaction.

in such a way that the active site of aFib is in close vicinity to the target nucleotide (Aittaleb et al., 2003). This architecture of the sRNP will thus be able to direct effective methylation of specific target residues.
Though several studies of guide sRNP assembly are available, comparatively little is known about the mechanism by which this modification is carried out.

Figure 3.2. Ribbon diagram of the C-Terminal domain of aNop5p superimposed on guide RNA. The GAEK motif is green and the ALFR motif in the Helix $\alpha 9'$ is shown in yellow. The remaining helices of the NOP domain are shown in white. The guide RNA strand is shown in Pink with the aromatic side chains of the bases projecting out of the backbone. The figure was generated using Swiss-pdb Viewer.

Particularly, how a large and complex substrate RNA like rRNA can get access to the active/catalytic site of Fibrillarin through pairing with a highly restrictive guide RNA is one of the key unanswered questions. This is in part because of the unavailability of substrate bound sRNP crystal structures for a long time. Some recent works however, were able to generate crystal structure of guide sRNP loaded with small oligomeric substrates which provided insights to the mechanism of box C/D sRNP mediated methylation of target RNA (Bleichert and Baserga, 2010; Bower-Phipps et al., 2012; Lapinaite et al., 2013; Xue et al., 2012; Ye et al., 2009). These crystal structure models identified a prominent protrusion consisting of a short helix called $\alpha 9'$ between helices
\( \alpha_9 \) and \( \alpha_{10} \) in *Pyrococcus furiosus* aNop5p C-terminal domain which plays a very important role by wedging between the two guide strands and separating them to two different paths which may prevent accidental pairing between the two guides themselves and liberate them for substrate binding (Lin et al., 2011; Xue et al., 2012; Ye et al., 2009). In fact, a very well conserved GAEK motif in helix \( \alpha_9 \) is responsible for prying apart the guide and the non guide strand and its interaction with the RNA is not sequence specific. Moreover, using UV cross linking and Electrospray Ionization mass spectrometric analysis, another study was able to identify an ALFR motif in between helices \( \alpha_9 \) and \( \alpha_{10} \) of *Pyrococcus furiosus* (AFLA in *Methanocaldococcus jannaschii*, see Figure 3.3) which was able to bind to sequences in and around stem II of the box C/D and box C’/D’ motifs, which proved that it was stem II of the Kink-Turn motif that is involved in the sequence specific recruitment of aNop5p (Ghalei et al., 2010). This motif also contacted regions in the single stranded guide region and hence was thought to help present the guide for substrate binding and play a role in guide-target base pairing (Ghalei et al., 2010). The above data were corroborated by NMR and X-ray and Neutron scattering experiments which showed extensive electrostatic interactions between the RNA guide regions and aNop5p that lead to a stabilization of the sRNP complex (Lapinaite et al., 2013). Hence the ALFR and the GAEK motifs, both of which are conserved across a wide range of species (See Figure 3.3), plays critical role in guide-target pairing.

The crystal structures and the NMR studies, majority of which support the di-sRNP model described before (see Chapter 1), envisages that the aNop5p coiled coil domain forms a platform with two RNAs lying at an angle of 45° off this platform.
The box C/D elements are found at the two extremities of this rod shaped structure. The four Fibrillarin molecules rest at the end of the aNop5p coiled coil domain at ends, two copies above the platform defined by the aNop5p proteins on the same side as the sRNA and two copies are on the opposite side. As such only two copies of Fibrillarin will be able to reach the guide RNA for a methylation competent conformation, though all of the four guide RNAs are completely able to bind substrate RNA. Moreover, the presence of a “linker region” between aNop5p N-terminal domain (NTD) and the coiled coil domain gives the N-terminal domain of aNop5p a very high degree of flexibility which forms an “open” and “closed”

**Figure 3.3. Partial multiple Sequence Alignment of Nop5p with representative archaeal and eukaryal homologs.** Sequence alignment of *Methanocaldococcus jannaschii* aNop5p was performed with Clustal Ω. The conserved GAEK and AFLR motifs are highlighted with orange. The representative homologs shown are (from Top to bottom)- *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Mus musculus*, *Homo sapiens*, *Xenopus laevis*, *Archaeoglobus fulgidus*, *Haloferax volcanii*, *Methanocaldococcus jannaschii*, *Pyrococcus furiosus*, *Sulfolobus solfataricus*. 

state of the complex (Gagnon et al., 2012; Oruganti et al., 2007). In a substrate unbound state, aNop5p NTD and Fibrillarin will be distal to the guide RNA, hence making the guide RNA accessible to substrate binding. Once the substrate is bound, Fibrillarin will adopt an “alternate conformation”, swinging into close proximity to site specifically methylate the guide-paired substrate. Both the crystal structure as well as NMR studies predict that substrate RNA cannot interact with the Fibrillarin which is most proximal to it (Lapinaite et al., 2013; Xue et al., 2012). This analysis have led to an interesting hypothesis that di-sRNP catalyze methylation reaction by a cross-sRNP mechanism (Xue et al., 2012). The same model predicts that in the substrate bound di-sRNP, the guide-target duplex is stably anchored by aNop5p and to a certain extent by L7Ae and this guide-target duplex is contacted and modified with Fibrillarin associated with the opposing aNop5p in the other half-mer RNP. This model also suggests that release of L7Ae facilitates substrate binding to the di-sRNP, hence predicting an asymmetric holoenzyme (Xue et al., 2012). However, subsequent NMR studies shows that it is important for L7Ae to be at the site of catalysis so that it can act in concert with the C-Terminal domain of aNop5p bound to the opposing end of the same RNA molecule to confine Fibrillarin to a restricted space on the target RNA and promote site-specific methylations at position 5 nucleotide upstream from the consensus sequence (Lapinaite et al., 2013). Hence, there are differences in observations with regards to the mechanism by which box C/D sRNP recognizes target RNAs.

Previous studies have shown other than box C/D core proteins there are some proteins which associate transiently with the box C/D sRNPs. For example, a pair of mouse proteins designated as p50/55 was shown to bind with the box C/D core
proteins via pull down experiments (Newman et al., 2000). p55 interacts with TATA-binding protein (TBP) and replication A protein (Kanemaki et al., 1997; Qiu et al., 1998). Moreover, both p55 and p50 have DNA helicase activity suggesting a coordination of snoRNA processing and snoRNP assembly with transcriptional events in the nucleus. Besides, p50 and p55 proteins, another RNA helicase Dbp4 was needed to unwind the pairing of two guide RNAs sR41 and U14 with the pre-ribosomal RNA (Kos and Tollervey, 2005). Also, it has been shown that Fibrillarin does not bind methylated duplexes (Lapinaite et al., 2013). The difference in affinities of Fibrillarin for methylated and unmethylated target duplexes will most probably trigger the release of the methyltransferase once the substrate has been modified. All the above observations lead to the speculation that post-modification, the guide sRNP undergoes a series of conformational changes which may help release substrate from the sRNP complex and prime the guide complex for the next round of substrate binding. The association of RNA helicases with the core proteins as well the reduced affinity of Fibrillarin for methylated substrates may help destabilize the guide-target RNA duplex and help unwind the target from the guide initiating catalytic turnover for the guide sRNP complex. It is not clear however whether the core proteins also disassemble from the guide RNP or just RNA member of the guide sRNP complex undergoes conformational change which enables the next round of substrate to bind.

The aim of the current study is to gain mechanistic insights into the catalytic processes of box C/D sRNP. The main focus of the study will be on the mechanism of target recruitment which is still an unresolved issue in the field. This study hopes to provide functional analyses to the structural observations made recently through NMR
and crystal structure studies. Also, whether a guide sRNP undergoes multiple turnover will be addressed in these studies.

3.2 MATERIALS AND METHODS

3.2.1 Generation of DNA templates for in vitro RNA synthesis.

PCR amplified templates were used for in vitro transcription. Different primer combinations and plasmids carrying the respective genes were used to PCR amplify the template for both guide and target generation (See Table 3.1). These PCR products were cleaned of unincorporated nucleotides and Taq enzyme by passing through Sephadex G25 (cut off range of 10 bases) spin columns. For each 100 µl PCR product, 1 ml of Sephadex G25 column was assembled in 1ml syringe tubes, whose mouth were plugged with sterile glass wool. Sephadex was added to the syringe by gentle tapping with a capillary tube to preclude air bubble formation and was spun at 2000 rpm for a minute in a GLC-1 centrifuge to remove stored water from the resin. The columns were equilibrated with 1 ml of autoclaved water and then the PCR products were passed through it.

3.2.2 Site-directed mutagenesis

Mutations in *Haloferax volcanii* pre-tRNA$^{\text{Met}}$ HVM●i36 (Gomes and Gupta, 1997) and *Methanocaldococcus jannaschii* aNop5p genes (Tran et al., 2003) were introduced using Quickchange™ site-directed mutagenesis kit from Stratagene. Complementary forward and reverse primers were designed which covered the site of
Table 3.1 List of RNA substrates and Primers used in Chapter 3

<table>
<thead>
<tr>
<th>RNA substrate</th>
<th>DNA Template</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-tRNA(^{\text{Met}})</strong> (Target RNA)</td>
<td>pHVM plasmid (pre-tRNA(^{\text{Trp}}) gene in Puc19 vector)</td>
<td>T7HVTRP5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAAACGACTCACTATAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCGGGGTTGGCTTAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HVMET3R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGG GCG TGT CCG AAC</td>
</tr>
<tr>
<td><strong>Pre-tRNA(^{\text{Met}})(^{\text{i36}})</strong> (Target RNA with 36 bases deleted from intron)</td>
<td>pHMV(^{\text{i36}}) plasmid (Pre-tRNA(^{\text{Met}})(^{\text{i36}}) in pUC19 vector)</td>
<td>T7HVTRP5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAAACGACTCACTATAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCGGGGTTGGCTTAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HVMET3R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGG GCG TGT CCG AAC</td>
</tr>
<tr>
<td><strong>sR-tMet (73 mer)</strong> (Guide RNA ;wild type)</td>
<td>Genomic DNA from Haloferax volcanii H26 Cells</td>
<td>T7HVMETCD5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAA TAC GAC TCA CTA TAG GCC GAT GAC GAC GGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HALOMETCDR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGC CTC GGC ACT CAT AGG GCT C</td>
</tr>
<tr>
<td><strong>sR-tMet (93 mer)</strong> (Guide RNA with 20 base extension on its 5’end)</td>
<td>Genomic DNA from Haloferax volcanii H26 Cells</td>
<td>T7HVMETCD21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAA TAC GAC TCA CTA TAG GTG CGG AGG GAA AAC GAA GTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HALOMETCDR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGC CTC GGC ACT CAT AGG GCT C</td>
</tr>
<tr>
<td><strong>Mature tRNA(^{\text{Met}})</strong></td>
<td>pHVM(^{\text{i}}) (mature tRNA(^{\text{Met}}) gene in pUC19 vector)</td>
<td>T7HVTRP5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAAACGACTCACTATAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCGGGGTTGGCTTAGC</td>
</tr>
<tr>
<td><strong>3’ stem loop of sR-h45</strong> (aCbf5 associated guide RNA; modifies 1940/1942 in 23S rRNA)</td>
<td>sR-45 3’ stem loop clones in KS(^{\text{Met}})(^{\text{Pro}}) plasmid</td>
<td>HVHA-F1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCT ACC CGG GTG CGT ACC TCA AGT CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1940/43HA-R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCG CAA GCT TGT GTC GCC CAG AAC ACT AAC GGC CG</td>
</tr>
</tbody>
</table>
Table 3.1 List of RNA substrates and Primers used in Chapter 3 (continued)

<table>
<thead>
<tr>
<th>RNA substrate</th>
<th>DNA Template</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Forward</strong></td>
</tr>
<tr>
<td><strong>Δ36i-ΔACA</strong></td>
<td>pHV M Δ36i-ΔACA</td>
<td>HVM-dACA-F</td>
</tr>
<tr>
<td>Target RNA with ACA at 3’ intron-exon junction of <em>H. volcanii</em> elongator pre-tRNA&lt;sub&gt;Met&lt;/sub&gt;Δi36</td>
<td>pUC19 vector</td>
<td>GCC TCG CAC CTG GGT GCG GAG ATC GTG G</td>
</tr>
<tr>
<td><strong>Δ36i-ΔCCC</strong></td>
<td>pHV M Δ36i-ΔCCC</td>
<td>HVM-di39-F</td>
</tr>
<tr>
<td>CCC deletion in intron of Δi36 version of <em>H. volcanii</em> elongator pre-tRNA&lt;sub&gt;Met&lt;/sub&gt;</td>
<td>pUC19 vector</td>
<td>CGC CGC ACT CAT AGG GTT TGC GAG GTC ATG CCG GC</td>
</tr>
<tr>
<td><strong>Δ36i-ΔAGA</strong></td>
<td>pHV M Δ36i-ΔAGA</td>
<td>HVM-dV-AGA-F</td>
</tr>
<tr>
<td>AGA deletion in variable arm (mature tRNA position 44-46) of <em>H. volcanii</em> elongator pre-tRNA&lt;sub&gt;Met&lt;/sub&gt;</td>
<td>pUC19 vector</td>
<td>CGC ACC TGG GAC ATG CGG TCG TGG GTT CCG AGC</td>
</tr>
<tr>
<td><strong>Δ36i-ΔD-Arm</strong></td>
<td>pHV M Δ36i-ΔD-Arm</td>
<td>T7HVM-dd</td>
</tr>
<tr>
<td>D-arm deletion (17 nt, position 10-25) of <em>H. volcanii</em> elongator tRNA&lt;sub&gt;Met&lt;/sub&gt;.</td>
<td>(Pre-tRNA&lt;sub&gt;Met&lt;/sub&gt;Δi36 in pUC19 vector)</td>
<td>TAA TAC GAC TCA CTA TAG CCC GGG TGG CCG CAC TCA TAG G</td>
</tr>
<tr>
<td><strong>Δ36i-ΔT-Arm</strong></td>
<td>pHV M Δ36i-ΔT-Arm</td>
<td>HVM-dT-F</td>
</tr>
<tr>
<td>T-arm (position 47-65) deletion of <em>H. volcanii</em> elongator tRNA&lt;sub&gt;Met&lt;/sub&gt; cloned in pHV plasmids.</td>
<td>(Pre-tRNA&lt;sub&gt;Met&lt;/sub&gt;Δi36 in pUC19 vector)</td>
<td>CCT GGG ACA TGC GGA GAC CCG GGC TGC AGG AAT TCG</td>
</tr>
</tbody>
</table>
mutation hence introducing the desired mutation in the gene. The changes introduced were deletion of sequences as well as introduction of additional sequences in the pre-tRNA\textsuperscript{Met}Δi36 gene (See Table 3.1 and Section 3.3.5). Similarly for aNop5p protein, Alanine substitutions were introduced at positions Gly284 of the conserved GAEK motif. Also, Leu289 and Phe290 were changed to Alanine to disrupt the ALFA motif in the protein (See Table 3.2 and Section 3.3.6)

