

Structural insight into rRNA methyltransferases and their function in nucleotide modification and antibiotic resistance

A Thesis

*Submitted in Partial Fulfillment of the
Requirements for the Degree of*

DOCTOR OF PHILOSOPHY

by

Sayan Saha
(166106111)

Under supervision of

Prof. Shankar Prasad Kanaujia



**Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India**

April 2024



Dedicated to my beloved parents







INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

**DEPARTMENT OF BIOSCIENCES AND
BIOENGINEERING**

DECLARATION

I do hereby declare that the research findings of this thesis entitled “**Structural insight into rRNA methyltransferases and their function in nucleotide modification and antibiotic resistance**” is the result of research work carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India, under the supervision of **Prof. Shankar Prasad Kanaujia**.

I also declare that the contents of this thesis have not been the basis for award of any other degree, diploma, fellowship or any other similar title of any University or Institution.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made, wherever the research findings of other researchers have been cited in this thesis.

Sayan Saha

Date: April 04, 2024

Place: IIT Guwahati





INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
DEPARTMENT OF BIOSCIENCES AND
BIOENGINEERING

CERTIFICATE

This is to certify that the work described in the thesis entitled “**Structural insight into rRNA methyltransferases and their function in nucleotide modification and antibiotic resistance**” is an authentic record of the results obtained from the research work carried out by **Sayan Saha** in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India, under my supervision and this work in part or as a whole has not been submitted elsewhere for the award of any other degree.

Shankar A Kanaujia

Date: April 04, 2024

Place: IIT Guwahati

Prof. Shankar Prasad Kanaujia

(Thesis Supervisor)

Acknowledgements

The work presented in this thesis was the outcome of the collective effort of many people. Throughout my tenure, I had the privilege of working with several individuals whose guidance and support made this achievement possible. I am grateful to each and every one of them and would like to express my sincere gratitude for their valuable contributions.

I am deeply grateful to my Ph.D. advisor, Prof. Shankar Prasad Kanaujia, for providing me with the opportunity to embark on a journey towards achieving a remarkable feat in the field of structural biology. Without his persistent guidance, unwavering support, and invaluable encouragement, the successful completion of my thesis would not have been possible. Unremitting passion for scientific research and his active involvement in acquiring and providing all the necessary resources have been instrumental in enabling me to accomplish this work. His words and actions have instilled in me the principle of "aiming high and staying grounded," which I will undoubtedly carry forward in my career. I extend my heartfelt appreciation to Prof. Kanaujia for his invaluable mentorship, which has been an inspiration to me and has helped me to grow both academically and professionally. Thank you, Sir.

I would like to express my sincere gratitude to the esteemed members of my doctoral committee, namely Dr. Priyadarshi Satpati, Dr. Shirisha Nagotu and Prof. Lal Mohan Kundu, for their invaluable contributions to my research. Their insightful suggestions and healthy discussions have been instrumental in shaping my research and advancing my understanding of the subject matter.

Furthermore, I wish to acknowledge the unwavering support of the past and present heads of the Department of Biosciences and Bioengineering, specifically Prof. Kannan Pakshirajan, Prof. Latha Rangan, and Prof. Rakhi Chaturvedi. Their tireless efforts in providing the necessary departmental facilities have been instrumental in facilitating my research endeavors.

Acknowledgements

I am pleased to express my sincere appreciation to the Department of Biosciences and Bioengineering at IIT Guwahati for providing me with invaluable infrastructural support and fostering an enabling academic environment. I am also grateful to the Central Instruments Facility (CIF), IIT Guwahati, for granting me access to sophisticated equipment, including the X-ray diffractometer (XRD), isothermal titration calorimetry (ITC) and matrix-assisted laser desorption ionization–time of flight (MALDI-TOF), which were essential to my research work.

I am immensely grateful to the Ministry of Education, Government of India, for providing financial assistance during my Ph.D. Program.

I express my heartfelt gratitude to the past and present members of the Structural and Computational Biology Laboratory (SCBL) for their support and contribution towards creating a conducive work environment. I am particularly grateful to Dr. Prerana Gogoi, Dr. Monika Chandravanshi, Dr. Suraj Kumar Mandal, and Dr. Angshu Dutta my senior fellows, for their invaluable intellectual contributions to my work. Their consistent support and encouragement have been instrumental in the successful completion of my work, and I remain deeply indebted to them for sharing their extensive experience. Their contribution has been pivotal in advancing my research, and I feel honored to have had the privilege of working with them. Additionally, I would like to express my sincere thanks to my fellow juniors, Pratik, Kalyan, Ritu, Arpana, Navneeth, Shimpy, Aditi, and Ankit for their invaluable cooperation and timely assistance, which have been essential to the progress of my work. Their contributions have been invaluable to the success of my work, and I am truly grateful for their support. Lastly, I would like to acknowledge and extend my gratitude to Dr. Ankita Gupta, Prerana, Smit, Ankita, and Arpita for all the timely support.

I am grateful for the support and encouragement provided by my esteemed friends and peers, who have stood by me through my professional and personal voyage. I would like to extend my sincerest appreciation to Arman Mohanty, Sujisha S, Prabir Das, Ananya Chouhan, Ananta Keshari Behera, Dr. Saddam Hussain, Dr. Adhiraj Nath, Dr. Suvankar Ghosh, Pratap Chandra, Vineet Anand, Dr. Ashutosh Kumar, Dr. Subbi rami Reddy Tadi,

Acknowledgements

Dr. Anurag Priyadarshi, Dr. Shah Ekramul Alom, Dilip, Dr. Vinay Bachu and Arupam Patra, for their unwavering support during challenging times.

I owe an immeasurable debt of gratitude to my parents, Mr. Dulal Chandra Saha and Mrs. Rina Saha, and my sisters, Debjani Das, my brother-in-law Mr. Atanu Das, my niece Ms. Anaika Das and Grandfather Late- Shanti Ranjan Saha for their unconditional love, unwavering encouragement, and unyielding perseverance. Their faith in me and numerous sacrifices have made this journey toward my dissertation possible, and their immeasurable worth has only increased with time. I am also immensely grateful to my uncle, Mr. Anup Saha, and my cousin sisters and brother, Puja Saha, Priyanka Saha, Anushikha Saha, and Rohit Saha, for their constant support and motivation and for taking care of our family while I pursued my Ph.D. Their unwavering presence and commitment have been instrumental in boosting my morale throughout this journey. Their love, support, and guidance have been an immense source of strength and inspiration, and I am forever indebted to them.

I would like to take this opportunity to express my sincere gratitude to everyone who contributed to this project. If I have inadvertently omitted anyone from my thanks, I apologize and assure you that your efforts are equally valued.

In conclusion, I express my sincere gratitude to the Almighty God for his blessing.

Table of Contents

Section	Content	Page No.
	Abstract	XIV
	Abbreviations and notations	XX
	List of tables	XXV
	List of figures	XXVI
	List of PDBs	XXX
	Chapter 1 - Introduction	1
1.	Introduction	2
1.1.	Ribosomal subunit composition	2
1.2.	Functional center of the ribosome	3
1.2.1.	Decoding center (DC)	4
1.2.2.	Peptidyl transferase center (PTC)	4
1.3.	Epigenetic modifications of rRNA	6
1.3.1.	Pseudouridylation and 2'-O-ribose methylation	8
1.3.2.	Base methylation	8
1.4.	Enzymes responsible for methylation of rRNA	13
1.5.	Methylation of rRNA in bacteria and eukaryotes	16
1.5.1.	Universally conserved rRNA methylation sites in the SSU	16
1.5.2.	Universally conserved rRNA methylation sites in the LSU	17
1.6.	Biological significance of rRNA methylation	19
1.6.1.	Ribosome biogenesis	19
1.6.1.1.	Ribosome biogenesis in prokaryotes	19
1.6.1.2.	Ribosome biogenesis in eukaryotes	20
1.6.2.	Antibiotic resistance	21
1.6.2.1.	Resistance associated with 16S rRNA	22
1.6.2.2.	Resistance associated with 23S rRNA	23
1.6.3.	Mutations in MTases can cause genetic disorders	27
1.6.4.	Cancer	27
1.7.	Evolutionary aspect of the rRNA methylation	27

Table of Contents

1.8.	Overview of rRNA modification in archaea	28
1.8.1.	Reported RNA modifications in <i>P. horikoshii</i> OT3	29
1.8.2.	Repertoire of putative RNA MTases in <i>P. horikoshii</i> OT3	31
1.9.	Objectives	33
	Chapter 2 – Materials and Methods	34
	General outline of Chapter 2	35
2.1.	Materials	36
2.1.1.	Reagents	36
2.1.2.	Substrate and cofactors	36
2.2.	Methods	36
2.2.1.	Sequence- and structure-based analyses	36
2.2.1.1.	Sequence-based analysis	37
2.2.1.1.1.	Identification of homologous proteins and subsequent sequence alignment	37
2.2.1.1.2.	Construction of phylogenetic tree	38
2.2.1.1.3.	Interaction studies	38
2.2.1.2.	Structure-based analysis	38
2.2.1.2.1.	Structure prediction	39
2.2.1.2.2.	Structural comparison with homologous proteins	39
2.2.1.2.3.	Structure-based study	39
2.2.1.2.4.	Surface analysis	40
2.2.2.	Molecular docking	40
2.2.3.	Recombinant construction for protein overexpression	41
2.2.3.1.	Competent cell preparation for gene cloning and protein overexpression	42
2.2.3.2.	Gene amplification	43
2.2.3.3.	Double digestion of the gene of interest and vector	43
2.2.3.4.	Ligation of the gene of interest and clone confirmation	44
2.2.3.5.	Site-directed mutagenesis	44
2.2.4.	Protein overexpression, solubilization and purification	45
2.2.4.1.	Overexpression of the selected rRNA MTases	45

Table of Contents

2.2.4.2.	Solubility of protein	46
2.2.4.3.	Affinity chromatography for protein purification	46
2.2.5.	Protein characterization	47
2.2.5.1.	Western blot analysis	47
2.2.5.2.	RNA analysis on native-PAGE	48
2.2.5.3.	Mass spectrometric (MS) analysis	48
2.2.5.4.	Energy dispersive X-ray (EDX) analysis	49
2.2.5.5.	UV-visible absorbance spectroscopy	49
2.2.5.6.	Circular dichroism (CD) spectroscopy	49
2.2.5.7.	Dynamic light scattering (DLS) analysis	50
2.2.5.8.	Fluorescence spectroscopy (FS) analysis	50
2.2.5.9.	Size exclusion chromatography (SEC) analysis	51
2.2.5.10.	Reversed-phase high-performance liquid chromatography (RP-HPLC) analysis	51
2.2.5.11.	Isothermal titration calorimetry (ITC) analysis	52
2.2.5.12.	Enzyme assay	53
2.2.6.	Crystallization of the selected rRNA MTases	54
2.2.7.	Data collection and processing	55
2.2.7.1	Data collection	55
2.2.7.2.	Data processing	56
2.2.7.3.	Structure determination	56
2.2.7.4.	Model building and structure refinement	57
2.2.7.5.	Cross-validation (R_{free} calculation)	57
2.2.8.	Structure validation	58
2.2.8.1.	PROCHECK	58
2.2.8.2.	MolProbity	58
2.2.9.	Domain dynamics analysis of rRNA MTases	59
2.2.10.	Structure and protein-ligand interaction visualization	59
	Chapter 3 – Study of archaeal $m^5\text{C}$ RNA MTases	60
	Abstract	61
	Graphical abstract	62

Table of Contents

3.1.	Introduction	62
3.2.	Methodology	65
3.2.1.	Data collection and sequence analysis	65
3.2.2.	Structure analysis	66
3.2.3.	Molecular docking	67
3.2.4.	hdANM (hinge-domain anisotropic network model) analysis	67
3.3.	Results	68
3.3.1.	Repertoire of putative RNA m ⁵ C in <i>P. horikoshii</i> OT3	68
3.3.2.	The archaeal RNA m ⁵ C MTases belong to the subfamilies IV to VI	70
3.3.3.	RsmB-like archaeal rRNA m ⁵ C MTases	74
3.3.3.1.	The interaction profile of the protein PH1078 involves similar interaction as <i>EcRsmB</i> and PH0851	74
3.3.3.2.	The active-site pocket of the protein PH1078 is archetypical to <i>EcRsmB</i> and PH0851 to accommodate the cognate substrate	76
3.3.3.3.	The NTE of the protein PH1078 (subfamily IV) and RsmBs from subfamily I is indistinguishable	80
3.3.4.	NSUN1-like archaeal rRNA m ⁵ C MTase	83
3.3.4.1.	The protein PH1374 is plausibly an rRNA m ⁵ C MTase	83
3.3.5.	Archaeal tRNA m ⁵ C MTases	86
3.3.5.1.	The genes <i>PH1536</i> and <i>PH1537</i> might form a bicistronic arrangement	86
3.3.5.2.	The proteins PH1536 and PH1537 form a functional complex	88
3.3.5.3.	The N-terminal region of the protein PH1537 resembles the N1 domain of the RsmF family	91
3.3.6.	A proposed evolutionary classification of RNA m ⁵ C MTases	92
3.4.	Discussion	94
3.5.	Conclusion	97
	Chapter 4 – Structural and functional characterization of	99

archaeal m_2^6A rRNA MTase	
	Abstract 100
	Graphical abstract 101
4.1.	Introduction 101
4.2.	Materials and methods 104
4.2.1.	Construction of wild-type and mutant expression plasmids 104
4.2.2.	Protein overexpression and purification 106
4.2.3.	Mass spectrometric (MS) analysis 108
4.2.4.	Dynamic light scattering (DLS) analysis 109
4.2.5.	Fluorescence spectroscopy (FS) analysis 109
4.2.6.	Crystallization of wild-type and mutant <i>PhDIMIT1</i> 110
4.2.7.	Data collection, processing, and structure determination 112
4.2.8.	Thermal denaturation studies using circular dichroism (CD) spectroscopy 119
4.2.9.	Isothermal titration calorimetry (ITC) analysis 120
4.2.10.	Methylation assays 121
4.2.11.	Reverse phase high-performance liquid chromatography (RP-HPLC) analysis 122
4.2.12.	Bioinformatic analysis 122
4.3.	Results 123
4.3.1.	PH1823 is DIMT1/KsgA ortholog present in <i>P. horikoshii</i> 123
4.3.2.	The protein <i>PhDIMIT1</i> is conformationally and compositionally stable 125
4.3.3.	The protein <i>PhDIMIT1</i> shows specificity towards various cofactors 127
4.3.4.	The overall structure and active-site pocket of the protein <i>PhDIMIT1</i> 129
4.3.5.	Hydrolysis of SAM to MTA occurs during crystal incubation 132
4.3.6.	Binding of cofactors causes local conformational changes in the NCD 135
4.3.7.	Invariant water molecules stabilize the critical orientation of 137

Table of Contents

	cofactors in the NCD	
4.3.8.	SAH-bound <i>PhD</i> IMT1 reflects a transmethylation state	141
4.3.9.	<i>PhD</i> IMT1 demonstrates an interdomain expansion upon cofactor binding	144
4.3.10.	Mutation of catalytic residues of the protein <i>PhD</i> IMT1 altered the affinity for substrate and cofactors	147
4.3.11.	The active site residues Asp117 and Asn132 are crucial in maintaining the structural integrity of the protein <i>PhD</i> IMT1	152
4.4.	Discussion	154
4.4.1.	Localized conformational rearrangement contributes to the opening and closing of the cofactor-binding pocket	154
4.4.2.	Cofactor selectivity is critical for substrate binding during the initial state of dimethylation	156
4.4.3.	Interdomain movements of <i>PhD</i> IMT1 facilitate substrate binding	158
4.5.	Conclusion and future perspectives	158
	Chapter 5 – Structural and functional characterization of archaeal <i>m</i>⁵U rRNA MTase	160
	Abstract	161
	Graphical abstract	162
5.1.	Introduction	162
5.2.	Materials and methods	165
5.2.1.	Construction of wild-type expression plasmids	165
5.2.2.	Protein overexpression and purification	166
5.2.3.	Mass spectrometric (MS) analysis	168
5.2.4.	Dynamic light scattering (DLS) analysis	168
5.2.5.	Circular dichroism (CD) analysis	168
5.2.6.	Crystallization and co-crystallization of the protein <i>PhR</i> ImCD	169
5.2.7.	Data collection, processing, and structure determination	170
5.2.8.	Isothermal titration calorimetry (ITC) analysis	173

Table of Contents

5.2.9.	UV-visible absorbance spectroscopy	173
5.2.10.	Energy-dispersive X-ray (EDX) analysis	173
5.2.11.	Bioinformatic analysis	174
5.3.	Results	175
5.3.1.	Arm ⁵ U MTase is horizontally acquired from Gram-positive bacteria	175
5.3.2.	The protein <i>PhRlmCD</i> is conformationally and compositionally stable	177
5.3.3.	The overall structure of the protein <i>PhRlmCD</i>	179
5.3.4.	Despite the absence of an iron-sulfur cluster, <i>PhRlmCD</i> retains its native structure	182
5.3.5.	Central domain loops regulate the structural rearrangement of <i>PhRlmCD</i>	185
5.3.6.	The <i>PhRlmCD</i> enzyme is a member of the classical Rossmann-fold MTase (RFM) family	187
5.3.7.	The TRAM and the central domains of <i>PhRlmCD</i> regulate the substrate selectivity	189
5.3.8.	<i>PhRlmCD</i> shows dual substrate specificity	192
5.4.	Discussion	194
5.5	Conclusion and future perspectives	199
	Chapter 6 – Structural and functional characterization of archaeal <i>m</i>¹ψ rRNA MTase	201
	Abstract	202
	Graphical abstract	203
6.1.	Introduction	203
6.2.	Materials and methods	206
6.2.1.	Construction of wild-type expression plasmids	206
6.2.2.	Protein overexpression and purification	208
6.2.3.	Size exclusion chromatography (SEC) analysis	209
6.2.4.	Dynamic light scattering (DLS) analysis	210
6.2.5.	Western blot analysis	210

Table of Contents

6.2.6.	Mass spectrometric (MS) analysis	210
6.2.7.	Crystallization of <i>PhNep1</i>	211
6.2.8.	Data collection, processing, and structure determination	212
6.2.9.	Thermal denaturation studies using circular dichroism (CD) spectroscopy	215
6.2.10.	Methyltransferase assays	216
6.2.11.	Bioinformatic analysis	216
6.3.	Results	217
6.3.1.	The putative protein PH1379 is <i>Nep1</i> ortholog present in <i>P. horikoshii</i>	217
6.3.2.	The protein <i>PhNep1</i> exhibits both conformational and compositional stability	220
6.3.3.	Overall structural fold and active-site pocket of the protein <i>PhNep1</i>	222
6.3.3.1.	The Architecture of <i>PhNep1</i>	222
6.3.3.2.	The cofactor-binding site of <i>PhNep1</i>	223
6.3.4.	Inter-subunit hydrogen bonds stabilize the <i>PhNep1</i> dimer	226
6.3.5.	Structural intricacies accentuate the mechanism of RNA binding by <i>PhNep1</i>	230
6.3.5.1.	Possible RNA-binding site and active center	230
6.3.5.2.	Interaction of <i>PhNep1</i> with rRNA	231
6.4.	Discussion	235
6.5	Conclusion and future perspectives	238
	Conclusion and summary	240
	Appendix-A - Supplementary data to Chapter 3	244
	Appendix-B - Supplementary data to Chapter 4	265
	Appendix-C - Supplementary data to Chapter 5	285
	Appendix-D - Supplementary data to Chapter 6	304
	Bibliography	317
	List of publications, conferences, and workshops	348



The logo of the Indian Institute of Technology Guwahati is a circular emblem. It features a central stylized 'S' or 'Om' symbol composed of three interlocking circles. The outer ring of the emblem contains the text 'Indian Institute of Technology Guwahati' in English and its Assamese equivalent 'ভাৰতীয় প্ৰযুক্তিগতী সংস্থান গুৱাহাটী'.

ABSTRACT

Abstract

The process of biosynthesis of ribosomes is called ribosome biogenesis. This involves various steps like synthesis, processing, and modification of both ribosomal RNA (rRNA) and proteins (Rps). During ribosome biogenesis, specific nucleotides in the rRNA are chemically modified to distinguish them from each other. These modified nucleotides act as a molecular adhesive to stabilize a functionally active rRNA conformation, facilitating the fine-tuning of the ribosome biogenesis process. In ribosomes, three major types of contacts are formed by rRNA modifications: The first type of contact involves the modified nucleotides located within the P site of the small ribosomal subunit in 16S rRNA. These modified residues directly interact with either mRNA and tRNAs or with hibernating factor pY by forming hydrophobic contacts. Loss of these modifications can impair the translational process. The second type of contact involves the 23S rRNA modifications located at the interface between the large and small ribosomal subunits. The third type of contact is the most abundant group comprising the remaining nucleotides buried inside and residing within the single-stranded rRNA segments or in highly distorted helices. These modifications are located in energetically unfavourable folds. Moreover, the machinery involved in the rRNA modification process can also be subdivided into two major groups: (1) stand-alone enzymes that are present in all domains of life, comprising base methylation (methylation at nucleobase of the targeted nucleotide) and (2) RNA-guided rRNA modifications that use ribonucleoprotein particles (RNPs) to guide and modify the target rRNA in a sequence-dependent manner. The second type of rRNA modification involves either methylation or non-methylation of the rRNA by the H/ACA box small nuclear ribonucleoproteins (snRNPs) and the C/D box snRNPs during the pseudouridylation (conversion of uridine to pseudouridine) and ribose methylation (methylation at the 2'-OH group of the ribose of the targeted nucleotide), respectively. It is important to note that bacterial ribosomes typically have base-methylated nucleotides, while eukaryotes and archaea predominantly consist of RNA-guided rRNA modifications.

All these methylations are carried out by a class of enzymes, namely methyltransferases (MTases), constituting a large enzyme superfamily within the transferase group. Most of these enzymes use S-adenosyl-L-methionine (SAM or AdoMet) as a cofactor to transfer

methyl groups to specific residues of DNA, RNA, proteins, and others, with the generation of S-adenosyl-L-homocysteine (SAH or AdoHcy) after catalysis. Based on the structure of the catalytic domain (fold), SAM-dependent MTases are divided into five distinct classes - Class I (Rossmann-like α/β), Class II (TIM $\beta/\beta/\alpha$ -barrel α), Class III (tetrapyrrole methylase α/β), Class IV (SPOUT α/β), and Class V (SET domain all beta). The most abundant groups among these classes are class I and IV MTases, distinguished by their typical catalytic fold: Rossmann fold for class I and SPOUT fold (α/β fold, forming a deep trefoil knot) for class IV. Apart from SAM, in some cases, the nucleotide methylation can be catalyzed by the folate-dependent MTases where N^5,N^{10} -methylenetetrahydrofolate (THF) function as a donor of the methyl group.

In this study, we have focused on nucleotide modification in RNA (rRNA and tRNA) and the enzymes responsible for it, RNA MTases. In the recent past, the rRNA modification sites and their respective MTases have been mapped for few model species, such as *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens*, for instances, m^6A , m^5C , m^5U , $m^1\psi$, m^7G , Nm, etc. In *E. coli*, 16S rRNA contains 11 modified sites, including 10 methylations and a pseudouridylation. Similarly, 23S rRNA contains 25 known modifications, of which 13 are methylated, nine are pseudouridine, a methylated pseudouridine, a dihydrouridine, and a hydroxycytidine. In contrast, the eukaryotic model organism *S. cerevisiae* has been reported to possess 44 pseudouridines, 55 ribose methylations, and 10 base methylations, while *H. sapiens* has 91 pseudouridines, 105 ribose methylations, and 10 base methylations. However, the number and types of rRNA modifications vary enormously across archaea. For instance, halophilic archaea possess fewer rRNA modifications. Specifically, the Nep1 class of rRNA MTase (eukaryotic homolog of rRNA MTase) was not found in the phylogenetically-related Methanogen class II and Haloarchaea, unlike the order of the Thermococcales in euryarchaeota. In archaea, the previous study shows that the Thermococcales order, which can grow at remarkably high temperatures (near the boiling point of water), contains a large amount of rRNA modification, including base methylation and snoRNPs mediated modification.

Hence, in this study, we have selected *Pyrococcus horikoshii* OT3, a hyperthermophilic, anaerobic archaeon from the order Thermococcales, to identify the number of well-characterized as well as putative RNA MTases along with their sites of modification. Based on the support of the literature and comprehensive data mining approach, probable open reading frames (ORFs), which function as MTases from *P. horikoshii* OT3, were identified. The results enable the identification of 60 different types of MTases, out of which 21 MTases are potentially involved as RNA MTase (10 rRNA and 11 tRNA MTases). Out of the 21 RNA MTases, 18 are responsible for base methylation, which includes m⁵C, m₂⁶A, m⁵U, m⁶A, m²G, m¹G, and m¹A. Two of them methylate pseudouridine (m¹ψ), and one is responsible for ribose methylation (Cytidine-2'-O-methyltransferase). Following the comprehensive identification of RNA modifications in *P. horikoshii* OT3, our study majorly focuses on characterizing the putative rRNA MTases and their corresponding modification sites. In pursuit of this goal, we have specifically targeted three distinct types of base methylation (m⁵C, m₂⁶A, and m⁵U) and one pseudouridine methylation (m¹ψ) associated with rRNA.

Regarding m⁵C modification, the genus *Pyrococcus* has the highest number (5) of m⁵C MTases. Among these five putative RNA m⁵C MTases, two proteins (PH0851 and PH1991) have already been identified as homologs of a bacterial rRNA MTase (RsmB) and eukaryal tRNA MTase (NSUN6), respectively. Our study aimed to characterize the remaining three putative RNA m⁵C MTases (PH1078, PH1374, and PH1537) based on their signature architecture and catalytic residues, likely involved in binding their cognate RNA substrates. The results suggest the existence of two RsmB-like proteins (PH0851 and PH1078) belonging to the same subfamily IV of m⁵C RNA MTase. However, the proteins PH1374 and PH1537 belong to the same subfamily V but bind to different substrates, rRNA and tRNA, respectively. The findings further indicate that archaeal RNA m⁵C MTases link those from bacteria and eukarya. The rRNA MTases m₂⁶A (PH1823, referred to as *PhD*IMT1), m⁵U (PH1259, referred to as *PhR*ImCD), and m¹ψ (PH1379, referred to as *PhN*ep1) are found in *P. horikoshii*. These MTases were further characterized by biophysical methods.

The crystal structures of a dimethyladenosine transferase 1 (DIME1, an ortholog of bacterial KsgA) from a hyperthermophilic archaeon *P. horikoshii* (*PhDIME1*, ORF id: PH1823) were elucidated in its apo and holo forms at a resolution range of 2.0 to 3.3 Å. The study found that the analog sinefungin (SFG) binds differently than molecules SAM and SAH within the DIME1-SFG complex. This insight sheds light on catalytic inactivation. The protein transitions from an open to closed state with different cofactors, but only SAM and SAH expand the domain interface, positioning the substrate correctly.

Additionally, this study unveils the first crystal structure of an rRNA-specific archaeal m⁵U (Arm⁵U, ORF id: PH1259) MTase, namely *PhRlmCD*. The results revealed the importance of the iron-sulfur cluster in RNA m⁵U MTase. Replacing it with disulfide bridges obstructs the substrate binding site. Structural analysis revealed unique features of *PhRlmCD*, influencing substrate selectivity. The protein interacts with two putative mini-rRNA fragments at a positively-charged cleft to facilitate dual methyltransferase activities like its ancestral protein.

Further, the crystal structure of nucleolar essential protein 1 (Nep1) from *P. horikoshii* (*PhNep1*, ORF id: PH1379) was determined in its apo and holo forms at a resolution range of 2.0 to 3.1 Å. The study found that *PhNep1* possesses the SPOUT domain, featuring a deep trefoil knot with two novel extensions. The cofactor-binding site features an ordered cofactor-binding pocket similar to other Nep1-class MTases. The *PhNep1* is a homodimer facilitated by inter-subunit hydrogen bonds and hydrophobic interactions. This study also showed that *PhNep1* only methylates consensus rRNA when pseudouridine (ψ) is at position 926 within helix 35 (h35) of 16S rRNA in *P. horikoshii*. The stability of the Nep1-RNA complex relies on conserved arginine residues.

In summary, the findings from this study provide a preliminary glimpse of archaeal RNA MTases, an area that has not received attention. The structural and functional analysis of the proteins DIME1, RlmCD, and Nep1 from the archaeon *P. horikoshii*, alongside comparisons with counterparts from the other two domains of life, provide evolutionary insights into rRNA methyltransferases. Furthermore, the study identified possible target

sites to develop new inhibitors akin to the antibiotic SFG, which have therapeutic properties to fight the antibiotic resistance-causing MTases found in pathogenic bacteria.



Abbreviations and Notations

Abbreviations

ACN	Acetonitrile
ADN	Adenosine
APBS	Adaptive Poisson-Boltzmann Solver
Arm ⁵ U	Archaeal rRNA m ⁵ U methyltransferase
BLAST	Basic local alignment search tool
CRBD	C-terminal RNA binding domain
CTD	C-terminal domain
CTE	C-terminal extension
CV	Column-volume
DC	Decoding center
DCI	Dynamic community identifier
DIMT1	Dimethyladenosine transferase 1
DLS	Dynamic light scattering
EDX	Energy dispersive X-ray
EDTA	Ethylenediaminetetraacetic acid
Erm	Erythromycin resistance methyl transferase
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
IW	Invariant water
hdANM	Hinge domain anisotropic network model
LB	Luria-Bertani
LGA	Lamarckian genetic algorithm
LSU	Large subunit
LUCA	Last universal common ancestor

Abbreviations and Notations

m ⁵ C	5'-Methylcytosine
m ⁶ ₂ A	N6,6-dimethyl adenosine
MALDI-ToF	Matrix-associated laser desorption/ionization-time of flight
ML	Maximum likelihood
mRNA	Messenger RNA
MSA	Multiple sequence alignment
MTA	5'-Methylthioadenosine
MTase	Methyltransferase
MR	Molecular replacement
NBD	Nucleotide-binding domain
NCD	N-terminal catalytic domain
NCBI	National Center for Biotechnology Information
Nep1	Nucleolar essential protein 1
NEB	New England Biology
NFW	Nuclease-free water
Ni-NTA	Ni-nitrilotriacetic acid
NJ	Neighbor-joining
NTD	N-terminal domain
NTE	N-terminal extension
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PIC	Protease inhibitor cocktail
PMSF	Phenylmethylsulfonyl fluoride
PTC	Peptidyl transferase center
QC	Query coverage
RMSD	Root mean square deviation
RlmC	Ribosomal RNA large subunit methyltransferase C
RlmCD	Ribosomal RNA large subunit methyltransferase CD
RlmD	Ribosomal RNA large subunit methyltransferase D

Abbreviations and Notations

RP _s	Ribosomal proteins
RP-HPLC	Reverse phase-high performance liquid chromatography
RRF	Ribosomal recycling factor
rRNA	Ribosomal RNA
RsmA	Ribosomal small subunit methyltransferase A
RtcB	Non-canonical tRNA ligase
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SASA	Solvent accessible surface area
SEC	Size exclusion chromatography
SFG	Sinefungine (Adenosyl-L-ornithine)
SI	Sequence identity
SSU	Small subunit
TFA	Trifluoroacetic acid
TFB1M	Mitochondrial transcription factor1M
TrmA	tRNA methyltransferase A
tRNA	Transfer RNA
β-ME	β-mercaptoethanol
ψ	Pseudouridine
Ψ ¹ m	N1-methylpseudouridine

NOTATIONS

~	Approximately
%	Percent
Å	Angstrom
a.u	Arbitrary unit
bp	Base pair
°C	Degree Celsius
EC ₅₀	Amount of enzyme required for 50% maximal activity
F _U	Fraction of unfolded molecule
g	Relative Centrifugal Force (RCF)
K _a	Association constant
kb	kilobase
kcal	Kilocalories
k _D	Dissociation constant
Da	Dalton
K _{av}	Distribution coefficient
kDa	kilodalton
K _M	Michaelis-Menten constant
K _{cat}	Turnover number
K _{cat} /K _M	Catalytic efficiency
L	Litre
M	Molar
mdeg	Change in ellipticity
mM	Millimolar
mL	Millilitre
M _w	Molecular weight
n	Binding stoichiometry
OD	Optical density
RLU	Net relative luminescence unit
s	Second
T _m	Melting point

Abbreviations and Notations

ΔG	Change in free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
V_e	Elution volume
$V_{\max}$	Maximum reaction velocity
V_t	Total volume
λ	Wavelength
μg	Microgram
μL	Microlitre
μM	Micromolar



List of Tables

List of Tables

Table 1.1.	List of MTases that are found in the small ribosomal subunits of model organisms <i>E. coli</i> , <i>S. cerevisiae</i> , and <i>H. sapiens</i> (including mitochondria)	10
Table 1.2.	List of MTases found in the large ribosomal subunits of model organisms <i>E. coli</i> , <i>S. cerevisiae</i> , and <i>H. sapiens</i> (including mitochondria)	11
Table 1.3.	List of rRNA MTases associated with antibiotic resistance.	25
Table 1.4.	The list of RNA MTases found in the archaeal model species <i>P. horikoshii</i> OT3	31
Table 3.1.	Details of the protein-protein docking results of targeted proteins (PH1537 and PH1536) and their homologs (PAB1947 and PAB1946)	89
Table 4.1.	List of oligonucleotide sequences used to construct the wild-type (<i>PhDIMIT1_WT</i>) and mutant (<i>PhDIMIT1_D117A</i> , <i>PhDIMIT1_N132A</i> , <i>PhDIMIT1_Y135A</i>) protein expression plasmids	105
Table 4.2.	X-ray crystallographic data collection and refinement statistics for the protein <i>PhDIMIT1_WT</i> (apo and holo forms)	113
Table 4.3.	X-ray crystallographic data collection and refinement statistics for the mutant proteins of <i>PhDIMIT1</i> (apo and holo forms)	116
Table 5.1.	X-ray crystallographic data collection and refinement statistics for the protein <i>PhRlmCD</i>	171
Table 6.1.	X-ray crystallographic data collection and refinement statistics for the protein <i>PhNep1</i> (apo and holo forms)	213

List of Figures

Figure 1.1.	Comparison of prokaryotic and eukaryotic ribosomal components	3
Figure 1.2.	The functional elements of the bacterial ribosome from <i>Thermus thermophilus</i> (PDB id: 6XHV)	6
Figure 1.3.	Types of rRNA modifications	7
Figure 1.4.	Bacterial modification (<i>E. coli</i>) sites in ribosomal RNAs	10
Figure 1.5.	The S _N 2-like methyl transfer reaction	14
Figure 1.6.	Structural features of Class I SAM-dependent MTases	15
Figure 1.7.	Conserved methylation sites on rRNA exist universally across both the SSU and LSU	18
Figure 1.8.	Schematic representation of the <i>rrnB</i> operon illustrating the nucleolytic processing of the primary transcript.	20
Figure 1.9.	Eukaryotic ribosome biogenesis	21
Figure 1.10.	The impact of nucleotide modifications in the decoding center (DC) of 16S rRNA and the peptidyl transferase center (PTC) of 23S rRNA on antibiotic resistance	26
Figure 2.1.	A schematic representation that illustrates the comprehensive approach employed to achieve the outlined objectives	35
Figure 2.2.	Schematic representation for molecular cloning	42
Figure 2.3.	A schematic overview of the PCR cycle illustrates its three key stages: denaturation, annealing, and extension	43
Figure 2.4.	A visual depiction illustrating the phase diagram of protein crystallization	55
Figure 3.1.	Distribution of RNA m ⁵ C MTases across the four major archaeal lineages	68
Figure 3.2.	Multiple sequence alignment of the conserved active-site region of all the major RNA m ⁵ C MTases, along with those from <i>P. horikoshii</i>	70

List of Figures

Figure 3.3.	The full-length pair-wise sequence comparison of RNA m ⁵ C MTases from archaea, bacteria, and eukaryotes	71
Figure 3.4.	Multiple sequence alignment of five archaeal RNA m ⁵ C MTases along with the RsmBs, RsmF, and NSUN1	73
Figure 3.5.	The protein-protein interaction network involving PH1078 and its homologs <i>EcRmB</i> and PH0851	75
Figure 3.6.	Comparison of the active site of the protein PH1078 with its closest homologs <i>EcRsmB</i> and PH0851	79
Figure 3.7.	Sequence conservation in the N-terminal extension (NTE) of the subfamily I (RsmBs) and subfamily IV (PH0851 and PH1078) with NusB proteins	82
Figure 3.8.	The protein PH1374 probably functions as a rRNA m ⁵ C MTase	85
Figure 3.9.	Operonic organization of the gene <i>archease</i> in prokaryotes	87
Figure 3.10.	Protein-protein docking between the proteins PH1536 and PH1537	90
Figure 3.11.	The N1 domain of PH1537 is highly conserved with bacterial RsmF proteins	92
Figure 3.12.	Evolutionary relationship of the RNA m ⁵ C MTases	93
Figure 4.1.	Cloning and site-directed mutagenesis of gene <i>PhDIMT1</i>	105
Figure 4.2.	Purification of <i>PhDIMT1</i> _WT and its mutant proteins	108
Figure 4.3.	Overview of the crystals of <i>PhDIMT1</i> _WT and the mutants in its apo and complex form	111
Figure 4.4.	PH1823 is an archaeal DIMT1 ortholog	124
Figure 4.5.	Conformation and compositional stability of the protein <i>PhDIMT1</i> _WT	126
Figure 4.6.	The binding affinity of the protein <i>PhDIMT1</i> for various cofactors	128
Figure 4.7.	The overall structure and active site of the protein <i>PhDIMT1</i>	131

List of Figures

Figure 4.8.	The stability assay for the molecule SAM	134
Figure 4.9.	Conformational flexibility of the protein <i>PhD1MT1</i> upon cofactor binding	136
Figure 4.10.	The significance of structurally invariant water molecules	140
Figure 4.11.	Mechanism of substrate binding	143
Figure 4.12.	Movement of the CRBD upon cofactor binding	146
Figure 4.13.	Analysis of mutant <i>PhD1MT1</i> proteins	151
Figure 4.14.	Structural integrity upon the active site residues mutations	153
Figure 5.1.	Cloning of gene <i>PhRlmCD</i>	166
Figure 5.2.	Purification of <i>PhRlmCD</i>	167
Figure 5.3.	Crystal and diffraction of <i>PhRlmCD</i>	170
Figure 5.4.	Genomic evidence suggests the acquisition of <i>rlmCD</i> genes in archaea from Gram-positive bacteria	177
Figure 5.5.	Conformation and compositional stability of the protein <i>PhRlmCD</i>	178
Figure 5.6.	Overall three-dimensional structure of the protein <i>PhRlmCD</i>	181
Figure 5.7.	<i>PhRlmCD</i> attains stability in the absence of an iron-sulfur cluster through the organization of disulfide bridges	184
Figure 5.8.	The loops 11 and 15 in <i>PhRlmCD</i> are critical for structural alteration	186
Figure 5.9.	<i>PhRlmCD</i> belongs to Class I SAM-dependent MTase	188
Figure 5.10.	The distinctive traits of the apo form of <i>PhRlmCD</i>	191
Figure 5.11.	<i>PhRlmCD</i> demonstrates the ability to recognize both mini-rRNA fragments, U875 _F and U2055 _F , derived from domains II and IV of the 23S rRNA of <i>P. horikoshii</i>	194
Figure 6.1.	Cloning of gene <i>PhNep1</i>	207
Figure 6.2.	Overexpression and purification of <i>PhNep1</i>	209
Figure 6.3.	Crystallization of the protein <i>PhNep1</i>	212

List of Figures

Figure 6.4.	PH1379 is an archaeal Nep1 ortholog	210
Figure 6.5.	Maintaining compositional and conformational stability of <i>PhNep1</i>	221
Figure 6.6.	The architecture of protein <i>PhNep1</i> and its active site	225
Figure 6.7.	Dimerization of <i>PhNep1</i>	229
Figure 6.8.	<i>PhNep1</i> demonstrates the ability to recognize mini-rRNA consensus motif derived from h35 of the 16S rRNA of <i>P. horikoshii</i>	234



List of PDBs

PhDIMIT1 (PH1823)

S. No.	Title	PDB id
1	Crystal structure of DIMIT1 from the thermophilic archaeon, <i>Pyrococcus horikoshii</i>	8X3W
2	Crystal structure of DIMIT1 in complex with 5'-methylthioadenosine and adenosine from <i>Pyrococcus horikoshii</i>	8X41
3	Crystal structure of DIMIT1 in complex with 5'-methylthioadenosine from <i>Pyrococcus horikoshii</i> (FormI)	8X44
4	Crystal structure of DIMIT1 in complex with 5'-methylthioadenosine from <i>Pyrococcus horikoshii</i> (FormII)	8X45
5	Crystal structure of DIMIT1 in complex with adenosylornithine (SFG) from <i>Pyrococcus horikoshii</i>	8X46
6	Crystal structure of DIMIT1 in complex with S-adenosyl-L-homocysteine (SAH) from <i>Pyrococcus horikoshii</i>	8X47
7	Crystal structure of the D117A mutant of DIMIT1 from <i>Pyrococcus horikoshii</i>	8X4G
8	Crystal structure of the D117A mutant of DIMIT1 in complex with 5'-methylthioadenosine from <i>Pyrococcus horikoshii</i>	8X4I
9	Crystal structure of the N132A mutant of DIMIT1 from <i>Pyrococcus horikoshii</i>	8X4L
10	Crystal structure of the Y135A mutant of DIMIT1 in complex with 5'-methylthioadenosine from <i>Pyrococcus horikoshii</i>	8X4O
11	Crystal structure of the Y135A mutant of DIMIT1 in complex with adenosylornithine (SFG) from <i>Pyrococcus horikoshii</i>	8X4P

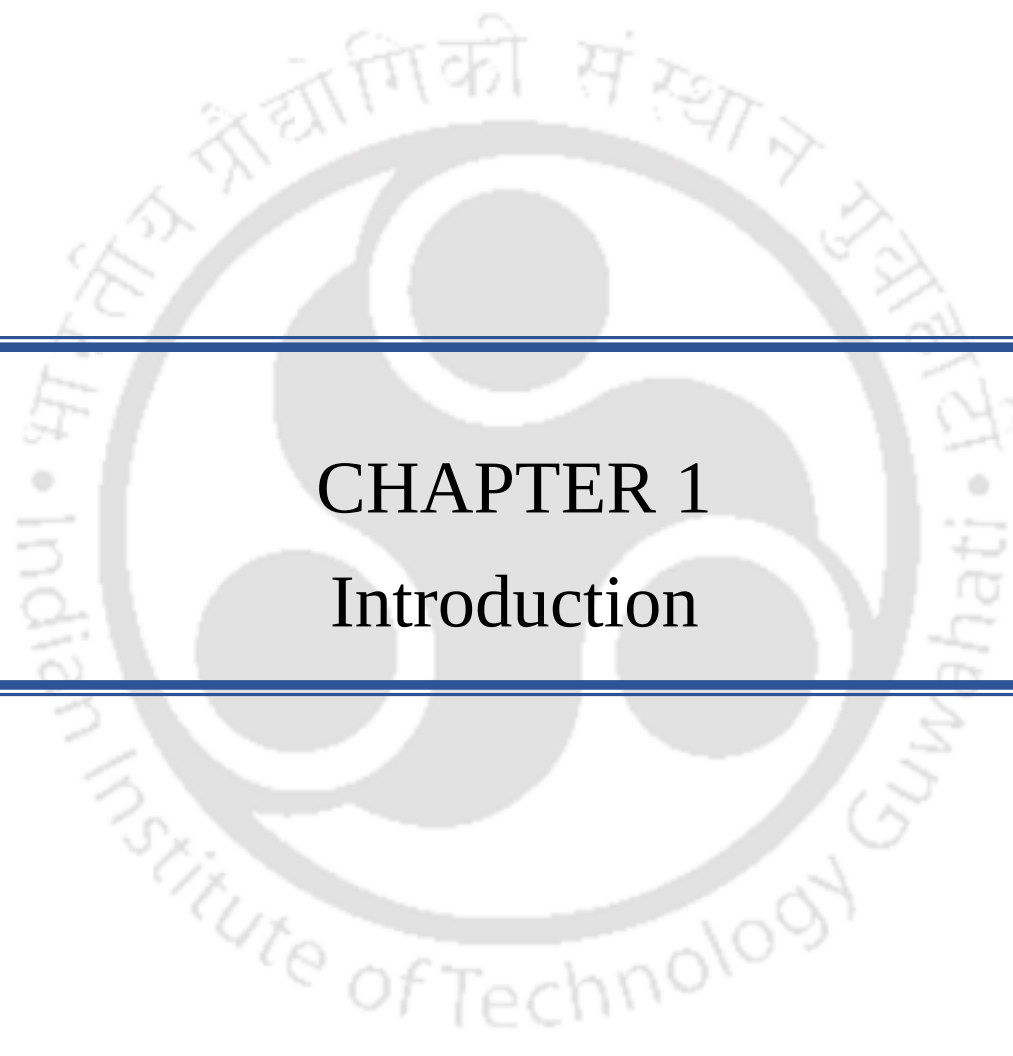
PhRlmCD (PH1259)

S. No.	Title	PDB id
1	The crystal structure of rRNA (uracil-C5)- methyltransferase from the thermophilic archaeon, <i>Pyrococcus horikoshii</i>	8Z62

PhNep1 (PH1379)

S. No.	Title	PDB id
1	The crystal structure of Nep1 from the thermophilic archaeon, <i>Pyrococcus horikoshii</i>	9JGB
2	Crystal structure of Nep1 in complex with adenosine from <i>Pyrococcus horikoshii</i>	9JGC
3	Crystal structure of Nep1 in complex with 5'-methylthioadenosine from <i>Pyrococcus horikoshii</i>	9JGD





CHAPTER 1

Introduction

1. Introduction

1.1. Ribosomal subunit composition

Ribosomes are ribonucleoproteins consisting of ribosomal RNAs (rRNAs) and proteins (Rps) that work together to translate genetic information from messenger RNA (mRNA) into proteins during protein biosynthesis in all organisms. Prokaryotic ribosomes are found in bacteria and archaea, comprising two subunits - a small subunit (SSU) 30S and a large subunit (LSU) 50S, together form a 70S ribosome. On the other hand, eukaryotic ribosomes are much larger and are found in animals, plants, fungi, and unicellular organisms with a nucleus. They consist of a 40S SSU and 60S LSU, which together form an 80S ribosome (Alberts et al., 2002; Ben-Shem et al., 2011). The difference in size between these two types of ribosomes is due to their sedimentation coefficients, which measure how fast they sediment in an ultracentrifuge. Eukaryotic ribosomes have a higher sedimentation coefficient than prokaryotic ribosomes, which means they sediment faster. It is important to note that sedimentation coefficients are not additive, which means the 30S and 50S subunits of an assembled ribosome have a sedimentation coefficient of 70S, not 80S. The higher the sedimentation coefficient, the larger or more spherical the structure (Teller, 1973).

All ribosomes share the same general architecture, with small and large ribosomal subunits composed of rRNA (~65%) and ribosomal proteins (~35%). The LSU is composed of three rRNAs in eukaryotes (25S, 5.8S, 5S in *Saccharomyces cerevisiae*) and two in prokaryotes (23S, 5S in *Escherichia coli*). On the other hand, the SSU has only one rRNA present in all kingdoms, specifically 18S and 16S in eukaryotes and prokaryotes, respectively. In bacteria, three rRNAs are as a single transcript, and in eukaryotes, four rRNAs are synthesized this way. These transcripts must be processed and cleaved at specific sites during the assembly process to produce mature rRNA species (Hage and Tollervey, 2004; Connolly and Culver, 2009). Although, eukaryotic ribosome subunits are generally larger than their prokaryotic counterparts. For instance, *S. cerevisiae* has a total of 5500 nucleotides and 83 Rps, while *E. coli* has 4567 total nucleotides and 55 Rps (Figure 1.1). Similarly, archaeal ribosomes consist of three types of rRNAs: the SSU 16S rRNA and the

LSU 23S and 5S rRNAs, which interact with 60-70 r-proteins, forming a complex network. (Melnikov et al., 2012; Ban et al., 2014; Bowman et al., 2020). The high-resolution crystal structures of prokaryotic ribosomes poorly fit the electron density observed in cryo-EM images of eukaryotic ribosomes. However, the overall shapes of both ribosomes are highly similar (Beckmann et al., 2001; Spahn et al., 2001; Chandramouli et al., 2008).

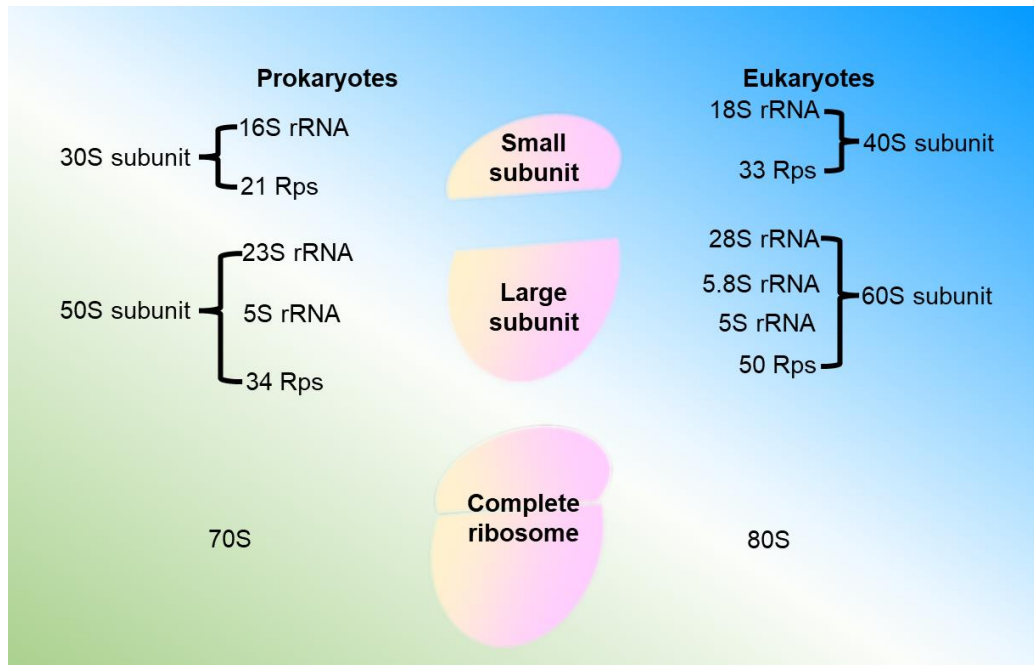


Figure 1.1. Comparison of prokaryotic and eukaryotic ribosomal components. The ribosomes of both prokaryotic and eukaryotic cells consist of two subunits of unequal size, each containing at least one rRNA molecule and multiple Rps.

1.2. Functional center of the ribosome

The ribosome functions as a platform that binds mRNA and transfer RNAs (tRNAs). Each tRNA has two functional ends: one end carries the amino acid (CCA-end), and the other end contains the anticodon that recognizes the mRNA codon. The tRNAs bind to the ribosome in three places: the aminoacyl (A) site, the peptidyl (P) site, and the exit (E) site. The incoming aminoacyl-tRNA (aa-tRNA) binds to the A site, while the peptidyl-tRNA carrying the nascent polypeptide chain binds to the P site (Figure 1.2). The deacyl-tRNA binds to the E site before it dissociates from the ribosome. These tRNA molecules are essential adaptors that connect the two primary functional centers of the ribosome: the

decoding center (DC) in the small subunit and the catalytic peptidyl transferase center (PTC) at the core of the large subunit (Ghosh et al., 2015; Tirumalai et al., 2021).

1.2.1. Decoding center (DC)

The decoding center (DC) is composed of five helices, namely h18, h44, h34, h24, and h31 of SSU rRNA (16S/18S rRNA). This has been indicated through various experimental methods such as chemical protection, cross-linking, and genetic studies (Rodnina et al., 2002; Noller, 2006). Detailed mapping of nucleotides within these helices that interact with A-, P-, or E-site tRNAs and mRNA has been achieved through crystallography studies (Ogle et al., 2001; Yusupov et al., 2001; Selmer et al., 2006). The functional significance of specific nucleotides in decoding has been elucidated through mutational analyses, impacting translation efficiency, fidelity, and susceptibility to error-inducing antibiotics (De Stasio et al., 1989; Recht et al., 1999; Kubarenko et al., 2006; Garcia-Ortega et al., 2008; Saraiya et al., 2008).

These investigations emphasize the significance of the sequence and structure of decoding centers in regulating tRNA selection. The structure of rRNA can significantly change to affect tRNA binding to the ribosome during the translation of other mRNAs (Lodmell and Dahlberg, 1997). It is believed that in 16S/18S rRNA, certain nucleotides in the rRNA alternate base pairing between one nucleotide or another, forming a "switch" that alters the rRNA's conformation. This process can affect the structure of the LSU and SSU, indicating that this conformational switch in the rRNA structure affects the entire ribosome's ability to match a codon with its anticodon in tRNA selection as well as decode mRNA (Gabashvili et al., 1999).

1.2.2. Peptidyl transferase center (PTC)

The Peptidyl transferase center (PTC) maintains a consistent structure across both the 50S and 70S ribosomal subunits. While the 50S subunit structures are largely similar in archaea and bacteria, there's a distinction in the E site of bacteria due to the absence of ribosomal protein L44E. The A and P sites likely predate the E site, possibly originating after the Last Universal Common Ancestor (LUCA) (Steitz 2008). The PTC is pivotal in protein synthesis, spanning approximately 165 bases with notable twofold pseudo-symmetry. In

this region, a single self-folding RNA aids in the chemical bonding of peptide formation between the amino acid moiety of aminoacyl-tRNA (aa-tRNA) and the growing polypeptide chain linked to the P-site tRNA (Sergiev et al., 2018). The PTC comprises a crucial helix, h69, which was captured in a 5.5 Å structure of the 70S ribosome of *Thermus thermophilus*. Conversely, h69 appears disordered in a 2.4 Å structure of the 50S subunit of *Haloarcula marismortui*. This suggests the dynamic role of h69 in subunit joining (Ban et al., 2000; Harms et al., 2001). Within the intact ribosome, h69 interacts with tRNAs, hinting at its roles in tRNA translocation and translation accuracy (Moazed and Noller, 1989; Yusupov et al., 2001; Gao et al., 2003). Recent genetic studies have highlighted the involvement of h69 in translation fidelity, with mutations in its loop region causing significant +1 frameshifting and stop codon readthrough (Hirabayashi et al., 2006). Furthermore, h69 is implicated in EF-G-catalyzed tRNA movement through cryo-EM analysis (Agrawal et al., 2000). In contrast, X-ray crystallography reveals the binding of h69 to the ribosome recycling factor (RRF), suggesting its critical role in ribosome assembly (Wilson et al., 2005). Subsequently, deletion of h69 in bacteria leads to a dominant lethal phenotype and compromised peptide release (Ali et al., 2006).

The elucidation of rRNA secondary structure represents a pivotal advancement in understanding the ribosome. These specialized helices within the rRNA secondary structure provide insights into how base pairs, double-helices, loops, bulges, and single-strands contribute to the framework of the ribosome (Figure 1.2). Ultimately, this aids in comprehending the mechanisms of global assembly and evolution of the ribosome (Glotz et al., 1981). A significant proportion of nucleotides within the 16S and 23S rRNA undergo post-transcriptional modifications situated in these pivotal centers of the ribosome. In *Saccharomyces cerevisiae*, these modifications account for ~60% of these nucleotides, while in *Escherichia coli*, the number rises to ~95% (Decatur and Fournier, 2002). Importantly, the majority of these modified nucleotides are found in conserved positions, highlighting their significance in ribosomal function and potentially in the synthesis process (Lestrade and Weber, 2006).

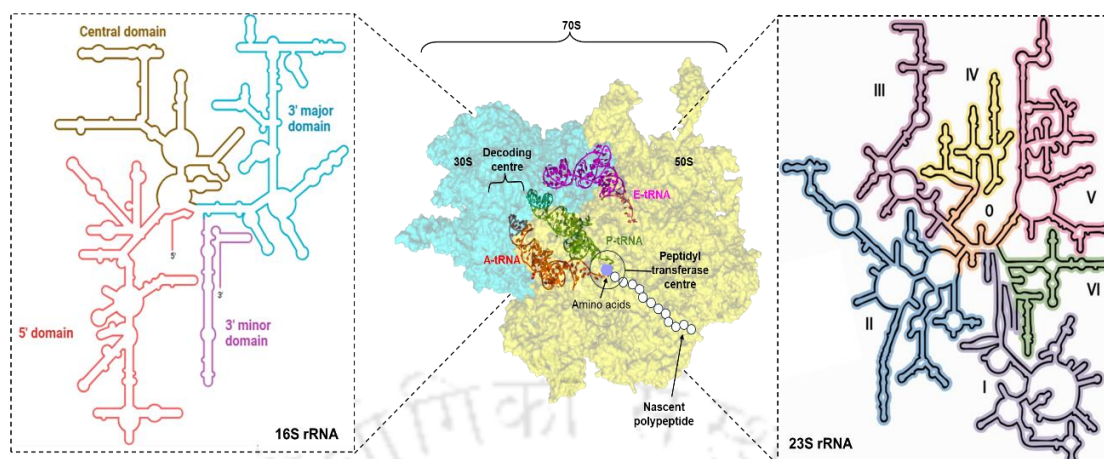


Figure 1.2. The functional elements of the bacterial ribosome from *Thermus thermophilus* (PDB id: 6XHV). The 30S subunit is depicted in cyan and the 50S subunit in yellow. The tRNAs are color-coded: red for the A site, green for the P site, magenta for the E site, and nascent polypeptide chain is represented by white circles. Two critical functional centers of the ribosome are identified. On the left and right sides of the panel, the secondary structure of 16S and 23S rRNA is delineated. The 16S rRNA molecule is composed of four domains: the 5' major and central domains (red and yellow), as well as the 3' major and minor domains (cyan and magenta). Meanwhile, 23S rRNA presents seven domains: domain 0 (orange), domain I (purple), domain II (blue), domain III (magenta), domain IV (yellow), domain V (pink), and domain VI (green).

1.3. Epigenetic modifications of rRNA

RNA modifications are widespread in all living organisms, contributing significantly to RNA metabolism and gene regulation, collectively known as "RNA epigenetics" (Schaefer et al., 2017). These modifications, which occur after transcription in various RNA types such as mRNA, rRNA, and tRNA, are essential for a multitude of cellular processes, particularly translation. Notably, they play a crucial role in ribosome biogenesis and fine-tuning ribosome function, including translation accuracy and rate (Sunita et al., 2007). In prokaryotes, the modified nucleotides in SSU typically occur on the surface and are added during late assembly stages. Conversely, modifications in the LSU occur earlier in the assembly process (Siibak and Remme, 2010). While in eukaryotes, rRNA modifications are

(A) RNA-dependent modification

The diagram shows a schematic of rRNA with regions 3-10, 4-10, and Box H or ACA (14-16 nt). An arrow labeled **MTase** points to the chemical structures of 2'-O-methylated ribonucleoside and N1-methyl-pseudouridine, both formed from SAM and releasing SAH.

(B) RNA-independent modification

The diagram shows a central oval labeled **rRNA modification**. An arrow labeled **MTase** points to the chemical structures of 6-methyladenosine (**m⁶A**), Methyladenosine (**m¹A**), and 5-methylcytosine (**m⁵C**), all formed from SAM and releasing SAH.

Figure 1.3. Types of rRNA modifications. (A) RNA-dependent modification requires single-strand RNA as a guide to identify the target site from modification. (B) RNA-independent modification takes place without the help of the guide RNA.

There are three primary types of rRNA modification: pseudouridylation (methylated or non-methylated), ribose methylation, and base methylation (Decatur and Fournier, 2002).

1.3.1. Pseudouridylation and 2'-O-ribose methylation

These two types of rRNA modifications are predominant in eukaryotic and archaeal RNA modification. Among them, pseudouridine (Ψ) emerges as a key modification, exerting its influence primarily in the realm of post-transcriptional mRNA modification, such as facilitating spliceosome-mediated RNA splicing and also contributing to ribosome processing. On the other hand, the 2'-O-methyl ribose nucleoside (Nm) plays a critical role in orchestrating the assembly of the 40S and 60S subunits in eukaryotic ribosomes (Kiss-László et al., 1996). The conversion of uridine to pseudouridine and nucleosides to 2'-O-methylated nucleosides is orchestrated by a complex ensemble known as the SnoRNA-protein complex (SnoRNP) (Figure 1.3A). This molecular machinery comprises two essential components: a small nucleolar RNA (snoRNA), meticulously tailored to be site and type-specific, which acts as the guiding force through base pairing with the substrate, and a complement of associated proteins that serve as the catalysts for the modification reaction. SnoRNA-protein complexes are termed snoRNPs (pronounced as “Snorps”). SnoRNAs are involved in various stages of eukaryotic ribosome biogenesis within the nucleolus. The two major families of snoRNAs that guide the modification of rRNA by methylated pseudouridine and 2'-O-methylation are box H (ANANNA, where N stands for any nucleotide) / ACA snoRNAs and box C (RUGAUGA, where R stands for any purine) / D(CUGA), respectively (Lowe and Eddy, 1999).

1.3.2. Base methylation

This form of modification involves the addition of methyl groups directly to the nucleobases of rRNA. Base methylation in rRNA is essential in prokaryotes to improve

protein translation efficiency (Krzyszosiak et al., 1987). Unlike other RNA modifications, which are guided by small RNA molecules, base methylation is carried out by a different class of MTases (majorly class I SAM-dependent MTases). These MTases possess the remarkable ability to locate and modify specific nucleotide targets without needing guidance from small RNA molecules (Figure 1.3B).

Base methylations are particularly common in prokaryotes and lower eukaryotes, where they play vital roles in various cellular processes. In contrast, higher eukaryotes, as well as certain bacteria and archaea, exhibit a greater prevalence of 2'-O-methylations of the ribose followed by pseudouridylations (Charette and Gray, 2000; Sergiev et al., 2018). This distinction underscores the diverse strategies employed across different organisms for RNA modification, reflecting the intricate evolution of molecular mechanisms tailored to meet specific biological needs.

The extent of RNA modifications correlates with the complexity of organisms, exemplified by varying numbers across different species. For instance, *E. coli* showcases 11 pseudouridines, 4 ribose methylations, and 19 base methylations (Figure 1.4 and Table 1.1-1.2). On the other hand, lower eukaryotes, particularly the model organism *S. cerevisiae* boast 44 pseudouridines, 55 ribose methylations, and 10 base methylations, while higher *H. sapiens* exhibit 91 pseudouridines, 105 ribose methylations, and 10 base methylations (Table 1.1-1.2) (Dubnau and Nomura, 1984; Gustafsson et al., 1996; Sergiev et al., 2018). Moreover, beyond the three domains of life, rRNA modifications are also prevalent in mitochondria. However, in the case of rRNA modification, mammalian mt-rRNAs exhibit fewer modifications compared to bacterial and cytoplasmic rRNAs. Presently, eight modifications have been identified, including three 2'-O-ribose methylations and five base methylations (Sergiev et al., 2018).

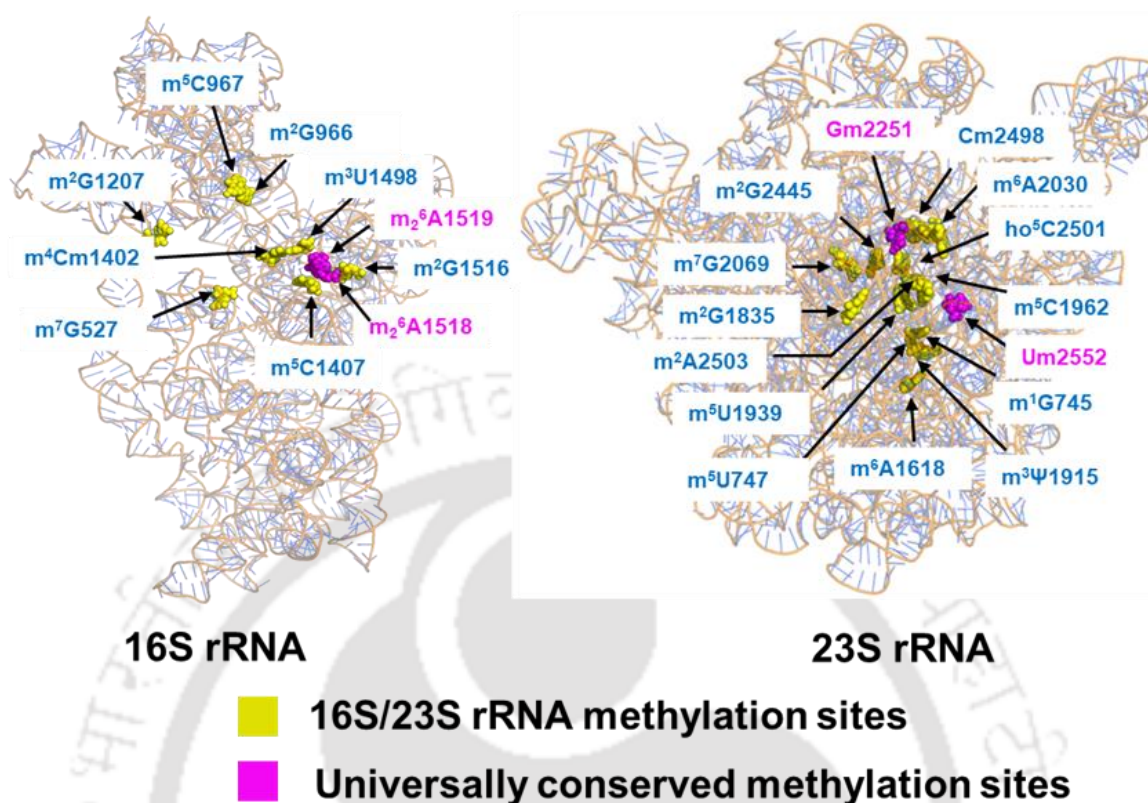


Figure 1.4. Bacterial modification (*E. coli*) sites in ribosomal RNAs. Spatial distribution of modified nucleotides in the structures of small (left) and large (right) ribosomal subunits from *E. coli*.

Table 1.1. List of MTases that are found in the small ribosomal subunits of model organisms *E. coli*, *S. cerevisiae*, and *H. sapiens* (including mitochondria) (Sergiev et al., 2018).

Methylated rRNA residues in the small ribosomal subunit							
<i>E. coli</i> (16S rRNA)		<i>S. cerevisiae</i> (18S rRNA)		<i>H. sapiens</i> (18S rRNA)		<i>H. sapiens</i> mitochondrial (12S rRNA)	
Nucleotide	Enzyme	Nucleotide	Enzyme	Nucleotide	Enzyme	Nucleotide	Enzyme
m ⁷ G527	RsmG	G574	Unmodified	G623	Unmodified	G252 (899)	Unmodified

Chapter 1. Introduction

							d
U788	Unmodified	Ψ999	snR31	Ψ1056	ACA8	m ⁵ U429 (1076)	Unknown
m ² G966	RsmD	m ¹ acp ³ Ψ1191	snR35; Emg1 (Nep1); Tsr3	m ¹ acp ³ Ψ1248	ACA13; hEMG1; TSR3	A573 (1220)	Unmodified
m ⁵ C967	RsmB	C1192	Unmodified	C1249	Unmodified	A574 (1221)	Unmodified
m ² G1207	RsmC	G1438	Unmodified	G1500	Unmodified	U690 (1337)	Unmodified
G1338	Unmodified	m ⁷ G1575	Bud23/Trm112	m ⁷ G1639	WBSCR2 / TRMT112	G775 (1422)	Unmodified
m ⁴ Cm1402	RsmH; RsmI	Cm1639	snR70	Cm1703	U43	m ⁴ C839 (1486)	Unknown
C1404	Unmodified	C1641	Unmodified	C1705	Unmodified	m ⁵ C841 (1488)	NSU N482
m ⁵ C1407	RsmF	C1644	Unmodified	C1708	Unmodified	C844 (1491)	Unmodified
m ³ U1498	RsmE	U1761	Unmodified	U1830	Unmodified	U916 (1563)	Unmodified
A1500	Unmodified	A1763	Unmodified	m ⁶ A1832	Unknown 3	A918 (1565)	Unmodified
m ² G1516	RsmJ	U1779	Unmodified	U1848	Unmodified	G934 (1581)	Unmodified
m ₂ ⁶ A1518 m ₂ ⁶ A1519	KsgA (RsmA)	m ₂ ⁶ A1781 m ₂ ⁶ A1782	Dim1	m ₂ ⁶ A1850 m ₂ ⁶ A1851	DIMT1L	m ₂ ⁶ A936 (1583) m ₂ ⁶ A937 (1584)	TFB1M

Table 1.2. List of MTases found in the large ribosomal subunits of model organisms *E. coli*, *S. cerevisiae*, and *H. sapiens* (including mitochondria) (Sergiev et al., 2018).

Methylated rRNA residues in the large ribosomal subunit			
<i>E. coli</i> (23S rRNA)	<i>S. cerevisiae</i> (25S rRNA)	<i>H. sapiens</i> (28S rRNA)	<i>H. sapiens</i> mitochondrial (16S rRNA)

Chapter 1. Introduction

Nucleotide	Enzyme	Nucleotide	Enzyme	Nucleotide	Enzyme	Nucleotide	Enzyme
U571	Unmodified	m ¹ A645	Rrp (Bmt1)	m ¹ A1322	NML17	C182 (1852)	Unmodified
m ¹ G745	RlmAI (RrmA, YebH); RlmAII	C877	Unmodified	C1594	Unmodified	G285 (1955)	Unmodified
m ⁵ U747	RlmC (RumB);	U879	Unmodified	U1596	Unmodified	A287 (1957)	Unmodified
m ⁶ A1618	RlmF (YbiN)	A1850	Unmodified	A2798	Unmodified	A764 (2434)	Unmodified
A1784	Unmodified	m ¹ A2142	Bmt	A3646	Unmodified	A834 (2504)	Unmodified
m ² G1835	RlmG (YgjO)	G2194	Unmodified	G3698	Unmodified	G876 (2546)	Unmodified
m ³ Ψ1915	RluD (YfiI); RlmH (YbeA)	Ψ 2258	snR191	Ψ 3762	U1993	C908 (2578)	Unmodified
G1935	Unmodified	m ⁵ C2278	Rcm1 (Bmt3)	m ⁵ C3782	NSUN5 (WBSCR20)	A928 (2598)	Unmodified
m ⁵ U1939	RlmD (YgcA, RumA)	U2282	Unmodified	U3786	Unmodified	U932 (2602)	Unmodified
G1954	Unmodified	U2297	Unmodified	U3801	Unmodified	m ¹ A947 (2617)	TRMT61B
m ⁵ C1962	RlmI (YccW)	G2305	Unmodified	G3809	Unmodified	C955 (2625)	Unmodified
m ⁶ A2030	RlmJ	A2372	Unmodified	A3876	Unmodified	A1023 (2693)	Unmodified

m ⁷ G2069	RlmKL (YcbY)	U2411	Unmodified	U3915	Unmodified	G1062 (2732)	Unmodified
Gm2251	RlmB	Gm2619	snR67	Gm4196	U3127	Gm1145 (2815)	MRM1
U2265	Unmodified	m ³ U2634	Bmt5	C4211	Unmodified	C1159 (2829)	Unmodified
A2274	Unmodified	A2643	Unmodified	m ⁶ A4220	Unknown3	A1168 (2838)	Unmodified
m ² G2445	RlmKL (YcbY)	G2814	Unmodified	G4391	Unmodified	G1262 (2932)	Unmodified
U2474	Unmodified	m ³ U2843	Bmt6	U4420	Unmodified	C1291 (2961)	Unmodified
Cm2498	RlmM (YgdE)	C2867	Unmodified	C4444	Unmodified	C1315 (2985)	Unmodified
ho ⁵ C2501	RlhA	m ⁵ C2870	Nop2 (Bmt4)	m ⁵ C4447	NSUN1 (NOP2, NOL1)	C1318 (2988)	Unmodified
m ² A2503	RlmN (YfgB)	A2872	Unmodified	A4449	Unmodified	A1320 (2990)	Unmodified
Um2552	RlmE	Um2921	snR52	Um4498	Unknown	Um1369 (3039)	MRM2
G2553	Unmodified	Gm2922	Spb1	Gm4499	Unknown	Gm1370 (3040)	MRM3
U2584	Unmodified	U2953	Unmodified	m ³ U4530	Unknown	U1401 (3071)	Unmodified

1.4. Enzymes responsible for methylation of rRNA

Methylation of rRNAs takes place during ribosomal biogenesis either by enzymes guided by an antisense snoRNA or by conventional protein enzymes. All enzymes responsible for known rRNA methylation are rRNA MTases, a large superfamily of membrane-bound enzymes found in the endoplasmic reticulum. The general mechanism of methyl transfer is an S_N2-like nucleophilic attack where the methionine sulfur serves as the nucleophile that

transfers the methyl group to the enzyme-substrate complex and finally transfers it to the substrate. SAM is converted to SAH during this process. The breaking of the SAM-methyl bond and the formation of the substrate-methyl bond happens nearly simultaneously (Figure 1.5). There are three different kinds of MTase based on their substrate choice, namely DNA MTase, RNA MTase, and protein MTase, all of which perform the same reaction shown below (Cheng and Roberts, 2001).

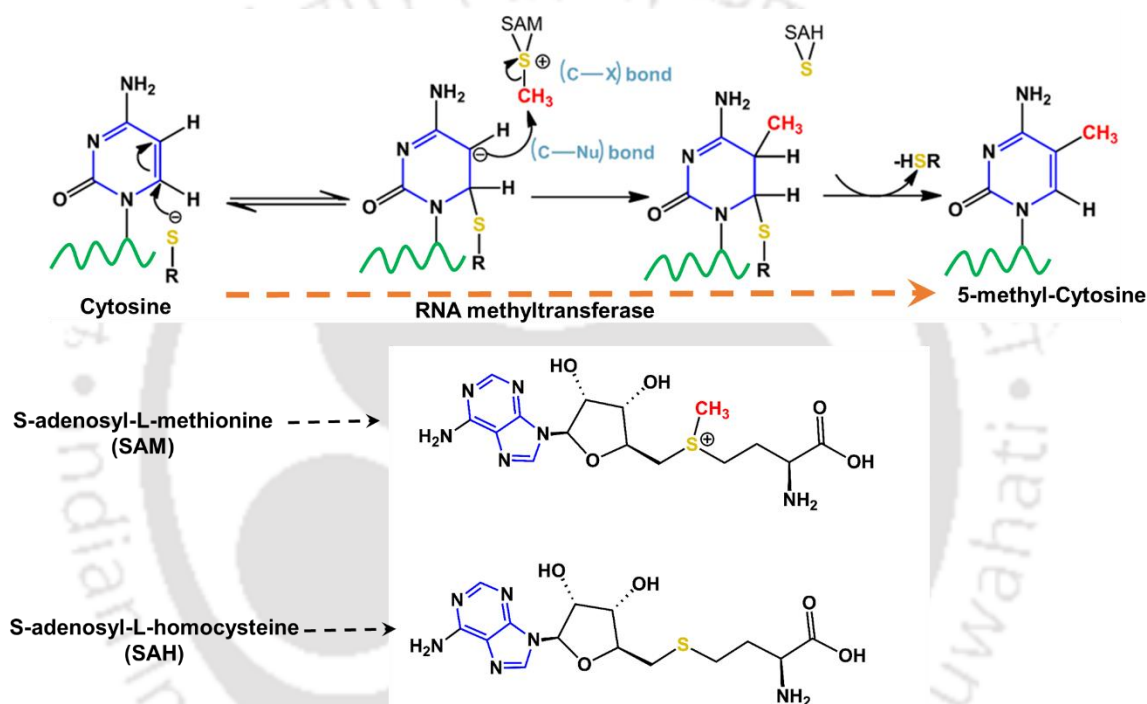


Figure 1.5. The S_N2 -like methyl transfer reaction. The breaking of the C–X bond and the formation of a new bond (C–Y or C–Nu) occur simultaneously through a transition state in which a carbon under nucleophilic attack is penta-coordinated and approximately sp^2 hybridized.

Based on the structure of the catalytic domain (fold), all MTases can be divided into 5 classes- Class I (Rossmann-like α/β), Class II (TIM β/α -barrel α/β), Class III (Tetrapyrrole methylase α/β), Class IV (SPOUT α/β) and Class V (SET domain all β) (Schubert et al., 2003). However, the majority of MTases belong to class I, which possess a Rossmann-like

fold. Rossmann fold is a super secondary structure having alternate $\beta\alpha\beta$ fold. This fold is found in enzymes that bind NAD^+ , FAD, and NADP^+ . The central core of the MTase structure is associated with seven β -strands. Wherein strand 7 is antiparallel to the other six strands and is inserted into the sheet between strands 5 and 6 ($6 \uparrow 7 \downarrow 5 \uparrow 4 \uparrow 1 \uparrow 2 \uparrow 3 \uparrow$) (Figure 1.6A). Most class I SAM-dependent enzymes have two or three domains. One domain is associated with catalytic activity, while the other one or two domains are mainly involved in RNA binding and positioning of targeted nucleotides in the catalytic groove. However, these MTases have poor conservation of SAM-binding residues, though each MTase binds SAM at the same equivalent position of the fold, and the cofactor adopts a similar binding conformation. However, the chemistry of the SAM-binding interaction varies tremendously. There are two conserved positions in class I MTases: one is glycine-rich sequence E/DXGXGXG (often referred to as motif I) between β_2 and α_3 , which interacts with the amino acid portion of SAM. The sequence alignment suggests that the glycine-rich region has signature motifs that vary among the different subclasses of enzymes, such as DNA MTases have the motif E/DXXXGXG, whereas RNA MTases contain YXGXXXG motif. The other one is an acidic loop between β_3 and α_4 (motif II), which interacts with the ribose hydroxyl groups of cofactors (SAM or SAH) (Schubert et al., 2003). A third region of sequence that contributes to the interaction with SAM is a linker between β_4 and α_4 (Schubert et al., 2003) (Figure 1.6B).

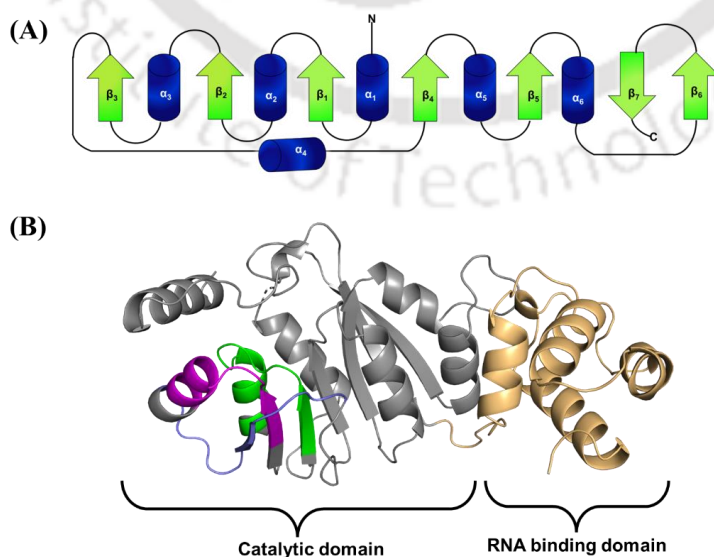


Figure 1.6. Structural features of Class I SAM-dependent MTases. (A) Schematic representation of the Class I SAM-dependent MTase topology, illustrating alpha helices depicted as deep blue cylinders and beta strands as green arrows. (B) The MTase fold depicts both the catalytic and RNA binding domains, which harbor the cofactor (SAM)- and substrate-binding sites. Notably, the primary motifs (I, II, and III) crucial for interaction with the cofactor are highlighted in green, magenta, and purple, respectively.

1.5. Methylation of rRNA in bacteria and eukaryotes

Methylation of rRNAs is a common characteristic found across all living organisms, and the occurrence of methylated rRNA residues at equivalent sites in both prokaryotes and eukaryotes suggests its evolutionary conservation. These conserved methylated residues are distributed across both the small subunit (SSU) and large subunit (LSU) of the ribosome.

1.5.1. Universally conserved rRNA methylation sites in the SSU

Within the rRNA of the small ribosomal subunit, two adenine residues, m_2^6A1518 and m_2^6A1519 (*E. coli* 16S rRNA numbering), undergo universal and specific N6-dimethylation. These modifications occur in helix 45 (h45), recognized as a critical 'contact hub' within the ribosome, positioned in close proximity to the P-site tRNA, mRNA, and the large ribosomal subunit (Poldermans et al., 1979; Polikanov et al., 2015). The introduction of these modifications is facilitated by the KsgA MTase enzyme. This methylation is critical as it facilitates contact between helices 44 and 45 near the decoding center of the ribosome, achieved through the formation of a hydrogen-bonding network that stabilizes this contact site as well as facilitates codon-anticodon base pairing (Figure 1.7A and Table 1.1) (Wimberly et al., 2000; Demirci et al., 2010). Inactivating or deleting the *ksgA* gene, responsible for dimethylating adenine residues at positions A1518 and A1519 of 16S rRNA, leads to subtle structural rearrangements within the decoding center, affecting the accuracy of translation initiation and elongation processes (Demirci et al., 2010; Van Buul et al., 1984). This slight alteration in the decoding center of the ribosome confers resistance to kasugamycin (Van Buul et al., 1984). While bacteria typically possess a single KsgA

MTase, eukaryotes boast multiple homologous enzymes responsible for modifying both cytoplasmic and mitochondrial ribosomal RNAs. In yeast, DIMT1, akin to KsgA, is crucial for methylating cytoplasmic ribosomes and is indispensable for cell viability. Mammalian mitochondria, on the other hand, harbor two paralogs of the KsgA MTases: TFB1M and TFB2M. TFB1M primarily methylates mitochondrial 12S rRNA, whereas TFB2M, despite primarily acting as a transcription factor, also retains rRNA MTase activity (Lafontaine et al., 1994; Cotney et al., 2009).

Apart from dimethylation, KsgA also engages with a late assembly intermediate of the 30S subunit, conducting a 'quality control check' by sequestering the intermediate until methylation occurs, after which the MTase dissociates (Connolly et al., 2008). On the other hand, the significance of DIMT1 MTase (eukaryotic counterpart of KsgA) during ribosome assembly extends beyond its methylation function, as lethality resulting from DIMT1 inactivation can be suppressed by expressing a catalytically deficient DIMT1 variant suggesting their other crucial role in the ribosome assembly process (Lafontaine et al., 1995).

1.5.2. Universally conserved rRNA methylation sites in the LSU

LSU features two highly conserved modified nucleotides, Gm2251 and Um2552 (using *E. coli* 23S rRNA numbering), situated in the P-loop and A-loop (helices 80 and 92, respectively) (Kiss-László et al., 1996; Lövgren and Wikström, 2001). Studies have elucidated that the methylation of Um2552 fits snugly between the neighboring bases G2553 and U2554, thereby preserving the active conformation of the G2553 base. This preserved conformation is vital for the accommodation of aminoacyl-tRNA molecules (Polikanov et al., 2015). Similarly, the 2'-O-methyl group of Gm2251 facilitates hydrophobic interactions with C2065 and U2449, thereby maintaining the active conformation of nucleotides engaged in base pairing with tRNAs at both the P- and A-sites (Polikanov et al., 2015) (Figure 1.7B-1.7C and Table 1.2). While the absence of Gm2251 in *E. coli* does not elicit any discernible phenotypic effects, the deficiency of Um2552 leads to a significant accumulation of LSU assembly intermediates (Lövgren and Wikström,

2001; Tan et al., 2002). In bacterial systems, the MTases RlmB and RlmE are responsible for introducing Gm2251 and Um2552, respectively (Caldas et al., 2000; Lövgren and Wikström, 2001). Conversely, their human cytoplasmic equivalents, Gm4196 and Um4498, undergo catalysis through a small nucleolar RNA (snoRNA)-guided mechanism and an unidentified process, respectively. Within mitochondria, modifications analogous to Gm2251 and Um2552 are introduced by mitochondrial rRNA MTase 1 and 2 (MRM1 and MRM2), respectively (Kiss-László et al., 1996).

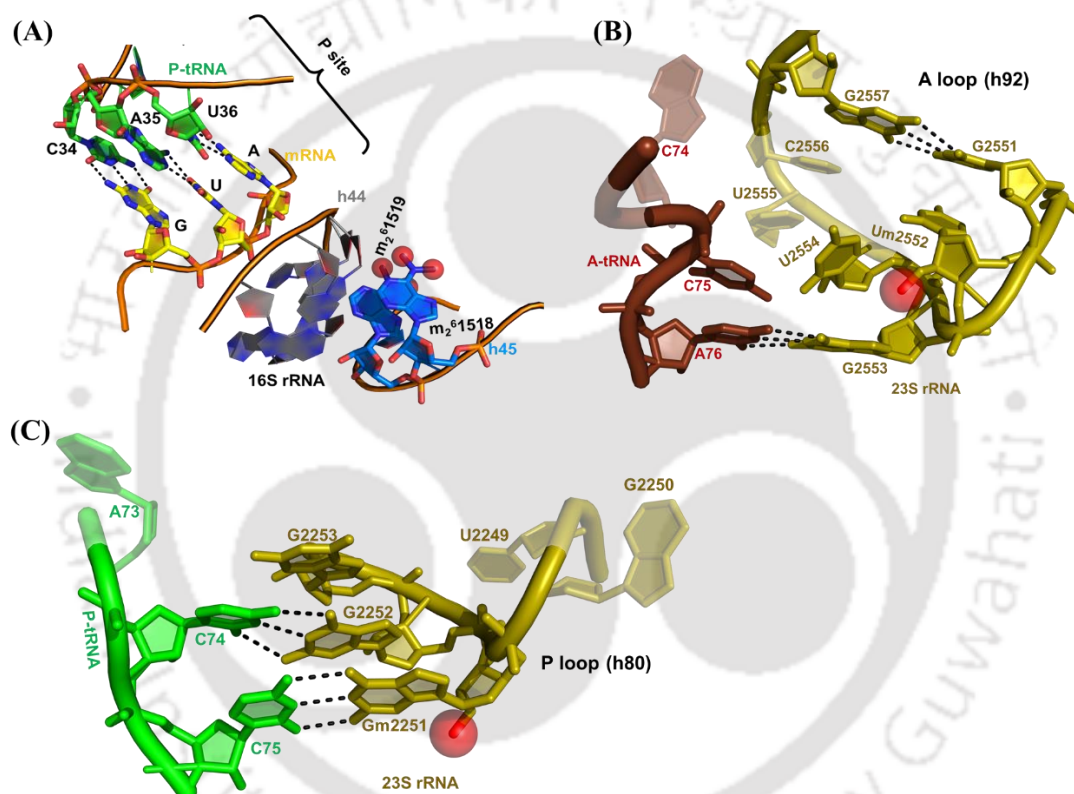


Figure 1.7. Conserved methylation sites on rRNA exist universally across both the SSU and LSU. (A) The positions of N6-dimethylated adenines, m_2^6A1518 and m_2^6A1519 , situated within h45 of the 16S rRNA (*E. coli* numbering), are proximal to the codon-anticodon double helix formed by mRNA (yellow) and the P-site tRNA (green) molecules. (B and C) The 2'-O-methylated nucleotides (olive) Um2552 (B) and Gm2251 (C) are situated in the A-loop (h92) and the P-loop (h80) of the 23S rRNA (*E. coli* numbering), respectively. Notably, nucleotides within the A- and P-loops are crucial for accommodating the aminoacyl-tRNA (A-tRNA, dark red) and ensuring the proper

positioning of the peptidyl-tRNA (P-tRNA, green) via Watson–Crick base pairing. All methylated positions of the targeted nucleotides are indicated by red spheres.

1.6. Biological significance of rRNA methylation

1.6.1. Ribosome biogenesis

Modification of nucleotides helps in the assembly of ribosome subunits. The difference between the biogenesis of prokaryotes and eukaryotes is discussed below.

1.6.1.1. Ribosome biogenesis in prokaryotes

Prokaryotic ribosome biogenesis is relatively simpler than the eukaryotic process. The pre *rrnB* transcript in prokaryotes is processed to obtain the mature rRNAs, which are synthesized as a single transcript (Nierhaus et al., 1991). A mature ribosomal rRNA comprises modified rRNA and ribosomal protein, which requires a high level of coordination as well as a host of additional factors for assembly.

The assembly of ribosomes involves a series of events:

- Ribosome biogenesis initiates with the transcription of the 16S, 23S, and 5S rRNA genes, which are synthesized as a single primary transcript from several rRNA operons. In *E. coli*, seven rRNA operons, namely *rrnA*, *rrnB*, *rrnC*, *rrnD*, *rrnE*, *rrnG*, and *rrnH*, play a significant role in cellular RNA synthesis. Numerous studies investigating rRNA expression have focused on the *rrnB* operon, selected as a representative model for studying rRNA operons (Figure 1.8) (Condon et al., 1995).
- Processing of rRNA, which includes nucleolytic processing of 16S, 23S, and 5S rRNA from primary transcript followed by chemical modification of rRNA. Most of the 16S rRNA modifications are added late during the maturation of the 30S subunit (Polikanov et al., 2015).

- Chemical modifications of ribosomal proteins (such as S11, L3, L7/L12, L16, L11, and L33) help them to stabilize through folding (Kaczanowska and Ryden-Aulin, 2007).
- Assembly of the 30S and 50S subunits with the help of metal ions and ribosomal proteins (Kaczanowska and Ryden-Aulin, 2007).

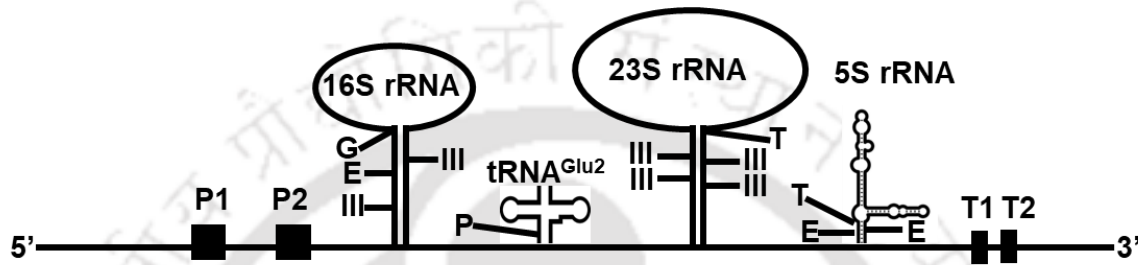


Figure 1.8. Schematic representation of the *rrnB* operon illustrating the nucleolytic processing of the primary transcript. The diagram depicts the rRNA and tRNA species, along with the promoters P1 and P2 and terminators T1 and T2. Processing sites targeted by RNase III (III), RNase G (G), RNase E (E), RNase P (P) and RNase T (T).

1.6.1.2. Ribosome biogenesis in eukaryotes

Eukaryotic ribosomes share a similar fundamental architecture as prokaryotes, but ribosome biogenesis in eukaryotes is much more complex than that in prokaryotes (Nazar, 2004). Ribosomal RNA undergoes cleavage and dozens of nucleotide modifications before assembly of ribosomal proteins into the mature ribosome. Several factors, including snoRNPs, GTPases, helicase, nucleases, transport proteins, and chaperones, are involved in eukaryotic ribosome biogenesis. The 60S or large subunit of eukaryotes is composed of the 28S (25S in yeast), 5S, and 5.8S rRNAs and 47 proteins in humans; the 40S or small subunit is composed of the 18S rRNA and 33 proteins. To generate the mature 18S, 28S and 5.8S rRNAs, a precursor 45S rRNA (35S in yeast) is transcribed by RNA pol I (Nazar, 2004). The pre-35S rRNA is transcribed in the nucleolus. Several factors, including U3 snoRNA, ribosomal proteins, and non-ribosomal factors associate with it to form a 90S

particle followed by covalent modification events such as 2'-O-methylation (Nm) and pseudouridylation (-Ψ) (Nazar, 2004) (Figure 1.9).

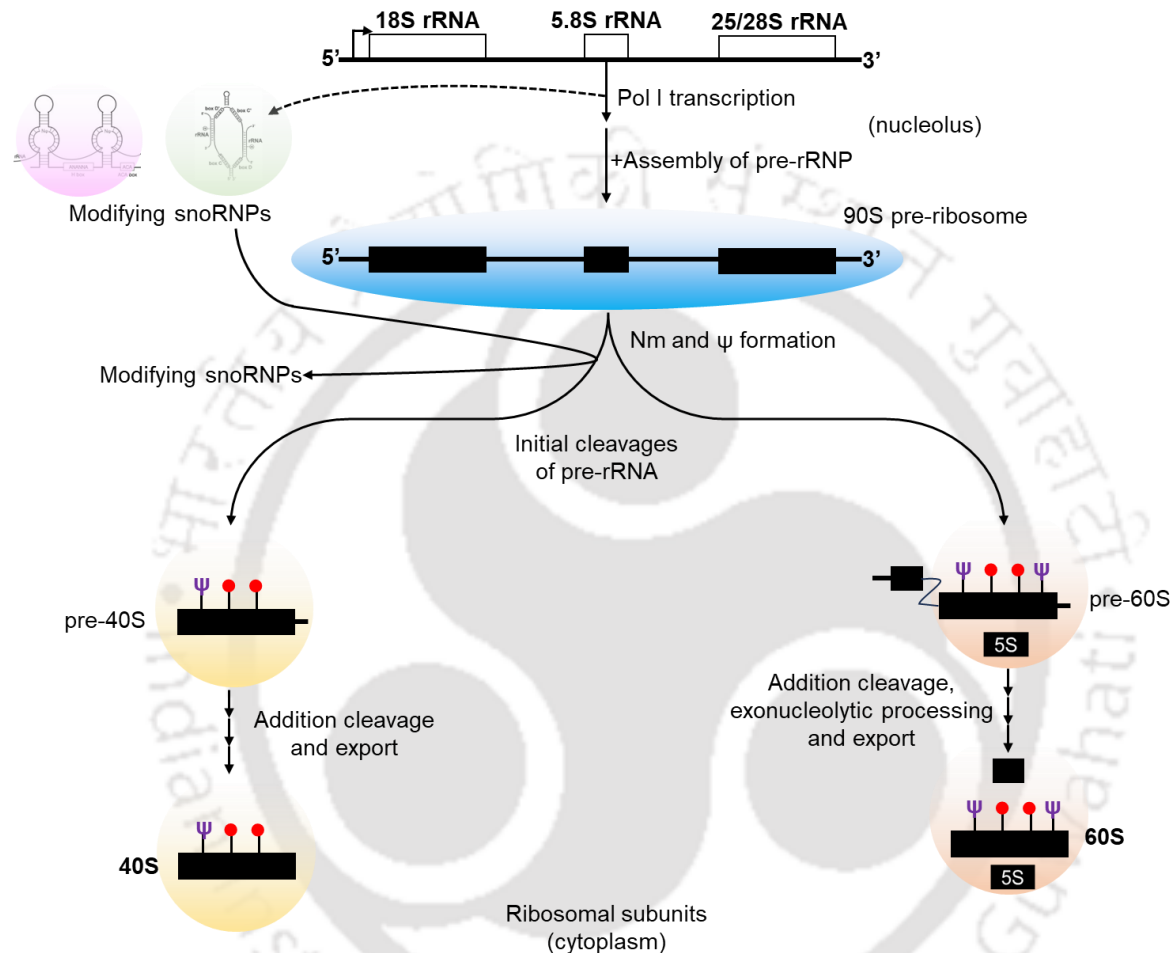


Figure 1.9. Eukaryotic ribosome biogenesis. Modifying snoRNPs act on pre-rRNA in the nucleolus during or after transcription.

1.6.2. Antibiotic resistance

The methylation of ribosomal RNA (rRNA) holds significant implications for human health. MTases can profoundly modify the drug-binding sites within ribosomes by adjusting the methylation status (the absence or presence of the methylation), which serves as one strategy for imparting antibiotic resistance (Morić. et al., 2010). The modified residues, predominantly methylated, exhibit varying degrees of resistance or susceptibility to antibiotics. These modifications play a crucial role in bacterial adaptation and resistance, often referred to as the "fitness cost," and serve as reservoirs for the evolution of resistance genes.

1.6.2.1. Resistance associated with 16S rRNA

Resistance to aminoglycosides in pathogenic bacteria can be caused by methylation in the DC region of ribosomes. Specifically, methylation occurs at the N7 (G1405, numbering according to *E. coli* 16S rRNA) and N1 (A1408) atoms of 16S rRNA, respectively. The N7 methylation at m⁷G1405 residue is catalyzed by ArmA and RmtAH MTases found in the SSU of pathogenic bacteria, resulting in resistance to aminoglycosides containing 4-6 substituted deoxystreptamine, such as kanamycin and gentamicin (Doi et al., 2007). Although this N7 methylation generally occurs in aminoglycoside-producing strains catalyzed by a large family of similar MTases, including KgmB, Sgm, FmrO, GrmA, GrmB, GrmO, Kmr, NbrB, and Srm1. However, it is believed that N7 methylation-causing MTases in pathogenic bacteria have been acquired through horizontal gene transfer (HGT) from them. (Doi et al., 2007; Grosjean, 2009).

On the other hand, modification of m¹A1408 in 16S rRNA can provide resistance to a wider range of aminoglycosides that contain either 4-6 or 4-5 substituted deoxystreptamine, such as neomycin and paromomycin. The modification is carried out by KamA, KamB, and KamC MTases for aminoglycoside producers, and CmnU for capreomycin has been identified in various strains (Grosjean, 2009). It was previously thought that pathogenic bacteria did not possess this type of resistance, but in 2007, a similar N1 A1408 MTase called NpmA was discovered in *E. coli* (Wachino et al., 2007). Although the evolutionary lineage of the NpmA enzyme is unclear, it is possible that these proteins originated from the tRNA-specific guanine N7 MTase, TrmB (Macmaster et al., 2010). Interestingly, the

MTase RsmF in *T. thermophilus* modifies three nucleotide residues (C1400, C1404, and C1407), indicating the relatively facile evolution of substrate specificity in these enzymes (Demirci et al., 2010). The modified m⁵C1404 residue resides close to the kanamycin-binding site but does not directly overlap with it, which excludes steric clashes as a resistance mechanism (François et al., 2005). Surprisingly, this modification does not interfere with the binding of gentamicin, despite its bulky methyl group (in contrast to the amino group in kanamycin) at the site of contact with the methyl group of m⁵C1404. It is believed that the m⁵C1404 modification confers resistance to kanamycin and tobramycin (Demirci et al., 2010).

On the other hand, there are a few cases where the absence of housekeeping methylation alters the decoding site of the ribosome, leading to resistance to certain antibiotics. For instance, SSU contains one universally conserved rRNA MTase known as ribosomal small subunit MTase A (RsmA or KsgA). Inactivation of the *ksgA* gene, eliminates the universal dimethylation of the adenine residues A1518 and A1519, and confers resistance against the antibiotic kasugamycin (Van Buul et al., 1984) (Figure 1.10A and Table 1.3).

1.6.2.2. Resistance associated with 23S rRNA

The most extensively studied modification involves the dimethylation of the 23S rRNA A2058 (numbering according to *E. coli* 23S rRNA) residue at the N6 atom by the erythromycin-resistant methyltransferase (Erm) family methylases. This modification confers resistance to a broad range of antibiotics, including macrolides, lincosamides, and streptogramins B. The 23S rRNA A2058 residue, situated in the PTC close to the peptide exit tunnel, interacts with antibiotics from all three classes. Dimethylation (or monomethylation in the case of lincosamides by TlrD) and related enzymes disrupt antibiotic binding by preventing the formation of hydrogen bonds with A2058 and causing steric hindrance with the methyl group. This resistance mechanism is prevalent among both macrolide producers and pathogenic strains targeted by these antibiotics (Skinner et al., 1983; Pernodet et al., 1996; Roberts et al., 1999; Bhujbalrao and Anand., 2019). It is thought that the *erm* gene family in bacteria, including pathogenic ones, has been acquired through HGT from producer strains. Moreover, methylation of the 23S rRNA G748

residue, situated within the LSU tunnel at the N1 atom concurrently with monomethylation of A2058, grants resistance to producer bacteria against tylosin, a 16-membered macrolide with bulky sugar substituents (Liu and Douthwaite, 2002b). Interestingly, modification of either A2058 or G748 alone does not confer this resistance. The formation of 23S rRNA m¹G748 is catalyzed by the MTase RlmA^{II} (TlrD), commonly found in Gram-positive bacteria. Conversely, Gram-negative bacteria typically harbor its homolog, RlmA^I, which modifies the neighboring 23S rRNA G745 residue at the N1 position (Liu and Douthwaite, 2002a). RlmA^I serves as a housekeeping enzyme, suggesting a potential common ancestor for both MTases. In *Bacillus subtilis*, the expression of the *rlmA*^{II} gene is regulated by transcription and translation attenuation, likely to mitigate the negative effects of methylation on adaptability in the absence of antibiotics. This raises questions about classifying such enzymes solely as housekeeping proteins (Yakhnin et al., 2019).