### Table 3.1 List of RNA substrates and Primers used in Chapter 3 (continued)

<table>
<thead>
<tr>
<th>Δ36i-GA\textsubscript{insertion}</th>
<th>pHVMΔi36 plasmid (Pre-tRNA\textsuperscript{Met}Δi36 in pUC19 vector)</th>
<th>HVM-i66aGGGAA-F</th>
<th>HVM-i66aGGGAA-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGGAA insertion after residue 66 in intron of <em>H. volcanii</em> elongator pre-tRNA\textsuperscript{Met}.</td>
<td>CAT GCC GGC CTC GCG GGA AAC CTG GGA CAT GCG</td>
<td>CGC ATG TCC CAG GTT TCC CGC GAG GCC GGC ATG</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Δ36i-CTA\textsubscript{insertion}</th>
<th>pHVMΔi36 plasmid (Pre-tRNA\textsuperscript{Met}Δi36 in pUC19 vector)</th>
<th>HVM-i69aCTA-F</th>
<th>HVM-i69aCTA-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTA insertion after residue 69 in intron of <em>H. volcanii</em> elongator pre-tRNA\textsuperscript{Met}.</td>
<td>CCG GCC TCG CAC CCT ATG GGA CAT GCG G</td>
<td>CCG CAT GTC CCA TAG GGT GCG AGG CCG G</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2 Protein Mutants, DNA templates and Primers used in Chapter 3

<table>
<thead>
<tr>
<th>Mutant aNop5p</th>
<th>DNA Template</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>aNop5p (GAEK-&gt;AAEK)</strong> Methanocaldococcus jannaschii aNop5p mutant with a change in Gly284 to Alanine</td>
<td>MJaNop5p in pET28a</td>
<td>MJN5-G284A-F</td>
<td>MJN5-G284A-R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCT TCA ACA ATA</td>
<td>GCA AAT AAA GCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAA GTT TTA GCT</td>
<td>TTC TCA GCA GCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCT GAG AAG GCT</td>
<td>AAA ACT TGT ATT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTA TTT GC</td>
<td>GTT GAA GC</td>
</tr>
<tr>
<td><strong>aNop5p (ALFA-&gt;AAAA)</strong> Methanocaldococcus jannaschii aNop5p mutant with changes in Leu289 and Phe290 to Alanine</td>
<td>MJaNop5p in pET28a</td>
<td>MJN5-LF289/90A-F</td>
<td>MJN5-LF289/90A-F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGT GCT GAG AAG</td>
<td>CCC ATC CTT AAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCT GCT GCT GCC</td>
<td>TGG GCA GCA GCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAT TTA AGG ATG GG</td>
<td>GCC TTC TCA GCA GCC</td>
</tr>
</tbody>
</table>

3.2.3 *In vitro* RNA synthesis

To generate 100 µl transcripts approximately 10 µg of PCR amplified DNA was added to the transcription mixture which contains 20 µl of 5X transcription buffer (200 mM Tris-Cl pH 8.0, 10 mM spermidine, 40 mM MgCl$_2$, 250 µg/ml BSA), 1 µl of 0.5 M MgCl$_2$, 2 µl of 1 M DTT, 1 µl of 0.1% Triton X-100, 20 µl of 20 % PEG$_{8000}$ and 1µl of 100 mM ATP, 1µl of 100 mM GTP, 1µl of 100 mM UTP and 100mM CTP) and then incubated with 2 µl of homemade T7 RNA polymerase (concentration unknown). All the components of the transcription reaction (except the enzyme) were bought to room temperature before the reaction. The final volume was made to 100µl with DEPC water. Reaction was incubated for 2 hours at 37°C. 2 µl of 0.5 M EDTA was added to stop the reaction. The RNA was resolved in a 6% denaturing polyacrylamide gel, observed
under a UV torch, eluted and purified by phenol/chloroform extraction and ethanol precipitation.

For synthesizing internal labeled *in vitro* transcripts, the reaction mixture used was the same as described above except that 20 µCi of [α-32P] CTP or [α-32P] ATP and 1µl of 40 mM cold CTP or ATP was added instead of 1 µl of 100mM CTP or ATP. The RNA was observed for eluting by autoradiography.

For synthesizing small transcripts, double stranded DNA templates were used containing the primer T7P which contains the T7 promoter sequence hybridized to a primer whose 3’ end contains a sequence complementary to the promoter sequence followed by the reverse complementary gene sequence (Table 3.1). The hybridized primers were PCR amplified and cleaned up using Sephadex G25 columns and this template was used to generate *in-vitro* transcripts using protocols described above.

### 3.2.4 3’-end labeling of RNA

In a 40 µl reaction, 50 picomole of cold RNA with 10 mM ATP, 5 mM DTT, 20% DMSO, approximately 100 µCi of [5'-32P] pCp, 4µl of 6x cocktail (pH 8.3) (300mM Na-HEPES pH 8.3, 120mM MgCl₂, 0.6mg BSA), 3 µl of T4 RNA Ligase (2U/µl) were added. The reaction was incubated for 3 hours at 37°C and stopped by adding 2 µl of 0.5 M EDTA (pH 8.0). The labeled RNA was eluted from denaturing polyacrylamide gel followed by phenol/chloroform extraction and ethanol precipitation.
3.2.5 5’-end labeling of RNA

50 pmols of RNA dissolved in DEPC treated water was heated at 90°C for 10 minutes to denature it. The RNA was chilled on ice for 2 minutes followed by the addition of 2µl 10X Dephosphorylation Buffer (Promega), 1µl DMSO and 5µl Shrimp Intestinal Phosphatase (SIP; 0.5 U/µl, Promega) to a 20µl final volume. The reaction was incubated at 37°C for an hour. The reaction was stopped by adding 80µl DEPC water, 10µl 10% SDS and 1µl 500mM EDTA. Phenol: chloroform extraction and ethanol precipitation were then performed. After precipitation, the RNA pellets were dissolved in 15 µl of DEPC water and heated at 90°C for 10 minutes. The RNA was chilled on ice for 2 minutes followed by the addition of 2µl 10X PNK Buffer, 1µl DMSO, 100 µCi of [γ-32P]ATP and 1µl T4 Polynucleotide Kinase (T4 PNK; concentration unknown) to a 20µl final volume. The reaction was incubated at 37°C for 60 minutes and stopped by adding 1µl 5mM EDTA and heating at 95°C for 2 mins. The labeled RNA was ran and eluted from denaturing polyacrylamide gel followed by phenol/chloroform extraction and ethanol precipitation.

3.2.6 In Vitro RNP-directed nucleotide 2’-O-Methylation and Thin Layer Chromatography

RNPs assembly and methylation reaction was performed using fixed concentrations of recombinant core proteins of M. jannaschii in presence of specific concentrations of labeled target RNA and cold guide RNA (sR-tMet 93mer) at 68°C for time specified in 20µl reactions containing 20mM Tris-HCl, pH 7.0, 150mM NaCl, 0.75mM DTT, 1.5mM MgCl2, 0.1mM EDTA, 10% glycerol (final concentration). There
are two main reasons why recombinant proteins from the thermophilic euryarchaeon \textit{Methanocaldococcus janaschii} were used in our assays. Halophilic proteins by nature have a high proportion of acidic amino acid residues and most halophilic proteins are inactivated when Na\textsuperscript{+} or K\textsuperscript{+} concentrations in the reaction solutions go below 2M (Madern et al., 2000). Moreover, recombinant proteins from Halophiles do not react well with the negatively charged RNA molecules in low salt concentrations. Therefore, recombinant proteins from \textit{Methanocaldococcus janaschii} were chosen for all our future assays. The concentrations of the guide RNA, Target RNA and core proteins are mentioned in the respective experiments. In some of the reactions, the guide RNA was also labeled. Concentrations of both the guide and target RNA were determined by scintillation counts, rather than spectrophotometric measurements. Reactions were stopped by 200 µl “Stop” buffer containing 0.5 M NH\textsubscript{4}OAc, 0.2 mM EDTA and 0.1% SDS, followed by phenol/chloroform extraction and ethanol precipitation. RNA samples were digested with Nuclease P1 or RNase T2/RNase A and the digested products were resolved on cellulose plates (Merk) using either one or two dimensional TLC. The solvents for TLC were isobutyric acid/0.5 N NH\textsubscript{4}OH (5:3, v/v) for the first dimension and Isopropanol/HCl/H\textsubscript{2}O (70:15:15, v/v/v) for the second dimension.

\textbf{3.2.7 Purification of His-tagged \textit{Methanocaldococcus janaschii} L7Ae and aFibrillarin}

Clones of His-tagged \textit{M. janaschii} L7Ae and aFibrillarin in pET28a vector were transformed in Rosetta DE3 strain of \textit{E. coli} (Tran et al., 2003). Single transformants were picked up and grown in a 500 ml culture. When the culture reached an OD\textsubscript{600} of 0.5, it was induced with freshly prepared 1mM IPTG for 4 hours at 37°C. Harvested cells
were resuspended in 5 ml of binding buffer (20 mM Tris pH 8.0, 2.5 mM Imidazole and 150 mM NaCl, 20% Glycerol). PMSF was added to the suspension. The suspension was then sonicated and soluble proteins were recovered from cell lysate by spinning at 10000 r.p.m. for 10 minutes at 4°C. The cell lysate cleared of cell debris were heated at 60°C for 10 minutes to precipitate *E.coli* proteins and again spun at 10000 rpm for 10 minutes at 4°C. Charged Ni-NTA resins were equilibrated in binding buffer. Supernatant was loaded onto pre-equilibrated Ni-NTA column. Non-specific proteins were removed by washing with at least 5 column volumes of wash buffer (20 mM Tris pH 8.0, 40 mM Imidazole and 150 mM NaCl, 20% Glycerol). Elution was carried out with buffer containing 20 mM Tris pH 8.0, 250 mM Imidazole and 150 mM NaCl, 20% Glycerol. After checking the eluted fractions by SDS-PAGE, fractions with maximum amounts of purified protein were pooled and dialyzed in presence of thrombin at 4°C overnight with 1L of dialysis buffer (20 mM HEPES pH 8.0, 100 mM NaCl, 3 mM MgCl2, 0.2 mM EDTA and 20% glycerol).

### 3.2.8 Purification of *M. jannaschii* aNop5p

The untagged clone of *Methanocaldococcus jannaschii* aNop5p in pET28a vector was transformed in *E. coli* Rosetta DE3 cells (Tran et al., 2003). A 500 ml culture was grown to an OD_{600} of 0.5 at 37°C and shifted to 16°C for 30 mins. The culture was induced with 200 µM IPTG and grown overnight at 16°C. Alternatively, 250 ml culture was grown to an OD_{600} of 0.5 at 37°C and then induced with 300 µM of IPTG for 2 hours at the same temperature. Cells were harvested and resuspended in 6 ml/gram of Buffer A (20 mM Tris.chloride pH 8.0 and 150 mM NaCl). Protease inhibitor cocktail (PepstatinA 2 µM, Leupeptin 0.6µM, Benzamidine 2 mM, Chymostatin 2 µg/ml,
metabisulphite 0.5 mM), along with PMSF 0.5 mM, DTT 0.5 mM (final concentrations) and DNase I (50 µg) and RNase A (50 µg) were added to the cell suspension. It is extremely crucial to add the protease inhibitors as aNop5p degrades very fast post purification, most probably because it is very sensitive to Protease cleavage. Also separate purification apparatus including pump and UV detector should be used to avoid DNasel and RNaseA contaminations. The suspension was then sonicated and the cells lysate was spun for 10 minutes at 10,000 r.p.m. at 4°C. A 20 ml (packed volume) fast flow SP-Sepharose column was equilibrated in Buffer A. The supernatant was loaded on this pre-equilibrated SP-Sepharose column. Extensive washing was carried out with at least 5 column volumes of Buffer B (20 mM Tris.Cl pH 8.0 and 500 mM NaCl). Bound proteins were eluted by a linear gradient of 500 mM to 1 M NaCl (start buffer: 20 mM Tris.Cl pH 8.0 and 500 mM NaCl; end buffer: 20 mM Tris.Cl pH 8.0 and 1 M NaCl). Eluted fractions were checked by SDS-PAGE. Fractions with maximum concentrations of protein were pooled and dialyzed overnight at 4°C against 1 liter of Dialysis buffer (20 mM Tris.chloride pH 7.0, 100 mM NaCl, 3 mM MgCl2, 0.2 mM EDTA and 20% glycerol).

3.2.9 Lead (II) mediated footprinting of RNP complexes

5' labeled or 3'end labeled guide or target RNA (as used in each experiment) was heated at 90°C for 5 minutes and gradually cooled to room temperature to obtain a homogeneous population of uniformly folded RNA. 2pmol of this refolded RNA was incubated in a 36µl reaction and kept at 68°C for 15 min in 20 mM Tris-HCl, pH 7.0, 150 mM NaCl, 0.75 mM DTT, 1.5 mM MgCl2, 0.1 mM EDTA, 10% glycerol and one or more recombinant proteins (concentrations of the proteins mentioned in each experiments) in
presence of 5µl of yeast total RNA. AdoMet (0.01 mM) or AdoHcy (0.42mM) was also included when required. Binding reactions were stopped by chilling on ice. Lead (II) induced cleavage reactions were performed at room temperature for 10 minutes by addition of 4µl freshly prepared 120mM lead acetate solution (0.045g of lead acetate powder dissolved in 100µl of autoclaved ddH₂O to make 120mM lead acetate solution, it is important to make this solution fresh everytime). Cleavage reactions were stopped by addition of 5µl of 0.5 M EDTA. Phenol: chloroform extraction and ethanol precipitation were then performed. RNA fragments were resolved on 12% denaturing, thin (0.1 mm spacer width) sequencing acrylamide gels and bands were detected using a phosphorimager. Addition of any dye is not encouraged. The migration patterns of the dyes present in the size markers indicated when gel running was complete. Also, 50 cm long gel plates were used instead of the 40cm ones normally used, for better resolution of the RNA ends. Lane analyses of footprinting gels were done by ImageQuant software. Image quant software essentially quanititates each band intensity across a specific lane and plots those band intensities along Y axis against their position in the particular lane along X-axis. The values of this plot can be copied to Microsoft EXCEL software by simple Copy Paste command. The band intensities across lanes were then summed up and compared across each lane to preclude loading errors. If the sum of the total lane signals are different between two lanes, then the fold difference between two lanes are quantitated by dividing the two summed values. Then the lane having lower/ higher intensity is adjusted either by multiplying or dividing all the points with that value. Once the two lanes are normalized, they are plotted against each other in Microsoft EXCEL software. A good plot is one where each peak of one lane is perfectly
superimposed on the peak of the other lane at the same position. If the peaks are not superimposed on each other, then some values on the top of the lane values on Microsoft EXCEL needs to be avoided. The changes brought about omitting the values can be observed in real time in the graph.

3.2.10 Iodine – induced cleavage of 5’ labeled CTPαS incorporated RNA in denatured condition

The *in vitro* transcripts were generated as described above. 1µl of 5 mM CTPαS was added along with 1µl of 100 mM ATP, 1µl of 100 mM GTP, 1µl of 100 mM UTP and 100mM CTP. The unlabeled transcripts were purified by running it on a denaturing gel. The purified CTPαS incorporated transcripts were labeled at the 5’ end. Approximately 50,000-100,000 cpm of transcripts was mixed with 1µl 20mM EDTA. The volume was made up to 20µl with 10M urea containing DEPC water. The mixture was heated at 90°C for 2 mins and then cooled down. 1µl of 2mM iodine solution (iodine was dissolved in 100% ethanol) was added to the 20µl solution (final iodine concentration was 0.1mM) at room temperature to start the cleavage. After 2 minutes, the cleavage reaction was stopped by adding 2µl of 2M β-mercaptoethanol. The cleaved transcripts were precipitated with 100% ethanol with 300 mM Sodium Acetate and 0.25% Linear Acrylamide and used in sequencing gel as a C-ladder.

3.2.11 Electrophoretic Mobility shift assay (EMSA)

Approximately 2 pmol of labeled transcripts were incubated at 68°C with varying amount or different concentrations of core proteins and other RNAs (as applicable) in 20µl reactions (20 mM Tris.Cl, pH 8.0, 150 mM NaCl, 0.75 mM DTT, 1.5 mM MgCl2, 0.1
mM EDTA, 10% glycerol, 100 mM Urea) for 15 minutes. Complexes were resolved on native 6% polyacrylamide gel in 0.5X TBE (100 mM Tris.Cl, pH 7.4, 10 mM EDTA, pH 8.0) buffer. The gel was run at 4°C at 100 volts (20 cm x 20 cm dimensions of gel). The bands were visualized by using a phosphorimager.

3.3 RESULTS AND DISCUSSION

3.3.1 In vitro functional analyses of box C/D sRNP

Most of the information pertaining to the molecular mechanism of box C/D mediated catalyses of their cognate target RNAs were obtained from crystal structures of box C/D sRNA formed with small oligomeric substrates (Ghalei et al., 2010; Xue et al., 2012; Ye et al., 2009). Because of the diverse nature of the substrates used, there were some discrepancies that arose from these studies. Hence, while studying the architecture of an archaean sRNP and its dynamics with the target RNA, efforts were made to keep both the guide and the target RNAs to be as physiologically relevant as possible so as to avoid artifacts coming out of these studies.

Previously, our lab was able to delineate the guide and target properties of pre-tRNA\textsuperscript{Trp}, a molecule in \textit{Haloferax volcanii} that methylates 2’-OH of a cytidine at position 34 and Uridine at position 39 with the help of a box C/D sRNP assembled on its intron (Singh et al., 2008; Singh et al., 2004). Studies from our lab also discovered a novel sRNA molecule which catalyzes the methylation of Cytidine 34 in pre-tRNA\textsuperscript{Met} (Joardar et al., 2012). Hence whereas pre-tRNA\textsuperscript{Trp} served as a model to study dual guide modifications (same guide RNA modifying two different targets), sR-tMet served as a model to study single guide modifications.
However, using the wild type 73mer sR-tMet and pre-tRNA\textsuperscript{Met} as a guide and target of choice respectively, were met with new challenges. First of all, in our lead (II)-mediated footprinting assays to study the architecture of sR-tMet sRNP (described later). Because of their close proximities to the terminus, the conformations as well the occupations of box C and box D by core proteins could not be studied properly. Hence, sR-tMet molecule (sR-tMet 93mer) having a 20 base extension on its 5’ end was generated as this would position the proximal box C consensus motifs towards a more internal location in the RNA molecule.

Also, on the substrate front, *in vitro* transcribed pre-tRNA\textsuperscript{Met} though could be modified at the correct positions, was found not to get spliced by partially purified endonuclease (low salt S100 cell extract from *Haloferax volcanii*) most probably due to the improper folding of the full length pre-tRNA\textsuperscript{Met} T7 transcript (Gomes and Gupta, 1997). Among the several derivatives of full length pre-tRNA\textsuperscript{Met} that were subsequently generated, one transcript called HVM\textDelta{i36} having a specific 36 base deletion in the intron was found to be maximally spliced. Hence this molecule was chosen as a possible substrate for our assays.

Since both the guide (sR-tMet 93mer) and the target RNA (HVM\textDelta{i36}) were modified from their wild type forms, it was necessary to see whether they act as a *bona fide* guide and target molecules in an *in vitro* reconstitution system. As such, an *in vitro* modification assay was performed using unlabeled sR-tMet 93 mer as guide and HVM\textDelta{i36} as target. As such, 2 pmole of [\textalpha\textsuperscript{32}P]CTP labeled HVM\textDelta{i36} was incubated with sR-tMet and 4 pmole of all the three recombinant *Methanocaldococcus janaschii* box C/D core proteins and 0.01 mM of AdoMet for 30 minutes at 68°C (Figure 3.4 C
and D). These modified RNAs were digested with Nuclease P1, which will cleave after every base and hence will produce a radiolabeled methylated cytidine (pCm) if the target cytidine 34 is modified. The appearance of pCm spot in Figure 3.4C, proved that the target HVMΔi36 can indeed be modified by sR-tMet93mer.

Figure 3.4 sR-tMet 93mer and HVMΔi36 can act as a model guide-target system to study box C/D mediated catalysis. Sequences and predicted secondary structures of (A) sR-tMet93mer and (B)
HVMΔi36. (C) 2 picomoles of [α-32P]CTP or (D) [α-32P]ATP labeled HVMΔi36 (Δi36 in the figure) was incubated with 1 picomole unlabeled sR-tMet 93mer guide molecule and 4 picomoles of recombinant *Methanocaldococcus janaschii* box C/D core proteins and 0.01 mM AdoMet. (C) TLC analysis of Nuclease P1 digested modified transcript (See Section 3.2.6) showed pCm formation. The “*” represents unknown spot which may represent undigested ribonucleotide. (D) Nearest neighbor analysis by RNase T2 digestion of modified [α-ATP32] HVMΔi36 transcript (See Section 3.2.6) showed CmAp formation.

To further confirm, it was the targeted cytidine at the 34th position and not any non-specific cytidine on the substrate molecule that was modified, RNase T2 digestions were performed on modified [α-ATP32] labeled transcripts. Nearest neighbor analyses by RNase T2 is a classic biochemical technique whereby RNase T2 cleaves at the 3’end of every phosphodiester bonds to produce mononucleotides with a 3’-Phosphate. So basically the enzyme transfers a 5’ phosphate of a ribonucleotide to the 3’end of the base preceding it. In the current transcript used for this nearest neighbor analyses only the phosphates preceding the adenosines are labeled. So ideally, labeled phosphate will be transferred to the bases which precede an Adenosine. RNase T2 is however unable to cleave RNA bases with 2’-O-methylations, in which case it produces a methylated dinucleotide with a 3’-phosphate. However, in the transcript, there are four other positions where a cytidine is followed by an adenosine.

Till date, other than Cm34, there are no reports of any other 2’-O-methylations in bases preceeding adenosines in pre-tRNA^Met from where our substrate is derived from. As can be seen from Figure 3.4 D, a methylated dinucleotide CmAp is formed, which showed it was most likely Cm34 that was the observed modified base in our experiments. These experiments proved that HVMΔi36 was indeed a *bona fide* substrate molecule for sR-
tMet 93 mer which can act as its cognate guide and hence this combination of RNAs were used in all our future functional and structural studies.

3.3.2 sR-tMet 93mer forms stable complexes with Methanocaldococcus janaschii recombinant box C/D core proteins in solution.

Previous studies identified conformational changes that occur in the dual guide box C/D RNA (intron of pre-tRNA\textsuperscript{Trp}) during its assembly to a sRNP and while it is catalyzing the modification of two sites in a truncated target (pre-tRNA\textsuperscript{Trp}\textsubscript{A67}) (Singh et al., 2008). So we planned to employ the same methodologies to study the structural changes that occur in single site guide RNA during its assembly and catalysis. Also, we wanted to compare the structural changes that occur over time in a single site and a double site physiologically relevant nearly full length target RNAs. To assess these conformational changes it was imperative that stable RNA-Protein complexes were formed during our studies. Electrophoretic Mobility Shift Assays (EMSA) provided us with the first clues about the stability of these complexes as well as gave us some structural clues regarding protein assembly on the novel single site guide RNA (sR-tMet) (Joardar et al., 2012).

Since it is already known that L7Ae is the first protein to assemble on a box C/D guide RNA (Joardar et al., 2012; Omer et al., 2002; Singh et al., 2008; Tran et al., 2003), it formed the starting point of our analyses on single site guide-sRNP formation. EMSA assays with 1 picomole of end labeled sR-tMet 73 mer and increasing concentrations of L7Ae proteins (0.01 picomole to 100 picomole) showed three distinct complexes-RNP1, RNP2 and RNP3 (Figure 3.5 A). This most probably indicated that
there are three binding sites for L7Ae on this RNA. This is a bit different from previous studies done on Sr-tMet 73 mer molecule which only showed two Srnp complexes with L7Ae reflecting two binding sites on sR-tMet 73 molecule (Joardar et al., 2012). However, on closer inspection it can be observed that only when L7Ae is present in excess over RNA (10 fold or 100 fold, see Fig 3.5 A, Lanes 10 and 11 respectively), is the third sRNP3 complex observed. Formation of this third complex in the presence of such overwhelming concentrations of protein may indicate non-specific binding of the protein to the RNA or non-specific protein-protein interactions. Lead (ii)-induced cleavage profiles of the guide RNA with such high concentration of the proteins did not reveal any additional protection though, which implied that the super shifted sRNP 3 complex that we observe may be most likely due to protein-protein interactions. In fact, limited lead (II)-induced cleavage reactions of sR-tMet 73mer with these high concentrations...
Figure 3.5 (A) EMSA of sR-tMet titrated with L7Ae. 1 picomole of 5’end labeled sR-tMet was incubated with increasing concentrations of L7Ae (.01picomole to 100 picomoles) at 68°C for 15 minutes and then loaded onto a 4% native gel made with 0.5X TBE. The ratio of RNA:L7Ae in each lanes are as follows- Lane 1 no protein, (2) 100:1, (3) 50:1 (4) 25:1, (5) 10:1 (6) 5:1 (7) 1:1 (8) 1:2 (9) 1:5 (10) 1:10 (11) 1:100. (B) EMSA of sR-tMet with box C/D core proteins. 1 picomole of 5’end labeled sR-tMet was incubated 2µM of box C/D core proteins L7Ae, 2 µM of aNop5p and 2µM aFibrillarin (lane 5) or 6µM Fibrillarin alone or in combination as indicated –Lanes (1) RNA only, (2) RNA +L7Ae, (3) RNA +aNop5p (4) RNA + aFibrillarin, (5) and (6) RNA + L7Ae + aNop5p + a Fib (in concentrations indicated before). Two sRNP complexes were formed- sRNP1 with box C/D RNA binding L7Ae and sRNP2, where we see a super-shifted complex with all the three core proteins. (C) EMSA of sR-tMet with box C/D core proteins in presence of 100 mM Urea. 1 picomole of 5’ end labeled sR-tMet is assembled into ribonucleoprotein complexes in presence of 2 µM of aNop5p (Lane 2), 2µM each of L7Ae,2 µM of aNop5p and 2µM aFibrillarin (Lane 3), 2µM each of L7Ae, aNop5p and aFibrillarin in presence of 100 mM Urea (Lane 4).

Concentrations of L7Ae did not show any additional sites of L7Ae binding (data not shown), which reinforces the notion of protein-protein interactions at such high concentrations of L7Ae.

From above studies, it seems probable that barring very high concentrations, L7Ae is binding to both box C’/D’ and box C/D as observed in various studies done before (Rashid et al., 2003; Tran et al., 2003). However, box C’/D’ in sR-tMet does not
fold into a typical K-loop (Joardar et al., 2012). Hence it seems that L7Ae can bind both
typical and atypical K-Loops again demonstrating a wider substrate recognition ability of
L7Ae compared to its eukaryotic homolog 15.5 kD protein (Gagnon et al., 2010;
Rozhdestvensky et al., 2003). However, it is unclear that whether binding of L7Ae to this
unusual internal C’/D’ motif allows the protein to recruit aNop5p and aFib on the RNA to
form a symmetrical sRNP complex. End labeled sR-tMet 93 mer-L7Ae preassembled
complex was incubated equimolar amounts of aNop5p and aFibrillarin. EMSA showed
that when all the three proteins were present together, a super shifted complex quite
distinct from the L7Ae-RNA complex was formed (Figure 3.5 B, see Lanes 5 and 6).
The concentration of the super shifted complex increases when the concentration of
aFibrillarin is in much excess over aNop5p (2 µM aNop5p to 6 µM aFib, compare Lanes
5 and 6). However, the results were inconsistent and not routinely reproducible.
Whenever aNop5p was present in the assembly reactions, under most circumstances, it
failed to enter the gel (Figure 3.5 B, Lane 3 and Figure 3.5 C, lanes 2 and 3). It is known
that aNop5p is predominantly positively charged and hence it may resistant to
electrophoretic migration towards the positive end of the gel. The retardation of the RNA
by aNop5p, may not reflect true binding of the protein, but rather non-specific
interactions of the predominantly positively charged protein aNop5p with the negatively
charged backbone of the RNA. Addition of 100 mM Urea in the assembly reactions
seemed to help migration of aNop5p-aFibrillarin complex through the gel, as can be
seen by an appearance of a sRNP complex (Figure 3.5 C, lane 4) when the denaturant
is present.
With these studies, though we could not map the exact regions where the box C/D core proteins contact the RNA, we were still able to determine that the box C/D sRNP complexes once assembled were very stable in solution and hence set the premise that these conditions can be used for subsequent studies.

3.3.3 Structural analyses of an in vitro assembled box C/D sRNP by Lead (II)-induced cleavage RNA foot printing technique.