The modification of 23S rRNA impedes the binding of several antibiotic classes, including Chloramphenicol resistance methyltransferase (Cfr), confers cellular resistance to phenicols, lincosamides, pleuromycins, oxazolidinones, and streptogramins A by methylation of A2503 at the C8 atom (Morić et al., 2010). The *ermB* gene was discovered upstream of the *cfr* gene, and the promoter of their shared operon was induced by erythromycin, leading to resistance against seven antibiotic classes. Furthermore, only one modified residue, namely the pseudouridylation of the U2504 residue by RluC, contributes significantly to antibiotic resistance. This modification enhances resistance to tiamulin, clindamycin, linezolid, chloramphenicol, hygromycin A, and sparsomycin to some extent. The enzyme RluC catalyzes the conversion of three uridine residues (955, 2504, and 2580) into pseudouridine, wherein the modification of only ψ2504 amplifies cellular resistance to the mentioned inhibitors of the peptidyl transferase reaction (Conrad et al., 1998; Toh and Mankin, 2008). These findings suggest that the modifications of nucleotide residues in the PTC often lead to antibiotic resistance (Figure 1.10B and Table 1.3). Thus, it seems ideal for developing inhibitors against antibiotic resistance-causing MTases to prevent modification of antibiotic binding sites located in the DC and PTC of ribosomes.

Table 1.3. List of rRNA MTases associated with antibiotic resistance.

MTase gene	rRNA	Methylated site	*	Resistant antibiotic
Antibiotic resistance causing MTase				
<i>efmM</i> , <i>armA/sgm/kgmB</i> , <i>npmA</i> , <i>tlyA</i> , <i>rsmE</i> ,	16S	1404, G1405, A1408, Cm1409, U1498	+	Aminoglycoside group
<i>erm</i>	23S	G2058	+	Erythromycin, MLSB
<i>cfr</i> , <i>tyl</i> , <i>rluC</i> ,	23S	G748, A2503, ψ2504,	+	Chloramphenicol, florfenicol
Housekeeping MTase				
<i>ksgA</i>	16S	A1518, A1519	-	Aminoglycoside group

* Resistance conferred by the absence (–) or presence (+) of the methylation.

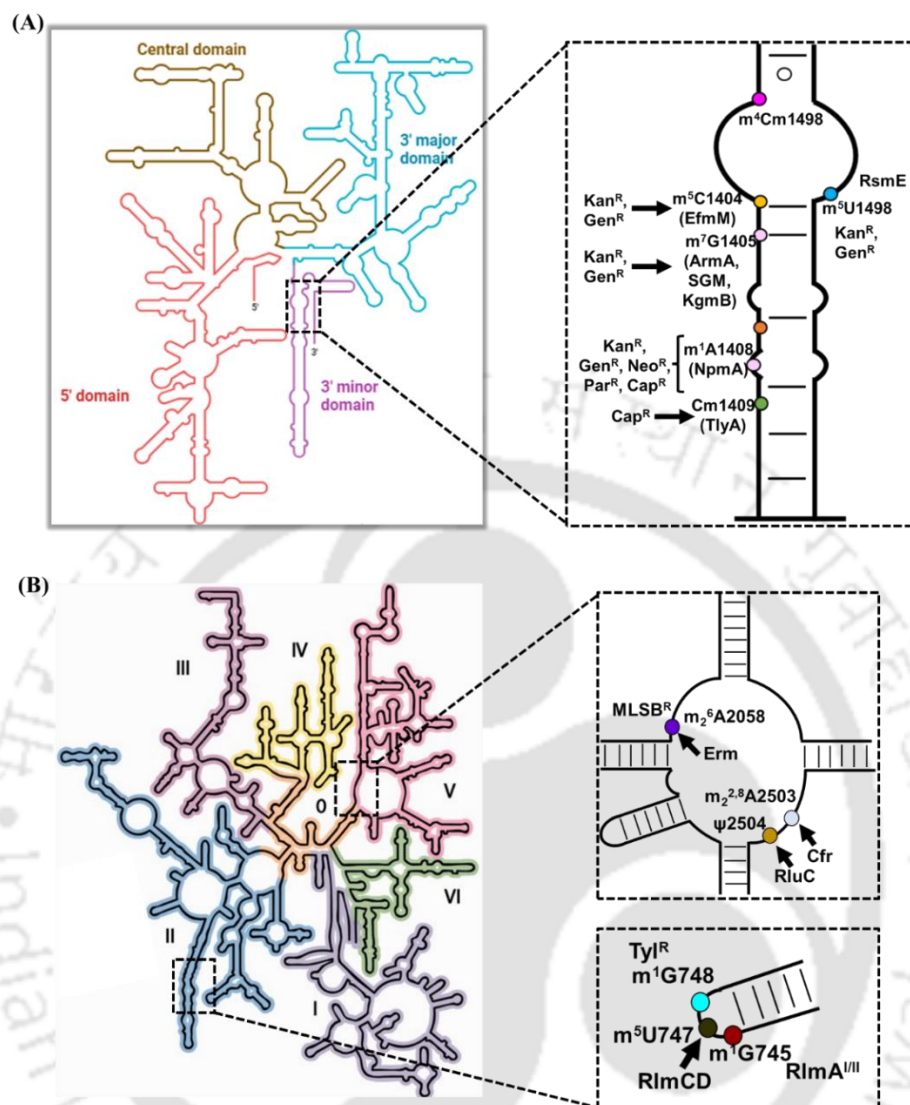


Figure 1.10. The impact of nucleotide modifications in the decoding center (DC) of 16S rRNA and the peptidyl transferase center (PTC) of 23S rRNA on antibiotic resistance. (A) The secondary structure of 3' minor domain (helix44; h44) in 16S rRNA, which constitutes a portion of the decoding center (DC). (B) The secondary structure of regions V and II in 23S rRNA, forms a segment of the peptide exit tunnel (PET) and the peptidyl transferase center (PTC). Modified nucleotides are depicted as circles.

1.6.3. Mutations in MTases can cause genetic disorders

A spectrum of human genetic disorders, termed ribosomopathies, arise from mutations in genes encoding ribosomal proteins or factors crucial for ribosome biogenesis, including those involved in rRNA methylation processes. For instance, Bowen-Conradi syndrome, characterized by severe developmental delays and growth failure leading to early infant mortality, is caused by a point mutation in the EMG1 MTase (Armistead et al., 2009). Prader-Willi syndrome, a neurological condition featuring intellectual disability, obesity, and muscle hypotonia, results from deletions in the 15q11–q13 locus containing snoRNAs crucial for RNA 2'-O-methylation (Sahoo et al., 2008). Additionally, mutations in genes like NOL1/NOP2/sun (Nsun) domain-containing genes encoding RNA MTases are linked to various neurodevelopmental disorders (Blanco and Frye, 2014). However, the precise impact of impaired rRNA methylation on the pathology of these disorders necessitates further investigation, given the dual role of some MTases in pre-rRNA processing.

1.6.4. Cancer

Growing evidence suggests a connection between methylation of rRNA (mRNA, tRNA, and rRNA) and cancer development. For instance, the m⁶A modification in mRNAs is linked to tumor proliferation in endometrial cancer (Liu et al., 2018), and m⁵C methylation of tRNAs by NSUN2 in skin cancer cells has been associated with tumorigenesis (Blanco et al., 2016). Likewise, dysregulated ribosome biogenesis is implicated in various cancers (Truitt and Ruggero, 2016). Moreover, the association between cancer and rRNA methylation is underlined by research demonstrating altered rRNA methylation patterns resulting from the inactivation of the p53 tumor-suppressor gene (Marcel et al., 2013). Future investigations may shed light on the specific roles of individual rRNA modifications in cancer. This holds particular interest as modulating ribosome biogenesis could offer an alternative means to halt cell proliferation and impede tumor formation (Brighenti et al., 2015).

1.7. Evolutionary aspect of the rRNA methylation

The evolution of rRNA MTases proceeded in both convergent and divergent ways. In convergent evolution, the same conserved nucleotide can be modified by completely

unrelated enzymes in different taxonomic groups, whereas closely related homologous enzymes can modify different targets for divergent evolution. Uridine methylation systems, particularly the generation of m⁵U residues in RNAs across diverse organisms, provide an excellent illustration of evolutionary relationships. Enzymes involved in m⁵U modification employ two biochemical pathways, utilizing either S-adenosyl methionine (SAM) or N⁵,N¹⁰-methylenetetrahydrofolate (THF) as the methyl group donor (Gupta and Griffiths, 2002; Koch, 2003). Bacteria employ both pathways to incorporate m⁵U residues into their RNA molecules. For instance, *Bacillus subtilis* and its likely ancestral counterparts utilize a single SAM-dependent RNA large subunit methyltransferase CD (RlmCD) to create m⁵U747 and m⁵U1939 residues in the 23S rRNA (Jiang et al., 2017). Concurrently, the formation of m⁵U54 in the TΨC-loops of tRNAs in this bacterium is catalyzed by the folate-dependent TrmFO MTases (Urbonavičius et al., 2005). Gene duplications in the ancestral lineage of *E. coli* and other proteobacteria resulted in three related SAM-dependent enzymes: RNA large subunit methyltransferase C, D (RlmC, RlmD), and tRNA methyltransferase A (TrmA). Two of these enzymes are involved in modifying m⁵U747 and m⁵U1939 in the 23S rRNA, while the third catalyzes the formation of m⁵U54 in the TΨC-loops of tRNAs (Lee et al., 2004, 2005; Walbott et al., 2008).

As we have discussed earlier, identifying the enzymes causing rRNA methylations has been completed for several model species like *E. coli*, *S. cerevisiae*, and *H. sapiens*. However, archaeal rRNA methylation remains unclear due to a lack of information. In addition to that, the number of rRNA modifications varies across domain archaea due to the reduced number of proteins and ribosome biogenesis factors in these organisms (Lecompte et al., 2002; Yutin et al., 2012; Ebersberger et al., 2014). For example, Nep1, the archaeal homologs of the eukaryotic MTase, is not found in the phylogenetically-related Methanogen class II and Haloarchaea, unlike organisms of the Thermococcales (Londei and Ferreira-Cerca, 2021).

1.8. Overview of rRNA modification in Archaea

In archaea, the order Thermococcales thrives in high temperatures and features a large amount of rRNA modification (Sleiman and Dragon, 2019; Coureux et al., 2020). Thus, in this thesis work, we focus on the identification and characterization of rRNA MTases from

an archaeon *P. horikoshii* OT3 from Thermococcales order. The structural, functional, and computational studies of rRNA MTases from the *P. horikoshii* OT3 would provide knowledge about the mode of methylation in the archaeal system. This would also help in understanding the similarities and differences present in domains other than archaea with respect to ribosome biogenesis, its assembly, and the role of these proteins in translation initiation. In addition, the growing resistance against antibiotics of pathogens is becoming a silent epidemic worldwide, which may kill more than 300 million people by 2050 (Bragginton and Piddock, 2014). Resistance by rRNA methylation is a frequent mechanism among macrolide and aminoglycoside antibiotic-producing actinomycetes strains. The proposed study would also provide insight into the molecular mechanism(s) of rRNA MTases and their association with antibiotic resistance. The acquisition of multidrug resistance (MDR) in pathogens acts as a hindrance to effective drug development. Hence, deciphering the structural aspects of rRNA MTases and their molecular mechanism would immensely help in understanding their potential role in antimicrobial resistance, thus making these proteins ideal targets for antimicrobial therapeutics. This would, in turn, help in the attenuation of virulence across the pathogenic organism.

1.8.1. Reported RNA modifications in *P. horikoshii* OT3

RNA modifications are a universal occurrence across all studied rRNAs (Piekna-Przybylska et al., 2007; Boccaletto et al., 2018). Nevertheless, their chemical nature, number, and position diversity can vary significantly among archaea (Grosjean et al., 2008; Dennis et al., 2015; Boccaletto et al., 2018; Coureux et al., 2020; Sas-Chen et al., 2020). As discussed above, these modifications can be broadly categorized into three types: base, ribose, and pseudouridine methylation. All the methylated residues in archaea, particularly in *P. horikoshii*, comprise two major groups: stand-alone enzymes, prevalent across all life domains, and RNA-guided modifications that employ RNP complexes to guide and modify the target rRNA in an RNA sequence-dependent manner (Lafontaine and Tollervey, 1998).

Among the various methylated ribonucleotide bases, 5-methylcytosine (m^5C) possesses a methyl group at the fifth position of cytosine. They are ubiquitously present in the domains of archaea. The organism *P. horikoshii* OT3, possesses highest number (five) of putative

RNA m⁵C MTases (ORF ids: PH0851, PH1078, PH1374, PH1537, and PH1991). Out of these five putative RNA m⁵C MTases, PH0851 has been reported as a homolog of the bacterial rRNA (RsmB) (Hikida et al., 2010). Moreover, although the structural characterization of the protein PH1374 shows similarity with the human p120 protein, its physiological substrate is unknown (Ishikawa et al., 2004). The crystal structure of another open reading frame (ORF), PH1915, has been determined, revealing a protein with three distinct domains. Structural homology searches and comparisons have identified two well-characterized domains: the PUA domain and the SAM MTase domain. Additionally, a third domain has been identified, showcasing a novel fold. This structural configuration bears a resemblance to the homolog of *E. coli* rRNA MTase (RlmI) (Sun et al., 2005).

Conversely, the structure of PH1948 closely resembles that of the antibiotic resistance-associated RNA MTase (erythromycin resistance causing MTase C'; ErmC') from *B. subtilis*. The dynamic and conserved loop spanning residues 179-191 in PH1948 exhibits considerable mobility, suggesting a role in binding and orienting the target RNA base, akin to the motion observed in ErmC'. This observation, coupled with a structural comparison to the rRNA (adenine-N6) MTase ErmC', highlights PH1948 as a SAM-dependent MTase. Notably, its N-terminus features a positively charged groove, serving as the likely binding site for the RNA substrate (Gao et al., 2005).

In addition to rRNA, *Pyrococcus* sp. features several tRNA MTases, including the TrmN/Trm14 family, Trm1, and NSUN6 protein. The TrmN/Trm14 exhibits specificity for the modification of the 6th position of guanosine in the tRNA (Phe) acceptor stem, a modification widely observed across the three domains of life. On the other hand, Trm1 and NSUN6 tRNA MTases (ORF id: PH1829 and PH1991), conserved in both eukaryotes and archaea, play a role in generating 2-methylated guanosine (m²G) and 5-methylcytosine (m⁵C) in the core structure of tRNAs, respectively (Nishimoto et al., 2008; Li et al., 2019). Beyond base methylation, a previously characterized RNA-guided modification has been identified in this archaeon, where Trm56 (ORF id: PH0461) catalyzes the 2'-O-methylation of the cytidine residue in archaeal tRNA (Kuratani et al., 2008).

1.8.2. Repertoire of putative RNA MTases in *P. horikoshii* OT3

The gene(s) encoding RNA MTases in major archaeal lineages, including *P. horikoshii* OT3, were identified and collected from GenBank and UniProtKB databases using the keywords “methyltransferase OR RNA methyltransferase OR rRNA methyltransferase OR tRNA methyltransferase”. The amino acid sequences of the identified ORFs were retrieved from the UniProtKB database.

Additionally, an alternative strategy was employed to identify homologs of well-characterized rRNA MTases within the archaeon *P. horikoshii* OT3. This involved a heuristic search analysis conducted across the non-redundant (nr) database, UniProt/SwissProt, and the Protein Data Bank (PDB) using the web tool BLAST (Basic Local Alignment Search Tool; Altschul et al., 1990). These MTases, commonly found in model organisms like *E. coli*, *S. cerevisiae*, and *H. sapiens*, served as references. Through this approach, homologous proteins were identified with a sequence identity (SI) cutoff of over 20%. This relatively lower SI cutoff facilitated the identification of distantly related homologous proteins that still retained similar functional characteristics.

Utilizing insights from literature and curated data across multiple databases, we have compiled a list of potential RNA MTases present in *P. horikoshii* OT3, encompassing both rRNA and tRNA modifications (Table 1.4).

Table 1.4. The list of RNA MTases found in the archaeal model species *P. horikoshii* OT3.

UniProt id	ORF id	Type of substrate	Type of modification	Modified nucleoside	Protein length (aa)
O58805	PH1078	rRNA	Base methylation	m ⁵ C	450
O58581	PH0851				315
O50082	PH1374			m ⁵ C	396
O59578	PH1915			m ₂ ⁶ A	290
O59487	PH1823			m ⁵ U	412
O58994	PH1259				

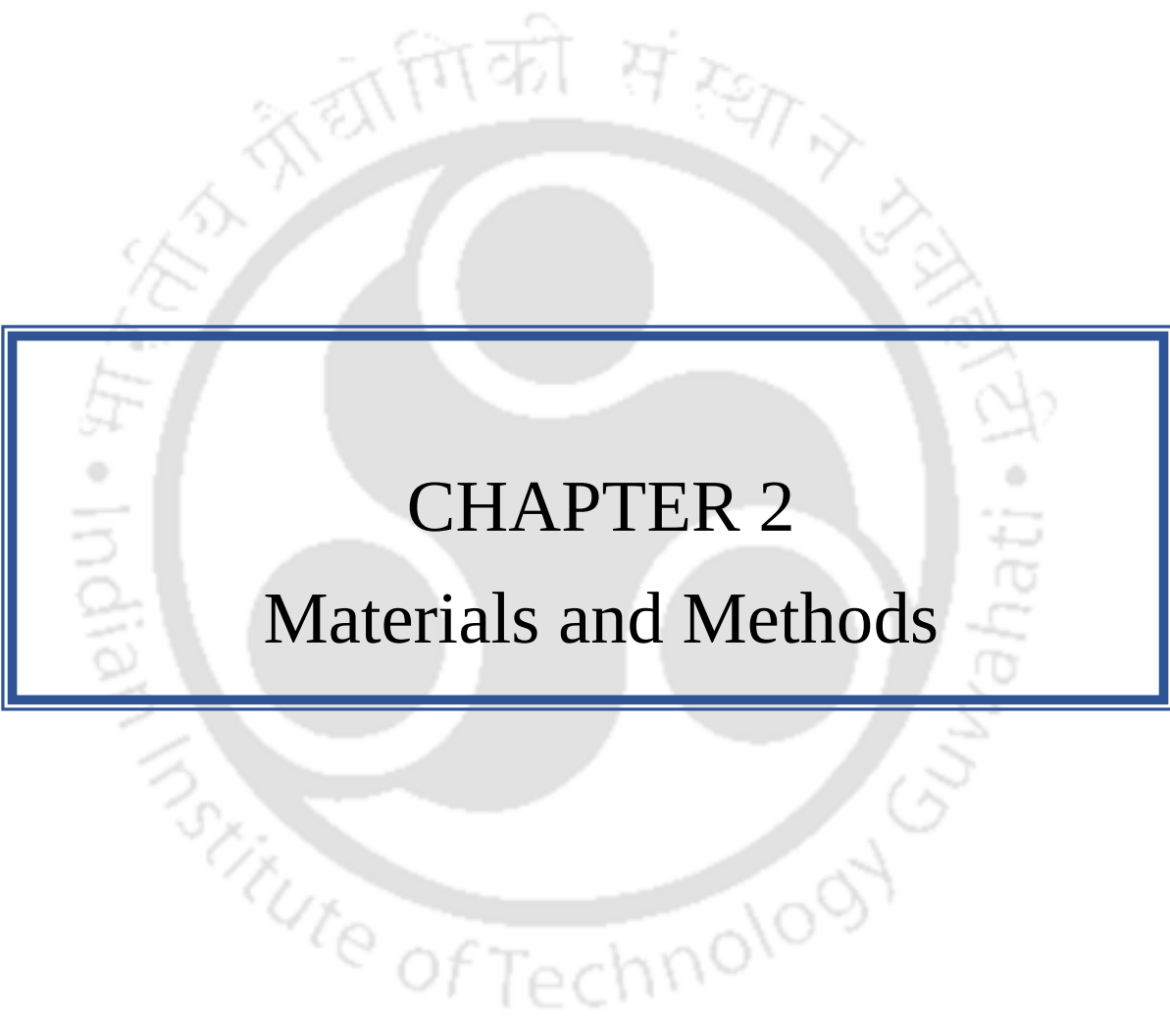
O58004	PH0266			m ⁶ A	247
O59611	PH1948				207
O59359	PH1642			m ² G	198
O50087	PH1379		Methylated pseudouridine	m ¹ ψ	229
O58214	PH0461	tRNA	Ribose methylation	Cytidine-2'-O	203
Q8U248	PH0644		Base methylation	m ² G	365
O59493	PH1829				381
O57712	PH1991			m ⁵ C	389
O59206	PH1537				312
O58864	PH1137			m ⁵ U	407
O50109	PH1401		Methylated pseudouridine	m ¹ ψ	203
O58442	PH0711		Base methylation	m ¹ G	419
O57855	PH0115				296
O50124	PH1416				330
O59562	PH1858		Base methylation	m ¹ A	253

Note: In this thesis, yellow and green highlighted RNA MTases were selected for computational and experimental work, respectively.

1.9. Objectives

To comprehend the selective mechanisms of various RNA MTases for complex substrates, the following objectives were formulated:

- 1) *In silico* characterization of the repertoire of probable m⁵C RNA MTases from archaea, particularly *P. horikoshii* OT3.
- 2) Molecular cloning, overexpression, and purification of m₂⁶A, m⁵U, and m¹ψ rRNA MTases from *P. horikoshii* OT3.
- 3) Crystallization and co-crystallization of the selected rRNA MTases to elucidate their three-dimensional structures and study their substrate-binding mechanism.
- 4) Biochemical and biophysical characterization of the above-mentioned proteins and their complexes.



CHAPTER 2

Materials and Methods

General outline of Chapter 2

This chapter provides a comprehensive overview of the bioinformatics methodology employed to investigate the repertoire of ribosomal RNA (rRNA) methyltransferases (MTases) within the archaeon *P. horikoshii* OT3. Additionally, the chapter delineates the design and methodology of various experiments undertaken for the structural and functional characterization of the chosen rRNA MTases from this particular microorganism. Within this context, Figure 2.1 illustrates a detailed work plan outlining the major experimental procedures, thereby serving as a visual representation of the overall investigative framework.

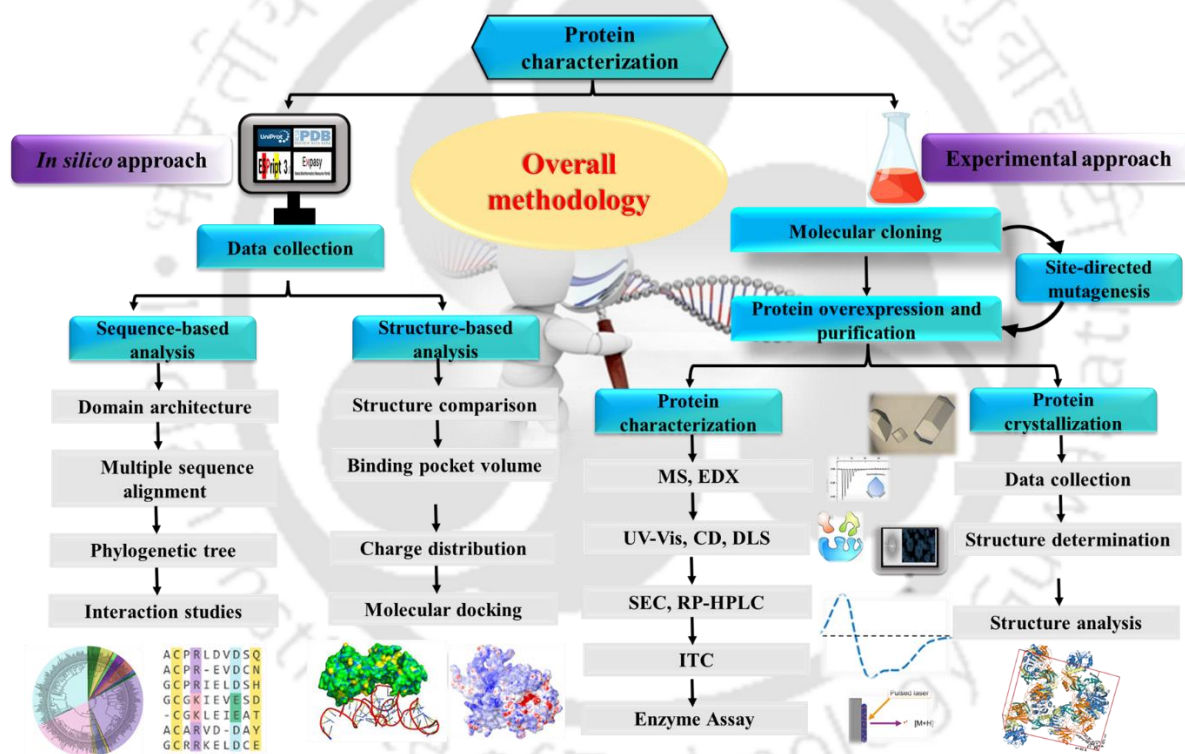


Figure 2.1. A schematic representation that illustrates the comprehensive approach employed to achieve the outlined objectives.

2.1. Materials

2.1.1. Reagents

The enzymes essential for molecular cloning, including restriction enzymes, shrimp alkaline phosphatase, and T4 DNA ligase, were acquired from New England Biolabs (NEB) and/or Thermo Fisher Scientific. Plasmid isolation, PCR clean-up, and gel extraction kits for recombinant cloning were sourced from QIAGEN (Germany), HiMedia (India), and GCC Biotech (India). Site-directed mutagenesis utilized the Q5 site-specific mutagenesis kit from NEB. Ni-NTA affinity resin and pierce centrifuge columns for purifying wild-type and mutant proteins were obtained from QIAGEN (Germany) and Thermo Fisher Scientific (USA), respectively. The Amicon Ultra-15 and Vivaspın Turbo 15 protein concentrators were purchased from Merck (USA) and Sartorius (Germany), respectively. Various crystallization buffers, microbatch and hanging-drop plates, and other protein crystallization-related chemicals and tools were sourced from Hampton Research (USA) and Molecular Dimensions (UK). For the enzyme assay, the MTase-Glo kit and 96-well white plates with a solid bottom were acquired from Promega (USA) and Corning (USA), respectively.

2.1.2. Substrates and cofactors

The various cofactors S-adenosyl-L-methionine (SAM), S-adenosyl-L-homocysteine (SAH), 5-methylthioadenosine (MTA), Adenosine (ADN), β -pseudouridine (PSE) and Sinefungin/adenosyl-L-ornithine (SFG) were obtained from Merck (USA). Meanwhile, various substrates (rRNA oligos) were synthesized at different scales by Integrated DNA Technologies (USA).

2.2. Methods

2.2.1. Sequence- and structure-based analyses

Upon solving the structures of the selected rRNA MTases, further functional characterization ensued through the sequence- and structure-based analysis methods

outlined below. The nucleotide sequences corresponding to each Open Reading Frame (ORF) encoding the selected rRNA MTases were extracted from the KEGG database (Kanehisa and Goto, 2000) and the Gene database of the National Centre for Biotechnology (NCBI) (Benson et al., 2000). Similarly, the respective proteins and their homologous sequences analyzed in this study were sourced from the UniProtKB database (UniProt Consortium, 2023). For comprehensive information on the total number of ORFs encoding potential rRNA MTases in *P. horikoshii* OT3, data collection involved the utilization of the following databases:

- The National Centre for Biotechnology (NCBI) (Benson et al., 2000)
- UniProtKB database (The UniProt Consortium, 2023)
- Structural-Biological Whole-Cell Project
- DDBJ database (Fukuda et al., 2021)

2.2.1.1. Sequence-based analysis

2.2.1.1.1. Identification of homologous proteins and subsequent sequence alignment

The well-characterized rRNA methyltransferases in model organisms such as *E. coli*, *S. cerevisiae*, and *H. sapiens* are employed to identify their homologs in the archaea, specifically, archaeon *P. horikoshii* OT3 through a heuristic search analysis against the non-redundant (nr), UniProt/SwissProt, and Protein Data Bank (PDB) databases, using the web tool BLAST (Basic Local Alignment Search Tool, Altschul et al., 1990). The protein BLAST was configured with specific parameters: max target sequence (100), expected threshold (10), word size (6), matrix (BLOSUM62), and gap costs (Existence: 11, Extension: 1). Post-BLAST search, homologous sequences were discerned based on sequence identity and query coverage. Subsequently, these sequences underwent multiple sequence alignments (MSA) facilitated by the Clustal Omega program (Sievers and Higgins, 2014). Clustal Omega employs seeded guide trees and hidden Markov model (HMM) iterations to generate alignments among multiple sequences, utilizing default parameters. To enhance clarity, the outputs of the MSA were visually represented using the web tool ESPript 3.0 (Gouet et al., 2003).

2.2.1.1.2. Construction of phylogenetic tree

Phylogenetic trees were generated to clarify the evolutionary connections among homologous protein sequences and deduce their shared ancestors. The protein sequences of chosen MTases and their counterparts were utilized to construct trees using the preferred software, MEGA, version X (Molecular Evolution and Genetic Analysis, Kumar et al., 2018). The phylogenetic trees were constructed in MEGAX through the following three steps:

(1) Retrieval of homologous protein sequences, followed by alignment using the ClustalW program embedded in the software MEGAX. Default alignment parameters were utilized (gap opening: 10 and gap extension penalties: 0.2). (2) Tree generation from the aligned sequences using the neighbor-joining (NJ) or Maximum-likelihood (ML) method. The significance of the tree was assessed through the bootstrapping method, providing a confidence level for the similarity between the formed clusters. (3) Refinement of the generated phylogenetic tree to enhance clarity in clustering and to extract relevant information about the evolutionary relationships among the studied proteins. The trees were further visualized in the web tool iTOL v6.3 (Letunic and Bork, 2021).

2.2.1.1.3. Interaction studies

Protein–protein interactions (PPIs) constitute an extensive and intricate network of reactions pivotal for the regulation and execution of diverse biological processes. Proteins usually function in conjunction with other proteins and are regulated. Molecular machines within a cell are often composed of numerous protein components, which are organized based on their protein-protein interactions. To identify the functionally associated proteins of selected MTases, a protein-protein interactome network was generated using the database STRING v11 with a default set of parameters (Szklarczyk et al., 2019). The interactions were subsequently selected based on the cut-off score. Following this selection process, the interactome networks were visually elucidated using Cytoscape 3.8.1 to ensure visual clarity (Shannon et al., 2003).

2.2.1.2. Structure-based analysis

2.2.1.2.1. Structure prediction

In the pursuit of predicting the theoretical three-dimensional structure of MTases, the technique of homology modeling was employed. This method involves constructing a theoretical model for the target protein by leveraging its primary structure (amino acid sequence) and utilizing a crystal structure of a homologous protein as a template. To achieve this, various online three-dimensional structure prediction programs, including Phyre 2.0 (Kelley et al., 2015), SWISS-MODEL (Biasini et al., 2014), RaptorX (Kallberg et al., 2012), I-TASSER (Yang et al., 2007), and more advanced AlphaFold (Jumper et al., 2021) were employed. All of these online programs identified the closest homologous protein as the optimal template for modeling the structures of the selected MTases. Subsequently, the predicted models underwent refinement through energy minimization using GalaxyRefine (Ko et al., 2012). The initially predicted structures and their refined counterparts were meticulously compared based on various geometrical and stereochemical parameters. Validation processes were conducted to ensure the reliability of the models, and those MTase models exhibiting superior geometrical and stereochemical parameters were selected for further structural analysis. This rigorous approach led to the finalization of theoretical models for the MTases, laying the groundwork for comprehensive structural investigations.

2.2.1.2.2. Structural comparison with homologous proteins

The structural homologs of the targeted MTases from the Protein Data Bank (PDB) were identified using the web server DALI (Berman et al., 2000; Holm, 2020). To assess the root mean square deviation (RMSD) among the selected structures and their homologs, calculations were conducted utilizing the 3dSS program (Sumathi et al., 2006) as well as PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

2.2.1.2.3. Structure-based study

In the Protein Data Bank (PDB), an extensive repository of structural information on RNA (rRNA/tRNA) methyltransferases (MTases) is available. Notably, these MTases within the same class exhibit significant structural similarity due to the presence of a conserved

catalytic domain. The distinctions among them arise from variations in their RNA binding domains. To establish the relationship between target MTases and their homologs, a structure-based distance tree was constructed using PyMOL and a custom Python script, which utilized the root-mean-square deviation (RMSD)-based distance matrix of targeted MTases. The RMSD matrix, formatted appropriately, underwent conversion to a Newick (.nwk) file using the dendroUPGMA program (Garcia-Vallvé et al., 1999). Subsequently, the generated .nwk file was input into MEGAX (Kumar et al., 2018) to produce the final structure-based evolutionary tree. This tree was instrumental in elucidating the structural relationships among the RNA binding proteins (RBPs). A cross-verification was carried out to validate these subclusters and discern finer structural features. The structures were superposed in PyMOL, providing a visual confirmation of the distinctions within the identified subclusters. This comprehensive analytical approach provides a robust framework for exploring the structural relationships and evolutionary patterns among MTases, integrating computational tools and visual inspection for a thorough understanding of their molecular architecture.

2.2.1.2.4. Surface analysis

The electrostatic nature of the active-site pockets in the chosen MTases and their homologs was assessed by employing the Adaptive Poisson-Boltzmann Solver (APBS) embedded in PyMOL (Baker et al., 2001). To enhance clarity, the resulting electrostatic potential charge distributions for each protein were colour-coded, with positively charged regions depicted in blue and negatively charged regions in red. The color is scaled from -10 to +10. Additionally, the determination of active site area and volume for both the selected proteins and their homologs was performed using the CASTp program, employing a default probe radius of 1.4 Å (Binkowski et al., 2003).

2.2.2. Molecular docking

The estimation of binding affinities between various cofactors and their complexes with targeted MTases involved the utilization of molecular docking analysis through the freely available AutoDock v4.2 program (Morris et al., 2009). The three-dimensional atomic coordinates of nucleotide/complexes were obtained from their protein-bound structures in

the RCSB Protein Data Bank (Berman et al., 2000). To facilitate cofactor accommodation, the rotameric orientations of active-site residues were adjusted using the program Coot to facilitate the binding in the active-site volume of selected MTases. (Emsley et al., 2010). Before each docking experiment, hydrogen atoms were added to the protein and ligand using AutoDock, and the Gasteiger charge algorithm (Gasteiger and Marsili, 1980) was used to assign partial charges to protein atoms. Blind docking of cofactors with the protein was conducted with a 126x126x126 grid size and 0.375 Å grid spacing centered on the protein's mass center. Rigid molecular docking was implemented with protein atoms kept rigid and cofactor atoms flexible about rotatable bonds. Each docking experiment involved 2000 runs of the Lamarckian genetic algorithm (LGA), and resulting docked ligand conformations were clustered with a RMSD cut-off of 2.0 Å. Coot was used to identify atomic interactions between the protein and docked ligands, and PyMOL was employed to prepared molecular interactions and structural figures.

For protein-protein and Protein-RNA docking, ClusPro, HDOCK, and MPRDock, a freely available online tool (Kozakov et al., 2017; He et al., 2019; Yan et al., 2020), was utilized. One protein molecule was submitted as a receptor, while the ligand-bound protein/RNA was submitted as a ligand, using default parameters. The three-dimensional structure of targeted fragments of substrate RNAs (rRNA/tRNA) was modeled and docked using the web server ModeRNA (Rother et al., 2011). The analysis considered the balanced results, providing optimum clusters, to obtain the plausible protein complex. Atomic interactions between proteins and docked ligands were identified using Coot, and PyMOL was used to prepare molecular interactions and structural figures.

2.2.3. Recombinant construction for protein overexpression

Molecular cloning, a pivotal technique in molecular biology, serves to generate novel recombinant DNA molecules by combining distinct nucleotide sequences. In this study, molecular cloning was achieved to construct a recombinant entity wherein the target rRNA MTase gene was seamlessly incorporated into a suitable vector for subsequent propagation. The fusion of vector and gene occurred in the presence of ligase enzyme, and the resulting construct was introduced into a host organism. This process initiated the replication of the

designed construct, yielding multiple copies of recombinant clones (Figure 2.2). The intricate details of each step involved in molecular cloning within the scope of this study are elucidated below.

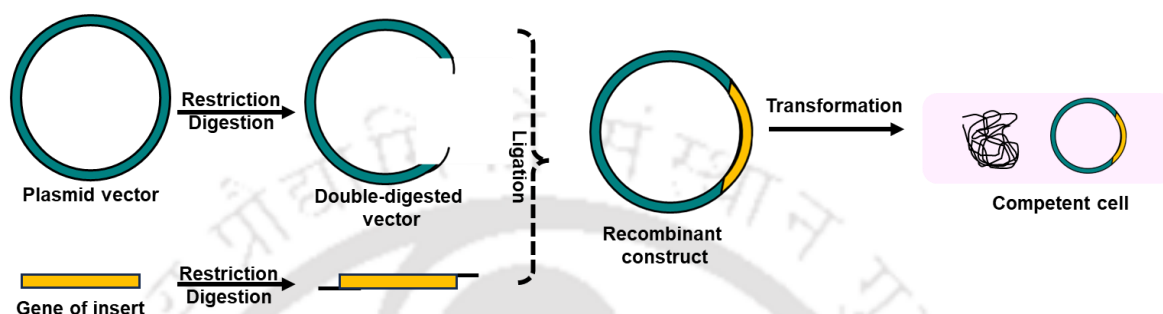


Figure 2.2. Schematic representation for molecular cloning.

2.2.3.1. Competent cell preparation for gene cloning and protein overexpression

The preparation of competent cells for various *Escherichia coli* strains employed the calcium chloride (CaCl_2)-based chemical method in molecular cloning and protein overexpression. Bacterial cells, in their log phase, were rendered competent by resuspending them in a 100 mM CaCl_2 solution, followed by a 30-minute incubation at 4°C. Subsequently, these competent cells were resuspended in an autoclaved mixture of 100 mM CaCl_2 and 15% glycerol, and after flash freezing in liquid nitrogen, they were stored at -80°C. The strains of *E. coli* utilized in this study for protein expression harbored a chromosomal T7 RNA polymerase for controlled protein expression under the influence of a T7 promoter. Induction of targeted protein overexpression was achieved by using Isopropyl β -d-1-thiogalactopyranoside (IPTG), an analog of lactose, which binds to the repressor.

In the pursuit of molecular cloning, protein overexpression, and solubility optimization, the following *E. coli* strains were employed in this study:

- DH5 α competent cells: Exhibiting high transformation efficiency, these cells played a crucial role in elevating the copy number of the cloned plasmid.
- Rosetta (DE3) competent cells: These cells were utilized for overexpressing genes under the regulatory control of a T7 promoter induced by IPTG.

2.2.3.2. Gene amplification

The amplification of the gene of interest involved the design of custom primers, considering parameters such as GC content, melting temperature (T_m), and restriction digestion sites. The OligoAnalyzer tool from IDT Technologies was utilized to scrutinize and validate all primer parameters. Amplification of the target genes was performed via polymerase chain reaction (PCR) using genomic DNA from *P. horikoshii* obtained from the Biological Research Center, NITE (NBRC, Japan), as a template. The PCR reaction encompassed the following steps: 1) initial denaturation (95°C, 4 min), 2) final denaturation (95°C, 30sec), 3) annealing (at the T_m of genes, 45 sec), 4) initial extension (72°C, 40 sec), and 5) final extension (72°C, 10 min) (Figure 2.3). The amplified product utilizes the 6xHis-tag downstream of the ribosome binding site (RBS) in the pET-28a(+) vector.

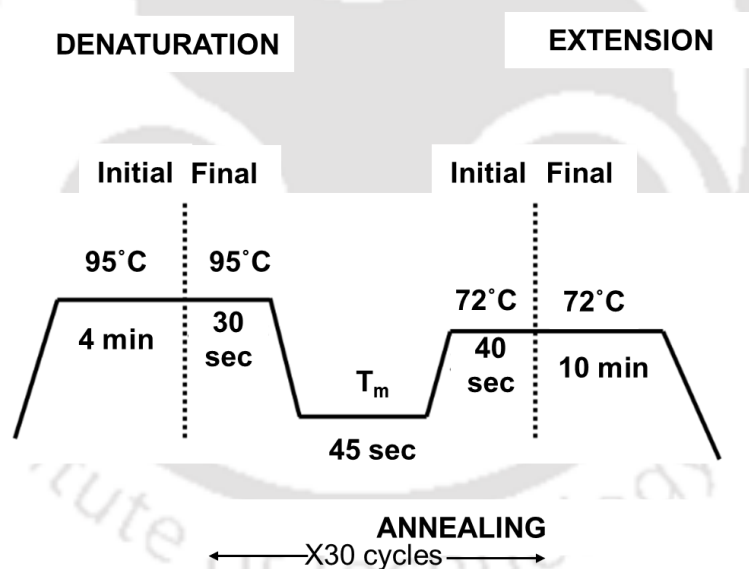


Figure 2.3. A schematic overview of the PCR cycle illustrates its three key stages: denaturation, annealing, and extension. Each accompanied by specific temperature and time parameters. The term " T_m " represents the melting temperature.

2.2.3.3. Double digestion of the gene of interest and vector

To facilitate directional cloning, the amplified genes underwent double digestion using specific restriction enzymes before insertion into the pET-28a(+) plasmid vector. *NdeI*, *NheI*, *EcoRI*, *HindIII*, and *XhoI* were selected as the restriction enzymes, generating 5' and 3' sticky ends in both the genes and the vector. Prior to the restriction digestion process, the genes were treated with the respective restriction enzymes and incubated at 37°C for 3 hours. Concurrently, the pET-28a(+) plasmid vector underwent the same treatment with the selected restriction enzymes and was incubated at 37°C for 3 hours. After the third hour, alkaline phosphatase was introduced to the vector reaction mixture and allowed to incubate for an additional 1 hour at 37°C to eliminate the 5'-phosphate group. Following the incubation period, the double-digested genes and vectors underwent purification using the PCR purification kit (QIAGEN) and gel extraction kit (QIAGEN), respectively.

2.2.3.4. Ligation of the gene of interest and clone confirmation

In the final stage of molecular cloning, the integration of the gene of interest into a pET-28a(+) vector involved a ligation reaction catalyzed by the T4 DNA ligase enzyme. This enzymatic process forms a covalent bond between the double-digested vector and the gene (insert), both possessing sticky ends. Various ratios of the digested vector to gene (1:1, 1:2, 1:3, 1:4, and 1:5) were mixed and allowed to incubate at 22°C for 4 hours in the presence of T4 DNA ligase. Following incubation, each ligation mixture was transformed into *E. coli* DH5 α competent cells to propagate the ligated recombinant clone.

Positive colonies were screened by cultivating transformed cells on antibiotic-supplemented Luria Bertani (LB) agar plates. For further confirmation, fully-grown colonies were inoculated into LB broth supplemented with selective antibiotics, and plasmid isolation was performed. After plasmid isolation, the anticipated cloned plasmids underwent incubation with corresponding restriction enzymes, followed by a 3-hour incubation at 37°C to verify the presence of the gene of interest. Once confirmation was achieved through double digestion, the clones underwent DNA sequencing for additional validation.

2.2.3.5. Site-directed mutagenesis

To achieve the highly flexible apo state of the protein *PhDIMT1* and unravel the crucial residues in its active site responsible for ligand binding, site-directed mutagenesis was conducted using the Q5 site-specific mutagenesis kit (NEB, USA). The selection of the mutagenesis target site was based on their functional role in catalysis as well as their conservation in the bacterial and eukaryotic homologs. Customized oligonucleotide primers, detailed in Chapter 5, were designed to introduce the desired mutations into the wild-type (WT) cloned plasmid for in vitro mutagenesis. The mutagenesis process involves three key steps:

- Amplification of the vector using a High-Fidelity (HF) DNA polymerase from a master mix formulation, with the WT cloned plasmid serving as the template.
- Incubation of the amplified products with specific enzymes, including kinase, ligase, and *DpnI*. This facilitated the circularization of the amplified products and subsequent removal of the template WT DNA.
- Transformation of highly efficient *E. coli* DH5 α competent cells with the circularized PCR product.

The confirmation of the desired point mutations in the target gene was accomplished through plasmid DNA sequencing, and the structural elucidation of the mutant proteins provided valuable insights into the ligand-binding mechanism of *PhDIMT1*.

2.2.4. Protein overexpression, solubilization and purification

2.2.4.1. Overexpression of the selected rRNA MTases

The selected rRNA MTases, cloned into the pET-28a(+) vector, were expressed by transforming *E. coli* Rosetta (DE3) competent cells with the recombinant constructs. To optimize the overexpression of *PhDIMT1*, *PhRlmCD*, and *PhNep1* proteins, various time intervals (ranging from 1 to 7 hours) at 37°C were tested. IPTG was employed as an inducer, and its concentration (0.1, 0.5, and 1.0 mM) was systematically optimized to avoid undesirable outcomes such as cell death or the formation of unproductive proteins.

Upon determining the optimal combination of expression host, temperature, IPTG concentration, and expression time, the solubility of the overexpressed proteins was assessed.

2.2.4.2. Solubility of protein

The targeted overexpressed protein was obtained through centrifugation, followed by resuspension in a lysis buffer for homogenization. To increase the solubility of the targeted proteins, two critical additives, 50 mM L-Arginine and L-Glutamic acid, were added to the lysis buffer. Subsequently, the homogenized suspension underwent sonication for cell disruption, applying 25% amplitude for 15 minutes with a sonication cycle of 3 seconds on and 5 seconds off. The resulting cell lysate of the thermophilic protein *PhDIMT1*, *PhRlmCD*, and *PhNep1* underwent a heat treatment in a boiling water bath at 60°C for 10 minutes to eliminate thermolabile host proteins. The cell lysate was then subjected to centrifugation at a maximum speed of 18894g for 40 minutes to separate the pellet and supernatant fractions. The solubility of the protein was assessed by running both fractions on a 10-12% SDS-PAGE gel. Upon identifying the solubilized protein in the supernatant fraction through SDS-PAGE analysis, it underwent further purification using chromatography techniques.

2.2.4.3. Affinity chromatography for protein purification

The immobilized metal affinity chromatography (IMAC) method was utilized to purify the target protein. To exploit the precise coordination between a transition metal ion (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+}) and polyhistidine residues, a 6xHis-tag was introduced at the C-terminus of the recombinant proteins. This tag facilitated coordination with immobilized Ni^{2+} ions on a matrix. Before purification, Pierce centrifuge columns were filled with Ni-NTA resins and then equilibrated using the appropriate buffers. During the purification process, the prepared Ni-NTA resins were incubated with the supernatant fraction of the protein of interest for approximately 4 hours. Subsequent to binding, the columns underwent washing with buffers containing varying imidazole concentrations (20-50 mM) to eliminate non-specifically bound proteins. In addition to imidazole, the wash buffer contained additives such as salt (NaCl), a reducing agent (β -mercaptoethanol), and a protease inhibitor

cocktails (PIC, HiMedia) to mitigate non-specific binding of host proteins. Elution of the target protein was achieved by using a higher imidazole concentration (250 mM) in the elution buffer. This disrupted the coordination bond between the Ni^{2+} ions in the resin and the histidine residues in the recombinant protein, allowing competitive binding and elution of the tagged proteins. The purity of the eluted fractions was assessed using 10% SDS-PAGE. Subsequently, the purified eluates were combined and subjected to gradient dialysis to remove imidazole completely. For the removal of endogenously bound metal ions in the protein *PhRlmCD*, dialysis was performed against 10 mM EDTA along with the appropriate buffer composition. The dialyzed proteins were concentrated using protein concentrators to achieve the desired concentration for subsequent biophysical and biochemical experiments.

2.2.5. Protein characterization

2.2.5.1. Western blot analysis

Western blotting, also known as immunoblotting, is a powerful laboratory technique used to detect specific proteins within a mixture of proteins. The process involves several sequential steps to effectively identify and visualize the target protein(s). The process begins with the preparation of a protein sample. In this particular case, the purified rRNA MTase (*PhNep1*) was diluted to a concentration of 50 μM in a buffer solution containing 50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol. This ensures that the protein remains stable and soluble throughout the experiment. Subsequently, the protein sample was separated based on molecular weight using SDS-PAGE. After electrophoresis for 2 hours at 100V, the proteins were now separated into distinct bands within the gel. A process known as immunoblotting or western blotting was performed to transfer the separated proteins from the gel onto a membrane. The gel was carefully placed onto a nitrocellulose membrane, and both were sandwiched between layers of filter paper soaked in a transfer buffer. An electric current was then applied, causing the proteins to migrate out of the gel and onto the membrane, where they became immobilized. Following the transfer, the membrane was blocked using skimmed milk to prevent the non-specific binding of antibodies and incubated with a primary antibody specific to the target protein.

In this case, a mouse anti-His₆ antibody was used at a dilution of 1:3000. The primary antibody binds specifically to the target protein (in this case, the His₆-tagged) on the membrane. After washing away any unbound primary antibody, the membrane was then incubated with a secondary antibody conjugated to a detection molecule, such as horseradish peroxidase (HRP). In this case, rabbit anti-mouse IgG antibody was used at a dilution of 1:5000. The secondary antibody binds to the primary antibody that was already bound to the target protein, followed by washing to remove the non-specific secondary antibody. Finally, the presence of the target protein was detected by adding a chemiluminescent substrate that reacts with the HRP enzyme to produce light. The emitted light was captured using specialized imaging equipment, such as a chemiluminescence detector, and the resulting image was analyzed using Image Lab software (Bio-Rad) to visualize the target protein.

2.2.5.2. RNA analysis on native-PAGE

To analyse RNA folding, the lyophilized rRNA substrate was dissolved in nuclease-free water to create a 1 μ M concentration. The dissolved RNA samples were then placed on a native 6% (19:1) polyacrylamide gel in TBE (Tris-Borate-EDTA) buffer at 25°C. The purpose of this procedure was to obtain a single band that would reflect the RNA's average structure based on its mobility within the gel.

2.2.5.3. Mass spectrometric (MS) analysis

The precise molecular weights of the proteins were determined through matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry. The selectively purified rRNA MTases were uniformly mixed with the matrix at protein-to-matrix ratios of 1:2, 1:3, and 1:5. Due to the theoretical molecular weight of the targeted rRNA MTases exceeding 10 kDa, sinapinic acid was utilized for protein analysis. The matrix was prepared by dissolving sinapinic acid in acetonitrile and 0.1% trifluoroacetic acid, maintaining a ratio of 30:70, resulting in a concentration of 10 mg mL⁻¹. To eliminate excess matrix crystals, the solution was bath sonicated for 30 minutes and centrifuged at 12000xg for 10 minutes. Subsequently, the prepared matrix was blended with protein samples at concentrations of 10, 15, and 25 μ M, adhering to the specified ratios. Spots of

the protein-matrix mixture (2 μ L) were deposited onto the grooves of a ground steel MALDI plate and allowed to air-dry. Following drying, the droplets were ionized using a Bruker autoflex speed (Bruker Daltonics, Germany), and the resulting data were analyzed using the integrated Flex control analysis software.

2.2.5.4. Energy dispersive X-ray (EDX) analysis

EDX microanalysis of the chosen protein was conducted using an 80T EDX detector (Oxford Instruments, Tokyo, Japan) installed in the TEM-Jeol 2100F transmission electron microscopy at the Central Instruments Facility (CIF) of the Indian Institute of Technology Guwahati, India. The EDX detector, featuring a super atmospheric thin window (SATW) and a 50 mm² silicon drift detector (SDD) at a 30° take-off angle, provided a resolution of 129 eV at the Mn K-L3 line. The INCA monitoring system (version 5.05, Oxford Instruments) was employed for EDX spectral processing. The software utilized a filtered least squares (FLS) approach for background removal, fitting, and peak deconvolution, including pulse pile-up correction. Quantitative data obtained after matrix correction using the XPP exponential model procedure were focused on the S/In ratio. The analysis involved indicating a 50 nm carbon layer coating on the sample, selecting all elements except sulfur and indium as "deconvolution elements," and using the "Quant Standardization (Extended Set)" provided by the software. S/N ratios were also calculated using the same procedure.

2.2.5.5. UV-visible absorbance spectroscopy

Absorbance spectra spanning from 200 to 700 nm were meticulously recorded to evaluate the presence of the Fe-S cluster of the selected protein. This analysis was performed using a double-beam LMSP-UV1900S spectrophotometer (LABMAN, India). The absorption peaks observed at around 325 and 410 nm served as indicators of iron-sulfur content in the selected protein. The identified peaks were generated by the subtraction of the spectra of the control buffer from the experimental data, thereby confirming the presence of the Fe-S cluster within the protein.

2.2.5.6. Circular dichroism (CD) spectroscopy

CD spectroscopy stands as a widely utilized biophysical technique for both protein characterization and the analysis of thermal stability. Within a wavelength range spanning 260 to 190 nm with a 1 mm optical path-length quartz cuvette using a circular dichroism spectrometer, JASCO J-1500 (JASCO, Japan) offers insights into the secondary structural elements of proteins, including α -helices, parallel and anti-parallel β -sheets, turns, and random coils. This study specifically focused on monitoring the secondary structural content and unfolding thermodynamics of the chosen proteins. Changes in the secondary structural content, particularly α -helices and β -sheets, were examined at wavelengths approximately 220 nm and 195 nm, respectively, across a temperature range. The resulting thermodynamic profile of protein unfolding with respect to temperature facilitated the calculation of the melting temperature (T_m) for the targeted rRNA MTases. The thermal stability of these selected proteins was systematically assessed over a temperature range spanning from 20 to 120°C.

2.2.5.7. Dynamic light scattering (DLS) analysis

The evaluation of the aggregation propensity of the targeted proteins involved the estimation of hydrodynamic diameters through dynamic light scattering (LITESIZER 500; Anton Paar, UK). A dilution of the purified protein sample was prepared in nuclease-free water (Himedia, India) to achieve a final concentration of approximately 1 μ M. Subsequently, the DLS spectra of the protein sample were generated by incrementally increasing the temperature from 0.1°C to 10°C in a stepwise fashion. The collection of spectra occurred over a predefined time interval within a temperature range spanning 20-70°C, utilizing a cuvette with a path length of 1 cm. In the final presentation, a two-dimensional graph illustrating the hydrodynamic diameter as a function of temperature, with respect to various wavelengths, was generated using the program Origin v9.6.

2.2.5.8. Fluorescence spectroscopy (FS) analysis

The pivotal role of aromatic amino acids, particularly tryptophan and tyrosine, in the active site pocket of *PhDIMT1* is notably enriched with aromatic residues. To evaluate the interaction between the protein and cofactors, an intrinsic protein fluorescence assay was conducted using a FluoroMax-4 fluorimeter (Horiba Scientific, Japan). Excitation of

samples at a fixed wavelength of 285 nm resulted in the emission of fluorescence signals spanning the range of 310-450 nm. The alterations in the fluorescence signals were closely tied to the changes in the protein conformation due to the binding or release of the cofactor. These changes were particularly evident in the local environment of aromatic residues within the protein. In this study, variations in the environment of these residues, as indicated by fluorescence patterns, were monitored in the presence and absence of various cofactors.

2.2.5.9. Size exclusion chromatography (SEC) analysis

Size exclusion chromatography (SEC) functions by exploiting the size exclusion effect of porous gels packed within a column to separate solutes of varying sizes. Typically conducted in peak mode, SEC involves the separation of solutes with different molecular sizes from a pulse of feed. As these solutes progress through the column, they exit separately, yielding distinct peaks. To determine the approximate molecular weight of targeted rRNA MTases, the elution volumes of protein with their corresponding molecular weights (M_w) are obtained by plotting “ K_{av} ” vs. “ $\log (M_w)$ ”. The distribution coefficient (K_{av}) is unique for each protein and determined by the following equation.

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

Where V_e denotes the column volume at which the targeted protein is eluted, while V_t encompasses the total volume both between and within the beads, and V_0 corresponds to the void volume. Utilizing the data from standard proteins in size exclusion chromatography (SEC) facilitates the determination of the molecular weight of the protein of interest by using A HiLoad™ 26/600 Superdex™ 200 pg column (GE Healthcare) was used for size exclusion chromatography (SEC) analysis on an AKTA pure chromatography system (Cytiva, Sweden).

2.2.5.10. Reversed-phase high-performance liquid chromatography (RP-HPLC) analysis

RP-HPLC, a liquid chromatography mode employing a non-polar stationary phase and polar mobile phases, served as the method for monitoring the time-dependent cofactor stability assay. The detection of organic compounds relies on alterations in analyte solubility in the mobile phase or changes in their binding to the stationary phase. To optimize the separation, various detection wavelengths in the UV range (203–365 nm) were explored. For this study, each sample was injected into a C18 column (YMC, Tokyo) pre-equilibrated with 0.1% trifluoroacetic acid (TFA) and run on the high-performance liquid chromatograph (Shimadzu-DGU-20A5R, Japan). A linear gradient of 0-5% acetonitrile (ACN) in 0.1% TFA was applied over 4 minutes at a flow rate of 1 mL/min, followed by maintaining an isocratic flow of 5% ACN in 0.1% TFA for 18 minutes. Standard samples of freshly prepared cofactors were then detected by absorbance at 254 nm. This comprehensive approach ensured efficient separation and precise monitoring of the analytes in the time-dependent assay.

2.2.5.11. Isothermal titration calorimetry (ITC) analysis

The study of protein-ligand interactions involves measuring the thermodynamic parameters using isothermal titration calorimetry (ITC). ITC is a biophysical method that provides information on stoichiometry (n), change in enthalpy (ΔH), change in entropy (ΔS), and equilibrium association constant (K_a) for protein-ligand binding. In this study, an incremental titration using MicroCal ITC200 (GE Healthcare, USA) was performed by adding a fixed volume of ligand to the protein solution at discrete time intervals. Incremental titration leads to a change in heat due to exothermic (heat released) or endothermic (heat absorbed) reactions taking place between the protein and ligand. The change in heat produces a peak, which is calculated by determining the area under the peak using the peak integration method. After integration, the data is normalized and fitted to one binding site model to calculate the stoichiometric ratio and change in affinity and enthalpy for the interaction. A control experiment is performed for each ITC experiment by titrating the ligand into a buffer solution in order to determine the heat of dilution. The heat of dilution obtained from the control experiment is subtracted from the reaction to get the actual heat of dilution associated with protein-ligand interaction. The subtracted heat change is further used to plot the graph against the protein-to-ligand molar ratio. The Gibbs

free energy for each binding experiment is calculated using the expression: $\Delta G = \Delta H - T\Delta S$, where ΔH is the change in enthalpy, T is the temperature of the adiabatic system, and ΔS is the change in the entropy for the reaction. This systematic approach provided comprehensive insights into the energetics and dynamics governing the studied protein-ligand interactions.

2.2.5.12. Enzyme assay

The MTase-Glo assay (Promega, USA) employs bioluminescence to detect the product S-adenosylhomocysteine (SAH) via a luciferase reaction, measured with the Tecan microplate reader (Tecan, Switzerland). In a 50 μL reaction mixture comprised of assay buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl_2 , 0.1 mg mL^{-1} BSA, and 1 mM DTT), a fixed amount of substrate rRNA, and cofactor, increasing quantities of the selected rRNA MTase are added. After a 30-minute incubation at 37°C, enzymatic activity (EC_{50}) is calculated using the formula:

$$Y = \text{Min} + \frac{\text{Max} - \text{Min}}{1 + \left(\frac{X}{\text{EC}_{50}}\right)^{\text{Hill coefficient}}}$$

Here, Y represents the difference in luminosity (net relative luminescence unit, RLU) compared to the control (no-enzyme). Max and Min denote the highest and lowest enzyme activity at a constant substrate concentration, the Hill coefficient indicates the slope of the curve, and X is the increasing enzyme concentration. Once the optimal enzyme concentration is determined, the Michaelis-Menten constant (K_M) of the substrate is assessed at a fixed cofactor concentration (10 μM) using the equation:

$$V_o = \frac{V_{\text{max}} [S]}{K_M + [S]}$$

Here, V_o and V_{max} represent the initial and maximum reaction velocities, and $[S]$ is the substrate concentration (μM). The reaction is halted using TFA (0.5%), and the V_{max} (RLU min^{-1}) is determined by dividing the RLU measured at the end of the reaction by the incubation time (30 min).

2.2.6. Crystallization of the selected rRNA MTases

This study employed X-ray crystallography as a method to elucidate the three-dimensional structures of the chosen rRNA MTases. To achieve high-quality protein crystals suitable for X-ray crystallography, a comprehensive screening process was conducted on the target protein solution using various crystallization buffers. This screening aimed to initiate crystal formation and involved multiple optimizations. Protein crystallization follows a thermodynamic progression with four zones: undersaturated, metastable, labile, and precipitation. Crystal formation occurs in the supersaturated stage within the metastable zone (Figure 2.4). Two primary methods were employed to reach the supersaturated state in protein crystallization: (1) vapor diffusion, using the hanging-drop method, and (2) microbatch-under-oil.

In the hanging-drop vapor-diffusion method, a drop containing a mixture of the protein solution and precipitant was positioned on a coverslip and equilibrated against the reservoir solution in an inverted position. Conversely, the microbatch-under-oil method involves submerging a mixture of the protein and precipitant under oil to slow down evaporation. In this study, a combination of paraffin and silicon oils with different densities was utilized to control diffusion rates. To obtain protein crystals, the solubility of the protein was assessed at two temperatures (4 and 20°C) using both hanging-drop and microbatch-under-oil methods. Following the identification of initial crystal hits, further optimization was carried out by separating the two crucial events in crystallization-nucleation and crystal growth-through various optimization techniques, including microseeding. In microseeding, submicroscopic crystals were generated and employed as seeds to introduce equilibrated new drops, enhancing the crystallization process.

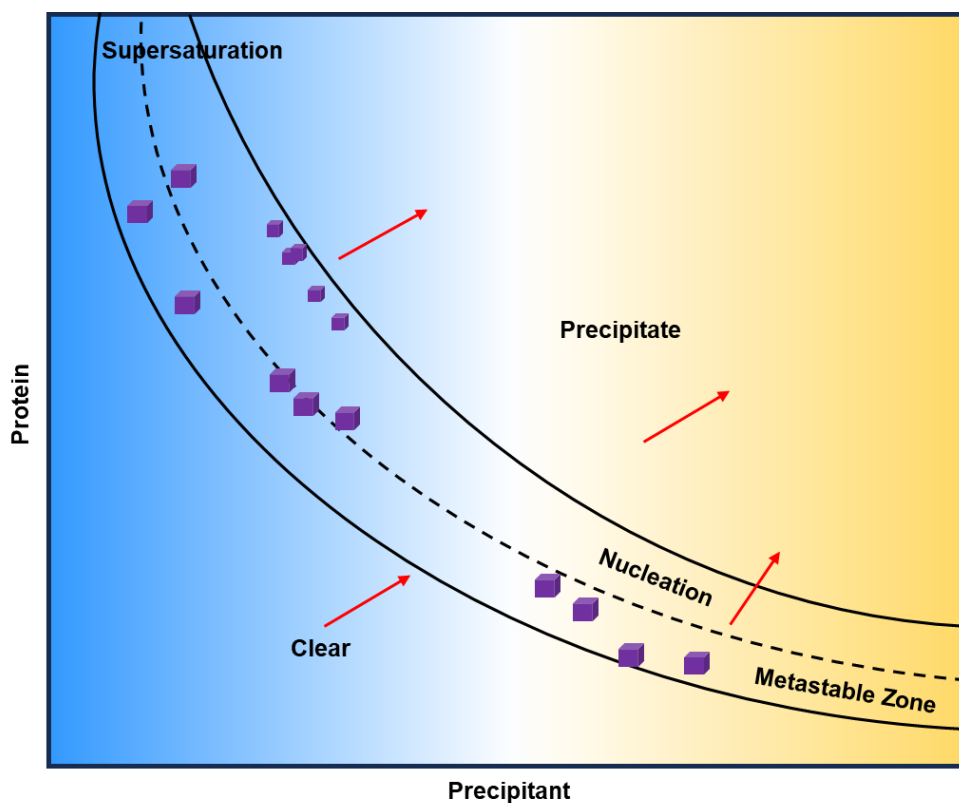


Figure 2.4. A visual depiction illustrating the phase diagram of protein crystallization.
(Image redrawn from Stewart et al., 2011)

2.2.7. Data collection and processing

2.2.7.1. Data collection

After obtaining a suitable protein crystal, it was exposed to a monochromatic X-ray beam to produce diffraction patterns. As the crystal contains a periodic arrangement of atoms, X-ray scattering occurs either constructively or destructively. Bragg's law defines the relationship between constructive interference and the crystal's geometry parameters. According to Bragg's law, the X-ray scattering depends on the incident beam wavelength (λ), the glancing angle (2θ), and a positive integer (n).

$$2d \sin\theta = n\lambda$$

For this study, all X-ray intensity data was collected at a fixed wavelength (λ) of the incident X-ray beam (1.5418 Å) and recorded on an image plate (IP) detector. To collect data with all possible diffraction peaks, the diffracted crystals' oscillation angle was kept at 1.0° (1.0 degree per image) during data collection. Depending on the crystal's space group, the desired number of images or frames were collected for the complete data set.

Since the data was collected at a fixed wavelength (λ), other parameters such as the crystal-to-detector distance (d) in a range of 150 - 210 mm and exposure time (300 s) were considered to improve the data collection process with reasonable statistics.

2.2.7.2. Data processing

During X-ray data processing, the raw diffraction data is transformed into Bragg reflections with measured intensities, which involves four steps, including calibration, integration, merging, and scaling. Different algorithms are available in software packages for each step. In this study, the iMosflm program was used for initial processing, while Aimless was used for scaling and merging. The program iMosflm is a part of the CCP4 suite and performs indexing to provide an index for each reflection and initial statistics on unit cell dimensions in a crystallographic symmetry. The program Aimless was used to calculate the possible point and space groups after data sets were subjected to scaling and merging. Finally, the obtained intensities were converted into structure factors using the module ctruncate, which is distributed in the CCP4 package.

2.2.7.3. Structure determination

The X-ray diffraction method provides information about structure amplitudes but not the phase angles, which are essential to calculate the electron density. To obtain the initial phase information for structural solutions of selected rRNA MTase, various methods are used, including molecular replacement (MR), single- and multiple-wavelength anomalous dispersion (SAD and MAD), and Molecular isomorphous replacement (MIR). In this study, molecular replacement was used to obtain the phase information for the structural solution of the selected rRNA MTase. The MR method uses a known molecular structure of the homologous protein to solve the structure of a new target protein using the program Phaser. A significant sequence similarity (>25%) between the known homologous protein and the

targeted unknown protein is required to obtain the initial phase by MR. Phaser utilizes multivariate statistics and maximum likelihood probability theory. Finally, the phase information obtained from the model structure was used to calculate the phases from the target protein intensities.

2.2.7.4. Model building and structure refinement

After obtaining the initial phase angles from the reflections of the target protein, a map of the electron density was created. This map helped in developing a model structure by tracing and modeling the secondary structural elements such as α -helices and β -sheets. The program used for this purpose was Coot, which stands for Crystallographic Object-Oriented Toolkit (Emsley et al., 2010). During the process of building the model, several chemical and physical parameters, including bond lengths, bond angles, torsion angles, and atomic overlaps, were analyzed using programs like PDB Goodies (Hussain et al., 2002) and PSAP (Protein Structure Analysis Package, Balamurugan et al., 2007).

In the current study, model refinement was carried out using the program Refmac5 (Vagin, 2004), which utilizes Bayesian statistics and maximum likelihood probability. The purpose of model refinement is to enhance the agreement between the calculated structure factors from the model and the observed structure factors from the experiment. After each refinement cycle, an agreement between the calculated structure factors (F_{calc}) was compared with that of the measured structure factors (F_{obs}) and can be represented as:

$$R = \Sigma(F_{\text{obs}} - F_{\text{calc}}) / (\Sigma F_{\text{obs}})$$

After refinement, the R-factor values and electron-density maps ($2F_o - F_c$ and $F_o - F_c$) were inspected and validated. Maps were contoured at 1.0σ and 3.0σ , respectively.

2.2.7.5. Cross-validation (R_{free} calculation)

Statistical cross-validation was used to evaluate the quality of the model using the R-factor (Brunger, 1992). The diffraction data were first split into two sets for cross-validation. A small complimentary test set of 5-10% of the diffraction data was removed from the complete data set, while the remaining 90-95% of the data formed the large working set

used during the refinement process. The accuracy of the predicted model was evaluated using the remaining 10% test data set by calculating the R-free value:

$$R_{free} = \frac{V \sum ||\mathbf{F}_{obs}| - k|\mathbf{F}_{calc}||_{max} [S]}{\sum |\mathbf{F}_{obs}|_{\mathbf{hkl} \in T}}$$

where $\mathbf{hkl} \in T$ represents all reflections belonging to test set T of unique reflections.

2.2.8. Structure validation

The accuracy of the three-dimensional structures of rRNA MTases was assessed through a process of structure validation. Generally, this process comprises three aspects: 1) the validity of the experimental measurements, 2) consistency in the atomic model, and 3) consistency of the chemical and physical properties of the atomic model. This study considered the following aspects for validating the structures of the selected proteins: (1) geometrical parameters and (2) root mean square deviation (RMSD) of the refined structures. Several online web servers and tools are available for structure validation, and in this study, PROCHECK (Laskowski et al., 1993) and MolProbity (Williams et al., 2018) were used for the selected rRNA MTases.

2.2.8.1. PROCHECK

After each cycle of refinement, the quality of the model structure was evaluated using PROCHECK (Laskowski et al., 1993). This program assesses the geometry of the model structure residue by residue. PROCHECK provides several pieces of information, including the Ramachandran plot, Gly & Pro Ramachandran and Chi1-Chi2 plots, main-chain and side-chain parameters, residue properties, main-chain bond length and angle distributions, and RMS distances from planarity and distorted geometry plots. The Ramachandran plot was used to estimate the geometry of the selected rRNA MTases.

2.2.8.2. MolProbity

MolProbity serves as a web-based tool dedicated to the validation of protein structures (Williams et al., 2018). Its functionality encompasses the comprehensive assessment of three-dimensional protein structures, addressing both global and local aspects. When applied to a given crystal structure, MolProbity undergoes optimization processes for polar and non-polar hydrogens, conducts thorough analyses of all-atom contacts, evaluates covalent geometry, and assesses torsion angles within the protein. Furthermore, MolProbity contributes to refining the local chemical environment by addressing structural resolution issues. This involves the identification of Ramachandran outliers and the correction of flipped-branched protein side chains. In the current investigation, MolProbity played a crucial role in detecting outliers related to Ramachandran angles and torsion angles within the studied protein structures.

2.2.9. Domain dynamics analysis of rRNA MTases

To examine the alterations in conformation occurring in the chosen rRNA MTases upon interaction with ligands, we employed the online tool DynDom (Hayward and Berendsen, 1998). The DynDom server facilitates the analysis of protein domain movements by computing both rotational angles and translational shifts between different conformations. It offers insights into the specific residues participating in hinge axes and the bending of hinges, particularly in the transition between the "open" and "closed" conformations of the rRNA MTases.

2.2.10. Structure and protein-ligand interaction visualization

The study utilized PyMOL) program to visualize the 3D structures of proteins and locate the active site of known and modeled structures by comparing them. Additionally, the program was used to create molecular graphics figures. Further, Coot was used to analyze the intermolecular interaction(s) between the cognate and docked ligand and the rRNA MTases (Emsley et al., 2010).

CHAPTER 3

Study of archaeal m^5C RNA MTases

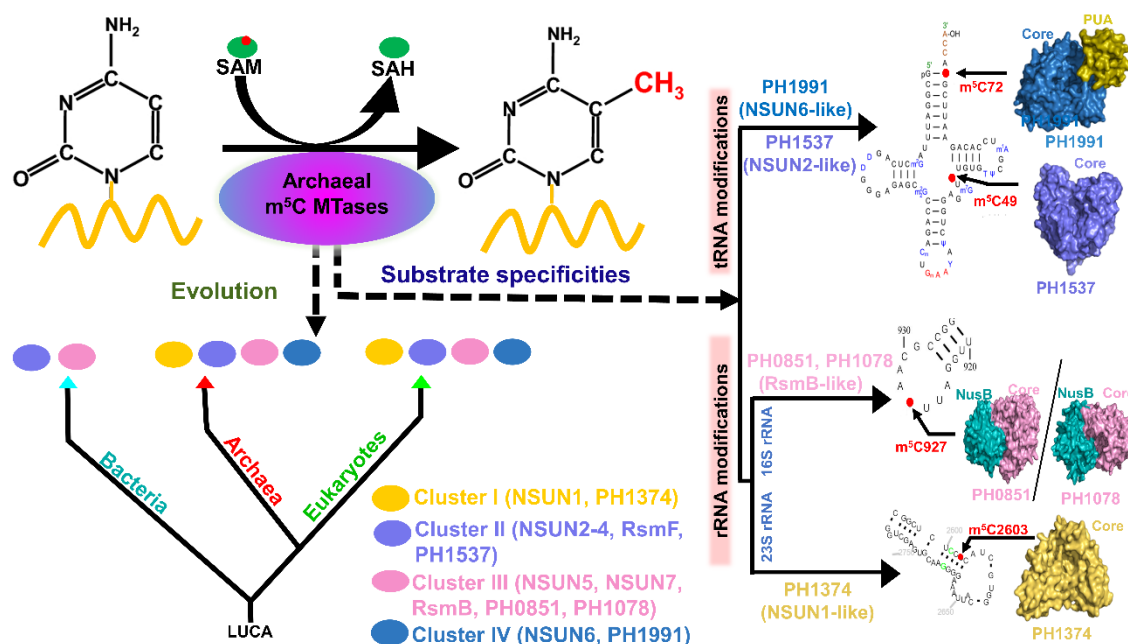
This chapter has been published as:

Saha, S. Mandal, S.K. and Kanaujia, S.P. (2024). Distinct characteristics of putative archaeal 5-methylcytosine RNA methyltransferases unveil their substrate specificities and evolutionary ancestries. *J. Biomol. Struct. Dyn.*, 1-18.

Abstract

5-methylcytosine methyltransferases (m⁵C MTases) are known to be involved in the modification of RNA. Although these enzymes have been relatively well characterized in bacteria and eukarya, a complete understanding of the archaeal counterparts is lacking. In this study, the identification and characterization of archaeal RNA m⁵C MTases were performed. As a case study, a hyperthermophilic archaeon, *Pyrococcus horikoshii* OT3, which possesses five putative RNA m⁵C MTases, was chosen. Among the five putative RNA m⁵C MTases, two proteins (PH0851 and PH1991) have been characterized as homologs of a bacterial rRNA MTase (RsmB) and eukaryal tRNA Mtase (NSUN6), respectively. The in-depth characterization of the remaining three putative RNA m⁵C MTases (PH1078, PH1374, and PH1537) in this study suggests the presence of the signature architecture and catalytic residues plausibly involved in the binding of their cognate RNA substrates. Additionally, the results also suggest the existence of two RsmB-like proteins (PH0851 and PH1078) belonging to the same subfamily IV of m⁵C RNA MTase. However, the proteins PH1374 and PH1537 belong to the same subfamily V but bind to different substrates, rRNA and tRNA, respectively. The findings further indicate that archaeal RNA m⁵C MTases link those from bacteria and eukarya.

Graphical abstract



3.1. Introduction

Ribosome, a protein synthesis machinery, is primarily made up of ribosomal RNAs (rRNAs) and proteins that decode mRNAs at the decoding center (DC). During ribosome biogenesis, various cellular processes, such as transcription of rRNA, pre-rRNA processing, chemical modification of rRNA, and simultaneous assembly of ribosomal proteins, play crucial roles (Nierhaus, 1991). The presence of chemically modified rRNA in the vicinity of the decoding and peptidyl transferase centers (DC and PTC, respectively) plays a crucial role in altering the active sites of the ribosome, stabilization of the rRNA scaffold, and contributing to the fine-tuning of translation regulation (Sergiev et al., 2018). Several antibiotics are known to target DC and PTC. Thus, pathogens are found to modify rRNA and/or tRNA near the antibiotics targeting sites as one of the strategies to be resistant to these antibiotics (Bhujbalrao and Anand, 2019). Primarily, there are three types of RNA modification, viz. base methylation, pseudouridylation (methylated or non-

methyated), and ribose methylation (Decatur and Fournier, 2002). Base methylation occurs in specific nucleobase of RNA without the help of small RNA molecules (snoRNA), unlike the other two RNA methylation wherein snoRNA aids the process (Kiss-László et al., 1996; Ni et al., 1997). Among the various methylated ribonucleotide bases, 5-methylcytosine (m⁵C) possesses a methyl group at the fifth position of cytosine. The m⁵C methylation is known to be crucial for RNA quality control, RNA stability, and translation accuracy (Trixl and Lusser, 2019). The m⁵C methylation is carried out by m⁵C methyltransferases. The m⁵C methyltransferases are capable of methylating both DNA and RNA molecules and contain Rossmann fold. They are ubiquitously present in all three domains of life, i.e., bacteria, eukarya, and archaea.

The m⁵C MTases are comprised of evolutionarily conserved motifs I to X, involved in the transfer of methyl group from S-adenosyl-L-methionine (SAM or AdoMet) to the substrate (DNA/RNA) (Bujnicki et al., 2004). The RNA and DNA m⁵C MTases utilize different nucleophilic cysteine residues in motifs VI and IV, respectively. Although both substrates possess a cysteine residue in motif IV, they participate in subtly different ways during catalysis. In RNA m⁵C MTases, the β -elimination (deprotonation) is initiated by the conserved catalytic cysteine (motif IV), while in DNA m⁵C MTases, this phenomenon is initiated by an aspartate residue (near the motif IV) or a water molecule oriented towards the active site of the enzyme. Thus, the two-cysteine mechanism of m⁵C modification is the hallmark of the RNA m⁵C MTases except for DNMT2 and RlmI (Kuznetsova et al., 2019). Moreover, in RNA m⁵C MTases, motif X is located near the N-terminal domain in contrast to DNA m⁵C MTases, where motif X is located at the C-terminus.

Notably, in eukaryotes and archaea, the RNA m⁵C modification is found majorly on rRNA and tRNA and, in a few cases, on mRNA, whereas in bacteria, it is confined only to rRNA (Edelheit et al., 2013). In eukaryotes, the RNA m⁵C MTases belong to the NSUN (NOL1/NOP2/sun domain) family; in humans, seven different RNA m⁵C MTases, viz. NSUN1-7 are known (Bohnsack et al., 2019). The biological function of NSUN subfamilies broadly includes cell differentiation, embryonic development, and regulation (Chi and Delgado-Olguín, 2013). The NSUN1, NSUN2, and NSUN5 subfamilies are

ubiquitously present in most eukaryotes, while other NSUN MTases are found only in higher eukaryotes. In humans, among the seven NSUN proteins, the NSUN1, NSUN4, and NSUN5 are specific to rRNA, whereas NSUN2, NSUN3, and NSUN6 methylate tRNA molecules. In addition, NSUN7 has been involved in modifying short, non-coding RNA, enhancer RNA (eRNA) linked to transcription regulation (Bohnsack et al., 2019). In bacteria, three RNA m⁵C MTases termed RsmB (or Fmu), RsmF (or YebU), and RlmI (or YccW) have been reported (Foster et al., 2003; Hallberg et al., 2006; Sunita et al., 2008). The protein RsmB methylates a free 16S rRNA molecule at the cytosine 967th position (C967), while RsmF modifies C1407 of 16S rRNA bound to the 30S subunit (Foster et al., 2003; Hallberg et al., 2006). Although the protein RlmI is reported to be an RNA m⁵C MTase, it lacks the additional cysteine residue, a hallmark of the two-cysteine mechanism followed by rRNA m⁵C MTases (Kuznetsova et al., 2019).

In general, the RNA m⁵C MTases are known to possess a ~270 amino acid residues long core conserved domain. In addition to the core domain, based on the presence of an RNA-binding domain either at the N-terminal extension (NTE) or C-terminal extension (CTE), RNA m⁵C MTases have been classified into six conserved subfamilies, I or Fmu, II or p120, III or YebU, IV, V, and VI. Among these, the members of subfamilies I and III comprise bacterial RNA m⁵C MTases, while the eukaryotic RNA m⁵C MTases belong to subfamily II (Reid et al., 1999). Compared to bacteria, archaea, being a prokaryote, also possess a relatively large number (up to five) of RNA m⁵C MTases, akin to eukaryotes (Li et al., 2019). For instance, *Pyrococcus horikoshii* OT3, a hyperthermophilic archaeon, possesses five putative RNA m⁵C MTases (ORF ids: PH0851, PH1078, PH1374, PH1537 and PH1991). Out of these five putative RNA m⁵C MTases, two proteins (PH0851 and PH1991) have been reported as homologs of the bacterial RsmB and eukaryotic NSUN6, respectively (Hikida et al., 2010; Li et al., 2019). Moreover, although the structural characterization of the protein PH1374 shows similarity with the human p120 protein, its physiological substrate is unknown (Ishikawa et al., 2004). The remaining two putative RNA m⁵C MTases (PH1078 and PH1537) remain uncharacterized.

Thus, in this study, a systematic *in silico* approach was employed to characterize the remaining three putative RNA m⁵C MTases (PH1078, PH1374, and PH1537) for a complete understanding of the archaeal m⁵C modification(s). This adequate information on the archaeal m⁵C MTases will help in their classification to compare and contrast their (probable) function with their respective homologs. Moreover, the results of this study are expected to establish the missing evolutionary link between prokaryotes and eukaryotes.

3.2. Methodology

3.2.1. Data collection and sequence analysis

The gene(s) encoding RNA m⁵C methyltransferases in major archaeal lineages, including *P. horikoshii* OT3, were identified and collected from GenBank and UniProtKB databases using the keywords “methyltransferase OR RNA m⁵C methyltransferase OR rRNA methyltransferase OR tRNA methyltransferase” (Clark et al., 2016; The UniProt Consortium, 2021). The amino acid sequences of the identified ORFs were retrieved from the UniProtKB database. Their homologous proteins were identified using the web tool BLAST, keeping a sequence identity (SI) cut-off of more than 20%; a lower sequence identity cut-off was used to identify the distantly related homologous proteins having similar functions (Altschul et al., 1990). From these, a total of 14 sequences were selected for heat map analysis. Multiple sequence alignment (MSA) of the targeted proteins with their homologs was performed using the program Clustal Omega (Sievers and Higgins, 2014). The MSAs were further rendered using the web tool ESPript 3.0, and the conserved regions in the sequences were identified using the web tool ConSurf (Gouet et al., 2003; Ashkenazy et al., 2016). The protein-protein interaction networks for the targeted proteins were generated using the STRING v11.0 database with a default set of parameters, keeping the top ten interactors as results (Szklarczyk et al. 2019). Subsequently, the interactome networks were further rendered using the program Cytoscape 3.8.1 for visual clarity (Shannon et al., 2003). The hydrophobicity of the NTE and CTE regions of selected RNA m⁵C MTases was determined using ProtScale (Gasteiger et al., 2005). The bar graphs signifying the percentage of residue conservation in the N-terminal extension (NTE) were calculated, and all the graphs were represented using the software OriginLab (OriginLab

Corporation, USA). The operonic arrangements of genes were located in the Gene database of the National Center for Biotechnology Information (NCBI) and were further verified using the web server Operon-mapper (Taboada et al., 2018). The evolutionary relationships of the targeted proteins with their homologs were established by constructing phylogenetic trees employing the Neighbor-Joining (NJ) method embedded in the program MEGAX with 1000 bootstrap replicates (Kumar et al., 2018). The trees were further visualized in the web tool iTOL v6.3 (Letunic and Bork, 2021). The Pfam and InterPro databases were used to identify and extract the domain arrangement of the rRNA MTases (El-Gebali et al., 2019; Mitchell et al., 2019).

3.2.2. Structure analysis

The three-dimensional structure of the proteins, wherever required, was modeled using the programs RaptorX, SWISS-MODEL, Phyre2, I-TASSER, and Alphafold (Källberg et al., 2012; Biasini et al., 2014; Kelley et al., 2015; Yang et al., 2015; Jumper et al., 2021). The obtained models were further refined using the program GalaxyRefine (Ko et al., 2012). The refined three-dimensional structure models were validated for their stereo-chemical attributes using the webserver SAVES v5.0 (Kumar et al., 2019). The structure-based homology search for the proteins was performed using the web server DALI (Holm, 2020). The root-mean-square deviation (RMSD) among the proteins was calculated using the web tool 3dSS (Sumathi et al., 2006). Subsequently, these pair-wise RMSD values were used to build a distance matrix in order to generate a structure-based evolutionary tree using the web tool DendroUPGMA (Garcia-Vallvé et al., 1999) and iTOL v6.3 (Letunic and Bork, 2021). The active-site pocket area and volume of the targeted RNA m⁵C MTases and their homologs were calculated using the program CASTp with a default probe radius of 1.4 Å (Binkowski et al., 2003). The electrostatic potential charge distribution of the targeted proteins and their homologs was computed using the PDB2PQR module embedded in the APBS tool (Baker et al., 2001) of the program PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC). All the molecular graphics were prepared using the program PyMOL. The packing geometries of the NTE and CTE regions of selected RNA m⁵C MTases were determined using the packing entropy webserver (Khade and Jernigan, 2022).

3.2.3. Molecular docking

The binding affinities of different putative cognate ligands with their respective RNA m⁵C MTases were estimated by molecular docking analysis performed using a freely available program, AutoDock v4.2 (Morris et al., 2009). The three-dimensional atomic coordinates of the ligands were extracted from their protein-bound structures available in the Protein Data Bank (PDB) and PubChem (Berman et al., 2000; Kim et al., 2021). Before each docking experiment, the protein and ligands were prepared using the program Coot (Emsley et al., 2010). The polar hydrogen atoms were added, followed by assigning partial charges to the targeted proteins using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). The grid size of 126x126x126 with a grid spacing of 0.375 Å was generated to perform the blind docking of the targeted proteins with their probable ligand molecules, SAM and cytidine 5'-monophosphate (hereafter referred to as cytosine). The rigid molecular docking, keeping the rotatable bonds of the ligand flexible, was performed for 2000 runs using the Lamarckian genetic algorithm (LGA). Upon docking, the ligand conformations were clustered with an RMSD cut-off of 2.0 Å. The atomic interactions between the protein and the docked ligand(s) were identified using the program Coot. Protein-protein docking calculations were performed using the web server ClusPro 2.0 (Kozakov et al., 2017). Out of the 30 clusters obtained as results, the cluster with the highest number of similar conformations of the protein complex and the lowest weighted score was selected for further analysis. The interface residues between the docked complex of the proteins were identified using the web server PISA (Krissinel and Henrick, 2007). The molecular interactions were identified using the program Coot and the structural figures used in this study were prepared using the program PyMOL.

3.2.4. hdANM (hinge-domain anisotropic network model) analysis

The dynamic community identifier (DCI) approach was used to identify dynamic communities in RNA m⁵C MTases, allowing specific protein regions to move as a cohesive unit using a DCI server (Kumar et al., 2022). The optimal dynamic community distribution for a given protein is determined by selecting the highest CH score. Subsequently, the domain dynamics of this class of proteins were tracked using an elastic network model with default parameters via the hdANM server (Khade et al., 2021).

3.3. Results

3.3.1. Repertoire of putative RNA m⁵C in *P. horikoshii* OT3

A search for putative RNA m⁵C methyltransferases in *P. horikoshii* OT3 in different databases using various keywords resulted in the identification of five ORFs, PH0851, PH1078, PH1374, PH1537, and PH1991. However, the heuristic approach revealed that these five genes are not ubiquitously found across the four major archaeal lineages, viz. Euryarchaea, TACK, DPANN, and Asgard. Despite the substantial differences in gene distribution and lack of specific common trends, the study indicated that the highly diverse group *euryarchaea* possesses the highest number of m⁵C MTases (five), specifically the class of *Thermococci*, which consist of three genera, *Thermococcus*, *Palaeococcus* and *Pyrococcus* (Figure 3.1 and Table A.1).

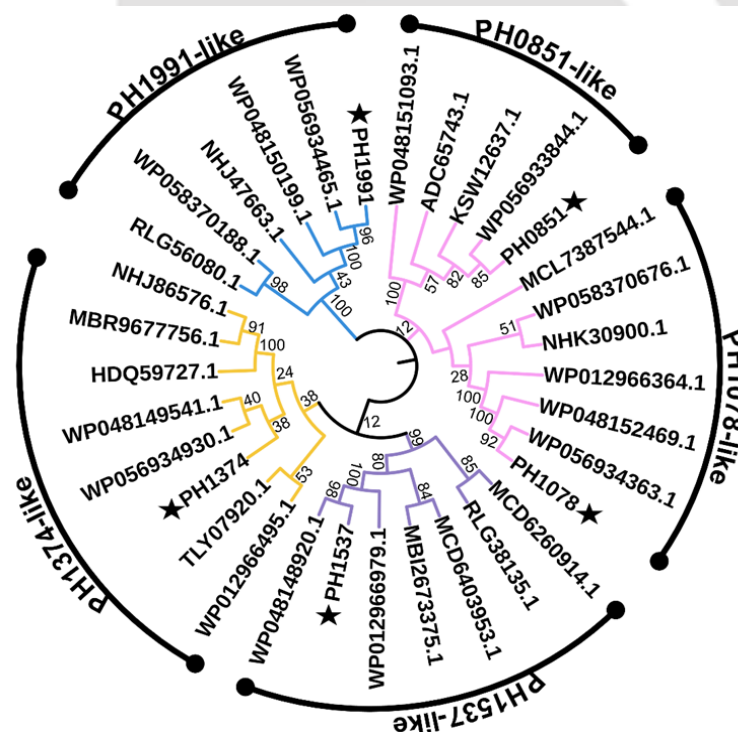


Figure 3.1. Distribution of RNA m⁵C MTases across the four major archaeal lineages. The tree was generated using thirty-three RNA m⁵C MTases, and details are provided in Table A.1. The analysis shows that archaeal RNA m⁵C MTases belong to five distinct

clades with reference to five ORFs of *P. horikoshii* OT3. RNA m⁵C MTases of *P. horikoshii* are marked with a black star.

In *P. horikoshii*, two proteins, PH0851 and PH1991, have previously been structurally characterized (Hikida et al., 2010; Li et al., 2019). The other three ORFs, PH1078, PH1374, and PH1537, have been annotated as “hypothetical fmu protein”, “hypothetical proliferating-cell nucleolar protein p120”, and “hypothetical nucleolar protein”, respectively, in the UniProtKB database. A homology search analysis of these hypothetical proteins suggests their similarity to RNA m⁵C MTases. The protein PH1078 shows homology with bacterial RsmBs (or Fmus) having sequence identity (query coverage) in the range of 24-29 (71-98)%. Similarly, the protein PH1374 shows similarity with eukaryotic NSUN1 proteins (QC: 63-95% and SI: 26-42%) and bacterial RsmFs (QC: 86-88% and SI: 36-41%). Interestingly, the homology search result suggests that the protein PH1537 shows the highest sequence similarity with a tRNA m⁵C MTase PAB1947 from *P. abyssi* (QC: 100% and SI: 87%) and bacterial RsmFs (QC: 87-92% and SI: 37-42%). Thus, to further confirm the proteins PH1078, PH1374, and PH1537 are RNA m⁵C MTases, their signature features at the sequence level were identified and compared by performing a multiple sequence alignment (MSA). The result exhibits the presence of two conserved cysteine residues in motifs IV and VI in all three archaeal proteins, indicating their possible two-cysteine mechanism for modification of cytosine base in RNA (Figure 3.2). Interestingly, motif X in all three archaeal proteins (PH1078, PH1374, and PH1537) is present at the N-terminal, suggesting them to be RNA m⁵C MTases. Moreover, two of the amino acid residues, lysine (Lys273 in PH0851, Lys275 in PH1078, Lys132 in PH1374 and Lys124 in PH1537) and aspartate (Asp337 in PH0851, Asp336 in PH1078, Asp194 in PH1374 and Asp186 in PH1537), located in the motifs I and IV, respectively, reported to be critical for catalysis or proper conformation of the active-site region (Bujnicki et al., 2004), are also found to be conserved (Figure 3.2).

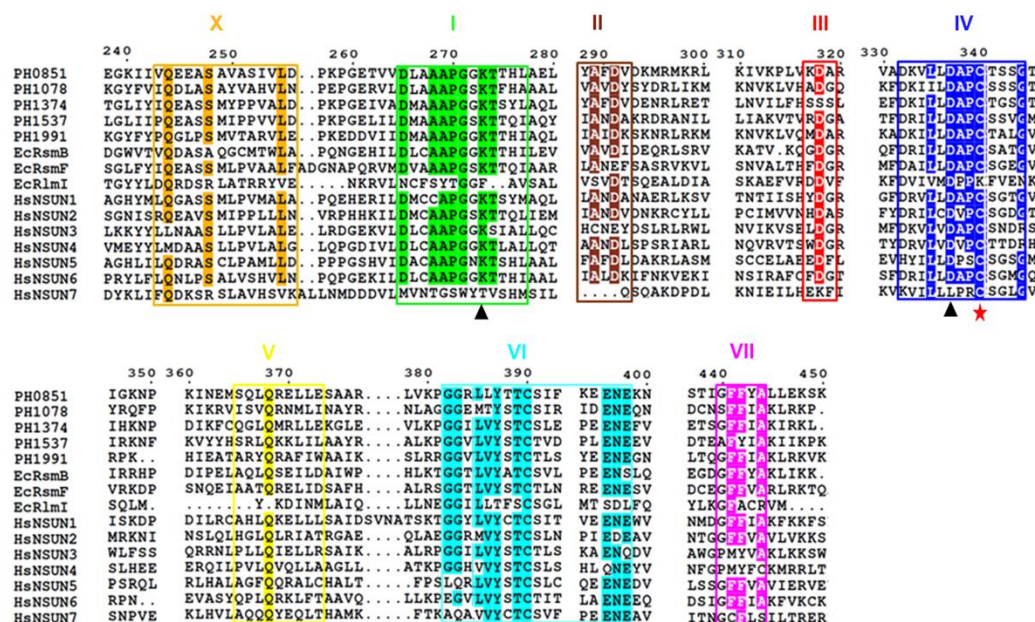


Figure 3.2. Multiple sequence alignment of the conserved active-site region of all the major RNA m⁵C MTases, along with those from *P. horikoshii*. The proteins used in MSA are PH0851 (O58581), PH1078 (O58805), PH1374 (O50082), PH1537 (O59206) and PH1991 (O57712) from *P. horikoshii*; RsmB (P36929), RsmF (P76273), RlmI (P75876) from *E. coli*; NSUN1 (P46087), NSUN2 (Q08J23), NSUN3 (Q9H649), NSUN4 (Q96CB9), NSUN5 (Q96P11), NSUN6 (Q8TEA1) and NSUN7 (Q8NE18) from *H. sapiens*; the UniProt id of each protein is provided in the parenthesis. The motifs X, I, II, III, IV, V, VI, and VII are boxed in orange, green, brown, red, blue, yellow, cyan, and magenta, respectively. The catalytic and two conserved cysteine residues are marked with an upward black arrowhead and red asterisk, respectively.

3.3.2. The archaeal RNA m⁵C MTases belong to the subfamilies IV to VI

As the three putative archaeal RNA m⁵C MTases (PH1078, PH1374, and PH1537) exhibit similarities with the bacterial (RsmB and RsmF from *E. coli*) and the eukaryotic (NSUN1-7 from *H. sapiens*) RNA m⁵C proteins, their domain conservation was analyzed. Pair-wise sequence comparison of the core domain of the RNA m⁵C MTases suggests that the core domain of the five archaeal MTases (PH0851, PH1078, PH1374, PH1537, and PH1991) are highly conserved among themselves and with that of the bacterial RsmB and RsmF proteins. Interestingly, all of them, except PH1991, show high conservation with that of the

eukaryal NSUN1 protein as well. The proteins PH0851, PH1078, and PH1374 possess a conserved core domain with that of the prokaryotic RNA m⁵C MTases, including the bacterial RsmB and RsmF proteins. However, the protein PH1537 does not show a high similarity with PH1991 and RsmB. Notably, the protein PH1991 shows core-domain conservation only with PH0851, PH1078, PH1374, and NSUN6 proteins (Figure A.1 and Table A.2). Further, to identify the closely associated proteins for the three uncharacterized archaeal RNA m⁵C MTases (PH1078, PH1374, and PH1537), their full-length sequences were compared. The result suggests that the protein PH1078 shows the highest similarity with bacterial RsmB (QC: 86%, SI: 26%) (Table A.2). The protein PH1374 shares a high similarity with the bacterial RsmB and RsmF and eukaryal NSUN1 and NSUN5. Similarly, the protein PH1537 exhibits closeness to bacterial RsmF and eukaryal NSUN1, NSUN3 and NSUN5 (Figure 3.3. and Table A.2).

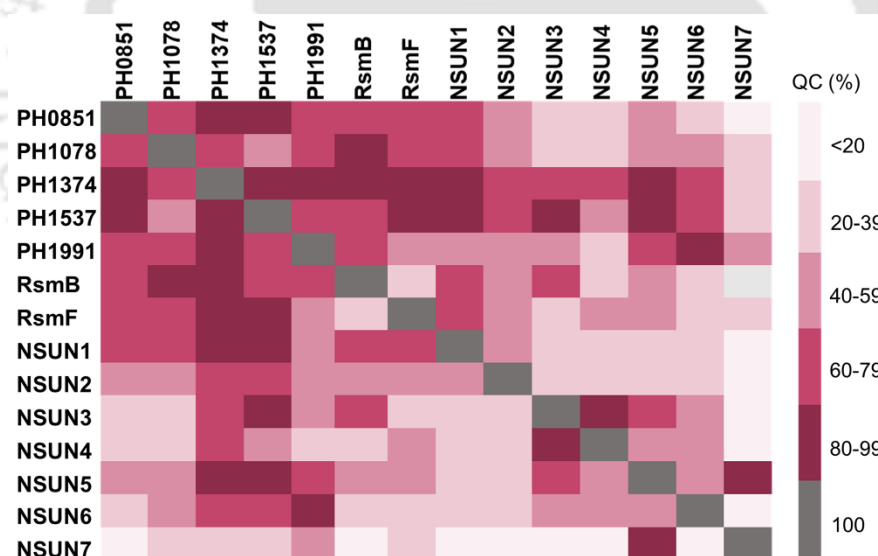
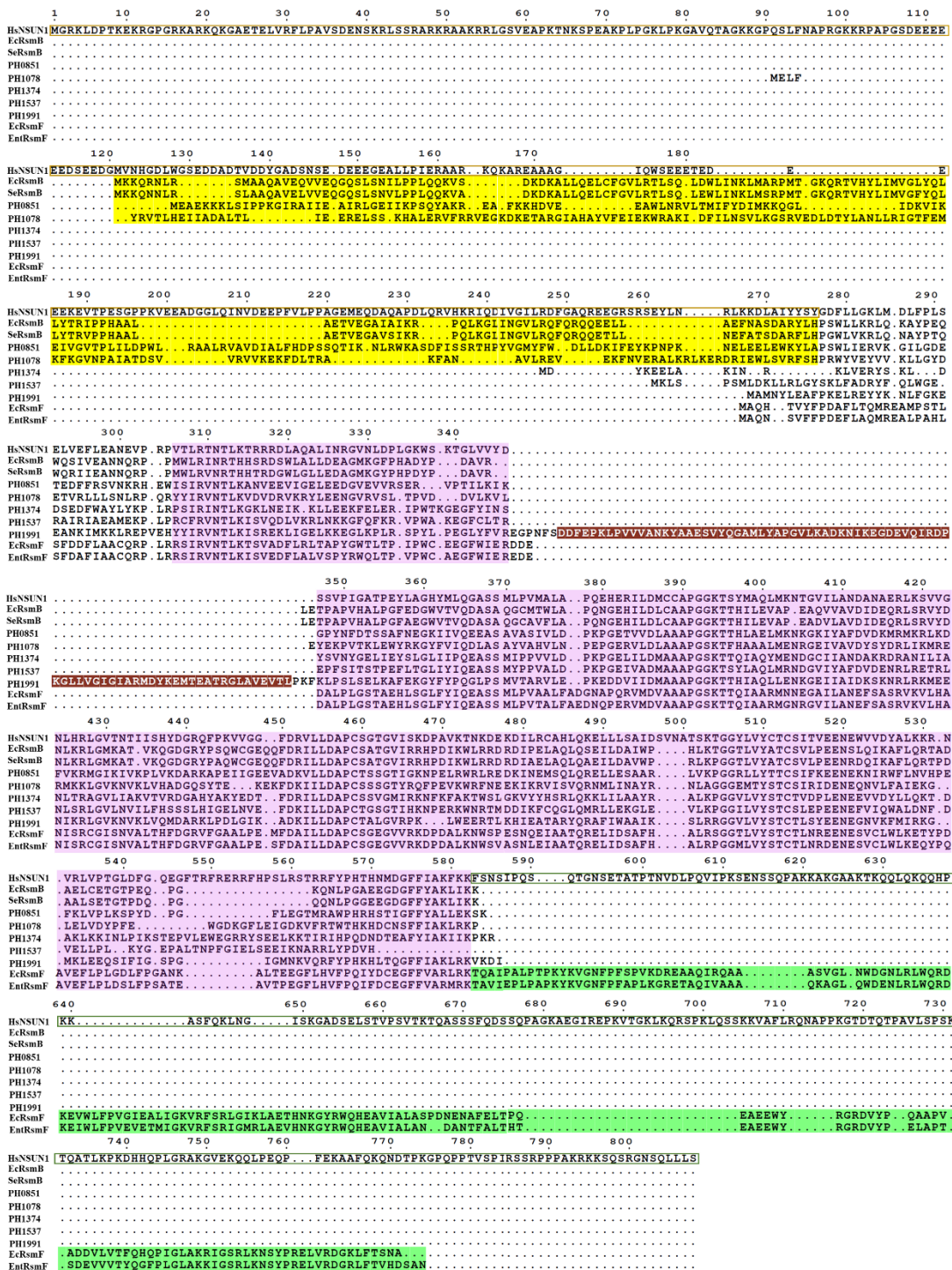


Figure 3.3. The full-length pair-wise sequence comparison of RNA m⁵C MTases from archaea, bacteria, and eukaryotes. The similarity between the pairs is depicted by the heat map. An increasing gradient of red shows protein sequences with increasing similarity.