Previous studies from the lab have been able to delineate the roles of guide and target RNAs by creating two derivatives of pre-tRNA$^{\text{Trp}}$ to investigate the conformation of a dual guide archaeal box C/D sRNP (Singh et al., 2008). These studies identified the structural changes in the guide (Intron of pre-tRNA$^{\text{Trp}}$) that occur when it interacts with target RNAs (pre-tRNA$^{\triangle 67}$) during the sequential modifications (Singh et al., 2008). In this study it was observed that L7Ae protected both C' and D' box motifs while exposing the corresponding guide regions. Upon addition of aNop5p to the assembly reactions, the D' guide paired up with the the target. Addition of aFibrillarin to the assembly reactions did not significantly alter the cleavage profiles, showing aFibrillarin made little impact on the sRNP assembly. These interactions among guide RNA, target RNA and the core proteins were studied by using limiting Lead (II) induced cleavage RNA footprinting technique (Gornicki et al., 1989; Lindell et al., 2002). It is one of the very few techniques available to study both RNA-RNA and RNA-Protein interactions. Lead (II) induced cleavage does not discriminate in RNA sequence. Regions of RNA that are paired or protected by proteins are refractory to lead (II) - induced cleavage, while single
stranded regions that are unprotected show enhanced cleavage. Using this technique, we then set out to compare and contrast structural changes that accompany a single site guide RNA like sR-tMet during its assembly with box C/D core proteins and catalysis of its cognate target RNA HVMΔi36 with a dual guide RNA. The representative lead (II)-induced cleavage patterns of sR-tMet having a length of 73 bases are shown in Figure 3.6 A and B. The sequence and the potential structure of sR-tMet 73 mer can be observed in Figure 3.4 A, except that the 5' end of this molecule is at +21 position from the 5'end of sR-tMet 93mer. As can be seen from the plot Figure 3.6 B, each peak in the plot along X-axis corresponds to a band in the gel. The plot runs from left to right which reflects reading of bands from the top to bottom of the gel; Any increase or decrease in peak height reflects an increase or decrease of band intensity at one particular position when compared between two lanes in the gel. This in turn corresponds to exposure or protection of that position on the RNA from lead (II)-mediated cleavage because of RNA-RNA or RNA-protein interactions. As can be seen from the Figure 3.6B, L7Ae clearly binds at the box C’, but not in box D’. However regions of the RNA which were proximal to the ends were difficult to analyze. These regions housed box C and box D. Hence we switched to modified version of sR-tMet having a 5’ extension of 20 bases. This RNA was already proven to be a bona fide guide by our methylation studies (See Section 3.3.1). A plot of lead (II)-induced cleavage pattern of sR-tMet 93mer is shown in Figure 3.7B. L7Ae binds to both box C’, box D and to certain extent to box C, but in this assay we could not detect L7Ae binding to box D (Panel ii). However, in contrast to what was observed for dual guide, no
Figure 3.6 Architecture of a single guide sRNP sR-tMet 73mer. Left: Representative phosphorimager scans of denaturing 12% acrylamide gels showing footprints after lead (II)-induced cleavage of 5’-end labeled RNAs in the presence of box C/D core protein/proteins as specified above each lanes. The left most lane represent cleavage pattern of a 5’end labeled RNA without any proteins. The right most lane represents a C-ladder which is made from 5’end labeled sR-tMet, transcribed in presence of CTPαS (See Materials and Methods) and cleaved by treatment with I$_2$/EtOH, being used as position markers to map the regions of RNA that shows protection. Right: Lane profile analyses were done using Image quant
software where the height of each peak represents the degree of lead (II)-induced cleavage of the RNA at that site. Linear schematics indicate the specific regions of the RNA corresponding to the regions in the gel scans and peaks in the plots as determined by sequencing reactions on the same template. The components of the assembly reaction on each panel are indicated on the upper right corner. The analogous positions in the gel and the cleavage profiles which show protection are indicated by square and circular dashed lines.

exposure of the corresponding guide regions were observed in this case (Singh et al., 2008). Also, addition of target RNA at this point, does not show any base pairing between the guide and target (Panel iii), proving L7Ae alone with the guide is unable to bind the target. Addition of aNop5p along with L7Ae and the guide, however, extended the protected regions to D' guide region (Panel iv), showing aNop5p was interacting with the single stranded guide RNA in presence of L7Ae. This observation seems to be in agreement with previous studies with *Pyrococcus furiosus* aNop5p where the ALFR motif in the NOP domain of aNop5p was shown to cross-link to the guide region of RNA (Ghalei et al., 2010). Moreover, in the same study it was speculated that extensive electrostatic interactions between aNop5p and the single stranded guide-spacer region of the RNA was responsible for stabilizing the guide sRNP complex. aNop5p alone was unable to bind to the guide region in absence of L7Ae (Data not shown). Hence the presence of L7Ae is required for aNop5p to bind to specific regions of the guide RNA. Again, this is not surprising as previous studies have shown that *Pyrococcus furiosus* aNop5p sandwiches between L7Ae and the guide RNA through the GAEK motif (Xue et al., 2012) for stable sRNP complex formation. No guide-target pairing was observed up on addition of target at this point (Panel v). Also, Fibrillarin does not contact the RNA
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guide</strong></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>L7Ae</strong></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>aNop5p</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>aFibrillarin</strong></td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- + + + +
- + + +
- + + +
- + + +
- + + +
- + + + U G
Figure 3.7 Lane profile analyses of Lead (II)-induced cleavage of assembly reactions of sR-tMet 93 mer. (A) Representative phosphorimager scans of denaturing 12% acrylamide gels showing footprints after lead (II)-induced cleavage of 3’-end labeled RNAs in the presence of box C/D core protein/proteins as specified above each lanes. The components of each assembly reaction are indicated above each lane. The lane marked “U” represents end labeled RNA untreated with lead(II)-induced cleavage. The lane labeled “G” is a sequencing reaction serving as size markers. (B) Lane profile analyses were done using Image quant software same as Figure 3.6.

even if the other core proteins are present (Panel F) or when it is alone (Data not shown). In all the assembly reactions, we were unable to detect any conformational changes in the D’ guide region, which is different from the observations on dual guide sRNAs, where both the D and the D’ guide showed structural changes upon protein binding. It has to be remembered that there are no known targets for the D’ guide for sR-tMet (Joardar et al., 2012). Hence, preliminarily, it seems that there are discernible differences in the catalytic mechanism of a single guide sRNP from the dual guide. These mechanistic differences may be dictated by the guide RNA themselves, which, other than the obvious lack of target complementarity in the guide spacer region, may have different structural cues which determine which RNAs will have dual guide properties and which RNAs will be a single guide.

3.3.4 Mechanism of substrate recruitment by box C/D sRNPs.

Several structural studies of partially and completely assembled guide sRNPs with and without small oligonucleotides as their targets are now available (Lapinaite et al., 2013; Xue et al., 2012; Ye et al., 2009). However, till date how the target RNA is recruited to an assembling guide sRNP is still an open question. Particularly, how a
large and complex substrate RNA like rRNA can get access to the active/catalytic site of Fibrillarin through pairing with a highly restrictive guide RNA is one of the key unanswered questions. It has been shown that structural features of the target RNA itself guides the level to which it will be methylated (Appel and Maxwell, 2007). For example, extending the length of guide-target duplex pairing has been shown to increase methylation levels of the target RNA. These nucleotides beyond the region base-paired to the sRNA guide sequence could be enhancing interaction of the target RNA with the sRNP core proteins and in turn enhancing catalytic turnover (Appel and Maxwell, 2007). Moreover, the guide sRNP complex was able to methylate target nucleotides within highly folded target RNAs which suggested an ability of this RNP complex to facilitate RNA unfolding (Appel and Maxwell, 2007). Also, previous work from our laboratory have shown that dynamic guide-target interaction occurs in a sRNP complex containing only two of the three core proteins- L7Ae and aNop5p (Singh et al., 2008). Hence, it may be speculated that guide: target RNA interactions could also be facilitated by initial interaction of single-stranded regions of each RNA brought about by one or more of the core proteins, followed by unzipping of the remaining target RNA structure for obtaining an optimal methylation competent conformation.

To probe how target RNA is recruited to an assembling guide sRNP, 2 picomoles of 5’ end labeled HVMΔi36 was cleaved by lead (II) - in the presence of 10 picomoles each of different combinations of core proteins, with or without the guide RNA sR-tMet 93mer. The analyzed plots of the lead (II)-induced cleavage patterns are shown in Figure 3.8 B and Figure 3.9.
As can be seen from Figure 3.8 Panel i, the guide and the target do not spontaneously pair sans proteins. Moreover, as Panel ii and Panel v show, L7Ae also is unable to bind target RNA with or without the presence of guide respectively. The lack of any K-Turn or K-loop motif on the target RNA may be the reason why L7Ae is unable to bind the target molecule.

However, the results became very interesting when aNop5p was introduced. When aNop5p was added to assembly reactions containing only the target, it was seen that aNop5p was binding to particular regions of the target RNA (Panel iii). This may have been due to non-specific electrostatic interactions between the positively charged aNop5p and the negatively charged phosphate backbone of the target RNA. To address the question of aNop5p binding to RNA were specific or not, various concentrations of aNop5p were titrated against a fixed concentration of RNA.

Profiles of lead (II)-induced cleavage of assembly reactions showed that aNop5p starts binding to RNA non-specifically when the protein amount is much more than the RNA amount (Figure 3.10). In fact, aNop5p seems to bind all over the RNA surface when present in an amount more than 5 molar fold excess than the RNA. Hence, in all our subsequent experiments, care was taken that aNop5p was not present in five molar excess amount than RNA.

Thus, aNop5p seemed to recognize specific regions of the target RNA. Interestingly, L7Ae and aFibrillarin seemed to have very little effect on aNop5p recognition pattern of the target RNA (Panel v and Panel vii). Also, L7Ae and a Fibrillarin together had no effect on substrate recognition by aNop5p (Panel viii).
Figure 3.8 Lead (II)-induced cleavage RNA footprinting of pre-tRNA^{Met}_{Δ36i}. (A) Representative phosphorimager scans of denaturing 12% acrylamide gels showing footprints after lead (II)-induced cleavage of 5'-end labeled HVMΔi36 RNAs in the presence of box C/D core protein/proteins and guide RNA. The components of the assembly reactions for each lane is as follows: Lane (1) Δ36i (2) Δ36i + sR-tMet (3) Δ36i + L7Ae (4) Δ36i + aNop5p (5) Δ36i + aFib (6) Δ36i + sR-tMet + L7Ae (7) Δ36i + sR-tMet + aNop5p (8) Δ36i + sR-tMet + aFib (9) Δ36i + L7Ae + aNop5p (10) Δ36i + L7Ae + aFib (11) Δ36i + aNop5p + aFib (12) Δ36i + L7Ae + aNop5p + aFib (13) Δ36i + sR-tMet + L7Ae + aNop5p (14) Δ36i + sR-tMet + aNop5p + aFib (15) Δ36i + sR-tMet + L7Ae + aNop5p + aFib (B) HVMΔi36 (Target) RNA interactions with the core proteins in absence of guide RNA sR-tMet. Selected lanes in Figure 3.6 A were analyzed by the Image Quant software and plotted as before.
Figure 3.8 shows the effect of addition of guide RNA on conformations of target RNA when present in with various combinations of core proteins. When unlabeled guide RNA was added to the assembly reactions containing the end labeled target RNA guide along with only aNop5p (Panel ii), no further protection from Lead (II)-induced cleavage was observed which showed that though aNop5p was able to bind to the target, it was unable to initiate the guide-target pairing necessary for methylation of the target RNA. L7Ae was also by itself unable to initiate guide-target pairing (Panel i). However, when both L7Ae and aNop5p were present together with the guide, we see a protection at the positions of 27 bases to 37 bases, which incidentally is the region of the target RNA that hybridizes with the guide RNA (Panel iv). Hence, though aNop5p by itself seems to be able to recognize specific regions of the target, only in the presence of L7Ae is it able to recruit the target for hybridization to an assembling guide RNA. This necessity for the presence of L7Ae is not surprising as multiple studies have shown that L7Ae binding to the guide RNA is absolutely indispensable for guide: target pairing, by both inducing conformational changes in the guide RNA which primes it for target binding and also for providing a RNA-Protein hybrid platform for aNop5p to bind to the guide RNA through its NOP domain (Gagnon et al., 2012; Gagnon et al., 2010; Ghalei et al., 2010; Lapinaite et al., 2013; Singh et al., 2008). Also, L7Ae stabilizes the aNop5p-RNA interactions by sandwiching aNop5p between itself and the RNA-phosphate backbones. Taken together, it seems that it is aNop5p by itself binds to specific regions of the target RNA and then loads the substrate in the L7Ae bound guide sRNP. Whether it is free aNop5p which is recruiting the target or aNop5p bound to a guide sRNP which is doing the same is not known, at least from these experiments.
Addition of aFibrillarin to the assembly reactions did not impact either the binding of aNop5p to the free RNA (Figure 3.8, panel vi) or to L7Ae complexed RNA (Figure 3.8, panel v). There are multiple lines of evidences which shows that aNop5p-aFibrillarin form stable complexes both in vitro (Lapinaite et al., 2013) and in vivo (Lechetier et al., 2009). In fact it is shown with Pyrococcus furiosus aNop5p, that upon aFibrillarin binding, the flexible helix α3 of aNop5p induced large domain rotations in the NOP domain which leads to a conformational shift in the aNop5p-aFib complex (Oruganti et al., 2007). This conformational change is thought to help Fibrillarin bind its co-factor S-Adenosyl Methionine and help position the catalytic center of Fibrillarin close to the RNA binding domain of aNop5p. However, there is no evidence that this dynamic change in the aNop5p NOP domain induces any change in its preference for RNA binding. Hence, it is not surprising that there were no changes observed in aNop5p RNA binding patterns upon Fibrillarin addition. Fibrillarin of course could not bind the target RNA by itself (Figure 3.7, panel iii and Figure 3.8 panel iv) as it lacks any RNA binding domain.

3.3.5 aNop5p recruits target RNA by recognizing single stranded loops and bulges in the RNA

Data from the last section hinted that it is aNop5p which loads the substrate RNA to an assembling guide sRNP. This is in direct agreement to previous studies which has shown that aNop5p gets cross linked to the target RNA (Gagnon et al., 2012). However,
those studies could not determine which regions or structural features of the target RNA aNop5p binds to. Hence, we were next interested to figure out what structural cues in a RNA aNop5p recognizes to recruit the target RNA to the guide sRNP enzyme. Figure 3.11 (Red lines) shows the positions where aNop5p contacts the target RNA. It can be observed on closer inspection that aNop5p preferentially contacts small single stranded loops and bulges in the substrate RNA. aNop5p is seen to bind at the apical loop of the substrate RNA (positions 54 to 56), the variable region (positions 81 to 83) and a small bulge near the position of modification Cm34 (positions 41 to 43). Hence, we investigated whether it is the presence of small bulges and loops in RNA that aNop5p predominantly binds to. We selected three more target
Figure 3.10. Titration of aNop5p with 5’end labeled pre-tRNA^{Trp} Δ67 RNA. 1 picomole of 5’end labeled pre-tRNA^{Trp} Δ67 was incubated with increasing concentrations of aNop5p protein. The molar ratios between RNA and Protein are indicated in the upper right corner of each panel.

RNAs: pre-tRNA^{Met}, mature tRNA^{Met} and the 3’ stem of sR-h45, which is a H/ACA guide RNA responsible for isomerizing U1940 and U1942 to Ψ1940 and Ψ1942 in Haloferax volcanii 23S rRNA (Blaby et al., 2011). Other than pre-tRNA^{Met}, the other two RNAs are not even substrates for sR-tMet 93mer. But since they retain small bulges and loops in their secondary structures, these RNAs should still act like a potential substrate for aNop5p binding. Lead (II)-induced cleavage profile of these RNA substrates incubated with aNop5p reinforced our hypothesis that it is indeed these single stranded bulges and loops that aNop5p recognize to bind targets. As can be seen from Figure 3.11 B, aNop5p bound to the same apical loop and the small protrusion in the variable arm of pre-tRNA^{Met} and mature tRNA^{Met} as it did for HVMΔi36. The fact that aNop5p recognized the same structural features of pre-tRNA^{Met} and the two forms derived from it show that these observations are not artifacts. Moreover, in the third substrate, the 3’ end of sR-h45 RNA, aNop5p conferred protection again to the apical loop of the RNA, which again reconfirms it’s preference for small loops and bulges in the RNA. To confirm our hypothesis, we then designed various derivatives of HVMΔi36 (Figure 3.12). In these substrates we systematically deleted the bulges and loops that we identified as aNop5p binding sites. In Δi36-ΔACA substrate, we deleted the small bulge ACA (positions 73 to 75). Curiously, aNop5p does not bind to this 3 nt. bulge, which may indicate that there may be a requirement for a minimum number of bases in
A

pre-tRNA^{Met}  3' stem loop of sR-h45  tRNA^{Met}

HVMΔi36
Figure 3.11 (A) Sequence and secondary structure of HVMΔi36 showing regions of RNA bound by aNop5p. The regions of RNA that are protected by aNop5p as determined by lead (II)-induced cleavage are shown in red lines. The region of the target RNA that hybridizes to the guide region is marked in blue.

(B) Sequence and predicted secondary structures of pre-tRNA^Met, the 3’ stem loop of sR-h45 and mature tRNA^Met. The regions of RNA protected by aNop5p have been shown in red lines like A.

a loop or bulge for aNop5p to bind. In Δi36-ΔCCC substrate, we deleted another 3 nt. bulge from the protruding hexa-nucleotide loop (positions 43 to 45) where aNop5p has been shown to bind. It has to be noted even when the three nucleotides are deleted, a very small bulge can still exist at this position because of two protruding uridines. Next, we deleted a trinucleotide from variable arm (positions 81 to 83) forming Δi36-ΔAGA.

However, like Δi36-ΔCCC, a very small bulge was retained because of a protruding cytidine and uridine from this position.

Though aNop5p seemed to bind to small bulges and loops in the RNA, interestingly there were three regions of the RNA having similar structural properties that aNop5p seemed to preclude from binding (Figure 3.11 A). One was a small ACA (positions 73 to 75) trinucleotide that we already discussed. The other two such regions were the D-Arm and the T-Arm of HVMΔi36. Both these structures have a loop of 8 and 7 nucleotides respectively. So if aNop5p indeed binds target molecules by recognizing bulges and loops, we should have seen aNop5p contacts at these positions as well.

However, tRNA molecules in solution does not adopt the clover leaf conformations but instead forms a 3D L-shaped structure (Quigley and Rich, 1976). Hence, the loops in the D-Arm and the T-Arm that we envisage in a clover leaf structure may not be actually
Figure 3.12 Sequence and potential secondary structures of substrates derived from HVMΔi36.

Like Figure 3.9, the regions showing protection by aNop5p are indicated by red lines.
available for aNop5p binding due to three dimensional folding of the RNA. To check this hypothesis, we constructed two more aNop5p binding substrates. In Δi36-ΔDArm and Δi36-ΔTArm, we deleted the D-Arm and the T-arm respectively from the original substrate. The rationale for creating these constructs were, deletion of either the D-Arm or the T-Arm, would prevent the tertiary interactions between the two arms and hence will free up the loop in at least one of the arms for aNop5p binding.

Next, instead of deleting bases we inserted sequences complementary to the nucleotides that form the penta-nucleotide bulge (positions 41 to 45). This created a substrate, HVMΔi36-GA_{insertion} where the bulge was replaced by a very long stem. Similarly, we inserted a trinucleotide sequence complementary to the bulge from positions 45 to 47, which again abolished the bulge and replaced it with a stem to form a substrate HVMΔi36-CTA_{insertion}.

Lead (II)-induced cleavage of the above substrates gave mixed results. As can be seen from Figure 3.12, for Δi36-ΔACA, the regions protected by aNop5p were absolutely identical to that of HVMΔi36, which again reinforced the notion that regions of protection conferred by aNop5p are actually real and not mere artifacts. As aNop5p did not bind to the ACA bulge, it was not surprising that deleting this trinucleotide motif will have little effect on aNop5p binding. However, aNop5p was still able to bind to the bulge that remained after 3 of the pentanucleotide in Δi36-ΔCCC substrate were detected. The same was observed for Δi36-ΔAGA, where aNop5p was able to bind to its preferred locations even when the AGA trinucleotide was deleted. This may be because aNop5p was still able to bind to the protrusions that remained after deletions of bases in the original bulges. In Δi36-ΔD-Arm, aNop5p bound to its expected positions, but contrary to
expectations, it was unable to bind to the loop on T-arm which was freed from tertiary interactions after D-Arm was deleted. More surprising results came from the remaining three substrates. In Δi36-T-arm substrate, we were unable to see aNop5p binding even to the observed positions for all other substrates. Rather it bound to a region (positions 24 to 28) that seemed to have no bulges or loops. Moreover, we were unable to witness any aNop5p binding to the substrates Δi36-GA\textsubscript{insertion} and Δi36-CTA\textsubscript{insertion}. This was highly unusual in the sense that other than the bulges that got replaced by the stem, the secondary structures that aNop5p seemed to recognize in the original substrate HVMΔi36 were still there. Hence it was highly surprising that we were unable to detect any aNop5p binding for any positions of these RNAs.

One concern especially in light of the lead (II)-cleavage profile results were that deletion of different bulges in the RNAs may have let to secondary structures that have rendered all the above RNA substrates incapable of modification by the box C/D core proteins. To address this concern, we did \textit{in vitro} methylation modification reactions as discussed before with each of the substrates. The substrates were body labeled with [α-P\textsuperscript{32}]- CTP and were incubated with equimolar amounts of unlabeled guide and the three core proteins. The modified RNAs were digested with Nuclease P1 and then digested samples were resolved by TLC. As can be seen from Figure 3.13 A, Δi36-ΔACA, Δi36-ΔAGA and Δi36-ΔCCC showed modification levels equivalent to the wild type substrates. This proved that deletion of the bulges did not make any of these substrates modification incompetent. The results were also at par with the lead (II)-cleavage data showed before where aNop5p bound to the same positions in all these substrates as the wild type.
Substrates $\Delta i36$-GA$_{\text{insertion}}$ and $\Delta i36$-CTA$_{\text{insertion}}$ showed modification levels almost half compared to the wild type. This is quite enigmatic as were unable to detect any aNop5p binding to these substrates. The low modification levels can be explained by the fact that introduction of long stems by introducing complementary sequences to nucleotides in the bulge have made the RNA structures very rigid and hence they lacked the dynamic flexibility that is required for box C/D mediated catalysis.

Figure 3.13 (A) Modification assays of substrates derived from HVM$\Delta i36$. $[\alpha^{32}\text{-P}]$ CTP labeled tRNA derivatives of HVM$\Delta i36$ were incubated with equimolar ratio of unlabeled guide RNA (sR-tMet93 mer) and proteins for pCm production. Mole pCm/mole of RNA were determined from TLC analyses of Nuclease
P1 digests of the products. The experiments were done in duplicates and the error bars are indicated above each bar. **(B) Single Turnover kinetics of aNop5p substrates.** [α-32-P] CTP labeled tRNA derivatives of HVMΔi36 were incubated with equimolar ratio of unlabeled guide RNA (sR-tMet93 mer) and proteins and aliquots were taken out at the time points indicated in the graphs. The reactions were stopped with “Stop” solution containing 0.1% SDS and 0.2mM EDTA, treated with Phenol: Chloroform and then EtOH precipitated. The modified samples were digested with Nuclease P1 and ran on TLCs and pmoles of pCm/mole of substrates were quantified as before.

The low modification levels of substrates Δi36-GA_{insertion} and Δi36-CTA_{insertion} raised an interesting possibility that for the substrates that show the methylation levels comparable to the wild type at the end point kinetics, may not actually behave like the wild type substrates in the initial burst of reactions. To preclude that possibility, time course assays of methylation modifications were performed under single turnover conditions which showed for substrates, like Δi36-ΔCCC that showed modification levels compared to the wild type, mimicked the reaction kinetics of the wild type substrate in Figure 3.13B. However, Δi36-GA_{insertion} and Δi36-CTA_{insertion} showed slower kinetics compared to the wild type from initial time points itself and were unable to ever catch up with the methylation levels compared to the wild type. Hence, at least for these two substrates deletion of the bulges did affect their methylation competency. This decrease in methylation levels may be due to effect of introduction of long stems in both of these RNA making the RNA more rigid.
3.3.6 Importance of ALFR and GAEK motifs in aNop5p recognition.

From the previous discussions it is quite evident that aNop5p contacts the target RNA and recruits it to an assembling guide sRNP for methylation. It also seems that aNop5p recognizes single stranded bulges and loops in the target RNA structure for

![Image](image_url)

**Figure 3.14 Mutating GAEK and ALFA motifs in aNop5p have differential effects on box C/D sRNP mediated catalysis.** $[\alpha-\text{P}^{32}]$ CTP labeled HVM$\Delta$i36 were incubated with wild type and mutant aNop5p proteins (as indicated below each spot in the Figure) in presence of cold sR-tMet 93 mer guide RNA with remaining core proteins and SAM. The modification reactions were allowed to proceed as described before (See Materials and Methods), digested with Nuclease P1 and resolved by TLC. pCm spot is observed for the wild type aNop5p protein and the AAEK mutant aNop5p protein along with the unmodified pC spot and an unidentified spot (marked by *).

recognition. NOP super family of proteins maintains a high degree of selectivity for sRNP binding despite having relaxed sequence requirements (Liu et al., 2007). This is because NOP domain which determines RNA binding is not a RNA binding module but
rather a RNP binding domain which binds to a composite RNA-Protein interface as compared to RNA alone. However, there are some studies that have pinpointed some key amino acid motifs within the NOP domain that help the protein interact with specific regions of RNA itself compared to a RNA-Protein platform (Gagnon et al., 2012; Ghalei et al., 2010; Hardin et al., 2009; Oruganti et al., 2007; Xue et al., 2012; Ye et al., 2009).

For example, a very well conserved GAEK motif in a short, protruding helix $\alpha 9'$ between helices $\alpha 9$ and $\alpha 10$ in the NOP domain of *Pyrococcus furiosus* aNop5p protein sandwiches between guide and non-guide strand and pries the two strands apart so that the guide RNA strand become more available for substrate binding. Moreover, UV-crosslinking experiments have shown another well conserved ALFR motif in helix $\alpha 9'$ of the aNop5p protein C terminal domain, bound to the single stranded sequence of guide RNA and was thought to play an important role in guide-target pairing. The same study also showed aNop5p cross-linking to target RNA. Hence looking at all the above evidences, the ALFR and the GAEK motifs seemed to be the key amino acid sequences which may help in target RNA recruitment for their ability to bind single stranded RNA sequences. With this rationale, we mutated both the ALFR and the GAEK motifs separately. We substituted the Gly284 of the GAEK motif to Alanine to make the aNop5p AAEK mutant. Similarly, we changed Leu289 and Phe290 to Alanines to obtain aNop5p AAAA mutant. In-vitro functional assays with these mutants gave differential activities. When the only the G284 residue of GAEK motif was changed, box C/D sRNP assembled with this mutant aNop5p could still carry out methylation modification of the substrate comparable to wild type. However, when both Leu289 and Phe290 were...
substituted with Alanine, no modifications were observed, showing the AFLA motif is required for box C/D mediated catalysis.

3.3.7 Box C/D sRNPs exhibit multiple turnovers.

The observation that various helicases like p50/55 and Dpb4 transiently associate with box C/D sRNP core proteins (Kos and Tollervey, 2005; Newman et al., 2000) raises the possibility that these proteins may help unwind the guide-target helix post-modification of the target. This hypothesis was further supported by the observation that αFibrillarin displayed lower affinity for methylated substrates compared to non-methylated ones (Lapinaite et al., 2013) which most probably trigger the release of the methyltransferase once the substrate has been modified. All the above observations lead to the speculation that post-modification, the guide sRNP undergoes a series of conformational changes led by unzipping of the guide-target helix by the helicases and reduced affinity of the core protein(s) for the methylated target which may help release substrate from the sRNP complex and prime the guide complex for the next round of substrate binding. It is not clear however whether the core proteins also disassemble from the guide RNP or just the RNA member of the guide sRNP complex undergoes a conformational change which enables the next round of substrate to bind.
Figure 3.15 Box C/D sRNPs exhibit multiple turnover kinetics. 1 picomole, 2 picomole and 5 picomoles of [$\alpha$$^{32}$-P] CTP $\Delta i36$ were incubated with 1 picomole of [$\alpha$$^{32}$-P] CTP labeled guide RNA (sR-tMet93 mer) and 10 picomoles of box C/D core proteins for pCm production. Aliquots were taken out at the time points indicated and mole pCm/mole of RNA was determined from TLC analyses of Nuclease P1 digests of the products. The experiments were done in duplicates and the error bars are indicated above each plot.

To address these issues, we carried out multiple turnover kinetic reactions with [$\alpha$$^{32}$-P] CTP labeled sR-tMet 93mer and HVM$\Delta i36$. With a fixed concentration of guide RNA (1 picomole), substrate concentrations were varied (1 picomole, 2 picomoles, 5 picomoles) with identical protein concentrations under all guide: target ratios. The concentrations of both guide and target RNAs were estimated from scintillation counting. The assembly reactions were allowed to proceed till various time points as indicated in the Figure 3.15 and the modified samples were digested by Nuclease P1 and resolved by Thin Layer Chromatography. The amount of modified nucleotides was calculated after factoring in the amount of unmodified labeled cytidine contributed by
the labeled guide. In this process, the intensities of all the individual ribonucleoside spots from TLC plates were summed up (say this value is X) and divided by the number of the labeled nucleotides (say this value is Y). For example, in the above case for both sR-tMet and HVMΔi36, [α32-P] CTP was used. So here, X was divided by the sum of 34Cs of Sr-tMet93mer and 35Cs of HVMΔi36. This value was then multiplied by the guide: target ratio. For example, in the case where the guide: target ratio was 1:5, Y was multiplied by 5/6. And this value was ascertained to be picomole of pCm produced per picomole of guide RNA. As can be seen from Figure 3.15, when target RNA was present more than two fold compared to the guide, more than one picomole pCm/picomole of guide RNA were produced, which reflected the fact that one molecule of guide sRNP was able to methylate more than one molecule of the target. This in turn proved multiple catalytic turnover of the box C/D sRNP.

The cleavage profiles in real time will reflect the conformations of the guide RNA at various stages of reaction. The conformations might change even when they have completed the reactions. The controls for obtaining these cleavage profiles in real time were done in presence of AdoHcy, a competitive inhibitor of SAM. This will prevent any trace amount of initial modification occurring due to the presence of contaminating AdoMet in aFibrillarin preparations used in the reactions. The results showed over time there is a gradual unwinding of the guide-target duplex at the D’ guide region of sR-tMet 93mer. Moreover, box C seems to get exposed over time as well. Both of these observations were accompanied with concomitant increase in methylated Cm34 as well, as observed from the autoradiograms of TLCs on the right panel.
Figure 3.16 Conformational changes in the single guide sRNA sR-tMet in real time during methylation reactions. **Left:** Standard methylation reactions were carried out with 3’end sR-tMet for the time points indicated in each panels. Plots of lead (II)-induced cleavage of the guide RNA at each of these time points are presented in these panels. The plot for the control reactions (red line) was
generated by incubating the guide (sR-tMet 93 mer) and target RNA (HVMΔi36) with core proteins in presence of AdoHcy for 15 minutes at 68°C. **Right:** TLC analyses of RNase T2 products of the methylation reactions for the same time points that were done with [α-32P] ATP-labeled HVΔi36 and unlabeled sR-tMet. The ratio of the guide and target used were identical to that used for the lead(II)-mediated cleavage reactions on the left. The numbers in the upper left corner of each TLC plate represents picomole of pCm produced per picomole of guide RNA, which are very similar to the amount of modification observed under multiple turnover conditions in Figure 3.15.

These reactions suggest that after the modification is complete there are significant changes in the guide sRNA most probably by the rearrangements of core proteins, as the guide-target duplex unwind after the reaction is complete.

Taken together, these data suggest the box C/D sRNP is capable of multiple turnovers, most probably by rearrangements of the core proteins accompanied by conformational changes in the guide RNA itself, which primes the sRNP to bind the next round of substrate.