Although the core domain of most RNA m⁵C MTases is conserved, they differ significantly due to the variation in their signature motif composition. RNA m⁵C family members have been categorized into six (I-VI) subfamilies depending upon the presence or absence of N-terminal extension (NTE), C-terminal extension (CTE), and insertion into the core sequence (Reid et al., 1999). For instance, the protein RsmB (or Fmu) contains an NTE, while RsmF (or YebU) possesses a CTE, involved in RNA binding, belonging to the subfamily I and III, respectively (Reid et al., 1999; Foster et al., 2003; Hallberg et al., 2006). Notably, unlike bacterial proteins, the eukaryotic protein NSUN1 possesses both NTE and CTE, is significantly longer reported to be involved in nucleolar localization, and belongs to the subfamily II (Reid et al., 1999; Kuznetsova et al., 2019).

To determine the subfamilies to which the archaeal RNA m⁵C MTases possibly belong, an MSA of known subfamilies was performed and compared with the targeted proteins. The result suggests that the proteins PH0851 (1-157) and PH1078 (1-159), akin to *Ec*RsmB (1-145) and *Se*RsmB (1-145), possess the NTE and lack the CTE (Figure 3.4). However, the bacterial and archaeal RsmBs are grouped into the subfamilies I and IV, respectively (Reid et al., 1999). Thus, the archaeal proteins PH0851 and PH1078 can be suggested to belong to subfamily IV. Notably, the other two archaeal proteins, PH1374 and PH1537, do not contain either of the extensions (i.e., NTE or CTE) except the core domain (divided into a small N1 region of about 306-346 amino acids and a larger MTase region of about 347–582 amino acids), hinting its closeness to the subfamily V. In addition, members of the subfamily VI possess an unusual insertion (between the motif N1 and X) referred to as the “PUA” domain in the core domain of RNA m⁵C MTases, that recognizes the RNA in the acceptor and D-stem of tRNA regions (Hikida et al., 2010). The MSA results reveal the presence of the “PUA” domain in the protein PH1991, indicating it to be a member of the subfamily VI (Figure 3.4). The result indicates that out of the five archaeal m⁵C RNA MTases, only PH0851 and PH1078 possess the NTE, akin to RsmBs. The proteins PH1374, PH1537, and PH1991 do not possess any extension. However, the protein PH1991 has an insertion sequence (PUA domain) between N1 (306-346) and motif X of the MTase region (347-582). The bacterial RsmFs only possess CTE.

Chapter 3. Study of archaeal m⁵C RNA MTases



horikoshii, O57712), *EcRsmF* (*E. coli*, P76273) and *EntRsmF* (*Enterobacter* sp, A4WBJ4); the organism name and UniProt id of each protein are provided in the parenthesis. The NTE, the core domain, and CTE are highlighted in yellow, pink, and green, respectively. The eukaryotic protein NSUN1 possesses both NTE (1-276) and CTE (583-808), indicated by dark yellow and green rectangular boxes. Moreover, the insertion between motifs N1 and X, referred to as “PUA” domain, is highlighted in brown.

3.3.3. RsmB-like archaeal rRNA m⁵C MTases

3.3.3.1. The interaction profile of the protein PH1078 involves similar interaction as *EcRsmB* and PH0851

The preliminary sequence analysis suggests that the archaeal proteins PH0851 and PH1078 belong to the archaeal subfamily IV and are close to bacterial subfamily I. Since the protein PH0851 has previously been suggested to function as a bacterial RsmB, the presence of additional RsmB (PH1078) in *P. horikoshii* OT3 requires further explanation (Hikida et al., 2010). Thus, to understand the exact role of the protein PH1078, a protein-protein interaction network analysis was performed and compared with that of the proteins PH0851 and its closest homolog, *EcRsmB*. The results exhibit the interaction of *EcRsmB* with the proteins *EcRsmA* (an rRNA m⁶₂A MTase or RsmA), *EcRlmC* (an rRNA m⁵U MTase), and *EcTruB* (tRNA pseudouridine synthase). Further, the interactome map analysis revealed the interaction of *EcRsmB* with tRNA (Met) cytidine acetyltransferase (TmcA), known to be involved in protein translation (Figure 3.5A). Moreover, the network analysis also reveals that a potassium transporter (TrkA), which interacts with *EcRsmB*, has been reported to regulate the gene *rsmB*.

Akin to *EcRsmB* and *EcRlmC*, both the archaeal proteins PH0851 and PH1078 show interaction with its archaeal ancestor rRNA m⁵U MTase protein PH1259, an archaeal homolog of *EcRlmC* (Figure 3.5A-3.5B). Interestingly, the interactome map analysis suggests that the proteins PH0851 and PH1078 also interact with PH1823, an archaeal

RsmA (Figure 3.5B-3.5C). The network analysis revealed that the two archaeal proteins, PH0851 and PH1078, also interact with PH1644, an archaeal homolog of *Ec*TruB. Also, similar to *Ec*RsmB, proteins PH0851 and PH1078 show interaction with PH0788, an archaeal tRNA(Met) cytidine acetyltransferase, TmcA (Figure 3.5B-3.5C). However, in contrast to *Ec*RsmB, the archaeal proteins PH0851 and PH1078 do not exhibit any interaction with PH1984, an archaeal homolog of TrkA from *P. horikoshii*. In summary, the protein PH1078 shares an almost identical protein-protein interaction network similar to that of the proteins PH0851 and *Ec*RsmB (Figure 3.5). Although the aforementioned interaction requires further experimental validation in order to filter out any false positive interactions.

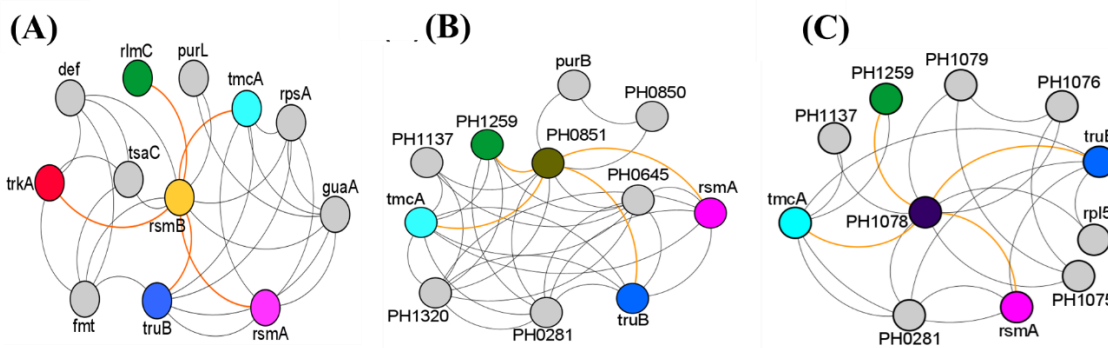


Figure 3.5. The protein-protein interaction network involving PH1078 and its homologs *Ec*RsmB and PH0851. The protein (A) *Ec*RsmB (yellow), (B) PH0851 (olive), and (C) PH1078 (violet) with a cut-off score of 7. The targeted proteins *Ec*RsmB (yellow), PH0851(Olive) and PH1078 (Violet) have been involved in common interaction with RsmA (magenta), RlmC/PH1259 (green), TmcA (cyan), TruB (blue) except only protein *Ec*RsmB interact with protein TrkA (red) and the protein-protein interactions of the targeted proteins are denoted as solid orange lines, while those of other proteins (grey) as solid grey lines. The other details of each interacting protein are provided in Table A.3.

3.3.3.2. The active-site pocket of the protein PH1078 is archetypical to *EcRsmB* and PH0851 to accommodate the cognate substrate

Earlier reports have suggested that the proteins *EcRsmB* and PH0851 bind to similar rRNA molecules (Hikida et al., 2010). Interestingly, sequence analysis indicates the similarity of PH1078 to *EcRsmB* and PH0851, indicating its role in m⁵C modification. In the process of m⁵C modification, the methyl group is donated by the SAM molecule, which acts as a cognate ligand for these MTases (Foster et al., 2003). To affirm the potential substrate(s) for the protein PH1078, its electrostatic potential charge distribution was computed and compared with that of PH0851 and *EcRsmB*. The results exhibit that the active site of the protein *EcRsmB* is made up of both positively and negatively charged amino acid residues. In *EcRsmB*, SAM binds to a negatively charged surface (Figure 3.6A). On the other hand, in PH0851, the SAM molecule binds near a positively charged surface (Hikida et al., 2010). Interestingly, akin to *EcRsmB*, the protein PH1078 also possesses a small patch of negatively charged surface in the vicinity of the active-site pocket, suggesting a possible binding site for the SAM molecule (Figure 3.6A).

To reaffirm the structural similarity of the protein PH1078 with *EcRsmB* and PH0851, their active-site area and volume were compared. The active-site area and volume of the protein PH1078 (1814 Å² and 2230 Å³) are comparable to its closest homolog *EcRsmB* (1393 Å² and 1987 Å³) and PH0851 (1682 Å² and 1222 Å³), suggesting a similar-sized rRNA binding to these proteins (Figure 3.6B). To further substantiate and estimate the binding of SAM and a cytosine molecule with the protein PH1078, molecular docking experiment(s) were performed and compared with that of PH0851 and *EcRsmB*. The molecular docking results suggest the binding of the SAM molecule at the active-site pocket of the protein PH1078 (model, -7.30 kcal mol⁻¹) akin to its closest homologs *EcRsmB* (PDB id: 1SQF) and PH0851 (PDB id: 2YXL, -9.20 kcal mol⁻¹). The SAM molecule docked to the protein PH1078 is coordinated by the residues Ala270, Ala271, Gly273, Lys275, Asp293 Asp320, and Asp336 from the conserved motifs I, II, III, and IV (Figure 3.6C-3.6D). As a positive control, cytosine was docked to the proteins *EcRsmB* and PH0851, which exhibited its binding at the active-site pocket with a binding energy of -6.04 and -7.39 kcal mol⁻¹, respectively (Figure A.2A-A.2B and Table A.4). To identify the

binding site of the substrate cytosine, the SAM-docked PH1078 was considered a receptor molecule. Expectedly, the cytosine base binds at the active site of the protein with a binding energy of -5.20 kcal mol⁻¹ (Table A.4). The docked cytosine interacts with the residues Asp336 and Cys339 from motif IV and Cys389 from motif VI of the catalytic domain, suggesting their participation in the catalysis of m⁵C formation (Figure 3.6E). Further, dynamic communities of the above three proteins are determined by using dynamic community identifiers. The results exhibit that the above three proteins (*EcRsmB*, PH0851 and PH1078) are clearly distinguished by the presence of two community RNA binding domains (NusB) with linker region and core domain with CH scores 769, 788, and 1047, respectively. Additionally, an elastic network model (ENM) was conducted to deduce the domain dynamics of the three aforementioned proteins. The findings revealed a similar conformational change in each case, transitioning from an open (state 1) to a closed (state 2) conformation. This transition was primarily associated with movements in the α_1 and α_2 helices of the RNA binding domain NusB and the α_{14} - β_6 region of the core domain. These dynamic changes suggest an induced-fit mechanism following substrate entrapment (Figure 3.6F). However, when examining the global motion in other RNA m⁵C MTases, specifically those containing a CTE like *EcRsmF* and *MmNSUN2*, three distinct communities (Core, Disorder, and PUA) were observed with CH scores 128 and 158 respectively (Figure A.3A-A.3B). This distinction likely arises from the presence of a different RNA binding domain (PUA) located in the CTE region (Figure 3.4). The initial movement (state 1) of *EcRsmF* reveals that the PUA domain is positioned at a distance from the active site pocket (Figure A.3C). This observation is consistent with the crystal structure of RsmFs and also in the model structure of NSUN2 proteins, where a disordered region separates the two domains - the core domain and PUA domain (Hallberg et al., 2006; Demirci et al., 2010). Thus, a conformational change occurs when the substrate binds and brings the PUA domain closer to the MTase domain (state 2), which probably enhances the RNA binding (Figure A.3C). This suggests an indirect involvement in the closure of the active site pocket.

Conversely, the networking model reveals distinct conformational changes when the PUA domain is positioned in the NTE region, as observed in NSUN6 proteins (*HsNSUN6*,

PhNSUN6) with four DCI community (Two RRM, core, and PUA domain) and CH score of 120 and 428 respectively (Figure A.3D-A.3F). These conformational changes are likely associated with the interaction between the PUA domain and the CCA end (acceptor arm) of the tRNA, guiding the target cytosine towards the active site (Liu et al., 2017).



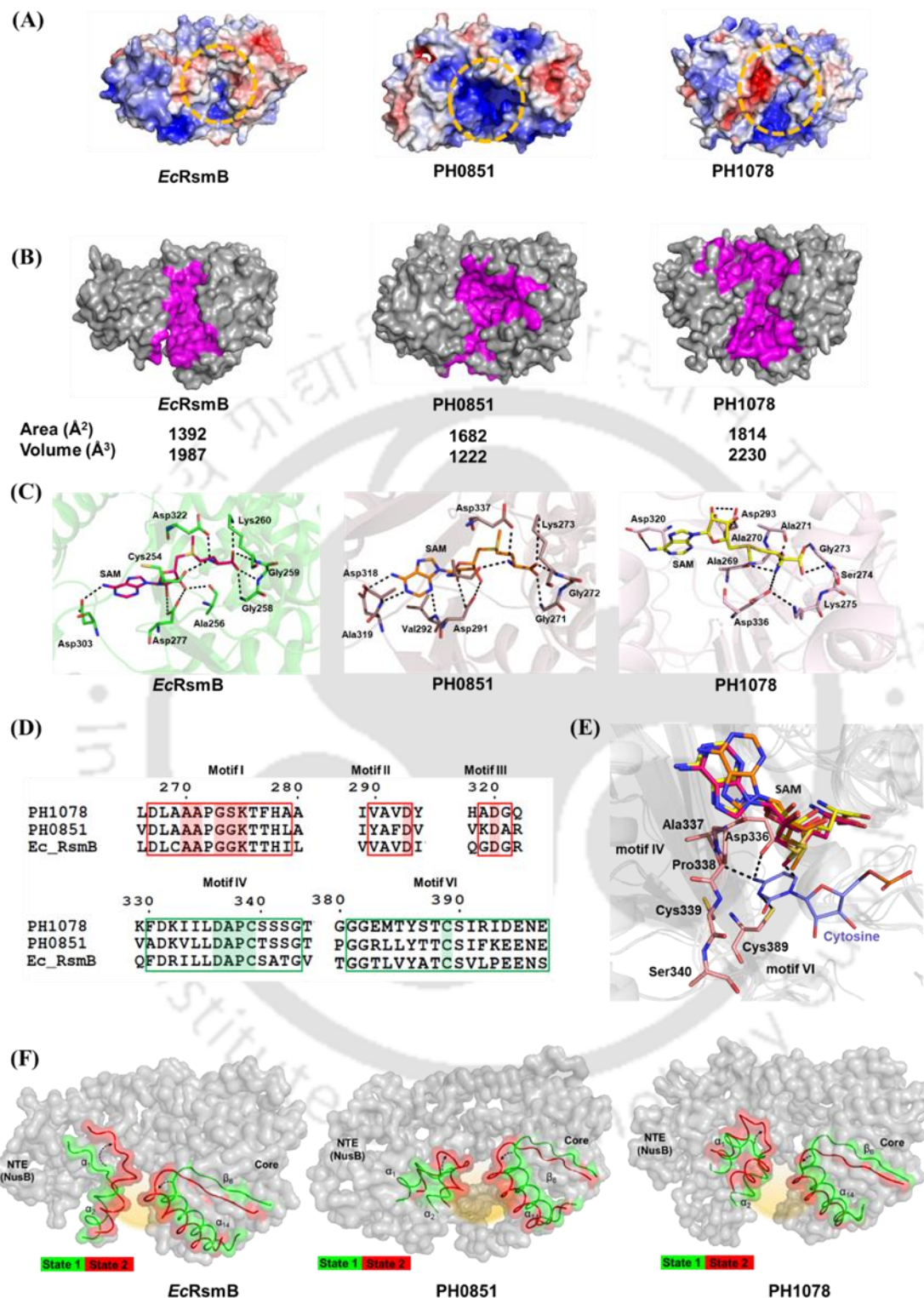


Figure 3.6. Comparison of the active site of the protein PH1078 with its closest homologs *EcRsmB* and PH0851. (A) Electrostatic potential charge distribution of *EcRsmB* (PDB id: 1SQF), PH0851 (PDB id: 2YXL), and PH1078 (model). The positively

and negatively charged surfaces are shown in blue (+5 kT/e) and red (-5 kT/e), respectively. The active-site region in all the proteins is marked with a yellow dashed circle. (B) Comparison of the active-site area and volume of the protein PH1078 with its closest homologs *EcRsmB* and PH0851. The active-site region of the selected proteins is highlighted in magenta. (C) The molecule SAM docked to the proteins PH0851 and PH1078. The interaction of SAM with the proteins *EcRsmB*, PH0851, and PH1078 is shown in magenta, orange, and yellow, respectively. (D) The conservation of the SAM (red) and cytosine (green) binding residues in the proteins *EcRsmB*, PH0851, and PH1078 is shown by multiple sequence alignment. (E) The binding of the cytosine base to the protein PH1078 near the SAM binding site. The molecule SAM bound to all three proteins (*EcRsmB*, PH0851, and PH1078) are superimposed together. (F) Conformational changes from open to closed forms were observed in all three proteins (*EcRsmB*, PH0851, and PH1078). The open and closed conformations were generated using the hdANM server. These conformational transitions are facilitated by rotations occurring in specific regions, namely α_1 and α_2 in the NusB domain, and α_{14} and β_6 in the core domain.

3.3.3.3. The NTE of the protein PH1078 (subfamily IV) and RsmBs from subfamily I is indistinguishable

The active-site analysis suggested that the proteins *EcRsmB*, PH0851, and PH1078 possess a similar binding site. However, the RsmBs belonging to the Fmu subfamily I differ from the subfamily IV due to their varying N-terminal sequence (Reid et al., 1999). The NTE of *EcRsmB* shows structural similarity with an RNA-binding protein NusB (Foster et al., 2003). Domain analysis of the proteins PH0851 and PH1078 confirms the presence of 157 and 159 residues long NTE, respectively (Figure A.4). To further affirm the presence of such similar NTE in the proteins PH0851, PH1078 and RsmBs, an MSA of the selected proteins along with NusBs was performed. The result reveals that the NTE of the proteins PH0851 and PH1078 possess 13 and 27 conserved residues, respectively. In the protein PH0851, ~8% of the NTE is conserved, while in the protein PH1078, ~17% of the NTE is conserved (Figure 3.7A-3.7B). A similar approach was applied to acquire further insight into the NTE conservation in RsmBs. The residue conservation in the NTE of the proteins

EcRsmB, *PaRsmB* (*P. aeruginosa*), *SeRsmB* (*Salmonella enterica*), *KpRsmB* (*K. pneumoniae*), *BsRsmB* (*B. subtilis*) and *ShbRsmB* (*S. boydii*) was found to be 19, 18, 19, 19, 17 and 19%, respectively, with an average of 18% (Figure 3.7A-3.7B). On the other hand, the residue conservation in NusBs, viz. *MtNusB* (*M. tuberculosis*), *EcNusB*, and *AaNusB* (*A. aeolicus*) are 13, 19, and 14%, respectively, with an average of 15% (Figure 3.7A-3.7B). Thus, this analysis clearly hints that the NTE of the protein PH1078 is similar to that of the protein NusB and the NTE of RsmB proteins rather than the NTE of the protein PH0851.

An in-depth analysis of the MSA result suggests that the residues Phe97, Val110, Val128, and Leu129 located in the NTE of the protein PH1078 are conserved in both the RsmB and NusB homologs. These conserved residues can probably participate in forming the hydrophobic core of the NusB domain to maintain its domain structure comprising eight α -helices packed into a globular bundle (Gopal et al., 2000). In contrast, these residues are absent in the protein PH0851. However, upon scrutinizing the packing entropies of the protein *MtNusB* in the context of the Kyte-Doolittle scale for amino acid hydrophobicities, along with the NTE (*EcRsmB*, PH0851, PH1078) and CTE (*EcRsmF* and *TtRsmF*) region of different proteins, distinctive patterns emerge. The results illustrate a specific region (cluster 1) on the entropy scale with high density across various parts of the hydrophobicity scale. This suggests that the majority of amino acid residues located in the *MtNusB*, as well as all the NTE regions of the aforementioned proteins, share a similar entropy level (~ 15 J. K⁻¹mol⁻¹), implying comparable packing geometries among their protein structures (Figure A.5). In contrast, proteins containing the CTE region exhibit higher packing entropy values (cluster 2) than those for the NTE regions and are dispersed across the packing entropies scale indicating distinct packing geometry (Figure A.5). Moreover, the NTE of the *MtNusB* (PDB id: 1EYV) shares a very low sequence identity (<20%) compared to the NTE of *EcRsmB*, PH0851, and PH1078, its overall tertiary structure is similar to the NTE of the latter proteins, with an RMSD of 1.5, 1.7 and 1.0 Å, respectively (Figure 3.7C). Altogether, this indicates that in spite of their low sequence similarity, the NTE of PH0851, RsmB, and PH1078 share a similar tertiary structure.

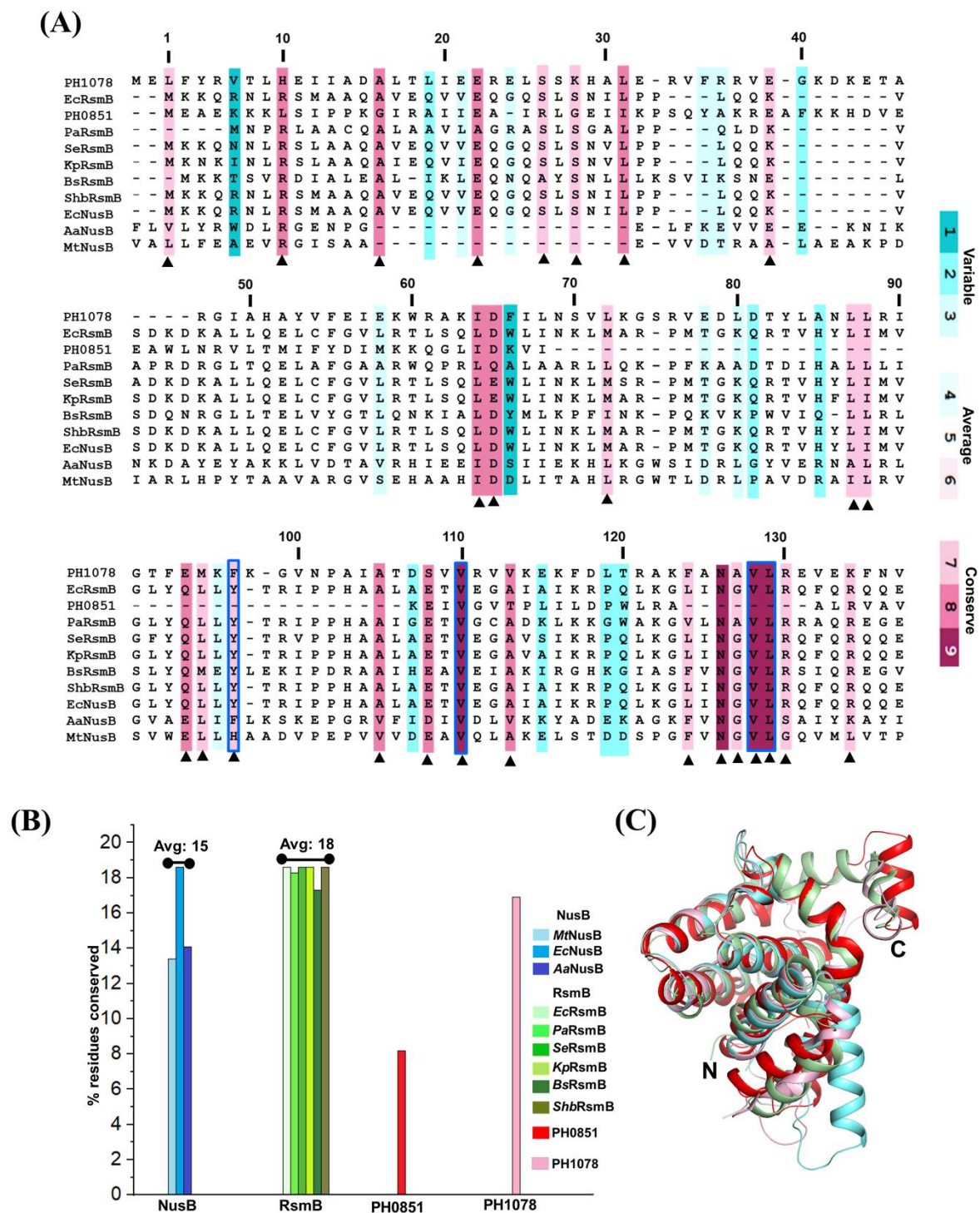


Figure 3.7. Sequence conservation in the N-terminal extension (NTE) of the subfamily I (RsmBs) and subfamily IV (PH0851 and PH1078) with NusB proteins. (A) Multiple sequence alignment for the NTE of PH1078 (*P. horikoshii*, O58805) with the NTE of

EcRsmB (*E. coli*, P36929), PH0851 (*P. horikoshii*, O58581), *PaRsmB* (*P. aeruginosa*, Q9I7A9), *SeRsmB* (*Salmonella enterica*, Q8Z1X1), *KpRsmB* (*K. pneumoniae*, B5XNC2), *BsRsmB* (*B. subtilis*, P94464), *ShbRsmB* (*S. boydii* Q31VY8), *EcNusB* (*E. coli*, P0A780), *AaNusB* (*A. aeolicus*, O66530), and *MtNusB* (*M. tuberculosis*, P9WIV1); the organism's name and UniProt id of each protein are provided in the parenthesis. The conserved residues at the NTE of the protein PH1078 are marked by triangles. The conserved hydrophobic core-forming residue is in a blue rectangular box. (B) Graphical representation of percentage residue conservation in protein NusBs and NTEs of RsmBs, PH0851 and PH1078. The average (avg) conservation of the residues is also provided. (C) Superposition of the overall structural fold of *MtNusB* (cyan) with NTE of *EcRsmB* (green), PH0851 (red), and PH1078 (pink).

3.3.4. NSUN1-like archaeal rRNA m⁵C MTase

3.3.4.1. The protein PH1374 is plausibly an rRNA m⁵C MTase

In the UniProtKB database, the protein PH1374 has been annotated as “hypothetical proliferating-cell nucleolar protein p120”. Further, it shows the closest homology (QC: 89%, SI: 41%) with human nucleolar p120 protein NSUN1 (*HsNSUN1*), in addition to *EcRsmB* (QC: 99%, SI: 32%) and *EcRsmF* (QC: 87%, SI: 39%). The preliminary analysis of this study suggests that the protein PH1374 belongs to the subfamily V, in contrast to the *HsNSUN1*, which belongs to the subfamily II of RNA m⁵C MTases (Reid et al., 1999). The protein *HsNSUN1* possesses an arginine/lysine-rich region at the NTE required for RNA binding in addition to the three repeats (referred to as p120R) at the CTE (Gustafson et al., 1998; Staub et al., 2004). To identify these characteristic features in the protein PH1374, an MSA of NSUN1 proteins from eukaryotes and *EcRsmB* and *EcRsmF* was performed. The MSA analysis reveals that the positively charged arginine/lysine-rich region and the three repeats observed in *HsNSUN1* are absent in all three proteins, PH1374, *EcRsmB*, and *EcRsmF* (Figure 3.8A). Thus, to identify a probable substrate for the protein PH1374, a structure-based evolutionary tree was generated using its homologous structures. The result suggests that the protein PH1374 clusters with *EcRsmB*

and *EcRsmF* despite sharing higher sequence similarity with the protein *HsNSUN1*, which clades with *NSUN5*, another human m⁵C rRNA MTase (Figure 3.8B).

Furthermore, the electrostatic potential charge distribution of the protein reveals that the active-site pocket of the protein PH1374 is comprised of both positively and negatively charged residues, similar to that of the *EcRsmB* and *EcRsmF* (Figure 3.8C). As expected, the negatively charged surface contributes to the SAM binding, while RNA molecules can be accommodated in the positively charged region. In contrast, the protein *HsNSUN1* is comprised of only a positively charged active site (Figure 3.8C). Moreover, the molecular docking of SAM indicates that it binds to the protein PH1374 ($-7.45 \text{ kcal mol}^{-1}$) and is coordinated by the residues Glu104, Lys132 and Asp194 from the conserved motifs X, I and IV, respectively (Figure 3.8D and Table A.4). Further, the protein PH1374 complexed with SAM was docked with the substrate cytosine at its active-site pocket with a binding energy of $-6.51 \text{ kcal mol}^{-1}$ (Table A.4). The docked cytosine interacts with the residues Cys194 and Cys247 from the motifs IV and VI, respectively, of the catalytic domain of the protein PH1374 suggesting their participation in the catalysis of m⁵C formation (Figure 3.8D-3.8F and Table A.4). As a positive control, the protein *EcRsmF* was firstly docked with SAM, followed by the substrate cytosine (Figure A.6 and Table A.4). Altogether, the results from this study suggest that the protein PH1374 is a probable m⁵C MTase capable of binding SAM utilizing rRNA as its substrate.

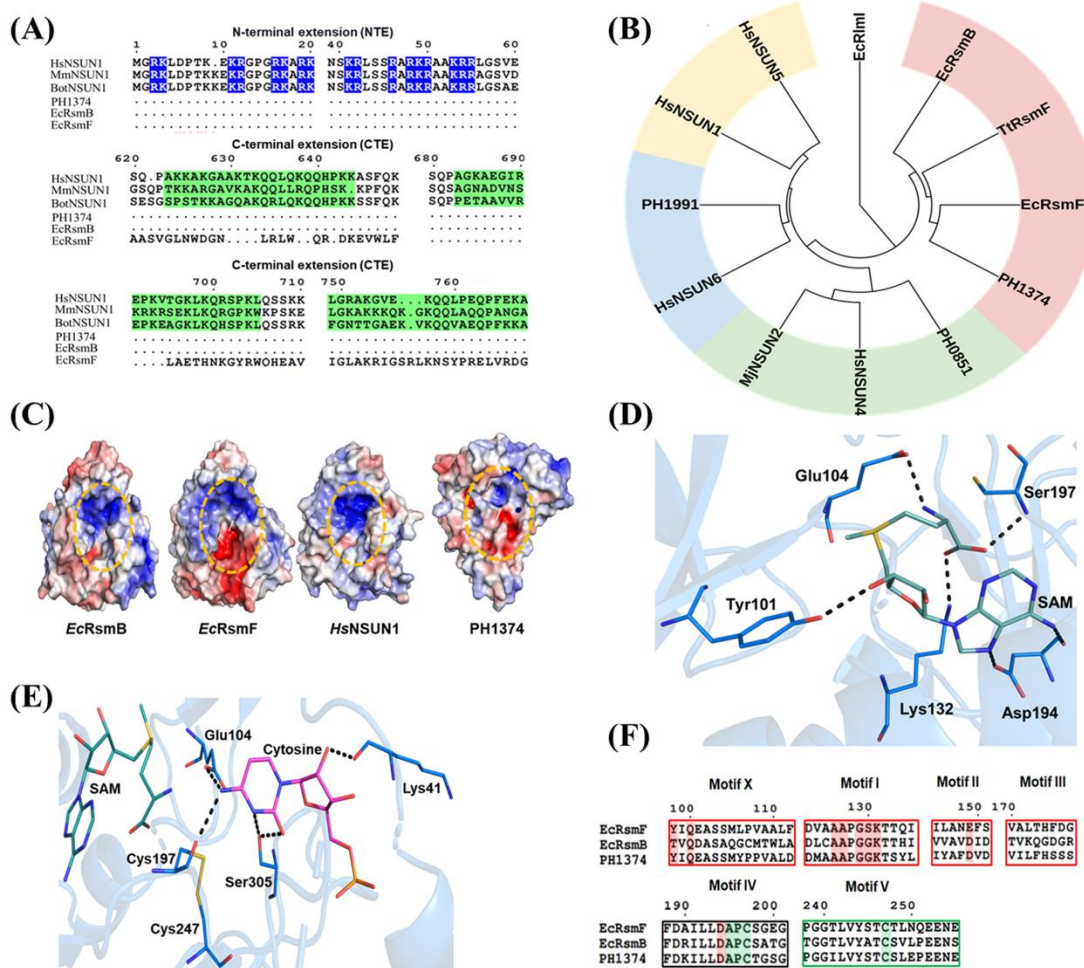


Figure 3.8. The protein PH1374 probably functions as a rRNA m⁵C MTase. (A) Multiple sequence alignment of the protein PH1374 with its homologs. The proteins used in MSA are *HsNSUN1* (*H. sapiens*, O58581), *MmNSUN1* (*M. musculus*, Q922K7), *BotNSUN1* (*B. taurus*, F1N3G2), PH1374 (*P. horikoshii*, O50082), *EcRsmB* (*E. coli*, P36929) and *EcRsmF* (*E. coli*, P76273), the organism name and UniProt id of each protein are provided in the parenthesis. (B) Structure-based distance tree of the protein PH1374 and its structural homologs. The protein structures used for the analysis are *EcRsmB* (PDB id: 1SQF), *TtRsmF* (PDB id: 3M6V), *EcRsmF* (PDB id: 2FRX), PH1374 (PDB id: 1IXK), PH0851 (PDB id: 2YXL), *HsNSUN4* (PDB id: 4FP9), *MjNSUN2* (PDB id: 3A4T), *HsNSUN6* (PDB id: 5WWR), PH1991 (PDB id: 5ZVH), *HsNSUN1* (model), *HsNSUN5* (PDB id: 2B9E) and *EcRlmI* (PDB id: 3C0K). (C) The electrostatic potential charge distribution of the protein PH1374 was compared with the protein *EcRsmB*, *EcRsmF*, and

HsNSUN1. The positively and negatively charged surfaces are shown in blue (+5 kT/e) and red (-5 kT/e), respectively. (D-E) The binding study of SAM and cytosine with the protein PH1374. (F) The conservation of the SAM (red) and cytosine (green) binding residues in the proteins *EcRsmF*, *EcRsmB*, and PH1374 is shown by multiple sequence alignment.

3.3.5. Archaeal tRNA m⁵C MTases

3.3.5.1. The genes *PH1536* and *PH1537* might form a bicistronic arrangement

The preliminary homology search analysis suggested that the gene *PH1537* belongs to the subfamily V, showing significant similarity with another archaeal gene *PAB1947*, which encodes a tRNA m⁵C MTase in *P. abyssi*. In *P. abyssi*, two genes, *PAB1947* and *PAB1946*, form an operon, of which the latter encodes an archease protein (Auxilien et al., 2007). Akin to its homologous protein, the genes *PH1537* and *PH1536* (encoding an archease protein) form an operon, as predicted by the web tool Operon Mapper (Taboada et al., 2018). The intergenic region, directionality, and functional relatedness are the three main characteristic features of the genes present in an operon. The two genes *PH1536* and *PH1537* are separated by ten nucleotides, similar to that of the *PAB1946-PAB1947* operon present in *P. abyssi*. The protein archease in *Thermotoga maritima* is reported to provide thermostability to its neighboring gene *gyrA* (a gyrase) (Canaves, 2004). In addition, the protein archease also provides specificity in RNA ligation and tRNA m⁵C modification (Auxilien et al., 2007; Desai et al., 2014).

In several organisms, such as *Pelobacter propionicus*, *Methanosarcina mazei*, and *Halobacterium salinarum*, the gene *archease* is reported to form an operonic arrangement with its neighboring gene *rtcB* (a non-canonical tRNA ligase) to modulate the RNA ligation. In *Thermococcus kodakarensis*, the *archease* operon is comprised of the *rtcB* gene along with an additional tRNA MTase gene (Desai et al., 2014). In *P. horikoshii*, the operon analysis suggests that the archease (*PH1536*) gene is associated only with *PH1537* in an operon (Figure 3.9A). The functional relatedness between the archease genes (*TM1083*, *Tk0361*, and *PH1536*) and the other gene(s) in the respective

microorganisms was further validated by a protein-protein interaction network analysis. The interactome analysis revealed that a bacterial archease, TM1083, interacts with TM1084 (a gyrase). Similarly, both the archaeal archease (TK0361 and PH1536) show interaction with RNA splicing ligase, RtcB (TK0360 and PH1602), and tRNA MTase (TK0358 and PH1537). Overall, the results from interactome mapping corroborate the operonic arrangement analysis (Figure 3.9B-3.9D and Table A.3).

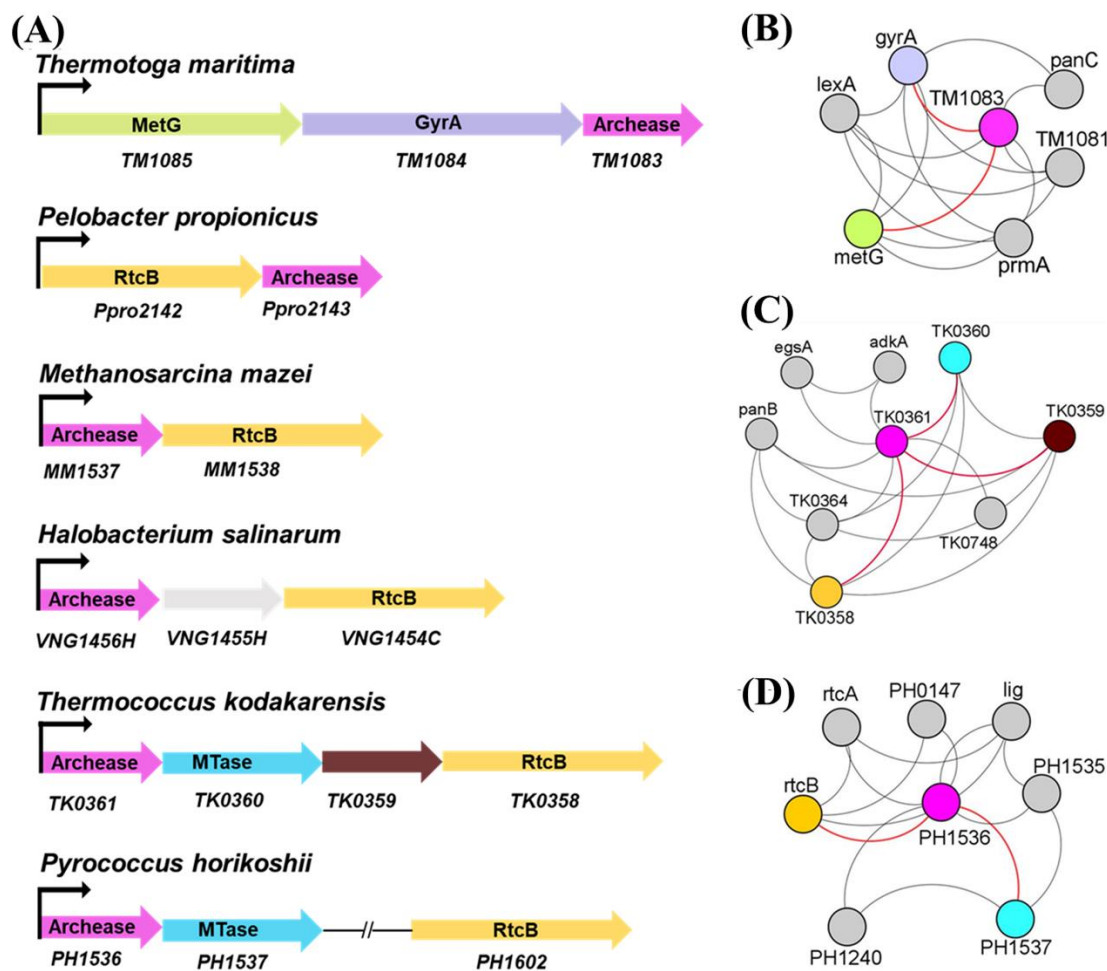


Figure 3.9. Operonic Organization of the *archease* Gene in Prokaryotes. (A) The operon organization of *archease* gene in bacteria (*T. maritima* and *P. propionicus*) and archaea (*M. mazei*, *H. salinarum*, *T. kodakarensis* and *P. horikoshii*). (B-D) The protein-protein interaction network of archease proteins with a cut-off score of 8; TM1083, PH1536, and TK0361 from *T. maritima*, *P. horikoshii* and *T. kodakarensis*, respectively.

The nodes of the archease proteins belonging to the same operon are shown in magenta. The nodes of the other interacting proteins, tRNA MTase, RtcB, and DNA gyrase subunit A, methionine-tRNA ligase, are represented by cyan, yellow, green, and purple, respectively. The targeted interactions of each operon are denoted as red dotted lines, while those of other proteins are indicated by black dotted lines. The other details of each interacting protein are provided in Table A.3.

3.3.5.2. The proteins PH1536 and PH1537 may form a functional complex

In *P. abyssi*, the protein PAB1947 (a tRNA MTase) modifies the tRNA at the 49th position for its maturation. The substrate specificity of PAB1947 is achieved by the archease by forming an archease-MTase complex (Auxilien et al. 2007). To figure out whether PH1536 (archease) and PH1537 (putative m⁵C MTase) may also form a complex, a protein-protein docking calculation was performed. For this, the dimeric form of the archease (PH1536, PDB id: 4N2P) was docked with the tRNA MTase (PH1537). The protein-protein docking results exhibit the complex formation of PH1536 and PH1537 with buried surface area (BSA) of 3524 Å² (Figure 3.10A). This indicates stable complex formation between the two proteins in accordance with the previous report (Chen et al., 2013). Further, the electrostatic potential charge distribution of the docked conformation of the protein PH1536 and PH1537 exhibits that the negatively charged surface of the protein PH1536 is neutralized by the positively charged surface of the protein PH1537 (m⁵C MTase), resulting in the formation of a stable complex (Figure 3.10B). Further, the result shows that the archease protein PH1536 masks a positively charged surface present at the peripheral surface of the protein PH1537, possibly involved in the binding of numerous RNA (probably tRNA) to avoid non-specific interactions (Figure 3.10B). In particular, the residues Asp13, Glu68, Glu115, Glu116, and Asp142 from chain A of the protein PH1536 interact with the residues Lys145, Lys164, Lys168, and Lys175 of the protein PH1537. Similarly, the residues Glu55, Glu57, Glu68, Glu69, Glu77, Glu116 and Asp142 from chain B of the protein PH1536 interact with the residues Arg215, Lys218, Lys219, Arg226 and Arg256 of the protein PH1537 (Figure 3.10C and Table 3.1). As a positive control, the

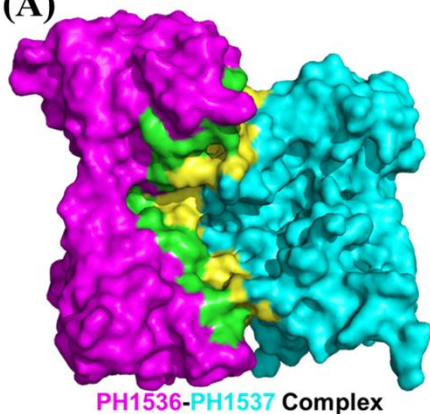
protein-protein docking between PAB1946 and PAB1947 was performed, resulting in the formation of a stable protein-protein complex with BSA of 2663 Å² (Figure A.7).

Table 3.1 Details of the protein-protein docking results of targeted proteins (PH1537 and PH1536) and their homologs (PAB1947 and PAB1946).

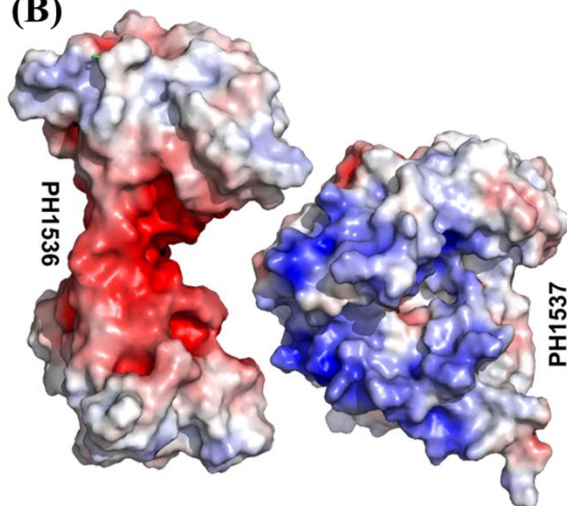
Receptor	Ligand	Number of Conformations	Distance (Å)	Interface residues between docked complex			
				Receptor		Ligand	
				Residue	Chain	Residue	Chain
PH1537	PH1536	62	2.7	Lys145	C	Glu115	A
			1.9	Lys145	C	Glu116	A
			2.7	Lys164	C	Glu68	A
			4.3	Lys168	C	Asp13	A
			1.8	Lys168	C	Asp142	A
			2.7	Lys175	C	Asp13	A
			1.8	Lys175	C	Asp142	A
			2.0	Arg215	C	Glu55	B
			2.2	Arg215	C	Glu57	B
			5.6	Arg215	C	Glu68	B
			1.7	Lys218	C	Glu69	B
			3.7	Lys219	C	Glu69	B
			1.7	Lys219	C	Glu68	B
			1.9	Arg226	C	Glu116	B
			1.8	Arg256	C	Glu77	B
PAB1947	PAB1946	60	2.6	Lys13	C	Glu108	A
			4.3	Lys168	C	Glu68	A
			2.8	Arg175	C	Glu67	A
			2.8	Arg215	C	Glu115	A

			4.2	Arg215	C	Glu114	A
			3.0	Arg219	C	Asp74	A
			6.3	Arg226	C	Glu67	A
			5.9	Lys256	C	Asp141	A
			2.7	Lys256	C	Asp141	B
			2.7	Lys260	C	Glu67	B
			2.8	Lys260	C	Glu68	B
			2.5	Lys281	C	Glu59	B
			4.6	Lys281	C	Asp57	B
			4.9	Arg287	C	Glu56	B
			3.6	Arg311	C	Asp74	B

(A)



(B)



(C)

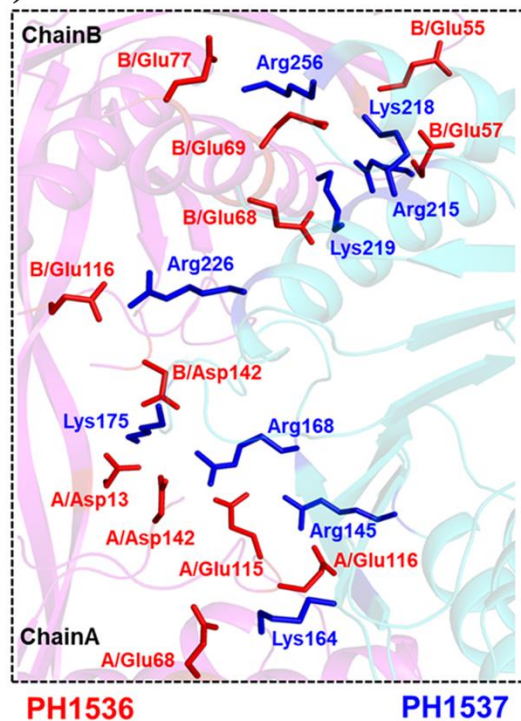


Figure 3.10. Protein-protein docking between the proteins PH1536 and PH1537. (A) Surface representation of the stable complex formed by the proteins PH1536 and PH1537. The archease (PH1536) and m⁵C MTase (PH1537) are represented in magenta and cyan, respectively. The amino acid residues located less than 6 Å away from their associate protein surface indicate the interface residues, which are shown in green and yellow, respectively, for the proteins PH1536 (magenta) and PH1537 (cyan). (B) Electrostatic potential charge calculation for the protein PH1536 (PDB id: 4N2P) and PH1537 (model). The positively and negatively charged surfaces are shown in blue (+5 kT/e) and red (-5 kT/e), respectively. (C) Interaction of the amino acid residues at the interface of the complex PH1536 and PH1537. The negatively charged residues at the interface of the protein PH1536 mask the positively charged residue of the protein PH1537. The negatively and positively charged residues in the protein PH1536 and PH1537 are represented as red and blue sticks, respectively.

3.3.5.3. The N-terminal region of the protein PH1537 resembles the N1 domain of the RsmF family

The preliminary homology search revealed the similarity of the protein PH1537 with *EcRsmF* (QC: 91%, SI: 38%), a member of the subfamily III. The members of the subfamily III possess three domains, viz. N1, catalytic, and the PUA (or RNA recognition) domain (Hallberg et al., 2006). The domain analysis of the protein PH1537 and its archaeal homologs revealed the absence of the PUA (or RNA recognition) domain (Figure 3.11A). It is noted that the bacterial and archaeal RNA m⁵C MTases possess similar catalytic domains, having all the signature motifs conserved (Figure 3.2 and A.1). The N1 domain (7-102) of *EcRsmF* comprises a $\beta\alpha\beta\beta\alpha\beta$ fold (Hallberg et al., 2006). In contrast, it is different than that of *EcRsmB*. Notably, the PUA domain and the disordered region of RsmF are absent in the protein PH1537 (Figure 3.11A). Multiple sequence alignment of the N1 domain of RsmF proteins along with PH1537 revealed that the N-terminal (10-100) of the protein PH1537 shares significant conservation (SI: 40%) with the N1 domain of the RsmF protein (Figure 3.11B).

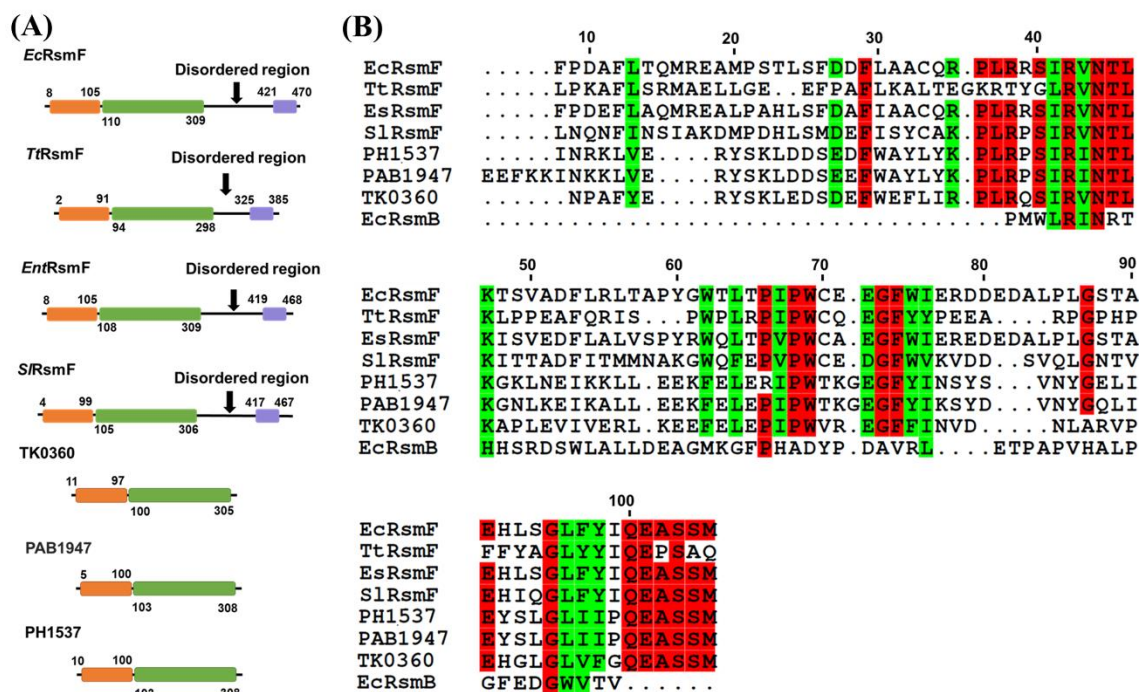


Figure 3.11. The N1 domain of PH1537 is highly conserved with bacterial RsmF proteins. (A) Comparison of the domain organization of the RsmF proteins from *E. coli* (EcRsmF, P76273), *Thermus thermophilus* (TtRsmF, Q5SII2), *Enterobacter* sp. (EntRsmF, A4WBJ4) and *Shewanella loihica* (SlRsmF, A3QDA2) with their archaeal homologs TK0360 (Q5JCY9), PAB1947 (Q9V106) and PH1537 (O59206)); the protein name (wherever required) and the UniProt id of each protein is provided in the parenthesis. The N1, catalytic, and PUA domains are represented by orange, olive, and purple rectangles, respectively. The disordered region is shown in the black line. (B) Multiple sequence alignment of these proteins indicates the conservation of the N1 domain. The identical and similar residues are highlighted in red and green, respectively.

3.3.6. A proposed evolutionary classification of RNA m⁵C MTases

The evolutionary analysis of RNA m⁵C MTases from all domains of life revealed four distinct clusters (Figure 3.12). Cluster I includes NSUN1 proteins from eukaryotes with the characteristics of longer N- and C-terminal extension. Among the archaeal proteins, PH1374 lies in cluster I, indicating its similarity to the core domain of the NSUN1 proteins. The cluster II members belong to the five different RNA m⁵C MTases, viz. bacterial RsmF,

archaeal PAB1947 and PH1537, and eukaryotic NSUN2, NSUN3, and NSUN4 proteins (Figure 3.12A). The archaeal proteins (PAB1947 and PH1537) are closer to the eukaryotic NSUN2 than the bacterial RsmFs. In cluster III, the eukaryotic NSUN5 and NSUN7, along with archaeal proteins PH0851 and PH1078, are grouped. Notably, archaeal proteins (PH0851 and PH1078) are closer to the bacterial RsmB than the eukaryotic proteins (NSUN5 and NSUN7) (Figure 3.12A). Cluster IV includes the eukaryotic NSUN6 and archaeal PH1991 protein (Figure 3.12A). Interestingly, the domain analysis of members of each cluster indicates a similarity in their core domain (Figure 3.12B). Furthermore, besides the core domain, each cluster contains a signature feature, making them unique. For example, the eukaryotic NSUN1, bacterial RsmF, and RsmB contain signature features p120R, PUA, and NusB, respectively. However, in cluster IV, the PUA domain lies within the core domain (Figure 3.12B).

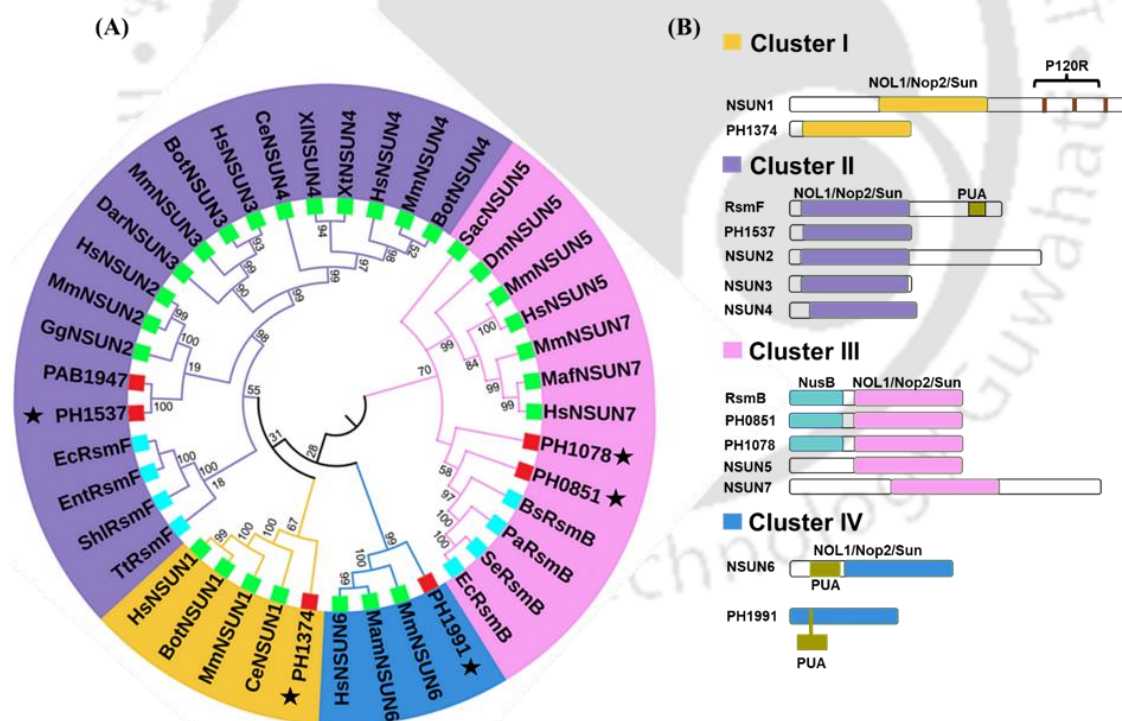


Figure 3.12. Evolutionary relationship of the RNA m⁵C MTases. (A) Phylogenetic tree showing the evolutionary relationship between different m⁵C MTases. The tree was generated using 41 RNA m⁵C MTases, and details are provided in Table A.5. Each RNA m⁵C MTase is labeled with a square box in green (eukaryote), red (archaea), and cyan

(bacteria). The result indicates that five archaeal RNA m⁵C MTases (PH0851, PH1078, PH1374, PH1537, and PH1991) belong to four distinct clusters. PH0851 and PH1078 cluster with RsmB, NSUN5 and NSUN7 (pink); PH1374 with NSUN1 (yellow); PH1537 clade with RsmF, NSUN2, NSUN3 and NSUN4 (purple); and PH1991 clade with NSUN6 (blue). The archaeal RNA m⁵C MTases are marked with a black asterisk. (B) Each cluster of RNA m⁵C possesses defining features of its domains. The core NOL1/NOP2/SUN domain in each member of RNA m⁵C MTases are highlighted with their respective cluster color. The other domains in the RNA m⁵C are depicted in different colors (e.g., p120R domain, brown; PUA domain, olive; NusB domain, teal).

3.4. Discussion

The RNA m⁵C MTases majorly target the translational apparatus modifying the rRNAs and tRNAs in the vicinity of the decoding and peptidyl-transferase centers in large numbers, influencing its translation accuracy and fidelity (Schaefer, 2015; Bohnsack et al., 2019). In eukaryotes, proteins NSUN1-7 modify a wide variety of RNAs, including rRNA, tRNA (cytoplasmic and mitochondrial), mRNAs, enhancer RNAs (eRNAs) and non-coding RNAs. However, in bacteria, RsmB and RsmF MTases containing NOL1/NOP2/SUN domains modify only 16S rRNA (Kuznetsova et al., 2019). This study shows that across the archaeal genome, the highly diverse group of Euryarchaea under the *Thermococci* class, the two genera *Pyrococcus* and *Paleococcus*, possess the highest number (five) m⁵C RNA MTases (Figure 3.1 and Table A.1). The genera of *Pyrococcus* specifically the species *P. horikoshii*, possesses two proteins, PH0851 and PH1991. Both these proteins have been structurally characterized as homologs of bacterial RsmB and eukaryotic NSUN6, which mainly modify two types of RNAs, i.e., rRNA and tRNA, respectively. The motif analysis of the remaining proteins PH1078, PH1374, and PH1537 suggested their involvement in m⁵C RNA modifications. Moreover, the presence of two cysteines in the motifs IV and VI suggests their two-cysteine mechanism of rRNA modification. All the subfamilies of RNA m⁵C MTases conserve their core domain and can be recognized by NTE or CTE or, in the case of subfamily VI, insertion into the core sequence.

A homology search of the protein PH1078 showed similarity with bacterial RsmBs. In addition, the presence of N-terminal extension (NTE) and the absence of C-terminal extension (CTE) in the protein PH1078 indicate its functional closeness to the bacterial RsmB. Based on these observations, both the archaeal proteins PH0851 and PH1078 can be included in the subfamily IV of rRNA m⁵C MTases. Previous studies have firmly established the interaction of RsmB with RsmA (modulation of gene expression), TruB (involvement in four-step catalytic mechanism), TmcA (protein translation), and RlmC (evolutionarily related) (Chatterjee et al., 2002; Bujnicki et al., 2004; Walbott et al., 2007). The comparative network analysis shows that PH0851 and PH1078 interact with their similar archaeal counterparts, suggesting that they are involved in a function similar to that of the bacterial RsmB. This clearly indicates the presence of two RsmB paralogs in archaea.

Owing to the presence of a conserved catalytic domain similar to *EcRsmB*, the proteins PH0851 and PH1078 exhibit functional similarity. However, the two proteins differ at the sequence level (QC: 39%, SI: 69%), indicating their variation in NTE. Interestingly, the NTEs of RsmBs are more similar to the NTE of PH1078 than those of the protein PH0851. However, the overall structures of the NTEs of these proteins are comparable. This suggests that the protein PH1078 can also function as an archaeal RsmB, similar to the protein PH0851.

It is well established that two functionally related proteins encoded by the same genome probably have arisen from a gene duplication event in the past. Such proteins have been recognized as in-paralogous to each other (Urbonavičius et al., 2008; Richardson and Watson, 2013). A similar phenomenon is observed in the case of human mitochondrial transcription factor (TFB1M), which is required for mitochondrial homeostasis. TFB1M catalyzes the dimethylation of two adjacent adenines located in helix45 (h45) of 12S rRNA. Furthermore, TFB2M, which is a paralog of TFB1M, also participates in biogenesis as a dimethyltransferase or transcription factor. The active-site pockets of TFB1M and PH1078 exhibit a clear binding site for SAM (negatively charged), but the corresponding region in TFB2M, as well as PH0851, is a highly positive, which might contribute to their

role as efficient transcription factors (Liu et al., 2019). Thus, based on these extensive analyses and available reports, it can be hypothesized that the two proteins, PH0851 and PH1078, are in paralogs. The hypothesis was further validated by Gaussian elastic network models exhibiting similar domain motion within a protein.

In conclusion, it can be proposed that in the archaeon *P. horikoshii* OT3, two RsmBs are present. However, the requirement of two RsmBs in *P. horikoshii* warrants further experimental studies. Surprisingly, eukaryotic paralogous proteins NSUN5 and NSUN7 belong to the same cluster as the proteins PH0851 and PH1078 (Cluster III). Due to the presence of NTEs, the domain arrangements of the proteins NSUN5 are similar to prokaryotic RsmBs. Moreover, the proteins NSUN7 display a peculiar characteristic of possessing CTEs in addition to NTEs, unlike NSUN5. Furthermore, both the eukaryotic proteins NSUN5 and NSUN7 lack the NusB domain. However, their core domains are highly conserved compared to their prokaryotic counterparts. As a consequence, the cluster III archaeal proteins (PH0851 and PH1078) are evolutionarily closer to their bacterial homologs rather than their eukaryotic counterparts.

The structure-based phylogeny revealed the surprising clustering of the protein PH1374 with bacterial RsmF. Despite lacking the signature extension (arginine/lysine-rich motif and p120R), the sequence-based phylogeny suggests that the ancestral protein PH1374 belongs to cluster I. However, both structure-based and sequence-based studies suggest that the protein PH1374 is probably associated with a similar substrate specificity to rRNA. Nevertheless, the proteins NSUN1 and RsmF are well-characterized rRNA m⁵C MTase, indicating the substrate of the protein PH1374 might be an rRNA molecule akin to both proteins (Table A.6). Based on the results from this study, it can be speculated that the primitive archaeal protein PH1374 might be the common ancestor between the eukaryotic NSUN1 and bacterial RsmF proteins. Our study leads to the conclusion that the RNA binding domain (NTE region) of RsmBs exhibits distinctly different geometric packaging compared to the RNA binding region (CTE region) of RsmFs. This disparity is likely attributed to the presence of differing secondary structural elements within these RNA binding domains. Specifically, the NusB domain is tightly packed due to the exclusive

presence of α helices in its secondary structural elements, whereas the CTE region is characterized by a combination of disordered regions and the presence of both α and β elements, notably in the PUA domain.

The protein PH1537, its archaeal homolog PAB1947, and bacterial homolog RsmFs are grouped in cluster II, which corroborates our previous analysis. The protein PAB1947, biochemically characterized as a tRNA m⁵C MTase forms a bicistronic operon with an archease protein in *P. abyssi* (Motorin et al., 2010). Interestingly, both the archaeal proteins, PAB1947 and PH1537, exhibit extensive non-specificity towards numerous RNA molecules as they do not possess any RNA-binding domain. However, negatively charged archease protein can neutralize the predominantly positively charged non-specific sites present on the surface of the protein PH1537 and its homolog PAB1947. Therefore, the archeases (PH1536 and PAB1946) prevent the tRNA methyltransferases (PH1537 and PAB1947) from making non-specific interactions with numerous non-specific tRNAs. Moreover, previous studies showed that the presence of the gene *archease* upstream of gene *mtase* is required for the stability of tRNA MTase and to ensure proper folding. Therefore, it can be hypothesized that the archease protein will probably serve the function of RNA-binding domain in archaeal tRNA m⁵C MTases such as the targeted protein PH1537 and its homolog PAB1947 (Auxilien et al., 2007). Similarly, in eukaryotes, the mitochondrial MTase NSUN4 also lacks the RNA-binding domain, the function of which is performed by a transcription control factor, MTERF4 (Spåhr et al., 2012). Altogether, the evolutionary study indicates that the primitive archaeal proteins PH1537, PAB1947, and eukaryotic NSUN4 are also grouped in cluster II.

3.5. Conclusion

In summary, this study reveals that two archaeal m⁵C MTases (PH1078 and PH1374) exhibit substrate specificity towards rRNA, while PH1537 would function as m⁵C tRNA MTase. Moreover, the five archaeal proteins are functionally and structurally homologous to either bacterial, eukaryotic, or both RNA m⁵C MTases. This finding has helped establish that the archaeal translation process is a connective link between the other two domains of life. The outcomes of this study shed light on the understanding of the evolution of m⁵C

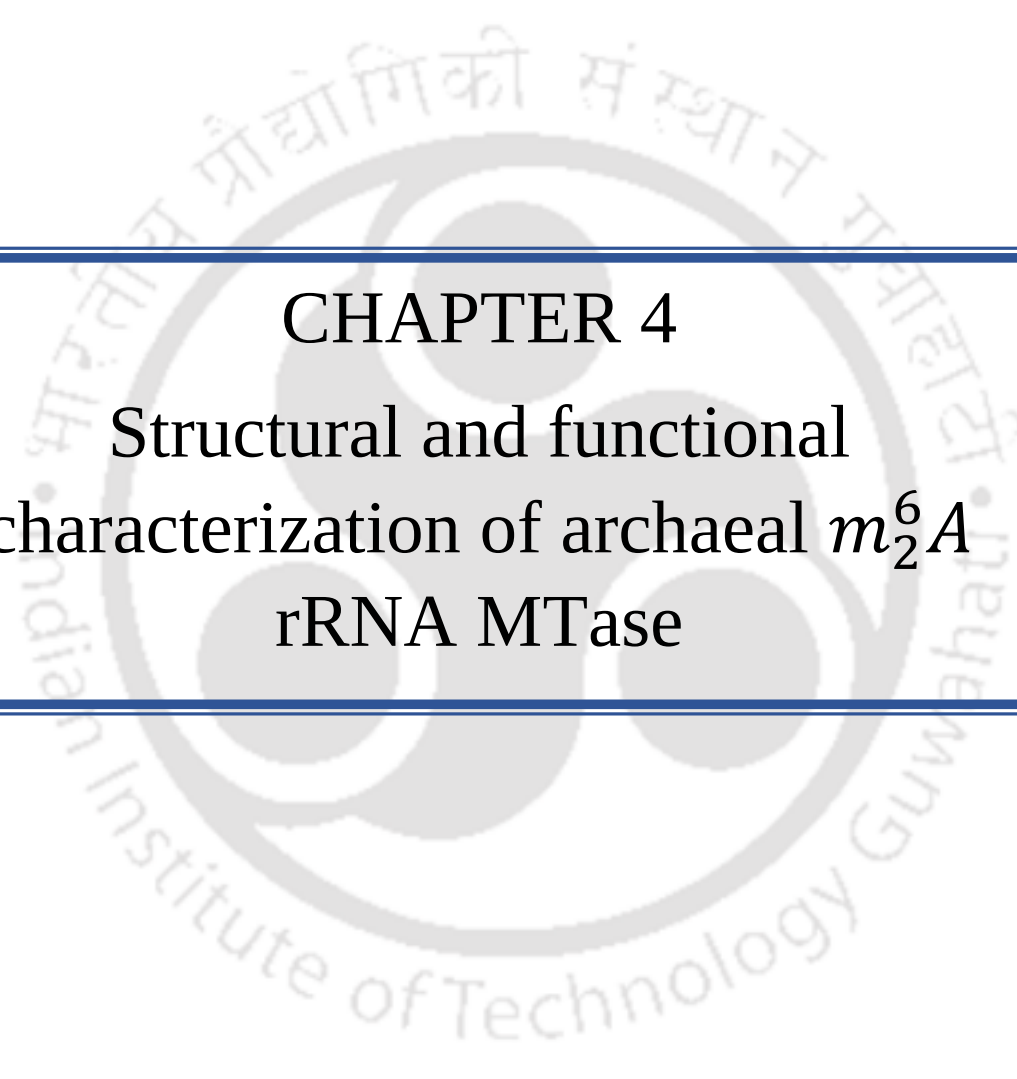
methylation. Furthermore, the work provides significant insights into the functional annotation of the RNA m⁵C MTases, which may pave the path for treating various epigenetics-associated diseases.

Appendix A. Supplementary data

Supplementary Tables, A.1-A.6

Supplementary Figures, A.1-A.7





CHAPTER 4

Structural and functional characterization of archaeal m_2^6A rRNA MTase

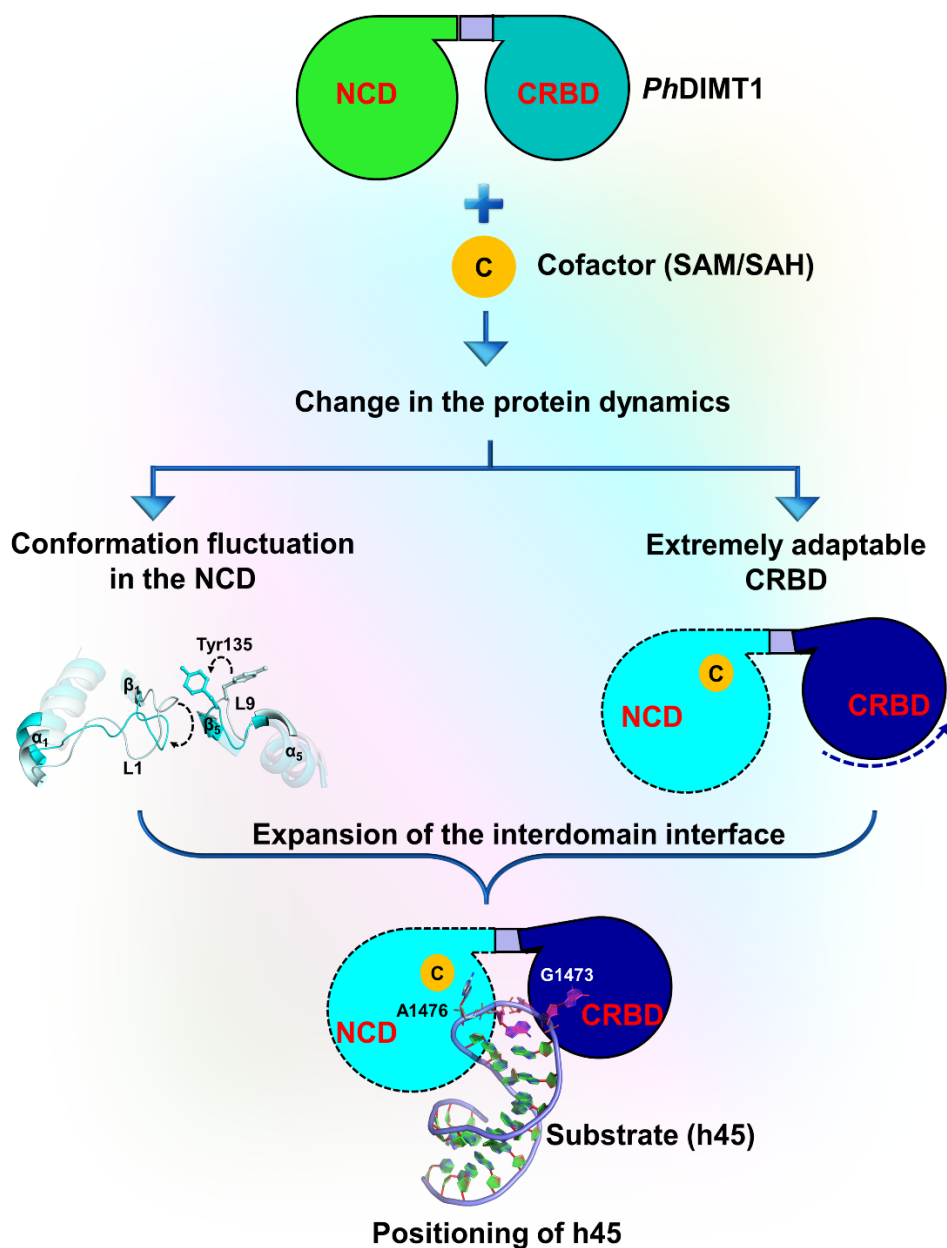
The manuscript for this chapter is under review as:

Saha, S. and Kanaujia, S.P. (2024). Structural and functional characterization of archaeal DIMT1 unveils distinct protein dynamics essential for efficient catalysis. *Structure*, Accepted.

Abstract

The protein dimethyladenosine transferase 1 (DIMT1, an ortholog of bacterial KsgA) is a universally conserved protein family pivotal in ribosome synthesis. It catalyzes rRNA maturation by the base dimethylation of two adjacent adenosine residues (A1518 and A1519 in *E. coli*) near the 3'-terminal end of the small subunit of rRNA. Since the process of ribosome synthesis differs significantly between prokaryotes and eukaryotes, understanding the structural and functional characteristics of the KsgA/DIMT1 protein is essential. Unlike well-characterized bacterial and eukaryal KsgA/DIMT1 proteins, a complete understanding of the archaeal counterpart is elusive. This study fills a gap by unveiling the crystal structures of the archaeal DIMT1 from a hyperthermophilic archaeon *Pyrococcus horikoshii* (*Ph*DIMT1) in its apo and holo forms. To the best of our knowledge, this study elucidates a unique α -amino tail movement of the analog sinefungin (SFG) within the DIMT1-SFG complex, distinguishing it from its cognate cofactors, S-adenosyl-L-methionine (SAM) or S-adenosyl-L-homocysteine (SAH) which reveals crucial insights into the mechanism of catalytic inactivation. Moreover, the study accentuates that the protein undergoes a transition from an apo (open) to a holo (closed) state in the presence of various cofactors. However, the occurrence of interdomain dynamicity, expanding the domain interface, is exclusive in the presence of SAM or SAH and proves to be critical for substrate positioning, revealing crucial insights into the mechanism of catalytic inactivation.

Graphical abstract



4.1. Introduction

Ribosomes are indispensable molecular machinery responsible for protein synthesis. Their biogenesis involves intricate processes, including the transcription of ribosomal

RNA (rRNA), pre-rRNA processing, rRNA chemical modification, and the simultaneous assembly of ribosomal proteins (Nierhaus et al., 1991). Among these processes, chemical modifications significantly contribute to the fine-tuning of protein translation. The chemical modifications are found mostly in the proximity of the decoding (DC) and peptidyl transferase center (PTC) of the ribosome. Several kinds of base methylation of the rRNAs such as N6-monomethyladenosine (m⁶A), N6,6-dimethyladenosine (m²6A), N1-methyladenosine (m¹A), 5-methylcytosine (m⁵C), 5-methyluridine (m⁵U), 3-methyluridine (m³U), and N2-methylguanosine (m²G) have been reported in the literature (Polikanov et al., 2015). Such rRNA methylations are performed by a diverse group of enzymes, referred to as rRNA methyltransferases (MTases) (Decatur and Fournier, 2002; Boccaletto et al., 2018). These enzymes possess a distinct capability to methylate nucleotides, whether they are buried within the core of the ribosome or exposed on the surface (Sergiev et al., 2007). The selective nature of these enzymes is attributed to the intricate tertiary structure of rRNA, which includes various bulges and loops, leading to distortions in the duplex structure.

Despite the complexity of the substrate rRNA, specific methylation sites have been conserved across all three domains of life. These conserved modifications offer valuable insights into the evolutionary path of rRNA methylation. One well-known example of a conserved modification is the dimethylation of two successive adenosine bases (A1518/936 and A1519/937, numbering as per the *E. coli*/*H. sapiens*) at the N6 position within the helix 45 (h45) of the 3'-terminal of the 16S/18S rRNA. In bacteria, the above modification is catalyzed by an enzyme, namely KsgA (also known as ribosomal RNA small subunit MTase A; RsmA) due to deletion of the gene (*ksgA*) leads to a modification in the ribosome binding site, resulting in resistance to the aminoglycoside antibiotic kasugamycin (Helser et al., 1971, 1972; Van Buul et al., 1984; Ochi et al., 2009). Conversely, in eukaryotes, the homolog enzyme named dimethyladenosine transferase 1 (DIME1) facilitates a similar modification (Lafontaine et al., 1994). These MTases require the cofactor S-adenosyl-L-methionine (SAM) as a methyl donor for their catalytic activity. The transfer of a methyl group from SAM to the substrate rRNA results in the formation of S-adenosyl-L-homocysteine (SAH) as the product (Van Buul and Van Knippenberg, 1985). This modification plays a crucial role in stabilizing the small ribosomal subunit (SSU) and facilitates ribosome

biogenesis by promoting the late-folding of helices h44 and h45 in rRNA (Demirci et al., 2010; Fischer et al., 2015). Thus, the dimethylase involved in the late quality-control checkpoint during the SSU biogenesis pathway requires methylating the assembled pre-30/40S particles and completing the final processing steps (Lafontaine et al., 1995; Connolly et al., 2008; Mangat and Brown, 2008).

Previous studies have suggested that the core rRNA MTase activity of KsgA is evolutionarily conserved in eukaryotes, as it can be complemented by its orthologs (DIMIT1) (O'Farrell et al., 2006). However, unlike DIMIT1, KsgA is unable to establish functional equivalence in eukaryotes due to an additional degree of complexity inherent in eukaryotic ribosome synthesis associated with a huge multifactor complex processome (Pulicherla et al., 2009). Additionally, two paralog KsgA/DIMIT1-like proteins, known as mitochondrial transcription factor B (TFB1M and TFB2M), have been reported within mammalian mitochondria. These proteins exhibit bifunctionality, serving both as transcription factors and rRNA MTases. Specifically, TFB1M has been recognized as a dimethylase that plays a crucial role in the modification of the 3'-end of mitochondrial 12S rRNA, explicitly adding the m⁶A modification (Cotney and Shadel, 2006; Metodiev et al., 2009; Sologub et al., 2009).

In addition to the housekeeping KsgA/DIMIT1 MTases, another group of rRNA N6,6-dimethyladenosine MTases exclusively contribute to antibiotic resistance by altering its binding site on the ribosome. For example, erythromycin resistance MTases (Erm) found in specific microorganisms confer resistance against the macrolide-lincosamide-streptogramin B (MLSB) antibiotics by methylating the N6 position of a specific adenine residue (A2058, in *E. coli*) in the 23S rRNA (Weisblum, 1995). Nevertheless, this superfamily (rRNA N6,6-dimethyladenosine) shares structural homology with a family of DNA N6-adenosine MTases, having a conserved N-terminal catalytic domain (NCD) with a Rossmann-fold and a C-terminal oligonucleotide-binding domain (Schluckebier et al., 1997; Yu et al., 1997; Schluckebier et al., 1999; O'Farrell et al., 2004).

The structural analysis of rRNA N6,6-dimethyladenosine MTases from *E. coli* (EcKsgA) and *Aquifex aeolicus* (AaKsgA) suggested that the protein does not bind

SAM efficiently in the absence of the substrate rRNA. On the other hand, KsgA from *Bacillus subtilis* (BsKsgA) and *Thermus thermophilus* (TtKsgA) demonstrate binding of the cofactor SAM (and its analog, 5'-methylthioadenosine (MTA)), in the absence of the substrate rRNA (Demirci et al., 2009; Bhujbalrao and Anand, 2019). Similarly, the crystal structure of DIMT1 from *H. sapiens* (HsDIMT1) demonstrates the binding of SAM in the absence of the substrate rRNA. In the case of archaea, X-ray crystallography has been employed to determine the structures of KsgA/DIMT1 in *M. jannaschii*. However, the dynamic behavior of archaeal proteins remains an enigma due to limitations in previous studies (O'Farrell et al., 2010).

In this study, the crystal structures of an archaeal DIMT1 (ORF id: PH1823, *Ph*DIMT1) from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 have been presented. The study encompasses both apo and various holo forms, facilitating a comprehensive comparison and contrast among them. The result suggests that molecular details of substrate positioning are critically maintained by the dynamic nature of the protein in the presence of cognate cofactors. This analysis aimed to deepen our insights into the dynamic characteristics of archaeal DIMT1. To achieve this goal, a comprehensive binding study was undertaken for *Ph*DIMT1_WT and its point mutants with various cofactors to capture different transition states, including catalytically inactive states, while assessing the flexible behaviors of the corresponding enzyme. Ultimately, utilizing a mechanistic model, we propose a novel intra-domain dynamicity solely associated with the cognate cofactor.

4.2. Materials and methods

4.2.1. Construction of wild-type and mutant expression plasmids

The recombinant construct of the full-length (873 bp) gene *Ph*DIMT1 (ORF id: PH1823) was obtained by molecular cloning in the pET-28a(+) plasmid vector. For cloning, the gene *Ph*DIMT1 was amplified using the polymerase chain reaction (PCR) method using the genomic DNA of *Pyrococcus horikoshii* OT3 as a template and the primers listed in Table 4.1 The *Ph*DIMT1 construct made use of the 6xHis-tag, which is positioned downstream of the ribosome binding site (RBS) in the pET-28a(+) vector. The PCR-amplified gene was double digested with *Nhe*I and *Hind*III restriction

enzymes before inserting into the pET-28a(+) vector excised with the same set of enzymes. The ligation of the inserted gene was performed using T4 DNA ligase (Figure 4.1A). Subsequently, *E. coli* DH5 α competent cells were transformed with the recombinant plasmid containing the gene *PhDIMT1* for plasmid amplification. The resulting wild-type construct was further used as a template to generate various mutants of *PhDIMT1* (*PhDIMT1_D117A*, *PhDIMT1_N132A*, and *PhDIMT1_Y135A*) with a Q5 site-directed mutagenesis kit (New England Biolabs, USA) using the primers mentioned in Table 4.1. The wild-type and mutant constructs of the *PhDIMT1* gene were confirmed by double digestion using the same set of restriction enzymes by incubating the vectors at 37°C for 2 hrs (Figure 4.1B-4.1D). This was further confirmed by plasmid sequencing.

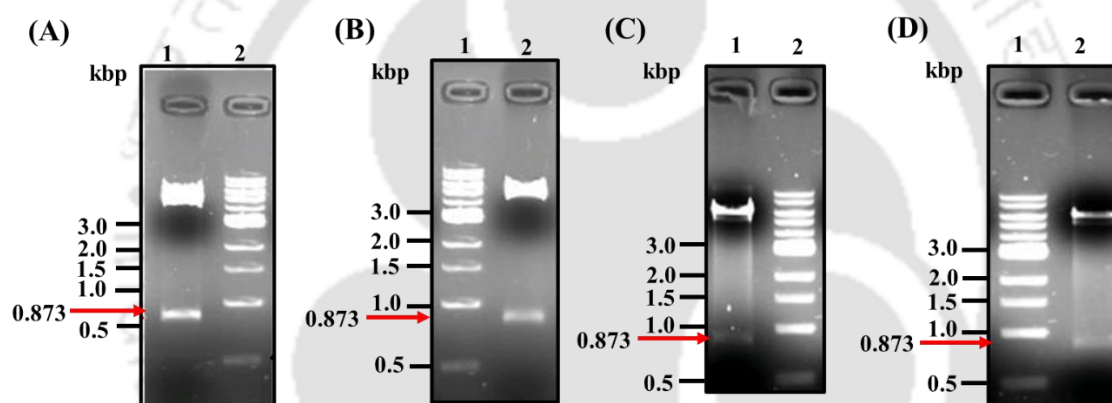


Figure 4.1. Cloning and site-directed mutagenesis of gene *PhDIMT1*. (A) Clone confirmation of *PhDIMT1_WT* by double digestion of the vector pET-28a(+) carrying the gene *PhDIMT1* by *NheI* and *HindIII* restriction enzymes (lane 1: insert pop out after double digestion of the plasmid pET-28a(+) lane 2: DNA ladder). Clone confirmation gel of all the *PhDIMT1* mutant constructs (B) *PhDIMT1_D117A* (lane 1: DNA ladder 2: insert pop out after double digestion) (C) *PhDIMT1_N132A* (lane 1: insert pop out after double digestion of the plasmid pET-28a(+) lane 2: DNA ladder) (D) *PhDIMT1_Y135A* (lane 1: DNA ladder 2: insert pop out after double digestion).

Table 4.1. List of oligonucleotide sequences used to construct the wild-type (*PhDIMT1_WT*) and mutant (*PhDIMT1_D117A*, *PhDIMT1_N132A*, *PhDIMT1_Y135A*) protein expression plasmids. Restriction sites are highlighted in bold, underlined, and in

uppercase. The altered codon for the alanine substitution is indicated as underlined, in bold, and in lowercase.

Plasmid name	Primer	Oligonucleotide sequence (5'-3')
<i>PhDIMT1</i> _WT	Forward (<i>NheI</i>)	GTAG <u>GCTAGC</u> ATGA GCTCCAGGATAAGGA
	Reverse (<i>HindIII</i>)	GCGCA <u>AAGCTT</u> CTAGCTTATTATTCCA TGCTT
<i>PhDIMT1</i> _D117A	Forward	AATTCAAGGAG <u>c</u> TGCCGTGAGAG
	Reverse	ATTTTAACATTGTTCCAGCTG
<i>PhDIMT1</i> _N132A	Forward	GGTTGTCTCC <u>gc</u> TATTCCATATAAGATATC
	Reverse	TTGTTAAATTTTCGGCCATTC
<i>PhDIMT1</i> _Y135A	Forward	CAATATTCCAG <u>gc</u> TAAGATATCATCGCCG
	Reverse	GGGCACGGCGGTGGCCTG

4.2.2. Protein overexpression and purification

The *E. coli* Rosetta (DE3) competent cells (Novagen, India) were transformed with the recombinant constructs for wild-type *PhDIMT1* (*PhDIMT1_WT*) and mutants (*PhDIMT1_D117A*, *PhDIMT1_N132A*, and *PhDIMT1_Y135A*) for their overexpression. The transformed cells were grown at 37°C in Luria-Bertani (LB) medium supplemented with 50 µg mL⁻¹ kanamycin and 34 µg mL⁻¹ chloramphenicol until an optical density (OD) of 0.6–0.8 at 600 nm was obtained. Subsequently, 3% alcohol was added to the culture 30 min before the induction with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were further grown for an additional 16 hrs at 20°C post-induction. The cells were harvested by centrifugation at 3836xg for 15 min and resuspended in a lysis buffer consisting of 50 mM HEPES pH 7.5, 5 mM imidazole,

300 mM NaCl, 5% glycerol, and 1X protease inhibitor cocktail (PIC) at room temperature. The cells were lysed by sonication with a cycle of 3 s on and 5 s off at 30% amplitude. As an initial purification step, the cell lysate was heated at 60°C for 15 min. The heated sample was centrifuged at 15644xg for 45 min at 4°C to separate out the soluble and insoluble fractions. Since the protein of interest was present in the soluble fraction, it was loaded onto a pierce centrifuge column (Thermo Fisher Scientific, USA) packed with nickel-nitrilotriacetic acid (Ni-NTA) affinity resin (Qiagen, Germany) pre-equilibrated with the lysis buffer. After 3 hrs of incubation at 4°C, the column was washed with ten column volumes each of buffer A consisting of 50 mM HEPES pH 7.5, 20 mM imidazole, 300 mM NaCl, 5% glycerol, and 1X PIC and buffer B consisting of 50 mM HEPES pH 7.5, 50 mM imidazole, 300 mM NaCl, 5% glycerol, and 1X PIC. To improve protein stability, 50 mM of L-arginine and L-glutamic acid were added to the elution buffer (Buffer C), which also contained 50 mM of HEPES (pH 7.5), 250 mM of imidazole, 300 mM of NaCl, 5% glycerol, and 1X PIC (Golovanov et al., 2004). The eluted fractions were pooled and subjected to gradient dialysis to remove imidazole against the final protein buffer D consisting of 50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol. The dialyzed proteins *PhD1MT1_WT*, *PhD1MT1_D117A*, *PhD1MT1_N132A*, and *PhD1MT1_Y135A* were concentrated in a Vivaspin Turbo 15 centricon (10 kDa cutoff; Sartorius, Germany) to concentrations of ~14, ~10, ~12, ~39 mg mL⁻¹, respectively (Figure 4.2).

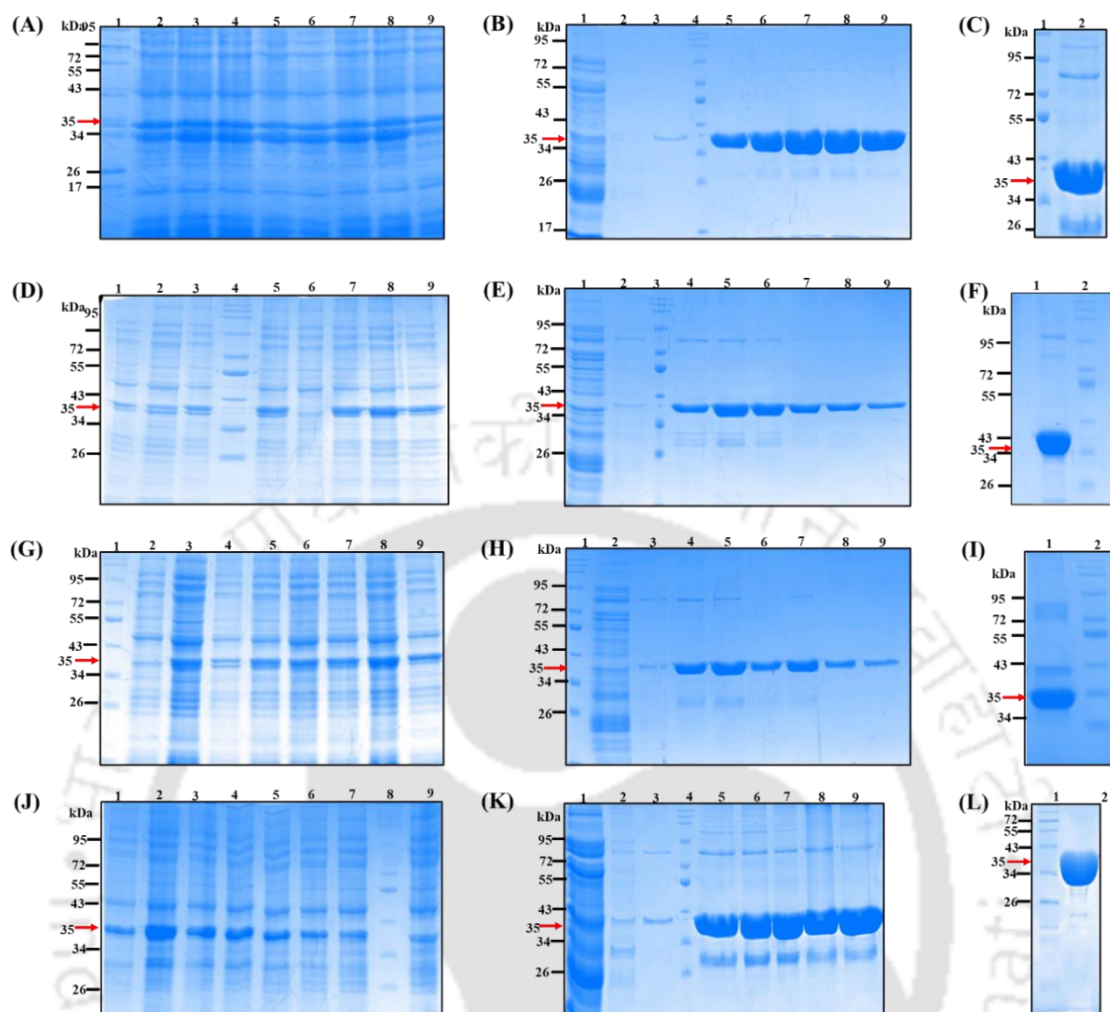


Figure 4.2. Purification of *PhDIMIT1_WT* and its mutant proteins. A) Overexpression of the protein *PhDIMIT1_WT* in *E. coli* Rosetta (DE3) at 1 mM IPTG at a different time interval (lane 1: protein marker, lane 2: uninduced or 0 hrs, lane 3-9: 1-7 hr, respectively) after induction. (B) SDS-PAGE analysis for the eluted fractions of the protein using Ni-NTA metal-affinity chromatography (lanes 1-3, Unbound, Flow through 1 and 2 & 5-9: eluted fractions of the protein respectively, lane 4: protein marker). (C) SDS-PAGE analysis for the concentrated protein (lane 1: protein marker, lane 2: concentrated protein). The gene and the protein of interest are marked by a red arrow. (D-F), (G-I), and (J-L) purification profiles of the mutant *PhDIMIT1_D117A*, *PhDIMIT1_N132A*, and *PhDIMIT1_Y135A*, respectively.

4.2.3. Mass spectrometric (MS) analysis

The molecular mass of *PhDIMIT1_WT* was estimated using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry analysis. The

homogeneously purified protein samples were mixed with the matrix (sinapinic acid) in a (protein:matrix) ratio of 1:2, 1:3, and 1:5. The sinapinic acid matrix was prepared to a final concentration of 10 mg mL⁻¹ dissolved in a solvent containing 100% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA) in a 30:70 ratio, respectively. Subsequently, the sinapinic acid was dissolved in the solvents by sonication for 30 min, followed by centrifugation at 12000xg for 10 min to eliminate the excessive matrix particles. The final prepared sinapinic acid matrix was mixed with the protein samples at 10, 15, and 20 μ M concentrations in the ratios specified above. A 2 μ L spot of the protein-matrix mixture was made on the ground-steel MALDI plate grooves and air-dried. Post-drying, the drops were ionized at a Bruker autoflex speed (Bruker Daltonics, Germany), and all data were analyzed using the built-in flex control analysis software. Graphs were plotted using the program Origin v9.6 for final representation.

4.2.4. Dynamic light scattering (DLS) analysis

To assess the aggregation tendency of the purified protein *PhD1MT1_WT*, hydrodynamic diameters were estimated using dynamic light scattering (LITESIZER 500; Anton Paar, UK). The purified protein sample was diluted in nuclease-free water (Himedia, India) with a final concentration of $\sim$ 1 μ M. Following this, the DLS spectra of the protein sample were obtained by ramping the temperature from 0.1°C to 10°C in a step-wise increment. Subsequently, the spectra were collected over a pre-defined time interval at a temperature range of 20-70°C in a cuvette with a 1 cm path length. For the final representation, the hydrodynamic diameter was plotted in a two-dimensional graph as a function of temperature concerning different wavelengths using the program Origin v9.6.

4.2.5. Fluorescence spectroscopy (FS) analysis

To estimate the binding of cofactor(s) such as S-adenosyl-L-methionine (SAM), S-adenosyl-L-homocysteine (SAH), and adenosyl-L-ornithine (SFG) to the protein, fluorescence spectroscopy experiments were performed. For this, the protein *PhD1MT1_WT* (10 μ M) was incubated with increasing concentrations (1-100 μ M) of the cofactors. The protein and ligand molecules were prepared in buffer D containing 50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol. Before each measurement, samples were incubated at 25°C for 1 hr and subjected to steady-state fluorescence

measurements. The measurements were performed in a 10 mm path-length quartz cuvette for protein and protein-cofactor complex at room temperature in a FluoroMax-4 fluorimeter (Horiba Scientific, Japan). Both the protein and protein-cofactor complexes were excited at a fixed excitation wavelength of 285 nm with a slit width of 5 nm and monitored for emission spectra from 300 to 500 nm. Three averaging scans were recorded and subtracted for each measurement with the blank for the actual fluorescence. Ligand binding was measured by estimating the quenching values of protein fluorescence attributed to the energy transfer. Further, bovine serum albumin (BSA) was used as a negative control, which revealed no discernible shifts in fluorescence spectra with increasing concentration (1-100 μ M) of the cofactor SAM.

4.2.6. Crystallization of wild-type and mutant *PhD*MT1

Preliminary crystallization screening of the purified protein *PhD*MT1_WT (~ 10 mg mL⁻¹) was performed using the buffers available in the crystallization screens from Hampton Research (Aliso Viejo, USA) by mixing 2 μ L each of the protein and buffer solutions (1:1 ratio) using the microbatch-under-oil technique at 4°C and 20°C. Initial crystal hits of *PhD*MT1 were obtained within 2-3 weeks in two conditions: (A) 0.2 M zinc acetate dihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 18% (w/v) PEG 8000, and (B) 0.2 M zinc acetate dihydrate pH 6.4 and 20% (w/v) PEG 3350 at 20°C. However, the morphology of the obtained crystals resembled thin plates that were stacked over each other. Thus, the crystals were further optimized by increasing the protein concentration (~ 14 mg mL⁻¹) and prolonging incubation time ($\sim$ nine months) in the microbatch-under-oil at 20°C. Simultaneously, the crystals of the binary complexes (*PhD*MT1_WT with SAM, SAH, and SFG) were obtained in the same conditions by co-crystallization employing the microbatch-under-oil technique at 20°C (Figure 4.3A- 4.3F).

As the protein *PhD*MT1_WT is hypothesized to methylate rRNA substrates containing architectures similar to the h45 (5'-UAGCCGUAGGGGAACCUACGGCUC-3') of 16S rRNA, the h45 rRNA was procured from integrated DNA technologies (Coralville, USA) at a 1 μ mol scale. Subsequently, ternary complexes (*PhD*MT1_WT with SAM and rRNA) were prepared by heating the rRNA sample at 75°C for 5 min and cooled on ice for 10 min before mixing with the protein solution. This was followed by heating

the protein-rRNA mixture at 60°C for 5 minutes, slowly cooling it down to room temperature, and allowing it to stand for 45 minutes before mixing with SAM. The ternary complex was set up for crystallization with the protein *PhD*IMT1_WT (~14 mg mL⁻¹ in buffer C) mixed with the unmodified h45 rRNA fragment and SAM at a molar ratio of 1:1.2:2. Although crystals of the ternary complex were obtained, they failed to produce collectable X-ray diffraction. In parallel, crystals of the apo form of the mutants (*PhD*IMT1_D117A, *PhD*IMT1_N132A) and complexes (*PhD*IMT1_Y135A•SAM and *PhD*IMT1_Y135A•SFG) were grown using the same crystal condition composition A. However, the *PhD*IMT1_D117A•SAM complex was crystallized in condition B using the microbatch-under-oil diffusion techniques at 20°C (Figure 4.3G-4.3K). X-ray intensity data were collected from all the crystals obtained at 20°C using a home-source X-ray diffractometer. However, the crystals of the mutant (*PhD*IMT1_Y135A) could not produce collectable X-ray diffraction.