### 3.4 CONCLUSIONS AND PERSPECTIVES

Majority of 2'-O-methylation of ribose sugars in RNAs of eukarya and archaea modifications are carried out by box C/D RNA-protein complexes(Henras et al., 2004; Kiss, 2002; Reichow et al., 2007). The guide RNA in the box C/D sRNP base pairs with the complementary substrates and pinpoints the site of modification by using a kind of molecular ruler (Cavaille et al., 1996; Tycowski et al., 1996). There are several studies done to understand the catalytic mechanism of this sRNP complex. Most of these
studies have been done from the perspective of guide RNA; how guide RNA assembles the core proteins and the architecture of the catalytically active guide sRNPs have been studied in detail. However, very little is known about mechanism of target RNA recognition and its’ recruitment to an assembled or assembling guide sRNP. In the present study, attempts have been made to illustrate the mechanism of target recruitment.

3.4.1 Box C/D core protein aNop5p recruits target to an assembling box sRNP.

Lead (II)-induced cleavage profile of end labeled target RNA with sequential addition of box C/D core proteins and unlabeled guide RNA showed that aNop5p is the only core protein which binds to the target RNA (Figure 3.8 and Figure 3.9). In fact, aNop5p can bind to the target RNA without the presence of any other proteins or guide RNA (Figure 3.8, panel iii). Titration of different concentrations of aNop5p against a fixed concentration of target RNA proved that the protection observed are not non-specific (Figure 3.10). The fact that aNop5p can bind RNA directly was a bit surprising as the C-terminal domain, with which aNop5p binds RNA, folds into a composite NOP domain which is not known to be a genuine “RNA binding” domain. Rather NOP domain is recognized for its ability to bind a RNA-Protein platform (Liu et al., 2007). The wide range of RNA substrates that aNop5p binds to do irrespective of its sequence have been attributed to the NOP domain because NOP domain recognizes structural cues of RNA –Protein interfaces rather than RNA sequences. However, recent UV cross linking studies and studies done on box C/D sRNP crystal structures have
indicated that aNop5p may bind directly to RNA, but all of those observations have studied aNop5p binding in presence of L7Ae (Ghalei et al., 2010; Xue et al., 2012; Ye et al., 2009). Our studies show that aNop5p binds to the target RNA directly without the help of any other core proteins and recruits the target RNA to an assembling guide sRNP.

3.4.2 aNop5p recognizes single stranded bulges and loops in the RNA

Once it was established that it was aNop5p which was recruiting target in an assembling guide box C/D sRNP, we wanted to see what sequences or structural cues the protein recognizes in the RNA. Lead (II)-induced cleavage profiles of pre-tRNA\textsuperscript{Met} HVMΔi36 substrate RNA revealed that it was single stranded bulges and loops in predominantly double stranded RNA molecule that the protein was binding to (Figure 3.11). To ascertain the above observation, we did some aNop5p binding studies on various naturally occurring substrate molecules—namely pre-tRNA\textsuperscript{Met}, mature tRNA\textsuperscript{Met} and the 3’ stem of sR-h45, the box H/ACA RNA associated with aCbf5 protein (Blaby et al., 2011). It has to be kept in mind that other than pre-tRNA\textsuperscript{Met}, the other two RNAs are not even substrates for guide box C/D sRNP mediated modifications. Lead (II)-induced cleavage profiles of all the three substrate RNAs however, showed that aNop5p did indeed bind to all the three RNAs and also retained its preference for binding to small bulges and terminal loops of these RNAs (Figure 3.11). There appeared to be no preference of RNA sequence for aNop5p to bind to. To offer further credibility to the hypothesis of aNop5p binding to single stranded bulges and loops in
RNA, we created various truncated versions of the original substrate pre-tRNA$_{\text{Met}}$ HVM$\Delta$i36, in which the bulges and loops where aNop5p were seen to bind, were either shortened abolished by introduction of complementary sequences opposite to those bulges (Figure 3.12). All of the substrates created were capable of being methylated by aNop5p comparable to wild type substrate levels, except for two RNA substrates $\Delta$i36-GA$_{\text{insertion}}$ and $\Delta$i36-CTA$_{\text{insertion}}$ (Figure 3.12). In both these substrates, there were two bulges that were abolished by introducing complementary sequences which base paired with the nucleotide sequences present in the bulge. This created a long double stranded stem in the RNA. It can be speculated that formation of this long stem made the RNA more rigid than the wild type. Hence the plasticity required in the RNA structure to enable the dynamic guide-target interactions that is needed in box C/D modification is absent in these two RNA structures, leading to lower turnover of these two substrates.

Lead (II)-induced cleavage of these substrates however gave mixed results. For five of the substrates, aNop5p showed preference for binding to the same regions as it did for the original substrate pre-tRNA$_{\text{Met}}$ HVM$\Delta$i36. For example, aNop5p was still able to bind to the bulge that remained after 3 of the pentanucleotide in $\Delta$i36-$\Delta$CCC substrate were deleted. Hence, aNop5p shows strong preference for binding to single stranded regions of any RNA. And this binding seems to be non-sequence specific. However, aNop5p did not show any binding for two substrates $\Delta$i36-GA$_{\text{insertion}}$ and $\Delta$i36-CTA$_{\text{insertion}}$. The correlation of any lack of detection of aNop5p binding to low levels of methylation observed before is quite interesting- an observation for which we don’t have any explanation at the moment.
3.4.3 The conserved ALFA motif but not the GAEK is essential box C/D mediated catalyses

*In vitro* modification reactions showed that the methylation competency of the resulting box C/D sRNP complex was severely diminished when the ALFA motif of *Methanocaldococcus* aNop5p was abolished. However, we do not see any defect in methylation modification competency of box C/D sRNPs assembled with AAEK mutant aNop5p. Previous crystal structural models have identified these two motifs, both of which are present in a protruding short helix called $\alpha 9'$ between helices $\alpha 9$ and $\alpha 10$ in *Pyrococcus furiosus* aNop5p to play a very important role in box C/D sRNP mediated modification (Lin et al., 2011; Xue et al., 2012; Ye et al., 2009). Hence the severe disruption of the methylation abilities of box C/D sRNPs assembled with AFLA mutant aNop5p protein underscored the importance of this motif for proper functionality of box C/D sRNPs. Currently, efforts are ongoing in the laboratory to biochemically assess the importance of these motifs to recruit the target in an assembling sRNP.

3.4.4 Other two box C/D core proteins L7Ae and aFibrillain have differential effects of aNop5p mediated target recruitment.

From our assays it is evident that aNop5p plays a pivotal role in substrate recruitment, we observed very different effects of the other two box C/D RNA core proteins have on aNop5p binding. Neither L7Ae nor aFibrillain is able to bind the target RNA either on their own (Figure 3.8, Panels ii and iv, Figure 3.7) or when they are present along with guide RNA (Figures 3.9, Panels i and iii). This is again not surprising;
given the fact that L7Ae has a very stringent requirement of a Kink Turn or a K-Loop motif for RNA binding, none of the substrate RNAs has that motif. aFibrillarin on the other hand does not have an independent RNA binding domain to which it can bind to. Though it must be pointed out the co-crystal structures of Fibrillarin and aNop5p complexed with box C/D sRNA have identified key positively charged amino acid residues in aFibrillarin that forms some contact points sRNAs (Deng et al., 2004).

Our data indicate that aFibrillarin does not influence aNop5p binding to sRNA. aFibrillarin interacts mainly with the N-terminal domain of aNop5p (Aittaleb et al., 2004), whereas aNop5p bind sRNA mainly through its C-terminal domain. Hence, it is not surprising that addition of aFibrillarin has not influenced the extent of aNop5p binding. Also, overwhelming evidences point to the fact that aNop5p and aFibrillarin associate prior to the their interaction with box C/D sRNA (Lechertier et al., 2009; Tran et al., 2003). Hence it is unlikely that aFibrillarin will majorly influence the extent of aNop5p binding to it.

However, though aNop5p is able to bind to a target by itself, effective guide-target interaction does not take place in the absence of L7Ae core proteins (Figure 3.9 panel iv). It has already been established the L7Ae is absolutely needed to have a catalytically efficient box C/D sRNP (Omer et al., 2002; Singh et al., 2008; Tran et al., 2003). Also, L7Ae upon binding to the box C/D and box C'/D' induces a conformational change in the RNA molecule that opens up the respective guide sequences for efficient hybridization to the target sequences (Singh et al., 2008). Hence binding of L7Ae at the box motifs are essential for the guide to be accessible for target binding. Moreover, though aNop5p can bind certain regions of RNA by itself, the NOP domain is still
essentially a sRNP recognition domain which will need a RNA-Protein platform to bind
(Liu et al., 2007). Hence taken together, it can be seen though aNop5p can recognize
and recruit the target RNA by itself, it still requires L7Ae for efficient formation of a
guide: target duplex.

3.4.5 Guide-Target unpairing is essential for box C/D sRNP mediated multiple
turnover reactions.

Time course studies of methylation modification reactions under substrate
excess conditions revealed that box C/D sRNP is a multiple turnover enzyme complex
(Figure 3.15). As can be seen from the figure, the ratio of target RNA to assembled
sRNP was increased, the picomoles of target RNA methylated were also increased.
This reflects that elevated concentrations of the substrate target RNA are driving the
reaction toward nucleotide methylation. This is in direct agreement with previous studies
which have also shown multiple catalytic turnover of box C/D mediated sRNP
complexes (Appel and Maxwell, 2007; Omer et al., 2002). Time course of lead(II)-
mediated cleavage profile of box C/D mediated modification reactions revealed that
over time, after more and more substrate RNA gets modified there is a gradual
unwinding of guide: target duplex with a concomitant exposure of box C’ and box D’ in
the guide RNA (Figure 3.16). Since the protections observed for box C’ and box D’ are
mostly because of core-proteins binding to the sRNAs (See figure 3.5 and figure 3.6),
the gradual exposure of the consensus box sequences are probably because of some
rearrangements of the proteins that are occurring during modification. Hence it seems
post-modification, the target RNA unwinds from box C/D sRNPs accompanied by some
rearrangements of the core proteins which primes the partially assembled guide sRNP for the next round of modification.

In summary, this study provides a brief mechanistic insight into the catalytic mechanism of box C/D sRNP and the following model may be envisaged. aNop5p, either alone or complexed with Fibrillin, bind to single stranded bulges and loops of target RNA. This aNop5p bound target is then hybridized to an assembling guide sRNP complex containing the guide RNA and L7Ae or guide RNA, L7Ae and aNop5p. If the guide: target sequences are complementary to each other, they should hybridize and the targeted nucleotide will be modified. Post modification, the guide and target strands separate, the core proteins rearrange themselves on the guide RNA and then prime it for next round of modification.
CHAPTER 4

DISCOVERY OF THE sRNA THAT GUIDES 2’-O-METHYLATION OF G1934 RESIDUE OF 23S rRNA IN HALOFERAX VOLCANII

4.1 Introduction.

Eukaryal nucleoli contain a subset of non-coding RNAs called small nucleolar RNAs (snoRNAs) that are responsible for processing of rRNAs and other RNAs (Bachellerie et al., 2002; Decatur and Fournier, 2003; Filipowicz and Pogacic, 2002; Henras et al., 2004; Kiss, 2002; Maxwell and Fournier, 1995; Weinstein and Steitz, 1999). Some of the most abundant post-transcriptional modifications in eukaryal ribosomal and other RNAs like 2’-O-methylations and pseudouridylations are guided by these snoRNAs. The guide RNAs that carry out 2’-O-methylations are called box C/D snoRNAs, whereas the the guide RNAs that carry out the pseudouridylations are called box H/ACA snoRNAs. Box C/D RNAs get their name due to the presence of a conserved motif called box C and box D at their termini and box C’ and box D’ located internally. The specific target nucleotide to be modified is determined by a stretch of 10-21 nucleotides upstream of the D and the D’ boxes in the guide RNA, which is complementary to the target RNA and it is the 5th. nucleotide of the target RNA in the guide target duplex which gets methylated. However, these guide RNAs have to take help of a set of proteins to carry out the above modifications. In eukaryotes, box C/D guide RNAs associate with Fibrillarin, which is the catalytic methyltransferase along with
other proteins that mostly play structural roles like 15.5 K protein, Nop56p and Nop58p. The RNA and the proteins form an asymmetric sRNP where 15.5 K, Nop58p and Fibrillarin bind the box C/D motif and Nop56p and Nop58 bind to the box C'/D' motif (Cahill et al., 2002; Szewczak et al., 2002).

Archaea also contains both box C/D and box H/ACA sno-RNA like RNAs (sRNAs) (Dennis and Omer, 2005; Gaspin et al., 2000; Henras et al., 2004; Kiss, 2002; Muller et al., 2008; Omer et al., 2000; Reichow et al., 2007; Tang et al., 2002). Like the eukaryal snoRNAs, archaeal sRNAs associate with its own set of cognate proteins - L7Ae (homolog of 15.5 K), aNop5p (ortholog of Nop56p and Nop58p) and aFibrillarin (homolog of Fibrillarin), the last one being the catalytic methyltransferase (Filipowicz and Pogacic, 2002; King et al., 2001; Omer et al., 2003; Omer et al., 2002). Though the architecture of an eukaryal snoRNP and an archaeal sRNP are mostly similar, there are significant differences between them as well, both at the levels of RNA as well as proteins.

Unlike in eukarya, boxes C' and D' in archaea are well conserved. However, eukaryal snoRNAs are generally bigger in size than the archaeal sRNAs (Omer et al., 2000). Majority of the archaeal sRNAs are “dual guide”; They are able to target methylations from both the guide sequences present in the RNA, however, only a fifth of eukaryal snoRNAs have that property (Speckmann et al., 2002). Also, there is a stringent requirement in the length of guide sequence in archaeal sRNAs, any alteration in the length of the guide sequences between the two box motifs adversely affects the activity of the sRNP complex (Tran et al., 2005). Another hallmark of archaeal sRNAs are the presence of signature folded RNA structures called K Turn for L7Ae binding.
A prototype K-Turn consists of a three base bulge containing an invariant Uridine at its third position, flanked on its 5' side by a Watson-Crick base paired Canonical stem (C stem) and on its 3' side by the Non-Canonical stem (NC stem) containing sheared G:A base pairs. This bulge provides a $60^\circ$ kink or bend to this RNA and hence the name. All known kink turns contain this asymmetric bulge between his two stems (Schroeder et al., 2010). Archaeal sRNAs also have another distinct RNA fold called K-Loop which like the K-turn contains the non-canonical stem, but unlike the K-turn, the canonical stem is replaced by a short terminal loop (Nolivos et al., 2005). L7Ae nucleates box C/D sRNP formation by binding to both K-Turn and K-Loop followed by the stepwise assembly of aNop5p and aFibrillarin (Omer et al., 2002; Singh et al., 2008; Tran et al., 2003).

Many computational algorithms have been designed to predict the presence of box C/D and box H/ACA sRNAs in the genomes of various organisms (Edvardsson et al., 2003; Lowe and Eddy, 1999; Schattner et al., 2004). These computational methods are mostly based on the characteristic structural features of the guide RNAs. Emergence of new techniques like RNA-seq also makes it possible to identify novel sno(s)RNAs in various organisms (Bernick et al., 2012a; Bernick et al., 2012b). Presence of some of these RNAs have been experimentally verified by classical biochemical techniques (Huttenhofer et al., 2001; Huttenhofer and Vogel, 2006). A combination of such in-silico and biochemical approaches have been able to identify a range of small non-coding RNAs in several archaea (Gaspin et al., 2000; Muller et al., 2008; Omer et al., 2000; Straub et al., 2009; Tang et al., 2002; Thebault et al., 2006). Such wealth of studies has highlighted some very interesting findings in the sRNA world.
of archaea. For example, in *Haloferax volcanii*, the intron of pre-tRNA$^{\text{Trp}}$ bears all the characteristic motifs of a bona fide box C/D sRNA and has been shown to sequentially modify *in trans* 2'-O-methylations of two residues in intron-containing pre-tRNA$^{\text{Trp}}$ that ultimately become C34 and U39 residues of spliced tRNA (Clouet d'Orval et al., 2001; Omer et al., 2000; Singh et al., 2008; Singh et al., 2004). Both the full length pre-tRNATrp and the excised intron can act as a functional guide sRNP. Similarly, 2'-O-methylation of C34 residue in *Haloferax volcanii* elongator tRNA$^{\text{Met}}$ is guided by a novel box C/D sRNA called sR-tMet (Joardar et al., 2012). This sRNA is unique in that it has a 3' extension which is not present in its homologs in other organisms and is also structurally striking as the box C'/D' motif of this guide RNA does not fold into either a classical K-Turn or K-Loop.

In this study we show that 2'-O-methylation of G1934 residue of *Haloferax volcanii* 23S rRNA is carried out by a bioinformatically predicted but experimentally unverified box C/D sRNA. Like, sR-tMet, this novel RNA also is structurally striking in the way that box C'/D' does not fold into either a typical or atypical K-loop. This is because unlike sR-tMet which has at least two tandem sheared G:A base pairs in the atypical K-loop formed at the C'/D' motif, this RNA has just one G:A pairing and no discernible pyrimidine-pyrimidine pairs are also observed. Our studies show that though archaeal L7Ae has quite a relaxed requirement for RNA motif recognition, it is unable to bind to the C'/D' motif of this RNA. This is in direct contradiction to the observations that core protein complexes on both the C/D motif and C'/D' motif needs to be juxtaposed along the length of the same sRNA molecule for a catalytically efficient sRNP (Tran et al., 2003). Hence, the architecture of sR-41sRNP is quite different from the archaeal
sRNPs and is more similar to the asymmetrically distributed eukaryotic sRNPs and hence provides a unique opportunity to study mechanism of catalysis of an asymmetrically arranged sRNPs in archaea.

4.2 Materials and Methods

4.2.1 Growth of *Haloferax volcanii* cells.

*Haloferax volcanii* H26 (a ΔpyrE2 strain) cells were kindly provided by Dr. Thorsten Allers. *Haloferax volcanii* ΔFib, where the endogenous box C/D methyltransferase Fibrillarin has been deleted and *Haloferax volcanii* ΔFib + pMDSFib strain, where the same enzyme is expressed from a plasmid borne copy in an enzyme deleted background, were provided by Parinati Kharel. The strains were grown in HV medium (Gupta, 1984) which contains 125 g NaCl, 45g MgCl$_2$.6H$_2$O, 10 g MgSO$_4$.7H$_2$O, 0.13 g CaCl$_2$.2H$_2$O (added after autoclaving), 3 grams of Yeast Extract, 5 g Tryptone, 10 g of KCl in 1litre medium. The medium was autoclaved for 20 minutes. The cells were grown at 42°C till stationary phase and harvested. *Haloferax volcanii* ΔFib + pMDSFib was grown in the presence of 1µg/ml Novobiocin.

4.2.2 Isolation of total RNA from *Haloferax volcanii* cells

Total RNA was isolated from different *Haloferax volcanii* strains using Tri Reagent (Molecular Research Center) following manufacturer’s protocol.

4.2.3 Primer Extension using limited dNTP (Maden et al., 1995)
5 picomole of $^{32}$P labeled primer specific for 1934 position of 23S rRNA (V23-1945-60R) was hybridized to 10 µg of total RNA from the strains mentioned in the figures in a 15 µl volume. The sequence of the primer was 5’- CGCTACCTTAAGAGGG-3’. The RNA-Primer solution was heated at 90°C for 10’, then snap chilled on ice for 10’. Reverse transcription of the RNAs were initiated in the presence of three different concentrations of dNTPs (0.75 mM, 0.075 mM and 0.0075 mM final concentrations) and 5 units of Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV-RT) and 5 µl of 5x MMLV-RT reaction buffer (250 mM Tris-HCl (pH 8.3), 375 mM KC1, 15 mM MgCl$_2$, 0.1 M DTT) for a total reaction volume of 25 µl. The Primer extension reactions were carried out for 30’ at 42 °C. The reactions were stopped by heating the tubes at 90°C for 3 minutes. 300 ng of RNase A was then added to the reaction mixture and incubated at 37°C for 15’. The products were resolved by 10% denaturing PAGE. To map the 5’ends of the RNA, the same procedure were followed except the primer extension reactions were done in the presence of 0.75 mM dNTPs.

4.2.4 Northern Hybridization

About 20 µg of total RNA from the strains mentioned were resolved by 6% denaturing PAGE (20 cm X 20 cm) and resolved at 250 volts for 2 h. The gel was stained with ethidium bromide to visualize the RNA (a gel picture was taken and saved at this point). The gel was then soaked in transfer buffer (1X TAE) for about 15 min. In the meantime, the nytran membrane was cut to the exact size of the gel, and a Whatman filter paper was cut to serve as a wick. The membrane was also kept soaked in transfer buffer. A tank was filled with transfer buffer and a wick was placed in it such that the two edges are dipped in the liquid. The gel was then carefully placed in this tank
and sealed on all sides with parafilm. The membrane was placed on top of the gel, which was then topped with a stack of paper towels and 500 g weight. This set-up was left O/N (12-16 hr) for transfer. Next morning, the membrane was UV-crosslinked and then kept for prehybridization for about 4 hr at 68°C (in hybridization bottles, kept rotating in hybridization chamber) in prehybridization solution (6X SSPE/1% SDS/5X Denhardt’s reagent/10 µg/ml denatured salmon sperm DNA). Following prehybridization, the membrane was subjected to hybridization O/N with the desired 5’-end labeled probe in hybridization solution (same as prehybridization solution except salmon sperm DNA was not added) at a temperature 5 degrees below the Tm of the probe. The probe used was the same as used for the Primer extension reactions (HV CD23R). The blot was then washed several times in 6X SSPE/0.5% SDS to wash off excess probe, and finally in 2X SSPE to wash off remaining SDS, and then exposed to phosphorimager screen. The blot was stripped with a large volume of stripping solution (10 mM Tris.Cl, pH 7.4, 0.2% SDS ) at 70°C for 2 hours. The blot was re-probed with 5’end labeled primer against 5S rRNA to assure that the loading in all the wells were equal.

**Composition of:**

**50X Denhardt’s reagent:** 1% Ficoll, 1% Polyvinylpyrrolidone (PVP), 1% BSA, filtered and stored at -20°C.

**20X SSPE:** 3M NaCl, 0.2 M NaH₂PO₄, 0.02 M EDTA, pH adjusted to 7.4 with 10 N NaOH. Sterilized by autoclaving.
20 X SSC: 3M NaCl, 300 mM Sodium Citrate; Sterilized by autoclaving. Adjusted to pH 7.0 with HCl.

50X TAE Buffer: 24.2 g of Tris Base, 5.7 ml. of Glacial Acetic Acid per 100 ml. of the buffer and 0.5 M EDTA (pH 8.0.)

4.2.5 Generation of DNA templates for in vitro RNA synthesis.

PCR amplified templates where used for in vitro transcription. Different primer combinations and plasmids carrying the respective genes were used to PCR amplify the template for both guide and target generation (See Table 4.1). These PCR products were cleaned of unincorporated nucleotides and Taq enzyme by passing through Sephadex G25 spin columns.

**Table 4.1 RNA Substrates, DNA templates and Primers used in Chapter 4.**

<table>
<thead>
<tr>
<th>RNA Substrates</th>
<th>DNA Template</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>sR-41 (Guide RNA)</td>
<td>Genomic DNA from <em>Haloflexus volcanii</em> H26 cells</td>
<td>Forward: T7HVCD23 5’TAA TAC GAC TCA CTA TAG GTG GCG ATG ACG AAG G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse: HVCD23R 5’GGT GGC TCG GAT GAA ATC</td>
</tr>
<tr>
<td>sR41-DPT (Small target complementary to D’ guide of sR-41)</td>
<td>Double stranded small target DNA generated by PCR</td>
<td>Forward: T7P 5’TAA TAC GAC TCA CTA TA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse: CD23-DPT 5’GGG TCA TAG TTA CCC CCG CCG TTG ACT ATA GTG AGT CGT ATT A</td>
</tr>
</tbody>
</table>
Table 4.1 RNA Substrates, DNA templates and Primers used in Chapter 4
(continued).

<table>
<thead>
<tr>
<th>RNA Substrates</th>
<th>DNA Template</th>
<th>Primers</th>
<th>Reverse CD23-DT</th>
</tr>
</thead>
<tbody>
<tr>
<td>sR41-DT (Small target complementary to D guide of sR-41)</td>
<td>Double stranded small target DNA generated by PCR</td>
<td>Forward T7P TAA TAC GAC TCA CTA TA</td>
<td>CCT TCC TAG ATT TCA TTG TCC TAT AGT GAG TCG TAT TA</td>
</tr>
</tbody>
</table>

4.2.6 *In vitro* RNA synthesis

For unlabeled transcripts, the protocol is exactly as followed in Section 3.2.3.

Two internal labeled short transcripts were generated. For synthesizing internal labeled *in vitro* transcripts with \([\alpha^{32}P]\) CTP, the reaction mixture used was the same as described above except that 20 µCi of \([\alpha^{32}P]\) CTP and 1µl of 60 mM cold CTP was added instead of 100mM CTP. For generating internal labeled transcripts with \([\alpha^{32}P]\) GTP, 20 µCi of \([\alpha^{32}P]\) GTP and 1µl of 35 mM cold GTP was added instead of 100mM GTP along with 3 µl of 100 mM GMP.

4.2.7 *In Vitro* RNP-directed nucleotide 2'-O-Methylation and Thin Layer Chromatography

RNPs assembly and methylation reaction was performed using fixed concentrations of recombinant core proteins of *M. jannaschii* in presence of specific concentrations of labeled small target RNAs (sR41-DPT and sR41-DT) and unlabeled
guide RNA (sR-41) at 68°C for 30 minutes in 20µl reactions. Reactions were stopped by 200 µl “Stop” buffer containing 0.5 M NH₄OAc, 0.2 mM EDTA and 0.1% SDS, followed by phenol/chloroform extraction and ethanol precipitation. RNA samples were digested with RNase T2/RNase A and the digested products were resolved on cellulose plates using two dimensional TLC. The solvents for TLC were isobutyric acid/0.5 N NH₄OH (5:3, v/v) for the first dimension and Isopropanol/H₂O/HCl (70:15:15, v/v/v) for the second dimension.

4.2.8 Electrophoretic Mobility shift assay (EMSA)

RNP assembly was performed using recombinant *M. jannaschii* core proteins. Approximately 1 pmol of labeled sR-41 transcript was incubated at 68°C with varying concentrations of L7Ae proteins (0.01 picomole to 20 picomole) in 20µl reactions (20 mM Tris.Cl, pH 8.0, 150 mM NaCl, 0.75 mM DTT, 1.5 mM MgCl₂, 0.1 mM EDTA, 10% glycerol) for 15 minutes. Complexes were resolved on native 6% polyacrylamide gel in 0.5XTBE buffer. The gel (20 cm X 20 cm) was run at 4°C at 100 volts. The bands were visualized by using a phosphorimager.

4.2.9 Lead (II) mediated footprinting of RNP complexes

The protocol is essentially the same as Section 3.2.9. Briefly, 1 picomole of 3'end labeled sR-41 was heated at 90°C for 5 minutes and gradually cooled to room temperature to obtain a homogeneous population of uniformly folded RNA. This refolded RNA was incubated in a 36µl reaction and kept at 68°C for 15 min in 20mM Tris-HCl, pH 7.0, 150mM NaCl, 0.75mM DTT, 1.5mM MgCl₂, 0.1mM EDTA and 10% glycerol, with increasing concentrations of L7Ae (0.001 picomole to 20 picomole).
Binding reactions were stopped by chilling on ice. Lead (II) induced cleavage reactions were performed at room temperature for 10 minutes and resolved by 10% urea-PAGE. Lane analyses of footprinting gels were done by Image Quant software. The Signal values obtained from the lane analysis in Image Quant were exported to Microsoft Excel, and linear plots were generated after normalizing the signals for each lane.

4.3 Results

4.3.1 G1934 residue of 23S rRNA in *Haloferax volcanii* is 2′-O-methylated.

Most of the modifications present in tRNA and 16S rRNA in *Haloferax volcanii* were mapped (Grosjean et al., 2008; Gupta, 1984; Gupta et al., 1983). As no modifications was found in 5S rRNA of closely related halophiles like *Halobacter halobium* and *Haloarcula marismortui*, it is thought that *Haloferax volcanii* 5S rRNA lack any modifications as well (Grosjean et al., 2008). Similarly no data were available for 2′-O-methylation modifications in *Haloferax volcanii* 23S rRNA. But there were some analyses performed to identify 2′-O-methylation modifications in 23S rRNA of *Haloarcula marismortui* (Kirpekar et al., 2005). These studies led to the identification of a 2′-O-methylated G1950 residue in the large subunit of *Haloarcula marismortui*. Hence we set out to check whether analogous position G1934 in *Haloferax volcanii* was also 2′-O-methylated as well. To map the presence of methylated guanosine at this position we took the help of primer extension technique using limited dNTP concentrations (Maden et al., 1995). Nucleotide modification 2′-O-methylation can be sterically challenging to the progress of Reverse Transcriptase causing it to pause or stop. The pause/stop effect is exacerbated at positions carrying a 2′-O-ribose
methylation when using diminished dNTP substrate concentrations (Maden et al., 1995). This phenomena was utilized to detect 2'-O-methylation in G1934 in *Haloferax volcanii* 23S rRNA. The primer extension results in concentration-dependent stops preceding the modified position and sometimes additionally at the position (Maden et al., 1995). However, RNA can fold into complex secondary structures to stop the progress of Reverse Transcriptase across the RNA template and thus result in concentration independent stops of the primer. To verify if the pause/stop of primer extension observed was actually because of the presence of 2'-O- methylation and not because of difficult contours of RNA, we performed this study on RNA isolated from wild type *Haloferax volcanii* strain H26 as well as strains where aFibrillarin protein, the catalytic methyltransferase component of a box C/D sRNP was deleted (H26ΔaFib strain). It has to be mentioned here that it was predicted Gm1934 modification was mediated by box C/D sRNP (discussed later) and deletion of aFib would abolish modification at the said position. Hence the concentration dependent pause/stop of primer at this position will be absent in this strain. As can be seen from Figure 4.1, for the wild type Haloferax strain, a strong pause/stop of the primer is noted at position 1934 in the lanes containing lowest dNTP concentration whereas this pause/Stop disappears at the analogous position in the H26ΔaFib. Hence, this analysis showed that Gm1934 of *Haloferax volcanii* 23S rRNA is 2'-O-methylated.

4.3.2 Identification of the sRNA predicted to modify Gm1934 of 23S rRNA of *Haloferax volcanii*
Using comparative genomic analysis it was hypothesized that 2’-O-methylation in ribose at position G-1950 in *Halobacterium halobium* is mediated by a box C/D sRNA and at least one suitable candidate called sR-41 was found which was thought to catalyze the modification at the said position (Grosjean et al., 2008). This RNA had all the hallmarks of box C/D sRNAs (Figure 4.2 A) and was found in 25 archaeal genomes (Grosjean et al., 2008).