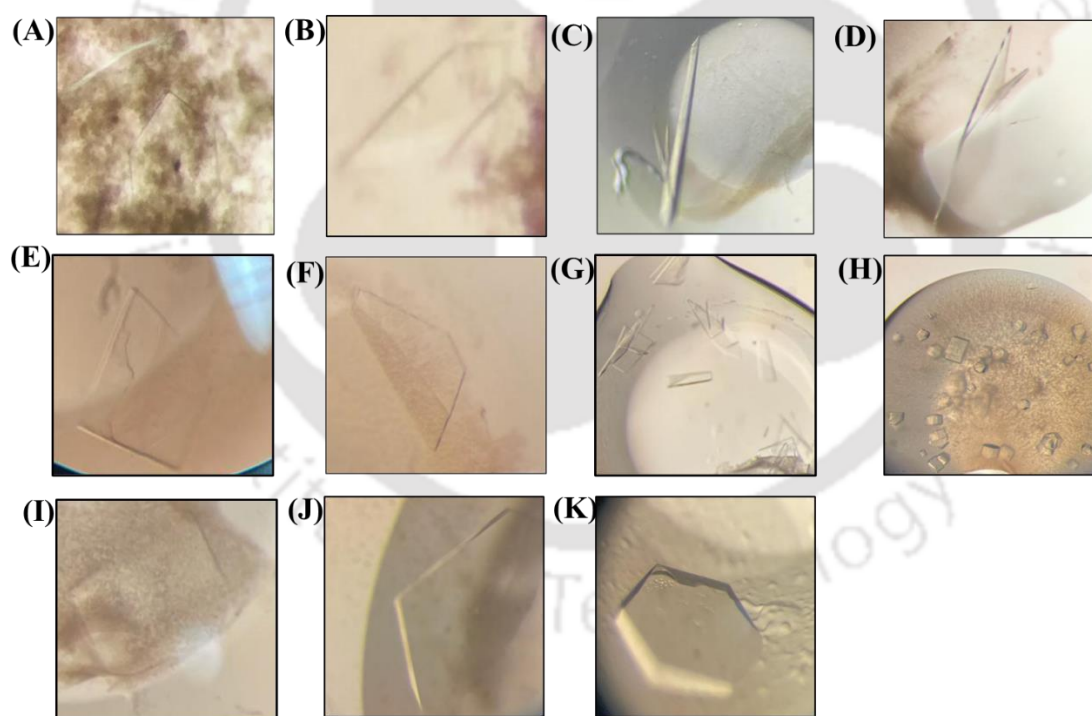


Figure 4.3. Overview of the crystals of *PhD*IMT1_WT and the mutants in its apo and complex form. Crystallographic hits were obtained for the wild-type protein *PhD*IMT1_WT in various states: (A) apo and holo forms bound with (B) SAM, (C) SAH, and (D) SFG, as well as (E, F) when soaked with the ligand SAM. Crystal image of the mutant *PhD*IMT1_D117A includes (G) the apo form and (H) the SAM-bound

form. Similarly, Crystal hits of the mutant *PhDIMIT1_N132A* are depicted in its (I) apo and holo forms of the mutant *PhDIMIT1_Y135A* bound with (J) SFG and (K) SAM.

4.2.7. Data collection, processing, and structure determination

X-ray intensity diffraction data for the *PhDIMIT1_WT* and mutant protein crystals were collected at -173°C using Cu K α radiation at a wavelength of 1.5418 Å generated by a home-source Rigaku MicroMax-007 HF diffractometer operated at 40 kV and 30 mA and an R-Axis IV++ image-plate detector available at the central instruments facility (CIF) of Indian Institute of Technology Guwahati, India. The X-ray intensity data sets for all the crystals were collected with 1° oscillation and 300 s exposure time while keeping the detector in a distance range of 150-210 mm. Processing and scaling of all of the X-ray intensity data were performed using the programs iMosflm (Battye et al., 2011) and AIMLESS (Evans and Murshudov, 2013), and the intensities were converted to structure factors using the ctruncate module in the CCP4 package (Evans, 2011; Winn et al., 2011). The structure solution of *PhDIMIT1_WT* was obtained using the molecular replacement method using the program Phaser (McCoy et al., 2007). The AlphaFold model of the protein PH1823 was used as a search template in the molecular replacement. Subsequently, wild-type binary complexes (*PhDIMIT1_WT* with SAM, SAH, and SFG) and the structures of the mutant proteins both in apo (*PhDIMIT1_D117A*, *PhDIMIT1_N132A*) and holo (*PhDIMIT1_D117A*•SAM, *PhDIMIT1_Y135A*•SAM, and *PhDIMIT1_Y135A*•SFG) forms were determined by molecular replacement using the three-dimensional atomic coordinates of *PhDIMIT1_WT* as a search model. Before structure refinement, a total of 5% of the reflections were kept aside from the data set for R_{free} calculations (Brunger, 1992). For each structure model, the refinement was performed using the program Refmac5 v5.8.0267 (Murshudov et al., 2011), followed by model building manually using the program Coot v0.9.4.1 (Emsley et al., 2010). Although a clear difference (F_o-F_c) electron density map was observed for the bound ligands after the first cycle of refinement, the protein atoms were firstly fitted into electron density contoured at 3.0 σ and 1.0 σ for the F_o-F_c and 2F_o-F_c maps, respectively. Subsequently, the electron densities observed for water and other molecules in the protein and crystallization buffer were modeled. Structure refinement was performed in each cycle of model

building using the same set of parameters. Finally, the electron densities observed for the bound ligands were fitted, and the models were refined. The geometrical parameters for all of the refined structures were checked and validated using the web tools PROCHECK (Laskowski et al., 1993) and MolProbity (Williams et al., 2018). The refinement and validation statistics for all refined models are provided in Tables 4.2-4.3. The three-dimensional atomic coordinates and structure factors of all of the refined structures have been deposited in the RCSB Protein Data Bank (Berman et al., 2002). The molecular graphic figures used in this study were prepared using the program PyMOL v2.3.4 (The PyMOL Molecular Graphics System, Schrodinger, LLC).

Table 4.2. X-ray crystallographic data collection and refinement statistics for the protein *PhDIMIT1_WT* (apo and holo forms). The values in parenthesis represent the statistics for the last resolution shell. MTA: 5'-methylthioadenosine, ADN: adenosine, SFG: adenosyl-L-ornithine, and SAH: S-adenosyl-L-homocysteine.

	<i>PhDIMIT1_WT</i>	<i>PhDIMIT1_WT</i> •MTA•ADN	<i>PhDIMIT1_WT</i> •MTA_FormI	<i>PhDIMIT1_WT</i> •MTA_FormII	<i>PhDIMIT1_WT</i> •SFG	<i>PhDIMIT1_WT</i> •SAH
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100	100	100
Space group	C2	C2	P22 ₁ 2 ₁	C2	C2	C222 ₁
Unit-cell parameters (Å, °)	<i>a</i> =117.86, <i>b</i> =80.17, <i>c</i> =84.64, β =115.50	<i>a</i> =120.38, <i>b</i> =77.26, <i>c</i> =94.22, β =119.95	<i>a</i> =79.78, <i>b</i> =85.34, <i>c</i> =106.81	<i>a</i> =120.66, <i>b</i> =80.08, <i>c</i> =85.74, β =116.24	<i>a</i> =119.15, <i>b</i> =80.34, <i>c</i> =84.11, β =115.46	<i>a</i> =81.48, <i>b</i> =100.27, <i>c</i> =77.89

**Chapter 4. Structural and functional characterization of archaeal m₂⁶A rRNA
MTase**

Resolution (Å)	76.39-2.01 (2.06-2.01)	81.64-2.35 (2.43-2.35)	85.34-2.60 (2.72-2.60)	76.90-2.50 (2.72-2.50)	75.94-2.20 (2.27-2.20)	63.23-2.70 (2.83-2.70)
No. of observed reflections	225549 (15354)	118049 (10963)	189170 (22136)	93512 (11504)	161859 (13702)	77412 (9992)
No. of unique reflections	47378 (3435)	31324 (3015)	23103 (2756)	22648 (2759)	36508 (3157)	8774 (1118)
Mn(I) CC(1/2)	0.995 (0.727)	0.998 (0.950)	0.994 (0.949)	0.980 (0.899)	0.994 (0.866)	0.937 (0.891)
Completeness (%)	99.9 (98.8)	99.9 (99.1)	100 (100)	99.8 (100)	100 (100)	97.3 (95.8)
$V_M (\text{\AA}^3 \text{Da}^{-1})$	2.52	2.65	2.54	2.60	2.54	2.24
Solvent content (%)	51.28	53.68	51.64	52.80	51.62	45.03
Mosaicity (°)	0.80	0.60	0.60	0.60	0.80	0.60
Mean I/ σ (I)	10.4 (2.7)	18.7 (6.6)	9.2 (3.3)	7.9 (2.9)	6.3 (2.0)	9.0 (3.4)
$R_{\text{merge}}^{\dagger}$ (%)	8.5 (49.4)	4.7 (16.8)	14.2 (52.8)	11.6 (42.2)	9.0 (50.0)	14.5 (63.0)
$R_{\text{pim}}^{\dagger\dagger}$ (%)	6.9 (39.9)	4.5 (15.7)	7.7 (29.2)	10.4 (35.6)	8.3 (42.8)	7.9 (32.6)
$R_{\text{meas}}^{\dagger\dagger\dagger}$ (%)	11.0 (63.8)	6.6 (23.1)	16.2 (60.4)	15.7 (55.5)	12.9 (66.1)	16.5 (71.1)

Chapter 4. Structural and functional characterization of archaeal m₂⁶A rRNA MTase

Multipli city	4.8 (4.5)	1.7 (1.9)	8.2 (8.0)	4.1 (4.2)	4.0 (4.0)	8.8 (8.9)
R _{work} /R _{fr} ee (%)	19.46/24. 56	16.48/20.5 4	20.52/27.26	23.11/28.21	21.30/2 6.10	18.43/27.6 0
Protein model						
No. of subunits in ASU	2	2	2	2	2	1
Protein atoms	4395	4378	4380	4389	4383	2156
Water molecul es	338	381	197	152	235	35
Ligand (number)	-	MTA (2), ADN (2)	MTA (2)	MTA (2)	SFG (2)	SAH (1)
Other molecul es	36	36	28	17	14	8
Deviation from ideal geometry						
Bond length (Å)	0.010	0.010	0.009	0.008	0.008	0.007
Bond angles (°)	1.455	1.516	1.585	1.546	1.420	1.389
Average B-factor (Å²)						
Protein atoms	25.95	25.33	25.39	33.24	27.33	47.68
Water molecul es	42.70	38.45	36.29	41.66	43.55	44.86

**Chapter 4. Structural and functional characterization of archaeal m₂⁶A rRNA
MTase**

Ligands	-	MTA (24.60), ADN (41.37)	MTA (35.76)	MTA (37.17)	SFG (28.90)	SAH (56.98)
Ramachandran plot (%)						
Favored	97.94	97.38	95.45	94.76	93.81	93.85
Allowed	2.06	2.62	4.55	5.24	6.19	6.15
PDB id	8X3W	8X41	8X44	8X45	8X46	8X47

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum overall reflections and $\sum_i$ is the sum over i measurements of reflection hkl .

^{††} $R_{\text{pim}} = \sum_{hkl} \sqrt{\frac{1}{n-1} \sum_{j=1}^n} |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$, where $I(hkl)$ is the intensity of reflection hkl .

^{†††} $R_{\text{meas}} = \sum_{hkl} \sqrt{\frac{n}{n-1} \sum_{j=1}^n} |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$, where $I(hkl)$ is the intensity of reflection hkl .

Table 4.3. X-ray crystallographic data collection and refinement statistics for the mutant proteins of *PhDIMT1* (apo and holo forms). The values in parenthesis represent the statistics for the last resolution shell. MTA: 5'-methylthioadenosine, and SFG: adenosyl-L-ornithine.

	<i>PhDIMT1_</i> D117A	<i>PhDIMT1_D1</i> 17A•MTA	<i>PhDIMT1_</i> N132A	<i>PhDIMT</i> 1_ Y135A• MTA	<i>PhDIMT1_Y1</i> 35A•SFG
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418	1.5418
Temperature (K)	100	100	100	100	100
Space	<i>C2</i>	<i>P6</i>	<i>C2</i>	<i>P3</i>	<i>C2</i>

**Chapter 4. Structural and functional characterization of archaeal m²6A rRNA
MTase**

group					
Unit-cell parameter s (Å, °)	$a=118.17$, $b=80.63$, $c=84.82$, $\beta=114.86$	$a=b=124.92$, $c=36.81$	$a=116.22$, $b=80.55$, $c=84.38$, $\beta=114.94$	$a=b=125$.54, $c=37.14$	$a=119.06$, $b=79.96$ $c=85.41$, $\beta=116.28$
Resolution (Å)	76.96-3.30 (3.56-3.30)	62.46-3.30 (3.57-3.30)	76.51-2.60 (2.72-2.60)	62.77- 3.00 (3.18- 3.00)	76.58-2.60 (2.72-2.60)
No. of observed reflections	47267 (9214)	38749 (7429)	102825 (12471)	415172 (18698)	91861 (9752)
No. of unique reflections	10317 (1995)	4814 (907)	21707 (2644)	6617 (1117)	20945 (2200)
Mn(I) CC(1/2)	0.954 (0.952)	0.984 (0.950)	0.992 (0.844)	0.995 (0.960)	0.939 (0.846)
Completeness (%)	93.6 (87.7)	93.2 (87.2)	99.5 (98.9)	94.9 (100.0)	94.3 (82.1)
V_M (Å ³ Da ⁻¹)	2.56	2.32	2.50	2.36	2.55
Solvent content (%)	52.04	46.98	50.90	47.97	51.76
Mosaicity (°)	0.80	0.45	0.40	0.50	1.10
Mean I/ σ (I)	6.1 (4.5)	9.1 (4.2)	6.5 (2.4)	13.7 (6.5)	5.8 (2.1)
$R_{\text{merge}}^{\dagger}$ (%)	18.7 (25.3)	13.3 (46.5)	11.5 (42.3)	9.7 (44.9)	16.5 (49.3)

**Chapter 4. Structural and functional characterization of archaeal m₂⁶A rRNA
MTase**

R _{pim} ^{††} (%)	15.2 (20.2)	7.2 (25.5)	8.2 (31.1)	4.3 (19.8)	13.8 (40.3)
R _{meas} ^{†††} (%)	24.2 (32.5)	15.1 (53.3)	14.2 (52.9)	10.7 (49.1)	21.6 (63.9)
Multiplici ty	4.6 (4.6)	8.0 (8.2)	4.7 (4.7)	11.4 (11.7)	4.4 (4.4)
R _{work} /R _{free} (%)	24.84/29.88	20.10/28.65	18.79/23.94	21.95/30. 44	26.23/28.91
Protein model					
No. of subunits in ASU	2	1	2	2	2
Protein atoms	4384	2186	4384	4364	4376
Water molecule s	13	6	115	21	102
Ligand (number)	-	MTA (1)	-	MTA (2)	SFG (2)
Other molecule s	-	1	21	4	4
Deviation from ideal geometry					
Bond length (Å)	0.006	0.004	0.007	0.012	0.008
Bond angles (°)	1.321	1.211	1.443	1.134	1.567
Average B-factor (Å²)					
Protein atoms	2.65	85.63	31.70	70.60	25.19
Water	6.78	71.53	40.90	71.33	41.57

**Chapter 4. Structural and functional characterization of archaeal m²6A rRNA
MTase**

molecules					
Ligands	-	MTA (116.82)	-	MTA (95.96)	SFG (48.68)
Ramachandran plot (%)					
Favored	90.11	86.52	96.46	90.82	94.03
Allowed	9.89	13.48	3.54	9.18	5.97
PDB id	8X4G	8X4I	8X4L	8X4O	8X4P

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum overall reflections and $\sum_i$ is the sum over i measurements of reflection hkl .

^{††} $R_{\text{pim}} = \sum_{hkl} \sqrt{\frac{1}{n-1} \sum_{j=1}^n} |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$, where $I(hkl)$ is the intensity of reflection hkl .

^{†††} $R_{\text{meas}} = \sum_{hkl} \sqrt{\frac{n}{n-1} \sum_{j=1}^n} |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$, where $I(hkl)$ is the intensity of reflection hkl .

4.2.8. Thermal denaturation studies using circular dichroism (CD) spectroscopy

Temperature-dependent conformational changes in the secondary structure of the wild-type and the mutant proteins of *PhDIMT1* were estimated by measuring the Far-UV spectra between the wavelength of 190 to 260 nm with a 1 mm optical path-length quartz cuvette using a J-1500 circular dichroism spectrometer JASCO J-1500 (JASCO, Japan). Each protein sample (5.0 μM) was dialyzed against the final protein buffer E consisting of 1 mM Tris-HCl pH 7.5 and 5 mM NaCl. The parameters response, sensitivity, and scan speed were fixed at default values 2 s, 100 meg, and 100 nm min⁻¹, respectively, for spectra measurement. Each experimental run includes the three-repeat scans followed by averaging to generate the final spectra with respect to a comparative blank. The final spectra with these sets of parameters were further used to probe the thermal denaturation profile of the wild-type and the mutant proteins of *PhDIMT1* over the temperature range of 20-110°C, heating rate of 2°C with 10°C of interval, and wavelength range of 190-260 nm. The denaturation analysis of the protein was

performed using the spectra manager software. To determine the melting point (T_m) using CD, ellipticity at 222 nm was measured over a range of temperatures. The apparent fraction of unfolded molecules (F_U) was determined using the following equation:

$$F_U = \frac{(Y_{obs} - Y)}{(Y_U - Y)}$$

where Y_{obs} is the value of ellipticity (θ) at 222 at each temperature, and Y_U and Y are the observed values in the baseline regions for the unfolded and folded forms, respectively (Singh and Manoj, 2017). The graph was plotted for fraction unfolded (F_U) against temperature utilizing the Boltzmann sigmoid (two-state model) equation using the program Origin v9.6.

4.2.9. Isothermal titration calorimetry (ITC) analysis

The binding affinity of the cofactors (SAM, SAH, and SFG), including MTA and substrate h45 with the protein *PhDIMIT1_WT*, was measured by isothermal titration calorimetry MicroCal ITC200 (GE Healthcare, USA) experiments. The protein, cofactors, and substrate rRNA were prepared in the same dialyzed buffer D containing 50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol to minimize the heat of dilution. The protein *PhDIMIT1_WT* (60-200 μ M) was filled in the sample cells, which were titrated with different cofactors (4-14 mM) and rRNA (1.5 mM) mentioned in Table B.1. Titrations were performed at 25°C with stirring at 250 rpm, and ligands filled in the syringe were injected 20 times each of 2 μ L (excluding 0.4 μ L pre-injection) of injection volume at an interval of 120 s. A similar protocol was followed to measure the binding affinity of different cofactors with the mutant proteins (*PhDIMIT1_D117A*, *PhDIMIT1_N132A*, and *PhDIMIT1_Y135A*) at 25 °C. Heat changes associated with the dilution of ligands in the buffer were also measured, which were subtracted from the heat of reactions to achieve the actual heat associated with protein-ligand binding. The subtracted data were analyzed using the program Origin v9.6. The thermodynamic parameters such as stoichiometric ratio (the number of binding sites, n), association constant (K_a), enthalpy (ΔH), and entropy (ΔS) changes were estimated by integrating the reaction heat using non-linear regression and one-site binding model. The thermodynamic parameter Gibbs free energy change (ΔG) was

calculated using the thermodynamic equation $\Delta G = \Delta H - T\Delta S$, where T is the reaction temperature. The estimated values for these thermodynamic parameters of each reaction are provided in Table B.1

4.2.10. Methylation assays

The assay was conducted using the MTase-Glo kit (Promega, USA) in 96-well white plates with a solid bottom. The MTase-Glo test is a bioluminescence-based assay that detects the product SAH through a luciferase reaction, which was measured utilizing the Tecan microplate reader (Zürich, Switzerland). In the assay, a 50 μ L reaction buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM ethylenediamine tetraacetic acid (EDTA), 3 mM MgCl₂, 0.1 mg mL⁻¹ bovine serum albumin (BSA), and 1 mM dithiothreitol (DTT)) containing a constant amount of substrate h45 (5 μ M) and cofactor (10 μ M) were combined with increasing amounts of the enzyme *PhDIMT1_WT*. The reaction was allowed to incubate for 30 min at 37°C, and the enzymatic activity (EC₅₀, which represents the amount of enzyme required for 50% maximal activity) was calculated using the equation:

$$Y = \text{Min} + \frac{\text{Max} - \text{Min}}{1 + \left(\frac{X}{\text{EC}_{50}}\right)^{\text{Hill coefficient}}}$$

where Y is the difference in luminosity (net relative luminescence unit, RLU) observed after subtracting the RLU values at each point of enzyme concentration from the control (no-enzyme), the Max and Min values represent the highest and lowest activity of the enzyme at constant substrate concentration, the Hill coefficient reflects the slope of the curve, and X is increasing enzyme concentrations. After the optimal enzyme concentration for a reaction was determined, the Michaelis-Menten constant (K_M) of the substrate h45 was assessed at increasing concentrations of h45 ranging from 1-5 μ M at fixed SAM concentration (10 μ M) using the following equation:

$$V_o = \frac{V_{\text{max}} [S]}{K_M + [S]}$$

where V_0 and $V_{\max}$ are the initial and maximum reaction velocities, and $[S]$ is the substrate concentration (μM). TFA (0.5%) was used to stop the reaction (Hsiao et al., 2016). The $V_{\max}$ (RLU min^{-1}) of the reaction was determined by dividing the RLU measured at the end of the reaction with the incubation time (30 min). Subsequently, the turnover number (K_{cat}) and catalytic efficiency (K_{cat}/K_M) of the enzyme *PhDIMT1*_WT were also calculated. The kinetic values of the mutants (*PhDIMT1*_D117A, *PhDIMT1*_N132A, and *PhDIMT1*_Y135A) were obtained using a similar protocol.

4.2.11. Reverse phase high-performance liquid chromatography (RP-HPLC) analysis

The ability of SAM hydrolysis was assayed in a reaction mixture containing 50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol. In condition I, the sample was incubated at 25°C for 24 hours, while in condition II, it was incubated at the same temperature for ~4 weeks. Each sample was subjected to RP-HPLC by injection onto a C18 column (YMC, Tokyo) equilibrated with 0.1% TFA followed by a linear gradient of 0-5% ACN in 0.1% TFA over 4 min at 1 mL min⁻¹ flow rate. After that, the isocratic flow of 5% ACN was maintained in 0.1% TFA for 18 min. Standard samples of freshly-prepared SAM (0.5 mM), SAH (0.5 mM), MTA (0.5 mM), and ADN (1 mM) were detected by absorbance at 254 nm.

4.2.12. Bioinformatic analysis

The open reading frames (ORFs) coding for rRNA MTases were collected from UniProtKB databases using specific keywords such as "RsmA," "KsgA," "Dim1," "DIMT1," and "dimethylase" (The UniProt Consortium, 2023). Their homologous proteins were identified using the web tool BLAST, keeping a sequence identity (SI) cut-off of more than 20% (Altschul et al., 1990). To establish evolutionary relationships between the targeted protein and its homologs, phylogenetic trees were constructed using the Neighbor-Joining (NJ) method in MEGAX, with 1000 bootstrap replicates (Kumar et al., 2018). The resulting trees were visualized using the web tool iTOL v6.3 (Letunic and Bork, 2021). Multiple sequence alignment (MSA) of the selected proteins and their homologs was performed using the web tool Clustal Omega (Sievers and Higgins, 2014). Further, MSA was rendered using the web tool ESPrpt v3.0 (Gouet et al., 2003). To identify the structurally conserved or invariant water (IW) molecules, the

flexible (mobile) structures of all the classes of rRNA N6,6-dimethyladenosine were superimposed over the reference (fixed) structure of that class employing the module align of the program PyMOL with the help of a home-built shell script. A water molecule of the reference structure having a similar pair in all the mobile structures of rRNA N6,6-dimethyladenosine MTases with low B-factor was included in the analysis. The B-factor of each IW molecule was extracted from its respective crystal structure. The interdomain movement of the proteins was predicted using the web server DynDom (Hayward and Berendsen, 1998). The 16S rRNA sequence of *P. horikoshii* was obtained from RNA central (RNAcentral Consortium, 2021). To identify the conserved h45 in *P. horikoshii*, MSA was performed using h45 sequences across three domains of life. The secondary structure of h45 was built using the program R2DT and validated with Mfold (Zuker, 2003; Sweeney et al., 2021). Subsequently, the three-dimensional structure of h45 was modeled and docked using the web server ModeRNA and HDOCK, respectively (Rother et al., 2011; Yan et al., 2020). The molecular interactions between the protein and the substrate rRNA (h45) were observed using the program Coot (Emsley et al., 2010).

4.3. Results

4.3.1. PH1823 is DIMT1/KsgA ortholog present in *P. horikoshii*

In the UniProtKB database, the ORF id PH1823 (UniProt id: O59487) from *Pyrococcus horikoshii* OT3, a hyperthermophilic archaeon, has been annotated as "probable ribosomal RNA small subunit MTase A". To further annotate the protein PH1823, a homology search was performed against the UniProtKB/SwissProt database using the program BLAST with the default set of parameters. The results show that protein PH1823 exhibits similarities with rRNA N6,6-dimethyladenosine MTases, specifically DIMT1 from *H. sapiens* (QC: 73% and SI: 41%), KsgA from *E. coli* (QC: 70% and SI: 33%), and Erm from *B. subtilis* (QC: 60% and SI: 29%); QC: query coverage and SI: sequence identity. Notably, these rRNA N6,6-dimethyladenosine MTases share structural similarities with DNA N6-adenosine MTases due to the presence of a Rossmann fold in the N-terminal catalytic domain (NCD). In the catalytic process of DNA N6-adenosine MTases, the catalytic pocket can only accommodate a single methyl group. This limitation arises from the presence of a bulkier proline residue

situated in the catalytic motif IV. Conversely, rRNA N6,6-dimethyladenosine MTases feature aliphatic amino acids, such as (iso)leucine, making them adept at dimethylating rRNA substrates (Schluckebier et al., 1995; Schluckebier et al., 1999).

To establish the inclusion of protein PH1823 in the rRNA N6,6-dimethyladenosine superfamily, its signature trait was identified and compared with rRNA and DNA N6-adenosine MTases using an MSA. The results show that PH1823 also contains an aliphatic amino acid (Ile133) instead of proline, indicating its capability for dimethylating the target adenine at the N6 position during catalysis (Figure 4.4A). This observation suggests that PH1823 functions as an archaeal rRNA N6,6-dimethyladenosine MTases. Additionally, a phylogenetic analysis was conducted to ascertain the evolutionary relationship between PH1823 and homologs of rRNA N6,6-dimethyladenosine MTases (Table B.2). The results indicate that PH1823 is more closely related to the KsgA/DIMT1 family than the Erm family (Figure 4.4B). Moreover, PH1823 has been clustered within the archaeal DIMT1 family, exhibiting greater similarity to eukaryotic DIMT1 than the bacterial KsgA family (Figure 4.4B). Consequently, the protein PH1823 has been designated as *PhDIMT1*.

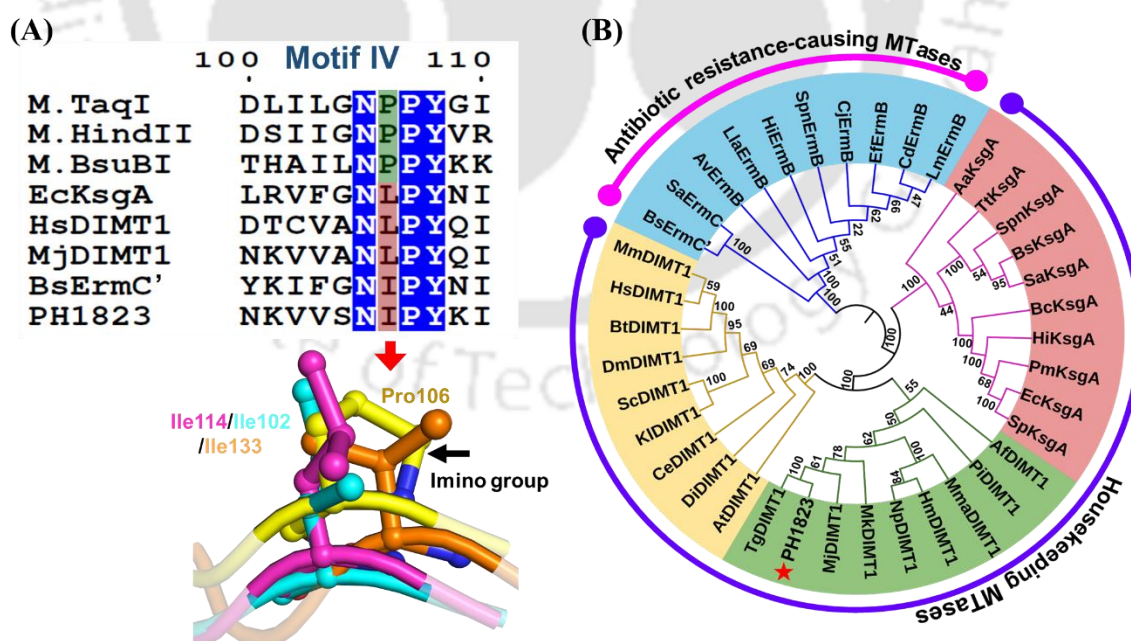


Figure 4.4. PH1823 is an archaeal DIMT1 ortholog. (A) MSA of the conserved active-site region of all the major DNA N6-adenosine and rRNA N6,6-dimethyladenosine MTases, along with PH1823. The MSA was performed using the

protein sequences of *M.TaqI* (*T. aquaticus*, P14385), *M.HindII* (*Haemophilus influenzae*, P44414), *M.BsuBI* (*Bacillus subtilis*, P33563), *EcKsgA* (*E. coli*, P06992), *HsDIMIT1* (*H. sapiens*, Q9UNQ2), *MjDIMIT1* (*Methanocaldococcus jannaschii*, Q58435), *BsErmC'* (*B. subtilis*, P13956) and PH1823 (*P. horikoshii*, O59487); the organism name and UniProt id of each protein are provided in the parenthesis. The 106th position (numbering according to protein *M.TaqI*) of proline and aliphatic amino acids are highlighted in green and brown, respectively. Superposition of motif IV of protein *M.TaqI* (PDB id: 1AQL, yellow) with *EcKsgA* (PDB id: 1QYR, magenta), *BsErmC'* (PDB id: 1QAN, cyan) and PH1823 (PDB id: 8X3W, orange). (B) A phylogenetic tree was constructed using the Neighbor-joining method with 1000 replicates. A phylogenetic tree showing the evolutionary relationship was generated using 38 different rRNA N6,6-dimethyladenosine MTases, including the protein PH1823; details are provided in Table B.2. The result indicates two distinct clusters of KsgA/DIMIT1- housekeeping and antibiotic resistance causing MTase (Erm). Further, these housekeeping MTases are divided into three clusters according to three different domains of life (bacterial, maroon; eukarya, green, and archaeal, yellow), where the archaeal DIMT1 family is closer to eukaryotic DIMT1 than the bacterial KsgA Family. The PH1823 is marked with a red asterisk.

4.3.2. The protein *PhDIMIT1* is conformationally and compositionally stable

To assess the conformational and compositional stability of the purified protein *PhDIMIT1_WT*, its secondary structural elements were estimated by performing circular dichroism (CD) experiments at 25°C. A Far-UV spectra analysis of *PhDIMIT1_WT* suggests that the protein comprises ~40% α -helices, ~19% β -strand, and ~41% other secondary structures (Figure 4.5A). To probe the effect of temperature on the protein, thermal ramping experiments were performed by monitoring the CD spectra analysis ranging from 20 to 110°C. The result suggests that the protein *PhDIMIT1_WT* maintains conformational stability up to 60°C (Figure 4.5B-4.5C).

Subsequently, the compositional stability of *PhDIMIT1_WT* was monitored by measuring hydrodynamic diameters with increasing temperature from 20°C to 70°C $\pm$ 0.1°C using dynamic light scattering (DLS). The result exhibits that the hydrodynamic diameter of *PhDIMIT1_WT* lies in the range of 4-6 nm and is stable up to 60°C,

Chapter 4. Structural and functional characterization of archaeal m₂⁶A rRNA MTase

indicating its homogeneity. The hydrodynamic diameter of the protein reached up to ~1 μ M at 70°C, suggesting protein aggregation (Figure 4.5D). Further, mass spectrometry analysis revealed the presence of a single high-intensity peak corresponding to a molecular weight of 35793.58 Da, equivalent to its theoretical molecular weight of ~35800 Da, indicating its compositional stability (Figure 4.5E).

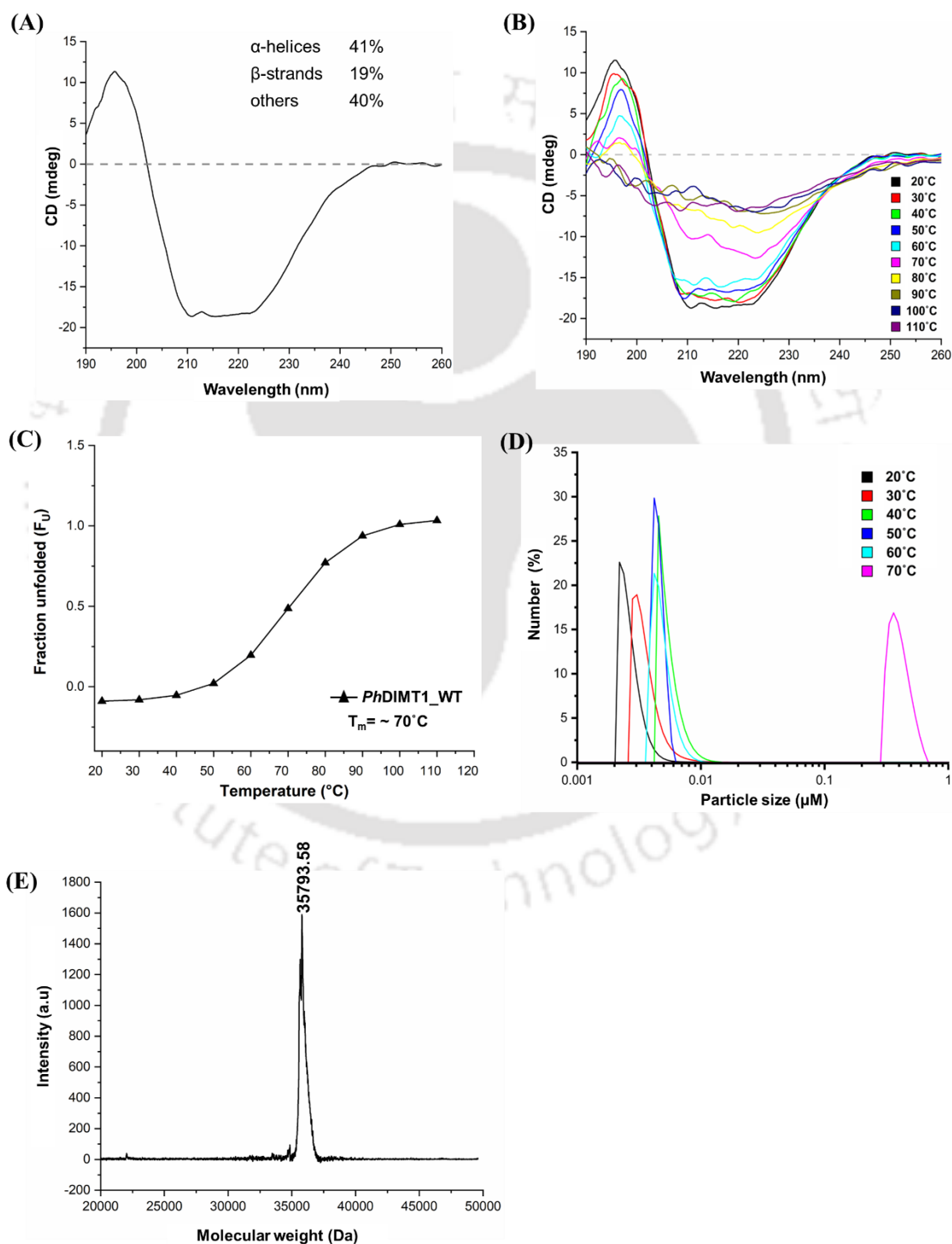


Figure. 4.5. Conformation and compositional stability of the protein *PhDIMIT1_WT*. (A) A Far-UV CD spectrum ($\lambda = 190\text{--}260\text{ nm}$) of *PhDIMIT1_WT* provides its secondary structural content. (B) The circular dichroism (CD) spectra were recorded over a wavelength range of 260–190 nm to investigate the thermal denaturation of *PhDIMIT1_WT*. Spectra were acquired at temperatures ranging from 20 to 110 °C with 10 °C intervals. The initial spectrum at 20°C is depicted by the black line, while subsequent spectra at higher temperatures are distinguished by various colors. The overlay allows the identification of the temperature (T_m) at which changes in the secondary structure of the protein occur. (C) Thermal unfolding curves for *PhDIMIT1* (depicted by filled triangles) were generated by monitoring the CD ellipticity at 222 nm. (D) The change in hydrodynamic diameter was obtained using DLS spectra collected during heating at the indicated temperatures. The colored traces show a hydrodynamic diameter range of 4–6 nm during the temperature of 20–60°C. However, it was $\sim 1\text{ }\mu\text{m}$ at 70°C, indicated by a magenta line. (E) Mass spectra of the protein *PhDIMIT1_WT* using MALDI-TOF analysis suggest its molecular weight of 35793.58 Da, equivalent to the theoretically calculated molecular weight ($\sim 35800\text{ Da}$).

4.3.3. The protein *PhDIMIT1* shows specificity towards various cofactors

The *in silico* analysis suggested that the protein *PhDIMIT1* is an archaeal homolog of the housekeeping SAM-dependent DIMT1. To corroborate the results further, the binding of the SAM (S-adenosyl-L-methionine), SAH (S-adenosyl-L-homocysteine), and SFG (adenosyl-L-ornithine) to the protein *PhDIMIT1_WT* was estimated using the fluorescence-based approach. The results depict that the emission maximum (λ_{max}) of the protein *PhDIMIT1_WT* occurs at 341 nm in the absence of cofactor(s). Meanwhile, with increasing concentration of the cofactors, the fluorescence intensity decreases (hypochromic shift), suggesting that cofactors can produce fluorescence quenching in the protein *PhDIMIT1_WT*. Further, the cofactors SAM and SFG cause a blue shift of $\sim 4\text{ nm}$ (341 to 337) in the λ_{max} , whereas that by the cofactor SAH is $\sim 5\text{ nm}$ (341 to 336) (Figure 4.6A–4.6C).

Chapter 4. Structural and functional characterization of archaeal m²6A rRNA MTase

To further estimate the binding affinity of SAM, SAH, and SFG to the protein *PhDIMT1_WT*, isothermal titration calorimetry (ITC) experiments were carried out. The ITC data revealed that the binding isotherms of all the cofactors with the protein are exothermic and energetically favorable (Figure 4.6D-4.6F). The ITC results indicate that *PhDIMT1_WT* shows a high affinity towards the product SAH compared to SAM and its analog SFG with K_d values of 1.67, 2.81, and 7.14 μM , respectively (Figure 4.6D-4.6F and Table B.1). Previous studies have reported that in most SAM-dependent MTases, SAH is a competitive inhibitor with a binding affinity slightly higher than SAM. This was probably required to regulate the ‘methylation index’ (ratio of SAM/SAH) of a cell with an adequate SAM supply and SAH removal (Melnik et al., 2000).

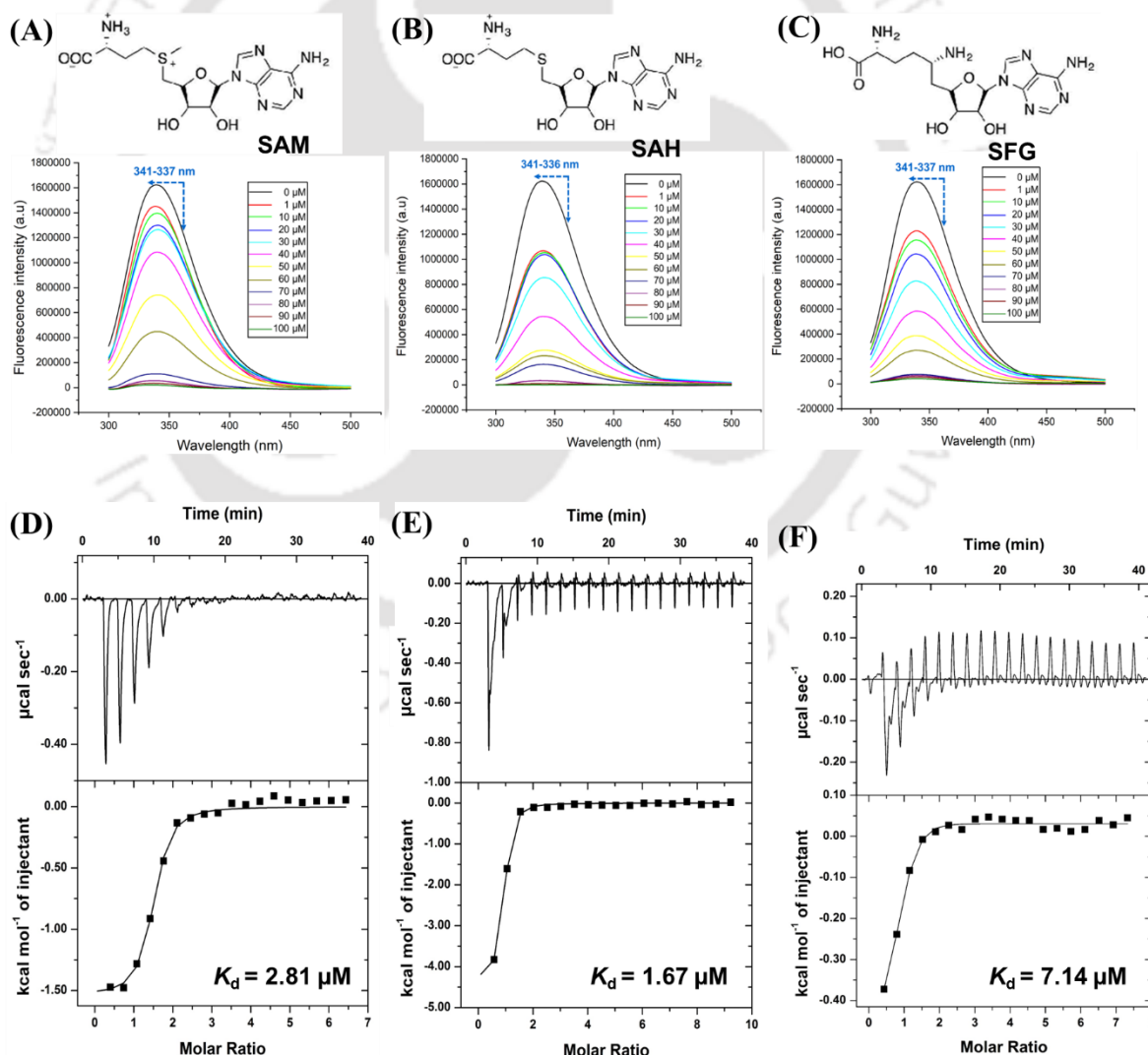


Figure 4.6. The binding affinity of the protein *PhD*IMT1 for various cofactors. (A-C) Fluorescence emission spectra of *PhD*IMT1_WT in the absence (black) and presence of different cofactors (SAM, SAH, and SFG). The blue shift of ~4 nm upon fluorescence quenching of *PhD*IMT1_WT by cofactors is represented with the broken arrows. The change in fluorescence spectra of *PhD*IMT1_WT due to the fluorescence quenching effect of SAM, SAH, and SFG at varying concentrations (1-100 μ M) have been represented in different colors. (D-F) A plot showing the respective binding of SAM, SAH, and SFG to the protein *PhD*IMT1_WT was obtained through isothermal titration calorimetry (ITC) data. The upper panels show the raw heat change upon ligand titration, while the integrated heat pulse subtracted with their heat of dilution is shown in the lower panels.

4.3.4. The overall structure and active-site pocket of the protein *PhD*IMT1

The crystal structure of the apo form of *PhD*IMT1_WT was determined at a resolution of 2.0 Å in the space group C2 containing two protomers in the asymmetric unit. The root-mean-square deviation (RMSD) of C α atoms between the two protomers of *PhD*IMT1_WT is ~0.3 Å, unlike *Ec*KsgA (RMSD: ~1.2 Å), indicating higher structural similarity (Table 4.2). Overall, akin to known rRNA N6,6-dimethyladenosine MTases, *PhD*IMT1 also consists of an N-terminal catalytic domain (NCD, 21-211 residues) having the Rossmann fold and a C-terminal RNA-binding (or head) domain (CRBD, 221-290 residues) (Figure 4.7A). A structure-based sequence alignment of *PhD*IMT1 with its homologs indicates the presence of an active-site pocket in the N-terminal end of the proteins. The active site of *PhD*IMT1 is made up of typical nine sequence motifs (I-VIII and X) required for its activity. Although the motifs I, II, III, IV, VIII, and X are highly conserved, motifs V, VI, and VII show variability (Figure B.1A). Motifs I, II, and III contain amino acid residues important for cofactor binding, whereas motifs IV and VIII contain residues important for catalysis (Schluckebier et al., 1999; O'Farrell et al., 2004)

Overall, the NCD of the protein *PhD*IMT1 adopts an α + β fold consisting of eight β -strands sandwiched between eight α -helices, while the CRBD consists of 4 α -helices. Topologically, the Rossmann fold consists of a series of alternating β -strand and α -helical segments wherein the parallel β -strands are connected by α -helices, resulting in

a β - α - β - α - β arrangement; overall, the strands are arranged as $\beta_4\beta_3\beta_2\beta_5\beta_6\beta_7$ where antiparallel strand β_8 inserted between β_6 and β_7 . This type of fold is a characteristic of SAM-dependent MTase fold (Figure 4.7B). Many structures of KsgA/DIMT1, including KsgA from *E. coli* (EcKsgA, PDB id: 1QYR), DIMT1 from *Homo sapiens* (HsDIMT1, PDB ids: 6W6C and 1ZQ9), TFB1M from *Homo sapiens* (HsTFB1M, PDB ids: 6AAX and 6AJK) and ErmC' from *Bacillus subtilis* (BsErmC', PDB ids: 2ERC, 1QAM, 1QAN and 1QAO) have previously been reported. Still, a comprehensive understanding of the entire active-site vicinity has remained elusive, primarily due to the inherent flexibility of the loop situated in motif X within the NCD of most of KsgA/DIMT1, with the exception of BsKsgA from *B. subtilis* (PDB ids: 6IFS and 6IFT). Fortunately, in our investigation, we have successfully crystallized this flexible loop in all forms of PhDIMT1 (Figure B.1B).

Interestingly, the co-crystals of the protein PhDIMT1_WT with SAM resulted in the binding of 5'-methylthioadenosine (MTA) in the negatively-charged SAM-binding site and adenosine (ADN) to the CRBD (Figure B.2A and Table 4.2). The adenine N6 and N1 atoms of MTA form hydrogen bonds with the residues Asp117 and Ala118 (motif III), respectively. The two ribose hydroxyl groups of MTA form hydrogen bonds with the residues Gln44 (motif X) and Glu93 (motif II). The sulfur tail of MTA was stabilized by forming a hydrogen bond with His45 (Figure 4.7C and Table B.3). SAM hydrolysis predominantly produces MTA as the primary degradation product. However, prolonged incubation of SAM resulted in an additional minor product, ADN (Parks and Schlenk, 1958). Thus, the SAM-binding pocket gets predominantly occupied with the primary hydrolysis product, MTA, leaving the minor product, ADN, with no alternative but to interact non-specifically with a small, negatively-charged pocket situated in the CRBD. The N6 atom of ADN forms hydrogen bonds with His245 and Leu247, while the N7 and N3 atoms interact with His245 and Ile288, respectively, in the CRBD (Figure B.2B and Table B.3).

The previous attempts to obtain a binary complex of the protein TtKsgA with SAM led to similar outcomes (Demirci et al., 2009). Structural comparison of PhDIMT1_WT and TtKsgA reveals spatial conservation of the active-site residues at both sequence and secondary-structure levels (Figure B.1A). The adenosine binding is conserved in SAM

and its various structural analogs (Schluckebier et al., 1995). Therefore, the MTA molecule in the cofactor-binding site is bound in a position similar to SAM in all rRNA N6,6-dimethyladenosine MTases (Figure 4.7D).

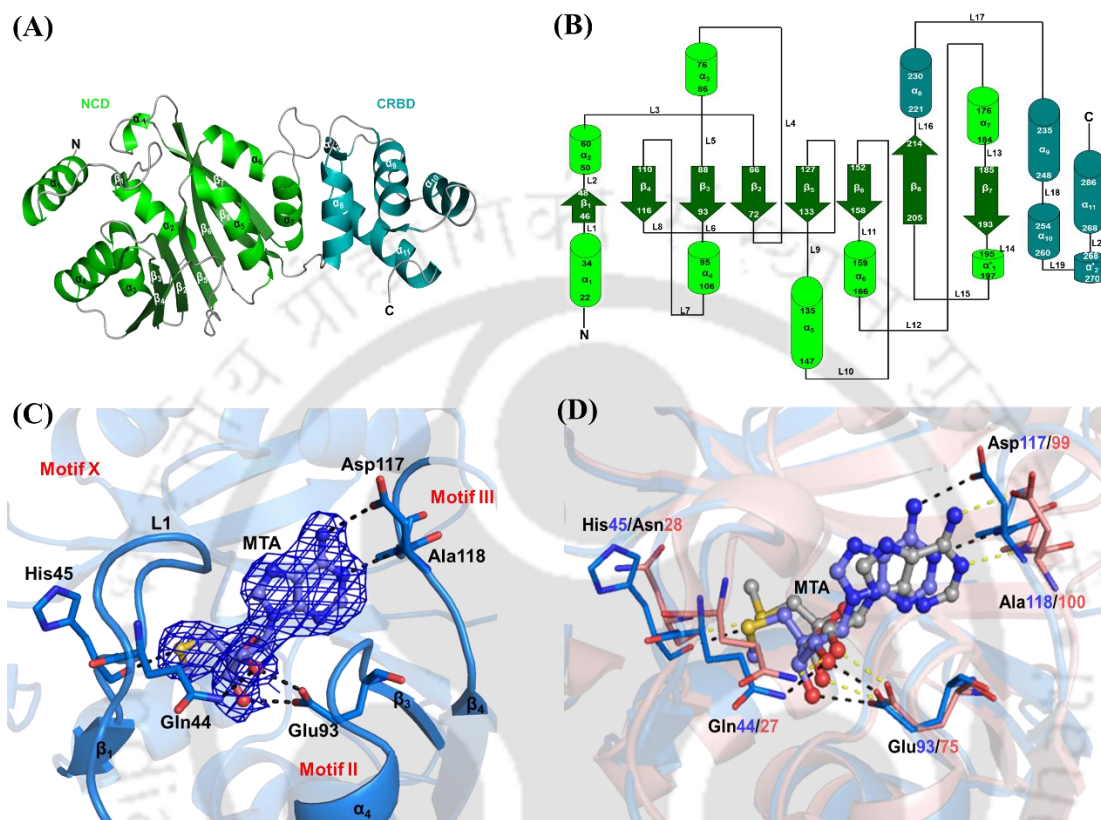


Figure 4.7. The overall structure and active site of the protein *PhDIMT1*. (A) The three-dimensional structure of *PhDIMT1*_WT (PDB id: 8X3W) highlights the secondary structural elements of the NCD (α -helices in green and β -strands in dark green) and the CRBD (α -helices in teal). Loops are shown in grey. (B) Topology diagram of *PhDIMT1*_WT (β -strands as arrows, α -helices as cylinders). The N-terminal domain (NCD) contains six parallel β -strands (β_2 - β_7) that form an extended β -sheet arranged in β - α - β - α - β inverted tandem repeat, found in Rossmann fold. (C) The bound MTA molecule to the protein has been shown as a ball-and-stick model along with its 2F_o-F_c electron density map contoured at 1 σ level. The active site residues interacting with MTA are shown as blue sticks. The interactions are represented as black dashed lines. (D) Superposition of the active site residues of *PhDIMT1*_WT•MTA•ADN (PDB id: 8X41, blue) and its homolog *TtKsgA*•MTA (PDB id: 3FUX, magenta). Hydrogen

bonds are displayed in black and yellow dashed lines for *PhDIMT1_WT* and *TtKsgA*, respectively.

4.3.5. Hydrolysis of SAM to MTA occurs during crystal incubation

The observation of the electron densities resembling MTA and ADN in the crystal structures of the SAM-incubated protein *PhDIMT1_WT* suggested that the molecule SAM was hydrolyzed. The question is whether the incubated SAM hydrolyzed to MTA and ADN prior to or during the co-crystallization trials. Furthermore, as the freshly-prepared SAM was titrated at room temperature for ITC experiments, what was the effect of temperature and incubation time on the SAM? To get answers to these queries, a time-dependent SAM hydrolysis assay was performed using reverse-phase high-performance liquid chromatography (RP-HPLC) experiments. The result indicates a single peak at 6.6 min when the sample was incubated for ~24 hrs at room temperature (condition I), while two peaks at 6.6 min and 15.1 min were obtained when the sample was incubated for almost four weeks at room temperature (condition II) (Figure 4.8A-4.8B). The freshly-prepared SAM was used as a reference solution, with a retention time of 6.6 min, indicating its non-hydrolysis under the assay condition I (Figure B.3A). This revealed that SAM remains intact for ~24 hours at 25°C.

The additional peak at 15.1 min (Figure 4.8B) in condition II was further investigated using a standard run of different nucleoside analogs, viz. MTA, ADN, and SAH, showed retention times of 15.1, 13.1, and 16.7 min, respectively (Figure 4.8C-4.8D, and Figure B.3B). The outcome of the experiment suggests that the second peak at 15.1 min corresponds to the molecule MTA. This indicates that the molecule SAM is hydrolyzed in condition II. Moreover, the intensity of the second peak (15.1 min) in condition II is significantly greater than that of the first peak (6.6 min), indicating that the majority of SAM has been hydrolyzed to MTA.

Further, the ITC experiments were performed to establish the differential binding affinities of two compounds, MTA and ADN, with *PhDIMT1_WT*, and their thermodynamic characteristics were compared to SAM. The binding profile of SAM and its analogs with the protein has been defined by a negative change in enthalpy and a positive change in entropy, implying an exothermic reaction (Figure B.3C). However,

the binding results signify that SAM had a higher affinity (K_d : 2.81 μ M) compared to its hydrolyzed equivalents, MTA (K_d : 5.30 μ M) and ADN (K_d : 5.78 μ M) (Figure 4.8E-4.8F and Table B.1).

Simultaneously, the SAM molecule incubated in condition II was utilized in soaking the apo crystal of *PhD1MT1*_WT and was incubated for ~2 days. The structure elucidation of these crystals revealed that MTA was bound to the SAM-binding site of the protein. Interestingly, these soaked crystals produced diffraction patterns in two distinct space groups, *P*22₁2₁ (Form I, PDB id: 8X44) and C2 (Form II, PDB id: 8X45) (Table 4.2).



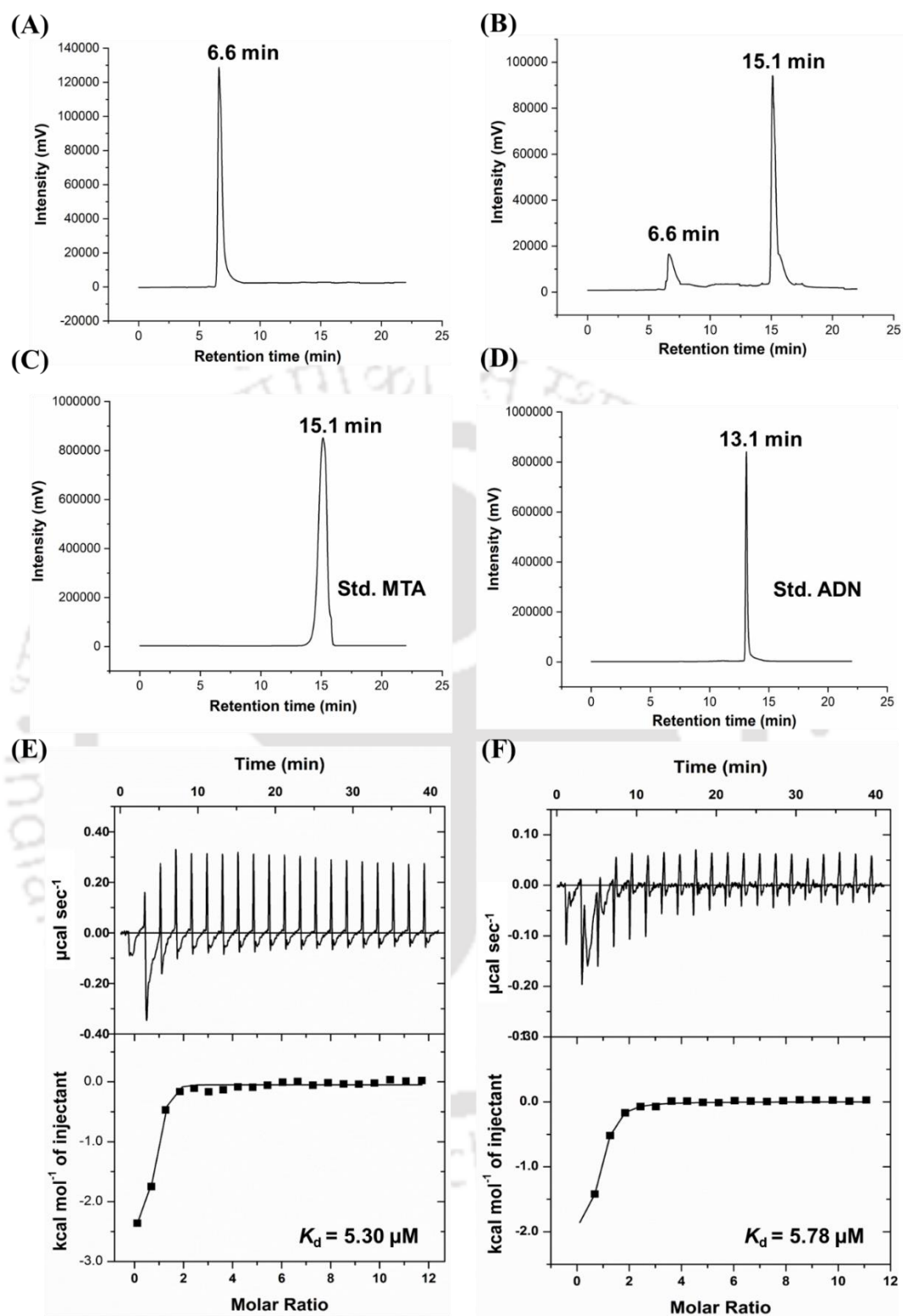


Figure 4.8. The stability assay for the molecule SAM. (A-B) The time-dependent SAM cleavage assay was performed. The two-reaction mixtures (solution I and II) were incubated for ~24 hrs and ~4 weeks at 25°C, followed by RP-HPLC analysis, and their retention time was recorded. The RP-HPLC profiling for the standard sample (C) MTA

and (D) ADN with their retention time. ITC data provides the binding of the protein *PhDIMT1_WT* with (E) MTA and (F) ADN.

4.3.6. Binding of cofactors causes local conformational changes in the NCD

In addition to the co-crystals of *PhDIMT1_WT* with SAM, the product SAH and the inhibitor SFG were also co-crystallized with the protein. The crystals of the protein bound to SFG (*PhDIMT1_WT*•SFG, PDB id: 8X46) and SAH (*PhDIMT1_WT*•SAH, PDB id: 8X47) diffracted to a resolution of 2.2 Å and 2.7 Å in two different space groups, C2 and C222₁, respectively (Table 4.2). The structural analysis of both complex structures revealed that the N6 atom of the adenine ring of the cofactors forms a hydrogen bond with the residue Asp117, while its N1 atom interacts with the main chain of the residue Ala118. Two hydroxyl oxygen atoms of the ribose are coordinated by the residues Gln44 and Glu93. As these molecules (SAM, its hydrolyzed form MTA, the product SAH, and the competitive inhibitor SFG) contain a common adenosine group, similar modes of interactions were observed in all these molecules.

An α -amino acid part of the SAH and SFG binds in a shallow groove pointing away from the adenosine group. In SFG, the α -amino acid moiety is ornithine, which forms a hydrogen bond with the carbonyl oxygen of the residue Asn132. The carboxyl oxygens of SFG are stabilized through hydrogen bonding with the residues His45 and Tyr135. Notably, the residue Tyr135 forms CH- π interaction with the carboxy tail of SFG (Figure 4.9A and Table B.3). On the other hand, the sulfur atom of SAH is stabilized by the carbonyl oxygen of the residues Gly43 while the carboxyl group of cofactor forming hydrogen bonds with residues Glu70 and Gly72 (Figure 4.9B and Table B.3). As a result, the orientation of the catalytic pocket made up of catalytic loops (α_1 - β_1 : L1, α_4 - β_3 : L6, β_4 - β_5 : L8, and α_5 - β_5 : L9 in the motif X, II, III and IV, respectively) of the cofactor-bound complexes significantly differs with that of the apo protein *PhDIMT1_WT* (Figure 4.9C). The loop L1 was found to exhibit the most significant structural change between the apo (unbound) and holo (cofactor-bound) forms of *PhDIMT1_WT*. In the apo form, this loop moves away from the active site, creating an open and flexible pocket (Figure 4.9C). This structural change causes the residue Gln44 within the loop L1 to bend in a way that partially obstructs the cofactor ribose position (Figure 4.9C).

On the other hand, a structural comparison between the holo forms of *Ph*DIMT1_WT and its AlphaFold model reveals a remarkable resemblance. This outcome suggests that Gln44 shifts sideways. This allows Gln44 to form a crucial hydrogen bond with Asp95 in motif II and a second one with Arg39 in motif X. These two interactions effectively restrict access to the ribose position for the residue Gln44 (Figure B.4A). Thus, the loop L1 acts like a flexible lid that tries to close the catalytic pocket after cofactor binding (Figure 4.9C). Moreover, comparative structural analysis revealed a similar conformational rearrangement occurring within the catalytic pocket of *Tt*KsgA (Figure B.4B). Moreover, analysis of the SAM (molecule incubated in condition II)-soaked *Ph*DIMT1_WT (Forms I and II) revealed that the dynamic loop L1 closely mimics the apo conformation, with RMSDs of 0.31 Å and 0.26 Å, respectively. Furthermore, when compared to the co-crystallized MTA form, the loop L1 in SAM-soaked forms of *Ph*DIMT1_WT remains static, reminiscent of the apo form (Figure 4.9D).

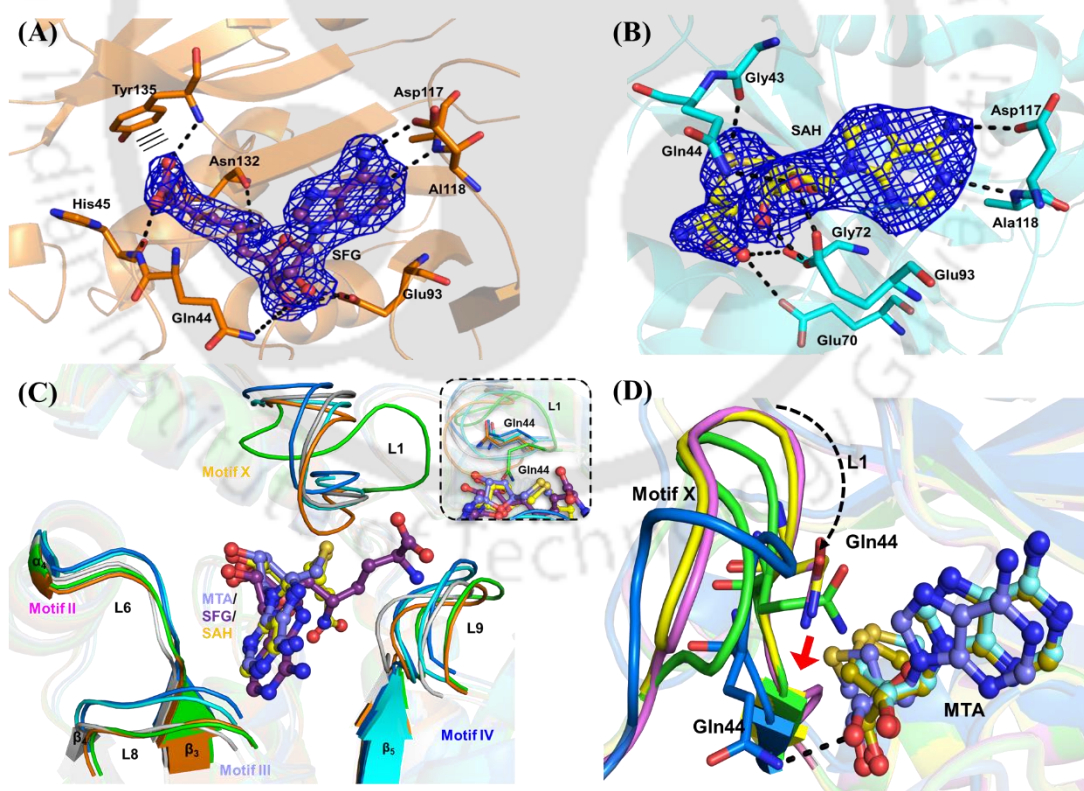


Figure 4.9. Conformational flexibility of the protein *Ph*DIMT1 upon cofactor binding. (A-B) Interactions of the cofactors SFG (PDB id: 8X46) and SAH (PDB id:

8X47) with active site residues of the protein *PhD*DIMT1_WT, respectively. Active site residues interacting with SFG and SAH are depicted as orange and cyan sticks, while cofactors SFG and SAH are shown as violet and yellow ball-and-stick model, respectively, along with their 2F_o-F_c electron density maps contoured at 1σ level. Hydrogen bonding interactions are represented as black dashed lines, while three parallel lines represent CH-π interactions. (C) Superposition of all complex forms (MTA, SFG, and SAH bound) and the AlphaFold model structure compared with the apo form of protein *PhD*DIMT1_WT to elucidate the functional role of catalytic loops. The most significant movement is observed in the flexible loop L1. The square box represents a magnified view of Gln44 in the apo and holo forms of *PhD*DIMT1_WT. The apo and AlphaFold model structures of *PhD*DIMT1_WT are highlighted in green and grey, respectively, while the backbone of the protein *PhD*DIMT1 complex with MTA (purple), SAH (yellow), and SFG (violet) are represented with the blue, cyan, and orange color. (D) Superposition of the active site residues of MTA-bound co-crystallized and soaked (Forms I and II) of *PhD*DIMT1_WT with its apo form. The active site residue of MTA-bound soaked forms I (PDB id: 8X44) and II (PDB id: 8X45) are highlighted as yellow and pink, while their cofactor MTA is shown as the sky and olive ball-and-stick model, respectively.

4.3.7. Invariant water molecules stabilize the critical orientation of cofactors in the NCD

Three structurally invariant water molecules (IW1, IW2, and IW3) located in the vicinity of the cofactor-binding site of protein *BsErmC*' have been previously reported (Schluckebier et al., 1999). To locate the presence of similar invariant water molecules in *KsgA*/*DIMT1* proteins, apo and complex structures were investigated (Table B.4). Akin to *BsErmC*', two invariant water molecules (IW1 and IW2) are present in the proteins *BsKsgA* and *MjD*DIMT1 in the proximity of the NCD. These water molecules interact with the cofactor (SAM and SAH) tail. The tail of the cofactors is further stabilized by hydrogen bonding through the residue Gly38 (in *BsErmC*' and *MjD*DIMT1) or Gly56 (in *BsKsgA*). Further, in *BsErmC*'/*MjD*DIMT1/*BsKsgA*, the IW1 molecule forms two additional hydrogen bonds with the hydroxyl oxygen and backbone nitrogen atoms of Thr45/45/63, while the IW2 molecule forms hydrogen bonds with Glu36/36/54 (Figure B.5A-B.5B). Interestingly, two similar water molecules are

present in the active site of the apo form of *PhD*MT1_WT near the glycine-rich loop of motif I (Figure 4.10A and Table B.4). Notably, in the SAH-bound *PhD*MT1_WT, these two molecules are displaced from their original positions, resulting in a 2.8 Å upward movement of the nitrogen atom of cofactor SAH. This upward shift prevents hydrogen bond formation with Gly72 (equivalent to Gly38 in *BsErmC'*/*Mj*MT1 or Gly28 in *BsKsgA*). This upward movement leads to the rise in the B-factor ($\sim 64 \text{ Å}^2$) of the nitrogen atom of SAH in *PhD*MT1_WT compared to that of *BsErmC'* (PDB id: 1QAN, $\sim 58 \text{ Å}^2$), *BsKsgA* (PDB id: 6IFT, $\sim 19 \text{ Å}^2$), and *Mj*MT1 (PDB id: 3GRR, $\sim 55 \text{ Å}^2$). Nevertheless, this displacement of two water molecules in the SAH-bound form of *PhD*MT1_WT occurs due to the loss of the critical contact between the amide nitrogen atom of Thr79 (equivalent to Thr45 in *BsErmC'*) and the IW2 molecule (Figure 4.10A and Figure B.5A-B.5B).

Thus, the stabilization of the α -amino acid moiety of SAH in the cofactor-binding pocket of the protein can be explained by two different mechanisms. In one, the presence of an S ^{δ} atom in SAH contributes to the stability of the tail of the cofactor through the formation of a chalcogen bond. The divalent S ^{δ} atom interacts with the carbonyl oxygen atom of Gly43, resulting in a strong chalcogen bonding interaction that is comparable in strength to conventional hydrogen bonds and remains unaffected by the surrounding solvent (Adhav et al., 2023). In the second, the invariant water molecule (IW3) forms a hydrogen bond with the O4' atom of the ribose sugar of SAH, contributing to its stability. The water molecule IW3 is anchored in its position by forming a hydrogen bond with active site residues of the protein. Particularly, in *BsErmC'*, the carbonyl atoms of Ile37 and Asn101 engage with the IW3 molecule, whereas in *PhD*MT1_WT, the coordination occurs via the residue Val71 (equivalent to Ile37) and Asn132 (equivalent to Asn101). The water molecule IW3 establishes two hydrogen bonds with the N ^{$\delta 2$} and carbonyl atoms of Asn132 and one with the carbonyl oxygen atom of Val71 (Figure 4.10B).

The superposition of SFG- and SAH-bound structures of *PhD*MT1_WT exhibited an unexpected movement of the α -amino acid tail of the former one despite the presence of IW1 and IW2. This tail movement of SFG is likely due to the presence of a smaller C ^{δ} atom compared to the S ^{δ} atom of SAM and SAH, respectively. Furthermore, the

presence of two atoms in close proximity, N^ε of SFG and O atom of the water molecule, unexpectedly leads to a steric hindrance, resulting in the displacement of IW3. Due to these unfavorable charge differences and steric clashes, the ribose group of SFG is partially destabilized, causing its α -amino acid group to rotate $\sim 87^\circ$ (N-C5'-N) compared to SAH bound to *PhD*MT1_WT (Figure 4.10C).

To identify the likely locations of dimethylation in h45 of 16S rRNA in *P. horikoshii*, the nucleotide sequences of h45 from three domains of life were examined. The results show that the nucleotide residues A1475 and A1476 in *P. horikoshii* correspond to A1518 and A1519 in *E. coli* (Figure B.6A). Following that, a three-dimensional model of the helix h45 of 16S rRNA in *P. horikoshii* was predicted (Figure B.6B). The predicted model of h45 has a tetraloop (5'-₁₄₇₃G-G-A-A₁₄₇₆-3') geometry without-out-in-out bases having a free energy change (ΔG) of -18.26 kcal mol⁻¹, similar to previously published structures of h45 from *H. sapiens* (Liu et al., 2019; Stephen et al., 2021). Furthermore, the purity of the synthesized h45 was determined (Figure B.6C). Thereafter, the binding affinity of h45 with the protein *PhD*MT1 was estimated by performing ITC experiments. The ITC data exhibit the binding of h45 to the protein *PhD*MT1 is exothermic and enthalpically favorable with a K_d of ~ 2.0 μ M (Figure B.6D and Table B.1). The archaeal ternary complex (*PhD*MT1_WT•SAM/•SAH•RNA) crystals were obtained. However, due to poor-quality diffraction, their three-dimensional structure could not be identified. As an alternative, the crystal structure of SAH-bound *PhD*MT1_WT was docked with h45 to form the ternary complex. Comparing this model with the crystal structure of SFG-bound *PhD*MT1_WT demonstrates that the extensive movement of the α -amino acid moiety of SFG results in a steric clash with the nucleotide residue A1476 of the substrate rRNA. This steric clash is stronger than the base-stacking interaction provided by the two aromatic residues, Tyr135 and Phe198, which are necessary for the correct positioning of substrate A1476. As a result, SFG binding blocks the target adenosine A1476, causing it to orient away from the active site and rendering the protein catalytically inactive (Figure 4.10D).

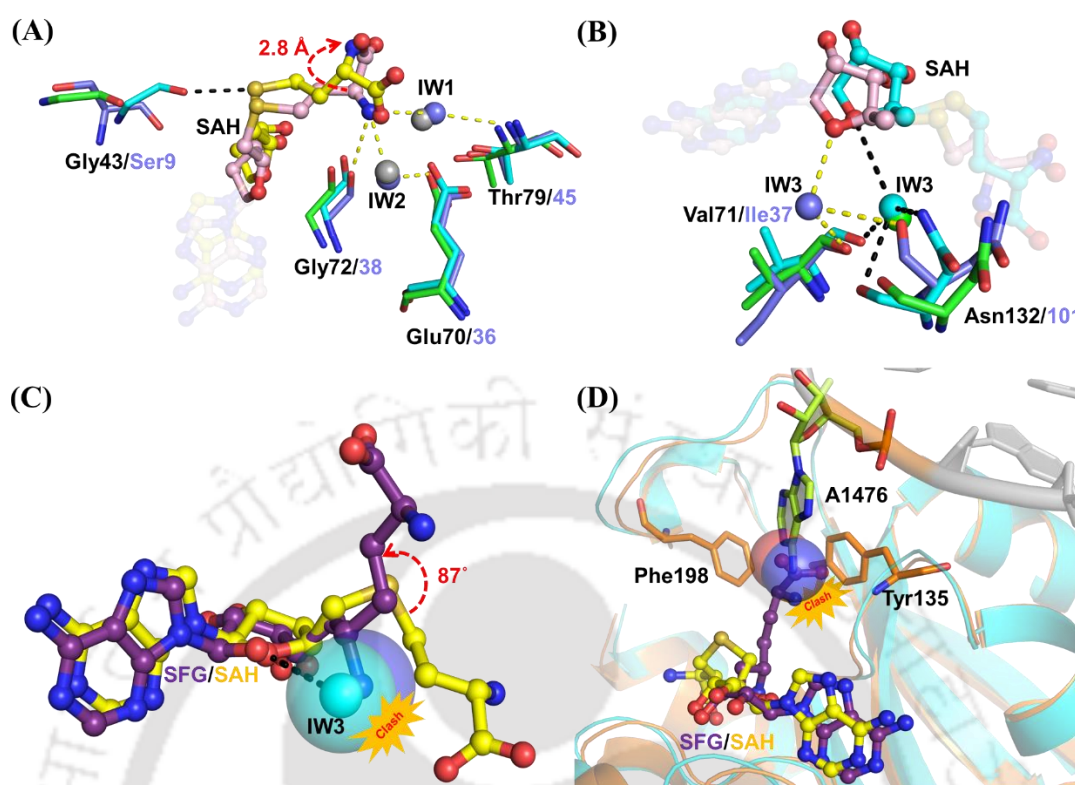


Figure 4.10. The significance of structurally invariant water molecules. (A-B) Superposition of SAH-bound *PhDIMT1_WT* structure (PDB id: 8X47) with its apo form (PDB id: 8X3W) and reference protein *BsErmC'* (PDB id: 1QAN) to illustrate the involvement of IW1, IW2, and IW3 during cofactor binding. The interacting active site residues of *PhDIMT1_WT* in apo (green) and holo (cyan), as well as *BsErmC'* (purple), are shown as sticks, while its cofactor SAH is shown as a yellow and pink ball-and-stick model, respectively. The dashed line represents hydrogen bonds and is highlighted in black and yellow for SAH-bound *PhDIMT1_WT* and reference protein *BsErmC'*, respectively. (C) The van der Waals radii of closely situated non-bonded N^ε atoms of SFG (PDB id: 8X46) and O atoms of IW3 cause steric conflict, resulting in 87° movement of the α -amino acid moiety as compared to SAH bound to *PhDIMT1_WT* (PDB id: 8X47). The spheres represent IW molecules present (cyan) and displaced (grey) in SAH-bound *PhDIMT1_WT* and *BsErmC'* (purple). (D) The superposition of SFG-bound *PhDIMT1_WT* structure (PDB id: 8X46) with the ternary model indicates the function of SFG as an inhibitor in the methyltransferase reaction by blocking substrate adenosine A1476.

4.3.8. SAH-bound *Ph*DIMT1 reflects a transmethylation state

In the quest to comprehend the catalytic mechanism of *Ph*DIMT1, the ternary complex was used as a mechanistic model. The docking results demonstrate that the negatively-charged h45 binds to a positively-charged site of the protein in the vicinity of the cofactor-binding pocket (Figure 4.11A). The interactions in h45 are predominantly governed by electrostatic forces with the positively-charged residues, viz. Lys136, Arg164, Arg172, Arg176, Lys203, Lys226, and Arg232 (Figure 4.11B). In the 5'-GGAA-3' tetraloop of h45, the first base guanine at position G1473 is conspicuously nestled within a small positive charge pocket, which is meticulously enclosed by a cluster of the key residues, Ser139, Phe143, Arg164, Arg176, and Lys226 (Figure B.7A-B.7B). Specific interactions, including the oxygen atom in C6 and N7 of guanine, the O4' atom of the ribose sugar, and the phosphate group of G1473, engage in the formation of hydrogen bonds with the Nⁿ² atom of Arg176, O^y atom of Ser139, and Nⁿ¹, Nⁿ² atoms of Arg164, respectively. Furthermore, G1473 forms a salient salt-bridge interaction with Lys226 of the CRBD (Figure B.7B). Therefore, the nucleobase G1473 plays a pivotal role in orchestrating these intricate molecular interactions, ultimately facilitating the formation of the ternary complex.

The above MSA results, preliminary indicate that A1475 likely functions as the second substrate (dimethylation site) of the enzyme *Ph*DIMT1 (Figure B.6A). To further validate this, a structural superposition was performed, aligning our mechanistic model with the experimental ternary complex of *Hs*TFB1M. The outcome indicates a preference for flipped-out terminal adenosine in the (5'-₁₄₇₃G-G-A-A₁₄₇₆-3') tetraloop, A1476 (corresponding to A937 in *Hs*TFB1M) was favored for modification by enzyme *Ph*DIMT1. Conversely, the flipped-in position of A1475 (A936 in *Hs*TFB1M) hinders entry into the catalytic pocket (Figure B.7C). Remarkably, in this conformation, the nucleobase A1475 facilitates the formation of two crucial interactions between its N7 and 2'-hydroxyl atoms with the N^{ε2} of Gln158 and O^{ε2} of Glu160, respectively (Figure 4.11C and B.7D). Thus, A1475 acts as a molecular hook outside the catalytic groove, stabilizing the tetraloop geometry. Although, analogous interactions were not observed in the ternary structure of *Hs*TFB1M (PDB id: 6AAX) (Figure B.7D). This phenomenon is likely attributable to differences in the orientation of conserved

residues, Gln177 (equivalent to Gln158) and Glu179 (equivalent to Glu160), between *Ph*DIMT1 and *Hs*TFB1M.

On the other hand, nucleobase A1476 is inserted into the catalytic pocket of *Ph*DIMT1, where the recognition of the target is accomplished by ensuring its correct positioning during the catalytic process. This is achieved through π - π stacking interactions involving two amino acid residues, Tyr135 and Phe198, which form an aromatic cage. Therefore, the nucleobases G1473 and A1476 of the tetraloop are crucial in coordinating between the NCD and the CRBD of the protein *Ph*DIMT1_WT. Remarkably, in our mechanistic model, the distance between the N6 atom of the target nucleobase A1476 and the S ^{δ} atom of SAM measures ~ 4.8 Å (Figure 4.11C), aligning with a prior report on methyl transfer, indicating a distance of ~ 4.7 Å (Zheng and Bruce, 1997). Additionally, this observation is consistent with the crystal structure of *Hs*TFB1M-SAM-h45 (PDB id: 6AAX) previously reported, wherein the distance between the methyl group of SAM and the N6 atom of the substrate adenosine (A973) was ~ 4.2 Å (Liu et al., 2019). These consistent observations strongly suggest that two pivotal factors crucial for initiating methyl transfer, as recognized by KsgA/DIMT1, are the maintenance of tetraloop geometry and precise positioning of the substrate.

Interestingly, a comparison of the active-site pocket of the apo and the ternary complex structures of *Ph*DIMT1_WT reveals significant differences in the spatial orientation of the loop L1. In the apo form of the protein, the loop L1 obstructs the substrate-binding pocket, whereas in the ternary complex model, the loop L1 undergoes two distinct segmented movements. In the first, the segment (residues 39-41) of the loop L1 comes closer to the substrate, directing an interaction between the O ^{$\delta 2$} of Asp40 and the O2' atom of A1476 (Figure 4.11D). The remaining segment (residues 42-48) of the loop L1 moves away from the substrate-binding pocket, creating a twist in it that aids in the formation of a hydrogen bond between the N6 atom of A1476 and the carbonyl oxygen atom of Gly43 (Figure 4.11D). Subsequently, a space is created, allowing the substrate to access the active site. Thus, segmented movements in the loop L1 are critical for substrate binding and positioning for the methyl transfer process.

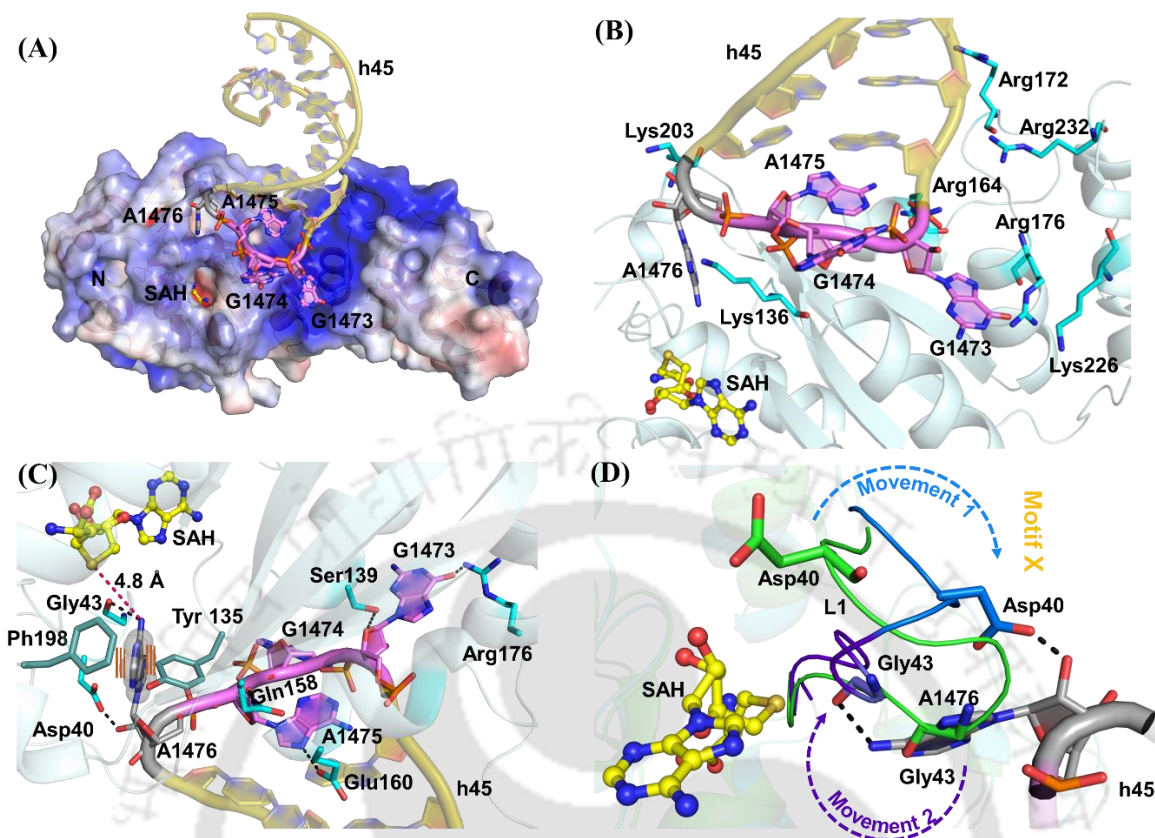


Figure 4.11. Mechanism of substrate binding. (A) Electrostatic potential charge distribution of SAH-bound *PhDIMT1_WT* (PDB id: 8X47) docked with h45. The negatively-charged 'GGAA' tetraloop of h45 binds to a positively-charged pocket near the acidic groove for the cofactor SAH. The positively- and negatively-charged surfaces are shown in blue (+5 kT/e) and red (-5 kT/e), respectively. Substrate h45 is represented as a cartoon in light olive. The first three and last bases in the h45 tetraloop are dark pink and grey, respectively. (B) Interactions between *PhDIMT1_WT* and 'GGAA' tetraloop of h45. The interacting amino acid residues are represented in the cyan sticks. (C) The diagram between the GGAA tetraloop of h45 and *PhDIMT1_WT* in the presence of SAH mimicking the mechanism of methyl transfer. The red dashed line represents the distance between the transmethylation reaction, while the black dashed lines represent the hydrogen bond formation. (D) The superposition of the ternary model with its apo form of *PhDIMT1_WT* (PDB id: 8X3W) with two segmental motions of the loop L1. Movement 1 brings the 39-41 segment (blue) into close proximity to the ribose sugar of A1476, while movement 2 twists the segment 42-48 (violet), forming interaction with the nucleobase of A1476.

4.3.9. *Ph*DIMT1 demonstrates an interdomain expansion upon cofactor binding

Previous studies have demonstrated the transition of the NCD from open to closed state upon cofactor binding to the proteins KsgA/DIMT1 (Schluckebier et al., 1995; Demirci et al., 2009; Bhujbalrao and Anand, 2019). A similar transition of the NCD is observed in the case of *Ph*DIMT1_WT as well. However, the conformational dynamics of the CRBD upon cofactor binding remain unexplained. To get insights into the CRBD movement, the open (apo) (PDB id: 8X3W) and closed (SAH-bound, PDB id: 8X47) structures of *Ph*DIMT1_WT were investigated. The analysis revealed that the CRBDs of the two structures vary significantly, with an RMSD of ~ 1.3 Å after aligning NCDs. The distance between the two residues Ser138 (α_5) from the NCD and Gly256 (α_{10}) from the CRBD in the closed state is ~ 40.1 Å, which is greater than that in the open conformations (~ 37.2 Å). This suggests that the CRBD undergoes a rigid-body translational displacement of ~ 3.0 Å (Figure 4.12A-4.12B). Further, the CRBD displays a rotation of $\sim 16^\circ$ away from the NCD from the open to the closed state of the protein (Figure B.8A). This rotation of the CRBD leads to the expansion of the interdomain interface, causing a transition from an open to a closed state of the protein (Figure 4.12B). A similar movement in the CRBD of the SAM-bound *Bs*KsgA (PDB id: 6IFT) compared to its apo form (PDB id: 6IFS) is observed (Figure B.8A-B.8B).

Notably, the expansion of the interdomain interface is observed in SAM/SAH-bound KsgA/DIMT1 structures, obtained exclusively through co-crystallization. This movement is not evident in the protein *Mj*DIMT1 crystals soaked with cofactors (PDB ids: 3GRR and 3GRY) as compared to apo structures (PDB id: 3FYC) due to comparable CRBD dynamicity with an RMSD of 0.3 Å. Furthermore, when comparing the closed forms to its apo form, it becomes evident that the distance between Ser107 (α_4) from the NCD to Met228 (α_{10}) in the CRBD remains almost constant (~ 35 Å.) This consistent observation emphasizes that the CRBD undergoes rigid-body movements (Figure B.9A-B.9B). Interestingly, unlike SAH, the movement of the CRBD in the MTA/SFG-bound of *Ph*DIMT1_WT (PDB ids: 8X44 and 8X46) structures is constrained due to the absence of an α -amino acid moiety in MTA. The loss of moiety in MTA causes the deprivation of interactions with active site residues, Glu70 and Gly70, which are crucial for achieving the closed state (Figure B.9C and Table B.3).

On the other hand, the α -amino acid moiety of SFG adopts an unusual orientation that contributes to obstructing the active site (Figure B.9D). That is, both MTA and SFG induce restricted movements of the CRBD. These observations indicate that the CRBD movement is inherent in the protein architecture and solely dependent on the type of the cofactor (SAM or SAH) bound. The residue Tyr135, located in the NCD, plays the role of a gate at the protein's cofactor-entry site. A similar observation has previously been reported in the case of the Mg²⁺-citrate binding protein belonging to the ABC transporter superfamily (Mandal and Kanaujia, 2021). In the open state of the NCD of *PhDIMT1_WT*, the residue Tyr135 repositions away from the active site, creating space for the cofactor to enter by forming a hydrogen bond between the hydroxyl atom of Tyr135 and the Nⁿ1 atom of Arg164 (Figure 4.12C).

On the other hand, cofactor encapsulation induces a conformational rearrangement in the NCD, effectively closing the active site with the shift of the residue Tyr135 towards the active site. Thus, the movement of residue Tyr135 from the active site is restricted, which disrupts the hydrogen bond between Tyr135 and Arg164 (Figure 4.12A, 4.12C). As a result, the residue Arg164 exerts a force on the CRBD, pushing it away from the NCD. This interaction is facilitated by the hinge residue Leu220. This atypical pushing leads to heightened flexibility in the helices of the CRBD, specifically in the region from α_8 to α_{10} (Table B.5). Thus, the analysis suggests that the interdomain networking initiates from the residue Arg164 in the NCD to the CRBD (Figure 4.12C). The increased flexibility might be the reason for the lack of a clear electron density in the region spanning from residues 246 to 249 of the CRBD in the SAH-bound *PhDIMT1_WT* structure (Table B.5). As a result, the loop L18 proximal to the region exhibits substantial conformational dynamicity (Table B.5). Specifically, the residues Lys226, within the α_8 helix, His231 and Lys234 in L17, as well as Asp242 and His245 in α_9 , Glu253 in L18 and Asn259 within α_{10} of SAH-bound *PhDIMT1_WT*, demonstrate the substantial structural changes when compared to their corresponding locations in the apo form of the protein (Figure 4.12C). The interface volume between the NCD and CRBD of *PhDIMT1_WT* measures $\sim 21 \text{ \AA}^3$ (Figure 4.12D). This study reveals that when SAH is bound, it leads to a significant expansion of the interdomain interface, which now occupies $\sim 53 \text{ \AA}^3$ (Figure 4.12E). This cofactor-mediated expansion is driven by the coordinated motion of the NCD and CRBD, where they

exhibit an anti-correlation movement (Figure 4.12F). This dynamic interplay was further substantiated when h45 was precisely docked with the binary complex (*Ph*DIMT1_WT•SAH) of *Ph*DIMT1_WT (Figure B.10A). Further, the tetraloop of h45 is vital for stabilizing the flexible residue Tyr135 through aromatic stacking interactions with substrate adenosine A1476 and also stabilizing the movement of the CRBD. The flexibility of the residues Arg164 and Lys226 diminishes by several molecular interactions associated with the first nucleobase of tetraloop G1473 of h45 (Figure B.10B). These interactions are instrumental in maintaining the stability and conformational dynamics of the protein. However, to further validate the structural stability of the CRBD upon binding with h45, it is necessary to elucidate the archaeal ternary complex.

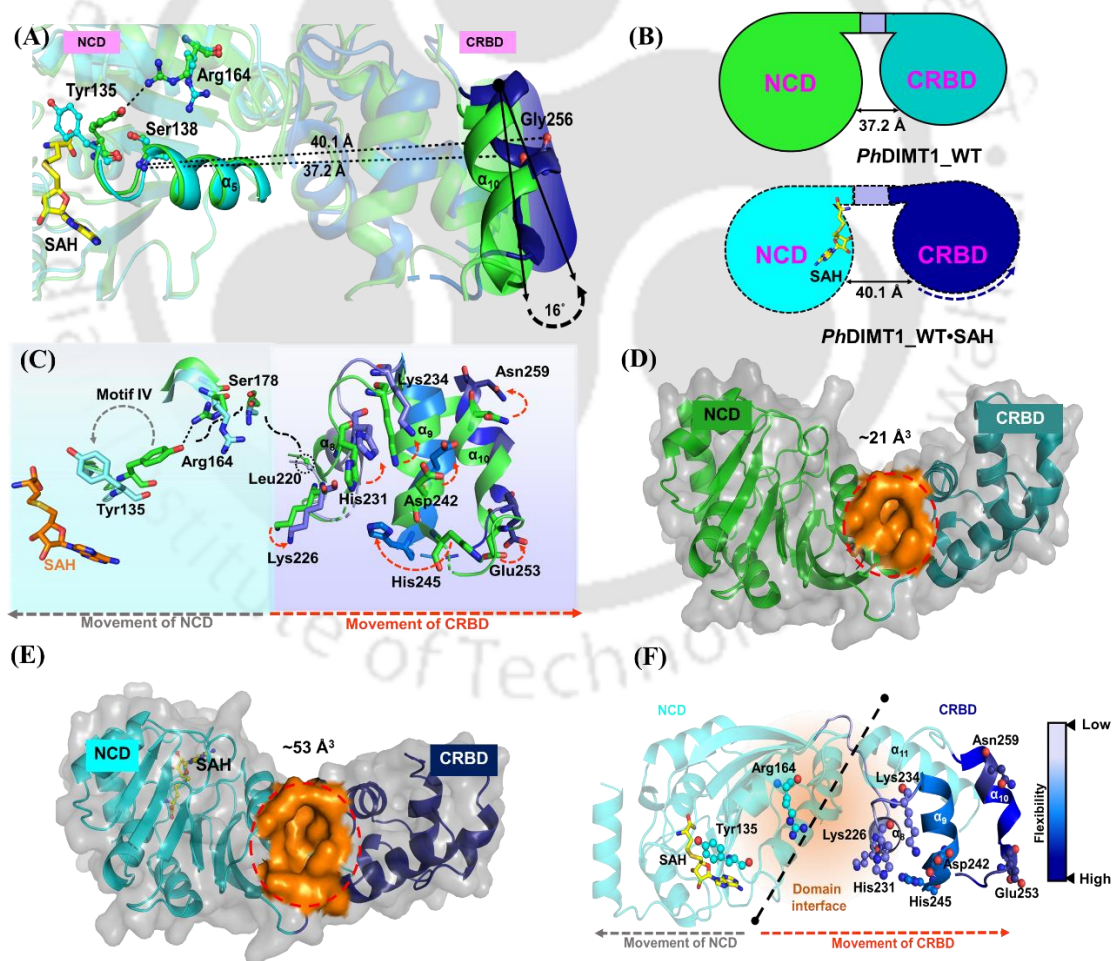


Figure 4.12 Movement of the CRBD upon cofactor binding. (A) Superposition of the apo (PDB id: 8X3W) (green) and SAH-bound (PDB id: 8X47) structures of *PhDIMIT1_WT* depicting the overall (dis)similarities, with an angular shift of 16° of the CRBD away from the active site. The conformation changes in the residues Tyr135 and Arg164 initiate the CRBD movement. The distances between Ser138 (NCD, cyan) and Gly256 (CRBD, blue) are represented by black dashed lines. (B) Schematic representation of the movement of the CRBD upon SAH (yellow stick) binding to the protein *PhDIMIT1_WT*. The conformational changes of the NCD (cyan) and CRBD (blue) upon SAH binding are depicted with dotted lines, while the movement of the CRBD is depicted by a blue dotted arrow. (C) Superposition of the apo (NCD and CRBD; green) and cofactor-bound (NCD/CRBD; cyan/blue) forms of protein *PhDIMIT1_WT* revealed a systematic conformational change upon cofactor binding, resulting in an open-to-closed transition. Cofactor binding closes the NCD by the movement of the gate residue Tyr135 towards the active site, triggering flexibility in the residue Arg164. This initiates an interaction network between Arg164 and Leu220 (hinge), enabling the CRBD to move away from the NCD. The cartoon representation for both (D) apo and (E) SAH-bound forms of the protein is shown, with the interdomain interface regions highlighted in orange, indicating volumetric expansion upon cofactor binding. (F) SAH binding to *PhDIMIT1_WT* induces flexibility in the CRBD, resulting in an expansion of the interdomain interface (orange background). The flexibility of the CRBD is represented by a blue gradient scale. The anti-correlated movement of the domains is represented by bold grey and orange dotted arrows.

4.3.10. Mutation of catalytic residues of the protein *PhDIMIT1* altered the affinity for substrate and cofactors

To investigate the specific roles of the active site residues, site-directed mutagenesis of three residues (Asp117, Asn132, and Tyr135) was performed. An analysis of the structures of these mutants (*PhDIMIT1_D117A*, *PhDIMIT1_N132A*, and *PhDIMIT1_Y135A*) with that of the wild-type protein (*PhDIMIT1_WT*) in its apo (unbound) and holo (cofactor-bound) forms indicates significant structural differences. A comparison of active-site volumes of apo (PDB id: 8X3W) and holo forms (PDB ids: 8X41, 8X46, and 8X47) of the wild-type protein reveals that the former one comprises

a larger volume (200 Å³) than the latter (126-147 Å³), hinting the closer of the protein upon the cofactor binding. Remarkably, the apo forms of the mutants, *PhDIMT1_D117A* (PDB id: 8X4G) and *PhDIMT1_N132A* (PDB id: 8X4L) comprise larger active-site volumes (261-280 Å³) than that of the wild-type protein (PDB id: 8X3W, 200 Å³), reflecting the super open form. Expectedly, the structures of the cofactor-bound mutants, *PhDIMT1_D117A*•MTA (PDB id: 8X4I), *PhDIMT1_Y135A*•MTA (PDB id: 8X4O), and *PhDIMT1_Y135A*•SFG (PDB id: 8X4P), demonstrate a reduction in the binding-pocket volume (190-223 Å³). The reduction in the active-site volume of the cofactor-bound mutants reflects that they are in an open state, as observed in the wild-type protein (Figure B.11).

To estimate and compared the binding affinities of the cofactors and the substrate (h45) between wild-type and the mutant proteins, isothermal titration calorimetry (ITC) experiments were performed. The ITC data revealed that the mutant *PhDIMT1_D117A* exhibits no binding to SAM and MTA, while a very weak binding was observed for the molecule SAH ($K_d = \sim 166 \mu\text{M}$). However, the binding of the substrate h45 is comparable with that of the wild-type protein (Figure 4.13A, Figure B.12A-B.12D, and Table B.1). This suggests that the residue Asp117 plays a crucial role in stabilizing the cofactor but has a negligible effect on the substrate binding. To provide additional clarification regarding the absence of binding, we conducted co-crystallization experiments with higher concentration of the cofactor SAM ($\sim 14 \text{ mM}$) with the mutant *PhDIMT1_D117A* ($\sim 280 \mu\text{M}$). Surprisingly, contrary to the thermodynamic data, the mutant was observed to be bound with MTA rather than SAM, similar to the wild type (Table 4.3). Although, the crystals of the mutant appeared relatively faster (within $\sim$ two months) than that of the wild-type protein (within $\sim$ nine months). In the mutant protein, the molecule MTA forms hydrogen bonds with residues His45, Glu93, and Ala118. However, unlike its wild-type counterpart, it does not interact with Gln44. Thus, this state exhibits a lower (four) number of interactions compared to that of the wild type (six) (Table B.3). Moreover, the substitution of Asp117 with alanine results in the loss of the salt bridge formed between Asp117 and Arg120 in the wild-type protein (Figure 4.13B and Table B.3). This suggests that, in the mutant *PhDIMT1_D117A*, the binding of MTA fails to facilitate the transition from an open to a closed state. Instead, it

becomes trapped in an intermediate transitional state, characterized as the cofactor-bound open form (Figure B.11).

Akin to the impact observed with Asp117, we explored the consequences of introducing an alanine mutation at the position of Asn132. Thermodynamic analysis indicated a significant reduction in the binding affinity of SAM and SAH to the *Ph*DIMT1_N132A mutant protein (Figure 4.13A, Figure B.12E-B.12F, and Table B.1). This decrease may be attributed to the displacement of structurally-invariant water molecules (IW1) (Table B.4). Validation of this phenomenon occurred during unsuccessful attempts to co-crystallize *Ph*DIMT1_N132A in the presence of SAM or SAH. Nevertheless, the apo form of the mutant could be successfully crystallized in space group C2, with a resolution of 2.6 Å (Table 4.3).

On the other hand, ITC data for this mutant revealed a complete loss of affinity for the SFG molecule (Figure B.12G). A further comparative structural analysis of *Ph*DIMT1_WT•SFG with the mutant exhibits a variance in torsion angles (Φ) between the substituted residue Ala132 (-67.6°) and the original Asn132 (-82.2°). This prevented the carbonyl oxygen atom of Ala132 from forming a critical hydrogen bond with the N^ε atom of SFG, resulting in the inability to stabilize this molecule within the pocket (Figure 4.13C). However, the ITC data for h45 to the corresponding mutant indicated no significant dissimilarity with respect to wild-type protein (Figure 4.13A, Figure B.12H, and Table B.1).

The ITC data analysis reveals a substantial reduction in the binding affinity of a cofactor for the mutant *Ph*DIMT1_Y135A protein in comparison to the wild-type *Ph*DIMT1_WT (Figure 4.12A, Figure B.12I-B.12J). This underscores the pivotal role of the residue Tyr135 in preserving the stability of the cofactor. The dynamic behavior of the *Ph*DIMT1 domains emphasizes the crucial bending of residue Tyr135 towards the NCD, a key step in facilitating domain closure upon cofactor encapsulation (Figure 4.13C). Therefore, in the absence of this residue, the active site pocket may become highly mobile. This phenomenon is likely due to the mutants' inability to crystallize in the apo state. Surprisingly, it has been observed that the mutant *Ph*DIMT1_Y135A, in the presence of a cofactor (MTA or SFG), can form a crystal (Table 4.3). However, the

structural superposition of the cofactor (SFG) bound form of *PhD*MT1 and the mutant with respect to *PhD*MT1_WT suggests that lacking Tyr135 in the mutant fails to close the NCD. It is evident that the catalytic loop L1 in the mutant is nearly static when compared to *PhD*MT1_WT•SFG (Figure 4.13D). Moreover, the orientation of loop L1 and active site volume of the mutant closely resemble that of the apo form of the wild-type *PhD*MT1, reflecting a transient state discussed above (Figure B.11).

On the other hand, the thermodynamic investigation provides additional insights, indicating that the mutant *PhD*MT1_Y135A exhibits weaker binding to the substrate compared to the wild-type protein. This is evident from the ~7-fold increase in the dissociation constant (K_d) in the presence of the substrate (Figure 4.13A, Figure B.12K, and Table B.1). The rationale for this observation can be elucidated by considering the involvement of Tyr135 in stabilizing the target adenine molecule. This stabilization is accomplished through a sandwich stacking interaction with another aromatic amino acid, Phe198 (motif VIII) (Figure 4.13C).

To further obtain the functional significance of the residues Asp117, Asn132, and Tyr135 in the substrate rRNA modification, *in vitro* MTase assays were conducted using the mutants *PhD*MT1_D117A, *PhD*MT1_N132A and *PhD*MT1_Y135A and the wild-type protein *PhD*MT1. Notably, the results revealed that the mutant *PhD*MT1_D117A showed no detectable catalytic efficiency, suggesting that residue Asp117 plays a crucial role in stabilizing cofactor SAM during catalysis of the protein *PhD*MT1_WT (Figure 4.13E and Table B.6). In contrast, the mutants *PhD*MT1_N132A and *PhD*MT1_Y135A display a significantly reduced catalytic efficiency (~89% and ~87%, respectively) compared to that of the wild-type enzyme (Fig. 4.13E and Table B.6). This indicates that these two mutations do not entirely inactivate the enzyme but affect its catalytic activity.

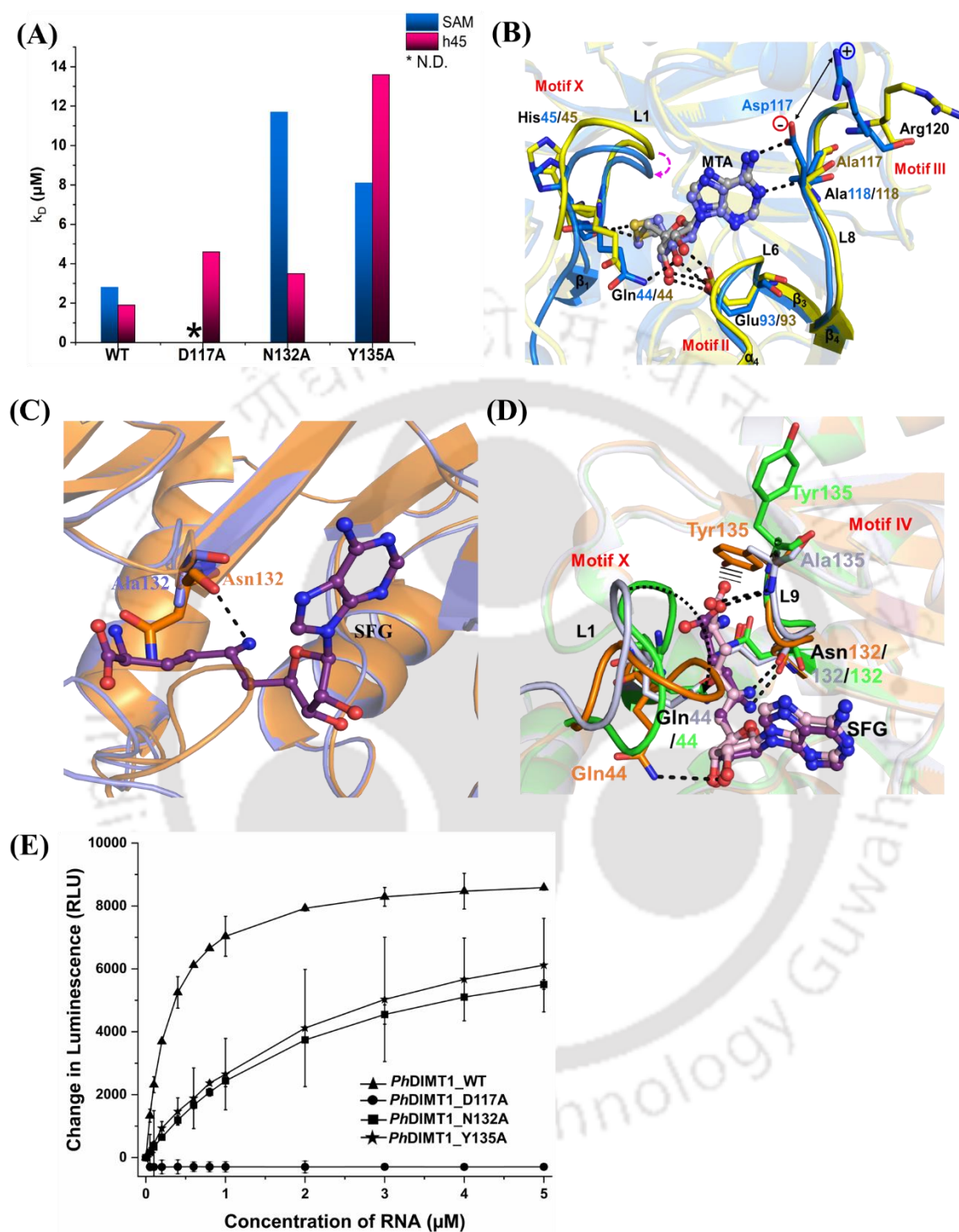


Figure 4.13. Analysis of mutant *PhDIMT1* proteins. (A) The binding affinities of SAM and the rRNA substrate h45 to the *PhDIMT1*_WT and its three different mutants (*PhDIMT1*_D117A, *PhDIMT1*_N132A, and *PhDIMT1*_Y135A) obtained using isothermal titration calorimetry (ITC). (B) Superposition of the active site residues of

the MTA-bound *PhDIMT1_WT* (PDB id: 8X41, blue) and *PhDIMT1_D117A* (PDB id: 8X4I, yellow) introduce a new state cofactor-bound open state. The cofactor MTA is shown as a ball-and-stick model, highlighted in purple and grey for the wild-type and the mutant proteins, respectively. The salt bridge between the residues Asp117 and Arg120 is represented by a double-headed black arrow. (C) Superposition of the active site residues of the SFG-bound *PhDIMT1_WT* (PDB id: 8X46, orange) and the mutant *PhDIMT1_N132A* (PDB id: 8X4L, purple) indicates the pivotal role of the residue Asn132 in the stabilization of the cofactor SFG. The cofactor SFG is shown as a violet ball-and-stick model. (D) Comparing the active site residues of *PhDIMT1_WT*•SFG (PDB id: 8X46, orange), *PhDIMT1_Y135A*•SFG (PDB id: 8X4P, light blue) with the apo form *PhDIMT1_WT* (green) revealed that the mutant complex holo form resembles its apo form. SFG bound to the mutant/wild-type protein is shown as a pink/violet ball-and-stick model. The spatial shift of the loop L1 in *PhDIMT1_WT* upon SFG binding is represented by a black dotted arrow. (E) *In vitro* MTase assay of *PhDIMT1_WT* (triangle) and the mutants *PhDIMT1_D117A* (circle), *PhDIMT1_N132A* (Square), and *PhDIMT1_Y135A* (asterisk).

4.3.11. The active site residues Asp117 and Asn132 are crucial in maintaining the structural integrity of the protein *PhDIMT1*

The structural analysis of the mutants *PhDIMT1_D117A* (PDB id: 8X4G) and *PhDIMT1_N132A* (PDB id: 8X4L) exhibits structural changes, mainly in the flexibility of the catalytic loops, L1, L6, and L9, resulting in the expansion of the active site compared to that of the wild-type protein (Figure 4.14A). This could be because the structures of these mutants were determined at low resolutions (3.3 to 2.6 Å) compared to that of the wild type (~2.0 Å). The structural analysis of these mutants revealed the loss of the invariant water molecule (IW1-IW3) (Table B.4). Thus, the residues Asp117 and Asn132 seem to be essential not only for the catalytic efficiency of *PhDIMT1_WT* but also for maintaining structural stability. The aforementioned structural observation was further corroborated through a comprehensive comparative thermal unfolding study involving *PhDIMT1_WT* and its mutants. The results revealed that all three mutants are critical for maintaining the structural integrity of the protein. The melting point (T_m) of mutant Y135A was found to be ~ 60°C comparable with wild-type protein, suggesting that the absence of the residue Tyr135, *PhDIMT1* somehow maintains structural

integrity (Figure 4.14B and Figure B.13). In contrast, the mutants *Ph*DIMT1_D117A and *Ph*DIMT1_N132A showed a drastic decrease in T_m of $\sim 30^\circ\text{C}$ (Figure 4.14C-4.14D, and Figure B.13). This indicates that the residues Asp117 and Asn132 are indispensable for maintaining the structural integrity of the protein *Ph*DIMT1_WT.

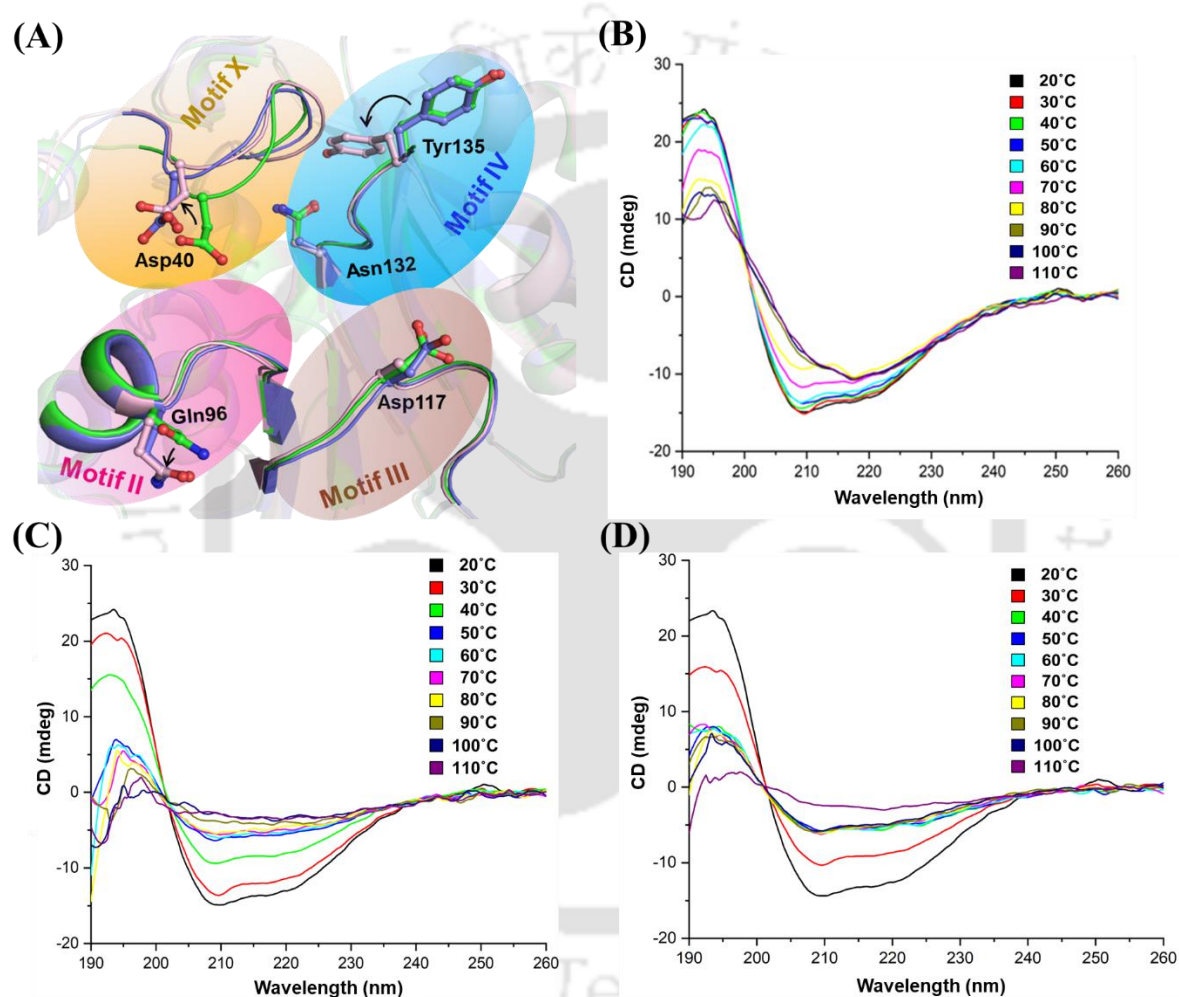


Figure 4.14. Structural integrity upon the active site residues mutations. (A) The superposition of the active site residues of the mutant proteins *Ph*DIMT1_D117A (PDB id: 8X4G, yellow), *Ph*DIMT1_N132A (PDB id: 8X4L, purple), and *Ph*DIMT1 (green). (B-D) The thermal melting analysis *Ph*DIMT1_Y135A, *Ph*DIMT1_N132A, *Ph*DIMT1_D117A was obtained from circular dichroism experiments.

4.4. Discussion

The structural and functional insights of KsgA/DIMT1 modification have been extensively studied in both bacterial and eukaryotic organisms (O'Farrell et al., 2004; Demirci et al., 2009; Tu et al., 2009; Shen et al., 2020). In contrast, the details of the archaeal counterpart have not been explored in detail. In this work, an in-depth investigation of the protein *Ph*DIMT1 in its wild-type and mutant forms (un)bound to cofactors provides mechanistic insights into the function of archaeal KsgA/DIMT1.

4.4.1. Localized conformational rearrangement contributes to the opening and closing of the cofactor-binding pocket

A previous study reported the binding of a cofactor to the protein KsgA only in the presence of substrate rRNA (h45) (Tu et al., 2009). This suggested that the binding of the substrate h45 causes an allosteric effect for the binding of the cofactor to the protein. In contrast, other studies showed the binding of cofactors even in the absence of the substrate rRNA (Demirci et al., 2009; O'Farrell et al., 2010). However, the lack of clear electron density of the catalytic loop L1 prevents the understanding of the entire active site. This is probably one of the reasons behind previous studies that could not provide insights into the protein dynamics during cofactor binding.

In this study, a comprehensive investigation of the thermodynamic properties of different cofactors that bind to the protein *Ph*DIMT1 in the absence of substrate was carried out. The results indicate that the active site of *Ph*DIMT1 is malleable, which is further supported during the comparison of the crystal structures of its unbound (apo) and cofactor-bound (holo) forms. The malleability in the active site is most likely associated with the catalytic loops of the protein. In the apo form of *Ph*DIMT1, the catalytic loop L1 partially and completely obstructs the sites for cofactors and substrates in the active site, respectively. In this state, the loop L9 is positioned in such a way that the gated residue, Tyr135, is distanced from the active site, creating an open-box-like cavity. When the cofactor binds, it induces dynamic changes in the protein, causing the segmented movements of the loop L1. These movements play a vital role in shifting the residue Gln44 away from the cofactor-binding pocket. Additionally, these segmented movements of the loops create spaces that facilitate substrate binding, promoting the stabilization of the target adenosine (A1476). This stabilization occurs through the

formation of hydrogen bonds among the loops and the substrate (A1476). At the same time, the rotation of the gated residue Tyr135 towards the active site further enhances the stability of substrate (A1476) through π - π stacking interactions. It is noteworthy that members of the KsgA/DIMT1 family display conformational changes upon cofactor binding (Demirci et al., 2009; Bhujbalrao and Anand, 2019). However, it is pertinent to highlight that these changes have been observed only in the case of co-crystallization, not in the cofactor-soaked crystal structures of KsgA/DIMT1.

Site-directed mutagenesis studies of *Ph*DIMT1 underscore the paramount importance of catalytic loops L6 and L9 in safeguarding the structural integrity of the protein. Specifically, the introduction of single-point mutations, transforming Asp117 (L6) and Asn132 (L9) to alanine in *Ph*DIMT1, resulted in a pronounced loss of structural stability. In contrast, the mutation of the gatekeeper residue Tyr135 (located in loop L9) to alanine in *Ph*DIMT1 exhibited a more subdued impact on protein stability, highlighting the resilience of this alteration. Moreover, a comparative enzymatic assay in this study unveiled a substantial reduction in catalytic efficiency for the mutants *Ph*DIMT1_D117A, *Ph*DIMT1_N132A, and *Ph*DIMT1_Y135A in comparison to the wild-type *Ph*DIMT1_WT. This decline in catalytic efficiency strongly indicates the direct involvement of these residues in the catalytic process. Further, the ITC data revealed a complete loss or a significant decrease in binding affinity of the mutant variants *Ph*DIMT1_D117A and *Ph*DIMT1_N132A, respectively, with the cofactor, compared to the substrate (h45). This was further validated by our structural analysis data, showcasing their ability to anchor the adenine base and α -amino tail of cofactors, respectively.

On the other hand, the ITC data of *Ph*DIMT1_Y135A corroborated our structural observations, emphasizing the importance of the Tyr135 residue in cofactor encapsulation. Simultaneously, this residue was also found to engage in sandwich stacking in collaboration with Phe198, contributing to the precise positioning of substrates such as A1476. This observation hints at the exclusive essentiality of Tyr135 for the enzymatic activity. Consequently, any mutation at these pivotal residues disrupts the conformational flexibility of *Ph*DIMT1, prompting the adoption of a cofactor-bound

open state by the protein. This phenomenon further signifies that cofactor encapsulation is intricately pre-encoded in the architecture of *PhD*MT1.

4.4.2. Cofactor selectivity is critical for substrate binding during the initial state of dimethylation

The methyl donors such as SAM are essential molecules in DNA and RNA methylation and play a crucial role in cellular function and regulation. Thus, their interaction with the MTases plays an important role. However, due to the instantaneous hydrolysis of SAM when incubated for a longer period (e.g., ~four weeks) at room temperature, the molecule could not be crystallized as a single entity even after several attempts. The shorter incubation periods (~four weeks) primarily yield a major hydrolyzed product, MTA. This observation was further supported by a time-dependent hydrolysis assay, where ~four weeks of SAM incubation resulted in the production of MTA. Interestingly, the molecule ADN was only observed after longer incubation periods (~nine months) in the binary complex (*PhD*MT1_WT with SAM). Therefore, it is speculated that ADN was produced from a small amount of SAM population in the crystal drop that survived the major hydrolysis and underwent minor hydrolysis.

Nevertheless, the products of SAM-hydrolysis, MTA, and ADN could be crystallized and demonstrate similar interactions due to structural conservation. SAM and its nucleoside analogs (MTA, ADN, SAH, and SFG) share a conserved adenosine component. However, they differ by the presence of different α -amino acid moieties (S ^{δ} of methionine or homocysteine in SAM or SAH and C ^{δ} of ornithine in SFG) linked at the 5' position of the conserved adenosine or the absence of this moiety (MTA and ADN). This lack of α -amino acid moieties may lead to stable but limited conformational changes. On the other hand, the variations in the α -amino acid constituents play a crucial role in determining their distinct binding orientations within a shallow groove located in the binding pocket of the protein *PhD*MT1. As a result, they have a significant influence on the dynamics of the protein. This diversity in binding interaction allows SFG to compete with SAM in the cofactor-binding pocket, effectively serving as an inhibitor for the MTase enzyme (Boeck et al., 1973; Fuller and Nagarajan, 1978). However, prior to this work, the precise binding interactions between SFG and the protein KsgA/DIMT1 were unknown.

The study reveals that SFG exhibits an unforeseen motion of its α -amino acid tail even when the invariant water molecules IW1 and IW2 are present. Despite the fundamental difference that exists between the C ^{δ} atom of SFG and the S ^{δ} atom of SAM or SAH, the primary factor driving this distinctive movement in SFG appears to be the presence of a smaller, positively charged N ^{ϵ} atom. The positively charged atom in SFG creates a hostile configuration, leading to a steric clash with the invariant water molecule IW3, displacing it from the active site. The displacement of IW3 destabilizes the α -amino acid tail and facilitates hydrogen bond formation between the positively charged N ^{ϵ} atom of SFG and Asn132. As a result, this reinforces the upward movement of the α -amino acid tail of SFG, blocking the substrate binding pocket to prevent substrate access in the active site, leading to the catalytically inactive protein.

On the other hand, KsgA/DIMT1 bound to SAM or SAH retains IW3 close to the cofactor-binding pocket by forming a hydrogen bond with the O4' atom in both cofactors. This interaction is majorly coordinated with the aid of the conserved active site residue Asn132, indirectly rendering for the cofactors' stabilization. However, the binding of SAH in *Ph*DIMT1 caused the displacement of two other invariant water molecules (IW1 and IW2), which resulted in a 2.8 Å upward movement of the nitrogen atom of the cofactor SAH. The molecule can still maintain its position by forming chalcogen bonding between its S ^{δ} atom and the carbonyl oxygen atom of Gly43 and between the carboxyl atoms of SAH and the carbonyl oxygen of Gly72. These interactions are of significant importance in maintaining the functional integrity and stability of cofactor SAH. The ternary complex generated by molecular docking the substrate h45 to its SAH-bound form showed a distance of 4.8 Å between the S ^{δ} of SAH and the N6 of substrate A1476, a measurement consistent with the prior experimental model (PDB id: 6AAX) of *Hs*TFB1M•SAM•h45 complex, which had a distance of 4.2 Å between the methyl group of SAMS and the N6 atom of the substrate (937A). This further indicates that the modeled ternary complex (*Ph*DIMT1•SAH•h45) closely resembles the 'biologically active configuration' or 'mechanistic model'. However, the proposed model might just serve as an initial enzymatic state involved in m⁶A modification in archaea and thus needs further experimental confirmation.

4.4.3. Interdomain movements of *PhDIMT1* facilitate substrate binding

A comparison of the active-site pockets of the apo (unbound) and holo (bound) forms of the protein *PhDIMT1* revealed insights into its conformational dynamics. Further, the residue Tyr135 plays the role of gatekeeper during the transition from an open to a closed conformation of the protein. The binding of the cofactor to the protein triggers the gatekeeper residue Tyr135 to close the active site. Interestingly, akin to KsgA from *B. subtilis* (*BsKsgA*), the protein *PhDIMT1* shows a rotation of 16° of its CRBD away from its NCD. Notably, the CRBD movement is specific to the cofactor-bound state of the protein. The movement can be observed only in the presence of cofactors SAM or SAH but not with MTA, ADN, or SFG. This specificity suggests that cofactors SAM and SAH induce unique conformational changes in the protein necessary for the catalytic function, indicating cofactor-induced domain movement. Further, the analysis of interdomain networking uncovered the disruption of hydrogen bonding between the residues Tyr135 and Arg164, a pivotal event that triggers a distinctive, anticorrelated motion of the NCD and CRBD. As a result, this anticorrelated movement leads to the enlargement of the interdomain interface, generating a favorable environment for the binding of the substrate h45. In summary, the conformational dynamics of *PhDIMT1* during cofactor binding, as well as the interdomain movements during the substrate binding, opens up avenues for understanding its enzymatic mechanism.

4.5. Conclusion and future perspectives

One of the noteworthy findings from our analysis underscores the preference for co-crystallization over soaking when investigating domain movements. Furthermore, this study delineates the structural and mechanistic aspects of the *PhDIMT1* protein through a comprehensive comparative analysis, juxtaposing *PhDIMT1* with other rRNA N6,6-dimethyladenosine MTases, including Erm. The extensive structural analysis highlights the critical role of the catalytic loops in their ability to encapsulate cofactors and in the positioning of the substrates into their active sites. This conformational dynamicity of *PhDIMT1* during cofactor binding suggests an induced-fit mechanism of substrate entrapment. Moreover, the protein *PhDIMT1* was co-crystallized with different cofactors, and their interactions within the binding sites were meticulously analyzed. This opens the possibility of developing novel analogs of those inhibitors with therapeutic properties to combat antibiotic resistance-causing MTases. Additionally, the

detailed analysis of structurally invariant water molecules provides valuable insights for identifying new drug target sites. Therefore, it is expected that the results obtained from this investigation will be helpful in addressing the critical issue of antibiotic resistance in the future.

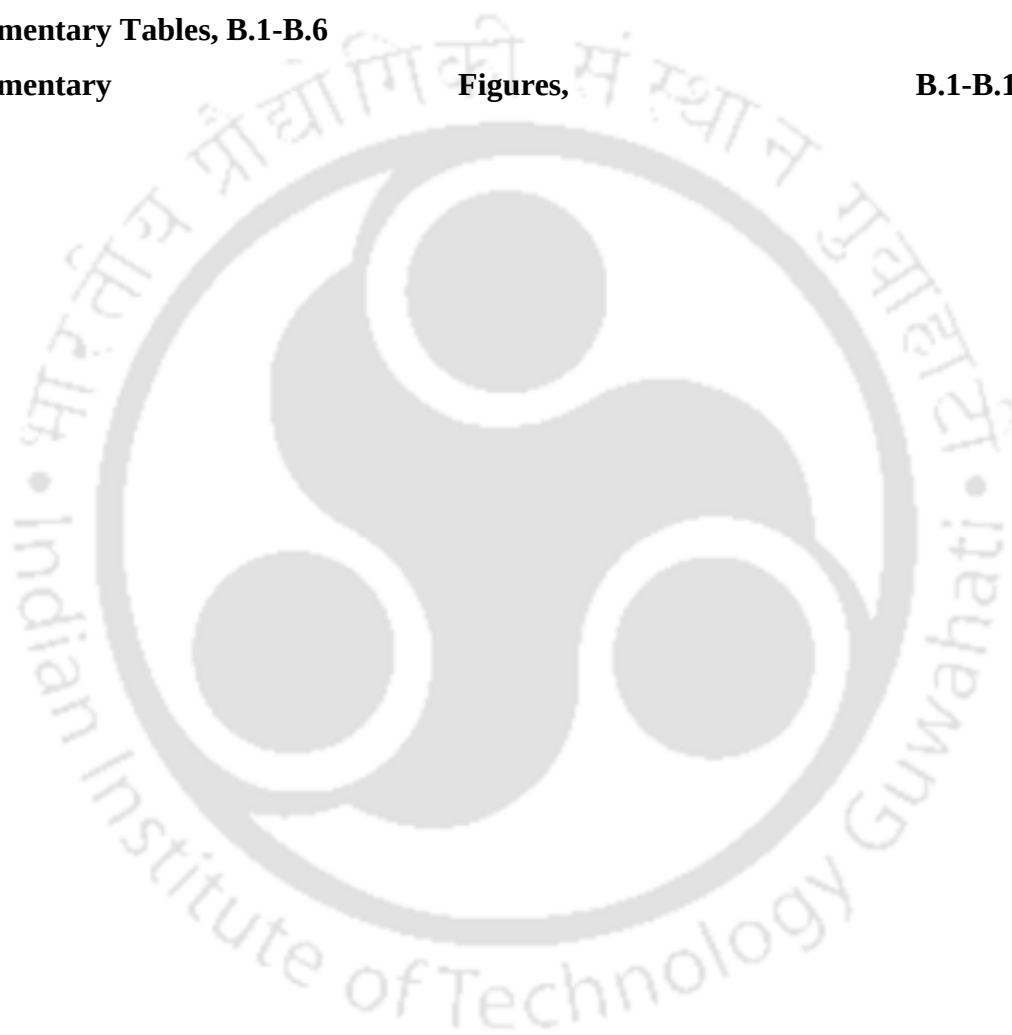
Appendix B. Supplementary data

Supplementary Tables, B.1-B.6

Supplementary

Figures,

B.1-B.13



CHAPTER 5

Structural and functional
characterization of archaeal m^5U
rRNA MTase

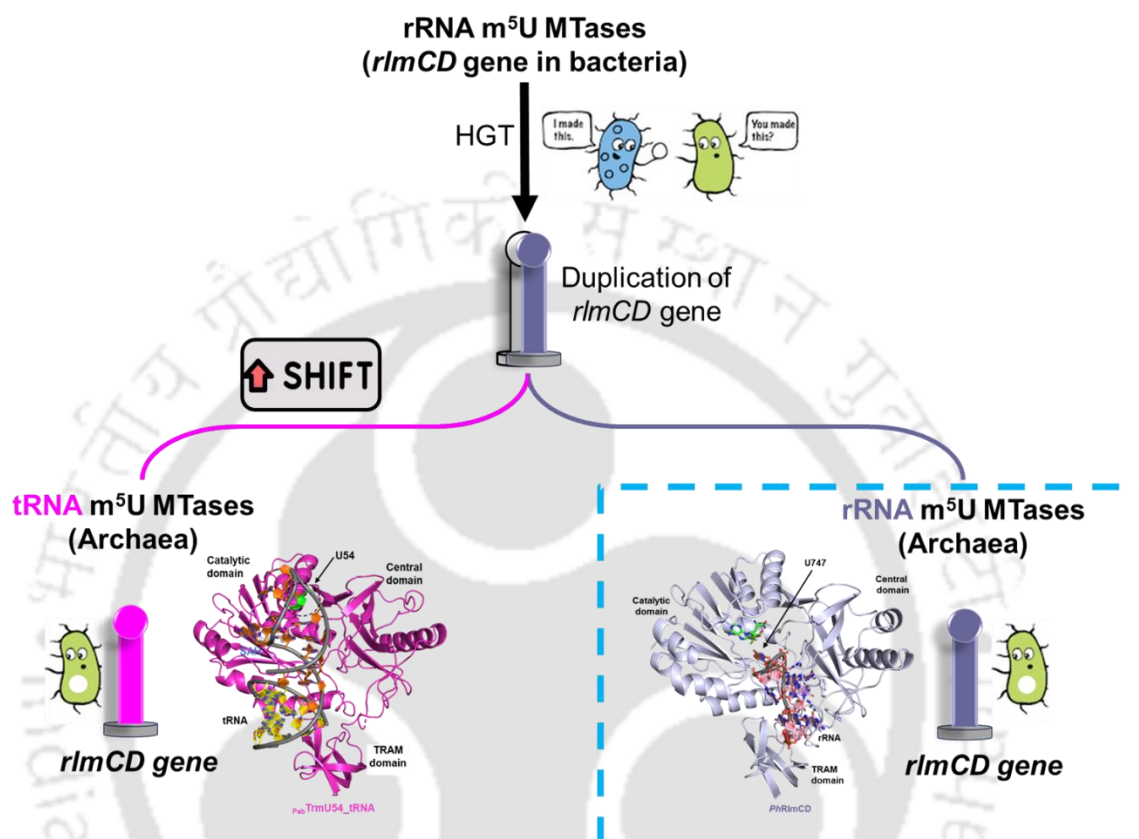
The manuscript for this chapter is under preparation as:

Saha, S. and Kanaujia, S.P. Decoding substrate selectivity of an archaeal RlmCD-like methyltransferase through its salient traits. *Biochemistry*, Accepted.

Abstract

5-methyluridine (m⁵U) rRNA modifications frequently occur at U747 and U1939 (numbering as per *Escherichia coli*) in domains II and IV of the 23S rRNA in Gram-negative bacteria, with the help of S-adenosyl-L-methionine-dependent (SAM)-dependent rRNA methyltransferases (MTases), RlmC and RlmD, respectively. In contrast, in Gram-positive bacteria, a single SAM-dependent rRNA MTase, RlmCD, modifies both the corresponding sites. In this study, a SAM-dependent RNA m⁵U (Arm⁵U) MTase was identified in the Thermococcales group of archaea. The results suggest it was horizontally transferred from Gram-positive bacteria. This study unveils the crystal structure of an rRNA-specific Arm⁵U MTase from the hyperthermophilic archaeon *Pyrococcus horikoshii* (*PhRlmCD*), closely resembling the RlmCD-like protein. The results further highlight a dynamic hinge movement induced by the replacement of the iron-sulfur cluster with disulfide bridges, obstructing the substrate binding site, suggesting the importance of the iron-sulfur cluster in the class of RNA m⁵U MTase. The in-depth structural analysis revealed the unique features of *PhRlmCD*, such as elongated positively charged loops in the central domain and rotational variations in the TRAM domain, influencing substrate selectivity. Furthermore, the results indicate a comparable interaction of the two putative mini-rRNA fragments with the protein at a positively charged cleft located at the interface of three domains, thereby facilitating dual methyltransferase activities akin to those of its ancestral protein.

Graphical abstract



5.1. Introduction

Post-transcriptional modifications are crucial in regulating the functions of non-coding RNAs, such as ribosomal RNA (rRNA) and transfer RNA (tRNA) (Machnicka et al., 2012). The modified nucleotides in rRNA act as a "molecular glue" to fix a functionally active RNA conformation, which contributes to the fine-tuning of the process of ribosome biogenesis (Woodson, 2008). These modified nucleotides are strategically positioned within the functional hubs of the ribosome, including the decoding (DC) and peptidyl transferase centers (PTC), as well as the interplay zones among ribosomal subunits (Sergiev et al., 2018). The ribosomal RNAs (rRNAs) undergo three distinct types of chemical modifications: (1) methylation of ribose sugars at the C2-position (Nm), (2) isomerization of uridines to pseudouridines (Ψ), and (3) base modifications

(e.g. methylation (mN), acetylation (acN), and aminocarboxypropylation (acpN)) (Sharma and Lafontaine, 2015). Among these, base methylation is more common and helps in creating hydrophobic properties of rRNA nucleotides, preserving the integrity of base stacking, blocking conventional hydrogen bonding, and promoting non-canonical Hoogsteen base pairing. This aids rRNAs in achieving and maintaining specific forms vital for both structural integrity and enzymatic activity (Sharma and Entian, 2022).

Among various base methylations in rRNAs, a common prokaryotic RNA (rRNA and tRNA) modification involves the conversion of uridine to 5-methyluridine (m⁵U), catalyzed by m⁵U methyltransferases (MTases).

The RNA m⁵U MTases are comprised of evolutionarily conserved motifs (I-VIII and X), involved in the transfer of methyl group from S-adenosyl-L-methionine (SAM) to the substrate (RNA) (Bujnicki et al., 2004). This process involves the attack of a cysteine residue from motif VI on the 6th position of uracil, forming a Michael adduct with the pyrimidine base. This activation allows the transfer of a methyl group from SAM, followed by deprotonation and β -elimination, releasing the methylated product and restoring the free enzyme. The rRNA m⁵C MTases have been evolved from RNA m⁵U MTases, possibly by developing a second cysteine residue in motif IV to aid in the breakdown of the covalent adduct (Kealey et al., 1994; Bujnicki et al., 2004).

In *E. coli*, RNA m⁵U modification occurs at three specific sites: (1) U54 in the T Ψ C-loop of tRNA, (2) U747, and (3) U1939 located in the loops of hairpin 35 within domains II and IV of 23S rRNA, respectively (Kealey et al., 1994; Agarwalla et al., 2002; Madsen et al., 2003). The tRNA m⁵U methyltransferase A (TrmA, also known as RumT) is known for catalyzing the methylation of U54 in bacterial tRNA and has been extensively characterized in *E. coli* (Björk and Isaksson, 1970; Keffer-Wilkes et al., 2020). This modification is also observed in the other two domains (archaea and eukarya) of life. In *Saccharomyces cerevisiae*, the modification is performed by the enzyme Trm2 (Nordlund et al., 2000). In *Pyrococcus abyssi* the enzyme _{Pab}TrmU54 has been reported to be responsible for the modification (Urbonavičius et al., 2008; Walbott et al., 2008).

In *E. coli*, two different genes, *ybjF*, and *ygcA*, encode potential rRNA m⁵U MTases, known to methylate U747 and U1939, respectively, of 23S rRNA (Gustafsson et al., 1996). The two genes *ybjF*, and *ygcA* are also designated as ribosomal RNA large subunit methyltransferase C (RlmC or RumB) and ribosomal RNA large subunit methyltransferase D (RlmD or RumA), respectively (Agarwalla et al., 2002; Madsen et al., 2003). On the other hand, in Gram-positive bacteria such as *Streptococcus pneumoniae* and *Bacillus subtilis*, a dual-function methyltransferase, RlmCD, and its homolog YefA, has been demonstrated to catalyze the methylation of U747 and U1939 in 23S rRNA (Desmolaize et al., 2011; Jiang et al., 2017).

The two m⁵U MTases (TrmA and RlmC) subfamily exhibit a two-domain (central and catalytic) organization. In contrast, the bacterial m⁵U MTases RlmD and RlmCD, eukaryotic Trm2, and archaeal m⁵U MTases contain an N-terminal extension in addition to the central and catalytic domains. The N-terminal extension shares sequence similarity with the "TRAM" domain, a term derived from the conserved sequence found in the Trm2 family of eukaryotes (Anantharaman et al., 2001, 2002; Urbonavičius et al., 2008). Despite these structural differences, it is noteworthy that all RNA (Uracil, C5)-MTases, including RlmC, RlmD, RlmCD, TrmA, and Trm2, belong to the same family of SAM-dependent MTases (Cluster of Orthologous Group COG2265). This shared membership suggests a common ancestral origin for these enzymes (Anantharaman et al., 2001).

The structural details both in apo and holo forms of bacterial rRNA and tRNA m⁵U MTases have been reported (Lee et al., 2004, 2005; Jiang et al., 2017; Jiang et al., 2018; Keffer-Wilkes et al., 2020). Notably, m⁵U MTases in archaea (Arm⁵U MTases) are found only in certain species, specifically within the order of Thermococcales. Interestingly, the Arm⁵U MTases have undergone horizontal transfer from bacteria and subsequent duplication, resulting in two distinct copies (Urbonavičius et al., 2008). These duplicated genes have further evolved to acquire distinct target specificities. A notable example can be observed in the genus *Pyrococcus* (e.g., *P. abyssi*), where the genome contains two genes encoding SAM-dependent m⁵U MTases for tRNA (PAB0719) and rRNA (PAB0760). The crystal structure of tRNA-specific m⁵U MTase

of PAB0719 (hereafter referred to as *pabTrmU54*) has been determined in complex with the product SAH (Walbott et al., 2008). However, no structural details of Arm⁵U rRNA MTases have been reported in the literature to date.

In this study, evolutionary and structural details of archaeal m⁵U MTases from *Pyrococcus horikoshii* OT3 (ORF id: PH1259) have been reported. The results of this study revealed that the gene PH1259 encodes an rRNA-specific m⁵U MTase and has likely originated from a single horizontal gene transfer event from Gram-positive bacteria (RlmCD) rather than Gram-negative bacteria (RlmD or RumA). Thus, it is referred to here as *PhRlmCD*. The crystal structure of *PhRlmCD* unveiled the existence of three distinct and well-organized (TRAM, central, and catalytic) domains. Notably, the structural investigation revealed crucial structural alterations induced by the substitution of the iron-sulfur cluster with disulfide bridges. Further, the comprehensive study introduces the signature structural trail of *PhRlmCD*, offering mechanistic insights into substrate selectivity as well as dual specificity.

5.2. Materials and methods

5.2.1. Construction of wild-type expression plasmid

The gene *PhrlmCD* (ORF id: PH1259, 1239 bp) was cloned into the pET28a(+) plasmid vector to create a recombinant construct. The gene was obtained by amplifying it from the *Pyrococcus horikoshii* OT3 genome as a template. To amplify the gene, forward (5'-CAGAC**GAATTC**ATGAGGGGGAAGATAACCAAGATCAACGAAAATG) and reverse (5'-CCGG**CTCGAG**TCACCTTCTAACGAGCTTAATTACA GCTTCAATATG-3') primers with *EcoRI* and *XhoI* (bold and underlined) restriction sites, respectively, were used in the polymerase chain reaction (PCR) (Figure 5.1A). The 6xHis-tag was utilized from the vector at the 5'-end of the cloned gene. The amplified gene was double-digested with *EcoRI* and *XhoI* restriction enzymes before being inserted into the pET28a(+) vector that had been excised with the same enzymes. The inserted gene was ligated using T4 DNA ligase. The *E. coli* DH5 α competent cells were transformed with the recombinant plasmid containing the gene *PhrlmCD* for plasmid amplification. The recombinant construct was confirmed by double digestion, followed by plasmid sequencing (Figure 5.1B).

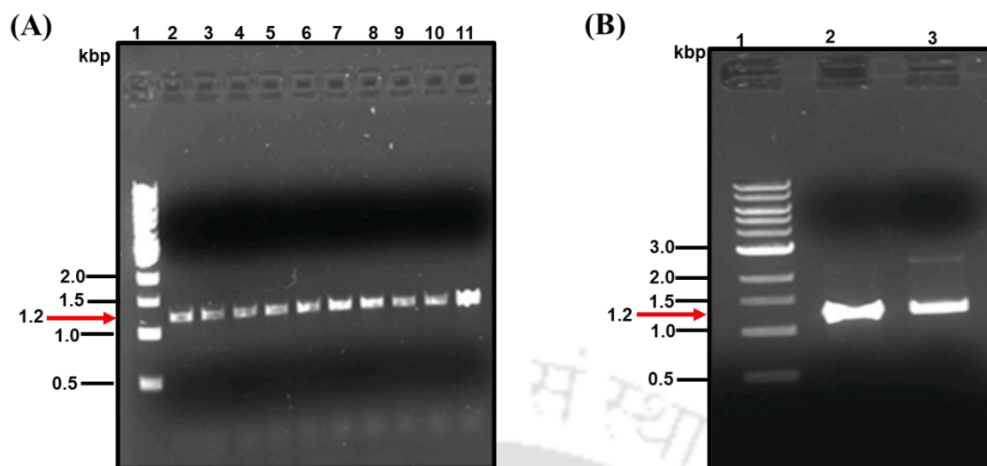


Figure 5.1. Cloning of gene *PhrlmCD*. (A) PCR amplification of *PhrlmCD* (1239 bp) gene (lane 1: DNA ladder, lane 2-11: PCR amplified product of the gene *PhrlmCD*). (B) Clone confirmation by double digestion of the vector pET-28a(+) carrying the gene *PhrlmCD* by *EcoRI* and *XhoI* restriction enzymes (lane 1: DNA ladder, lane 2: gene of interest, lane 3: insert pop out after double digestion of the plasmid pET-28a(+)).

5.2.2. Protein overexpression and purification

The *E. coli* Rosetta (DE3) competent cells (Novagen, India) were transformed with the recombinant construct of *PhrlmCD* for protein overexpression. The transformed cells were grown in Luria-Bertani (LB) medium at 37°C with 50 µg mL⁻¹ kanamycin and 34 µg mL⁻¹ chloramphenicol until the optical density (OD) of cells at 600 nm reached 0.6–0.8, 30 minutes before the induction of cells with 1 mM Isopropyl-1-thio-β-d-galactopyranoside (IPTG), 3% (v/v) alcohol was added to the culture (Figure 5.2A). The overexpressed cells were harvested by centrifugation at around 4000g for 15 minutes and resuspended in lysis buffer A, which contained 50 mM HEPES pH 7.5, 5 mM imidazole, 300 mM NaCl, 5% glycerol, 5 mM Dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), 50 mM L-Glutamic acid and L-Arginine. The cells were further grown for an additional 12 hours at 25°C with shaking at 180 rpm. The cells were then harvested by centrifugation at ~4000g for 15 minutes at 4°C and stored at -20°C overnight. Subsequently, cell pellets were resuspended in a lysis buffer A (50 mM HEPES pH 7.5, 5 mM imidazole, 300 mM NaCl, 5% glycerol, 5 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 50 mM L-Glutamic acid and L-Arginine mixture) on ice. The cells were then lysed by sonication

at 30% amplitude using a 3 second on and 5 second off cycle for 15 minutes. As an initial purification step, the cell lysate was heated at 60°C for 15 min. The heated cell lysate was subsequently subjected to centrifugation at ~18500g for 1 hour at 4°C to separate the soluble and insoluble fractions. A clear supernatant fraction was loaded onto a pre-equilibrated pierce centrifuge column (Thermo Fisher Scientific, USA) packed with Ni-nitrilotriacetic acid (Ni-NTA) affinity resin (Qiagen, Germany) for protein binding. After 3 hours of incubation at 4°C, the column was washed in a gradient manner with 10 column volumes each of buffer B (50 mM HEPES pH 7.5, 10 mM imidazole, 300 mM NaCl, 5% glycerol, 5 mM DTT, and 1 mM PMSF), and buffer C (buffer B with 20 mM imidazole) followed by a final wash using 10 column volumes of buffer D (buffer B with 50 mM imidazole). The recombinant protein(s) bound to the Ni-NTA resin was then eluted with 250 mM imidazole in buffer B. The eluted fractions were subjected to gradient dialysis to remove imidazole against the final protein buffer E (50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol) (Figure 5.2B). The dialyzed protein *PhRlmCD* was concentrated in a Vivaspın turbo 15 centricon (10 kDa cut-off; Sartorius, Germany) up to a concentration of ~18 mg mL⁻¹(Figure 5.2C).

For a metal-free *PhRlmCD*, a similar approach of protein purification was followed, except that at the time of dialysis, the buffer was supplemented with 10 mM ethylenediamine tetraacetic acid (EDTA) in two consecutive steps. However, the last two steps of dialysis were performed in the buffer E in the absence of EDTA. The EDTA-chelated protein (referred to as *PhRlmCD*_EDTA) was concentrated to ~6 mg mL⁻¹.

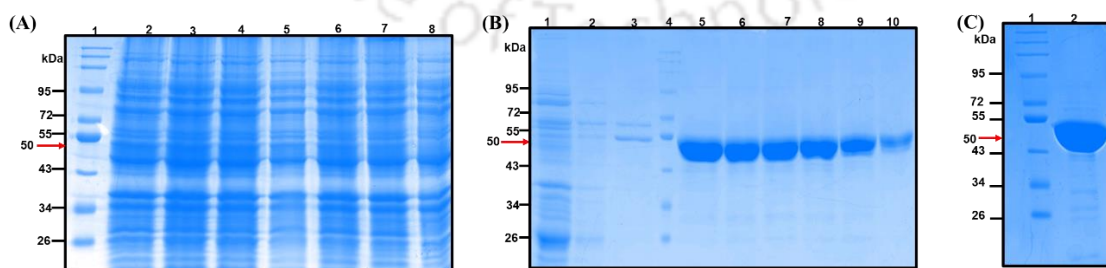


Figure 5.2. Purification of *PhRlmCD*. (A) Overexpression of the protein *PhRlmCD* in *E. coli* Rosetta (DE3) competent cell at 1 mM IPTG at a different time interval (lane 1:

protein marker, lane 2-7: 1-7 hr, lane 8: uninduced or 0 hrs) after induction. (B) SDS-PAGE analysis for the eluted fractions of the protein using Ni-NTA metal-affinity chromatography (lanes 1-3, Unbound, Flow through 1 and 2 & 5-10: eluted fractions of the protein respectively, lane 4: protein marker). (C) SDS-PAGE analysis for the concentrated protein (lane 1: protein marker, lane 2: concentrated protein).

5.2.3. Mass spectrometric (MS) analysis

The molecular weight of the protein *PhRlmCD* was determined using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry. For the measurement, the purified *PhRlmCD* was mixed with the sinapinic acid matrix at ratios of 1:2, 1:3, and 1:5. The sinapinic acid was dissolved in a mixture of acetonitrile and 0.1% trifluoroacetic acid in a 30:70 ratio, sonicated for 30 minutes, and then centrifuged at ~12000xg for 10 minutes to eliminate excess crystals. The prepared matrix was then mixed with varying concentrations (1, 5, and 10 mg mL⁻¹) of the protein *PhRlmCD*. Using a Bruker autoflex speed mass spectrometer (Bruker, Germany) equipped with a 2 kHz smart beam-II solid-state laser for enhanced speed, 2 µL droplets of the protein-matrix mixture were deposited onto the grooves of a steel MALDI plate and kept to air dry. The droplets were then ionized within the mass spectrometer, and the data obtained were analyzed using the built-in Flex control analysis software.

5.2.4. Dynamic light scattering (DLS) analysis

Hydrodynamic diameter of the protein *PhRlmCD* was measured employing a LITESIZER 500 (Anton Paar UK) equipped with a cuvette having a path length of 1 cm. The purified protein was appropriately diluted in nuclease-free water to a final concentration of ~1 µM. Dynamic light scattering (DLS) experiments were then conducted for the prepared solution, utilizing side scatter angles as determined by the instrument's specifications. The collected data were utilized to construct a two-dimensional graph using the program Origin v9.6.

5.2.5. Circular dichroism (CD) analysis

The circular dichroism (CD) spectra of *PhRlmCD* in the far-ultraviolet range (190 to 260 nm) were measured using a Jasco J-1500 spectrometer (Jasco, Germany). The

measurements were taken in a 1 mm quartz cuvette. The protein (5 µM) sample was dialyzed in buffer F (1 mM Tris–HCl pH 7.5 and 5 mM NaCl). During spectral measurements, the response time was fixed at 2 second, sensitivity at 100m degrees (mdeg), and scan speed at 100 nm min⁻¹. Each measurement was averaged over three scans, and the blank signal was obtained using the same buffer as a reference and was subtracted. CD spectra analysis and secondary structural content were obtained using JASCO's Spectra Manager software. An optimized spectral profile was first determined to analyze the thermal denaturation profile of *PhRlmCD*. Subsequently, thermal unfolding curves were monitored by subjecting the protein *PhRlmCD* sample to CD signal measurements over a temperature range of 10 to 110 °C, with a heating rate of 2 °C per minute. Changes in the secondary structure of *PhRlmCD* were monitored between 190 and 260 nm wavelengths. To ascertain the melting point (T_m) via circular dichroism (CD), the ellipticity at 208 nm was recorded across a temperature range. The apparent fraction of unfolded molecules (F_U) was determined using the formula:

$$F_U = \frac{(Y_{obs} - Y)}{(Y_U - Y)}$$

Where Y_{obs} represents the observed ellipticity value (θ) at 208 nm for each temperature, while Y_U and Y denote the observed values in the baseline regions for the unfolded and folded forms, respectively (Singh and Manoj, 2017). Subsequently, a graph was plotted, displaying the fraction unfolded (F_U) against temperature. The graph was represented using the Boltzmann sigmoid equation.

5.2.6. Crystallization and co-crystallization of the protein *PhRlmCD*

Preliminary screening to crystallize the purified *PhRlmCD* (~18 mg mL⁻¹) was performed using various kits such as crystal screen 1, crystal screen 2, PEG/Ion 1 & 2, Natrix and Index 1 & 2 kits (Hampton Research, USA). The crystallization setup was executed using the microbatch-under-oil technique by mixing 2 µL each of protein and buffer (1:1 ratio) at 4 and 20 °C. Initial crystal hits of the protein *PhRlmCD* were obtained in three different buffer conditions, (1) 0.5 M ammonium sulfate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 1.0 M lithium sulfate monohydrate, (2) 0.1 M

BIS-TRIS pH 5.5, 2.0 M ammonium sulfate, and (3) 0.1 M HEPES sodium pH 7.5, 2% (v/v) PEG 400, 2.0 M ammonium sulfate at 20 °C (Figure 5.3).

All three crystallization conditions were utilized for co-crystallization of the protein *PhRlmCD* with cofactors such as SAM and SAH. The co-crystallization of the protein with cofactors was performed using different approaches. (1) The protein was co-crystallized with varying concentrations (12–50X) of the cofactors by incubating overnight separately at 4 and 20 °C. (2) A mixture of the protein with varying concentrations (12–50X) of the cofactors was heated at 310K for 5–10 minutes, followed by incubation on ice for 15 min. (3) The prepared ligands were directly added at the time of setting up the crystal drop at varying concentrations. Prior to each crystal setup, the mixture was centrifuged and filtered (0.2 µm). Although multiple small crystals were obtained using the above co-crystallization experiments, they failed to produce collectible X-ray diffraction.

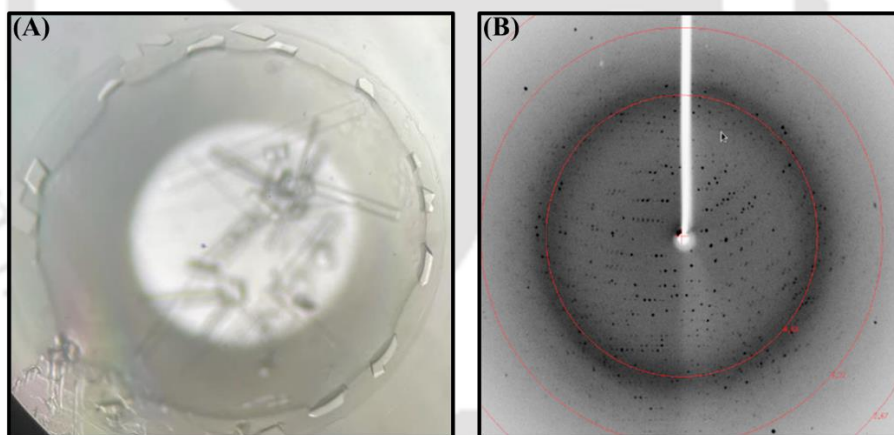


Figure 5.3. Crystal and diffraction of *PhRlmCD*. (A) Overview of the obtained crystals of *PhRlmCD*. (B) The diffraction pattern produced by the crystals.

5.2.7. Data collection, processing, and structure determination

X-ray diffraction intensity data the protein crystals of *PhRlmCD* were collected at a home source utilizing the Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) available at the central instruments facility of the Indian Institute of

Technology Guwahati, India, at -173°C. During the data collection, each frame was exposed for 300 seconds and rotated at 1° oscillation keeping the detector at a distance of 190 mm. Although multiple small co-crystals were obtained, they failed to produce collectable diffraction data. The X-ray intensity data were processed and scaled using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013). The intensity data were converted to structure factors utilizing the modules in the CCP4 package (Evans, 2011; Winn et al., 2011). The structure solution of the protein *PhRlmCD* was obtained using the molecular replacement method using the program Phaser (McCoy et al., 2007). The AlphaFold model of the protein *PhRlmCD* (AlphaFold id: AF_AFO58994F1) was utilized as the search model during the molecular replacement. For R_{free} calculations, a total of 5% reflections were set aside (Brünger, 1992). The structural models were refined using the program Refmac5 (Murshudov et al., 2011) followed by model building using the program Coot v0.9.4.1 (Emsley et al., 2010). The protein atoms were fitted into the electron density maps contoured at 3.0σ and 1.0σ for $F_o - F_c$ and $2F_o - F_c$ maps, respectively. Subsequently, the electron densities corresponding to water molecules and other components present in the protein and crystallization buffer were fitted. Geometrical and chemical parameters of the refined structures were assessed and validated using the programs PROCHECK and MolProbity (Laskowski et al., 1993; Williams et al., 2018). The refinement and validation statistics for the refined models are provided in Table 5.1. The three-dimensional atomic and structure factors of the refined protein *PhRlmCD* will be deposited in the RCSB Protein Data Bank. (PDB) (Berman et al., 2002).

Table 5.1. X-ray crystallographic data collection and refinement statistics for the protein *PhRlmCD*. The values in parenthesis represent the statistics for the last resolution shell.

	<i>PhRlmCD</i>
Wavelength (Å)	1.5418
Temperature (K)	100
Space group	$P2_12_12_1$
Unit-cell parameters (Å,°)	$a=91.89, b=96.61, c=122.40$
Resolution (Å)	75.83-3.20 (3.42-3.20)
No. of observed reflections	149497 (26915)

**Chapter 5. Structural and functional characterization of archaeal m⁵U rRNA
MTase**

No. of unique reflections	18162 (3314)
Mn(I) CC(1/2)	0.986 (0.898)
Completeness (%)	100 (100)
$V_M (\text{\AA}^3 \text{ Da}^{-1})$	2.32
Solvent content (%)	47.00
Mosaicity (°)	0.50
Mean I/ σ (I)	8.3 (3.5)
$R_{\text{merge}}^{\dagger}$ (%)	21.2 (50.4)
R_{pim} (%)	11.6 (27.6)
R_{meas} (%)	24.2 (57.6)
Multiplicity	8.0 (8.1)
$R_{\text{work}}/R_{\text{free}}$ (%)	16.92/23.72
Protein model	
No. of subunits in the asymmetric unit	2
Protein atoms	6634
Water molecules	110
Other molecules	12
Deviation from ideal geometry	
Bond length (Å)	0.012
Bond angles (°)	1.658
Average B-factor (Å²)	
Protein atoms	28.89
Water molecules	21.322
Ramachandran plot (%)	
Favoured	93.11
Allowed	6.89
PDB id	8Z62

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum overall reflections and $\sum_i$ is the sum over i measurements of reflection hkl .

5.2.8. Isothermal titration calorimetry (ITC) analysis

The binding affinity of the cofactors (SAM, SAH, and SFG) was estimated by performing isothermal titration calorimetry (ITC) experiments using the MicroCal iTC200 (Malvern, UK). The protein and ligand (cofactor) solutions were prepared in the buffer E (50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol). For each experiment, the protein *PhRlmCD* (68-200 μ M) filled in the sample cell was titrated with cofactors (3-10 mM) (Table C.1). During each titration, the sample cell was agitated at speeds of 250/450 rpm, and each cofactor was injected 24 times with an injection volume of 1.5 μ L at an interval of 150 second. The thermodynamic parameters of the protein *PhRlmCD* were measured at two different temperatures, 25 and 60 °C. For each measurement, heat changes induced by the dilution of the cofactor in the buffer were determined and subtracted from the heat of the reactions. Finally, the corrected data was analyzed using the program Origin v9.6, embedded in MicroCal (GE Healthcare). The thermodynamic parameters (such as association constant (K_a), enthalpy (ΔH) and entropy (ΔS) changes, and stoichiometric ratio (the number of binding sites, n)), were as estimated by integrating the reaction heat using non-linear regression and a one-site binding model. The Gibbs free energy change (ΔG) was calculated using the thermodynamic equation $\Delta G = \Delta H - T\Delta S$, where T represents the reaction temperature. The estimated values for these thermodynamic parameters of each reaction are provided in Table C.1.

5.2.9. UV-visible absorbance spectroscopy

The absorption peaks at ~ 325 and ~ 410 nm are indicative of the Fe-S-containing protein. Thus, the presence of the Fe-S cluster in the protein *PhRlmCD* was assessed by recording absorbance spectra between 200 and 700 nm using a double-beam LMSP-UV1900S spectrophotometer (LABMAN, India). Both the types of protein (*PhRlmCD* and *PhRlmCD*_EDTA) were utilized for the measurement. The peak corresponding to the Fe-S cluster was determined by subtracting the spectra of the control buffer E (50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol) from the experimental data.

5.2.10. Energy-dispersive X-ray (EDX) analysis

Energy-dispersive X-ray (EDX) microanalysis was performed on both the proteins (*PhRlmCD* and *PhRlmCD*_EDTA) using Jeol 2100F transmission electron microscope

(TEM) equipped with an 80T EDX detector (Oxford Instruments, Japan) and a GeminiSEM 300 scanning electron microscope (Zeiss, Germany). To prepare protein samples for analysis, carbon coating was applied using carbon film (300 mesh) from Electron Microscopy Sciences (Hatfield, USA). The analysis was carried out under vacuum conditions with a 20 kV acceleration voltage, a 10 mm working distance, 40000X magnification, and a 100 nm beam size. Probe currents ranging from 0.75 to 1.50 nA were measured in a Faraday cup, enabling spectral acquisition rates of 10000–15000 counts per second (cps) with a dead time of ~10%. The total acquisition time, typically not exceeding 40 seconds, with a total count range of ~310000–320000, was collected using the program INCA. For EDX spectral data processing, the background subtraction, peak fitting, deconvolution, and pulse pile-up correction were performed using a filtered least squares method using the program INCA. The quantitative data were obtained after matrix correction, following the XPP exponential model procedure. As the presence of numerous light elements in the samples hinders the overall signal-to-noise (S/N) ratio, the sample was coated with a 50 nm carbon layer, selecting all elements except indium as deconvolution elements.

5.2.11. Bioinformatic analysis

The list and amino acid sequences of rRNA MTases were collected from the UniProtKB database using specific keywords such as "RlmC," "RlmD," "RlmCD," "RumA," or "m⁵U MTase" (The UniProt Consortium, 2023). The homologs of the rRNA MTases of *P. horikoshii* were identified using the web tool BLAST, keeping a sequence identity (SI) cut-off of more than 20% (Altschul et al., 1990). To establish the evolutionary origin of archaeal proteins, phylogenetic trees were constructed with respect to three domain life using the maximum-likelihood (ML) method in MEGAX, with 1000 bootstrap replicates (Kumar et al., 2018). The resulting trees were visualized using the web tool iTOL v6.3 (Letunic and Bork, 2021). The multiple sequence alignment (MSA) of the *PhRlmCD* and its homologs was performed using the web tool Clustal Omega (Sievers and Higgins, 2014). Further, outcomes of MSA were rendered using the web tool ESPript v3.0 (Gouet et al., 2003). The interdomain movement of the proteins was predicted using the web server DynDom (Hayward and Berendsen, 1998). The binding affinities of the protein *PhRlmCD* with SAM were estimated by molecular docking using a freely available program, AutoDock v4.2 (Morris et al. 2009). The 23S

rRNA sequence of *P. horikoshii* was obtained from RNA Central (RNAcentral Consortium, 2021). The conserved hairpin 35 motif of domain II and IV in *P. horikoshii* were identified by comparing the secondary structure of *E. coli* 23S rRNA. Subsequently, the three-dimensional structure of the required fragments of substrate rRNAs was modeled and docked using the web server ModeRNA and HDOCK, respectively (Rother et al., 2011; He et al., 2019; Yan et al., 2020). The molecular interactions between the protein and the substrate rRNA were observed using the program Coot (Emsley et al., 2010).

5.3. Results

5.3.1. Arm⁵U MTase is horizontally acquired from Gram-positive bacteria

A previous study showed a comprehensive genomic analysis of 410 complete genome sequences retrieved from the NCBI database, which revealed a total of 456 homologs of RNA m⁵U MTases, comprising 31 RlmC-like, 303 RlmD-like, 49 Trm2-like, 64 TrmA-like, and 9 Arm⁵U MTases (Urbonavičius et al., 2008). In archaea, two distinct copies of Arm⁵U MTases were identified in members of the Thermococcales group, encompassing *P. abyssi*, *P. furiosus*, *P. horikoshii*, and *Thermococcus kodakarensis*. In contrast, *Nanoarchaeum equitans*, a member of Nanoarchaea, possesses a single copy of Arm⁵U MTase (Huber et al., 2002).

An in-depth investigation of Arm⁵U MTase in *P. horikoshii* revealed the presence of two genes, *PH1259* and *PH1137*. In the UniProtKB database, PH1259 and PH1137 were annotated as "23S m⁵U rRNA MTase" and "m⁵U tRNA MTase", respectively. A homology search for PH1259 revealed its similarity with two copies of Arm⁵U MTases from *P. abyssi*, viz. PAB0719 (*Pab*TrmU54, QC: 99%, SI: 44%) and PAB0760 (QC: 99%, SI: 75%). Previous studies have demonstrated that PAB0719 (*Pab*TrmU54) is specific to tRNA, while PAB0760 methylates the U747 site in 23S rRNA fragments and function as a homolog of RlmC-like MTases (Urbonavičius et al., 2008; Auxilien et al., 2011). Interestingly, a comparison of rRNA m⁵U MTases (RlmC, RlmD in Gram-negative and RlmCD in Gram-positive bacteria) across domains of life revealed significant differences among them. In particular, both RlmD-like and RlmCD-like MTases possess the unique TRAM domain at their N-terminal, distinguishing them

from RlmC-like MTases, which lack the TRAM domain (Table C.2). The presence of the TRAM domain in the rRNA-specific Arm⁵U MTases (e.g. PAB0760 and PH1259) enables them to identify both the sites (U747 and U1939) within the 23S rRNA.

To further confirm the dual site-specificity of Arm⁵U MTases, a phylogenetic analysis was performed for the protein PH1259 and its homologs. The results exhibited the two distinct groups, (1) tRNA-specific (e.g. TrmA, Trm2, and archaeal MTases; variants B_{a1}, B_{a2}, and B_{a3} respectively) and (2) rRNA-specific (e.g. RlmC, RlmD, RlmCD, and archaeal MTases; variants A_{b1}, A_{b2} and A_{b3}, respectively). Notably, further investigation revealed that the Arm⁵U MTases share a closer evolutionary relationship with rRNA m⁵U MTases from Gram-positive bacteria (e.g. RlmCD, variants A_{b3}) rather than those from Gram-negative bacteria (e.g. RlmC variants A_{b2} and RlmD A_{b1}) (Figure 5.4A). This observation indicates that the Arm⁵U MTases likely evolved from Gram-positive bacteria through horizontal gene transfer (HGT) (Figure 5.4B). The possible evolutionary trajectory seems to be that the ancestral bacterial gene specific to rRNA modification (e.g. RlmCD, variant A_{b3}) shifted its function to tRNA modification during the gene duplication event. This functional shift may be attributed to the absence of the original selective pressure on the duplicated gene copy, resulting in the gene acquiring a different function.

In *P. horikoshii*, such an evolutionary process might have resulted in the presence of two closely related RNA m⁵U MTases (PH1259 and PH1137) (Figure 5.4B). The protein PH1259 is claded with protein PAB0760 compared to its counter copy PH1137, which is claded with _{pab}TrmU54 (PAB0719). Moreover, the protein PH1259 exhibited the highest similarity with its ancestral homologs (RlmCD-like, rRNA m⁵U MTases), therefore referred to as *PhRlmCD*.

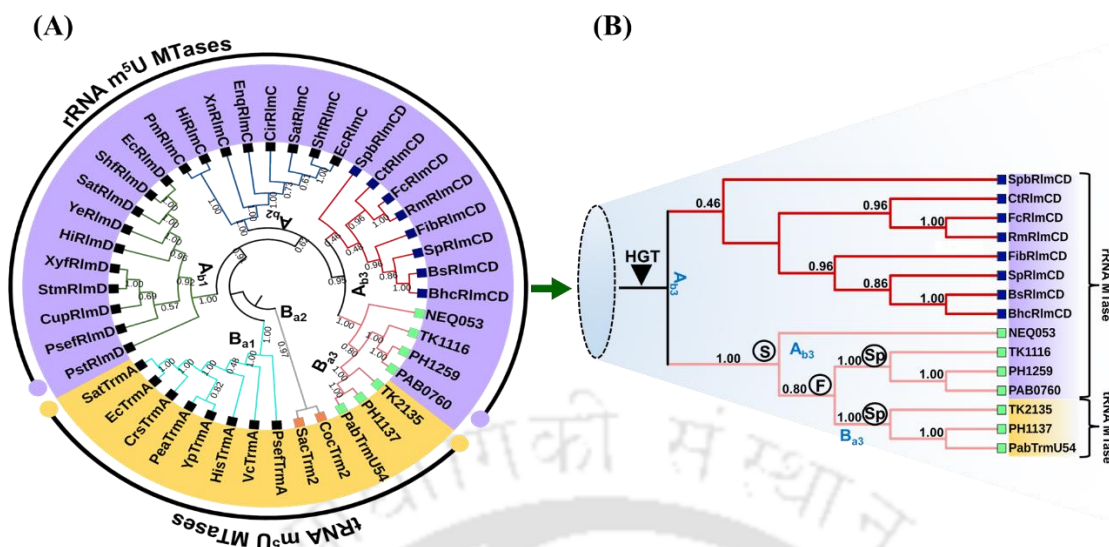


Figure 5.4. Genomic evidence suggests the acquisition of *rlmCD* genes in archaea from Gram-positive bacteria. (A) A phylogenetic tree was constructed using the maximum-likelihood method with 1000 replicates. A phylogenetic tree showing the evolutionary relationship was generated using 43 different RNA m⁵U MTases, including the protein PH1259 (*PhRlmCD*); details are provided in Table C.3. Further, RNA m⁵U MTases are divided into four different clusters in accordance with their existence in various life forms (Gram-negative bacteria, black; Gram-positive bacteria, navy blue; eukarya, salmon; and archaeal, green). (B) A probable scenario for the functional evolution of the archaeal genes coding for SAM-dependent Arm⁵U MTases was acquired from a Gram-positive bacterial donor via a single horizontal gene transfer (HGT). The black triangles indicate that the *rlmCD* gene was horizontally transferred while the encircled symbols ‘S’, ‘F’, and Sp’ are denoted for the gene speciation of the Thermococcales and Nanoarchaea lineages, the functional shift to form m⁵U tRNA MTases from the m⁵U rRNA MTases, and for the diversification of the Thermococcales lineage.

5.3.2. The protein *PhRlmCD* is conformationally and compositionally stable

For successful protein crystallization, it necessitates dual forms of stability: (i) conformational stability and (ii) compositional stability (Deller et al., 2016). To evaluate the stability of the purified protein *PhRlmCD*, both its structural integrity and composition were measured through circular dichroism (CD) experiments and dynamic

light scattering (DLS), respectively. The protein *PhRlmCD* was analyzed for conformational stability through far-UV CD spectra at 25 °C. The results showed that the protein is composed of ~25% α -helices, ~40% β -strands, and ~35% other secondary structures (Figure 5.5A). Subsequently, DLS measurements demonstrated a single peak with an average hydrodynamic diameter of ~11 nm, suggesting a monodispersed solution consisting of *PhRlmCD* particles of uniform size, shape, and mass (Figure 5.5B). Additionally, mass spectrometry analysis of *PhRlmCD* further affirmed its compositional stability, showing a single high-intensity peak corresponding to a molecular weight of ~50608 Da, closely matching the theoretical molecular weight of 50639 Da (Figure 5.5C).

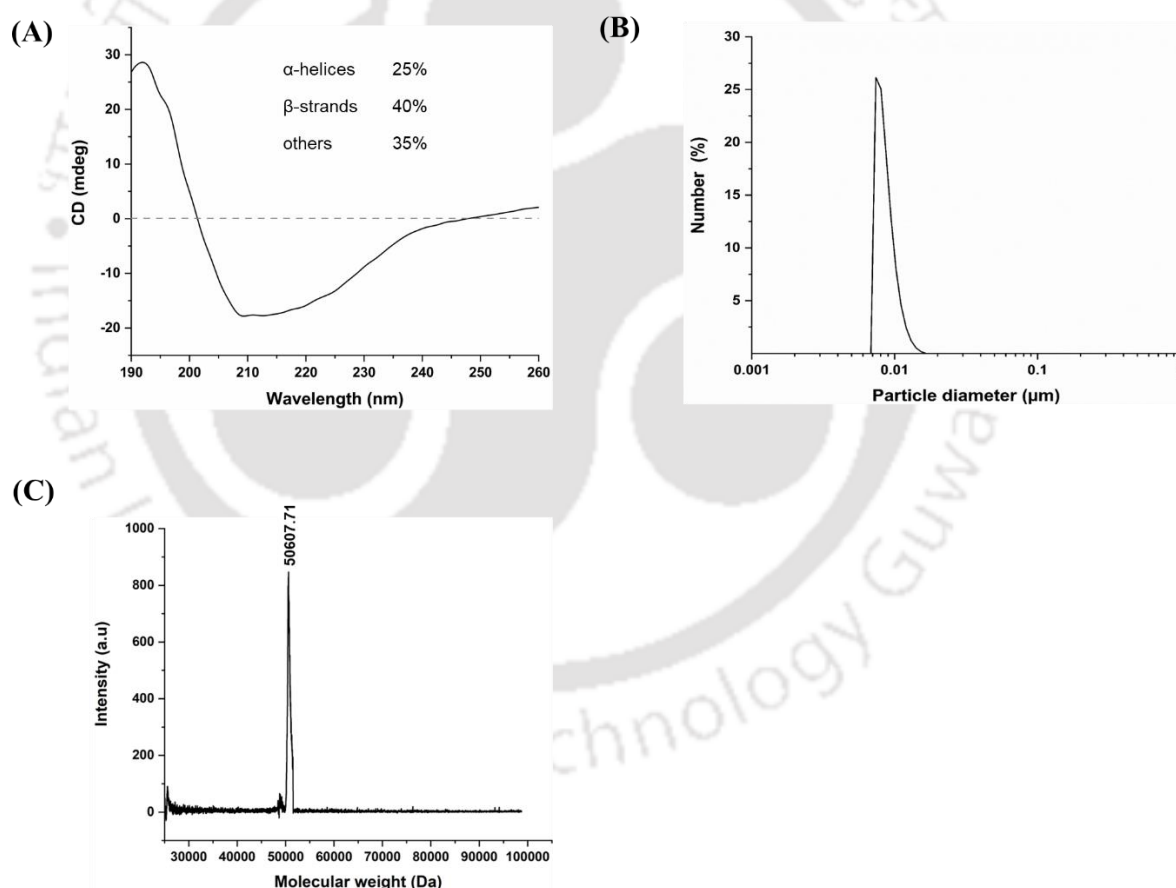


Figure 5.5. Conformation and compositional stability of the protein *PhRlmCD*. (A) A Far-UV CD spectrum ($\lambda = 190\text{-}260$ nm) of *PhRlmCD* provides its secondary structural content. (B) DLS measurements of *PhRlmCD* show that the solution is monodisperse, indicating that the protein exists in solution as a monomeric with an average hydrodynamic diameter of ~7.4 nm, free of aggregates. (C) Mass spectra of the

protein *PhRlmCD* using MALDI-TOF analysis suggest its molecular weight of ~50608 Da, equivalent to the theoretically calculated molecular weight (50639 Da).

5.3.3. The overall structure of the protein *PhRlmCD*

The crystal structure of protein *PhRlmCD* was determined at a resolution of 3.2 Å in the space group $P2_12_12_1$ with two protomers in each asymmetric unit (Table 5.1). The protomers in the asymmetric unit relate by a two-fold non-crystallographic symmetry (Figure C.1). A structural comparison of both the protomers in the asymmetric unit exhibits a complete similarity with a root mean square deviation (RMSD) of 0.22 Å. Similarly, structural comparison of *PhRlmCD* with its homologs, *SpRlmCD* and *EcRlmD* exhibits an RMSD of 1.7 and 3.0 Å, respectively.

Sequence analysis of *PhRlmCD* indicated the existence of a distinctive TRAM domain within the protein (Table C.2). The structural examination of *PhRlmCD* confirms the presence of the N-terminal TRAM domain, characterized by five β strands. These five β -strands are arranged in an antiparallel fashion ($\beta_1\uparrow\beta_2\downarrow\beta_3\uparrow\beta_4\downarrow\beta_5\uparrow$), forming a β -barrel (or Greek key) structural motif, also known as an oligonucleotide binding fold (OB-fold) (Figure 5.6A-5.6B). Despite a low sequence conservation in the TRAM domains, a structure-based multiple sequence alignment (MSA) of *PhRlmCD* with its homologs reveals a prevalent occurrence of equivalent residues in the N-terminal region of *PhRlmCD* (Figure 5.6C). Overall, a total of about 15-23% of the TRAM-domain residues are positively charged, hinting at their pivotal role in negatively-charged substrate binding.

In addition to the TRAM domain, the protein *PhRlmCD* comprises two other domains, viz., a central domain (residues 111-233) and a C-terminal catalytic domain (residues 251-412). The TRAM and central domains are linked by α/β (residues 54-110), and the central and catalytic domains are connected by β/β (residues 234-250) linkers. The central domain features a six-stranded β -sheet ($\beta_7\text{--}\beta_{12}$) juxtaposed with two α -helices (α_2 and α_3) (Figure 5.6A-5.6B). A previous study reported that rRNA m⁵U MTases comprise an iron-sulfur cluster pocket shaped by an extended loop ($\beta_5\text{--}\alpha_1$) comprising three cysteine residues and another loop ($\beta_9\text{--}\alpha_2$) containing the fourth cysteine (Lee et

al., 2004). However, despite possessing four conserved cysteine residues (Cys63, Cys69, Cys72, and Cys139), the central domain of the protein *PhRlmCD* does not harbor any Fe-S cluster. Instead, it forms a cystine by establishing disulfide bonds between Cys63-Cys69 and Cys72-Cys139 (Figure 5.6C).

On the other hand, the C-terminal catalytic domain features a seven-stranded β -sheet (β_{15} - β_{21}) arranged as $\beta_{15}\uparrow\beta_{16}\uparrow\beta_{17}\uparrow\beta_{18}\uparrow\beta_{19}\uparrow\beta_{20}\downarrow\beta_{21}$, along with five α -helices (α_4 - α_8) organized in an interleaved manner, resulting in a $\beta\alpha\beta\alpha\beta$ structure (Figure 5.6A-5.6B). This pertains to the conserved catalytic core domain that encompasses highly conserved catalytic motifs (X, I, II, III, IV, VI, and VIII) responsible for the formation of an active-site pocket (Figure 5.6C). The residues conserved within these catalytic motifs are hypothesized to play a crucial role in cofactor binding and are essential for catalytic activity (Bujnicki et al., 2004).



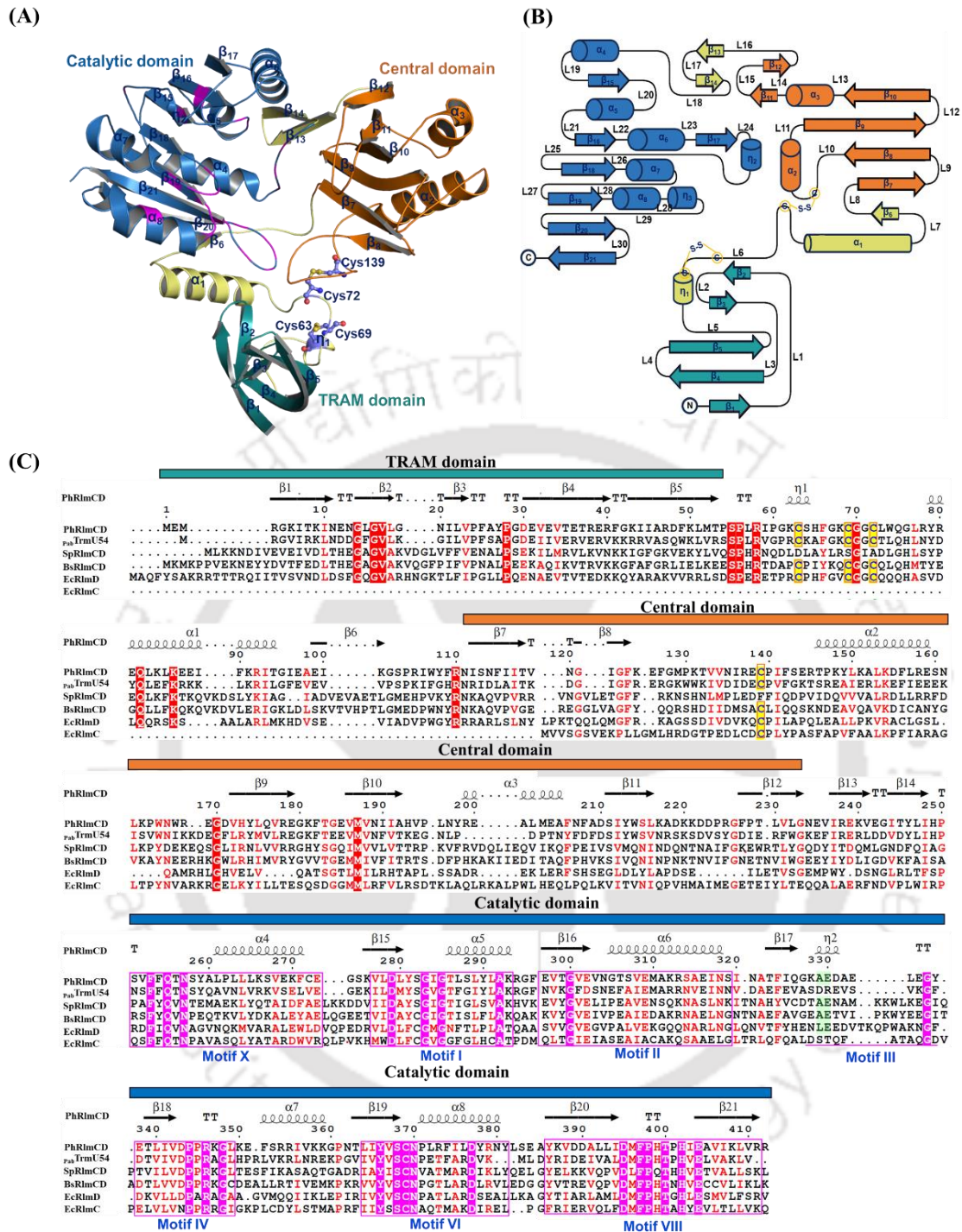


Figure 5.6. Overall three-dimensional structure of the protein *PhRlmCD* (PDB id: 8Z62). (A) The individual protomer of *PhRlmCD* comprises three distinct regions, represented in cartoon format. The N-terminal TRAM domain (teal), the central domain (orange), and the C-terminal catalytic domain (blue). These domains are interconnected by linker regions colored in pale yellow. The four conserved cysteines (purple stick) form disulfide bridges among them. (B) The planar model provides an overview of

protein topology, with secondary structural elements designated using classic numbering to provide insight into the architectural arrangement. (C) The structure-based MSA of *PhRlmCD* and its homologs show domain organization and conserved functional motifs. The proteins used in the study are *PhRlmCD* (*P. horikoshii*, O58994), *pabTrmU54* (*P. abyssi*, Q9UZR7), *SpRlmCD* (*S. pneumoniae*, Q97R12), *BsRlmCD* (*B. subtilis*, O31503), *EcRlmD* (*E. coli*, P55135), and *EcRlmC* (*E. coli*, P75817); organism name and their respective UniProt id of each protein is provided in the parenthesis. Secondary structure elements α -helices, β -strands, 3_{10} -helix, and β -turns are indicated by squiggles, arrows, η , and TT symbols. The long rectangular box outlines the boundary of each domain, which is highlighted with its corresponding color. The conserved residues in TRAM and the central domain are marked in red, while the four conserved cysteines are indicated in yellow. The sequences within the magenta box represent the seven conserved catalytic motifs present in the SAM-dependent RNA m⁵U MTases.

5.3.4. Despite the absence of an iron-sulfur cluster, *PhRlmCD* retains its native structure

The protein *PhRlmCD* possesses the central domain with four conserved cysteine residues (Cys63, Cys69, Cys72, and Cys139) arranged in the coordination motif C-X(5,7)-C-X(2)-C, rendering it an optimal candidate for the formation of an iron-sulfur cluster (Figure 5.6C). Notably, post-induction, the cell pellets containing the recombinant protein showed a distinct color change. The pellets transitioned from Off-white (uninduced) to dark yellow or light brown (induced) color at a protein concentration of ~200 μ M. The observed color alteration is a distinctive trait typically found in proteins containing iron-sulfur clusters (Figure 5.7A). Thus, the consecutive assembly of iron-sulfur clusters in *PhRlmCD* was investigated by scanning protein solutions in the range of 300-600 nm using UV-visible absorption spectroscopy. The result exhibited that the UV-visible spectrum of *PhRlmCD* contained a small shoulder at 347 nm and a clear absorption broad peak at 410 nm, indicating the presence of [4Fe-4S] or [3Fe-4S] clusters. However, both peaks were completely eliminated in the case of EDTA-treated protein *PhRlmCD* (*PhRlmCD*_EDTA) (Figure 5.7A).

To further affirm the presence of an iron-sulfur cluster in the protein *PhRlmCD*, an energy-dispersive X-ray (EDX) experiment was performed. The EDX results revealed that the protein solution is comprised of iron (Fe) atoms (~1.13% of the total weight) in addition to C, N, and O atoms (Figure 5.7B and Figure C.2A). Conversely, EDX microanalysis of *PhRlmCD*_EDTA indicates a negligible presence of Fe atoms (~0.07% of the total weight), suggesting the complete elimination of the iron-sulfur cluster (Figure 5.7C and Figure C.2A).

The previously published structure of RNA m⁵U MTase from *E. coli* (*EcRlmD*, PDB ID: 1UWV) hypothesized the necessity of the Fe-S cluster in proper protein folding. Interestingly, the study observed the production of insoluble proteins in the absence of the cluster or the conserved cysteine (Lee et al., 2004). The structural superposition of the crystal structure of *EcRlmD* with the AlphaFold model of its homologs *BsRlmCD*, *PabTrmU54*, and *PhRlmCD* revealed that all the homolog model structures maintain the coordination of the four conserved cysteines indicating the ability to form the Fe-S cluster (Figure 5.7D). However, the cluster was oxidized during the crystallization of *PabTrmU54* (PDB id: 2JJQ), which increased flexibility in a localized region within the central domain. The enhanced flexibility in the region spanning from residues 58 to 68, specifically in the β_5 - α_1 loop (L6), might have caused an unclear electron density (Walbott et al., 2008). As a result, two conserved residues, Cys61 and Cys67, dissociated from their coordination due to the absence of the cluster, leading to a disorder region (Figure C.2B). In this study, too, the electron density for the expected Fe-S cluster is missing in the structure of *PhRlmCD*. This hints that the iron-sulfur cluster, if at all present, in *PhRlmCD* underwent oxidation during the crystallization process, leading to the release of iron and subsequent decomposition of the cluster (Imlay, 2006), as observed in the protein *PabTrmU54*. However, notably, the protein *PhRlmCD* exhibited no disorder in the central domain primarily due to the strategic placement of two disulfide bridges. A clear electron density for the two disulfide bridges between the residues of Cys63 to Cys69 and Cys72 to Cys139 was observed (Figure C.3A-C.3B). Interestingly, the two disulfide bridges form different dihedral angles ($C_\beta-S_\gamma-S_\gamma-C_\beta$; χ_3) (Katz and Kossiakoff, 1986). In contrast, a structural comparison of crystal structure and the AlphaFold model of *PhRlmCD* revealed a difference in the state of the disulfide bonds. The AlphaFold model of *PhRlmCD*

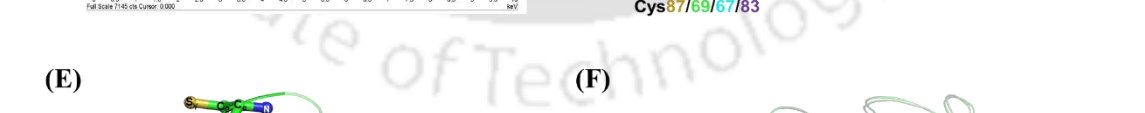


Figure 5.7. *PhRlmCD* attains stability in the absence of an iron-sulfur cluster through the organization of disulfide bridges. (A) Inset is an image of cell pellets after induction with (sample 1) or without (sample 2) IPTG. UV-visible scanning of purified *PhRlmCD* showing peaks at 347 and 410 nm (40 μ M, orange) indicates an iron-sulfur cluster, while these peaks were completely eliminated during scanning of purified protein sample *PhRlmCD*_EDTA (40 μ M, violet). Energy dispersive X-ray (EDX) graph for (B) *PhRlmCD* and (C) *PhRlmCD*_EDTA demonstrating the total elemental content of the protein solution. The Y-axis represents the arbitrary intensity of elemental content. (D) Overlaying the crystal structure of *EcRlmD* (PDB id: 1UWV) with the AlphaFold-derived structures of its homologs *BsRlmCD* (Violet), *PabTrmU54* (cyan) and *PhRlmCD* (green) indicate that the model structures have ideal coordination of conserved cysteine residue to form the iron-sulfur cluster. (E-F) A comparison of the AlphaFold model (in green) and the crystal structure of *PhRlmCD* (in grey) reveals that the experimental model shows a change in the coordination of conserved cysteine residues (shown in purple sticks). This change results in the formation of right- and left-handed ordered disulfide bridges between Cys63-Cys69 and Cys72-Cys139, respectively.

5.3.5. Central domain loops regulate the structural rearrangement of *PhRlmCD*

A structural comparison of *PabTrmU54*, *SpRlmCD*, and *PhRlmCD* showed a noticeable elongation of the loops L11 and L15 than those of *EcRlmD* (Figure 5.8A-5.8B). This is in support of the hypothesis that the Arm⁵U MTases are more similar to Gram-positive bacteria than to Gram-negative bacteria. In *PhRlmCD*, both the loops L11 (residues 159-173) and L15 (residues 217-227) are clearly resolved (Figure 5.8C). Although these loops are of similar lengths in *PabTrmU54*, *SpRlmCD* exhibits both loops of equal lengths (13 aa) (Walbott et al., 2008; Jiang et al., 2017).

In *EcRlmD*, two similar loops (3 and 6 residues, respectively) play a direct role in RNA recognition (Lee et al., 2005). The extended lengths of these loops in *PhRlmCD* hint at a broader range of substrate RNA selectivity, akin to the protein RlmCD in Gram-positive bacteria (Jiang et al., 2017). An analysis of the average temperature factor (B-factor) of the protein *PhRlmCD* revealed that the domains of this protein (TRAM 25 Å², central 25 Å², and catalytic 22 Å²) domains are comparatively stable except those

two loops L11 (α_2 - β_9 , 39 Å²) and L15 (β_{11} - β_{12} , 65 Å²) located in the central domain (Figure 5.8D). This phenomenon indicates that the flexibility in the loop regions might associated with the conformational changes during the RNA binding.

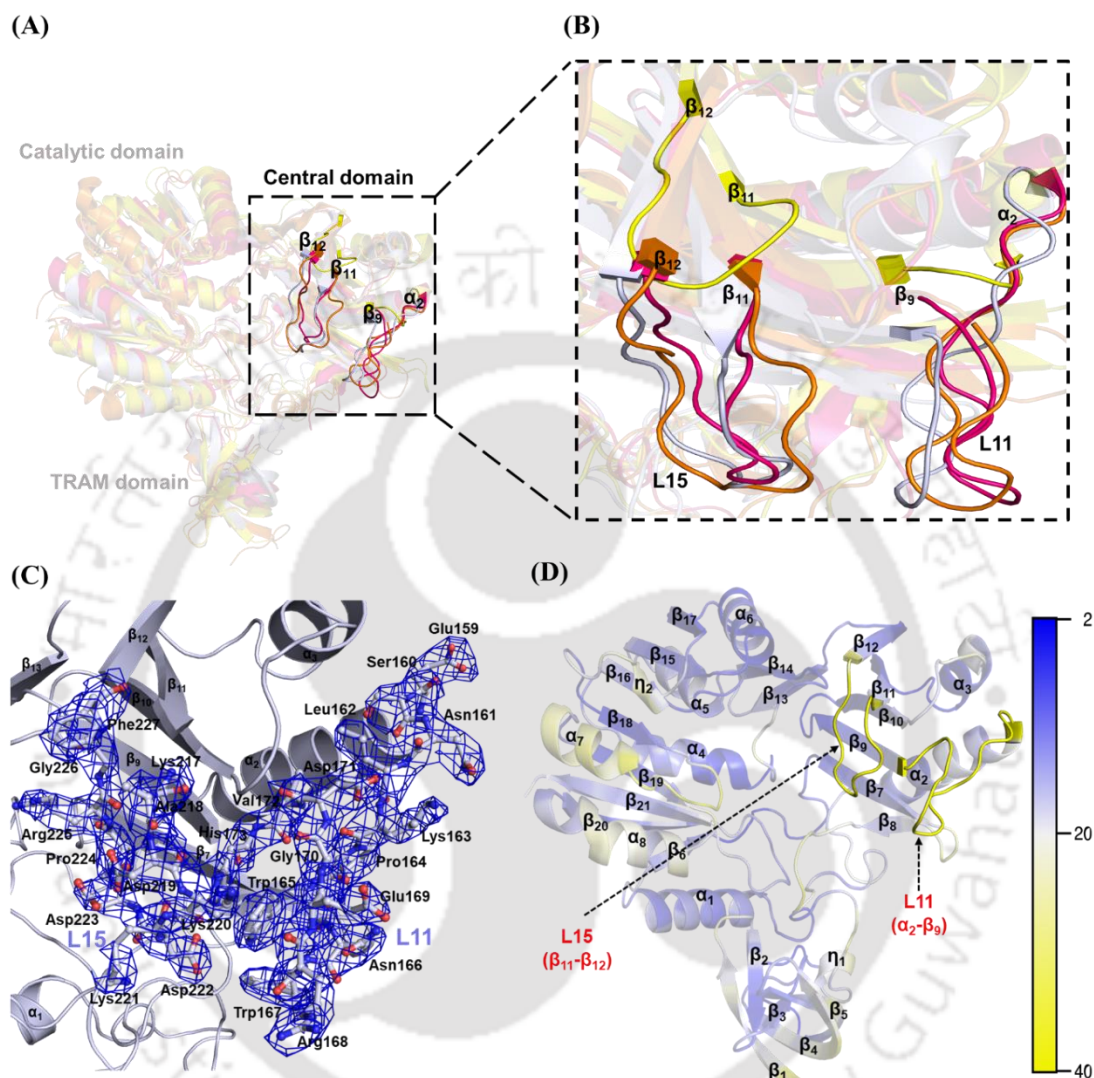


Figure 5.8. The loops 11 and 15 in *PhRlmCD* are critical for structural alteration.

(A-B) Comparative analysis of *PhRlmCD* (grey) with its closest homologs: *SpRlmCD* (PDB id: 5XJ1, orange), *PabTrmU54* (PDB id: 2JJQ, magenta), and *EcRlmD* (PDB id: 1UWV, yellow) reveals major structural differences lies in the central domain of these proteins. (C) Structural details of Loops 11 and 15 within the central domain of *PhRlmCD* are illustrated in stick representation, accompanied by their respective electron density map. The 2F_o-F_c map is contoured at 1 σ , providing insights into the structural integrity and precision of these loops. (D) The thermal parameter distribution (B-factor) for the protein *PhRlmCD*. A color gradient ranging from blue-white-yellow

represents the normalized differences in B-factors, displaying the flexibility in structural elements, with blue denoting lower values indicating stabilization while the transition to yellow indicates higher positive values indicating destabilization.

5.3.6. The *PhRlmCD* enzyme is a member of the classical Rossmann-fold MTase (RFM) family

Akin to other rRNA m⁵U MTases, the catalytic domain of *PhRlmCD* adopts the conserved class I SAM-dependent RFM fold (254-412). To substantiate the structural findings, the binding affinities of cognate cofactor SAM and its product SAH to the protein *PhRlmCD* were determined through isothermal titration calorimetry (ITC) experiments. Notably, the ITC experiments performed at room temperature (25 °C) could not produce any binding isotherm even after multiple repeats (Figure C.4A-C4B and Table C.1). Thus, to perform the ITC experiments at a physiological temperature (~90°C) of the hyperthermophilic archaeon *P. horikoshii* OT3, firstly its stability at such a high temperature was monitored using circular dichroism (CD). Thermal ramping experiments in the temperature range of 20°C to 120°C indicated that the protein *PhRlmCD* remains stable up to 87°C (Figure 5.9A and Figure C.4C). Further, overlapping the temperature profiles of *PhRlmCD* at 20°C and 60°C revealed that the secondary structure elements of the protein remain unchanged even at the higher temperature (60°C) (Figure C.4D). The ITC experiments performed at 60 °C exhibited a stronger binding affinity towards the molecule SAM compared to the product SAH and its analog SFG with K_D values of ~15, 100, and 222 μ M, respectively (Figure 5.9B, Figure C.4E-C.4F, and Table C.1). This suggests that, at the physiological conditions of the organism, the protein *PhRlmCD* shows structural rearrangement for the cofactor binding.

Although the cofactors such as SAM and SAH could not be co-crystallized with the protein *PhRlmCD*, their binding to the protein prompted the use of blind molecular docking to identify the binding site. The results of the molecular docking showed the binding of cofactors to the protein *PhRlmCD* in a similar pose to that of other RNA m⁵U MTases, such as *SpRlmCD* (PDB Id: 5ZTH), *PabTrmU54* (PDB Id: 2JJQ), and *EcRlmD* (PDB Id: 2BH2) (Figure 5.9C and Table C.4). The molecule SAM exhibits an estimated binding free energy of -7.15 kcal mol⁻¹, corroborating the experimental ΔG value obtained through ITC (Table C.1).

The exocyclic amino group and N1 atom of the adenine moiety of SAM establish hydrogen bonds with the carboxylate group and the backbone amide of Glu330 and Ala329, respectively. The ribose sugar ring of SAM engages in stacking interactions with the active site residue Pro345. The two hydroxyl groups of SAM establish hydrogen bonds with the carboxyl group of Glu302, as observed in other SAM-dependent MTases (Alian et al., 2008). Further, the terminal carboxylate of the methionine moiety forms hydrogen bonds with the hydroxyl group of Ile284 and the amide side chain of Gln255 (Figure 5.9D and Table C.4).

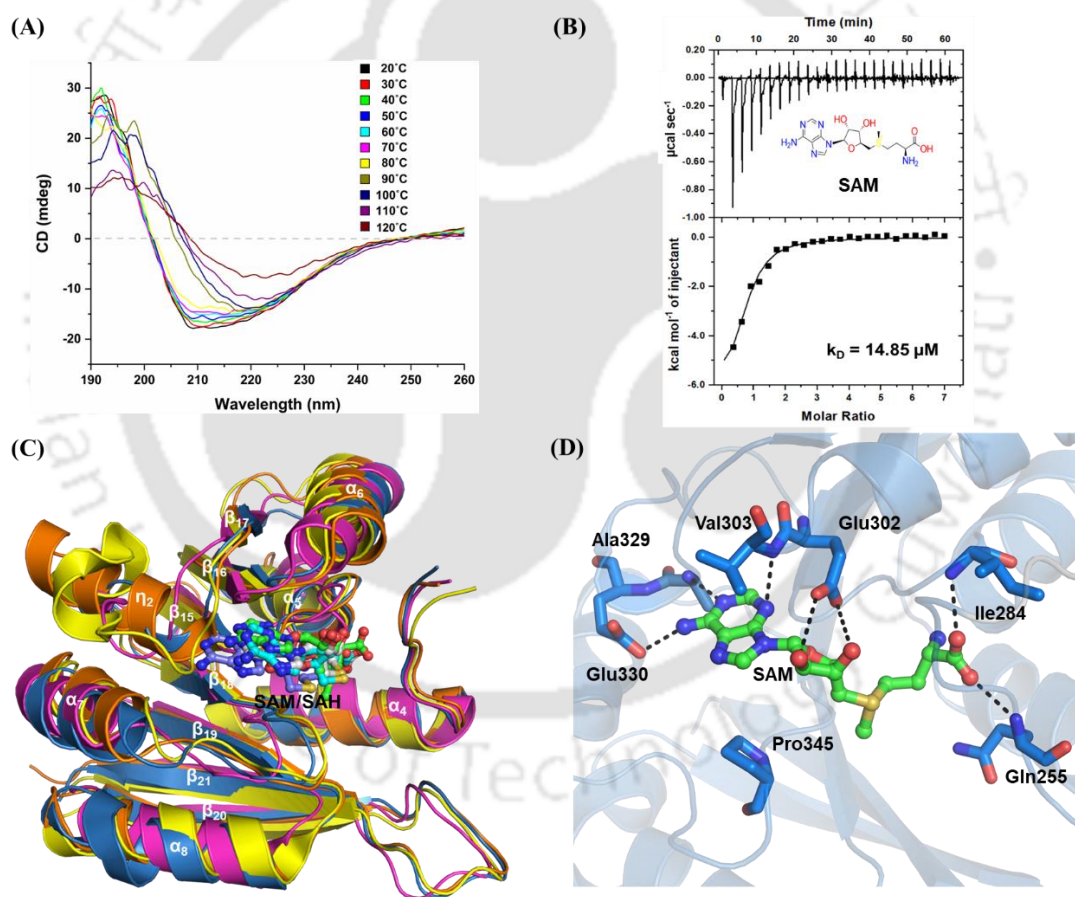


Figure 5.9. *PhRlmCD* belongs to Class I SAM-dependent MTase. (A) A thermal melting experiment of *PhRlmCD* revealed that changes in secondary structure with increasing temperature begin at the melting point (T_m) of protein at 80°C. The initial spectra (at 20°C) are illustrated by a black line, then followed by spectra at other

temperatures in different colors (B) Plot showing the binding of SAM to the protein *PhRlmCD* obtained through isothermal titration calorimetry (ITC) data. The upper panels show the raw heat change upon ligand titration, while the integrated heat pulse subtracted with their heat of dilution is shown in the lower panels. (C) Superposition among the catalytic domain of *PhRlmCD* with its closest homologs: *SpRlmCD* (orange), *PabTrmU54* (magenta), and *EcRlmD* (yellow), revealing the characteristic class I SAM-dependent MTase fold. SAH molecules bound to *SpRlmCD*, *PabTrmU54*, and *EcRlmD* are highlighted in grey, purple, and cyan sticks, respectively, while SAM is represented as green sticks. (D) Representation of the SAM molecule docked to *PhRlmCD*, with interactive active site residues depicted in blue lines.

5.3.7. The TRAM and the central domains of *PhRlmCD* regulate the substrate selectivity

In *PabTrmU54* (PDB id: 2JJQ), the negatively charged tRNA is accommodated at the interface of three domains. However, only the T ψ C stem-loop and the aminoacyl stem of the tRNA are used to recognize the substrate (Urbonavičius et al., 2008; Walbott et al., 2008). To investigate the binding of the substrate to the out-paralog protein *PhRlmCD*, a tertiary model of a truncated (omitting both the anticodon and D stem-loops) tRNA^{Asp} from *P. horikoshii* was generated (Figure C.5A-C.5B).

To validate the generated model further, molecular docking of the truncated tRNA with *PabTrmU54* was performed as a positive control. The molecular docking calculations showed the binding of the truncated tRNA to a positively charged narrow groove formed at the interface of the domains (Figure C.5C). The residues Asp11, Lys39, Arg41, Ser216, Ser218, Ser300, Asn301, Glu302, and Arg343 are observed to interact with C2, G71, G1, G51, G64, C61, C62, U54, and C4 nucleotides of the tRNA molecule. Further, the residues, Arg217 and Arg343, form a salt bridge with the tRNA nucleotides G52 and C4 (Figure C.5D and Table C.5). A structural superimposition of *PabTrmU54* and *PhRlmCD* exhibits the conservation of similar residues, Asp219, Lys220, Lys221, Asn304, and Arg346 in the latter protein (Figure C.5E). Although these structural similarities suggest the binding of tRNA to the protein *PhRlmCD*, a hinge loop (L10, between α_2 and β_8 , residues 120-141) of *PhRlmCD* hinders the binding

of the aminoacyl stem of the substrate (Figure 5.10A). In particular, the residues Phe127 and Pro130 within L10 obstruct the stacking interaction with the purine bases, G65 and G1, of tRNA. In contrast, the corresponding residues Lys127, Trp128, Trp129, and Lys130 in *Pab*TrmU54 are strategically positioned away from the epicenter to avoid steric hindrance with the truncated tRNA (Figure 5.10B).

The structural comparison of the apo form of *Ph*RlmCD with those of rRNA-specific m⁵U MTases, *Ec*RlmD (PDB id: 1UWV) and *Sp*RlmCD (PDB id: 5XJ1), affirms the unique hinge conformation of *Ph*RlmCD, involved in blocking substrate RNA (tRNA/rRNA). Further, in both apo and holo forms, the hinge region adapts a similar conformation (Figure C.6A-C.6B). Such a conformation of this hinge is crucial for substrate selectivity, that is formed due to a unique left-handed disulfide bridge between Cys72 and Cys139 in *Ph*RlmCD, not observed in other structures including its AlphaFold model (Figure 5.10B and Figure C.6A). Notably, the AlphaFold model of *Ph*RlmCD (AlphaFold Id: AF_AFO58994F1) represents a conformation similar to that of homologs and differs significantly from its crystal structure (Figure C.6A). As the crystal structure of *Ph*RlmCD represents a closed form of the protein, its open state was modeled taking its AlphaFold model as a template (Figure C.6C).

The open state of the protein *Ph*RlmCD was utilized for the molecular docking calculations of substrates. The electrostatic potential charge distribution analysis of *Ph*RlmCD revealed a distinction from *Pab*TrmU54. This is reflected in the binding of the truncated tRNA at a site other than the positively charged cavity at the interface of domains (Figure 5.10C). A comparison of docked *Ph*RlmCD with the truncated tRNA and the crystal structure of *Pab*TrmU54 revealed that the residues Arg40 and Phe41 located in the loop L4 (β 4- β 5) of the TRAM domain cause steric hindrance with the nucleotide residues C3 and C4 of the substrate (Figure 5.10D). This is possibly due to the flexibility observed in the TRAM domain of *Ph*RlmCD, akin to its homologs *Ec*RlmD and *Pab*TrmU54. The TRAM domain in *Ph*RlmCD shows an angular movement of 22° and 5°, respectively, with that of *Ec*RlmD and *Pab*TrmU54 (Figure C.7A). In *E. coli* RlmD, the mini-tRNA exhibits no interaction with the TRAM domain, while it fits in the binding site without any steric hindrance with the TRAM domain of *Pab*TrmU54 (Figure C.7B). In summary, the TRAM domain seems to play a crucial role in the substrate binding.

Besides the TRAM domain, the central domain of *PhRlmCD* exhibited a steric hindrance with the truncated tRNA. The residue Lys221 located in loop L15 of the central domain, restricts the entry of nucleotide C63 situated in the T ψ C stem loop of the substrate. In contrast, a smaller residue Ser218 at that position in *PabTrmU54* causes no hindrance, allowing the substrate to fit into the binding site (Figure 5.10D). This highlights an adaptive structural design of *PabTrmU54* in accommodating the substrate, in contrast to *PhRlmCD*.

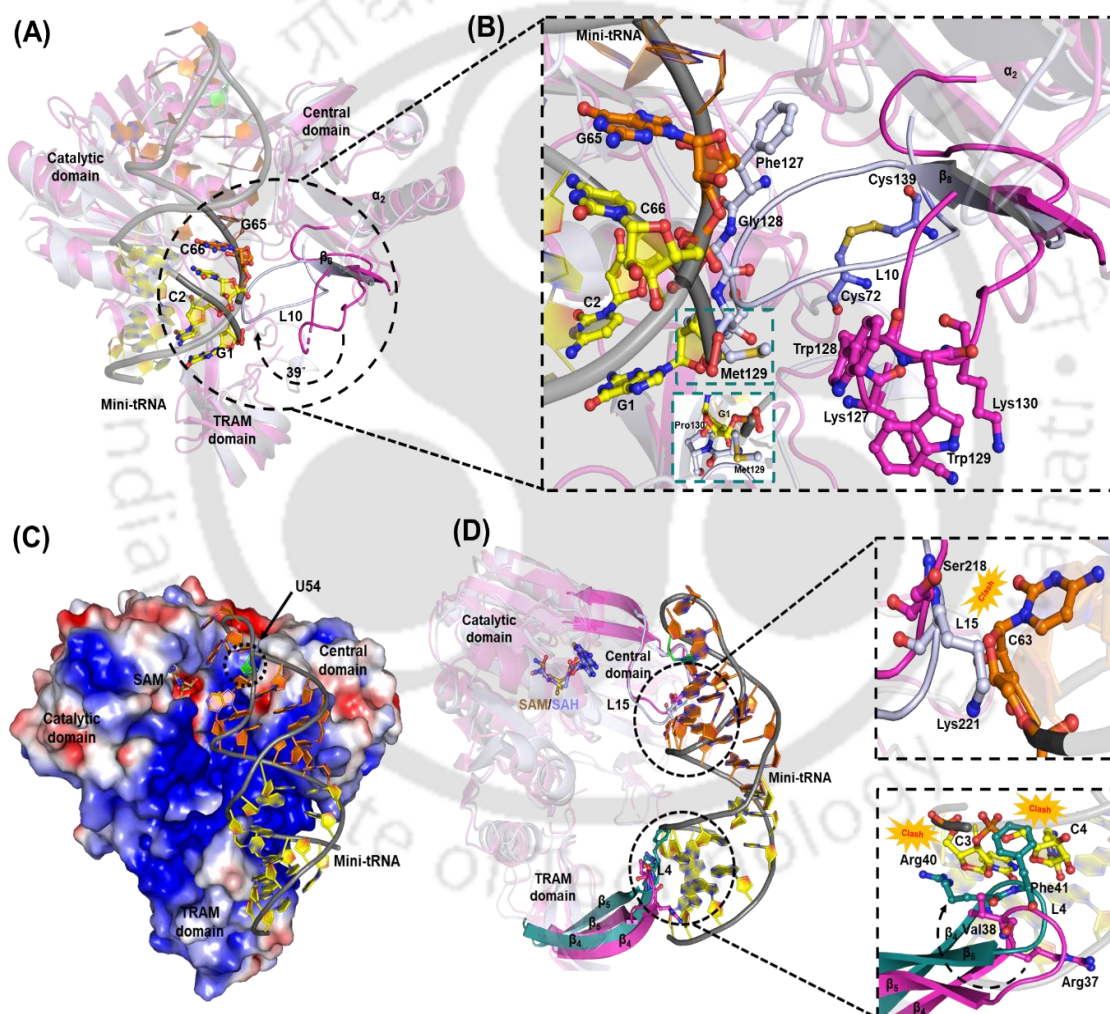


Figure 5.10. The distinctive traits of the apo form of *PhRlmCD*. (A) Overlay of mini-tRNA bound *PabTrmU54* (magenta) onto the closed state (grey) of *PhRlmCD*, revealing dynamic hinge movements that obstruct the substrate binding pocket. (B) A detailed view highlights steric clashes originating from the dynamic hinge region of

PhRlmCD with the mini-tRNA molecule. (C) Representation of electrostatic charge distribution on *PhRlmCD* docked with mini-tRNA, depicting positively-charged surfaces in blue (+3 kT/e) and negatively-charged surfaces in red (-3 kT/e). (D) Comparison of mini-tRNA bound open state of *PhRlmCD* with *pabTrmU54*, indicating steric clashes between the central and TRAM domains of *PhRlmCD* with T ψ C (orange) and aminoacyl arm (yellow) of tRNA. The dashed curved arrow illustrates the varying degrees of rotation between the TRAM domain of *PhRlmCD* and *pabTrmU54*.

5.3.8. *PhRlmCD* may have dual substrate specificity

In Gram-positive bacteria, the m⁵U MTases, viz. *RlmCD*, a homolog of *EcRlmC* and *EcRlmD*, methylates at the 5th position of uracil within both domain II (U747) and domain IV (U1939) of the hairpin 35 motif in 23S rRNA (Shoji et al., 2015). In archaea, particularly in *P. horikoshii*, the corresponding sites of modification (m⁵U) are domain II (U875) and domain IV (U2055) within the hairpin 35 motif of 23S rRNA. For further analysis the corresponding regions of mini-rRNA (U875_F and U2055_F; UX_F, X= the corresponding position, _F stands for fragment rRNA) were taken into consideration (Figure C.8A-C.8B). While the previous studies have indicated that the recognition of mini-rRNA by m⁵U MTases is not only dependent on their sequential composition but also on unique conformation. Therefore, mini-rRNAs from *S. pneumoniae* that were previously crystallized were used as templates (PDB id: 5ZQ8, 5ZTH; Jiang et al., 2017 and 2018). The predicted loops of mini-rRNA (U875_F and U2055_F) from *P. horikoshii* were modeled while maintaining a 'U' shaped geometry and internal base stacking (Figure C.8C-C.8D).

To study the dual substrate specificity of rRNA-specific Arm⁵U MTases, the open state of the protein *PhRlmCD* was docked with mini-rRNA fragments to form two different complexes (*PhRlmCD*•SAM•U875_F/U2055_F). The molecular docking results revealed that both the mini-rRNA fragments occupy the narrow groove at the interface of the domains. In the complex forms, the nucleotides U875 and U2055 are flipped out toward the negatively charged pocket of the cofactor and face the methyl group of SAM. (Figure C.9A-C.9B). The residues Gln255, Arg346 (motif IV), and Cys369 (motif VI) coordinate the interaction of the substrate through U875. Additionally, the residue

Phe253 (motif X) forms a stacking interaction with the nucleotide U875 (Figure 5.6C, Figure 5.11A, and Table C.5). A similar set of interactions was observed in the case of U2055, in addition to the residue Asp343 (motif IV) that helps in holding the substrate in the binding site (Figure 5.6C, Figure 5.11B, and Table C.5).

The methylating sites (U875 and U2055) of the rRNA fragments predict a similar set of interactions, while the neighboring nucleotides suggest different interactions. The analysis of these interactions revealed that the protein *PhRlmCD* is associated with two partially joined stacking networks with fragment U875_F (Arg1110-G876-Phe124-C877-A878-G874-U873-A872 and Arg110-G876-Phe124-C877-Lys125-Trp167-G879-U880 and U881) which formed as a “λ”-shape stacked geometry between the RNA-protein complex (Figure 5.11C). However, the protein *PhRlmCD* was unable to form a base stacking network with the other fragment U2055_F. This was possibly due to the orientation of two nucleotides, A2056 and A2057, located away from the residue Phe124, restricting the formation of π-π stacking interactions (Figure 5.11D). Besides aromatic stacking, the interaction of the two mini-rRNA fragments with the protein *PhRlmCD* is facilitated through the formation of multiple hydrogen bonds. Specifically, within the U875_F fragment, the nucleotides G876 and C877 form hydrogen bonds with the residues Asn114 and Lys125, respectively. Likewise, in the 2055_F fragment, the nucleotides U2056 and C2057 at similar positions engage in hydrogen bonding with the residues Asn114 and Lys125, respectively (Figure C9C-C9D). Notably, the U875_F and 2055_F fragments exhibit distinct interactions with the active-site residues of *PhRlmCD*. In U875_F, hydrogen bonds and salt bridges are formed between nucleotides U880, U873, and G874 and active site residues Lys131, Lys221, and Asp222, respectively. Meanwhile, U2055_F involves nucleotides A2054 and A2052 interacting with active site residues Lys347 and Arg373. (Figure C.9C-C.9D and Table C.5).

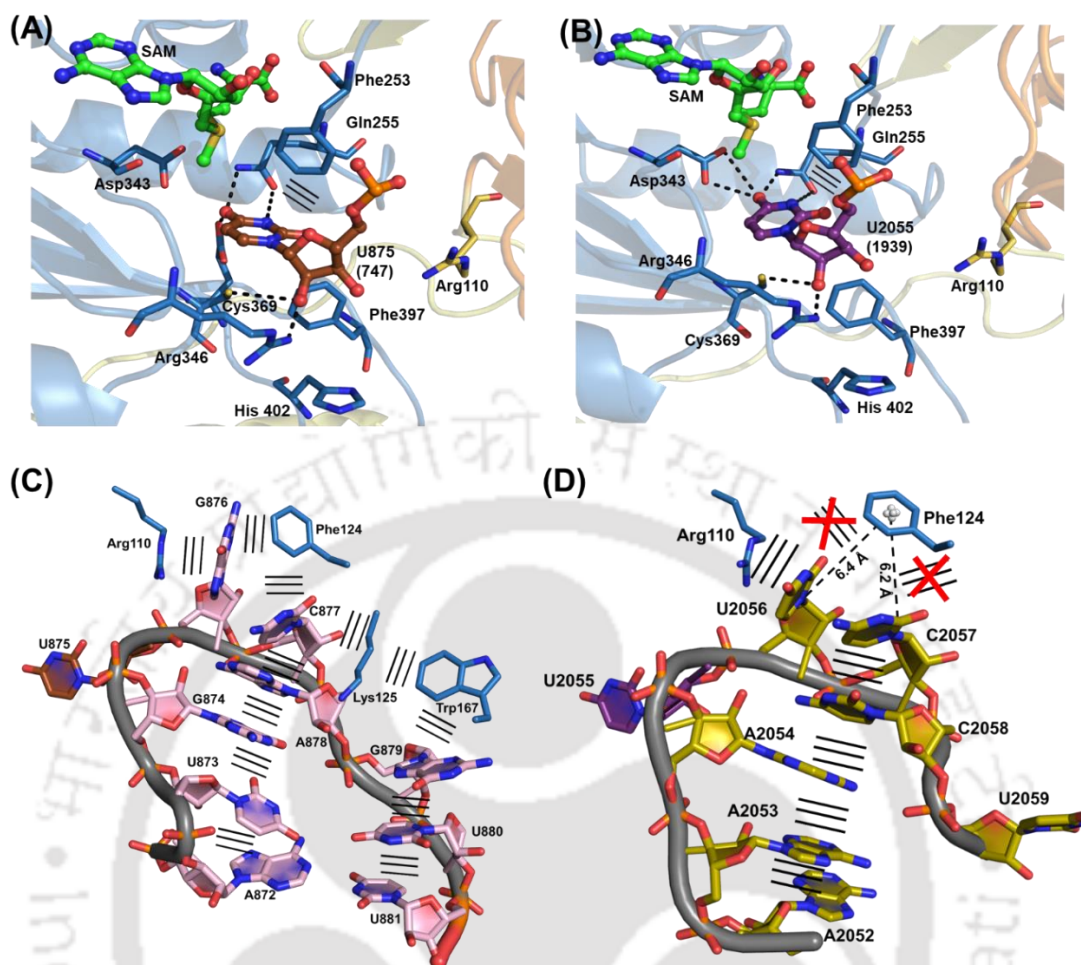


Figure 5.11. A docking simulation suggests that *PhRlmCD* may recognize both mini-rRNA fragments, U875_F and U2055_F, derived from domains II and IV of the 23S rRNA of *P. horikoshii*. (A-B) Depiction of the interaction between the target nucleotide of U875 or U2055 and the adjacent residues of *PhRlmCD*. Hydrogen bonding interactions are illustrated as black dashed lines, while the three parallel lines denote face-to-edge stacking. (C-D) Illustration of aromatic stacking interactions in *PhRlmCD* (marine) with U875_F (pink) and *PhRlmCD* (marine) with U2055_F (olive). The target nucleotides in the fragments of U875 and U2055 are highlighted as brown and violet sticks, respectively.

5.4. Discussion

SAM-dependent methyltransferases (MTases) are divided into five distinct classes (classes I-V) based on their structural and biochemical properties. The Class I MTases

are the most abundant group identified so far, which includes m⁵U MTases, characterized by a Rossmann fold motif (Schubert et al., 2003). This class I SAM-dependent m⁵U MTases are ubiquitously found in prokaryotes and a few eukaryotes. Their role in modifying a diverse array of non-coding RNA, such as rRNA and tRNA, has well been studied (Nordlund et al., 2000; Lee et al., 2004; Jiang et al., 2017; Keffer-Wilkes et al., 2020). In archaea, two distinct RNA-specific Arm⁵U MTases have been identified, of which a tRNA-specific Arm⁵U MTase from *P. abyssi* (PabTrmU54) has been characterized (Walbott et al., 2008). In this study, the rRNA-specific Arm⁵U MTase from *P. horikoshii* (PhRlmCD) was structurally characterized.

Previous research suggests that RNA m⁵U MTases trace their origins back to a common ancestor resembling the ancient protein RlmD (or RumA) found in Gram-negative bacteria MTase (Urbonavičius et al., 2008; Walbott et al., 2008). In these bacteria, this ancestral protein underwent duplication, resulting in multiple copies with varying substrate specificities. While the original RlmD enzyme is dedicated to methylating U1939 of 23S rRNA, its paralogous counterparts, RlmC and TrmA, target different RNA sites: U747 in 23S rRNA and U54 in tRNA, respectively. Additionally, it is proposed that in certain archaea, such as Thermococcales and *N. equitans*, the gene responsible for RNA m⁵U methyltransferases was acquired directly from Gram-negative organisms through a single horizontal gene transfer (HGT) event (Urbonavičius et al., 2008).

In this study, archaeal RNA m⁵U MTases likely originated during the last universal ancestor (LUA), bridging the gap between Gram-negative and Gram-positive bacteria. The ancestral gene was transferred to Gram-negative bacteria, which was further duplicated as two SAM-dependent rRNA (RlmD, Ab₂ and RlmC, Ab₁) and a tRNA (TrmA, Ba₁) m⁵U MTases. The protein RlmC differs significantly from RlmD in its rRNA-recognition mechanisms by losing the rRNA-capturing domain TRAM (Lee et al., 2004, 2005). In contrast, in Gram-positive bacteria, a single gene responsible for SAM-dependent rRNA m⁵U methyltransferase (RlmCD, Ab₃), capable of efficiently catalyzing both sites on 23S rRNA, functioning equivalently to both RlmC and RlmD (Jian et al., 2017; Sergiev et al., 2018).

In contrast, in Gram-positive bacteria, N⁵,N¹⁰-methylenetetrahydrofolate (THF)-dependent mechanisms are employed to modify the conserved site at m⁵U54 in the T Ψ C loop of tRNA (Urbonavičius et al., 2005). The divergence in the cofactor utilization between SAM and THF-dependent methyltransferases (MTases) initiated from the most primitive bacterial class, i.e., the single-layer Gram-positive bacteria (Gupta and Griffith, 2002; Koch, 2003). The results of this study indicated that the RlmCD-type proteins were transferred from the primitive Gram-positive bacteria to archaea. Over the evolution, the ancestral bacterial gene transferred to archaea got duplicated into tRNA- and rRNA-specific m⁵U MTase (Urbonavičius et al., 2008; Walbott et al., 2008; Auxilien et al., 2011).

The structural elucidation of the protein *Ph*RlmCD corroborated the hypothesis of the evolutionary analysis. This is reflected by the presence of longer loops L11 and L15 of its central domain as observed in other Arm⁵U MTases (e.g., *Pab*TrmU54 and *Ph*RlmCD). Notably, these longer loops present in RlmCD are distinctly different from RlmD. The extension of these loops helps them in substrate selectivity (Jiang et al., 2017). It has previously been demonstrated that the longer loops of the *Pab*TrmU54 (an Arm⁵U MTase) impede the binding of the 5' end of the mini-rRNA substrate of RlmD (Walbott et al., 2008). Thus, it is tempting to propose that substrate selectivity is achieved through the moderation of the loop size. The shorter loop is the signature of an RlmD-like protein, while a longer loop is of the RlmC-like protein. Based on these results, it can be hypothesized that the RlmCD-like proteins alter the length of the central-domain loops to accommodate the substrate in the active site pocket depending on the modification site, U747 (domain II) or U1939 (domain IV) of 23S rRNA.

Interestingly, both the archaeal paralogs rRNA- and tRNA-specific RNA m⁵U MTases contain such a loop of a similar length. This implies that other determinants of these proteins help distinguish the substrates. The results of this study indicate that substrate selectivity is governed by the unique composition of the loop L15. It was observed that the positively-charged residue Lys221 in the archaeal rRNA-specific enzyme *Ph*RlmCD causes a steric clash with the tRNA molecule in the binding site. Meanwhile in the tRNA-specific enzyme *Pab*TrmU54, a smaller polar residue Ser218 helps fit the

substrate in the binding site. This suggests that the sequence composition, in addition to the structural conformation, helps in the selectivity of the substrates of these enzymes. Besides the loops in the central domain, substrate selectivity is determined by the domain TRAM (Lee et al., 2005). The results of this study suggest that the rotational variations among the TRAM domains of the RNA m⁵U MTases may serve as a critical factor in substrate selectivity. In *PabTrmU54*, the positioning of the TRAM domain acts as a molecular ruler, aiding in the specific placement of the tRNA molecule during modification in loop T ψ C by restricting the mobility of the aminoacyl acceptor stem. A deviation in the TRAM orientation seems to cause a steric clash in the substrate binding (Walbott et al., 2008). In *PhRlmCD*, the TRAM domain rotated by 5° compared to that of the *PabTrmU54*. This alternation in the TRAM domain of *PhRlmCD* instigates loop L4 mobility causing a steric hindrance to the substrate tRNA.

The structural elucidation of the rRNA-specific Arm⁵U MTase (i.e, *PhRlmCD*) suggests that it may modify two specific sites, U875 (domain II) and U2055 (domain IV) of 23S rRNA, exhibiting a dual methyltransferase activity. Further, a meticulous analysis of the protein-rRNA docked complex of two mini-rRNA fragments (U875_F and U2055_F) bound to the protein *PhRlmCD* revealed that the conserved active-site residues Phe253, Gln255, Asp343, Arg346, and Cys369 directly interact with the target nucleotides U875 and U2055. This hints at its role as dual substrate binding specificity.

Based on the structural elucidation of this study, it is tempting to propose that the protein *PhRlmCD* employs distinct mechanisms to recognize dual substrates. In the case of the fragment U875_F, aromatic stacking and hydrogen bonding with the catalytic residues of *PhRlmCD*, akin to *SpRlmCD* (Jiang et al., 2018), help recognize the substrate. Although both the proteins *SpRlmCD* and *PhRlmCD* form a λ -shaped stacking geometry network with the domain II fragments (U747_F and U875_F), their stacking interactions differ. In *SpRlmCD*, the nucleotides A749 and A751 form π - π stacking interactions with the residues Phe146 and His151, beside Phe146 and His151 intramolecular π - π stacking interaction (Jiang et al., 2018). In contrast, in *PhRlmCD*, the corresponding residues Lys125 (Phe146) and Trp167 (His141) establish a λ -shaped stacking network through π -cation (C877-Lys125 and Lys125-Trp167) and π - π stacking (Trp167-U880) interactions.

Further, in *SpRlmCD*, the mini-rRNA fragment from domain IV of 23S rRNA (U1939_F) establishes a solitary stacking network with the protein. Meanwhile in *PhRlmCD*, the corresponding U2055_F fragment lacks such a stacking network with the protein. This, in turn, unables the nucleotides U2056 and C2057 to stack with the catalytic residue Phe124 of *PhRlmCD*. This indicates that despite the substrate sequence conservation in domain IV of two proteins, *PhRlmCD* and *SpRlmCD*, the substrates are distinguished by the distinction between the active-site residues.

An iron-sulfur cluster, coordinated by four conserved cysteine residues, has been identified in RlmD and other RNA methyltransferases, playing a crucial role in enhancing their stability (Lee et al., 2004, 2005; Walbott et al., 2008). In *PhRlmCD*, the presence of four conserved cysteine residues and results of the various biophysical studies, hinted at the binding of an iron-sulfur cluster in the purified protein *PhRlmCD*. However, such an Fe-S cluster could not be observed in the crystal structure of *PhRlmCD*. It has previously been reported that the Fe-S cluster is often oxidized during the crystallization process of the protein (Imlay, 2006). Although the protein *PhRlmCD* did not contain the Fe-S cluster in its crystallized form, it maintained its structural integrity by forming two distinct disulfide bridges between Cys63 with Cys69 (right-handed) and Cys72 with Cys139 (left-handed). The presence of a right-handed disulfide bridge helps the protein to adapt a lasso conformation. In contrast, the left-handed bridge induces a distinctive hinge movement within the α_2 - β_8 region of *PhRlmCD*. This, in turn, obstructs the substrate binding pocket. To the best of our knowledge, the crystal structure of *PhRlmCD* reported in this study represents a closed form of an RNA m⁵U MTase not reported previously.

Furthermore, the structural differences between the crystal structure and the AlphaFold model of *PhRlmCD* suggest that the presence of a left-handed disulfide bond induces the change in the coordination of conserved four cysteine residues. This is in corroboration with the crystal structure of *PabTrmU54* in form II (PDB id: 2VS1), where a single disulfide bridge is formed between the Cys67 and Cys70, which dissociate the coordination among the four conserved cysteines required to maintain the Fe-S cluster. Subsequently, only one disulfide bond is formed instead of two, inducing high

flexibility in the dynamic hinge loop (L10) and potentially causing unclear electron density.

Nevertheless, Fe-S clusters act as cofactors that can alter the structure and function of proteins. In addition to contributing to structural and catalytic processes, they also serve as sensors of iron and oxygen levels, linking the intracellular environment with proteins involved in metabolic pathways. Due to their labile nature and sensitivity to oxygen, these cofactors are prone to decomposition (Imlay, 2006). Several proteins rely on Fe-S clusters, including helicase-nuclease DNA2, glycosylases, helicases, and base excision repair enzymes, which have been found to bind to Fe-S cluster cofactors likely due to the conditions on early Earth, where abundant iron and sulfur-made it easy to incorporate Fe-S cluster cofactors into enzymes (Fuss et al., 2015; Prakash and Doublé, 2015; Puig et al., 2017; Baranovskiy et al., 2018). However, increasing oxygen levels over time led to more iron oxidizing into less usable forms, exacerbating the instability and oxygen sensitivity of Fe-S clusters. This suggests that Fe-S cluster cofactors were positively selected through evolution due to their crucial and irreplaceable role, which other, more stable cofactors cannot fulfil (Rouault, 2015). Thus, ease of oxidation of the cluster in *PhRlmCD* is a way to abolish its methyltransferase activity in the cell, and it might have certain advantages for ribosomal functions under oxidative stress conditions. A similar phenomenon was observed when oxidation of *EcRlmD* led to cluster decomposition and subsequent precipitation of the protein (Agarwalla et al., 2004). These observations suggest that the iron-sulfur cluster is dispensable and the substrate binding mechanism can be altered by such molecular switching.

5.5. Conclusion and future perspectives

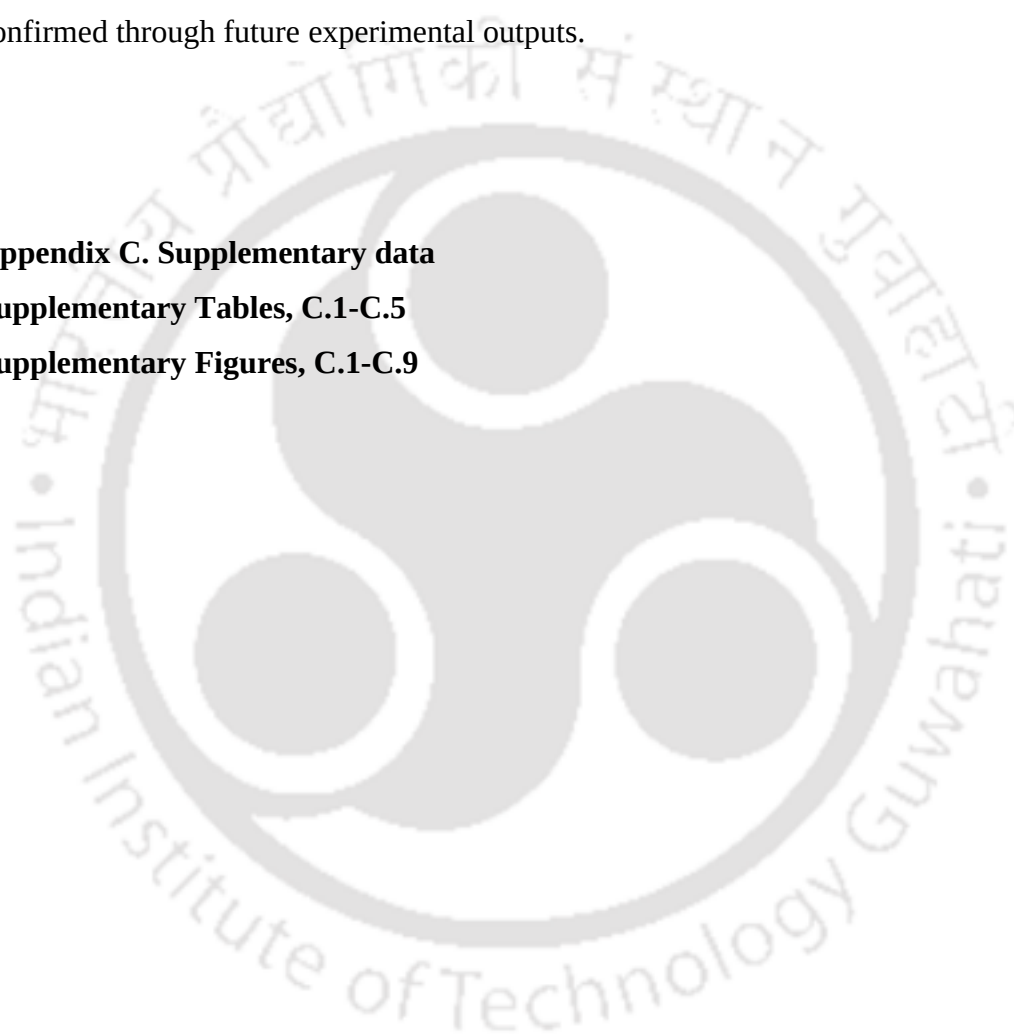
In summary, our evolutionary insights disrupt conventional beliefs regarding the presence of m^5U MTase genes in the archaeal genome. This assertion gains support from both the structural attributes and substrate binding proficiency of *PhRlmCD*, which resemble Gram-positive RlmCD-type proteins rather than the canonical ancestors RlmD or RumA. This paradigm shift necessitates a revaluation of the prior proposition concerning the existence of the genes in the archaea. For instance, the study delves into the unique segmented hinge arrangement in *PhRlmCD* and how the respective segment plays a crucial role in the accommodation of substrate. This led to the proposition of

“Hinge movement” that would involve a mechanism of the binding-site opening. Such a mechanism has not been observed in any known other m⁵U MTases. However, further studies should verify the proposed mechanism. Moreover, an in-depth analysis offering mechanistic insights into *PhRlmCD*'s substrate selection and binding abilities has been elaborated and compared with its closest homologs. This enhances our understanding of the enzyme's functional characteristics within the broader context of evolution. Nevertheless, it is essential to emphasize that the proposed binding studies must be confirmed through future experimental outputs.

Appendix C. Supplementary data

Supplementary Tables, C.1-C.5

Supplementary Figures, C.1-C.9



CHAPTER 6

Structural and functional
characterization of archaeal $m^1\psi$
rRNA MTase

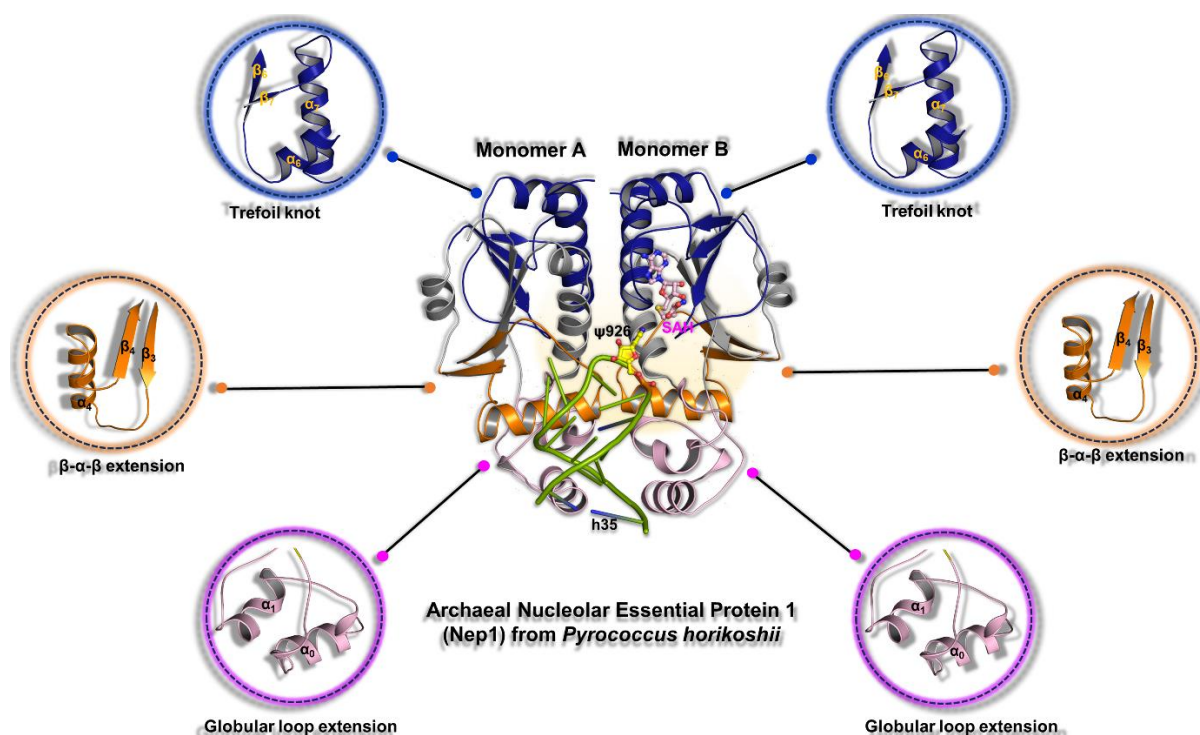
The manuscript for this chapter is under preparation as:

Saha, S. and Kanaujia, S.P. Unveiling the structural and mechanistic insights of the ribosome assembly factor Nep1: a N1-specific pseudouridine methyltransferase (*Under preparation*).

Abstract

The Nucleolar Essential Protein 1 (Nep1) is a crucial element in the formation of small ribosomal subunits in both eukaryotic and archaeal cells. Nep1 has a characteristic fold of SAM-dependent SPOUT rRNA methyltransferase (MTase) and catalyzes the pseudouridine (Ψ) methylation at specific SSU rRNA sites. Mutations in the Nep1 proteins result in a severe developmental disorder in humans and reduced growth in yeast. These consequences suggest the participation of Nep1 in ribosome biogenesis. In this study, we present high-resolution (~2.0 Å) crystal structures of Nep1 homolog from the archaeobacterium *Pyrococcus horikoshii* (*PhNep1*) in both its apo and holo (Adenosine, ADN; 5-methylthioadenosine, MTA; bound form) states. The structural analysis of *PhNep1* reveals an α/β fold featuring a deep trefoil knot akin to the SPOUT domain, with two novel extensions—a globular loop and a β-α-β extension. Moreover, the cofactor-binding site of *PhNep1* exhibits a preformed structure pocket, which is topologically equivalent to cofactor-binding sites observed in other SPOUT-class methyltransferases. Further, the study unveiled that the biological structure of *PhNep1* exists as a homodimer upheld by inter-subunit hydrogen bonds and hydrophobic interactions. Additionally, our findings demonstrate that *PhNep1* exhibits specific methylation activity towards consensus RNA, mainly when pseudouridine (ψ) is located at position 926 within helix 35 (h35) of 16S rRNA in *P. horikoshii*. Notably, during catalysis, the stability of the Nep1-RNA complex relies on conserved arginine residues located between the interface of monomeric subunits.

Graphical abstract



6.1. Introduction

Ribosome biogenesis is a complicated process that involves several steps, including the transcription of ribosomal RNA (rRNA), pre-rRNA processing, chemical modification of rRNA, and the assembly of ribosomal proteins. Chemically modified nucleotides present in rRNA act as glue, which helps to restore the active rRNA conformation needed during the final assembly of the ribosome (Woodson, 2008). During the post-transcriptional step, non-coding RNAs (ncRNA), rRNA, and transfer RNA (tRNA) undergo a variety of chemical modifications such as isomerization of uridine to pseudouridine (pseudouridylation) or addition of methyl, carbonyl, amino, or thio groups (Kaczanowska and Ryden-Aulin, 2007). The two most frequently observed modifications in rRNAs include 2'-O-methylations of ribose moieties and the conversion of uridine into pseudouridine, although other modifications, such as base methylations, are also present (Decatur & Fournier, 2002). Base methylation in RNA takes place at specific nucleobases without the involvement of other factors, while the

other two processes require small nucleolar RNA (snoRNA) to facilitate the site-specific rRNA modifications that include methylation and pseudouridylation (Kiss-László et al., 1996; Ni et al., 1997). Among the three types of rRNA methylations, base methylation is predominant in bacteria, while snoRNA-guided methylation is predominant in eukaryotes and archaea for rRNAs. The two major families of snoRNAs guiding the modification of rRNA are box H (ANANNA, where N represents any nucleotide)/ACA snoRNAs in coordination with pseudouridine kinase Cbf5 (ribonucleoprotein, RNP component of box H/ACA snoRNA), mediate the formation of pseudouridines (Ye, 2007), and Box C (RUGAUGA, where R represents any purine)/D(CUGA), of snoRNAs, along with the Nop1 (RNP component of C/D box), facilitate the introduction of ribose 2'-O-methylations (Tollervey et al., 1993).

Nucleolar essential protein 1 (Nep1), also known as essential for mitotic growth 1 (Emg1), is a highly conserved protein found in eukaryotes and most of the archaea (Liu and Thiele, 2001). In humans, mutation of the Nep1 protein (D86G) is associated with Bowen-Conradi syndrome, a lethal genetic disorder characterized by low birth weight, microcephaly, joint deformities, and failure to thrive (Armistead et al., 2009). While in yeast, it was observed that pre-rRNA processing of the 18S-rRNA precursor was hindered at sites A0, A1, and A2 in temperature-sensitive Nep1 mutants at restrictive temperatures, highlighting its crucial role in ribosome biogenesis (Venema and Tollervey, 1999; Eschrich et al., 2002). Furthermore, studies have demonstrated that restoration of growth in temperature-sensitive Nep1 mutants was accomplished through the overexpression of the synthetase gene, which encodes S-adenosyl-L-methionine (SAM)-a common cofactor for enzyme methyltransferase (MTase). This finding suggests the involvement of Nep1 in rRNA methylation (Eschrich et al., 2002).

The process of rRNA methylation itself is energetically expensive, as significant cellular resources are invested in the production of the respective MTases of methyl donor cofactors, highlighting the importance of this process in biology (Chiang et al., 1996). Most of the MTases modify the nucleic acids by methylation, with the help of the cofactor SAM, which is converted to S-adenosyl-L-homocysteine (SAH) after donating the methyl group (Cantoni, 1975). SAM-dependent MTases are categorized into five distinct superfamilies, with Nep1 proteins falling under class IV SPOUT

MTases. These are recognized by a catalytic domain featuring an unusual α/β fold, which forms a deep trefoil knot, known as the knotting loop, within the structure (Anantharaman et al., 2002).

Interestingly, the SPOUT family is not limited to rRNA methylation; its members also catalyze modifications in tRNA molecules. The two tRNA MTases that served as the founding members of this class are SpoU (tRNA methyltransferase H, TrmH), which catalyzes 2'-O-methylation at G18 (Gm18), and TrmD, which catalyzes N1-methylation at G37 (m¹G37) (Anantharaman et al., 2002; Tkaczuk et al., 2007). Over the past few years, several additional members of this superfamily, such as Trm10, Trm56, TrmY, ribosomal RNA small subunit methyltransferase E (RsmE), ribosomal RNA large subunit methyltransferase H (RlmH), Nep1 and others have been characterized. These enzymes have been associated with various substrate modifications, including m¹ψ, m³ψ, m¹G, m¹A, and 2'-O-methylation on diverse cellular RNAs (Krishnamohan and Jackman, 2018; Strassler et al., 2022).

Nep1 MTases target the N1-methylation of pseudouridine Ψ914/1191 located in the highly conserved consensus motif (5'-C/UUCAAC-3') that is found in the 16S/18S rRNA of both archaea (numbering according to 16S rRNA of *Methanocaldococcus jannaschii*) and eukaryotes (numbering according to 18S rRNA of *Saccharomyces cerevisiae*) (Wurm et al., 2010; Meyer et al., 2011). While this motif repeats several times in eukaryotic and archaeal rRNA sequences, experimental evidence has confirmed that Nep1 MTases exclusively methylate the target site in helix 35 (h35) of small subunit (SSU) RNA. (Wurm et al., 2010; Meyer et al., 2011).

Studies have shown that Nep1 proteins are found in the nucleolus of eukaryotes (yeast and humans), where they methylate the targeted pseudouridine at the N1 position. After assembly of the ribosome subunit and its export into the cytoplasm, another modification (3-amino-3-carboxypropyl modification; acp) occurs in the N3 position of this targeted pseudouridine resulting in m¹acp3-ψ. In contrast to eukaryotes, archaea lack a nucleolus, suggesting Nep1 modification likely occurs in the cytoplasm. It is unknown whether and how the 3-amino-3-carboxypropyl modification occurs in most archaea. However, the archaeon *Haloferax volcanii* contains an N3-(3-amino-3-

carboxypropyl)-uridine (acp³U) nucleotide at an equivalent position (Kowalak et al., 2000).

Besides the MTase activity, the Nep1 MTase serves as a positive regulator. It competes with snR57 (snoRNA) for binding to pre-18S rRNA near helix 47 (h47), thereby facilitating the incorporation of Rps19 (Ribosomal proteins). This process likely involves inducing a conformational switch in the 18S rRNA, which is essential for the maturation of 18S rRNA (Buchhaupt et al., 2006). However, the study had already demonstrated that during the maturation of 18S rRNA in yeast, the absence of methylation activity (not the loss of snoRNA binding) at the consensus rRNA site within h35 leads to a decreased overall growth rate and the accumulation of 18S rRNA precursors (Liang et al., 2009).

Despite the extensive structural and functional insights available for ScNep1 (PDB id: 3OII) and MjNep1 (PDB id: 3BBE) from *S. cerevisiae* and *M. jannaschii*, respectively (Taylor et al., 2008; Thomas et al., 2011), there is still a lot of uncertainty, especially about the Nep1 MTase found in archaea. To gain further knowledge, the structural and functional characteristics of an archaeal Nep1 MTase from *Pyrococcus horikoshii* OT3 (*PhNep1*, ORF id: PH1379) have been investigated. Crystal structures of *PhNep1* have been determined in three different forms: apo protein, bound to adenosine (ADN), and bound to a 5'-methylthioadenosine (MTA) analog of the cofactors SAM or SAH. The crystal structures of *PhNep1* reveal a domain architecture reminiscent of the SPOUT-class of MTases but with two novel extensions (a globular loop and β-α-β extension). Moreover, our analysis also showed that *PhNep1* features a preformed cofactor-binding pocket. Subsequently, our study also highlights that dimeric *PhNep1* represents the biological form of this protein, maintained by inter-subunit hydrogen bonds and hydrophobic interactions. Additionally, the study highlights the importance of conserved arginine residues located in the interface of monomeric subunits of *PhNep1* during the substrate (rRNA) binding.

6.2. Materials and methods

6.2.1. Construction of wild-type expression plasmid

The gene *PhNep1* (ORF id: O50087, 690 bp) from the *Pyrococcus horikoshii* OT3 genome was cloned into the pET-28a(+) plasmid vector to generate a recombinant construct. Amplification of the gene utilized forward (5'-CGC**GCTAGC**ATGGAAGAAAAAAGAGACTTCACTTAATAATAGCCG-3') and reverse (5'-CCGC**CTCGAG**CTAACTCTTAACTTTTAAATAACTTCCC-3') primers, incorporating *NheI* and *XhoI* (bold and underlined) restriction sites, respectively, in the polymerase chain reaction (PCR) (Figure 6.1A). The 6xHis-tag was utilized from the vector at the 5'-end of the cloned gene. After amplification, the gene was double-digested with *NheI* and *XhoI* restriction enzymes before being inserted into the pET-28a(+) vector that had been excised with the same enzymes. The inserted gene was ligated using T4 DNA ligase. To amplify the plasmid, the *E. coli* DH5α competent cells were transformed with the recombinant plasmid containing the gene *PhNep1*. The integrity of the recombinant construct was validated through double digestion, followed by plasmid sequencing (Figure 6.1B).

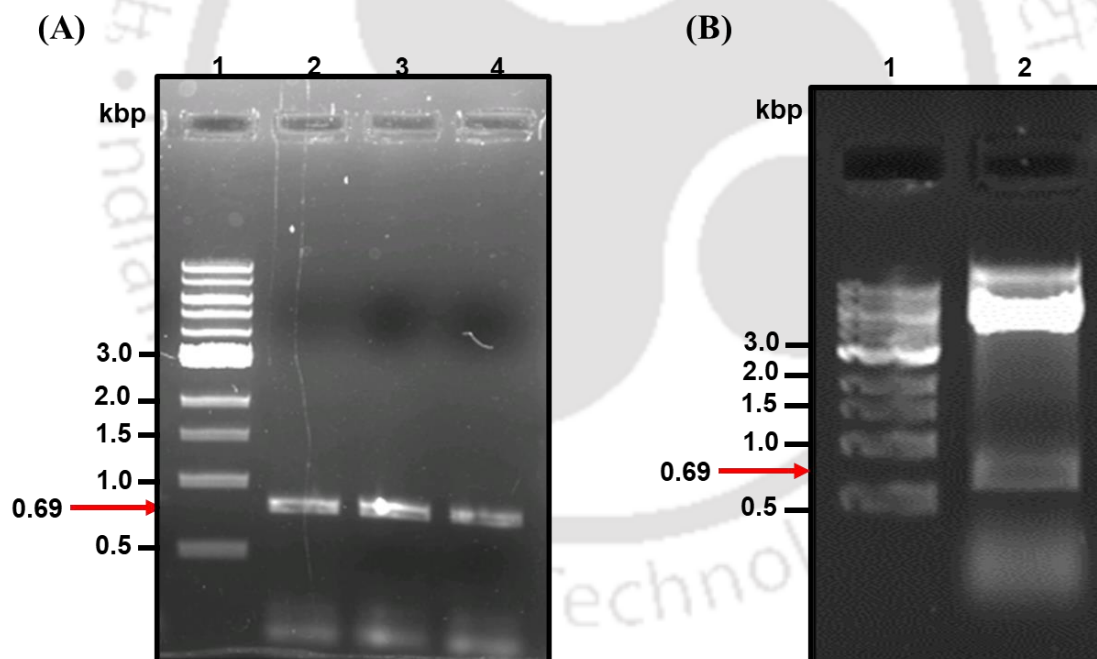


Figure 6.1. Cloning of gene *PhNep1*. (A) PCR amplification of *PhNep1* (690 bp) gene (lane 1: DNA ladder, lane 2-4: PCR amplified product of the gene *PhNep1*). (B) Clone confirmation by double digestion of the vector pET-28a(+) carrying the gene *PhNep1* by *NheI* and *XhoI* restriction enzymes (lane 1: DNA ladder, lane 2: insert pop out after double digestion of the plasmid pET-28a(+)).

6.2.2. Protein overexpression and purification

The *E. coli* Rosetta (DE3) competent cells (Novagen, India) were transformed with the recombinant construct of *PhNep1* to overexpress the protein. The transformed cells were grown in Luria-Bertani (LB) medium with 50 µg mL⁻¹ kanamycin and 34 µg mL⁻¹ chloramphenicol at 37°C until the optical density (OD) of cells at 600 nm reached 0.6–0.8. 30 minutes before induction with 1 mM Isopropyl-1-thio-β-d-galactopyranoside (IPTG), 3% (v/v) alcohol was added to the culture (Figure. 6.2A). The cells were further grown for an additional 12 hours at 25°C with shaking at 180 rpm. Afterward, the cells were harvested by centrifugation at ~4000g for 15 minutes at 4°C and stored at -20°C overnight. Next, the cell pellets were resuspended in a lysis buffer A (50 mM HEPES pH 7.5, 5 mM imidazole, 300 mM NaCl, 5% glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 50 mM L-glutamic acid L-arginine and 5mM DTT) on ice. The cells were then lysed by sonication at 30% amplitude using a 3-second on and 5-second off cycle for 15 minutes. As an initial purification step, the cell lysate was heated at 60°C for 15 minutes. The heated cell lysate was subsequently subjected to centrifugation at ~18500g for 1 hour at 4°C to separate the soluble and insoluble fractions. A clear supernatant fraction was loaded onto a pre-equilibrated pierce centrifuge column (Thermo Fisher Scientific, USA) packed with Ni-nitrilotriacetic acid (Ni-NTA) affinity resin (Qiagen, Germany) for protein binding. After 3 hours of incubation at 4°C, the column was washed in a gradient manner with 10 column volumes (CV) each of buffer B (50 mM HEPES pH 7.5, 10 mM imidazole, 300 mM NaCl, 5% glycerol, 5 mM DTT, and 1 mM PMSF) and buffer C (buffer B with 20 mM imidazole), followed by a final wash using 10 CV of buffer D (buffer B with 50 mM imidazole). Finally, the recombinant protein(s) bound to the Ni-NTA resin was eluted with 250 mM imidazole in buffer B (Figure 6.2B). In the initial batches of *PhNep1*, the eluted fractions were pooled and subjected to desalting to remove imidazole against the final protein buffer E consisting of 50 mM HEPES pH 7.5, 300 mM NaCl, and 5% glycerol. After that, *PhNep1* was concentrated in a Vivaspin turbo 15 centricon (10 kDa cut-off; Sartorius, Germany) up to a concentration of ~8 mg mL⁻¹. Since the protein was found to be co-purified with bound cofactor (Adenosine, ADN), before conducting any in vitro experiments, the later batches of *PhNep1* were first subjected to desalting followed by overnight gradient dialysis in buffer E to remove

imidazole, as well as endogenously bound cofactors. After the dialysis, the protein was concentrated up to ~20 mg mL⁻¹ (Figure 6.2C).

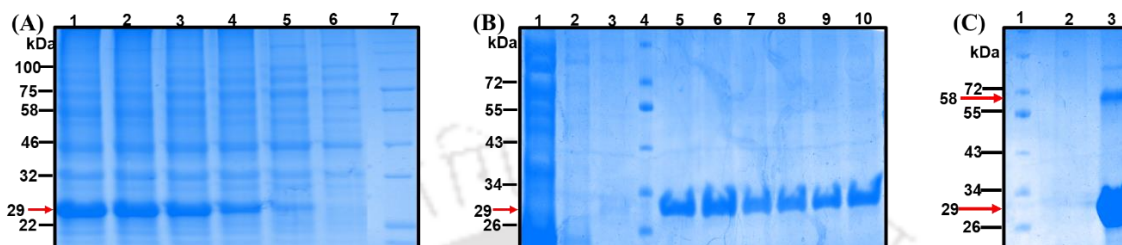


Figure 6.2. Overexpression and purification of *PhNep1*. (A) Overexpression of the protein *PhNep1* in *E. coli* Rosetta (DE3) competent cell at 1 mM IPTG at a different time interval (lane 1-5: 5-1 hr, lane 6: uninduced or 0 hrs and lane 7: protein marker) after induction. (B) SDS-PAGE analysis for the eluted fractions of the protein using Ni-NTA metal-affinity chromatography (lanes 1-3, Unbound, Flow through 1 and 2 & 5-10: eluted fractions of the protein respectively, lane 4: protein marker). (C) SDS-PAGE analysis for the concentrated protein (lane 1: protein marker, lane 2: collector, and lane 3: concentrated protein).

6.2.3. Size exclusion chromatography (SEC) analysis

Size exclusion chromatography (SEC) was conducted using a HiLoad™ 26/600 Superdex™ 200 pg column (GE Healthcare) connected to an AKTA pure chromatography system (Cytiva, Sweden). The column was initially equilibrated with storage buffer E and then calibrated with standard proteins (Sigma-Aldrich) for gel filtration at a 1 mL/min flow rate. Subsequently, affinity-purified *PhNep1* was injected into the above column. Based on the size, protein fractions progress through the column and exit separately, resulting in distinct peaks. To estimate the relative molecular weight (M_w) of *PhNep1*, the distribution coefficient (K_{av}) of *PhNep1*, as well as the standard protein, was determined using the following equation.

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

Where V_e represents the elution volume of the targeted protein, V_t indicates the total volume encompassing both interstitial and intraparticle volumes, and V_o corresponds to the void volume. After that, their corresponding distribution coefficient (K_{av}) was plotted against " $\log (M_w)$ ". Data from standard proteins in size exclusion chromatography (SEC) were utilized to determine the relative molecular mass of *PhNep1*.

6.2.4. Dynamic light scattering (DLS) analysis

The two (monomeric and dimeric) fractions of *PhNep1*, separated by SEC, were evaluated for their tendency to aggregate using dynamic light scattering (LITESIZER 500; Anton Paar, UK). The fractions, at a concentration of $\sim 1 \mu\text{M}$, were diluted in nuclease-free water (Himedia, India). DLS experiments were conducted for the prepared solution, using side scatter angles specified by the instrument. The collected data were used to create a two-dimensional graph with Origin v9.6.

6.2.5. Western blot analysis

To determine the oligomerization state of purified *PhNep1*, it was diluted to a concentration of $50 \mu\text{M}$ in a solution containing 50 mM HEPES (pH 7.5), 300 mM NaCl, and 5% glycerol. Subsequently, the protein was subjected to SDS-PAGE on 12% polyacrylamide gels for 2 hours at 100V. After electrophoresis, the protein was transferred efficiently onto a nitrocellulose membrane using a transfer buffer containing 24 mM Tris, 192 mM glycine, and 20% methanol at 4°C , with the transfer process sustained at 100V for 1.5 hours. Following the transfer, the membrane was incubated with primary antibodies (mouse anti-His₆) and secondary antibodies (rabbit anti-mouse IgG) at dilutions of 1:3000 and 1:5000, respectively. The secondary antibody utilized was conjugated with horseradish peroxidase, facilitating band detection via chemiluminescence upon the addition of the substrate (Bio-Rad, USA). Finally, immunoblots were analyzed using Image Lab Software (Bio RAD).

6.2.6. Mass spectrometric (MS) analysis

The molecular mass of *PhNep1* was determined through matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry analysis.

Sinapinic acid (SA) matrix was prepared at a final concentration of 10 mg mL⁻¹, dissolved in a solvent containing 100% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA) in a 30:70 ratio, respectively. The sinapinic acid was dissolved in the solvents via sonication for 30 minutes, followed by centrifugation at 12000xg for 10 minutes to remove excess matrix particles. As the previous unsuccessful attempts to estimate the molecular weight of *PhNep1*'s dimer, a sandwich application method was employed. In this method, the sample was placed between layers of the matrix. This allowed the top matrix layer to be excited before the sample, ensuring that when the sample was exposed to the laser, photons were released from the matrix (Sinapic acid, SA). As the laser did not directly contact the sample, the protein could maintain its oligomerized state (Liti, 2016). The sandwich application involved adding a layer containing 0.5 µL of matrix, followed by a layer of the sample volume of 1 µL, and finally, another layer containing 0.5 µL of matrix. Each layer was gently air-dried before applying the next layer. After drying, the drops were ionized in a Bruker autoflex speed (Bruker Daltonics, Germany), and the data were analyzed using the built-in flex control analysis software. Graphs were generated using Origin v9.6 for final representation.

6.2.7. Crystallization of *PhNep1*

Crystallization screening was conducted with initial batches of *PhNep1* (without dialysis) using the Crystal Screen and PEG/Ion kits (Hampton Research, USA) by mixing 2 µL each of the protein and buffer solutions (a 1:1 ratio) using the microbatch-under-oil technique at 4 and 20°C. Crystal hits of the protein *PhNep1* were obtained in two different buffer conditions- (1) 0.1 M sodium chloride, 0.1 M HEPES pH 7.5, 1.6 M ammonium sulfate, and (2) 0.2 M potassium sodium tartrate tetrahydrate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 2.0 M ammonium sulfate (Figure 6.3A-6.3B). As the protein was co-purified with bound ADN, it was displaced by cognate cofactor S-adenosyl-L-methionine (SAM) by soaking this crystal and was incubated for ~3 days (Figure 6.3C).

On the other hand, the later batches of *PhNep1* were subjected to overnight dialysis and screened for crystallization using the Crystal Screen, PEG/Ion (Hampton Research, USA), BCS, JCSG plus, and Wizard Cryo (Molecular Dimensions, UK) kits. A crystal

appeared in a new crystallization buffer consisting of 0.1 M Tris pH 8.0, 0.2 M Ammonium sulfate, and 20% v/v polyethylene glycol (PEG) smear broad.

Mini-rRNA fragments were procured from Integrated DNA Technologies (IDT; Coralville, USA) at a 1 μmol scale. These fragments contained architectures similar to h35 (5'-UUGGAUψCAACGCCGG-3' and 5'-UUGGAUUCAACGCCGG-3'). The purity of the mini-rRNA fragments was analyzed on a native polyacrylamide gel in a TBE (Tris-Borate-EDTA) buffer at 25°C to analyze the purity. Ternary complexes (*PhNep1* with SAM/SAH and rRNA) were prepared by heating the rRNA sample at 75°C for 5 minutes and then cooling it on ice for 10 minutes before mixing it with the protein solution. This was followed by heating the protein-rRNA mixture at 50°C for 5 minutes, slowly cooling it down to room temperature, and allowing it to stand for 45 minutes before mixing with SAM/SAH. The ternary complex was set up for crystallization with the protein *PhNep1* (~20mg mL⁻¹ in buffer E) mixed with each h35 rRNA fragment and SAM at a molar ratio of 1:1.2:2. Although crystals of the ternary complex were obtained, they were unable to produce collectable X-ray diffraction.

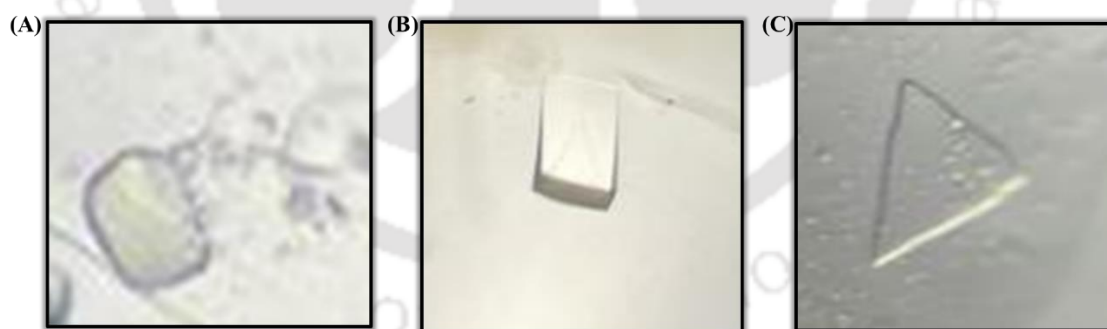


Figure 6.3. Crystallization of the protein *PhNep1*. (A) Protein crystals of *PhNep1* in its apo form. (B) Protein crystals of *PhNep1* complexed with adenosine (ADN). (C) Displacement of adenosine (ADN)-bound *PhNep1* crystal with S-adenosyl-L-methionine (SAM) through soaking for ~3 consecutive days.

6.2.8. Data collection, processing, and structure determination

X-ray diffraction intensity data from the protein crystals of both apo and bound forms of *PhNep1* were collected using the Rigaku MicroMax-007 HF diffractometer, operated at 40 kV and 30 mA, available at the central instruments facility of the Indian Institute of Technology Guwahati, India. The data was collected at a temperature of -173 °C. During the data collection, each frame was exposed for 300 seconds and rotated at 1° oscillation while the detector was kept at a distance of 190 mm. The X-ray intensity data were processed and scaled using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013). The intensity data were then converted to structure factors using the modules in the CCP4 package (Evans, 2011; Winn et al., 2011). The molecular replacement method was used to obtain the structure solution of the protein *PhNep1* using the program Phaser (McCoy et al., 2007). The crystal structure of archaeal Nep1 from *Methanocaldococcus jannaschii* was used as the search model during the molecular replacement. For R_{free} calculations, 5% of the reflections were set aside (Brünger, 1992). The structural models were refined using the program Refmac5 (Murshudov et al., 2011), followed by model building using the program Coot v0.9.4.1 (Emsley et al., 2010). The protein atoms were fitted into the electron density maps contoured at 3.0σ and 1.0σ for $F_o - F_c$ and $2F_o - F_c$ maps, respectively. Subsequently, the electron densities corresponding to water molecules and other components present in the protein and crystallization buffer were fitted. Geometrical and chemical parameters of the refined structures were assessed and validated using the programs PROCHECK and MolProbity (Laskowski et al., 1993; Williams et al., 2018). The refinement and validation statistics for the refined models are provided in Table 6.1. The three-dimensional atomic and structure factors of all the refined structures of protein *PhNep1* have been deposited in the RCSB Protein Data Bank (Berman et al., 2002).

Table 6.1. X-ray crystallographic data collection and refinement statistics for the protein *PhNep1* (apo and holo forms). The values in parenthesis represent the statistics for the last resolution shell. MTA: 5'-methylthioadenosine, ADN: adenosine.

	<i>PhNep1</i>	<i>PhNep1</i>•ADN	<i>PhNep1</i>•MTA
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100

**Chapter 6. Structural and functional characterization of archaeal m¹ψ rRNA
MTase**

Space group	<i>H32</i>	<i>H32</i>	<i>H32</i>
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> =113.81, <i>c</i> =84.64, γ =120	<i>a</i> = <i>b</i> =113.41, <i>c</i> =102.18, γ =120	<i>a</i> = <i>b</i> =113.24, <i>c</i> =102.82, γ =120
Resolution (Å)	71.01-3.10 (3.31- 3.10)	70.81-2.01 (2.06- 2.01)	70.97-2.20 (2.27-2.20)
No. of observed reflections	62217 (11144)	246585 (16745)	160893 (13625)
No. of unique reflections	4773 (847)	17065 (1245)	13053 (1127)
Mn(I) CC(1/2)	0.996 (0.955)	0.999 (0.965)	0.995 (0.952)
Completeness (%)	100 (100)	100 (100)	100 (100)
V_M (Å ³ Da ⁻¹)	2.32	2.32	2.32
Solvent content (%)	47.00	47.00	47.00
Mosaicity (°)	0.70	0.50	0.80
Mean I/ σ (I)	16.40 (5.40)	20.2 (6.00)	13.30 (5.10)
$R_{\text{merge}}^{\dagger}$ (%)	13.00 (41.20)	9.40 (46.60)	0.13 (0.53)
$R_{\text{pim}}^{\dagger\dagger}$ (%)	5.30 (17.00)	3.70 (19.20)	5.50 (23.10)
$R_{\text{meas}}^{\dagger\dagger\dagger}$ (%)	14.10 (44.60)	10.10 (50.50)	14.20 (58.10)
Multiplicity	13.0 (13.20)	14.40 (13.40)	12.30 (12.10)
$R_{\text{work}}/R_{\text{free}}$ (%)	16.39/24.51	15.79/20.93	18.54/24.99
Protein model			
No. of subunits in ASU	1	1	1
Protein atoms	1871	1877	1871
Water molecules	18	141	90
Ligand (number)	-	ADN	MTA
Other molecules	4	11	5
Deviation from ideal geometry			
Bond length (Å)	0.012	0.012	0.012
Bond angles (°)	1.667	1.672	1.668
Average B-factor (Å²)			
Protein atoms	30.26	19.84	29.60
Water molecules	36.80	39.36	40.65

**Chapter 6. Structural and functional characterization of archaeal m¹ψ rRNA
MTase**

Ligands	-	20.52	24.55
Ramachandran plot (%)			
Favored	94.64	97.31	96.43
Allowed	5.36	2.69	3.57
PDB id	9JGB	9JGC	9JGD

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum overall reflections and $\sum_i$ is the sum over i measurements of reflection hkl .

^{††} $R_{\text{pim}} = \sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{j=1}^n |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$, where $I(hkl)$ is the intensity of reflection hkl .

^{†††} $R_{\text{meas}} = \sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{j=1}^n |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl,j}$, where $I(hkl)$ is the intensity of reflection hkl .

6.2.9. Thermal denaturation studies using circular dichroism (CD) spectroscopy

To investigate temperature-induced conformational changes in the secondary structure of both wild-type and mutant proteins of *PhNep1*, Far-UV spectra were measured within the wavelength range of 190 to 260 nm. This was accomplished using a J-1500 circular dichroism spectrometer (JASCO, Japan), employing a 1 mm optical path-length quartz cuvette. Each protein sample (5.0 μM) was dialyzed against the final protein buffer F, consisting of 1 mM Tris-HCl pH 7.5 and 5 mM NaCl. The experimental parameters, including response time, sensitivity, and scan speed, were set to default values (2 s, 100 meg, and 100 nm min⁻¹, respectively) for spectrum measurement. Each experimental run comprised three repeated scans, which were subsequently averaged to generate the final spectra relative to a comparative blank. These spectra, obtained under the specified parameters, were then utilized to investigate the thermal denaturation profiles of *PhNep1*. This investigation covered a temperature range of 20-110°C, with a heating rate of 2°C per 10°C interval, and spanned the wavelength range of 190-260 nm. Denaturation analysis was conducted using the spectra manager software. To ascertain the melting point (T_m) using Circular Dichroism (CD), ellipticity at 208 nm was monitored across a range of temperatures. The apparent fraction of unfolded molecules (F_U) was calculated using the equation:

$$F_U = \frac{(Y_{\text{obs}} - Y)}{(Y_U - Y)}$$

Where Y_{obs} represents the ellipticity value (θ) at 208 nm at each temperature, while Y_U and Y denote the observed values in the baseline regions for the unfolded and folded forms, respectively (Singh and Manoj, 2017). Finally, a graph depicting the fraction unfolded (F_U) against temperature was plotted using the Boltzmann sigmoid (two-state model) equation, using Origin v9.6 software.

6.2.10. Methyltransferase assays

The assay was performed using the MTase-Glo kit (Promega, USA) in 96-well white plates with a solid bottom. This kit utilizes a bioluminescence-based approach to detect the product SAH via a luciferase reaction, which was quantified using the Tecan microplate reader (Zürich, Switzerland). In the assay setup, a 50 µL reaction buffer containing 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM ethylenediamine tetraacetic acid (EDTA), 3 mM MgCl₂, 0.1 mg mL⁻¹ bovine serum albumin (BSA), and 1 mM dithiothreitol (DTT) was prepared. The Michaelis-Menten constant (K_M) of the substrate (h35) was determined at increasing concentrations of rRNA ranging from 1-10 µM while keeping the cofactor (SAM) concentration fixed at 10 µM. This was achieved using the following equation.

$$V_o = \frac{V_{\text{max}} [S]}{K_M + [S]}$$

Where V_o and V_{max} are the initial and maximum reaction velocities, and $[S]$ is the substrate concentration (µM). TFA (0.5%) was used to stop the reaction (Hsiao et al., 2016).

6.2.11. Bioinformatic analysis

The list and amino acid sequences of Class IV- SAM-dependent SPOUT MTases were collected from the UniProtKB database using specific keywords such as "Nep1," "SPOU," "TrmD," "TrmH," or "Trm10," "Trm56", "RsmE", "RlmH" and so on (The UniProt Consortium, 2023). Furthermore, the homologs of *PhNep1* from *P. horikoshii*

were identified using the web tool BLAST, employing a sequence identity (SI) threshold of more than 20% (Altschul et al., 1990). To elucidate evolutionary relationships among *PhNep1* (ORF id: PH1379), its homologs, and members of the SPOUT superfamily, phylogenetic trees were constructed using the maximum-likelihood (ML) method in MEGAX, with 1000 bootstrap replicates (Kumar et al., 2018). The resulting trees were visualized using the iTOL v6.3 web tool (Letunic and Bork, 2021). Subsequently, multiple sequence alignment (MSA) of *PhNep1* and its homologs was conducted using the Clustal Omega web tool (Sievers and Higgins, 2014), and the outcomes were rendered using ESPript v3.0 (Gouet et al., 2003). To identify structurally conserved or invariant water (IW) molecules, flexible structures of all available Nep1 proteins were superimposed onto the target structure using the align module of the PyMOL program, assisted by a custom shell script. The 16S rRNA sequence of *P. horikoshii* was retrieved from RNA Central (RNAcentral Consortium, 2021). For identification of conserved h35 in *P. horikoshii*, MSA was performed using h35 sequences across three domains of life. Subsequently, the secondary structure of h35 was constructed using the R2DT program and validated with Mfold (Zuker, 2003; Sweeney et al., 2021). The binding affinities of the protein *PhNep1* with SAH were estimated by molecular docking using a freely available program, AutoDock v4.2 (Morris et al. 2009). The three-dimensional structure of h35 was then modeled and docked using ModeRNA and HDock web servers, respectively (Rother et al., 2011; Yan et al., 2020). Molecular interactions between the protein and substrate rRNA (h35) were analyzed using the Coot program (Emsley et al., 2010).

6.3. Results

6.3.1. The putative protein PH1379 is Nep1 ortholog present in *P. horikoshii*

The UniProtKB database has preliminary annotated the ORF id PH1379 (UniProt id: O50087) from *Pyrococcus horikoshii* OT3, a type of hyperthermophilic archaeon, as a "probable ribosomal RNA small subunit methyltransferase Nep1." To get more information on this protein, a homology search was conducted on the UniProtKB/SwissProt database using the BLAST program and default parameters. The results indicate that protein PH1379 exhibits homology to the Nep1 group of MTases found in archaea (*M. jannaschii*, *MjNep1*; QC: 93%, SI: 45%) and eukaryotes (S.

cerevisiae, ScNep1; QC: 96%, SI: 27%); QC: query coverage and SI: sequence identity. The Nep1 group of proteins are annotated as rRNA MTases and belongs to the class IV SAM-dependent SPOUT MTases, a characteristic of 'α/β knot fold' (Tkaczuk et al., 2007). These SPOUT MTases are associated with a variety of substrate methylations to form methylated nucleotides such as m¹Ψ, m³Ψ, m¹G, m¹A, and 2'-O-methylation (Xm, where X = A, G, C, or U) on diverse RNAs (rRNA and tRNA) (Table D.1).

To validate that the protein PH1379 belongs to the Nep1 MTases, a phylogenetic analysis was conducted among various SPOUT relatives (Table D.1). The results indicate that the SPOUT MTases superfamily is divided into four major clades with numerous subclades mostly containing at least one functionally characterized enzyme (Figure 6.4A and Table D.1). All the Nep1 and Trm10 proteins located in the clade I and IV respectively, are well distinguished by other mix proteins clades of SPOUT MTases. Moreover, the tree indicates that PH1379 has been clustered within the archaeal Nep1 family, exhibiting similarity to its eukaryotic counterpart (Figure 6.4A). Additionally, the domain architecture of PH1379 is assessed and compared with other SPOUT relatives. The results show that PH1379 contains the SPOUT domain in the C-terminus, similar to its Nep1 relatives (Table D.1).

The previous studies indicate that the rRNA-consensus motif (5'-C/UUCAAC-3') appears three times in the eukaryotic and twice in the archaeal rRNA; however, in eukaryotes, two helix 35 (h35) and helix 47 (h47) and archaea only h35 was experimentally confirmed (Wurm et al., 2010). Thus, Nep1 MTase can modify at 1191 and 1575 near h35 and h47 of 18S rRNA, respectively, in *S. cerevisiae*, while in archaea, it was 914 of 16S rRNA (number according to 16S rRNA of *M. jannaschii*) (Figure D.1).

Moreover, this consensus sequence is methylated only if the targeted site contains pseudouridine instead of uridine. Subsequently, in eukaryotes and archaea, the targeted pseudouridine first methylated at the N1 position, followed by hypermodification at N3. (Brand et al., 1978, Kowalak et al., 2000). Thus, to identify the probable locations of the rRNA-consensus motif in h35, 16S rRNA in *P. horikoshii* was examined along with known target sites identified in eukaryotes and archaea. The results show that the

nucleotide residues at 926 in *P. horikoshii* correspond to 914 and 1194 sites in *M. jannaschii* and *S. cerevisiae* (Figure D.1). Following that, in vitro methylation assays were conducted using different types of RNA substrates. As a negative control, nonspecific RNA oligonucleotide identical to helix 45 (h45) was also used. The result revealed that the protein PH1379 methylates the RNA oligonucleotides containing a pseudouridine at the position corresponding to 926 in *P. horikoshii* of 16S rRNA (Figure 6.4B). However, the enzyme is unable to methylate almost similar oligonucleotide fragments when the corresponding site contains uridine instead of pseudouridine. Thus, the protein PH1379 was found to exhibit similar characteristics with other well-characterized Nep1 MTases and referred to as *PhNep1*.

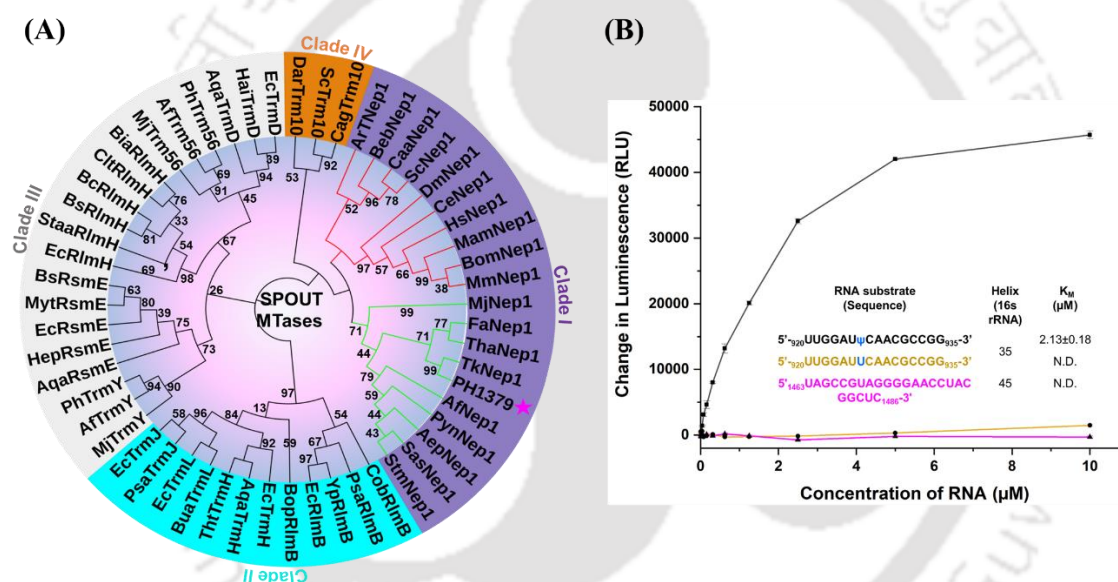


Figure 6.4. PH1379 is an archaeal Nep1 ortholog. (A) A phylogenetic tree was constructed using the maximum likelihood (ML) method with 1000 bootstrap replicates. A phylogenetic tree showing the evolutionary relationship was generated using 54 different SPOUT MTases, including the protein PH1379; details are provided in Table D.1. The result indicates that SPOUT MTases are divided into four cluster-Nep1, two mix classes (RlmB-TrmH-TrmL-TrmJ, TrmY-RsmE-RlmH-Trm56-TrmD), Trm10. Additionally, the protein PH1379 belongs to the Nep1 cluster. The protein PH1379 is indicated by a magenta asterisk. (B) In vitro MTase assay of PH1379 with different RNA substrates. The result exhibits that PH1379 was unable to methylate the nonspecific rRNA fragment (h45) (red) and only methylate the rRNA substrate of h35 containing a pseudouridine at the position corresponding to 926 (black) with K_M of ~2

μM. However, no methylation was recorded for the same sequence if the pseudouridine was replaced with uridine (magenta). The target nucleotide (926) of PH1379 is marked in blue.

6.3.2. The protein *PhNep1* exhibits both conformational and compositional stability

Prior to crystallization, it is essential to evaluate the stability of the protein in regard to its (i) composition and (ii) conformational state (Deller et al., 2016). To evaluate these two characteristics, purified *PhNep1* underwent various biophysical analyses, including dynamic light scattering (DLS) and circular dichroism (CD) experiments. The dimeric fraction of *PhNep1*, isolated through size exclusion chromatography (SEC), was subjected to monitoring its hydrodynamic diameters. The objective was to evaluate the compositional stability of the dimeric fraction of *PhNep1*. Subsequent DLS measurements of the dimeric *PhNep1* revealed a single peak with an average hydrodynamic diameter of 8.2 nm, suggesting a monodispersed solution with uniform size and shape and free from aggregates. As a control, the monomeric fraction of *PhNep1* was also measured. The results indicated a single peak with an average hydrodynamic diameter of almost half (3.9 nm) of that of the dimeric fraction (Figure 6.5A). This comparative assessment of DLS spectra confirms the level of homogeneity in the dimeric fraction of *PhNep1*.

Furthermore, in order to estimate the conformational stability of *PhNep1*, circular dichroism (CD) experiments were conducted to observe its secondary structural elements and thermal stability. Analysis of the far-UV CD spectra revealed that the protein is comprised of ~37% α-helices, ~14% β-strand, and ~49% other secondary structures (Figure 6.5B). The effect of temperature on the structural integrity of *PhNep1* was monitored through thermal ramping experiments conducted across a wide range from 20 to 110°C. It was observed that *PhNep1* maintains its conformational stability up to ~70°C (Figure 6.5C-6.5D).

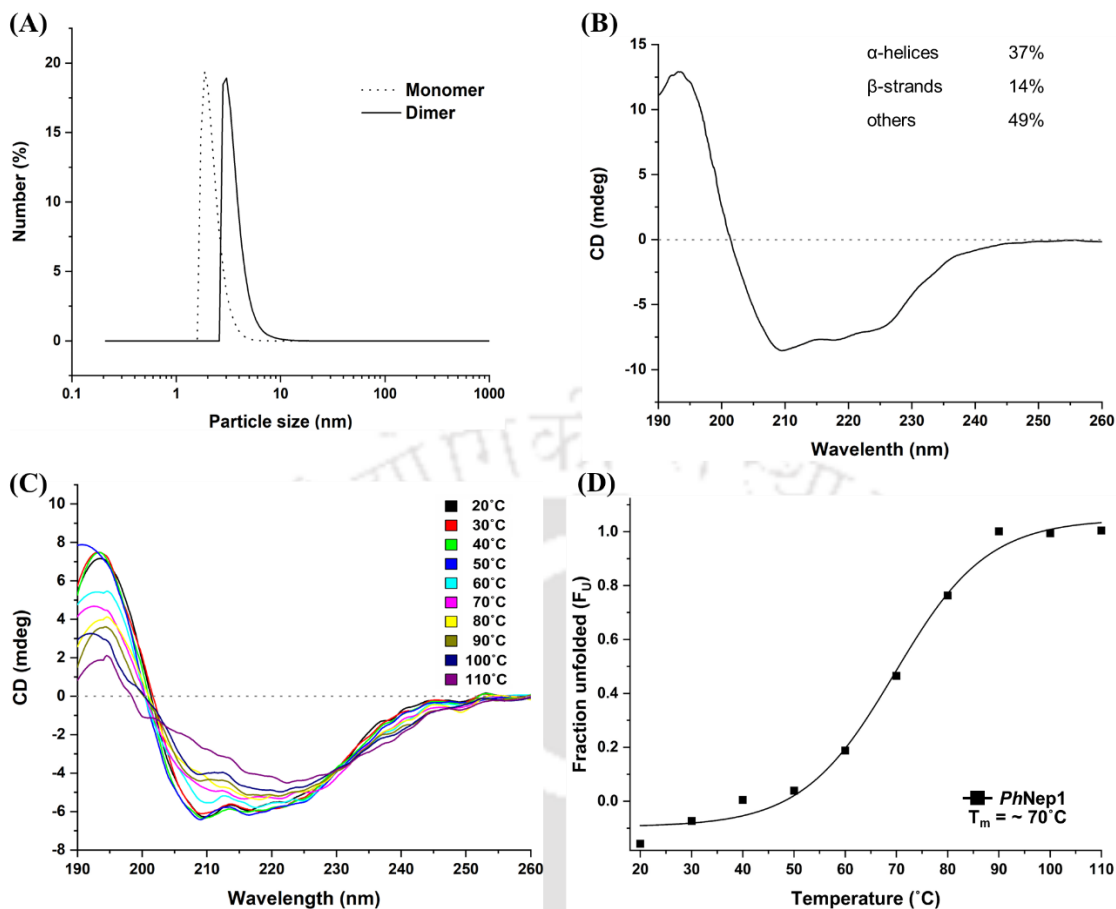


Figure 6.5. Maintaining compositional and conformational stability of *PhNep1*. (A) Dynamic Light Scattering (DLS) measurements of the dimer fraction of *PhNep1* reveal a monodisperse solution, suggesting the presence of the protein in a dimeric state with an average hydrodynamic diameter of 8.2 nm (black solid line) and notably devoid of aggregates. As a control experiment, the monomeric fraction of the protein was also assessed to demonstrate an average diameter of 3.9 nm (black dotted line). (B) A Far-UV CD spectrum ($\lambda = 190$ -260 nm) of *PhNep1* elucidates its secondary structural content. (C) Thermal denaturation experiments on *PhNep1* unveil changes in secondary structure with increasing temperature. The initial spectra (at 20°C) are represented by a black line, followed by spectra at other temperatures depicted in different colors. The overlay facilitates the identification of the temperature (T_m) at which changes in the secondary structure of the protein occur. (D) The thermal unfolding curve for *PhNep1* (depicted by a filled square) was generated by monitoring the CD ellipticity at 208 nm. The curve illustrates that the melting temperature of *PhNep1* is $\sim 70^\circ\text{C}$.

6.3.3. Overall structural fold and active-site pocket of the protein *PhNep1*

6.3.3.1. The Architecture of *PhNep1*

The protein underwent overnight dialysis prior to crystallization, resulting in the formation of an apo crystal for *PhNep1*, with a resolution of 3.1 Å. The crystal structure of the protein *PhNep1* belongs to the *H32* space group and contains one protomer in the asymmetric unit (Table 6.1). Similar to other reported Nep1 MTases, *PhNep1* also consists of a long N-terminal domain (NTD, 1-147 residues) and a small C-terminal domain (CTD, 151-227). The architecture of *PhNep1* adopts an $\alpha+\beta$ fold, characterized by seven twisted β -strands encompassing both domains. Six of these strands form a parallel arrangement, while one strand (β_3) is oriented anti-parallel, sandwiched between α -helices. This arrangement follows a $\beta_7\uparrow\beta_5\uparrow\beta_6\uparrow\beta_1\uparrow\beta_2\uparrow\beta_3\downarrow\beta_4\uparrow$ pattern, with the core topology of the five parallel β -strands ($\beta_7\uparrow\beta_5\uparrow\beta_6\uparrow\beta_1\uparrow\beta_2\uparrow$) recognized as the minimal conserved core in SPOUT domain-containing MTases (Ahn et al., 2003; Buchhaupt et al., 2006). As for the α -helices, *PhNep1* possesses a total of eight, where seven (α_0 , α_1 , α_2 , α_3 , α_4 , α_6 , α_7) are situated along one side of the β -sheet, while one α -helix (α_5) is located on the other side (Figure 6.6A-6.6B).

The NTD of Nep1 proteins contains two unique extensions, setting it apart from the typical core motif seen in other class IV SAM-dependent SPOUT MTases (Tkaczuk et al., 2007; Leulliot et al., 2008). The first extension is an elongated globular loop (residues 17–65), devoid of regular structural features except for the presence of ordered and distorted α -helices α_0 and α_1 , respectively. Interestingly, α_0 is majorly absent in most of the Nep1 proteins except certain archaeal species like *A. fulgidus* and *P. horikoshii* (Figure D.2A). Across all available structures of Nep1 MTases, this loop notably contains a high abundance of positively charged residues, suggesting their significant involvement in substrate stabilization during catalysis. Moreover, the globular loop demonstrates a higher positive charge in archaeal Nep1 proteins (32-37%) compared to their eukaryotic counterparts (22-25%) (Figure 6.6A).

On the other hand, the second insertion comprises a specific ~40 amino acid segment forming a β – α – β structural motif. This segment comprises anti-parallel β -strands β_3 and

β₄ connected by an α-helix α₄, situated between β₂ and α₅ within the protein structure (Figure 6.6A-6.6B). Notably, this β–α–β extension harbors a cluster of highly conserved residues not only in *PhNep1* but also in its homologs from diverse species (Figure D.2A), suggesting potential involvement in RNA binding. Thus, it emerges as a distinctive signature motif for the Nep1-family of proteins.

In contrast, the small CTD of *PhNep1* encompassing β-strands β₅ to β₇ and α-helices α₆ and α₇ adopts a distinctive deep trefoil knot structure. Here, β₅ and α₇ are nestled beneath connecting loops L13 (β₆–β₇, residues 183-197) and L14 (β₅–α₆, residues 200-227), respectively (Figure 6.6A-6.6C). The knot region in Nep1 is held together by a complex network of hydrophobic interactions, forming a densely packed hydrophobic core. The side chains of hydrophobic residues interact predominantly with each other, creating the initial layer of this core by tightly packing β₆ (Pro178, Val180, and Val182) against the hydrophobic surface of α-helix α₇ (Ile211, Val212, Val215, and Val223). The inner core consists of the turn near the β₇ (Ile199 and Leu201) and α₆ (Phe168, Leu171, and Leu172), which are snugly packed due to hydrophobic interaction with Ile211, Val215, and Val223 of α₇. Additionally, the contact of Pro178 with the side chains of residues Leu171 and Leu172 of α₆ further helps in stabilizing this core form of the knot (Figure 6.6C). Most of the residues involved in this hydrophobic interaction network are either conserved or substituted with other hydrophobic residues in Nep1 sequences from various species. This knot is crucial for both the formation of the co-factor-binding site and a component of the dimer interface (See below).

6.3.3.2. The cofactor-binding site of *PhNep1*

Instead of undergoing desalting followed by dialysis, the initial batches of *PhNep1* only underwent desalting during protein purification, a rapid and efficient method with minimal protein loss (Görisch, 1988). Interestingly, crystals obtained from these batches revealed additional electron density in the active-site pocket, distinct from that of the amino acid residues. This extra density was identified as an endogenously bound cofactor, adenosine (ADN), bound to the active site of the protein, likely co-purifying with *PhNep1* (Table 6.1). As mentioned in the previous study, a co-purified cofactor can often be displaced by a cognate cofactor with equal affinity or higher affinity (Müller, 2017). Thus, the cognate cofactor (SAM) molecule was utilized in soaking for

ADN-bound crystal and was incubated for ~3 days. This resulted in the binding of 5'-methylthioadenosine (MTA) in the place of ADN (Table 6.1). This is likely due to the hydrolysis of SAM predominantly producing MTA as the primary degradation product (Parks and Schlenk, 1958).

Nevertheless, the adenine N6 atom of MTA/ADN forms hydrogen bonds with the residues Tyr202 and Glu204, while N1 and N7 form hydrogen bonds with the residues Leu201 and Leu206, respectively. The two ribose hydroxyl groups of MTA/ADN form interactions with Gly183, while Met156 and an invariant water molecule (IW34) form hydrogen bond only with the 2' and 3' hydroxy groups of the cofactor, respectively (Figure 6.6D, Figure D.2B, and Table D.2). Moreover, the root-mean-square deviation (RMSD) of C_α atoms between the apo *PhNep1* and *PhNep1*•ADN/MTA is ~0.18 Å, indicating no structural alteration is necessary to accommodate the cofactor. This phenomenon was also observed in other homologs of the protein, indicating that cofactor binding sites are preformed in the Nep1 MTases (Taylor et al., 2008; Thomas et al., 2011).

However, to identify the mechanism of accommodating the α-amino acid part of the cognate cofactors, SAH was docked with the apo form of *PhNep1* using a blind molecular docking strategy. The structural comparison between the holo form (*PhNep1*•MTA) and the SAH docked apo form of *PhNep1* demonstrates a remarkable resemblance. The outcome suggests that the molecule SAH occupies the identical position in the cofactor binding pocket of apo *PhNep1*, indicating that no structural alterations are required for cofactor binding. Additionally, the findings reveal that the α-amino acid moiety of the cofactor is stabilized through coordination with three water molecules (IW34, IW33, and IW38) and Gly188 (Figure D.2C and Table D.2). In archaeal Nep1, this invariant water molecule is exclusively present in protein forms bound to the cofactor, where it plays a crucial role in forming hydrogen bonds critical for stabilization. In eukaryotes, the equivalent invariant water molecule is missing. However, the protein ScNep1 possesses an acidic residue, Asp214, which is present, capable of forming hydrogen bonds with the O3' and N atoms of the cofactor SAH (Table D.3).

Chapter 6. Structural and functional characterization of archaeal m¹ψ rRNA MTase

The structural comparison among *Ph*Nep1 (PDB id: 9JGB), *Af*Nep1 (PDB id: 307B), *Mj*Nep1 (PDB id: 3BBD), and *Sc*Nep1 (PDB id: 3OII) highlights remarkable preservation of active-site residues, evident in both sequence and secondary structure (Figure 6.6E and Table D.3). This conservation is facilitated by the presence of a trefoil knot in the CTD of the proteins, which establishes a spatial hydrophobic core crucial for the formation of the co-factor-binding site.

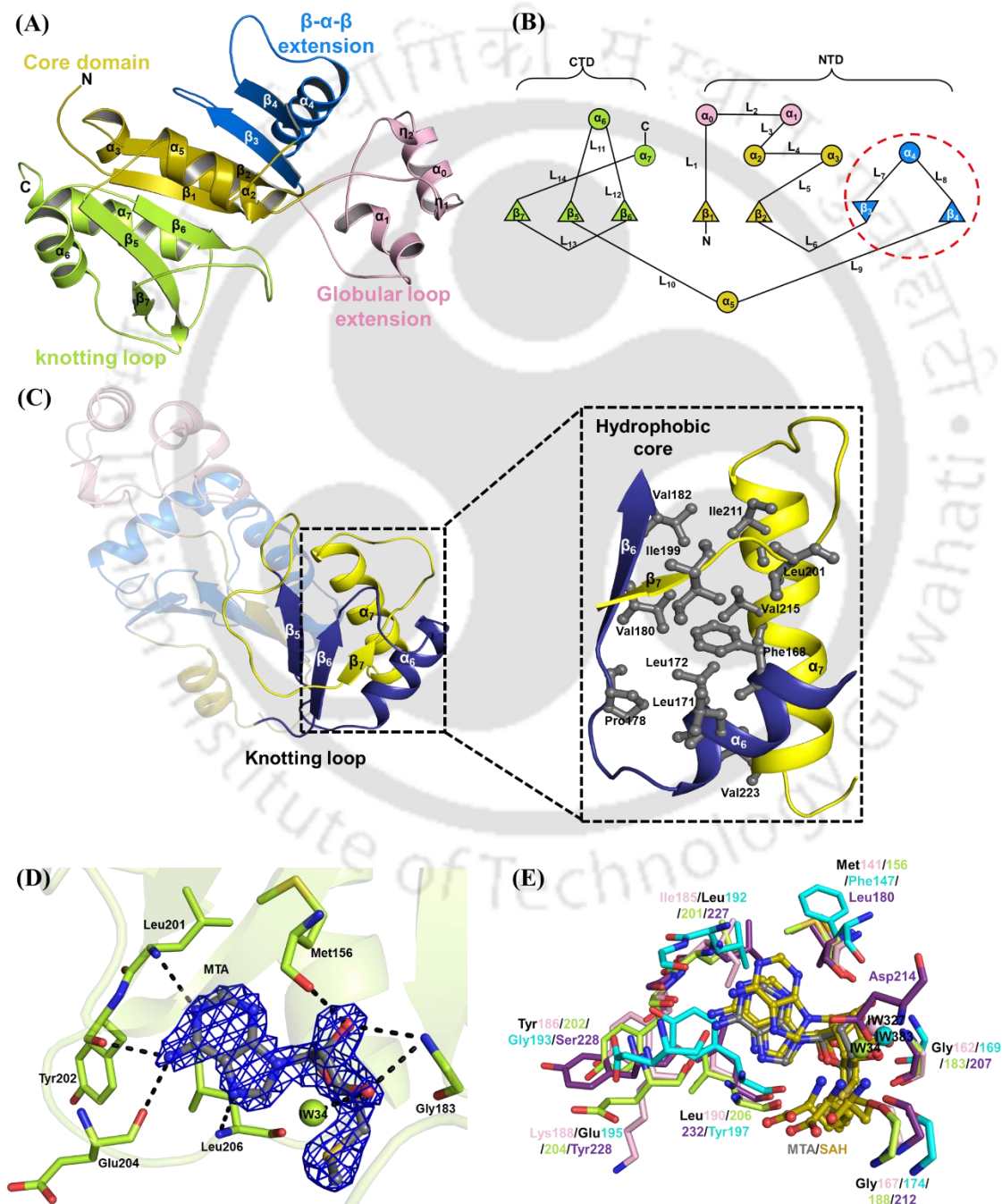


Figure 6.6. The architecture of protein *PhNep1* and its active site. (A) The three-dimensional structure of *PhNep1* delineates the secondary structural elements of the N-terminal domain (NTD) with the globular and β-α-β extensions depicted in pink and blue, respectively, and the core domain is highlighted in olive. Meanwhile, the C-terminal domain (CTD) is represented in limon. (B) A topology diagram of *PhNep1* illustrates β-strands as triangles and α-helices as circles, with the β-α-β extension demarcated by a dotted red line. (C) The trefoil knot in the CTD of *PhNep1* with β₅, β₆, and their connecting loop, including α₆, are shown in navy blue, while β₇ and α₆, threaded through the ‘knotting loop,’ are highlighted in yellow. The magnifying view of the core of the trefoil knot in *PhNep1*, where the side chains of the hydrophobic amino acids are colored in a grey stick, contributes to its formation of a hydrophobic core. The magnifying view of the hydrophobic core of the trefoil knot. The side chains of amino acids that contribute to the formation of the hydrophobic core are colored in grey and shown as stick models. (D) The MTA molecule bound to the protein is visualized as a ball-and-stick model, accompanied by its 2F_o-F_c electron density map contoured at the 1σ level. Active site residues interacting with MTA are depicted as limon lines, with the interactions shown as black dashed lines. (E) Superposition of the active site residues of *PhNep1*•MTA (PDB id: 9JGD, limon) and its homologs *AfNep1* (PDB id: 307B, cyan), *MjNep1* (PDB id: 3BBD, pink), and *ScNep1* (PDB id: 3OII, violet) revealing the characteristic class IV SAM-dependent SPOUT MTases.

6.3.4. Inter-subunit hydrogen bonds stabilize the *PhNep1* dimer

In the SDS-PAGE analysis of concentrated *PhNep1* protein, we observed two distinct bands at ~29 and ~58 kDa. The first band likely corresponds to the monomeric state of the protein, while the second band may represent either the oligomeric state of *PhNep1* or an impurity encountered during the purification process (Figure D.3A). To confirm this, the purified protein was further analyzed via Western blotting using anti-His₆ antibodies, which confirmed the identity of the second sharp band at ~58 kDa as the oligomeric state of the protein (Figure 6.7A). Subsequently, the oligomerization ability of the *PhNep1* in solution was evaluated using SEC. The result exhibits two (high and low) intensity peaks observed at 221 and 243 mL, respectively (Figure D.3A). By determining these two fractions’ distribution coefficient (k_{av}) along with SEC standards

proteins, we have estimated their molecular weight ~55 and ~26 kDa for the corresponding fractions 221 and 243 mL, respectively (Figure D.3B). This indicates that in solution, *PhNep1* predominantly exists as a dimeric entity, akin to its homologs (Taylor et al., 2008; Thomas et al., 2011). To accurately determine the molecular weights of the monomeric and dimeric states of *PhNep1*, MALDI-TOF analysis was conducted using the sandwich method. The analysis revealed two high-intensity peaks corresponding to molecular weights of 29,330 and 58,456 Da (Figure D.3C). The first peak corresponds to the monomeric state of *PhNep1*, which is consistent with its theoretical molecular weight of 29,333 Da, while the second peak represents its dimeric state. As a positive control, the molecular weight of the previously reported dimeric protein, 5-methylthioribose 1-phosphate isomerase (M1Pi), was also evaluated using a similar approach (Figure D.3D) (Gogoi et al., 2019).

It is surprising that the structure of *PhNep1* only contains a single protomer in the asymmetric unit. This is because one protomer in the dimeric system has the same conformation as the other, resulting in zero RMSD between them. As a result, asymmetric units of these types of dimeric proteins are made up of only one protomer (Iida et al., 2021). According to the PISA analysis of *PhNep1*, it has a homo-2-mer multimeric state or biological assembly. The results also reveal that the conformational variances between the two monomeric subunits of *PhNep1* are confined to solvent-exposed surface loop L7 (α_4 - β_3). Particularly, within the L7 loop, the conserved motif (₁₀₆RLPR₁₀₉) regions exhibit close contact between the two monomeric subunits (Figure 6.7B and Figure D.2A). Therefore, dimerization of *PhNep1* leads to a significant reduction in the solvent-accessible surface- accessible surface area (SASA) by ~1750 Å², compared to its Nep1 relatives, which exhibit values of ~ 1300, 1209 and 1500 Å² for *MjNep1*, *AfNep1*, and *ScNep1*, respectively. The magnitude of this buried surface area is indicative of biologically significant and stable protein-protein interactions (Nooren and Thornton, 2003). The dimer interface is associated with the residues located on helices α_1 , α_2 , α_3 , and α_4 of NTD and α_7 of CTD of *PhNep1* (Figure 6.7B). Meanwhile, the interface in eukaryotic Nep1 comprises a combination of electrostatic interactions or salt bridges, hydrogen bonds predominantly at the edges, and a few hydrophobic interactions at the core (Thomas et al., 2011). However, *PhNep1* primarily relies on inter-subunit hydrogen bonds and hydrophobic interactions for its association.

Specifically, the two hydrogen bonds are observed between the side chains of Asn75 (α_2) from both monomers, as well as between the side chains of Asn75 and Asn213 from both monomers. Additionally, there are two hydrogen bonds formed between the carboxylate group of the Glu84 (α_3) side chain of one monomer and the hydroxyl group of Tyr202 (near α_7) (Figure 6.7C). This results in a total of six inter-subunit hydrogen bonds in this region, contributing to the formation of the dimer interface. Within this intermolecular interface, the cluster of helical bundles is established through two helices, α_2 and α_7 , of each monomer, characterized by two sets of hydrophobic interactions. In the first set, the side chains of Ile71, Leu74, and Trp209 from one monomer are tightly packed against the corresponding monomer. Meanwhile, the second set, which consists of hydrophobic residues Val217 and Trp221, repeats this similar interaction pattern (Figure D.3E). Besides the primary dimer interface, six more inter-subunit hydrogen bonds are established through interaction with two additional helices bundles (α_1 and α_4) situated in the N-terminal domain (NTD) of *PhNep1*. Four of these six involve the side chain of Thr48 (α_1) in one monomer and the side chains of Asn112 and Asn113 (α_4) in the other monomer. The remaining two hydrogen bonds form between the side chains of Asn112 from both monomers (Figure 6.7C).

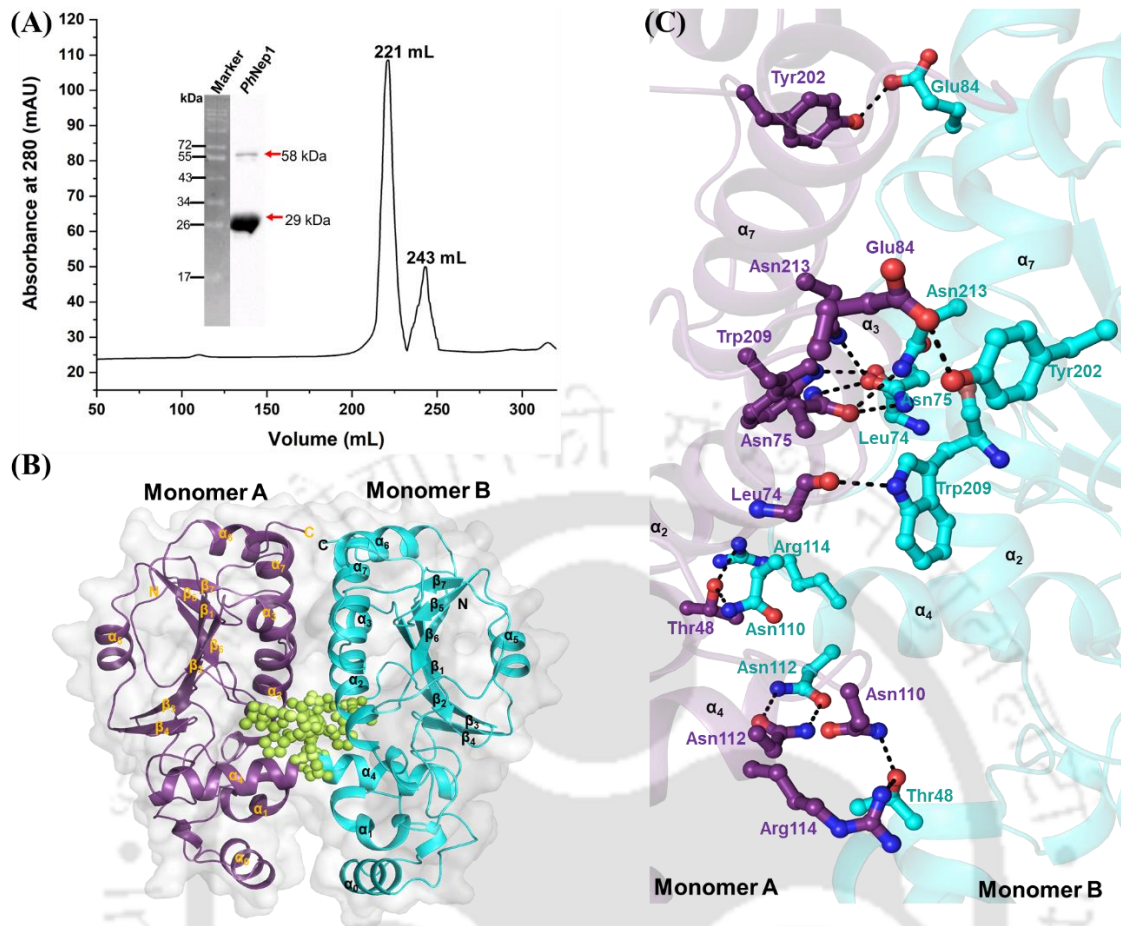


Figure 6.7. Dimerization of *PhNep1*. (A) Western blot analysis of *PhNep1* using anti-His₆ antibodies. Lane 1 displays the standard molecular weight marker, while lane 2 shows the oligomerization state of purified *PhNep1* detected by anti-His₆ antibodies. The oligomerization state of the protein was further validated by size-exclusion chromatography profile of proteins using a HiLoad™ 26/600 Superdex™ 200 pg column (GE Healthcare). The chromatographic profile revealed that the predominant protein population elutes as dimers, with a minor fraction existing as monomers. (B) The PISA server predicts that *PhNep1* forms a dimeric assembly. The monomers (A and B) are highlighted in violet and cyan, respectively. The close contact area of the conserved motif (106RLPR109) is shown in the light green sphere. (C) The dash lines represent the hydrogen bonding interactions across the dimer interface. The interacting residues of each monomer are labeled with residue numbers and the same color as the corresponding chain.

6.3.5. Structural intricacies accentuate the mechanism of RNA binding by *PhNep1*

6.3.5.1 Possible RNA-binding site and active center

Class I MTases are the most abundant group identified so far and are generally composed of two domains- a catalytic domain (Rossmann fold) and RNA binding domains (RBDs) as the RNA recognition motif (RRM) (Schubert et al., 2003; Dominguez, & Allain, 2005). However, the class IV SPOUT domain containing *Nep1* rRNA MTases contains no specific RNA binding domain. Instead, these MTases are comprised of highly positively charged residues. For example, on average, the protein *PhNep1* and its archaeal homologs (*MjNep1* and *ScNep1*) contain 37 basic residues (Arg or Lys) in their total sequence length. However, many of these residues, especially lysine residues, are not generally conserved and probably appear to stabilize the surface charge of the *Nep1* proteins and maintain salt bridges.

On the other hand, the previous studies reported that four highly conserved arginine residues, Arg54/88, Arg95/129, Arg98/132, and Arg102/136 in *MjNep1/ScNep1* were found to be engaged in rRNA binding and mutation of any of this residue results in a loss of the enzymatic function of protein (Taylor et al., 2008; Wurm et al., 2010; Krishnamohan and Jackmen, 2018). In *PhNep1*, these four conserved arginine residues are present on the interface region of both the monomers. Arg65 (Arg54/88) is positioned on the monomer B while Arg106 (Arg95/129), Arg109 (Arg98/132) and Arg113 (Arg102/136) are associated with monomer A (Figure D.4A). The Arg65 is part of conserved sequences (Arg-Pro-Asp-Ile), located near the α_2 of *Nep1* protein, while the other three conserved arginine residues are located around and within α_4 part of the β - α - β extension (Figure D.2A).

Furthermore, analysis of the electrostatic surface potential of dimeric *PhNep1* reveals highly positively charged areas distributed within the interfaced regions between the two monomers (Figure D.4B). This suggests that, in addition to the two regions (α_2 and α_4 of the β - α - β extension), another region, namely an elongated globular loop spanning residues 17–65, resides in the interface of *PhNep1*. The positively charged residues within the sequence of this loop region exhibit poor conservation. These three interface

regions of *PhNep1* seem to collectively form a positively charged cleft, likely primed for RNA binding.

6.3.5.2. Interaction of *PhNep1* with rRNA

The previous study showed that the crystal structure of the *ScNep1* dimer forms a ternary complex with the cofactor SAH and the consensus mini-rRNA of h35 in 18S rRNA of *S. cerevisiae* (Thomas et al., 2011). In which the targeted nucleotide (U1191) in h35 of the 18S rRNA is observed to be flipped out and directed towards the S^δ atom of the cofactor (SAH). This alignment is probably facilitated by interactions with conserved active-site residue with consensus nucleotides of mini-rRNA of *S. cerevisiae* (Jedrzejewski et al., 2023).

On the other hand, *PhNep1* has the ability to recognize the corresponding consensus mini-rRNA (target nucleotide ψ926) located in h35 of 16S rRNA of *P. horikoshii* (Figure D.1). Subsequently, crystal hits of the archaeal ternary complex (*PhNep1*•SAH•h35) were successfully obtained. However, due to poor-quality diffraction, the three-dimensional structure of these complexes could not be resolved. As the RMSD of C_α atoms between the cofactor bound *ScNep1* (*ScNep1*•SAH) and ternary complex (*ScNep1*•SAH•h35) is ~0.3 Å indicating higher structural similarity with minimum induced fit mechanism observed during the substrate binding for *Nep1* protein. This has intrigued us to predict the molecular details of the enzyme *PhNep1* through an alternative approach. First, a three-dimensional model of the mini-rRNA (h35 of 16S rRNA) in *P. horikoshii* was predicted using homology modeling, with the crystal structure of mini-rRNAs from *S. cerevisiae* (PDB id: 3OIN) serving as the template (Figure D.5A-D.5B). Secondly, the crystal structure of cofactor-bound *PhNep1* cofactor bound structure (PDB id: 9JGD) was docked with mini-rRNA present in h35 of 16S rRNA of *P. horikoshii* to generate a likely similar biologically relevant ternary complex of *PhNep1*, known as a structural model (Figure D.5C-D.5D). The results of molecular docking reflect that positively charged areas located between the interface regions of the two monomers create catalytic groove of *Nep1* proteins. This groove allows the accommodation of the negatively charged mini-rRNA fragment from the 16S rRNA of *P. horikoshii* (Figure 6.8A). These observations further substantiate that the biologically active oligomeric state of *Nep1* proteins is a dimer. Although, each

monomeric unit of Nep1 contains an α/β knotted hydrophobic core, which creates a pocket for its cofactors (Figure 6.6D, Figure D.2B-D.2C). Thus, this catalytic groove and cofactor binding pocket together form the active site core of these proteins.

The active site residues of *PhNep1* establish numerous crucial contacts with its mini-rRNA consensus motif (5'-₉₂₅C/UψCAAC₉₃₀-3'), particularly with the neighboring bases of the targeted sites, thereby conferring specificity through hydrogen bonding interactions. For instance, the N3 and N4 atoms of the third base of the consensus rRNA motif, C927, are contacted through bidentate hydrogen bond interactions with N^{η1} of conserved Arg113 of monomer A. Additionally, N4 atoms of this nucleotide interact with O^{γ1} atom Thr48 of monomer B. The N1 and N6 atoms of the next two successive adenine residues (A928 and A929) in the consensus sequence mediate the backbone interactions between the nitrogen and oxygen atoms of amides and the carboxyl group of the conserved Arg106 of monomer A, respectively. Furthermore, the N6 atom of A928 forms a hydrogen bond with the carboxyl O atom of residue Glu104 of monomer A. Finally, the last base (C930) of this consensus motif forms salt bridges between the phosphate backbone of the base and the N^ζ atom of Lys124 of monomer A (Figure 6.8B and Table D.3). In addition to the consensus rRNA motif, nucleobases C931, C932, and C933 engage in interactions with active site residues of *PhNep1*. Specifically, the N2 atoms of nucleobase C931 interact with the N^{η2} atom of conserved Arg113 in monomer A, while the O4' and O2' atoms of nucleobases C932 and C933 interact with the N^{ε2} and N^{η2} atoms of His51 and Arg55 in monomer B, respectively (Figure 6.8B and Table D.3).

In our structural model, the targeted conserved nucleobase ψ926 is flipped out from the consensus loop and binds in a pocket between the two monomers adjacent to the bound SAH molecule. The atoms of the target nucleobase (ψ926) interact with the conserved arginine residue mentioned earlier. For instance, the O4 and O4' atoms of ψ926 form interactions with the N^{η2} and N^{η1} atoms of Arg65 in monomer B, respectively. These interactions are crucial not only for maintaining and positioning the target nucleobase in the active site of Nep1 but also for the Arg65 to initiate the methyl transfer reaction by donating the proton to the O4 atom ψ926 (Figure 6.8C and Figure D.6). In this position, the distance between the N1 atom of the target nucleobase (ψ926) and the S^δ atom of

SAH measures ~4 Å (Figure 6.8A, 6.8C and Figure D.6). This observation is consistent with a previous report on methyl transfer (Zheng and Bruice, 1997). Akin to ScNep1, PhNep1 also contains the conserved residue Arg65, which is stabilized by interactions involving its N^{η2} and N^ε atoms and the O^{δ1} and O^{δ2} atoms of conserved Asp67 in monomer B. The crucial role of Asp67 in stabilizing the essential interaction with Arg65 is emphasized by its significance. Mutations in Asp67, equivalent to Asp86 in HsNep1, where it is substituted with glycine, result in Bowen-Conradi syndrome, a severe developmental disorder leading to early childhood mortality (Armistead et al., 2009). Additionally, the N^{η2} atoms of conserved Arg106 and Arg109 in monomer A form hydrogen bonds with the O4 atom of ψ926. While the O2' atom of ψ926 interacts with the N^{η2} atom of Arg109. The conserved Arg109 is firmly held in place by hydrogen bonding involving the N^{η1} and N^{η2} atoms of Arg132 and the O^{δ1} and O^{δ2} atoms of conserved Asp78 in monomer A (Figure 6.8C and Table D.3). The conserved residue Asp78 of PhNep1 (equivalent to Asp101 in ScNep1) is proposed to serve as the primary proton acceptor during the methyl transfer process (Figure D.6). Furthermore, our analysis indicates that this conserved Asp78 residue is situated ~7 Å away from the N1 atom of the targeted nucleobase (ψ926), a scenario reminiscent of ScNep1 (PDB id: 3OIN). In the enzyme ScNep1, the conserved Ser233 plays a crucial role in mediating proton transfer. However, in certain archaeal species like AfNep1 (PDB id: 3O7B), the equivalent conserved position has Thr198 instead of serine. In contrast, PhNep1 does not possess any of this residue in that equivalent position to transfer protons from Asp78 (Figure D.2A). The detailed structural analysis of PhNep1 showed that no such neighboring polar residues are also available between Asp78 and the N1 atom of ψ926. Furthermore, aside from the conserved arginine residues, the O2 atom of ψ926 engages in hydrogen bonding with the amide nitrogen atom of Ala208 in monomer A (Figure 6.8C).

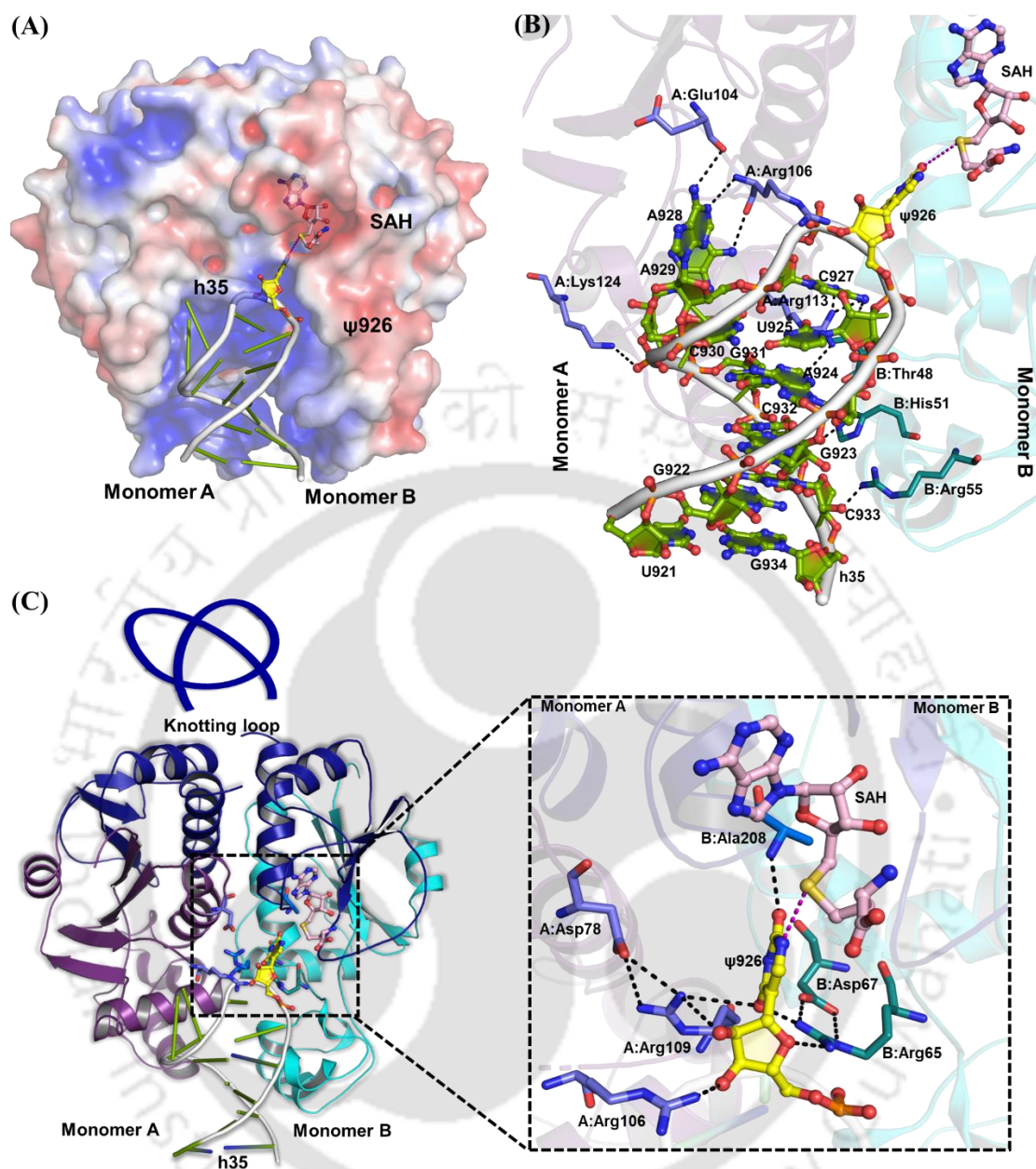


Figure 6.8. Molecular docking suggests how *PhNep1* recognize rRNA h35 of 16S rRNA of *P. horikoshii*. (A) Representation of electrostatic charge distribution on *PhNep1* docked with mini-rRNA consensus motif, depicting positively-charged surfaces in blue (+5 kT/e) and negatively-charged surfaces in red (-5 kT/e). The red dashed line represents the transmethylation distance. The targeted $\psi 926$ and SAH are highlighted in yellow and pink stick. (B) RNA bound in the cleft between the two monomers. Residues are colored according to the monomer from which they originate (violet-monomer A; cyan-monomer B). (C) The figure depicts the interaction between the target nucleotide of $\psi 926$ with the active site residues of *PhNep1*. The deep trefoil

knot in both the monomers is highlighted in deep blue, while hydrogen bonding interactions are illustrated as black dashed lines.

6.4. Discussion

SAM-dependent MTases are categorized into five evolutionarily distinct classes (classes I–V) based on their structural and biochemical characteristics. The most abundant group among these classes is class I MTases, distinguished by a typical Rossmann fold motif (Schubert et al., 2003). Following closely behind is the family of class IV SPOUT MTases, which feature a catalytic domain with a unique α/β fold, forming a deep trefoil knot within the structure (Anantharaman et al., 2002). While most of the SPOUT superfamily members have similarities, the Nep1 MTases are clearly distinguished from them by their unique sequence features and the positioning of the SPOUT domain. The size exclusion chromatography (SEC) profile of purified archaeal Nep1 from *P. horikoshii* (*PhNep1*) indicates that the protein exists predominantly as a homodimer. However, a minor peak corresponding to the monomeric form is observed, which is typical for an oligomeric protein undergoing dissociation during column traversal. The elution position of such a protein is determined by the weighted average size of the molecules within the fraction (Woodbury et al., 2002). The active form of *PhNep1*, existing as a homodimer, was isolated and utilized for all in vitro studies, including crystallization.

The crystal structure of *PhNep1* unveils its characteristic core domain with distinctive features: a two-signature insertion globular loop and a β - α - β extension. The *PhNep1* protein has an elongated globular loop that spans from residue 17 to 65. This loop is more structurally organized when compared to a similar extension found in eukaryotic Nep1. The reason for this difference in structure is most likely due to the presence of well-packed helices α_0 , which are mostly absent in eukaryotic Nep1. This structural disparity is one of the primary reasons for the differences in the Nep1 structure between archaea and eukaryotes. Meanwhile, the protein displays an unusual insertion of two anti-parallel β -strands (β_3 and β_4) separated by α -helices (α_4) that interact with the core domain of five β -sheet to form the α/β -sandwich topology reminiscent of Rossmann-like fold (Thomas et al., 2011). Nevertheless, these extensions comprise highly positive charge residues within the interface between the two monomers, forming a catalytic

groove for substrate (rRNA) binding. In addition, a deep trefoil knotting loop situated in the C-terminal half of *PhNep1* acts as the signature for the SPOUT-class of MTases. The knot forms a deep crevice rich in various aliphatic and aromatic residues, establishing the hydrophobic core. Specifically, in this core region, Ile199 of β_7 , Leu201, Leu206 in the loop L14 (α_7 - β_7), and Ile211, Val215 of helices α_7 form cofactor binding pocket to accommodate adenine moiety of the bound cofactor. Comparative analysis between the apo form and holo forms of *PhNep1* reveals no discernible structural alterations upon cofactor binding. This observation was reinforced by a detailed investigation of the interactive residues between *PhNep1* and its homologs with the cofactor. The conservation of the hydrophilic core across all Nep1 proteins suggests that the cofactor binding site is inherently encoded within the architecture of *PhNep1* and its homologs. However, upon detailed comparative analysis between archaeal and eukaryotic Nep1, contrasting results were found. In all available archaeal Nep1, an invariant water molecule is present, forming two critical hydrogen bonds to stabilize the α -amino tail of the cofactor. However, in eukaryotes, the equivalent invariant water molecule is absent.

As observed in other structurally characterized Nep1 relatives, *PhNep1* also forms a dimer. The dimer interface involves various noncovalent interactions, including inter-subunit hydrogen bonds, salt bridge formation, and hydrophobic interaction. Residues located on helices α_1 , α_2 , α_3 , and α_4 of the N-terminal domain (NTD), as well as α_7 of the C-terminal domain (CTD), play a crucial role in holding the two monomers in close contact through inter-subunit hydrogen bond interactions. Additionally, the close proximity of the L7 loop and the hydrophobic packing associated with parallel helices of each monomer's α_2 and α_7 in the N- and C-terminal regions promote dimer formation, resulting in a significant decrease in solvent accessible surface area (SASA) compared to other Nep1 relatives. Interestingly, some of the dimeric populations of *PhNep1* were found to retain their oligomeric state in the presence of ~3.2% SDS, as determined by SDS-PAGE analysis (Figure 6.7A and Figure D.3A). This phenomenon is probably attributed to the high stability exhibited by hyperthermophilic proteins, which possess unusually low exposed surface areas, leading to unfavorable interactions with SDS. (Gentile et al., 2002).

Our in vitro MTase assay demonstrates that *PhNep1* selectively targets its substrate by methylating RNA sequences resembling nt 920–935 of h35 of *P. horikoshii* 16S rRNA only when the target nucleobase contains a pseudouridine (ψ) in place of uridine (U) at position 926. This methylation specifically takes place at the N1 position of the targeted nucleobase. Previous research has indicated that, besides N1 modification, such as in *S. cerevisiae* and archaeon *Haloferax volcanii*, a hypermodification (3-amino-3-carboxypropyl; acp3-ψ) occurs at the N3 position of the targeted pseudouridine (Brand et al., 1978; Kowalak et al., 2000). However, this hypermodification at the N3 position of the target pseudouridine remains elusive in the archaeal class Thermococci, which consists of three genera: *Thermococcus*, *Palaeococcus*, and *Pyrococcus*.

Based on our observations, we propose that 926 sites within h35 of the 16S rRNA of *P. horikoshii* contain at least one N1-methyl-pseudouridine at the corresponding site. Additionally, it is possible that the N3 position of the target pseudouridine may undergo hypermodification (Kowalak et al., 2000). Suppose the hypermodification occurs at the N3 position of the target pseudouridine; it should occur after the methylation of pseudouridine at the N1 position (m¹acp³ψ). If this assertion is inaccurate, then *PhNep1* would be unable to identify the fragment solely featuring pseudouridine (ψ) at position 926. Instead, it would necessitate a mini-rRNA containing N3 hypermodified pseudouridine (acp³-ψ) at position 926.

Our structural model indicates that RNA binds within the cleft formed between the two monomers, similar to its homologs protein ScNep1 and MjNep1 (Taylor et al., 2008; Wurm et al., 2010). A meticulous analysis of the protein-rRNA docked complex predict that four conserved arginine residues at the interface of the monomeric subunits play a significant role in RNA anchoring by forming multiple hydrogen bonds. Among these residues, three are predominantly involved in anchoring the RNA: Arg106, Arg109, and Arg113 in monomer A and Arg65 in monomer B.

Additionally, our mechanistic model supports the assertion from a previous report that base flipping is a prerequisite during the modification of nucleobase by rRNA MTases (Bhujbalrao et al., 2022). Based on insights gleaned from sequence conservation and detailed structural analysis, our study delineates a mechanism for the N1-methylation of

the target pseudouridine (m¹ψ926) catalyzed by *PhNep1*. We propose the critical role of Asp78, a conserved residue equivalent to Asp101 in *ScNep1*, serves as the primary proton acceptor during catalysis. In both cases, the aspartic residue was located (~7 Å) distantly from the N1 atom of the target pseudouridine. As a result, this necessitates a mediator residue to relay proton transfer from N1 of ψ to conserved aspartic acid. Unlike other *Nep1* relatives (*ScNep1*, *MjNep1*, and *AfNep1*), *PhNep1* lacks the residue of serine or threonine for this role. Furthermore, a detailed structural analysis of *PhNep1* confirms the absence of adjacent polar residues between N1 of ψ and conserved aspartic acid. Therefore, it is posited that during catalysis in *PhNep1*, a water molecule, particularly the -OH group, serves as a conduit connecting the N1 of ψ and the conserved proton-accepting residue, Asp78. This phenomenon was supported by the previous report (Hassanali et al., 2013). Further, the experiments necessitate the validation of our hypothesis.

Moreover, our analysis also indicates that conserved residue Arg65 forms hydrogen bonds with the O4 oxygen of the target nucleobase (Ψ926). This suggests that Arg65 (equivalent to Arg54, Arg84, and Arg88 in *MjNep1*, *HsNep1*, and *ScNep1*, respectively) likely facilitates the deprotonation of N1 and stabilizes the negative charge in the intermediate (Figure D.6). The critical stabilization of residue Arg65 is ensured by a conserved residue Asp67 (equivalent to Asp54, Asp86, and Asp90 in *MjNep1*, *HsNep1*, and *ScNep1*, respectively) through the formation of salt bridges. Mutation of these conserved aspartic acid residues in humans results in severe growth defects leading to Bowen-Conradi syndrome (Armistead et al., 2009).

6.5. Conclusion and future perspectives

In summary, the study extensively explores the structural and mechanistic characteristics of the *PhNep1* protein by conducting a thorough comparative analysis with its *Nep1* counterpart. This study emphasizes the importance of hydrophobic core packing in their SPOUT domain for forming a trefoil knot, which is crucial for accommodating cofactors and facilitating dimer formation. Furthermore, the study suggests a mechanism for N1-methylation at the target ψ926, catalyzed by *PhNep1*, which indicates that Asp78 mediates proton acceptance through a water molecule. In the future, exploring the proton transfer mechanism in archaea devoid of any mediator

polar residue for transferring the proton from N1 position ψ to conserved Asp78 will be compelling. Therefore, validating our hypothesis with experimental evidence becomes imperative.

Appendix D. Supplementary data

Supplementary Tables, D.1-D.3

Supplementary Figures, D.1-D.6





Conclusion and Summary

Conclusion and Summary

The present study is aimed at investigating RNA methyltransferases (MTases) and their respective sites of modification from a hyperthermophilic, anaerobic archaeon, *Pyrococcus horikoshii* OT3. The preliminary analysis suggests the existence of 21 MTases that may be involved in RNA MTase, including 10 rRNA and 11 tRNA MTases. The study focused on two major classes of SAM-dependent rRNA MTases, Class I (which modifies m⁵C, m₂⁶A, and m⁵U) and Class IV (which modifies m¹ψ). The work presented in this thesis can be broadly divided into two parts. The first part deals with the computational characterization of archaeal m⁵C RNA MTases of *P. horikoshii* OT3. The second part focuses on the structural and functional characterization of three m₂⁶A, m⁵U, and m¹ψ rRNA MTases from the same organism, using various experimental approaches, including X-ray crystallography. The bioinformatic analysis indicated that the genera *Pyrococcus* possesses the highest number of RNA m⁵C MTases (five). From which two proteins (PH0851 and PH1991) have been reported as bacterial RsmB and eukaryotic NSUN6 homologs. Hence, this study provides an in-depth characterization of the remaining three putative RNA m⁵C MTases (PH1078, PH1374, and PH1537) to suggest the presence of the signature architecture(s) and catalytic residues plausibly involved in the binding of their cognate ligands, SAM and substrate RNAs. Moreover, the archaeal proteins PH0851 and PH1078 belong to the same subfamily (subfamily IV) owing to the presence of the NusB domain in the N-terminal extension (NTE) and a conserved catalytic domain. This phenomenon indicates the existence of two RsmBs in the archaeon *P. horikoshii*. Additionally, the other paralogs, PH1374 and PH1537, belong to the same subfamily (subfamily V) due to the absence of the extensions (i.e., NTE or C-terminal extension, (CTE)) but the former is specific to rRNA and homolog of eukaryotic protein NSUN1 whereas the latter is most likely homolog of eukaryotic protein NSUN2, a tRNA MTase. Furthermore, the findings from this study lead to a novel evolutionary classification for RNA m⁵C MTases from Cluster I to IV. Overall, the work provides significant insights into the functional annotation of the RNA m⁵C MTases, which may pave the path for treating various epigenetics-associated diseases.

In the second part of the thesis, three rRNA methyltransferases, namely m₂⁶A (referred to as *PhD1MT1*, ORF id: PH1823), m⁵U (referred to as *PhRlmCD*, ORF id: PH1259), and m¹ψ (referred to as *PhNep1*, ORF id: PH1379), are structurally and functionally

characterized using various experimental approaches.

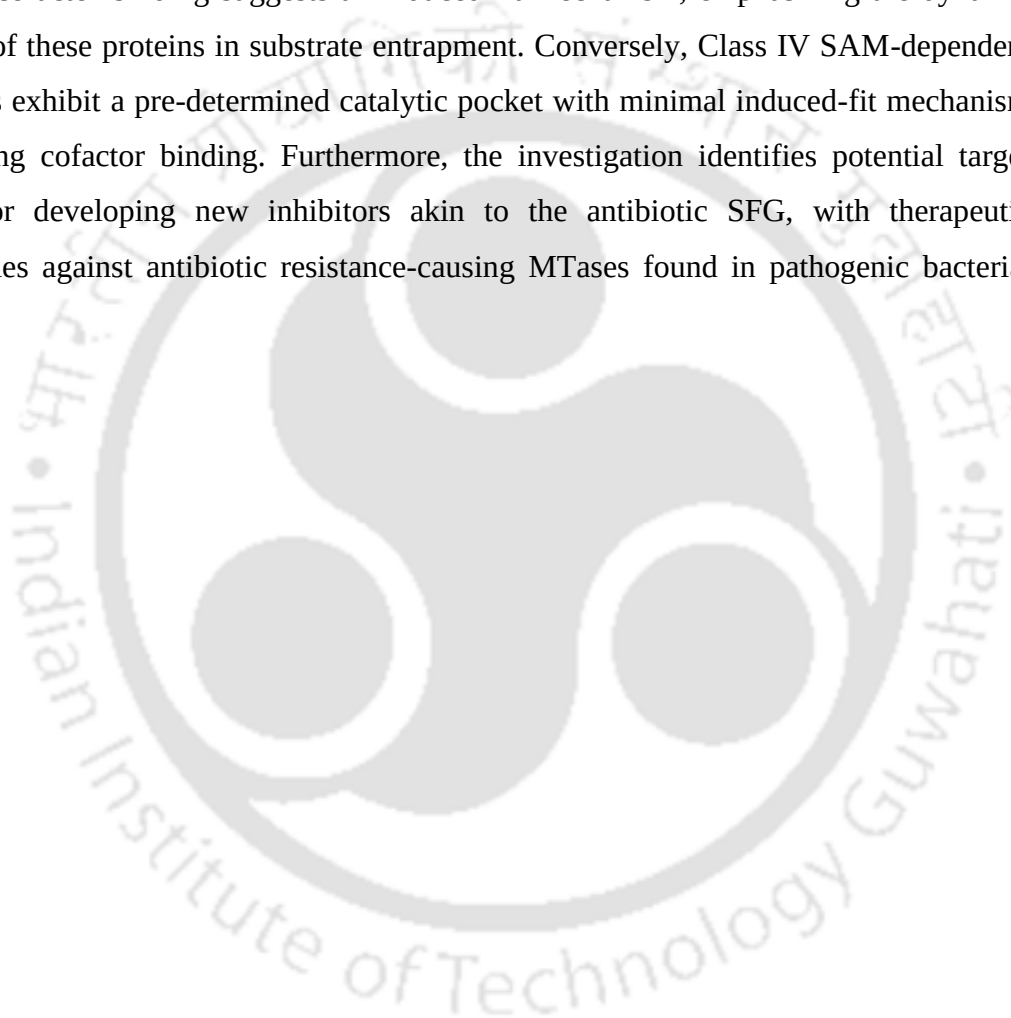
In this study, crystal structures of a dimethyladenosine transferase 1 (DIMIT1) from a hyperthermophilic archaeon *P. horikoshii* (*Ph*DIMIT1, ORF id: PH1823) were studied in both apo and holo forms with a resolution range of 2.0 to 3.3 Å. The study found that the antibiotic sinefungin (SFG) moves differently than similar molecules SAM and SAH within the DIMIT1-SFG complex, shedding light on catalytic inactivation. Therefore, only the cognate cofactors (SAM and SAH) can expand the domain interface to facilitate the substrate's positioning in the catalytic pocket.

The second part of the thesis also reports the first crystal structure of *Ph*RlmCD (ORF id: PH1259), an rRNA-specific archaeal m⁵U MTase. Through structural and functional investigations, it became evident that the iron-sulfur cluster plays a crucial role in its functionality. Introducing disulfide bridges instead of this cluster hindered substrate binding at the active site. In addition, the protein has unique features (elongated positively charged loops in the central domain and rotational variations in the TRAM domain) that affect its substrate selectivity. Analysis of protein-RNA complex interaction suggests its potential for dual methyltransferase activities, as observed in its ancestral protein, by interacting with two putative mini-rRNA fragments.

Finally, this investigation elucidates the crystal structure of nucleolar essential protein 1 (Nep1) derived from the archaeon (*Ph*Nep1, ORF id: PH1379), presenting both its apo and holo forms at a resolution ranging from 2.0 to 3.1 Å. The study reveals that *Ph*Nep1 harbors the SPOUT domain, characterized by a profound trefoil knot adorned with two novel extensions. The cofactor-binding pocket of this protein class is inherently structured by the protein's architecture, requiring no conformational changes upon cofactor encapsulation. The study also found that *Ph*Nep1 is a homodimer held together by inter-subunit hydrogen bonds and hydrophobic interactions. Additionally, the study showed that *Ph*Nep1 only methylates consensus rRNA when pseudouridine (ψ) is at position 926 within helix 35 (h35) of 16S rRNA in *P. horikoshii*. The stability of the Nep1-RNA complex is reliant on conserved arginine residues.

Altogether, the findings of this investigation provide a comprehensive overview of archaeal RNA MTases in *P. horikoshii*, suggesting diverse mechanisms of substrate binding. Structural and functional analyses of DIMIT1, RlmCD, and Nep1 from this

archaeon, alongside comparisons with counterparts from other domains of life, offer evolutionary insights into rRNA MTases. The comprehensive structural analysis of archaeal MTases brought to light that cofactor binding is independent of substrate rRNA binding. Moreover, the study delineates disparities between Class I (for example, *PhD1MT1*) and IV (for example, *PhNep1*) SAM-dependent MTases. Class I MTases demonstrate the importance of catalytic loops in facilitating cofactor encapsulation and substrate positioning within their active sites. The observed conformational flexibility during cofactor binding suggests an induced-fit mechanism, emphasizing the dynamic nature of these proteins in substrate entrapment. Conversely, Class IV SAM-dependent MTases exhibit a pre-determined catalytic pocket with minimal induced-fit mechanism following cofactor binding. Furthermore, the investigation identifies potential target sites for developing new inhibitors akin to the antibiotic SFG, with therapeutic properties against antibiotic resistance-causing MTases found in pathogenic bacteria.



APPENDIX A
Supplementary data to Chapter 3

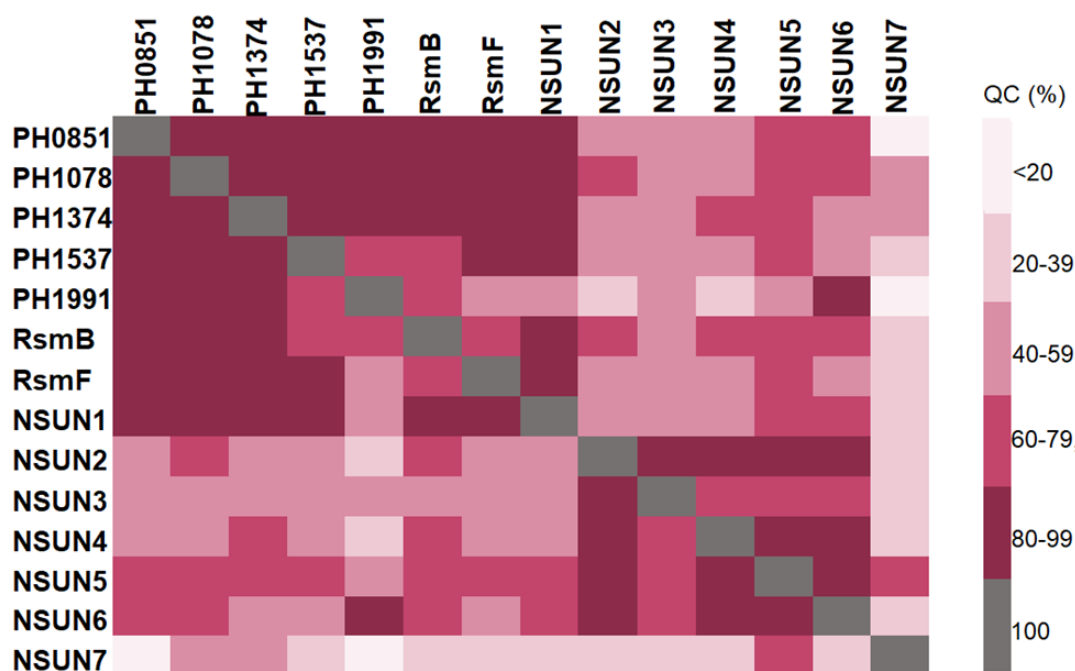


Figure A.1. The heat map depicting the pair-wise sequence similarity between the core domain of the RNA m^5C family of MTases. An increasing gradient of red shows the protein sequences with increasing similarity. QC: query coverage.

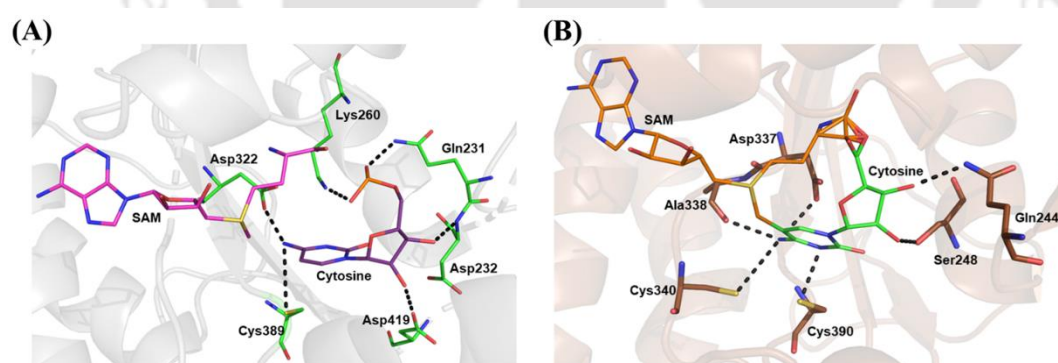


Figure A.2. A Comparative docking study of *EcRsmB* and its archaeal homolog **PH0851**. The active site in the crystal structure of (A) *EcRsmB* (PDB id:1SQF) and (B) PH0851 (PDB id: 2YXL) with docked cytosine reveal the arrangement and key interacting residues in the active site.

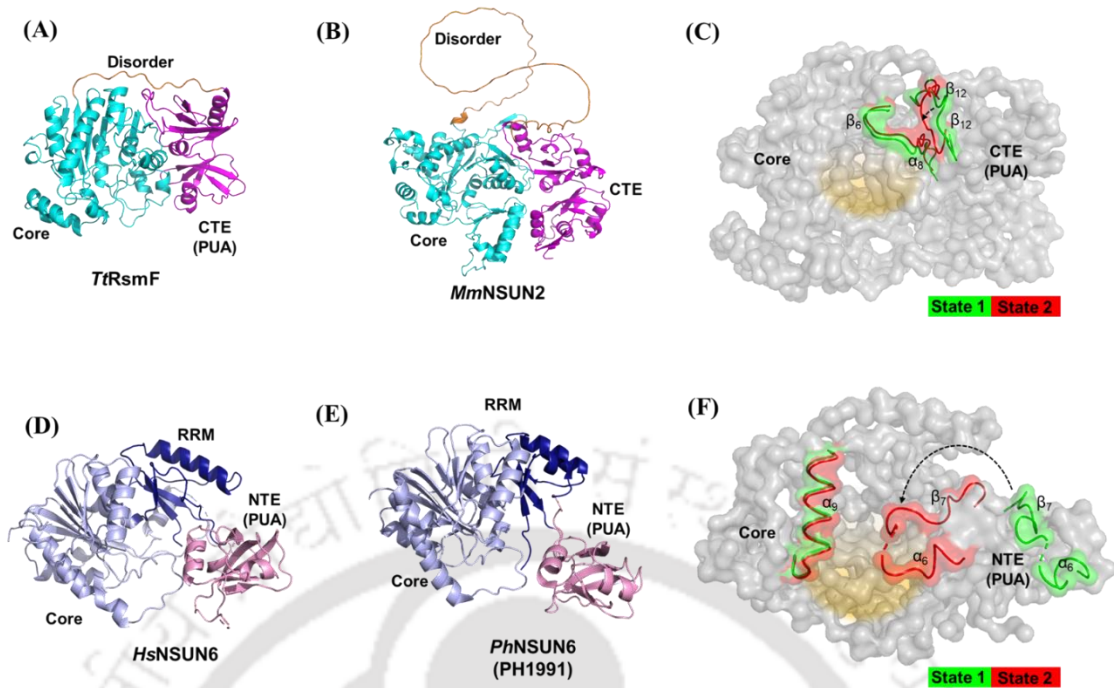


Figure A.3. Dynamic communities within proteins were identified, and their global motion was visualized using an elastic network model. This analysis focused specifically on RNA m^5C MTases that harbor the PUA domain in either the C-terminal extension (CTE) (A-C) or the N-terminal extension (NTE) (D-F).

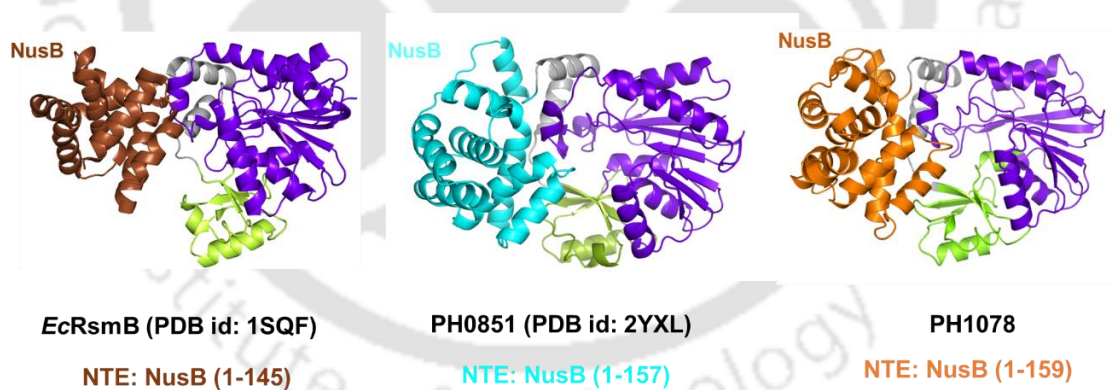


Figure A.4. Domain architecture of RsmB-like proteins. Comparison of the NusB domain organization in the proteins *EcRsmB*, PH0851, and PH1078 represented in brown, cyan, and orange, respectively.

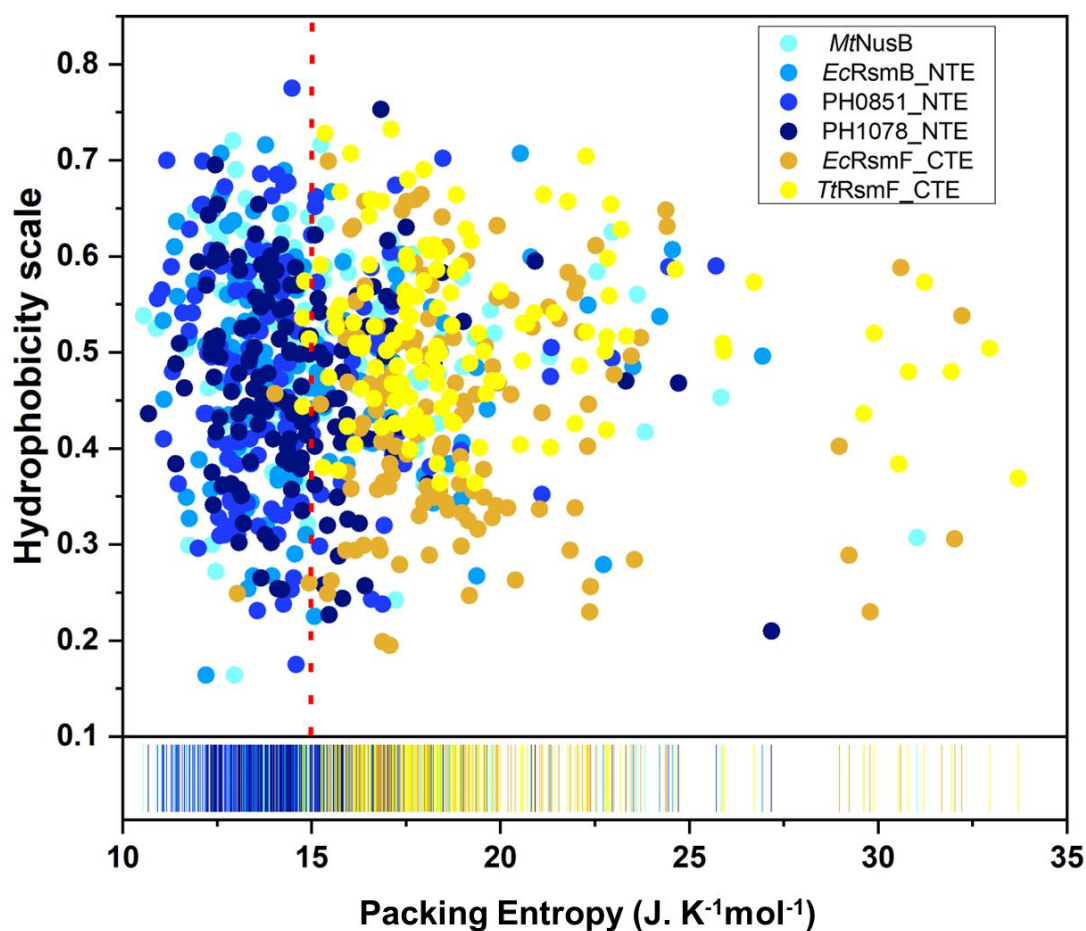


Figure A.5. The comparative distribution of packing entropies for the NTE (N-terminal extension) and CTE (C-terminal extension) regions of RNA m⁵C MTases, along with their respective hydrophobicity profiles. The data illustrates two discernible clusters representing the distinct domains (NTE and CTE) present within various RNA m⁵C MTases. Data points for protein *MtNusB* and the NTE regions of *EcRsmB*, *PH0851*, and *PH1078* are depicted in a blue gradient, while data points corresponding to the CTE regions of *EcRsmF* and *TtRsmF* are represented in a yellow gradient.

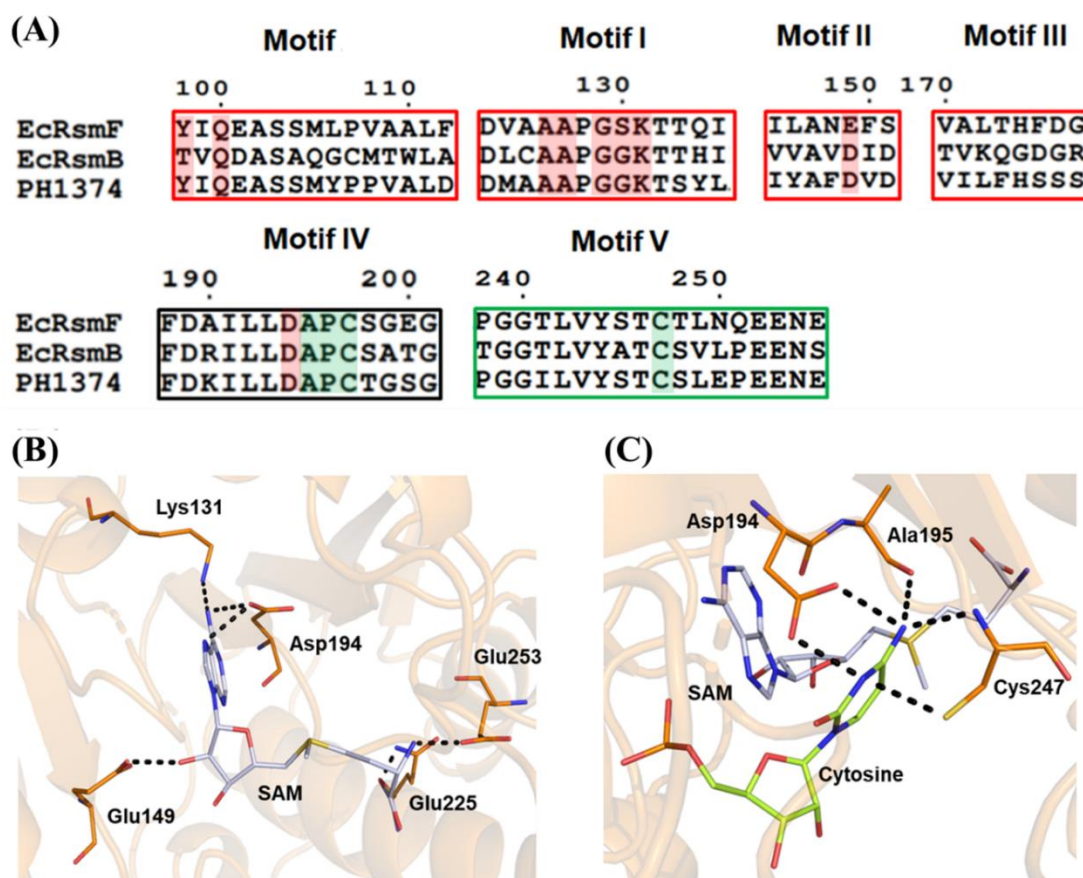


Figure A.6. A molecular docking investigation of *EcRsmF* with its cofactor SAM and target nucleotide. (A) Multiple sequence alignment of the *EcRsmB*, PH0851, and PH1078 proteins indicates that AdoMet and Cytosine binding residues are conserved, and the catalytic residues that interact with AdoMet and Cytosine are highlighted by red and green, respectively. (B-C) The crystal structure of *EcRsmF* (PDB id: 2FRX) docked with (B) AdoMet and (C) cytosine, revealing the arrangement and key interacting residues at the active site.

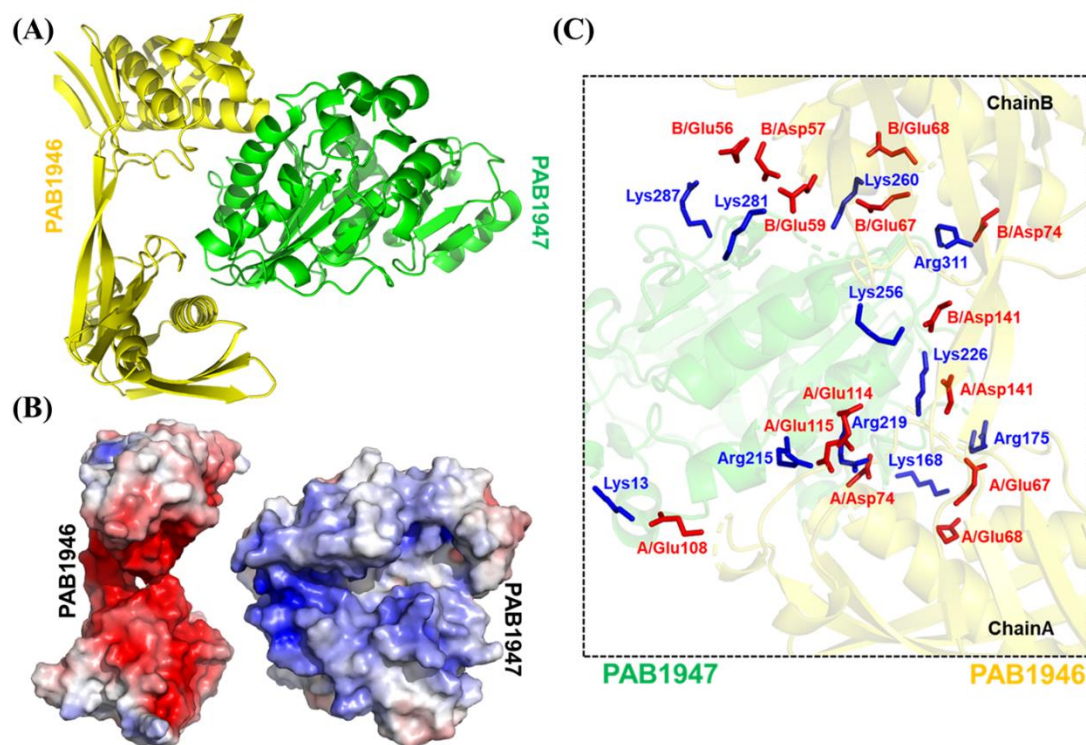


Figure A.7. Protein-protein docking between the proteins PAB1946 and PAB1947.

(A) Cartoon representation of the stable complex formed by the proteins PAB1946 and PAB1947. The archease (PAB1946) and tRNA m⁵C MTase (PAB1947) are represented in yellow and green, respectively. (B) Electrostatic potential charge calculation for the protein PAB1946 and PAB1947 (modeled). The positively and negatively charged surfaces are shown in blue (+5 kT/e) and red (-5 kT/e), respectively. (C) Interaction of the amino acid residues at the interface of the complex PAB1946 and PAB1947. The negatively charged residues at the interface of the protein PAB1946 mask the positively charged residue of the protein PAB1947. The negatively and positively charged residues are represented as red and blue stick models, respectively.

Table A.1. List of RNA m⁵C MTases distributed in the archaeal genome.

S. No.	Group	Organism name	QC (%)	SI (%)	Accession No
PH0851-like					
1.	Euryarchaea	<i>Thermococcus barophilus</i>	99	70	WP_056933844.1
2.		<i>Ferroglobus placidus</i>	97	58	ADC65743.1
3.		<i>Palaeococcus ferrophilus</i>	98	60	WP_048151093.1
4.	TACK	<i>Pyrodictium occultum</i>	99	61	KSW12637.1
5.		<i>Nitrososphaerota archaeon</i>	-	-	-
6.		<i>Korarchaeota archaeon</i>	-	-	-
7.	DPANN	<i>Nanoarchaeota archaeon</i>	-	-	-
8.		<i>Woesearchaeota archaeon</i>	-	-	-
9.	Asgard	<i>Asgard group archaeon</i>	-	-	-
PH1078-like					
11.	Euryarchaea	<i>Thermococcus barophilus</i>	99	78	WP_056934363.1
12.		<i>Ferroglobus placidus</i>	97	46	WP_012966364.1
13.		<i>Palaeococcus ferrophilus</i>	100	69	WP_048152469.1
14.		<i>Pyrodictium occultum</i>	73	39	WP_058370676.1

Supplementary data to chapter 3

15.	TACK	<i>Nitrososphaerota archaeon</i>	98	30	MCL7387544.1
16.		<i>Korarchaeota archaeon</i>	-	-	-
17.		<i>Nanoarchaeota archaeon</i>	-	-	-
18.	DPANN	<i>Woesearchaeota archaeon</i>	-	-	-
19.	Asgard	<i>Asgard group archaeon</i>	85	31	NHK30900.1
PH1374-like					
21.		<i>Thermococcus barophilus</i>	98	80	WP_056934930.1
22.	Euryarchaea	<i>Ferroglobus placidus</i>	95	48	WP_012966495.1
23.		<i>Palaeococcus ferrophilus</i>	97	80	WP_048149541.1
24.		<i>Pyrodictium occultum</i>	-	-	-
25.		<i>Nitrososphaerota archaeon</i>	95	41	TLY07920.1
26.	TACK	<i>Korarchaeota archaeon</i>	-	-	-
27.		<i>Nanoarchaeota archaeon</i>	92	45	MBR9677756.1
28.	DPANN	<i>Woesearchaeota archaeon</i>	93	44	HDQ59727.1
29.	Asgard	<i>Asgard group archaeon</i>	95	44	NHJ86576.1
PH1537-like					
31.		<i>Thermococcus barophilus</i>	-	-	-
32.	Euryarchaea	<i>Ferroglobus placidus</i>	94	41	WP_012966979.1
33.		<i>Palaeococcus ferrophilus</i>	99	68	WP_048148920.1

Supplementary data to chapter 3

34.	TACK	<i>Pyrodictium occultum</i>	-	-	-
35.		<i>Nitrososphaerota archaeon</i>	88	46	MCD6260914.1
36.		<i>Korarchaeota archaeon</i>	96	43	RLG38135.1
37.	DPANN	<i>Woesearchaeota archaeon</i>	95	53	MBI2673375.1
38.		<i>Nanoarchaeota archaeon</i>	95	52	MCD6403953.1
39.	Asgard	<i>Asgard group archaeon</i>	-	-	-
PH1991-like					
41.	Euryarchaea	<i>Thermococcus barophilus</i>	98	83	WP_056934465.1
42.		<i>Ferroglobus placidus</i>	-	-	-
43.		<i>Palaeococcus ferrophilus</i>	98	76	WP_048150199.1
44.	TACK	<i>Pyrodictium occultum</i>	69	37	WP_058370188.1
45.		<i>Nitrososphaerota archaeon</i>	-	-	-
46.		<i>Korarchaeota archaeon</i>	92	42	RLG56080.1
47.	DPANN	<i>Nanoarchaeota archaeon</i>	-	-	-
48.		<i>Woesearchaeota archaeon</i>	-	-	-
49.	Asgard	<i>Asgard group archaeon</i>	97	40	NHJ47663.1

Abbreviations: (-): Absence of RNA m⁵C MTases; QC: Query coverage; SI: Sequence identity

Table A.2. Pair-wise sequence alignment among the core domain and full-length of the archaeal, bacterial, and eukaryal RNA m⁵C MTases. The percentage values of the query coverage (sequence identity) of the core domain and full-length protein, respectively, are provided.

Protein (UniProt id)	PH085 1 (O58581)	PH1078 (O58805)	PH1374 (O50082)	PH1537 (O59206)	PH1991 (O57712)	RsmB (P36929)	RsmF (A8A13)	NSUN1 (P46087)	NSUN2 (Q08J23)	NSUN3 (Q13686)	NSUN4 (Q7Z6M)	NSUN5 (Q63ZY6)	NSUN6 (Q8TEA1)	NSUN7 (Q8NE18)
PH0851 (O58581)	100 (100), 100 (100)	98 (40), 69 (39)	85 (40), 85 (40)	87 (36), 95 (36)	99 (40), 65 (40)	98 (34), 65 (33)	98 (31), 65 (29)	97 (33), 60 (32)	57 (29), 47 (26)	59 (36), 38 (33)	55 (34), 34 (33)	69 (34), 58 (29)	66 (36), 38 (36)	16 (38), 9 (38)
PH1078 (O58805)	-	100 (100), 100 (100)	98 (36), 65 (33)	99 (34), 59 (34)	94 (44), 62 (44)	90 (31), 86 (26)	85 (33), 63 (32)	98 (32), 68 (31)	61 (28), 54 (25)	51 (37), 35 (30)	56 (30), 36 (29)	73 (29), 56 (29)	62 (37), 41 (35)	42 (23), 26 (24)
PH1374 (O50082)			100 (100), 100 (100)	88 (43), 88 (43)	85 (38), 95 (38)	86 (34), 99 (32)	87 (39), 87 (39)	89 (41), 89 (41)	46 (35), 76 (31)	48 (35), 73 (31)	60 (31), 63 (31)	62 (36), 86 (32)	57 (33), 64 (33)	32 (25), 33 (25)

Supplementary data to chapter 3

PH1537 (O59206)				100 (100), 100 (100)	73 (37), 73 (37)	70 (33), 70 (33)	90 (38), 91 (38)	87 (37), 87 (37)	56 (33), 71 (29)	48 (31), 87 (37)	47 (33), 49 (33)	64 (27), 86 (25)	56 (42), 66 (38)	24 (31), 29 (30)
PH1991 (O57712)					100 (100), 100 (100)	70 (31), 64 (31)	57 (37), 53 (37)	57 (36), 53 (36)	39 (30), 45 (28)	52 (31), 52 (30)	37 (31), 39 (29)	52 (31), 61 (30)	91 (31), 93 (30)	12 (34), 45 (21)
RsmB (P36929)						100 (100), 100 (100)	61 (36), 36 (36)	99 (34), 66 (35)	65 (31), 44 (32)	58 (32), 69 (24)	60 (31), 38 (30)	71 (30), 48 (29)	66 (32), 39 (32)	30 (33), 18 (33)
RsmF (A8A133))							100 (100), 100 (100)	93 (36), 60 (36)	53 (32), 47 (28)	46 (37), 33 (38)	45 (32), 41 (30)	65 (32), 59 (29)	49 (35), 30 (35)	28 (29), 36 (23)
NSUN1 (P46087)								100 (100), 100 (100)	49 (29), 42 (23)	51 (34), 20 (35)	51 (33), 20 (34)	68 (34), 34 (32)	65 (33), 28 (32)	26 (29), 9 (29)

Supplementary data to chapter 3

NSUN2 (Q08J23)									100 (100), 100 (100)	83 (29), 25 (29)	87 (30), 31 (27)	95 (31), 35 (28)	83 (33), 34 (31)	21 (36), 5 (36)
NSUN3 (Q13686)									100 (100), 100 (100)		73 (43), 92 (35)	71 (27), 62 (26)	72 (31), 48 (31)	28 (27), 17 (27)
NSUN4 (Q7Z6M 4)											100 (100), 100 (100)	82 (29), 49 (30)	83 (32), 44 (31)	24 (37), 19 (32)
NSUN5 (Q63ZY 6)												100 (100), 100 (100)	86 (32), 44 (32)	66 (33), 99 (25)
NSUN6 (Q8TEA 1)													100 (100), 100 (100)	10 (39), 5 (39)
NSUN7														100

Supplementary data to chapter 3

(Q8NE1 8)														(100), 100 (100)
--------------	--	--	--	--	--	--	--	--	--	--	--	--	--	---------------------

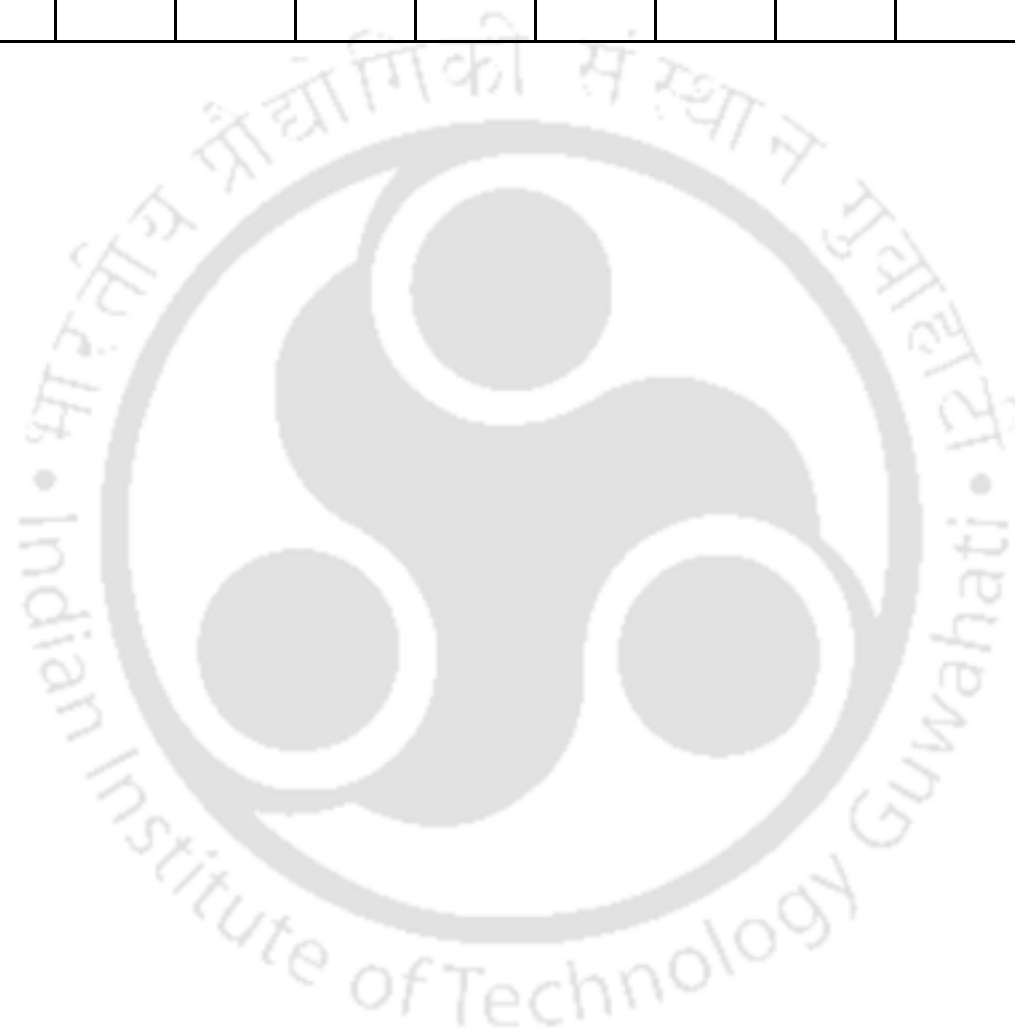


Table A.3. Details of the interacting partners of the protein *EcRsmB*, PH0851, PH1078, TM1083, PH1536, and TK0361, obtained in the protein-protein interaction network.

Reference protein name	Details of the interacting protein		
	Gene	UniProt id	Protein name
RsmB	<i>rlmC</i>	P75817	Ribosomal large subunit methyltransferase C
	<i>rsmA</i>	P06992	Ribosomal large subunit methyltransferase C
	<i>truB</i>	P60340	tRNA pseudouridine synthase B
	<i>tmcA</i>	P76562	tRNA (Met) cytidine acetyltransferase
	<i>trkA</i>	P0AGI8	The Potassium Transporter Trk
	<i>def</i>	P0A6K3	Peptide deformylase
	<i>tsaC</i>	P45748	Threonylcarbamoyl-AMP synthase
	<i>purL</i>	P15254	Phosphoribosylformylglycinamidine synthase subunit PurL
	<i>guaA</i>	P04079	GMP synthase
	<i>rpsA</i>	P0AG67	30S ribosomal protein S1
	<i>fmt</i>	P23882	Methionyl-tRNA formyltransferase
PH0851	<i>PH1259</i>	O58994	Ribosomal large subunit methyltransferase C
	<i>rsmA</i>	O59487	Ribosomal large subunit methyltransferase C
	<i>truB</i>	O59357	tRNA pseudouridine synthase B
	<i>tmcA</i>	O58518	tRNA (Met) cytidine acetyltransferase

Supplementary data to chapter 3

	<i>PH1137</i>	O58864	tRNA (uracil (54)-C (5))-methyltransferase
	<i>PH0281</i>	O58019	ABC2_membrane domain-containing protein
	<i>PH0850</i>	O58580	Uncharacterized protein
	<i>PH0645</i>	O58379	Hypothetical GTP-binding protein
	<i>purB</i>	O58582	Adenylosuccinate lyase
	<i>PH1320</i>	O59046	OBG-type G domain-containing protein
PH1078	<i>PH1259</i>	O58994	Ribosomal large subunit methyltransferase C
	<i>rsmA</i>	O59487	Ribosomal large subunit methyltransferase C
	<i>truB</i>	O59357	tRNA pseudouridine synthase B
	<i>tmcA</i>	O58518	tRNA (Met) cytidine acetyltransferase
	<i>PH1137</i>	O58864	tRNA (uracil (54)-C (5))-methyltransferase
	<i>PH0281</i>	O58019	ABC2_membrane domain-containing protein
	<i>PH1075</i>	O58802	Homoserine dehydrogenase
	<i>PH1076</i>	O58803	Cytochrome c oxidase subunit 1
	<i>PH1079</i>	O58806	Cytochrome c oxidase subunit 1
	<i>rpl5</i>	O59431	60S ribosomal protein L5
TM1083	<i>gyrA</i>	O33926	DNA gyrase subunit A
	<i>metG</i>	O33925	Methionine--tRNA ligase
	<i>panC</i>	Q9X0G6	Pantothenate synthetase

Supplementary data to chapter 3

	<i>lexA</i>	O33927	LexA repressor
	<i>TM1081</i>	Q9X0H0	Putative anti-sigma factor antagonist TM_1081
	<i>prmA</i>	P0A6K3	Ribosomal protein L11 methyltransferase
PH1536	<i>rtcB</i>	O59245	RNA-splicing ligase RtcB
	<i>PH1537</i>	O59206	m ⁵ C methyltransferase
	<i>PH1240</i>	O58943	Methyltranf_PUA domain-containing protein
	<i>PH1535</i>	O59204	Uncharacterized protein
	<i>rtcA</i>	O59198	RNA 3'-terminal phosphate cyclase
	<i>PH1047</i>	O58750	Uncharacterized protein
	<i>lig</i>	O59288	DNA ligase
TK0361	<i>TK0359</i>	Q5JCZ0	Methyltranf_PUA domain-containing protein
	<i>TK0360</i>	Q5JCY9	tRNA (uracil (54)-C (5))- methyltransferase
	<i>TK0748</i>	Q5JHD2	Biotin biosynthesis protein BioY
	<i>egsA</i>	O59144	Glycerol-1-phosphate dehydrogenase
	<i>adkA</i>	O59443	Adenylate kinase
	<i>panB</i>	P9WIL7	3-methyl-2-oxobutanoate hydroxymethyltransferase
	<i>TK0362</i>	Q5JCY7	Uncharacterized protein
	<i>TK0364</i>	Q5JCY5	Dolichol-phosphate mannosyltransferase

Table A.4. Details of the molecular docking results of the proteins *EcRsmB*, PH0851, PH1078, PH1374, and *EcRsmF* with their cognate ligand SAM and the substrate cytosine. EFBE, estimated free energy of binding; NHBOND, number of the hydrogen bond.

Target protein/ ORF id	Ligand	EFBE (kcal mol ⁻¹)	NHBOND	Interactions	
				Ligand atom	Protein atom(s)
<i>EcRsmB</i>	Cytosine	-6.04	6	OP1	Gln231 N ^{ε2}
				O3'	Asp232 N
				OP2	Lys260 N ^ζ
				N4	Asp322 O ^{δ1}
				N4	Cys389 S ^γ
				O2'	Asp419 O ^{δ2}
PH0851	SAM	-9.20	10	N	Ala267 O
				OXT	Gly271 N
				O	Lys273 N, & N ^ζ
				O3'	Asp291 O ^{δ1}
				O2'	Asp291 O ^{δ2}
				N3	Val292 N
				N6	Asp318 O ^{δ1}
				N1	Ala319 N
	Cytosine	-7.39	6	N	Asp337 O ^{δ2}
				O3'	Gln244 N ^{ε2}
				O2'	Ser248 O ^γ
				N4	Asp337 O ^{δ1}
				N4	Ala338 O
				N4	Cys340 S ^γ
PH1078	SAM	-7.30	10	N3	Cys390 S ^γ
				N	Ala269 O
				N	Ala270 O
				N	Ala271 O
				O	Gly273 N

Supplementary data to chapter 3

				OXT	Ser274 N
				OXT	Lys275 N
				O3'	Asp293 O ^{δ1}
				O2'	Asp293 O ^{δ2}
				N6	Asp320 O ^{δ1}
				N	Asp336 O ^{δ2}
	Cytosine	-5.20	4	N4	Asp336 O ^{δ1}
				N4	Ala337 O
				N3	Cys339 N
				N3	Cys389 S ^γ
PH1374	SAM	-7.45	5	O3'	Tyr101 O
				N	Glu104 O ^{ε2}
				O	Lys132 N ^ζ
				N6	Asp194 O
				OXT	Cys197 N
	Cytosine	-6.51	5	O2'	Lys41 O
				N4	Glu104 O ^{δ2}
				N4	Cys197 O
				O2	Ser305 O ^γ
				N3	Ser305 O ^γ
EcRsmF	SAM	-6.53	6	N6	Lys131 N ^ζ
				O2'	Glu149 O ^{ε2}
				N6	Asp194 O ^{ε2}
				N7	Asp194 O ^{ε2}
				O	Gln225 N ^{ε2}
				N	Glu253 O ^{ε1}
	Cytosine	-6.37	4	N4	Asp194 O ^{δ1}
				N3	Asp194 O ^{δ2}
				N4	Ala195 O
				N4	Cys247 N
				N3	Cys247 S ^γ

Table A.5. List of RNA m⁵C MTases used for the construction of the phylogenetic tree.

S. No.	Protein name	UniProt id	Organism name	Representation in phylogenetic tree
1.	NSUN1	P46087	<i>Homo sapiens</i>	HsNSUN1
2.		F1N3G2	<i>Bos taurus</i>	BotNSUN1
3.		Q922K7	<i>Mus musculus</i>	MmNSUN1
4.		Q9TYV5	<i>Caenorhabditis elegans</i>	CeNSUN1
5.	PH1374	O50082	<i>Pyrococcus horikoshii</i>	PH1374
6.	RsmF	Q5SII2	<i>Thermus thermophilus</i>	TtRsmF
7.		A3QDA2	<i>Shewanella loihica</i>	ShlRsmF
8.		A4WBJ4	<i>Enterobacter sp</i>	EntRsmF
9.		P76273	<i>Escherichia coli</i>	EcRsmF
10.	PH1537	O59206	<i>Pyrococcus horikoshii</i>	PH1537
11.	PAB194 7	Q9V106	<i>Pyrococcus abyssi</i>	PAB1947
12.	NSUN2	Q5ZLV4	<i>Gallus gallus</i>	GgNSUN2
13.		Q1HFZ0	<i>Mus musculus</i>	MmNSUN2
14.		Q08J23	<i>Homo sapiens</i>	HsNSUN2
15.	NSUN3	Q4KMK0	<i>Danio rerio</i>	DarNSUN3
16.		Q8CCT7	<i>Mus musculus</i>	MmNSUN3
17.		Q0P5D8	<i>Bos taurus</i>	BotNUS3
18.		Q9H649	<i>Homo sapiens</i>	HsNSUN3
19.	NSUN4	Q95XR2	<i>Caenorhabditis elegans</i>	CeNSUN4
20.		Q5M7E3	<i>Xenopus laevis</i>	XINSUN4
21.		Q66KI9	<i>Xenopus tropicalis</i>	XtNSUN4
22.		Q96CB9	<i>Homo sapiens</i>	HsNSUN4
23.		Q9CZ57	<i>Mus musculus</i>	MmNSUN4
24.		Q0V8R7	<i>Bos taurus</i>	BotNSUN4
25.	NSUN5	P53972	<i>Saccharomyces cerevisiae</i>	SacNSUN5

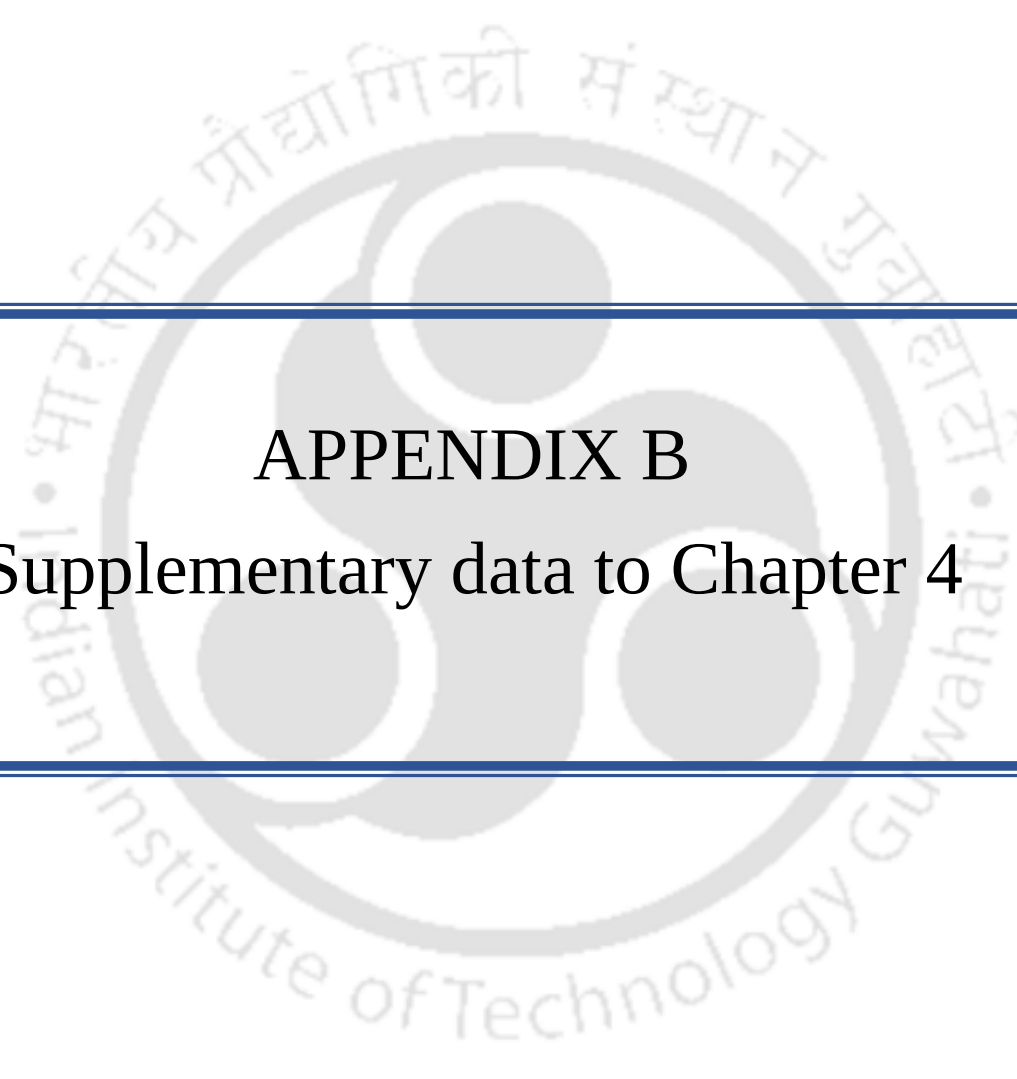
26.		Q9VDZ4	<i>Drosophila melanogaster</i>	DmNSUN5
27.		Q8K4F6	<i>Mus musculus</i>	MmNSUN5
28.		Q96P11	<i>Homo sapiens</i>	HsNSUN5
29.		Q14AW5	<i>Mus musculus</i>	MmNSUN7
30.	NSUN7	Q4R7Q1	<i>Macaca fascicularis</i>	MafNSUN7
31.		Q8NE18	<i>Homo sapiens</i>	HsNSUN7
32.	PH1078	O58805	<i>Pyrococcus horikoshii</i>	PH1078
33.	PH0851	O58581	<i>Pyrococcus horikoshii</i>	PH0851
34.		A0A164SH2 7	<i>Bacillus subtilis</i>	BsRsmB
35.	RsmB	Q9I7A9	<i>Pseudomonas aeruginosa</i>	PaRsmB
36.		A0A759NY K5	<i>Salmonella enterica</i>	SeRsmB
37.		P36929	<i>Escherichia coli</i>	EcRsmB
38.	PH1991	O57712	<i>Pyrococcus horikoshii</i>	PH1991
39.		Q7TS68	<i>Mus musculus</i>	MmNSUN6
40.	NSUN6	F6SWB4	<i>Macaca mulatta</i>	MamNSUN6
41.		Q8TEA1	<i>Homo sapiens</i>	HsNSUN6

Table A.6. Four clusters containing different RNA m⁵C MTase and their RNA substrates.

Cluster	Name	RNA substrate	Characterization type	Reference
I	NSUN1	rRNA	Biochemically	(Sharma et al., 2013)
	PH1374	rRNA	Bioinformatics	(This study)
II	RsmF	rRNA	Biophysically	(Hallberg et al., 2006)
	PH1537	tRNA	Bioinformatics	(This study)
	PAB1947	tRNA	Biochemically	(Auxilien et al., 2007)
	NSUN2	tRNA	Biochemically	(Brzezicha et al., 2006)

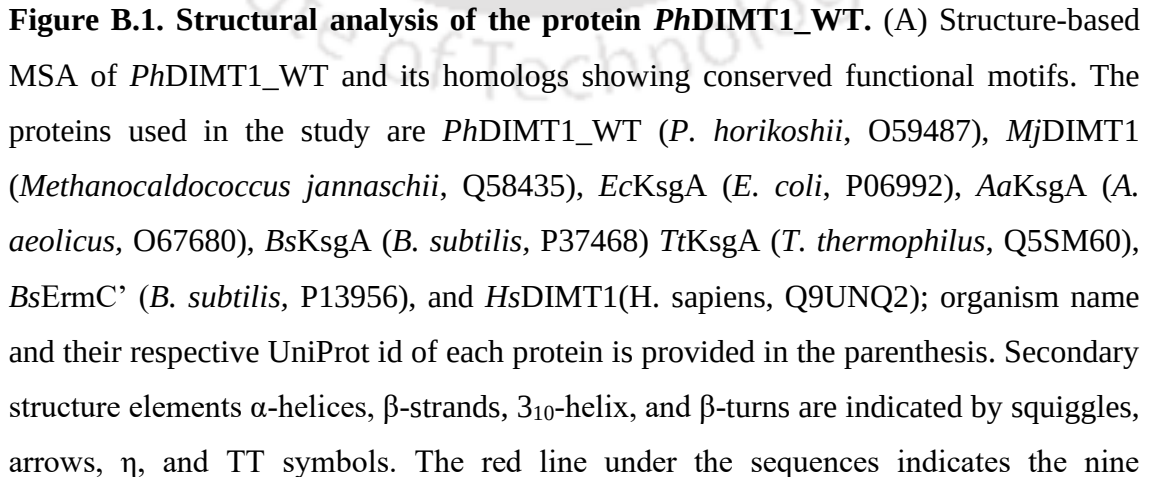
	NSUN3	tRNA	Biochemically	(Nakano et al., 2016)
	NSUN4	rRNA	Biophysically	(Spåhr et al., 2012)
III	RsmB	rRNA	Biophysically	(Foster et al., 2003)
	PH0851	rRNA	Biophysically	(Hikida et al., 2010)
	PH1078	rRNA	Bioinformatics	(This study)
	NSUN5	rRNA	Biochemically	(Schosserer et al., 2015)
	NSUN7	eRNA	Biochemically	(Aguilo et al., 2016)
IV	NSUN6	tRNA	Biophysically	(Liu et al., 2017)
	PH1991	tRNA	Biophysically	(Li et al., 2019)





APPENDIX B

Supplementary data to Chapter 4



conserved functional motifs common for the SAM-dependent N6,6-dimethyladenosine MTases. The conserved motifs X, I, II, III, IV, and VIII are boxed in yellow, green, magenta, purple, blue, and cyan, respectively. (B) The $2F_o - F_c$ electron density map (blue mesh) contoured at 1σ was used to highlight the active site residues within motif X of the *Ph*DIMT1_WT (PDB id: 8X3W). The conserved motifs in the above-mentioned color scheme are found in the NCD domain of the *Ph*DIMT1_WT structure.

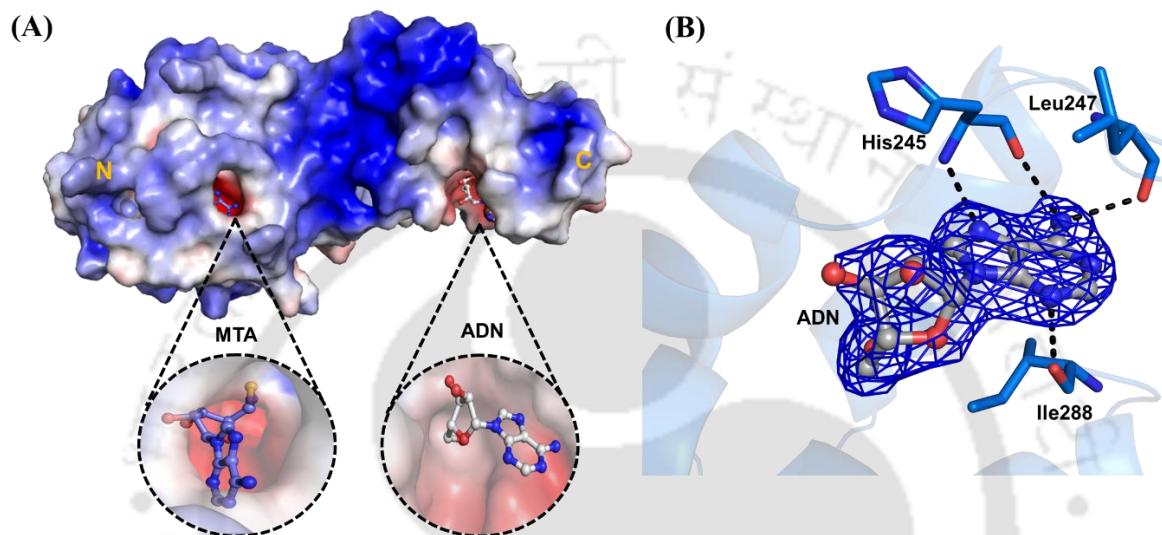


Figure B.2. Structure of the protein *Ph*DIMT1_WT complexed with MTA and ADN (PDB id: 8X41). (A) Electrostatic potential charge distribution of the protein *Ph*DIMT1_WT. The positively- and negatively charged surfaces are shown in blue (+5 kT/e) and red (-5 kT/e), respectively. The positively charged MTA outcompetes the relatively smaller molecule ADN for binding in the negatively charged cofactor-binding site in the NCD. Thus, the molecule ADN resides non-specifically in the other negatively-charge pocket in the CRBD. (B) The bound ADN in the CRBD is shown as a ball-and-stick model along with its $2F_o - F_c$ difference electron density map contoured at 1.0σ level. The amino acid residues interacting with ADN are shown as stick models.

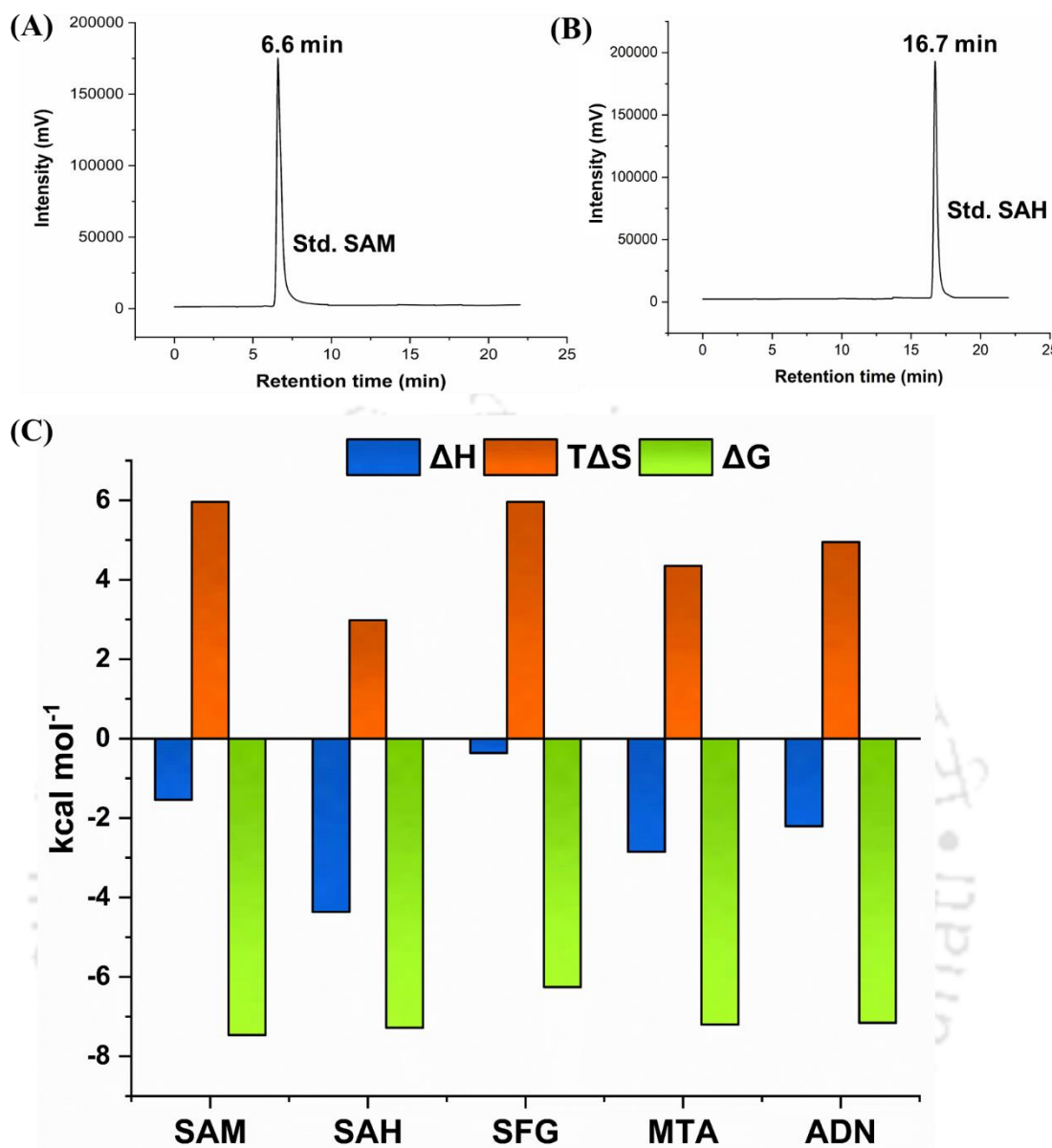


Figure B.3. Typical RP-HPLC curves of cofactor molecules (SAM and SAH) and comparison of their thermodynamic properties with other nucleoside analogs in the binding study of *PhDIMT1_WT*. (A-B) Standard RP-HPLC runs for the cofactors SAM and SAH. The uppermost trace shows the chromatogram of an equimolar concentration of the standards SAM and SAH detected peak with retention times of 6.6 and 16.7 min, respectively. (C) Thermodynamic parameters of cofactor SAM and its nucleoside analogs determined using ITC. Thermodynamic change in enthalpy (ΔH , blue), entropy (ΔS , orange), and free energy (ΔG , green) for the binding of SAM, SAH, SFG, MTA, and ADN to the protein *PhDIMT1_WT* are represented as histograms.

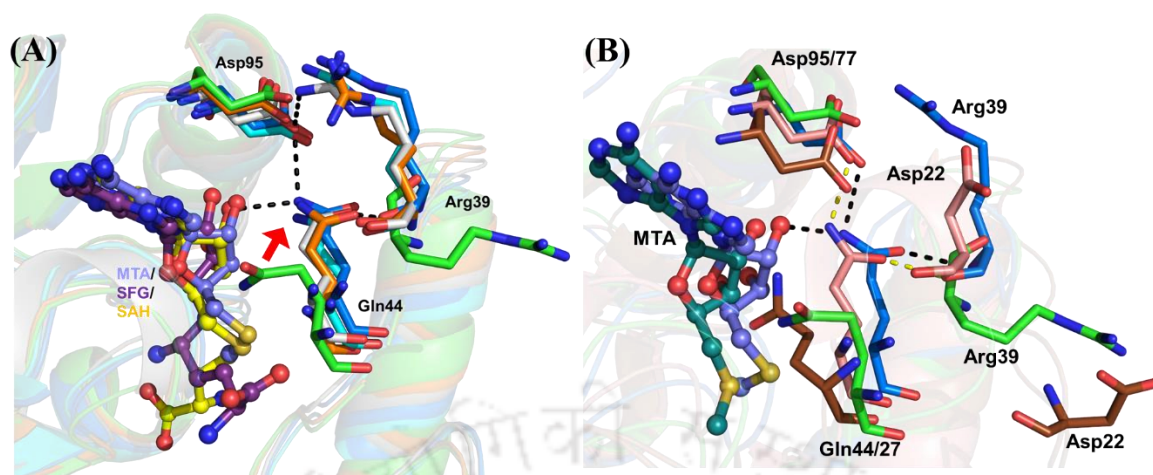


Figure B.4. Comparison of crystal structure and AlphaFold model of *PhDIMT1_WT*. (A) In the apo form of *PhDIMT1_WT*, the ribose sugar-binding pocket is partially obstructed by Gln44. However, in their holo forms, Gln44 moves sideways to establish a hydrogen bond with the O2' atom of the cofactor. The mobility of Gln44 is coordinated with that of two other residues, namely Arg39 (motif X) and Asp95 (motif II). The apo and AlphaFold model structures of *PhDIMT1_WT* are highlighted in green and grey, respectively, while the backbone of the protein *PhDIMT1* complex with MTA (purple), SAH (yellow), and SFG (violet) are represented with the blue, cyan, and orange color. (B) Superposition of the active-site pocket of *PhDIMT1_WT* with its homolog *TtKsgA* reveals that the movement of the residue Gln44/27 of *PhDIMT1_WT/TtKsgA* in holo forms is confined compared to the apo state. Active site residues are shown as sticks in apo (green/brown) and holo (blue/salmon) forms, while the cofactor MTA is shown as a ball-and-stick model in purple/teal.

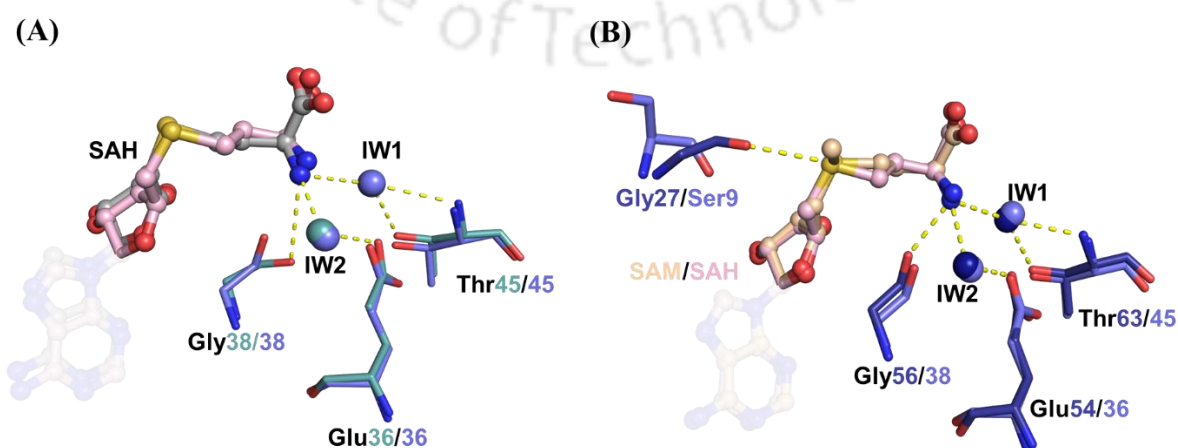


Figure B.5. Structural superposition of *Mj*DIMT1 (PDB id: 3GRR) and *Bs*KsgA (PDB id: 6IFT) with the reference protein *Bs*ErmC' (PDB id: 1QAN) to elucidate the involvement of IW1, IW2, during cofactor binding. The interacting active site residues in (A) *Mj*DIMT1 (teal) and (B) *Bs*KsgA (dark blue) are represented as sticks, while their cofactors are shown in grey and wheat ball-and-stick models, respectively. *Mj*DIMT1 (Teal), *Bs*KsgA (Dark Blue), and *Bs*ErmC' (Purple) IW molecules are shown by the spheres, while the yellow dashed line denotes hydrogen bonding.

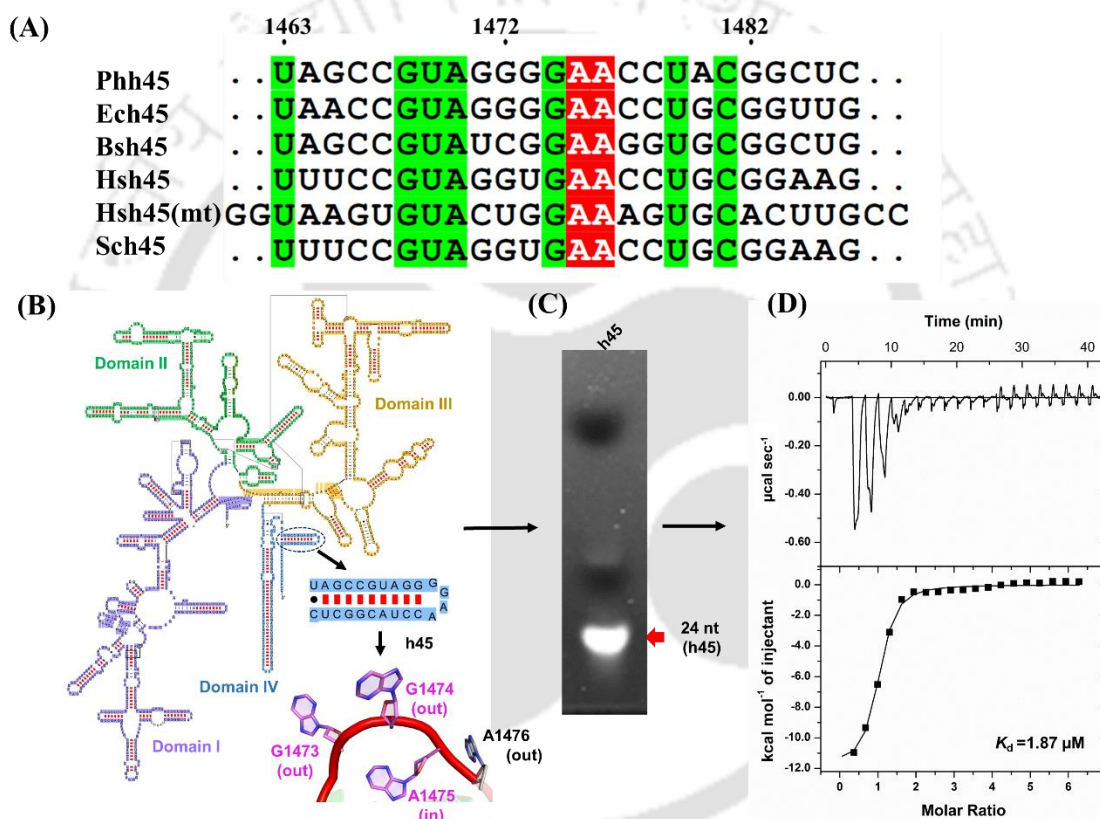


Figure B.6. Analysis of the conserved helix h45 of domain IV of SSU. (A) Multiple sequence alignment of h45 of *P. horikoshii* OT3 (Phh45) along with *E. coli* (Ech45), *B. subtilis* (Bsh45) of 16S rRNA, and *H. sapiens* (Hsh45), *Saccharomyces cerevisiae* (Sch45) of 18S rRNA as well as mitochondrial 12S rRNA of *H. sapiens* {Hsh45(mt)}. The dimethylation sites and conserved nucleotides of h45 are highlighted in red and green, respectively. (B) Secondary structure of 16S rRNA of *P. horikoshii*. The 24-nucleotide-long construct of h45 was modeled. (C) The purity of synthesized molecule h45 was determined by performing 9% native PAGE. (D) *Ph*DIMT1_WT shows high binding affinity towards the substrate h45 measured by ITC.

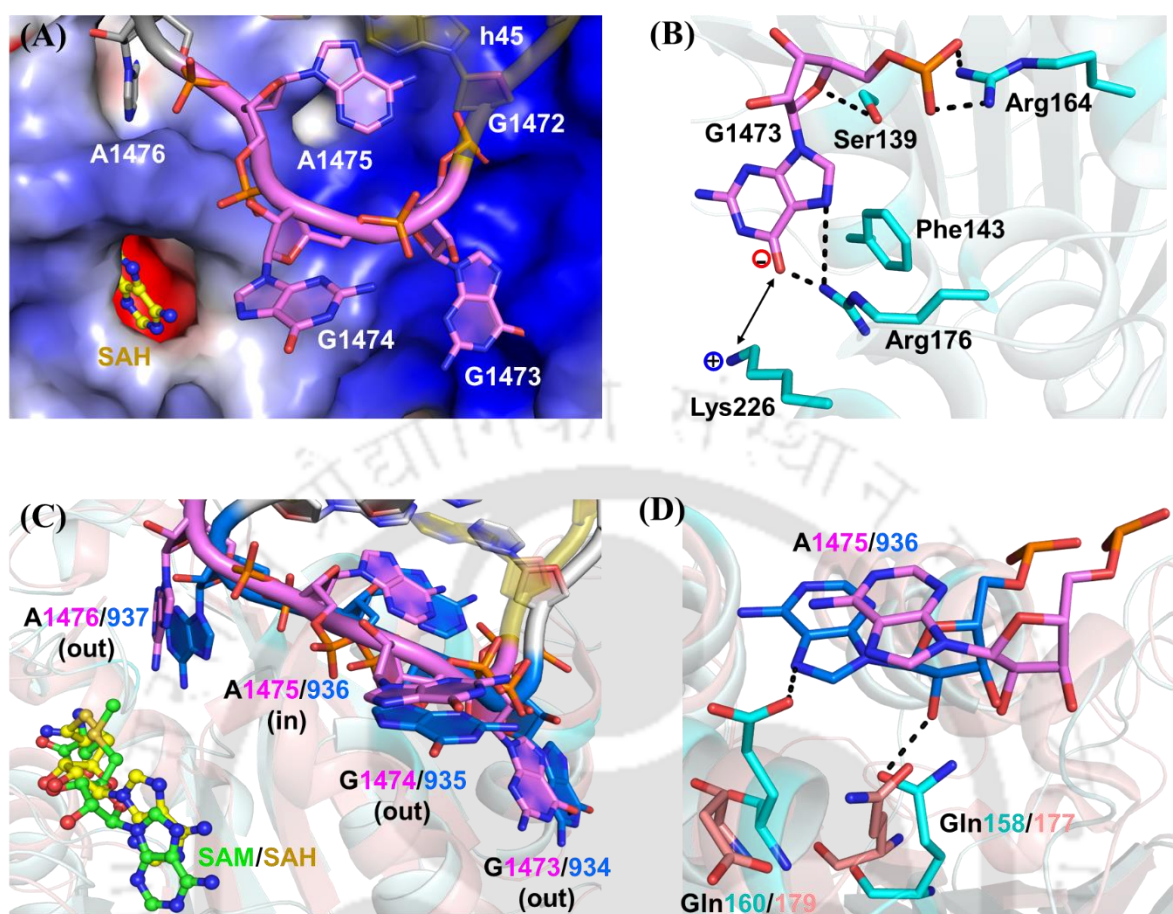


Figure B.7. The docked h45 to the *PhDIMT1_WT*•SAH structure. (A) The electrostatic potential charge distribution of *PhDIMT1_WT* indicates that the nucleobase G1473 is tethered within a small positively charged pocket. In contrast, the cofactor SAH is bound to a negatively charged protein pocket and is represented as a yellow ball-and-stick model. (B) The interactions between G1473 (magenta) and the amino acid residues (cyan) are shown as dashed lines. The two-way arrow depicts a salt-bridge-like interaction. (C) Overlay of *PhDIMT1_WT* (cyan), SAH (yellow), and h45 (pink) onto *HsTFB1M_WT* (salmon), SAM (green), and h45 (marine), illustrating the precise alignment of the GGAA tetraloop in our proposed mechanistic model. (D) Comparative interaction profile for the third nucleobase A1475 within the tetraloop of h45.

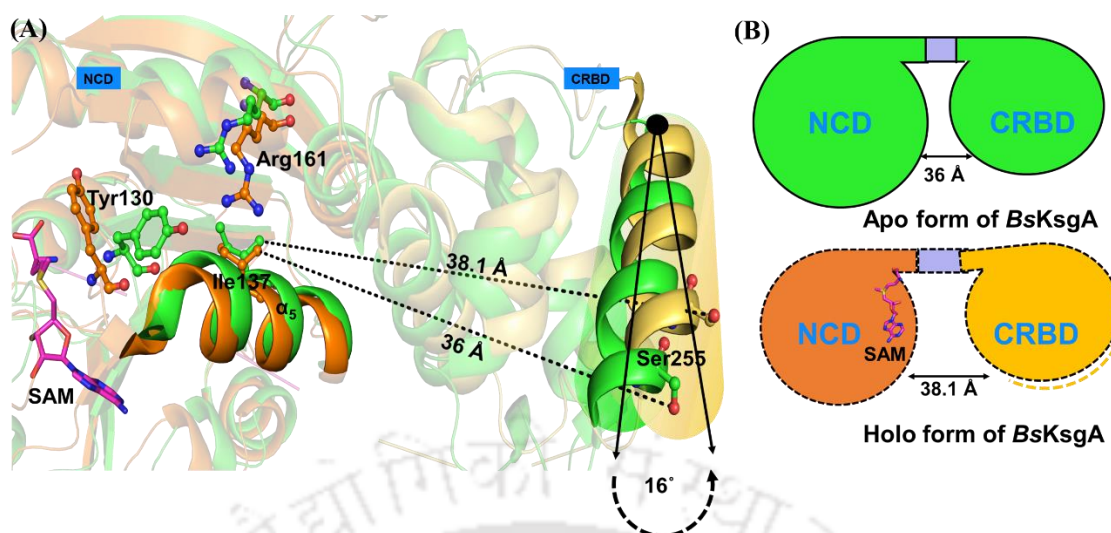


Figure B.8. The interdomain movement upon the cofactor binding to *BskSgA*. (A) Superposition of the apo (PDB id: 6IFS) and SAM-bound (PDB id: 6IFT) form of *BskSgA*. The overall (dis)similarities were depicted with an angular shift of 16° of the CRBD away from the active site. The conformational changes in the residues Tyr130 and Arg161 (orange ball-and-stick model) initiate the CRBD movement. The distances between Ile137 and Ser255 of the two forms are represented as black dashed lines. The NCD and the CRBD of the SAM-bound state are shown in orange and yellow, respectively, while those in the apo state are in green. (B) Schematic representation of the movement of the CRBD of the SAM-bound *BskSgA*. The conformational changes of the NCD (orange) and CRBD (yellow) upon SAM (magenta stick) binding are depicted with dotted lines. The CRBD movement away from the NCD is depicted as a yellow dotted arrow.

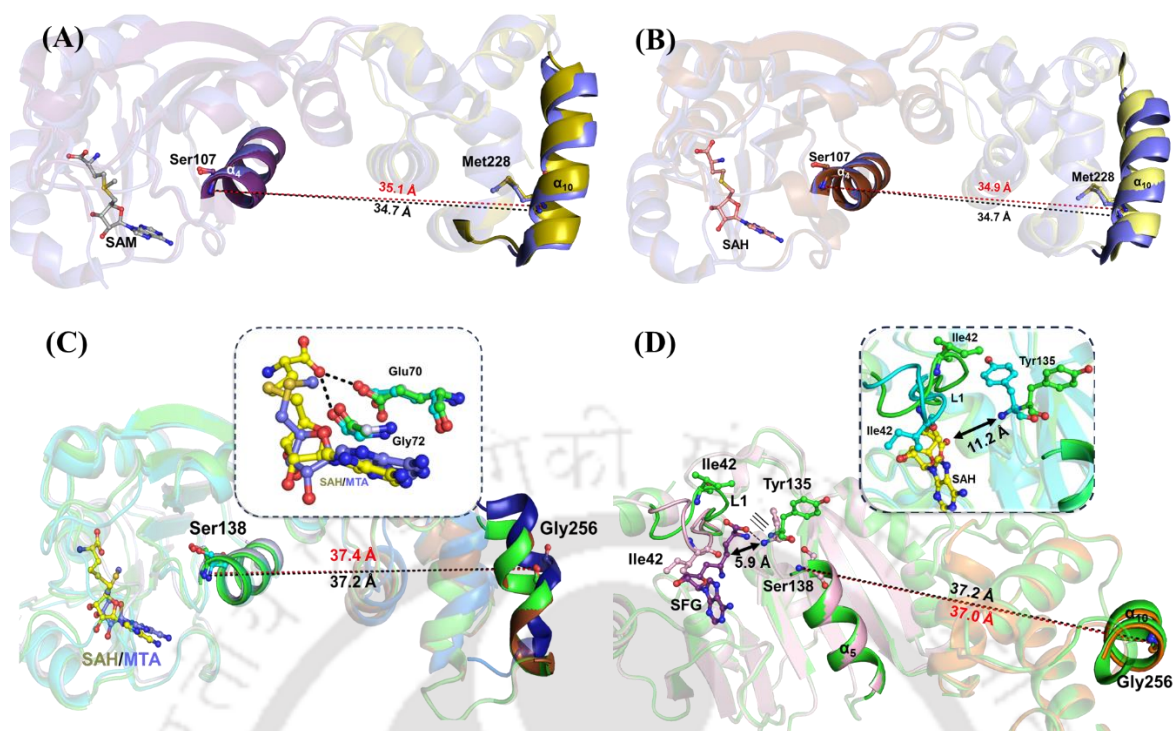


Figure B.9. The movement between domains is only observed in KsgA/DIMT1 structures obtained via co-crystallization and is limited to the binding of cognate cofactors (SAM/SAH). The comparison of the apo form (PDB id: 3FYC) with (A) the SAM-bound (PDB id: 3GRY) and (B) SAH-bound (PDB id: 3GRR) structures of *Mj*DIMT1 indicates that there was no significant inter-domain movement. The distances between Ser107 (α 4) of the NCD (purple/violet/brown) and Met228 (α 10) of the CRBD (purple olive/yellow) of apo/SAM/SAH-bound structures, respectively, are represented by black dashed lines. (C) MTA-bound *Ph*DIMT1_WT (PDB id: 8X44) shows no inter-domain movement, unlike the SAH-bound (PDB id: 8X47) form compared to its apo form. The distances between Ser138 (NCD) and Gly256 (CRBD) are marked as black dashed lines. The cofactors (SAH and MTA) and amino acid residues interacting with them are shown as ball-and-stick models (zoomed-in panel). (D) Superposition of the apo form (PDB id: 8X3W) with SFG-bound (PDB id: 8X46) and SAH-bound (PDB id: 8X47) forms of *Ph*DIMT1. Compared to SAH, the α -amino acid moiety in SFG undergoes unusual conformational changes. These changes constrain the active site pocket of *Ph*DIMT1_WT and are marked by double-headed arrows. The NCD and CRBD in the MTA, SFG, and SAH-bound forms are depicted in grey/brown, pink/orange, and cyan/blue, respectively, while its apo form is in green.

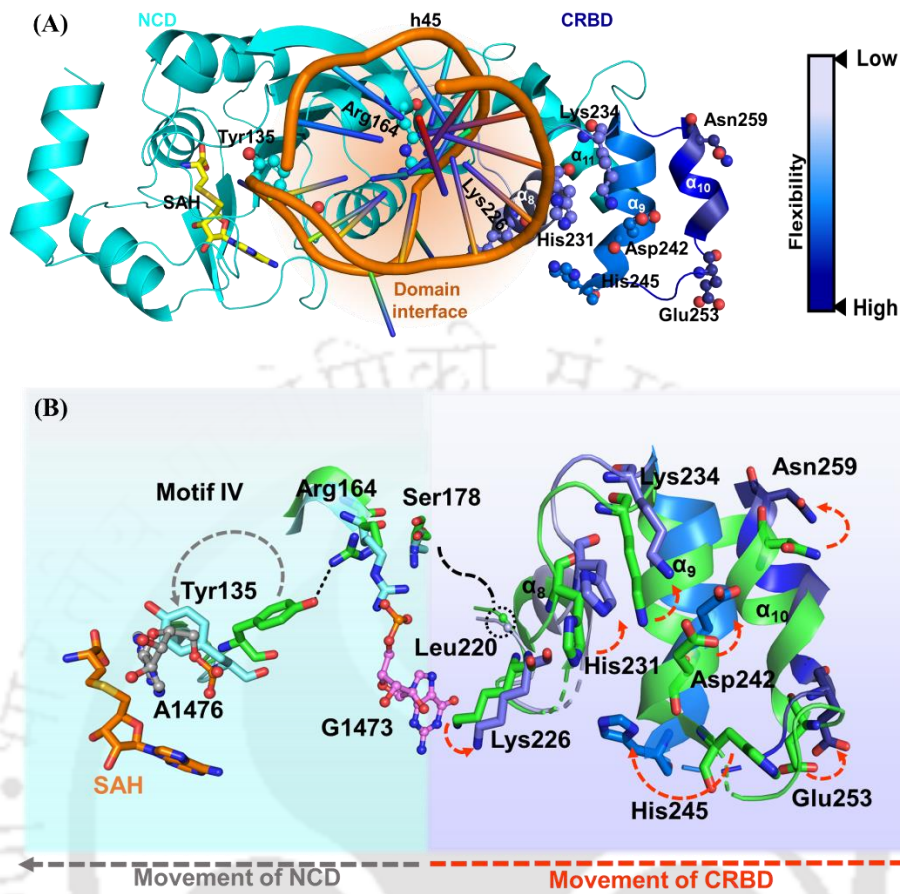


Figure B.10. The CRBD movement upon cofactor binding. (A) The ternary complex *PhDIMT1*_WT•SAH•h45 model hints that the rRNA substrate h45 binds in the expanded domain interface of the protein. (C) The restriction imposed on the residues Tyr135 and Arg164 upon the substrate h45 binding was caused by the nucleobases A1476 and G1473. For clarity, only the nucleobase G1473 has been shown.

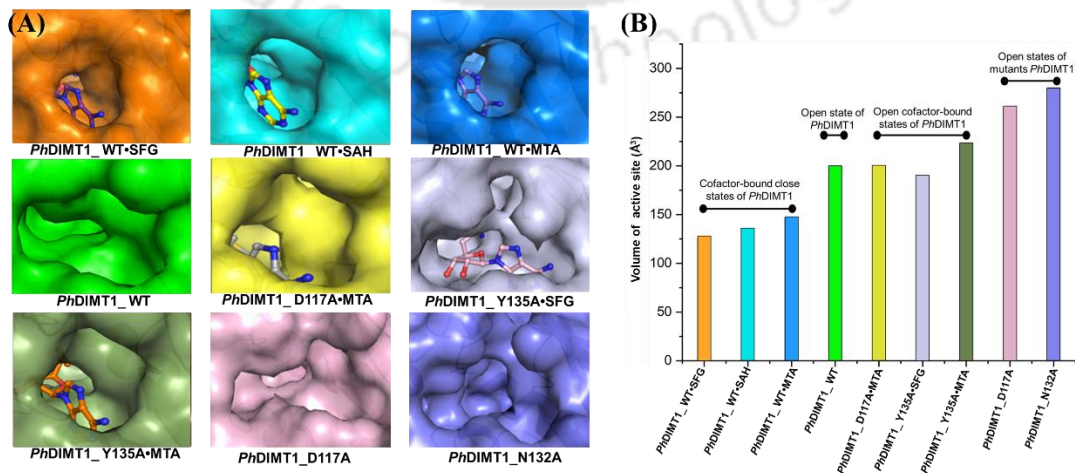


Figure B.11. Comparison of the active-site volume of the *Ph*DIMT1_WT and its mutants. (A) The active-site volumes of wild type and its mutant proteins in their apo and holo forms are shown as surface. (B) The active-site volumes of these proteins are represented quantitatively in the form of a histogram.

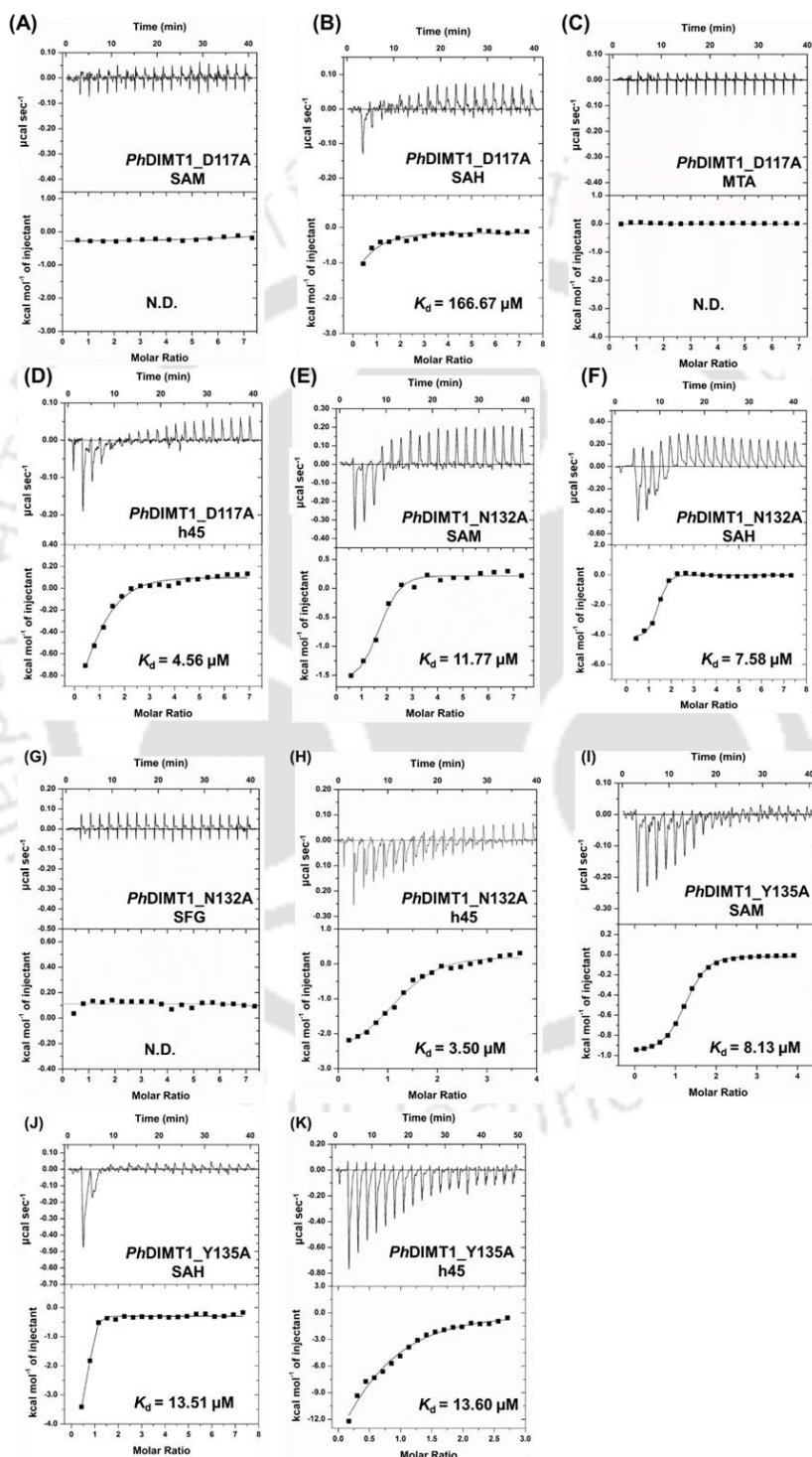


Figure B.12. Thermodynamic data of the *Ph*DIMT1 mutants. The thermodynamic data obtained from ITC for the mutants *Ph*DIMT1_D117A with (A) SAM, (B) SAH,

(C) MTA, and (D) h45, *PhDIMT1_N132A* with (E) SAM, (F) SAH, (G) SFG, and (H) h45 and *PhDIMT1_Y135A* with (I) SAM, (J) SAH, and (K) h45.

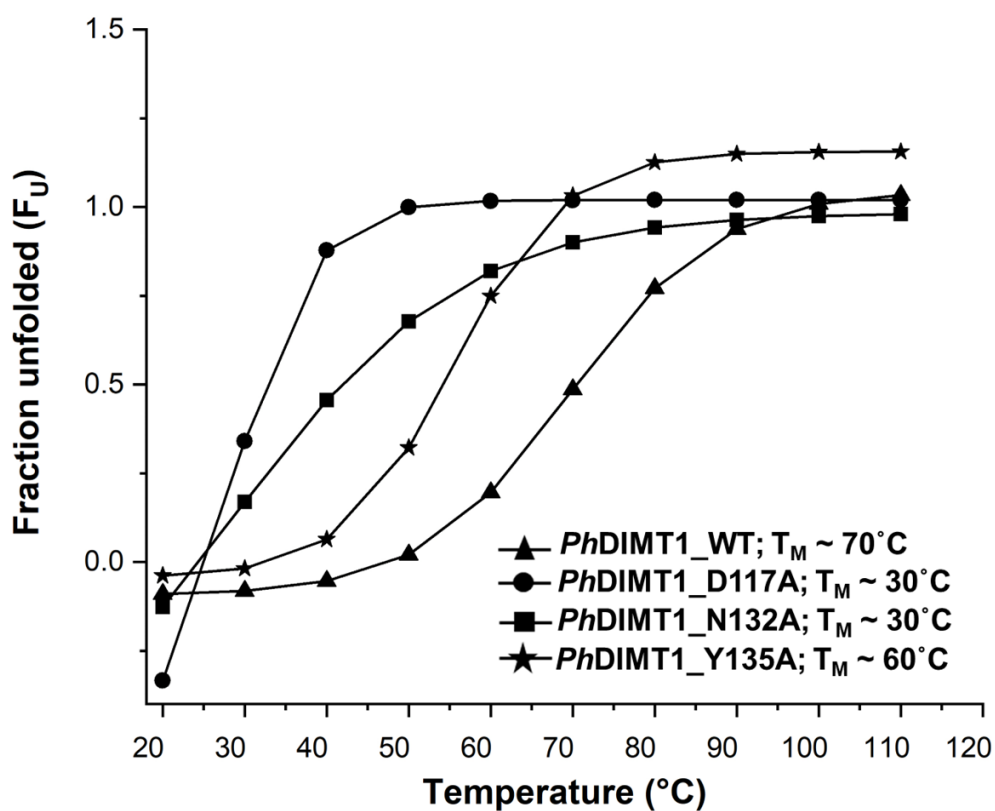


Figure. B.13. Thermodynamic data illustrating the thermal stability of *PhDIMT1* and its mutants. The unfolding curves were derived from the thermal ramping experiment, measuring CD ellipticity at 222 nm.

Table B.1. Thermodynamic parameters of cofactors (SAM, SAH, MTA, ADN, SFG, and rRNA (h45)) binding with *PhDIMT1*_WT and mutants (*PhDIMT1*_D117A, *PhDIMT1*_N132A, and *PhDIMT1*_Y135A). N.D.: not detected. † Due to the high value of the error of the fit of titration, the stoichiometry ratio was fixed as n=1 during the one-site-binding fitting model.

Protein name	Protein conc. (μM)	Cofactor/h45	Cofactor/h45 conc. (mM)	Stoichiometry ratio (n)	Association constant x 10 ⁵ (k _a , M ⁻¹)	Dissociation constant (k _D , μM)	ΔH	TΔS	ΔG
							(kcal mol ⁻¹)		
<i>PhDIMT1</i> _WT	80	SAM	4	1.38±0.02	3.55±0.84	2.81	-1.54±0.04	5.96	-7.46
		SAH		0.74±0.01	5.98±1.36	1.67	-4.36±0.12	2.98	-7.28
		MTA		0.63±0.03	1.89±0.50	5.30	-2.85±0.17	4.35	-7.20
		ADN		0.71±0.05	1.73±0.46	5.78	-2.21±0.24	4.95	-7.16
	280	SFG	14	0.72±0.07	1.40±1.37	7.14	-0.40±0.05	6.06	-6.46
	60	h45	1.5	0.93±0.01	5.34±0.79	1.87	-11.89±0.30	-4.08	-7.81
<i>PhDIMT1</i> _D117A	80	SAM	4	N.D.					
		MTA							
		†SAH		1.00±0.00	0.06±0.02	166.67	-2.72±0.30	2.26	-4.98
		†h45		1.00±0.00	2.19±2.00	4.56	-0.74±0.09	5.98	-6.72

Supplementary data to chapter 4

<i>PhDIMIT1</i> _N132A	80	[†] SAM	4	1.00±0.00	0.85±0.81	11.77	-1.62±0.29	5.09	-6.71
		[†] SAH		1.00±0.00	1.32±0.56	7.58	-5.43±0.57	1.43	-6.86
		SFG		N.D.					
	60	h45	1.5	1.12±0.05	2.85±1.13	3.50	-2.25±0.14	5.19	-7.44
<i>PhDIMIT1</i> _Y135A	80	SAM	4	1.19±0.03	1.23±0.33	8.13	-0.98±0.03	5.96	-6.94
		[†] SAH		1.00±0.00	0.74±0.58	13.51	-3.38±0.54	3.00	-6.38
	60	[†] h45	1.5	1.00±0.00	0.73±0.19	13.60	-12.53±0.00	-4.55	-7.98

Table B.2. List of RNA dimethylase used for the construction of the phylogenetic tree.

S. No.	Protein name	UniProt id	Domain of life	Organism name	Representation in phylogenetic tree
1.	KsgA	P06992	Bacteria	<i>Escherichia coli</i>	<i>EcKsgA</i>
2.		P37468		<i>Bacillus subtilis</i>	<i>BsKsgA</i>
3.		Q5SM60		<i>Thermus thermophilus</i>	<i>TtKsgA</i>
4.		B5BL27		<i>Salmonella paratyphi A</i>	<i>SpKsgA</i>
5.		A0KAD1		<i>Burkholderia cenocepacia</i>	<i>BcKsgA</i>
6.		P66663		<i>Staphylococcus aureus</i>	<i>SaKsgA</i>
7.		B8ZNY9		<i>Streptococcus pneumoniae</i>	<i>SpnKsgA</i>
8.		O67680		<i>Aquifex aeolicus</i>	<i>AaKsgA</i>
9.		B4F2I2		<i>Proteus mirabilis</i>	<i>PmKsgA</i>
10.		Q4QMZ9		<i>Haemophilus influenzae</i>	<i>HiKsgA</i>
11.	DIMIT1	Q9UNQ2	Eukaryotes	<i>Homo sapiens</i>	<i>HsDIMIT1</i>
12.		Q9D0D4		<i>Mus musculus</i>	<i>MmDIMIT1</i>
13.		Q2KHT8		<i>Bos taurus</i>	<i>BtDIMIT1</i>
14.		P41819		<i>Saccharomyces cerevisiae</i>	<i>ScDIMIT1</i>
15.		Q9VAQ5		<i>Drosophila melanogaster</i>	<i>DmDIMIT1</i>
16.		O22268		<i>Arabidopsis thaliana</i>	<i>AtDIMIT1</i>
17.		Q09522		<i>Caenorhabditis elegans</i>	<i>CeDIMIT1</i>
18.		Q54QK7		<i>Dictyostelium discoideum</i>	<i>DiDIMIT1</i>
19.		P78697		<i>Kluyveromyces lactis</i>	<i>KlDIMIT1</i>
20.	DIMIT1	O59487	Archaea	<i>Pyrococcus horikoshii</i>	<i>PhDIMIT1</i>
21.		Q58435		<i>Methanocaldococcus jannaschii</i>	<i>MjDIMIT1</i>
22.		C5A594		<i>Thermococcus gammatolerans</i>	<i>TgDIMIT1</i>
23.		Q3IPM0		<i>Natronomonas pharaonis</i>	<i>NpDIMIT1</i>
24.		Q5V588		<i>Haloarcula marismortui</i>	<i>HmDIMIT1</i>
25.		A1RRK0		<i>Pyrobaculum islandicum</i>	<i>PiDIMIT1</i>
26.		O28491		<i>Archaeoglobus fulgidus</i>	<i>AfDIMIT1</i>
27.		Q8PU18		<i>Methanosarcina mazei</i>	<i>MmaDIMIT1</i>
28.		Q8TWU7		<i>Methanopyrus kandleri</i>	<i>MkDIMIT1</i>
29.	Erm	P13956	Bacteria	<i>Bacillus subtilis</i>	<i>BsErmC'</i>
30.		P13957		<i>Staphylococcus aureus</i>	<i>SaErmC</i>
31.		P20173		<i>Enterococcus faecalis</i>	<i>EfErmB</i>
32.		H2DJ10		<i>Clostridioides difficile</i>	<i>CdErmB</i>
33.		A0A060QNN3		<i>Streptococcus pneumoniae</i>	<i>SpnErmB</i>

34.		A0A6M3 THI1		<i>Lactococcus lactis</i>	<i>LlaErmB</i>
35.		A0A291 L8D6		<i>Haemophilus influenzae</i>	<i>HiErmB</i>
36.		K4NRN0		<i>Listeria monocytogenes</i>	<i>LmErmB</i>
37.		A0A6M3 THJ0		<i>Aerococcus viridans</i>	<i>AvErmB</i>
38.		A0A0C4 M7A4		<i>Campylobacter jejuni</i>	<i>CjErmB</i>

Table B.3. List of interaction(s) among the various cofactors with different forms of the protein *PhD1MT1*. NHBOND: number of hydrogen bonds.

Structure (PDB id)	Cofactor	NHBOND	Interactions		
			Cofactor atom	Protein atom	Distance (Å)
<i>PhD1MT1_</i> <i>WT•MTA•A</i> <i>DN</i> (8X41)	MTA	6	N6	Asp117 O ^{δ1}	2.8
			N1	Ala118 N	3.1
			O2'	Gln44 N ^{ε2}	2.7
			O2'	Glu93 O ^{ε2}	2.5
			O3'	Glu93 O ^{ε1}	2.7
			S ^δ	His45 O	3.3
	ADN	4	N6	Leu247 O	3.0
			N6	His245 O	2.9
			N7	His245 N	3.0
			N3	Ile288 O	3.3
<i>PhD1MT1_</i> <i>WT•MTA</i> <i>FormI</i> (8X44)	MTA	4	N6	Asp117 O ^{δ1}	3.1
			N1	Ala118 N	3.0
			O3'	Glu93 O ^{ε1}	2.6
			O2'	Glu93 O ^{ε2}	2.3
<i>PhD1MT1_</i> <i>WT•MTA</i> <i>FormII</i> (8X45)	MTA	4	N6	Asp117 O ^{δ1}	2.9
			N1	Ala118 N	3.0
			O3'	Glu93 O ^{ε1}	3.0
			O2'	Glu93 O ^{ε2}	2.3
<i>PhD1MT1_</i>	SFG	10	N6	Asp117 O ^{δ1}	2.8

Supplementary data to chapter 4

WT•SFG (8X46)			N1	Ala118 N	3.0
			O3'	Glu93 O ^{ε1}	2.6
			O2'	Glu93 O ^{ε2}	2.3
			O2'	Gln44 N ^{ε2}	3.5
			N ^ε	Asn132 O	2.8
			N ^ε	Gly72 O	2.5
			N	Asn132 O ^{δ1}	3.1
			O	His45 N	3.0
			OXT	Tyr135 N	2.7
<i>Ph</i> DIMT1_ WT•SAH (8X47)	SAH	7	N6	Asp117 O ^{δ1}	2.8
			N1	Ala118 N	3.2
			O3'	Glu93 O ^{ε1}	2.7
			O2'	Glu93 O ^{ε2}	2.4
			O2'	Gln44 N ^{ε2}	2.9
			OXT	Glu70 O ^{ε2}	3.1
			OXT	Gly72 O	3.2
<i>Ph</i> DIMT1_ D117A•MT A (8X4I)	MTA	4	N1	Ala118 N	3.0
			O3'	Glu93 O ^{ε1}	2.7
			O2'	Glu93 O ^{ε2}	2.2
			S5'	His45 O	3.1
<i>Ph</i> DIMT1_ Y135A•MT A (8X4O)	MTA	5	N6	Asp117 O ^{δ1}	2.7
			N1	Ala118 N	3.3
			O3'	Glu93 O ^{ε1}	2.7
			O2'	Glu93 O ^{ε2}	2.4
			S5'	His45 O	2.9
<i>Ph</i> DIMT1_ Y135A•MT A (8X4P)	SFG	7	N6	Asp117 O ^{δ1}	3.5
			N1	Ala118 N	3.0
			O3'	Glu93 O ^{ε1}	2.8
			O2'	Glu93 O ^{ε2}	2.0
			N ^ε	Asn132 O	2.7
			N ^ε	Gly72 O	2.9
			N	Asn132 O ^{δ1}	3.1

Table B.4. List of three invariant water molecules (IW1, IW2, and IW3) in KsgA/DIMT1 protein corresponding to the reference protein *BsErmC*'.

Protein	PDB (chain) id	IW1 (B-factor in Å ²)	IW2 (B-factor in Å ²)	IW3 (B-factor in Å ²)
<i>BsErmC</i> '	1QAM (A)	910 (24.30)	1011 (25.44)	905 (30.46)
	1QAM (A)	935 (42.65)	950 (31.19)	949 (31.06)
<i>MjDIMT1</i>	3FYC (A)	328 (31.29)	285 (25.66)	Absent
	3GRR (A)	280 (37.18)	281 (34.33)	Absent
<i>BsKsgA</i>	6IFS (A)	305 (16.69)	Absent	Absent
	6IFT (A)	411 (17.11)	483 (11.64)	Absent
<i>PhDIMT1</i>	8X3W (A)	11 (19.16)	6 (20.95)	30 (37.72)
	8X41 (A)	20 (16.24)	6 (15.88)	100 (23.46)
	8X46 (A)	46 (27.25)	39 (36.12)	Displaced
	8X47 (A)	Displaced	Displaced	8 (44.98)
	8X4G (A)	Absent	Absent	8 (10.96)
	8X4L (A)	Absent	12 (33.10)	20 (41.20)
	8X4I (A)	Absent	Absent	Absent
	8X4O (A)	Absent	Absent	Absent
	8X4P (A)	5 (34.73)	Absent	Absent
<i>HsDIMT1</i>	6W6C (B)	442 (27.90)	460 (45.87)	408 (39.25)
	1ZQ9 (B)	4009 (23.34)	4078 (23.34)	Displaced
<i>TtKsgA</i>	3FUT (A)	350 (9.41)	365 (12.76)	664 (14.18)
	3FUX (A)	295 (19.87)	289 (19.52)	Displaced
<i>AaKsgA</i>	3FTC (A)	389 (46.36)	258 (28.17)	415 (43.98)
	3FTF (A)	Displaced	632 (20.00)	Displaced
<i>EcKsgA</i>	1QYR (B)	271 (28.08)	280 (37.91)	Absent

Table B.5. Summary of the average movements observed in the CRBD residues (based on the C_{α}) of SAH-bound *PhDIMT1*_WT. N.D.: No electron density.

Residue No.	Position	RMSD (Å)	Average movement (Å)
221	α_8	1.16	1.26
222		1.27	
223		1.29	
224		1.21	
225		1.35	
226		1.42	
227		1.45	
228		1.03	
229		1.10	
230		1.41	
231	$\alpha_8 - \alpha_9$ (L17)	1.32	1.37
232		1.02	

Supplementary data to chapter 4

233		1.19	
234		1.98	
235		1.79	
236		1.39	
237		1.90	
238		2.28	
239		1.94	
240		1.90	
241		2.45	
242		2.43	
243		2.24	
244		1.68	
245		4.49	
246		N.D.	
247		N.D.	
248		N.D.	
249		N.D.	
250		6.12	
251	$\alpha_9 - \alpha_{10}$ (L18)	4.88	
252		3.26	
253		2.46	
254		2.70	
255		2.74	
256		2.65	
257		2.36	
258		2.79	
259		3.66	
260		2.94	
261		1.64	

Table B.6. The kinetic values of the wild-type protein (*PhDIMT1_WT*) and mutants (*PhDIMT1_D117A*, *PhDIMT1_N132A*, and *PhDIMT1_Y135A*) were calculated with respect to the substrate h45. The assay was performed in triplicate, followed by non-linear regression analysis using Michaelis-Menten kinetics. Each protein K_{cat} value was calculated based on the V_{max} divided by total enzyme concentration ($E_{tot} = 5 \mu M$). All the reactions were performed at 37°C for 30 min. N.D.: not detected.

Protein name	V_{max}^* (RLU min ⁻¹)	K_m^* (μM)	K_{cat}^* (min ⁻¹)	K_{cat}/K_m (M ⁻¹ S ⁻¹)
--------------	---	------------------------	-------------------------------------	---

Supplementary data to chapter 4

<i>PhD1MT1</i> _WT	303±10.10	0.29±0.04	60.60±2.02	3.48x10 ⁶
<i>PhD1MT1</i> _D117A	N.D.			
<i>PhD1MT1</i> _N132A	267±18.64	2.29±0.34	53.40±3.72	0.38x10 ⁶
<i>PhD1MT1</i> _Y135A	318±10.80	1.12±0.10	63.60±2.16	0.93x10 ⁶



APPENDIX C
Supplementary data to Chapter 5

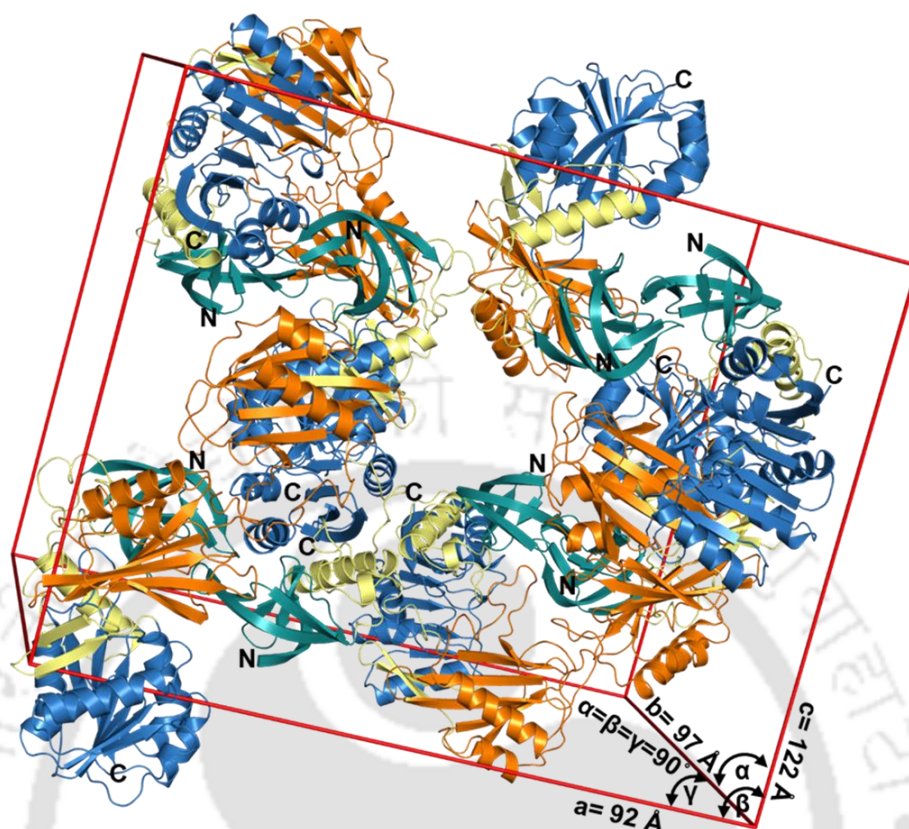


Figure C.1. Crystal packing arrangement of *PhRImCD*, which has been resolved within the $P2_12_12_1$ space group (PDB id: 8Z62). Within the crystal unit cell, an intriguing assembly is observed, where eight protomers originating from four asymmetric units are intricately packed. These protomers exhibit a unique 180° rotational relationship with respect to each other, contributing to the overall crystal lattice.

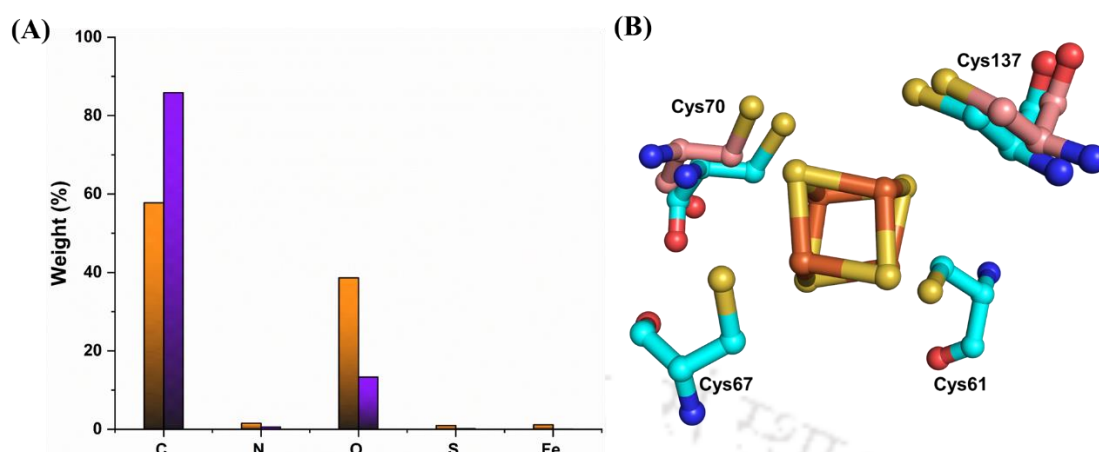


Figure C.2. Essential role of coordinated cysteine residues in preserving the iron-sulfur cluster in RNA m^5U MTases. (A) The histogram illustrates a comparative shift in the elemental composition between *PhRlmCD* (orange) and *PhRlmCD_EDTA* (violet), confirming the presence of an iron-sulfur cluster in *PhRlmCD*. (B) The superposition of the crystal structure (PDB ID: 2JJQ, salmon) and the AlphaFold model (cyan) of *PabTrmU54* highlight the necessity of coordinating conserved cysteine residues for maintaining the integrity of the cluster.

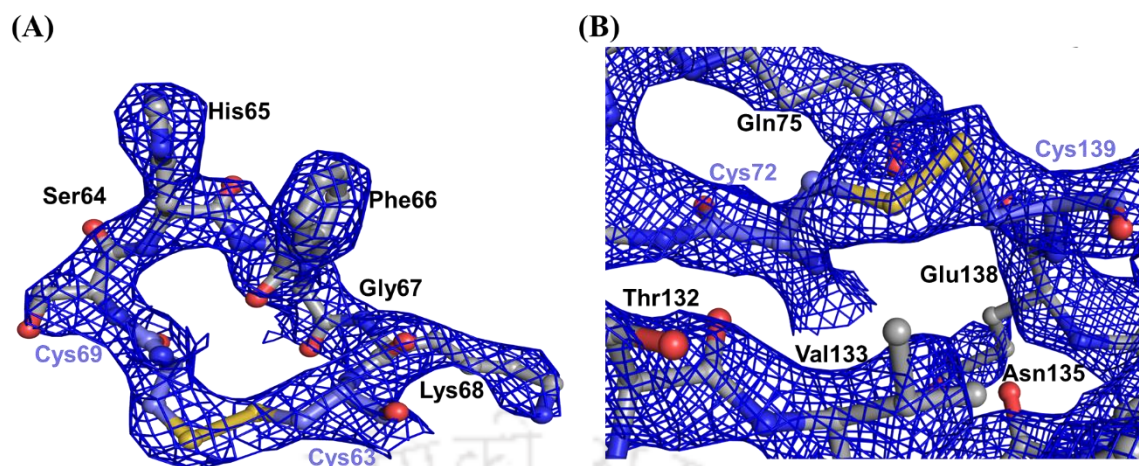


Figure C.3. The electron density of disulfide bridges in *PhRlmCD* is distinctly evident. The stereo view of the $2F_o - F_c$ map is depicted with a contour level set at 1σ . The continuous electron density clearly illustrates the covalent bonding between (A) Cys63 and Cys69, as well as (B) Cys72 and Cys139, revealing the formation of disulfide bridges.

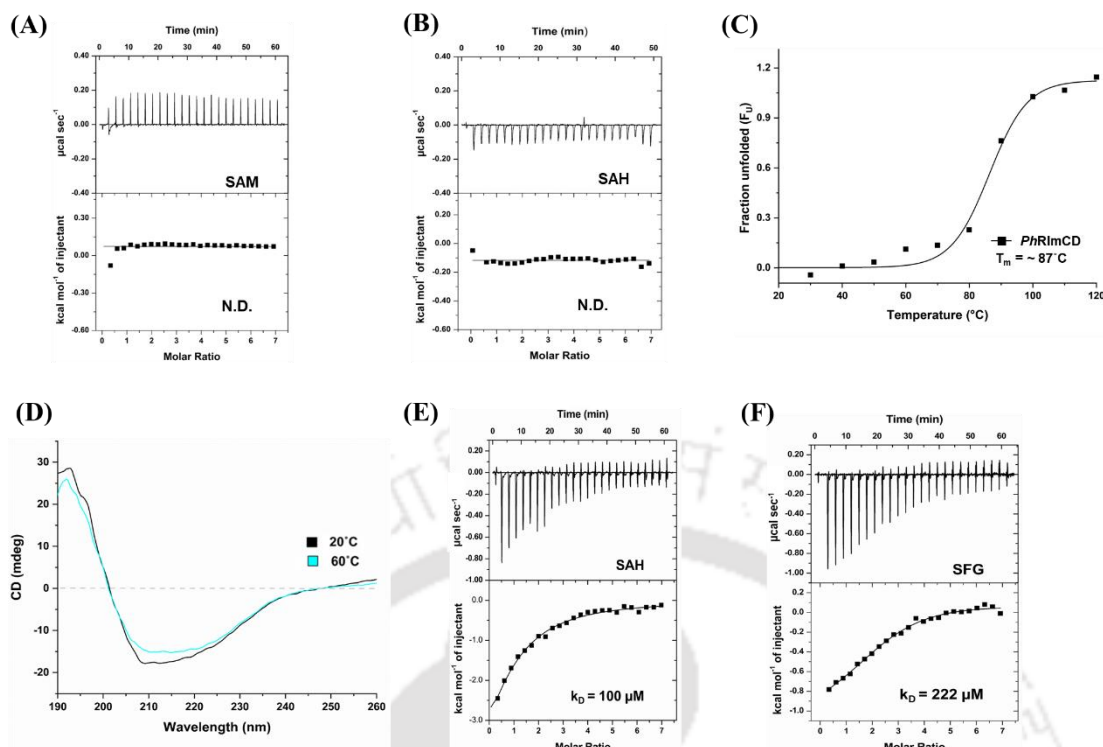


Figure C.4. The binding isotherm illustrates the interaction of *PhRlmCD* with various cofactors. Thermodynamic data of the protein *PhRlmCD* was acquired from ITC at 25°C in the presence of cofactors (A) SAM and (B) SAH. (C) Thermal unfolding curves for *PhRlmCD* (depicted by filled square) were generated by monitoring the CD ellipticity at 208 nm. (D) Alignment of CD spectra recorded at 20°C (Black) and 60°C (Cyan). Thermograms were generated at an elevated temperature (60°C) for the protein *PhRlmCD* with the cofactor (E) SAH and (F) SFG.

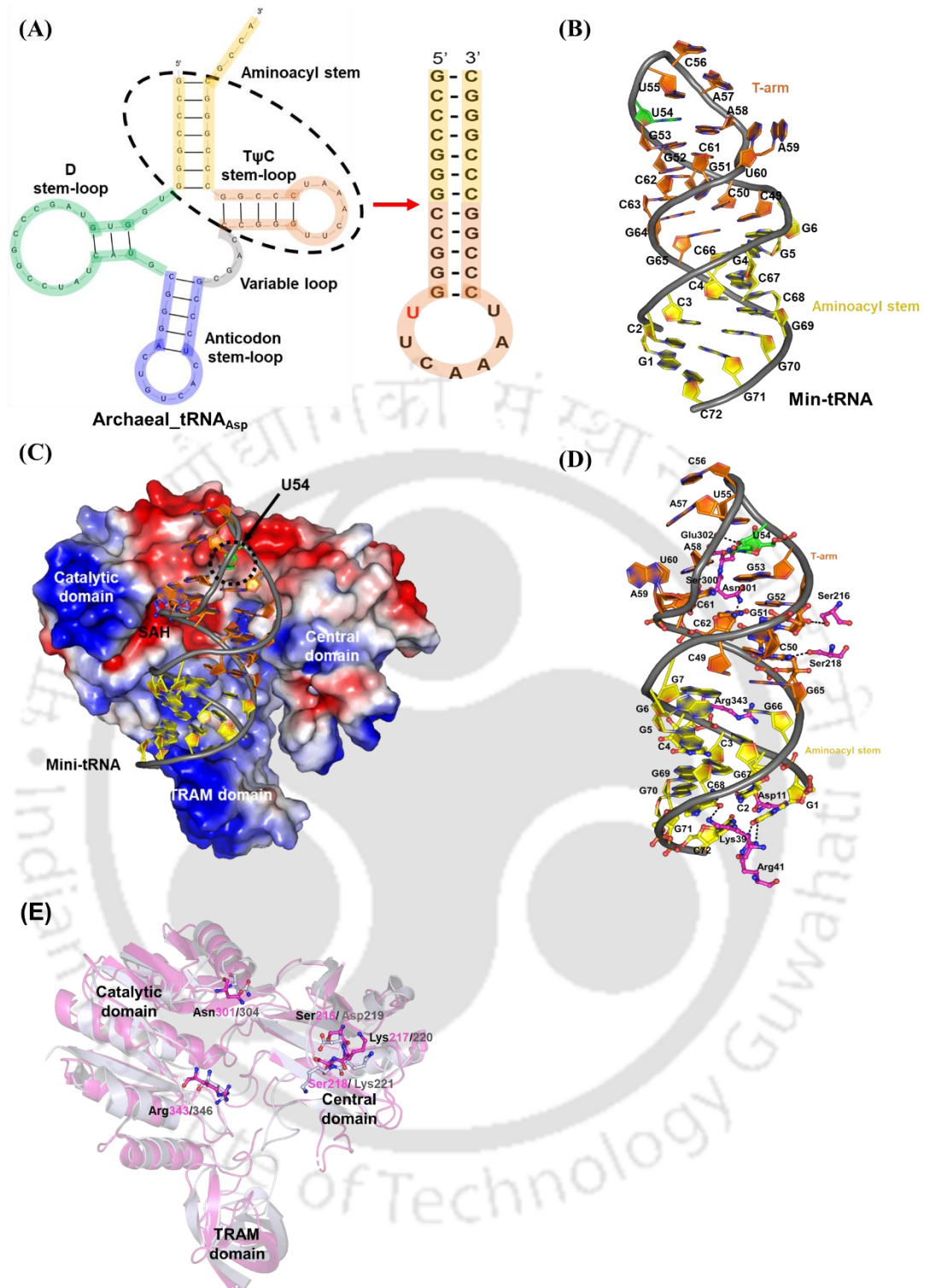


Figure C.5. Interactions between *PabTrmU54* and mini-tRNA in the presence of cofactor. (A) Cloverleaf structures of archaeal tRNA^{Asp} and a 31-mer mini-tRNA substrate encompassing the aminoacyl arm stem and TψC stem-loop. (B) The 31-nucleotide-long construct of mini-tRNA was modeled. (C) Visualization of electrostatic charge distribution on *pabTrmU54* (PDB id: 2JJQ) docked with mini-tRNA, with

positively-charged surfaces in blue (+3 kT/e) and negatively-charged surfaces in red (-3 kT/e). (D) Depiction of interactions between *pab*TrmU54 and mini-tRNA, highlighting active site residues in magenta sticks. (E) Superposition of *pab*TrmU54 (magenta) with *Ph*RlmCD (grey), revealing residual conservation in the active site pocket.

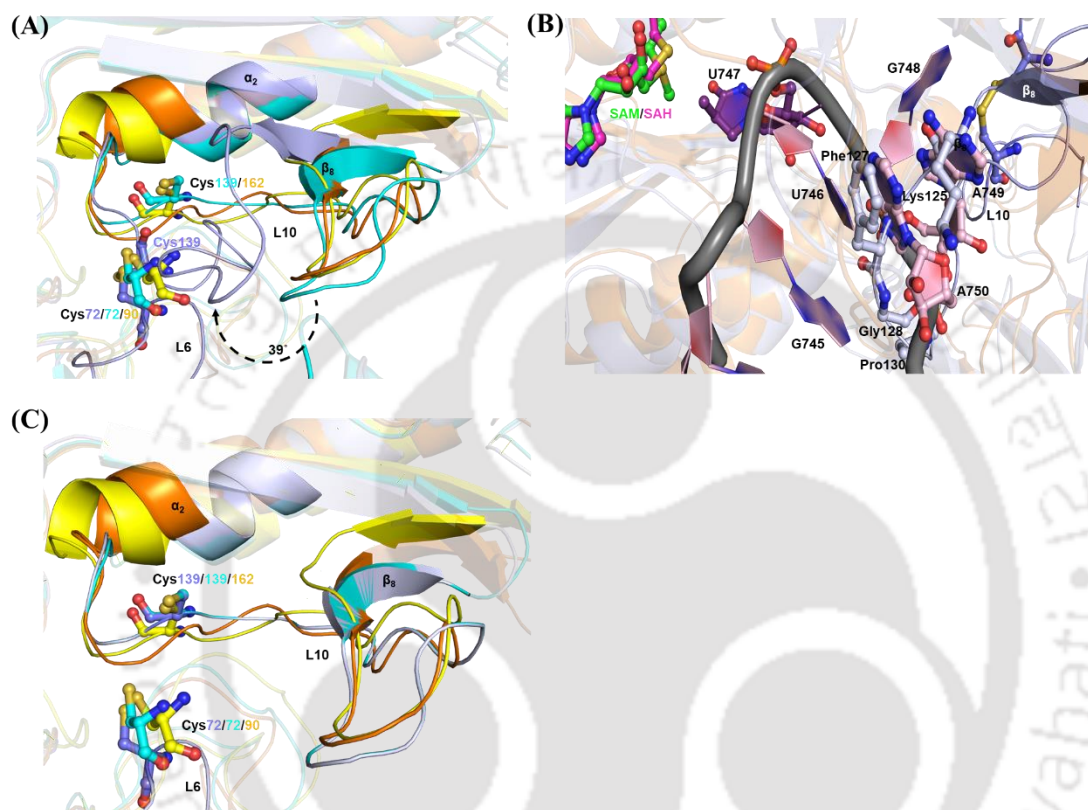


Figure C.6. Hinge movements were facilitated by disulfide bridges. (A) The juxtaposition of the crystal structure of *Ph*RlmCD (closed state) (grey) with its AlphaFold model (cyan) and its closest homologs, *Sp*RlmCD (PDB id: 5ZTH) and *Ec*RlmD (yellow, PDB id: 2BH2), reveals a 39° hinge movement (α_2 - β_8). This motion is attributed to the formation of a disulfide bridge between Cys72 and Cys139 in the *Ph*RlmCD structure. (B) Comparative structural analysis between the crystal structure of *Ph*RlmCD and the ternary complex (Protein•SAH•mini-rRNA) structure of *Sp*RlmCD suggests that the hinge region hinders the accommodation of substrate rRNA in the catalytic pocket. (C) The overlay of the crystal structure of *Ph*RlmCD in the open state with the AlphaFold model and its closest homologs reveals restricted hinge movement due to the absence of a disulfide bond.

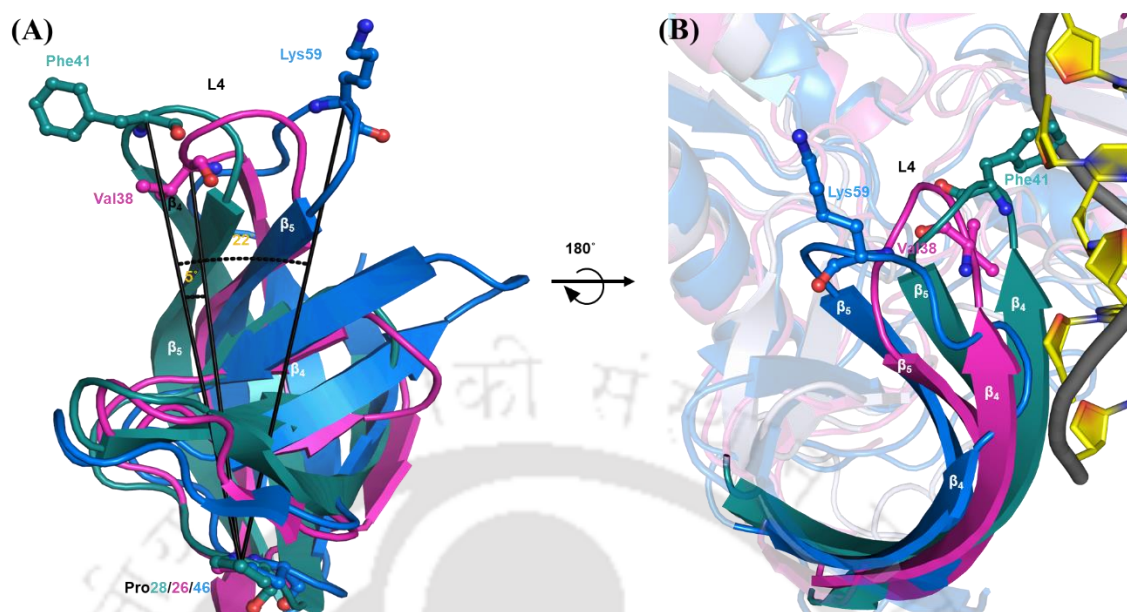


Figure C.7. Rotation differences in TRAM domains are observed based on the type of substrate. (A) Superposition of the TRAM domains of *PhRlmCD* (Teal) with *pabTrmU54* (magenta) and *EcRlmD* (marine) reveals distinct conformations of the domain in *PhRlmCD* compared to *pabTrmU54* and *EcRlmD*. Specifically, conformational changes result in approximately a $\sim 5^\circ$ and $\sim 20^\circ$ tilt in loop L4 of the TRAM domain in *PhRlmCD* and *EcRlmD*, respectively, compared to *pabTrmU54*. (B) Upon rotating the overlay structure by 180° in the presence of a mini-tRNA substrate indicates that the $\sim 5^\circ$ tilt of loop L4 towards the substrate limits the entry of mini-tRNA into its catalytic groove. In contrast, the $\sim 20^\circ$ tilt away from mini-tRNA of Loop L4 in the TRAM domain of *EcRlmD* suggests that this loop is not directly associated with substrate specificity.

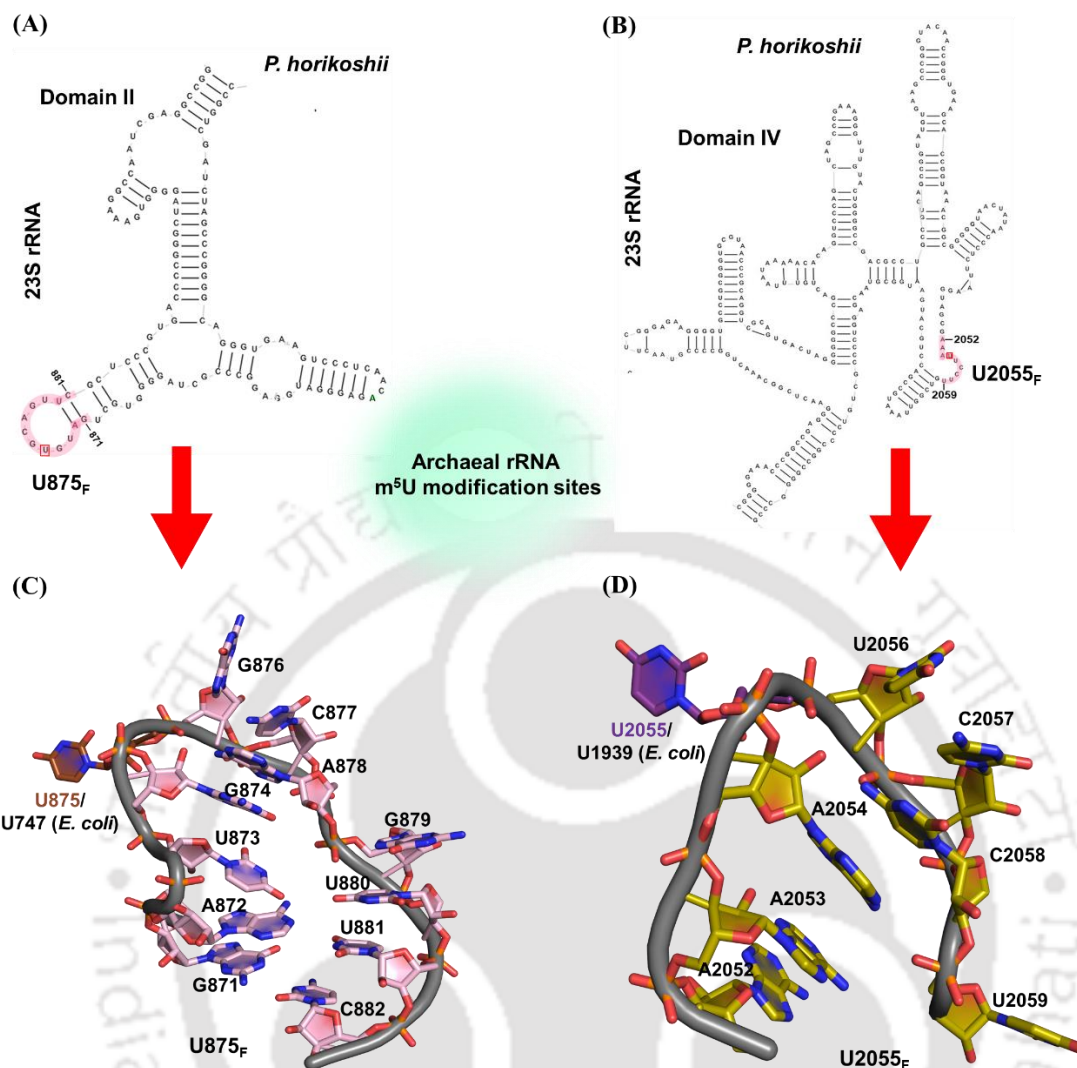


Figure C.8. The 35-hairpin motif is located in domains II and IV of the 23S rRNA in *P. horikoshii*. (A, B) The consistent secondary structure of the 35-hairpin motifs in domains II and IV of 23S rRNA is evident in the archaeon *P. horikoshii* respectively. (C, D) The three-dimensional arrangement is depicted in the model of the likely modified loops within domains II and IV of 23S rRNA in this archaeon, respectively.

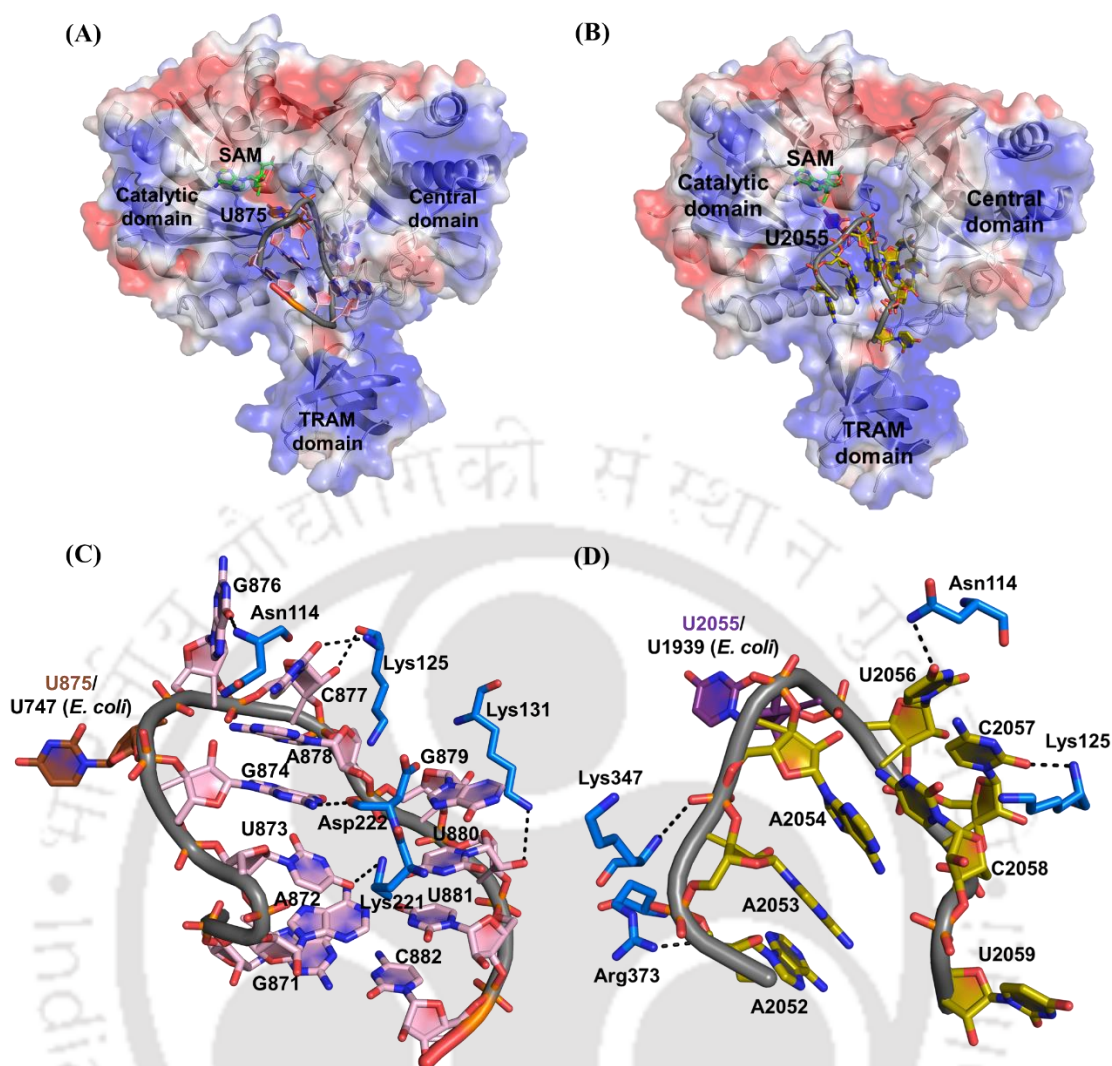


Figure C.9. Mini-rRNA fragments have been docked to the open state of *PhRlmCD*. Visualization of electrostatic charge distribution on *PhRlmCD* docked with fragments (A)U875_F and (B) U2055_F, with positively-charged surfaces in blue (+3 kT/e) and negatively-charged surfaces in red (-3 kT/e). The interaction studies of the active site residue of *PhRlmCD* (marine) with neighbouring target nucleotide of mini-rRNA fragments (C) U875_F and (D) U2055_F.

Supplementary data to chapter 5

Table C.1. Thermodynamic parameters of cofactors (SAM, SAH, and SFG) binding with *PhRlmCD*. S-Adenosyl-L-methionine, SAM; S-Adenosyl-L-homocysteine, SAH; and Adenosyl-L-ornithine, SFG; N.D., not detected.

Protein name	Protein conc. (μM)	Cofactor name	Cofactor conc. (mM)	Experimental temperature (°C)	Stoichiometry ratio (n) [†]	Association constant x 10 ⁴ (K _a , M ⁻¹) [†]	Dissociation constant (k _D , μM)	ΔH [†]	TΔS	ΔG
								(kcal mol ⁻¹)		
PhRlmCD	177	SAM	8.85	25	N.D.					
		SAH								
	68	SAM	3.43	60	0.72±0.06	6.73±1.50	14.85	-6.61±0.71	0.67	-7.28
		SAH			0.83±0.18	1.00±0.14	100	-7.54±1.86	-1.29	-6.25
	200	††SFG	10	1.00±00	0.448±0.09	222	-2.10±0.18	3.10	-5.20	

[†]The values are provided along with the error of the fit of titration.

^{††}Due to the high value of the error of the fit of titration, the Stoichiometry ratio was fixed as n=1 during the one-site-binding fitting model.

Table C.2. Table showing domain organization among the various life forms comprising the genes coding for the SAM-dependent RNA m⁵U MTase (RlmC, RlmD, RlmCD, TrmA, Trm2, and Arm⁵U).

S. No.	Uniprot id	m ⁵ U MTase s	Substrate type	Domain organisation [†]		
				TRAM domain	Central domain	Catalytic domain
1.	P44083	RlmC	rRNA	-	+	+
2.	Q9CPH1			-	+	+
3.	D3VBF7			-	+	+
4.	A0A4R0FQI6			-	+	+
5.	A0A482PFR6			-	+	+
6.	Q8ZQJ5			-	+	+
7.	P75817			-	+	+
8.	Q83S14			-	+	+
9.	Q3IDM6	RlmD		+	+	+
10.	B2FPX2			+	+	+
11.	Q9PAY7			+	+	+
12.	Q46ZH7			+	+	+
13.	Q3K8E9			+	+	+
14.	Q4QNK7			+	+	+
15.	A1JJR0			+	+	+
16.	Q8ZME1			+	+	+
17.	Q83QD5			+	+	+
18.	P55135			+	+	+
19.	O31503	RlmCD		+	+	+
20.	M7NF39			+	+	+
21.	Q97R12			+	+	+

Supplementary data to chapter 5

22.	V2YDS 8			+	+	+
23.	A0A2N9 PD12			+	+	+
24.	A0A151 B3D5			+	+	+
25.	A0A379 MRI2			+	+	+
26.	A0A1V5 WFT6			+	+	+
27.	A0A7Z2 TFJ3	TrmA	tRNA	-	-	+
28.	B0UVN 9			-	-	+
29.	C3LPZ5			-	-	+
30.	Q8ZAA 0			-	-	+
31.	Q6CZB4			-	-	+
32.	A7ML9 5			-	-	+
33.	P23003			-	-	+
34.	P22038			-	-	+
35.	A8N040	Trm2	tRNA	+	+	+
36.	P33753			+	+	+
37.	O58994	Arm ⁵ U	rRNA	+	+	+
38.	O58864		tRNA	+	+	+
39.	Q9UZK 1		rRNA	+	+	+
40.	Q9UZR 7		tRNA	+	+	+
41.	Q5JE68		rRNA	+	+	+
42.	Q5JHF7		tRNA	+	+	+
43.	Q74N58		rRNA	+	+	+

† The presence and absence of the domain are indicated by ‘+’ and ‘-’ respectively.



Table C.3. List of RNA m⁵U MTases used for the construction of the phylogenetic tree.

S. No.	Protein name	UniProt id	Organism name	Representation in phylogenetic tree
39.	RlmC	P44083	<i>Haemophilus influenzae</i>	HiRlmC
40.		Q9CPH1	<i>Pasteurella multocida</i>	PmRlmC
41.		D3VBF7	<i>Xenorhabdus nematophila</i>	XnRlmC
42.		A0A4R0FQI6	<i>Enterobacter quasihormaechei</i>	EnqRlmC
43.		A0A482PFR6	<i>Citrobacter rodentium</i>	CirRlmC
44.		Q8ZQJ5	<i>Salmonella typhimurium</i>	SatRlmC
45.		P75817	<i>Escherichia coli</i>	EcRlmC
46.		Q83S14	<i>Shigella flexneri</i>	ShfRlmC
47.	RlmD	Q3IDM6	<i>Pseudoalteromonas translucida</i>	PstRlmD
48.		B2FPX2	<i>Stenotrophomonas maltophilia</i>	StmRlmD
49.		Q9PAY7	<i>Xylella fastidiosa</i>	XyfRlmD
50.		Q46ZH7	<i>Cupriavidus pinatubonensis</i>	CupRlmD
51.		Q3K8E9	<i>Pseudomonas fluorescens</i>	PsefRlmD
52.		Q4QNK7	<i>Haemophilus influenzae</i>	HiRlmD
53.		A1JJR0	<i>Yersinia enterocolitica</i>	YeRlmD
54.		Q8ZME1	<i>Salmonella typhimurium</i>	SatRlmD
55.		Q83QD5	<i>Shigella flexneri</i>	ShfRlmD
56.		P55135	<i>Escherichia coli</i>	EcRlmD
57.	RlmCD	O31503	<i>Bacillus subtilis</i>	BsRlmCD
58.		M7NF39	<i>Bhargavaea cecembensis</i>	BhcRlmCD
59.		Q97R12	<i>Streptococcus pneumoniae</i>	SpRlmCD
60.		V2YDS8	Firmicutes bacterium ASF500	FibRlmCD
61.		A0A2N9PD12	<i>Flavobacterium columnare</i>	FcRlmCD
62.		A0A151B3D5	<i>Clostridium tepidiprofundum</i>	CtRlmCD
63.		A0A379MRI2	<i>Rikenella microfusum</i>	RmRlmCD
64.		A0A1V5WFT 6	Spirochaetes bacterium	SpbRlmCD

65.	TrmA	A0A7Z2TFJ3	<i>Pseudomonas fluorescens</i>	PsefTrmA
66.		B0UVN9	<i>Histophilus somni</i>	HisTrmA
67.		C3LPZ5	<i>Vibrio cholerae</i>	VcTrmA
68.		Q8ZAA0	<i>Yersinia pestis</i>	YpTrmA
69.		Q6CZB4	<i>Pectobacterium atrosepticum</i>	PeaTrmA
70.		A7ML95	<i>Cronobacter sakazakii</i>	CrsTrmA
71.		P23003	<i>Escherichia coli</i>	EcTrmA
72.		P22038	<i>Salmonella typhimurium</i>	SatTrmA
73.	Trm2	A8N040	<i>Coprinopsis cinerea</i>	CocTrm2
74.		P33753	<i>Saccharomyces cerevisiae</i>	SacTrm2
75.	Arm ⁵ U	O58994	<i>Pyrococcus horikoshii</i>	PH1259
76.		O58864	<i>Pyrococcus horikoshii</i>	PH1137
77.		Q9UZK1	<i>Pyrococcus abyssi</i>	PAB0760
78.		Q9UZR7	<i>Pyrococcus abyssi</i>	_{Pab} TrmU54
79.		Q5JE68	<i>Thermococcus kodakarensis</i>	TK1116
80.		Q5JHF7	<i>Thermococcus kodakarensis</i>	TK2135
81.		Q74N58	<i>Nanoarchaeum equitans</i>	NEQ053

Table C.4. List of interaction(s) of the cofactors (SAM or SAH) with different RNA m⁵U MTases. NHBOND: number of hydrogen bonds.

Protein (Cocrystal/docked)	Cofactor	NHBOND	Interactions		
			Cofactor atom	Protein atom	Distance (Å)
SpRlmCD (Cocrystal)	SAH	8	N1	Ala361 N	2.80
			N3	Leu334 N	3.50
			O2'	Glu333 O ^{ε1}	3.20
			O3'	Glu333 O ^{ε2}	3.20
			N	Tyr312 O	3.10

Supplementary data to chapter 5

			O	Asp381 O ^{δ2}	3.30
			O	Gln283 O ^{ε1}	3.40
			OXT	Tyr293 OH	3.0
<i>Pab</i> TrmU4 (Cocrystal)	SAH	9	N6	Asp326 O ^{ε1}	2.96
			N1	Asp326 N	2.97
			N3	Ser300 O ^γ	3.13
			O2'	Asp299 O ^{δ1}	2.67
			O3'	Asp299 O ^{δ2}	2.61
			N	Asp340 O ^{δ2}	2.90
			O	Thr283 O ^{γ1}	2.87
			O	Gln252 N ^{ε2}	3.05
			OXT	Thr283 O ^{γ1}	3.14
<i>Ec</i> RlmD (Cocrystal)	SAH	8	N6	Asn342 O ^{δ1}	3.10
			N1	Ala343 N	2.80
			O2'	Glu333 O ^{ε2}	2.60
			O3'	Glu333 O ^{ε1}	2.70
			N	Asp340 O ^{δ2}	3.10
			N	Phe294 O	3.0
			O	Asn299 N ^{δ2}	2.9
			O	Gln265 N ^{ε2}	3.00
<i>Ph</i> RlmCD (Docked)	SAM	7	N6	Glu330 O ^{ε2}	3.10
			N1	Ala329 N	2.70
			N3	Val303 N	3.20
			O2'	Glu302 O ^{ε1}	3.50

			O3'	Glu302 O ^{ε2}	3.50
			O	Gln255 N	3.20
			OXT	Ile284 N	3.20

Table C.5. List of interaction(s) among the various substrates with *Pab*TrmU54 and *PhRlmCD*. NHBOND: number of hydrogen bonds.

Protein	Substrate	NHBOND	Interactions		
			Nucleotide No. (atom)	Amino acid No. (atom)	Distance (Å)
<i>Pab</i> TrmU54	mini-tRNA	11	C2 O2'	Asp11 O ^{δ1}	2.40
			G1 O6	Arg41 N ^{η2}	2.00
			C72 N4	Arg41 N ^{η2}	2.00
			G1 N2	Arg41 N ^{η1}	3.50
			G71 O6	Lys39 N ^ζ	2.80
			C4 OP1	Arg343 O	2.90
			G64 O2'	Ser218 O ^γ	3.20
			G51 O2'	Ser216 O ^γ	3.20
			U54 O2'	Glu302 O ^{ε1}	3.00
			C61 O2'	Ser300 O ^γ	2.40
			C62 O2'	Asn301 N ^{δ2}	2.50
<i>PhRlmCD</i>	mini-rRNA (U875 _F)	11	G876 O6	Asn114 N	2.90
			C877 O2	Lys125 N	3.00
			C877 O2'	Lys125 O	2.90
			U880 O2'	Lys131 N ^ζ	3.50

			U873 O4	Lys221 N ^ζ	3.20
			G874 N1	Asp222 O	1.90
			U875 O4	Gln255 N ^{ε2}	2.60
			U875 N3	Gln255 O ^{ε1}	3.10
			U875 O4	Glu404 O ^{ε1}	2.00
			U875 O2'	Cys369 S ^γ	3.50
			U875 O2'	Arg346 N ^{η2}	2.20
<i>PhRlmCD</i>	mini-rRNA (U2055 _F)	10	U2056 O4	Asn114 N ^{δ2}	2.80
			C2057 O2	Lys125 N	2.40
			A2054 OP1	Lys347 N	2.90
			A2052 O2'	Arg373 N ^{η2}	3.00
			U2055 O4	Asp343 O ^{δ1}	3.20
			U2055 O4	Asp343 O ^{δ2}	3.30
			U2055 O4	Gln255 N ^{ε2}	2.20
			U2055 N3	Gln255 O ^{ε1}	3.40
			U2055 O2'	Cys369 S ^γ	2.90
			U2055 O2'	Arg346 N ^{η2}	2.10

Institute of Technology

APPENDIX D
Supplementary data to Chapter 6

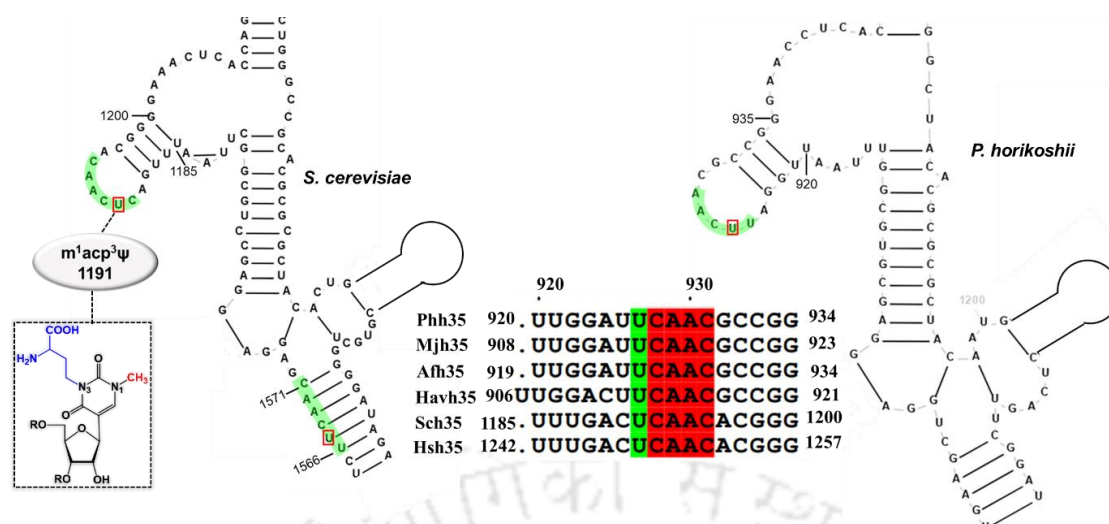
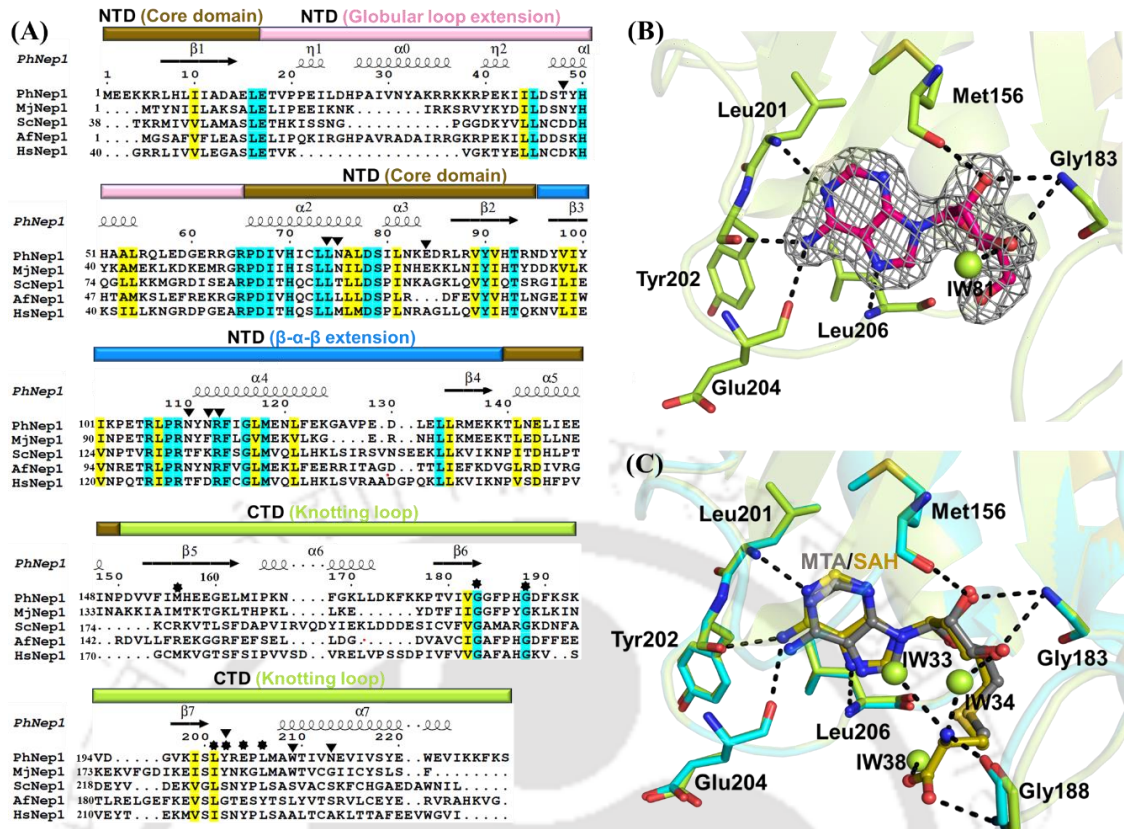


Figure D.1. Analysis of the conserved putative Nep1 MTase target sites in SSU.

Left: The secondary structure of a segment of yeast 18S rRNA harboring two Nep1 binding RNA-consensus motifs 5'-C/UUCAAC-3' (highlighted in green) in *S. cerevisiae*. Each targeted uridine is initially converted to N1-specific pseudouridine to facilitate recognition by Nep1 MTases. Additionally, the pseudouridine at position 1191 undergoes hypermodification, as illustrated in the inset. Middle: Comparative MSA of helix 35 (h35) from various organisms, including *P. horikoshii* OT3 (Phh35), *M. jannaschii* (Mjh35), *A. fulgidus* (Afh35), *H. volcanii* (Havh35) in 16S rRNA, as well as *S. cerevisiae* (Sch35) and *H. sapiens* (Hsh35) in 18S rRNA. Highly conserved nucleotides and the modification site within h35 are highlighted in red and green, respectively. Right: The secondary structure of the equivalent fragment from *P. horikoshii* OT3 of 16S rRNA is shown with the location of the only conserved copy of the RNA consensus motif highlighted in green. Modified nucleotides within the secondary structures are denoted by red boxes.



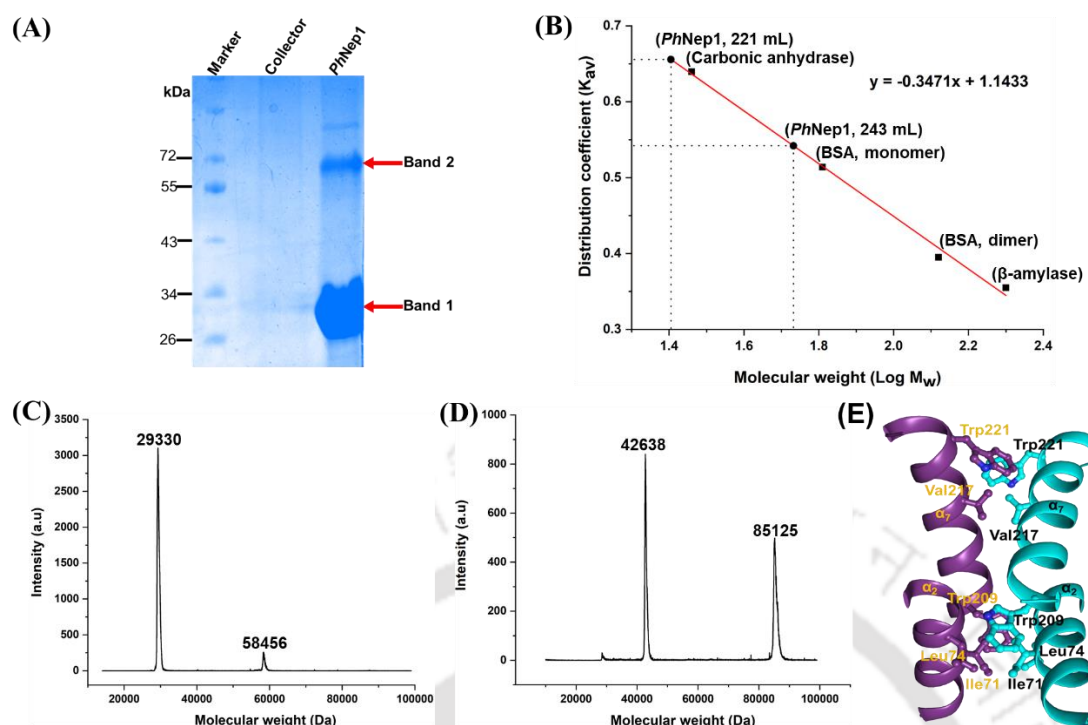


Figure D.3. The dimeric assembly of *PhNep1*. A) SDS-PAGE analysis of concentrated *PhNep1* protein reveals two predominant sharp peaks in the purified protein. Lane 1 shows the protein marker, while lanes 2 and 3 depict the collector solution and concentrated protein, respectively. (B) Eluted fractions of *PhNep1*, alongside SEC standards, were run on a HiLoad™ 26/600 Superdex™ 200 pg column to determine the relative molecular mass of the protein. A molecular weight determination curve was constructed by plotting the distribution coefficient against the molecular weight of the SEC standards: β -amylase (200 kDa), bovine serum albumin (BSA dimer, 132 kDa), BSA monomer (66 kDa), and carbonic anhydrase (29 kDa). In the graph, the apparent molecular weights of the recombinant protein *PhNep1* were determined using the distribution coefficients (K_{av}) of two fractions (221 and 243 mL), indicated by circles, while SEC standards are represented by square symbols. (C, D) Mass spectra analysis of dimeric proteins (*PhNep1*: C; M1Pi: D) using MALDI-TOF analysis. (E) The arrangement of two sets of parallel helices fostering hydrophobic interactions is illustrated. In monomer A, helices α_2 and α_7 are depicted in violet, while in monomer B, they are presented in cyan. The residues participating in hydrophobic packing interactions have their side chains visualized in a ball-and-stick format. Amino acid labels are colored differently to distinguish residues from different monomers.

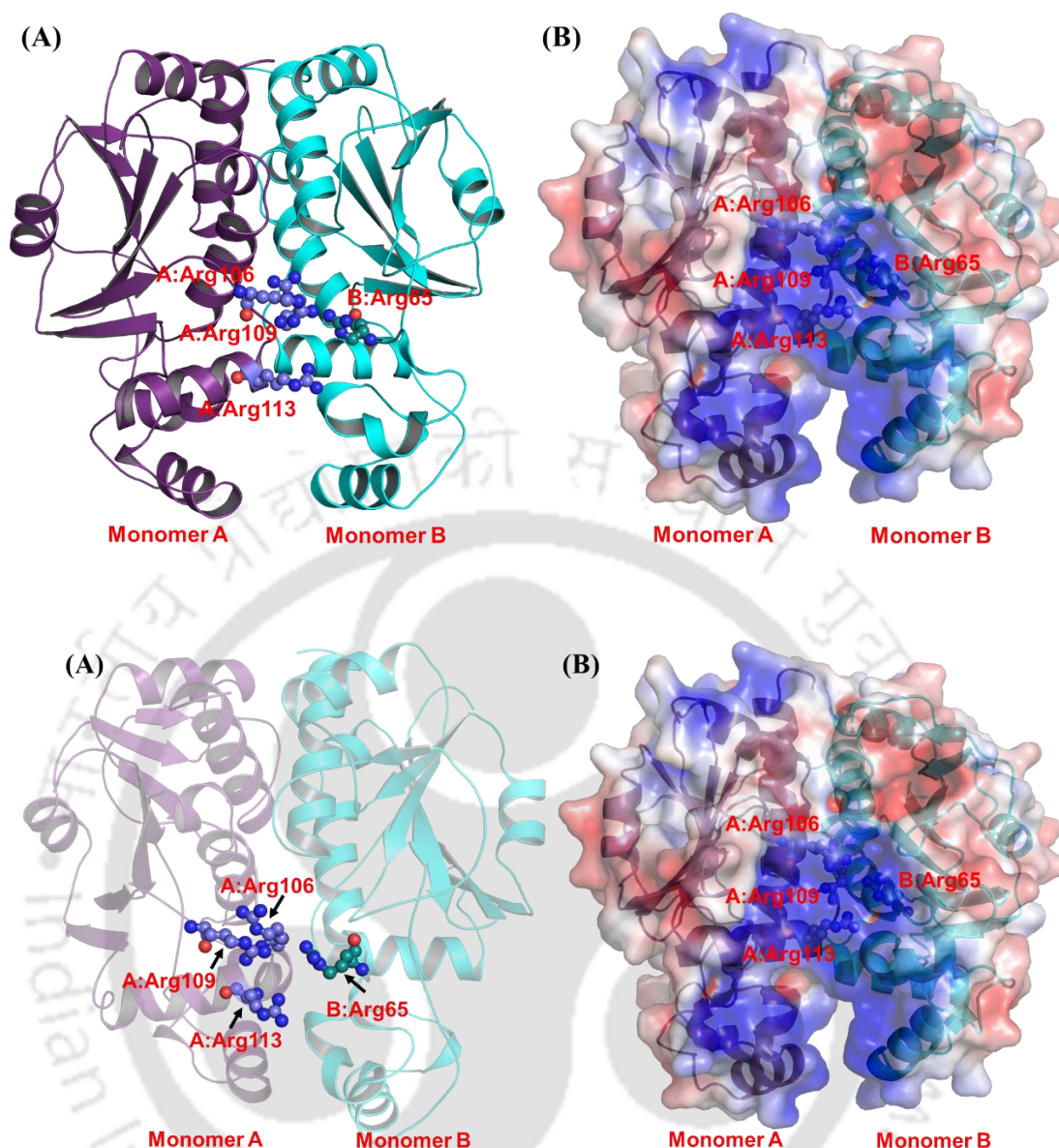


Figure D.4. The conserved arginine residues present at the interface of the *PhNep1* dimer are crucial during RNA binding. (A) The interface of *PhNep1* Dimer contains conserved arginine residues (Arg106, Arg109, and Arg113) from monomer A (violet) and Arg65 from monomer B (cyan). (B) The electrostatic charge distribution on *PhNep1* depicts that these conserved arginine residues formed a positive charge cleft in the interfaced region of the dimeric protein for the negatively charged RNA to bind.

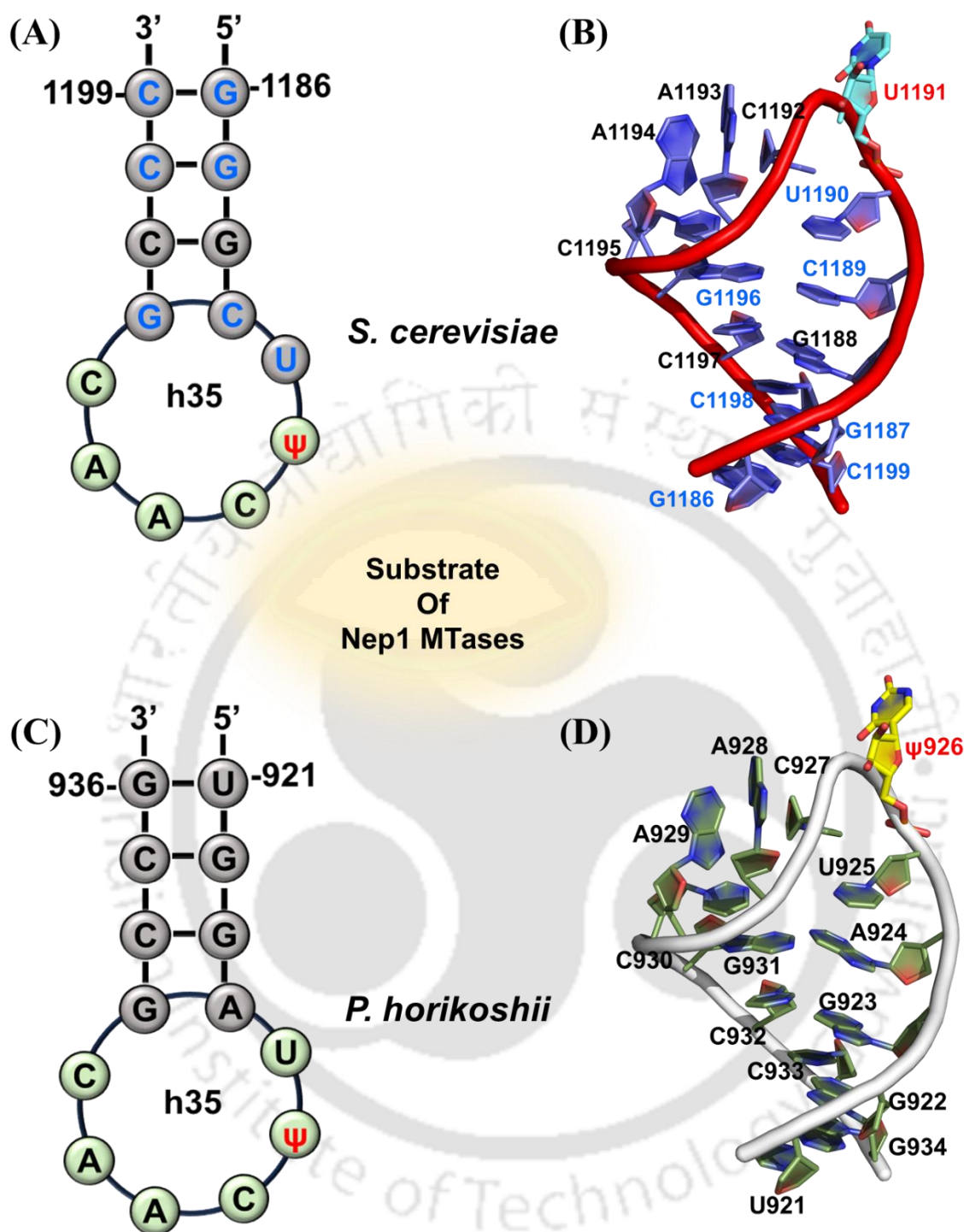


Figure D.5. The mini-rRNA consensus motif (5'-C/UUCAAC-3') located at helix 35 (h35) of 18S/16S rRNA of *S. cerevisiae*/*P. horikoshii*. (A) Secondary structure representation of the mini-rRNA consensus motif of h35 illustrates the spatial arrangement of nucleotides within the h35 region, highlighting its distinctive folding pattern. The change in the nucleotide sequence is highlighted in blue, and the target pseudouridine residue is highlighted in red. (B) Crystal structure of this RNA consensus motif located at h35 of 18S rRNA from *S. cerevisiae* (PDB id: 3OIN). (C) The

secondary structure of the mini-rRNA consensus motif of h35 of 16S rRNA of *P. horikoshii* represents the characteristic folding pattern similar to its eukaryotic counterpart. (D) The corresponding tertiary structure model of mini-rRNA consensus motif from the archaeon.

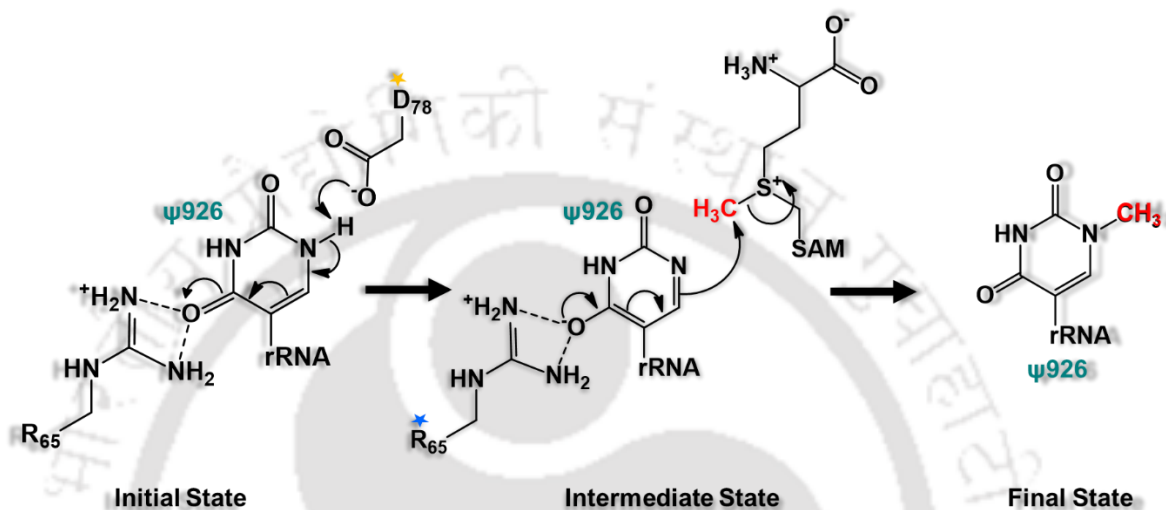


Figure D.6. A probable mechanism for the N1-methylation of pseudouridine, catalyzed by *PhNep1*. The methylation reaction proceeds through three steps: initial, where Asp78 serves as a proton acceptor; an intermediate step, where conserved Arg65 facilitates the deprotonation of N1; and finally, the formation of methyl bond in the N1 position target ψ926 of h35 located in the 16S rRNA of *P. horikoshii*.

Table D.1. List of different members of SPOUT MTases used in phylogenetic tree construction.

Clade	Protein	Substrate (modification)	UniProt id	Organism	Representation in phylogenetic tree	Domain architecture
I	Nep1	rRNA (m ¹ ψ1191 ^a /914 ^b)	O50087	<i>Pyrococcus horikoshii</i>	PH1379	NTD–SPOUT
			Q57977	<i>Methanocaldococcus jannaschii</i>	MjNep1	
			O29524	<i>Archaeoglobus fulgidus</i>	AfNep1	
			Q5JI44	<i>Thermococcus kodakarensis</i>	TkNep1	
			Q97WJ0	<i>Saccharolobus solfataricus</i>	SasNep1	
			A3DNG9	<i>Staphylothermus marinus</i>	StmNep1	
			B1YD95	<i>Pyrobaculum neutrophilum</i>	PynNep1	
			Q9HJ48	<i>Thermoplasma acidophilum</i>	ThaNep1	
			Q9YES9	<i>Aeropyrum pernix</i>	AepNep1	
			A0A1V0N230	<i>Ferroplasma acidiphilum</i>	FaNep1	
			Q06287	<i>Saccharomyces cerevisiae</i>	ScNep1	

Supplementary data to chapter 6

II			Q92979	<i>Homo sapiens</i>	HsNep1	
			O35130	<i>Mus musculus</i>	MumNep1	
			Q9W4J5	<i>Drosophila melanogaster</i>	DmNep1	
			Q9XX15	<i>Caenorhabditis elegans</i>	CeNep1