![Figure 4.1 Autoradiograms showing primer extension analysis to detect 2’-O-methyl modification at position G1934 of Haloferax volcanii 23S rRNA.](image)

The primer extension analysis was done on total RNA isolated from *Haloferax volcanii* H26 strain and H26ΔaFibrillarin strain. The primer (V23-1945-60R) binds at position 1945 of 23S rRNA. The sequence of the 23S rRNA is indicated on the left. The position of the primer pause/Stop is indicated. The
decreasing dNTP concentration gradient (0.75 mM, 0.075 mM and 0.0075 mM final concentrations) is indicated at the top.

The D’ guide of this RNA matched perfectly with positions in 23S rRNA and the terminal box C/D motifs also folded into a typical K-turn. However, no complementary target sequence was found for the D guide region and also the internal C’ and D’ box was not predicted to fold into either a typical or atypical K-turn or K-loop. To validate the presence of this RNA, total RNA was isolated from Haloferax volcanii H26 cells, Haloferax volcanii ΔFib cells and a third strain from where the genomic copy of Fibrillarin was deleted and the protein was instead expressed from a plasmid. The isolated total RNA was hybridized to a primer complementary to the region indicated in Figure 4.2 A (red line). The appearance of a band in the Northern Hybridization data thus validated the presence of the predicted RNA in Haloferax cells. The amount of the box C/D sRNA does not seem to decrease in absence of Fibrillarin. Primer extensions were then done on total RNA isolated from Haloferax volcanii H26 cells to map the 5’ ends of this guide RNA using the same primer used in our Northern Hybridizations. This result determined the 5’ ends of the sR-tMet RNAs to be at the predicted G shown at position 1 in the Figure 4.2A. Collectively, these data were able to detect the presence of the predicted box C/D sRNA sR-41 in Haloferax volcanii.

**4.3.3 Only the D’guide of sR-41 is functional**

To determine whether sR-41 was able to methylate the G1934 of 23s rRNA in Haloferax volcanii, *in vitro* modification assays were performed with unlabelled *in vitro*
transcribed sR-41 and labeled small targets having sequences complementary to the D
(sR41-DT) guide and D’ guide (sR41-DPT) regions in presence of all the box C/D core
proteins and SAM. TLC separation of RNAses T2 digests of the modified RNA showed
that the labeled target RNA which was complementary to the D’ guide region of the sR-
41 was 2’-O- methylated (Figure 4.3).
Figure 4.2 sR-41 is present in *Haloferax volcanii*. (A) Sequence and predicted secondary structure of box C/D sRNA sR-41. The 5' and the 3' ends of the sRNA are indicated in the figure. Boxes C, C',D and D' are highlighted in pink. The terminal box C and box D folds into a canonical K-turn structure, however, the internal box C' and box D' neither form a K-turn nor a K-loop as in observed in most of archaeal box C/D sRNAs. (B) Northern Hybridization of sR-41 in *Haloferax volcanii*. RNA gel blot of total RNA separated by 6% denaturing PAGE is hybridized to 5' ³²P-labeled oligonucleotide complementary to the 22-base sequence of the sRNA (HVCD23R). The three lanes represent total RNA isolated from *Haloferax volcanii* Wild type strains, *Haloferax volcanii* ∆Fib cells and an *in-trans* expressed Fibrillarin strain (∆Fib + pMDSFib). The lower panel represents 5S rRNA which serves as a loading control. (C) The 5'end of Haloferax box C/D sRNA agrees with the predicted sequence. 5' ³²P-labeled primer used for the Northern hybridization reactions were also used in reverse transcriptase extension reactions with total RNA from *H. volcanii*. The products were separated by 12% denaturing PAGE. Sequencing reactions of sR-41 template DNA obtained by PCR amplification of gene was used as size markers. The 5'end of the sRNA matched the bioinformatically predicted 5'end at the G as seen in A.

(Tran et al., 2005). However, we still checked whether this region can function as guide as well. We used an antisense RNA whose sequence is complementary to the D guide region in our *in-vitro* modification reactions, but no modification was observed (data not shown). The D guide region of sR-41 (region between box C' and box D) is probably not functional as it is just 10 bases long, which is less than the 12 base optimum size for an archaeal guide/spacer sequence (Tran et al., 2005). Hence like sR-tMet (Joardar et al., 2012), sR-41 is a single guide RNA as well, whose D' guide is the only functional one.

4.3.4 *Haloferax* sR-41 box C'/D' does not fold into either a K-Turn or K-Loop.

The C'/D' box of *Haloferax volcanii* is not predicted to fold into either a K-turn or a K-loop (Figure 4.2A). This is very different from box C'/D' motif of the other two known box C/D sRNA in *Haloferax volcanii*; box C'/D' of pre-tRNA^Trp^ RNA folds into a typical
K-loop motif and an atypical K-loop is formed by the same motifs in sR-tMet. Both K-turn and K-loop have been shown to be recognized for substrate binding by L7Ae protein (Gagnon et al., 2010; Joardar et al., 2012; Rashid et al., 2003; Singh et al., 2004; Tran et al., 2003). Moreover, L7Ae is also shown to bind to atypical K-loop structure as evidenced by its binding to sR-tMet (Joardar et al., 2012).

Figure 4.3 D’ guide region of sR-41 can modify target RNA. [α-32P] GTP labeled RNA antisense to the D spacer region was incubated either alone or in presence of unlabeled sR-41 with all the 3 box C/D core proteins and SAM followed by RNase T2 digestion and TLC analyses. The mono and di-nucleotide products are indicated. The appearance of GmGp in RNase T2 digests indicates the methylation of target G in our small antisense RNA targets.
Hence sR-41 was incubated with increasing concentration of L7Ae to determine if box C'/D' is not able to form motifs having any semblance with K-turn or K-loop, whether L7Ae will still be able to bind it. Gel shift analyses suggested that L7Ae is able to form only one ribonucleoprotein complex with sR-41 (Figure 4.4). Even, in the presence of 20 fold molar excess of L7Ae, only one sRNP complex was formed. This suggests that one molecule of sR-41 can bind only one molecule of L7Ae, probably at to the canonical K-turn formed by its box C/D motif.

To further probe the region where L7Ae binds, Lead (II)-induced footprinting was performed on 5'end labeled sR-41 in the presence of increasing concentrations of L7Ae (Figure 4.5). These analyses reveal that L7Ae only binds to the K-turn at box C/D but not at box C'/D'. This is not surprising as box C'/D' motif of H. volcanii sR-41 does not show typical features of either a K-turn or a K-loop. Especially the two tandem, sheared G•A base-pairs and a pyrimidine-pyrimidine pair in its non-canonical stem that has been shown to be so crucial for L7Ae binding seems to be absent in the box C'/D' sequence (Gagnon et al., 2010; Joardar et al., 2012). It is however worth noting that it is the D' guide of sR-41 which is functional and not the D guide region. Hence it seems at least in sR-41 both box C/D and box C'/D' need not occupied by core proteins for their activity. This is in direct contrast to what has been observed for most archaeal box C/D RNA molecules in which the box C/D and box C'/D' complexes must be efficiently “juxtaposed” along the length of sRNA molecule for its proper functionality (Rashid et al., 2003; Tran et al., 2003).
4.4 Discussion

4.4.1 sR-41, the third guide box C/D RNA in *Haloferax volcanii*

Compared to the general archaeal populations, haloarchaea contains significantly fewer number of box C/D guide RNAs (Omer et al., 2000). Previously, only

---

**Figure 4.4 Electrophoretic Gel Mobility Shift Assays of sR-41 with increasing concentrations of L7Ae.** 1 picomole of 5’end labeled sR-tMet 73mer was incubated with increasing concentrations of L7Ae (.001 picomole to 20 picomoles) at 68°C for 15 minutes and then loaded onto a 4% native gel made with 0.5X TBE. Only one sRNP complex seemed to form even when the protein was 20 fold excess of RNA. Lane 1 is devoid of any proteins, whereas lanes 2 to 8 represent an increasing L7Ae concentration gradient of 0.001 picomole to 20 picomoles; Lane (2) 0.001 picomole (2) 0.05 picomole (4) 0.1 picomole (5) 0.5 picomole (6) 1 picomole (7) 10 picomoles (8) 20 picomoles.
two box C/D sRNAs were identified—the intron of pre-tRNA\textsuperscript{Trp}, a “dual guide” molecule which modified Cm34 and Um39 of pre-tRNA\textsuperscript{Trp} and sR-tMet which modified Cm34 of pre-tRNA\textsuperscript{Met}. Using comparative structural analyses, bioinformatically a third box C/D sRNA called sR-41 was predicted to be present in the repertoire of guide RNAs in haloarchaea (Grosjean et al., 2008). Through Northern hybridization technique we were able to detect the presence of this guide RNA in \textit{Haloferax volcanii} and primer extension studies mapped the 5’end of this molecule which also matched with \textit{in silico} prediction. Holomogs of this sRNA was found to be present in 25 archaeal species (Grosjean et al., 2008). This RNA had both the consensus terminal box C/D as well as the internal box C’/D’ sequences. Though box C/D sequence folds into a consensus K-Turn motif, box C’/D’ does not seem to form either a K-turn or a K-loop. (Nolivos et al., 2005; Rashid et al., 2003; Tran et al., 2003).

4.4.2 Only the D’ guide of sR-41 is functional.

\textit{In silico} analyses identified target for only the D’-guide region and no targets for D guide region (Grosjean et al., 2008). However, because of the presence of consensus box C/D and C’/D’ sequences as well as the presence of a bona fide K-turn at the terminal box C/D motif there was a hypothesis that the D guide region might be functional as well. However, modification assays using \textit{in vitro} transcribed unlabeled sR-41, box C/D core proteins and labeled small targets complementary to both the D guide and D’ guide regions of sR-41 revealed that though the D’guide region was able to methylate the correct nucleotide at the small target, the D guide region was unable to do so. Previous studies have revealed the stringent requirement of at least 12 bases in the guide spacer region between box C/D and box C’/D’ in archaeal guide RNAs for
optimum functionality (Tran et al., 2005). The requirement of this conserved spacing is highlighted if we consider the functional guides in the two haloarchaeal box C/D sRNAs known to date. In the case of the dual guide box C/D sRNA in the intron of pre-tRNA$^{Trp}$,

![sR-41 box C/D RNA interaction with L7Ae](image)

**Figure 4.5** L7Ae has affinity only for the sole K-Turn present in sR-41 box C/D sRNA. Left: Lane profile analyses of Lead (II)-induced cleavage of assembly reactions of sR-41 with L7Ae. Lane profile analyses were done using Image quant software. Linear schematics indicate the specific regions of the RNA corresponding to the regions in the gel scans and peaks in the plots as determined by Iodine ladder or sequencing reactions on the same
the D guide is 16 bases long and the D’ guide is 14 bases long and both the guides are functional (Clouet d'Orval et al., 2001; Singh et al., 2008). In the case of sR-tMet, the functional D guide is 14 bases long and the non-functional D’ spacer region is 10 bases long (Joardar et al., 2012). In the case of sR-41, the D’ spacer region is 12 bases long which is functional and the D spacer region which is non-functional is 10 bases long. Hence, the stringent requirement of 12 bases for a guide spacer region is further emphasized by sR-41.

4.4.3 Asymmetrically assembled sR-41 is still functional

Electrophoretic Gel Mobility Shift Assays as well as Lead (II)-induced foot printing reactions revealed that L7Ae binds only to the terminal box C/D motifs and not the internal box C’/D’ motifs (Figure 4.4 and Figure 4.5). This is in direct contrast to a host of studies that have underscored the importance of a symmetric assembly of the core proteins on the RNA (Omer et al., 2002; Rashid et al., 2003; Singh et al., 2008; Tran et al., 2003). When protein binding to either to the terminal box C/D or the internal box C’/D’ was mutated, the methylation activity from both the D guide as well D’ spacer region was sufficiently disrupted. The requirement that efficient methylation of both complexes requires their “juxtaposition” on the same RNA points to multiple protein-protein interaction and a cross talk between the two complexes. This cross talk was thought to be mediated by interactions between the coiled coil domains of aNop5p assembled on the two motifs (Tran et al., 2003). However, assembly reactions with halfmer box C/D complexes were found to be methylation competent at high salt.
concentrations, although the methylation occurs quite non-specific manner (Hardin and Batey, 2006). Also, cross linking studies have pointed out that eukaryal homolog of 15.5 kD protein L7Ae binds asymmetrically only to the terminal box C/D motif (Cahill et al., 2002; Szewczak et al., 2002). It was hypothesized that since eukaryal box C’/D’ motif cannot fold into a typical K-turn, 15.5 kD is unable recognize the alternatively folded structure (Charron et al., 2004). Box C’/D’ in sR-41 is also not thought to fold into either a typical K-turn or K-loop structure because of the absence of the conserved G:A base pairing that is hallmark of these structures. A recent study have also revealed that there is a spatio-functional coupling between the terminal box C/D and the internal box C’/D’ motifs (Qu et al., 2010). Occupation of the terminal box C/D by the core proteins is absolutely needed to direct methylation from the D’ guide. Hence, the architecture of Halofex volcanii sR-41 box C/D sRNP seems to be closer in conformation to eukaryal snoRNPs than an archaeal one. Till date, no information regarding the catalytic mechanism of an symmetrically arranged eukaryal box C/D snoRNPs are available, because of unavailability of any assembly systems or crystal structures. Hence, sR-41 guide sRNP provides an unique opportunity to study mechanism of modification in an asymmetrically arranged box C/D sRNP molecule.


127. Liang, X. H., Xu, Y. X., and Michaeli, S. (2002). The spliced leader-associated RNA is a trypanosome-specific sn(o) RNA that has the potential to guide pseudouridine formation on the SL RNA. Rna 8, 237-246.


turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box

174. Sahoo, T., del Gaudio, D., German, J. R., Shinawi, M., Peters, S. U.,
Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small

175. Schattner, P., Decatur, W. A., Davis, C. A., Ares, M., Jr., Fournier, M. J.,
snoRNAs: analysis of the Saccharomyces cerevisiae genome. Nucleic Acids Res
32, 4281-4296.

interactions in a riboswitch stabilize the structure of a kink turn. Structure 19,
1233-1240.

structural database for k-turn motifs in RNA. Rna 16, 1463-1468.


5, e1000507.

180. Shigi, N., Suzuki, T., Terada, T., Shirouzu, M., Yokoyama, S., and
Watanabe, K. (2006). Temperature-dependent biosynthesis of 2-
thioribothymidine of Thermus thermophilus tRNA. J Biol Chem 281, 2104-2113.


methyltransferase provides insights into its tRNA specificity. Nucleic Acids Res 36, 4929-4940.


VITA

Graduate School
Southern Illinois University

Kunal Chatterjee
kunalmicrobio@gmail.com

Sammilani Mahavidyalaya
Calcutta University, Calcutta
Bachelor of Science (Microbiology Honors) August 2004

Department of Microbiology,
Calcutta University, Calcutta
Master of Science in Microbiology, August 2006

Awards:
Recipient of rank certificate for securing 3rd position in the Masters degree in Microbiology, Calcutta University.

Dissertation Title:
A tale of two methylation modifications in archaeal RNAs.

Major Professor: Dr. Ramesh Gupta
Publications:

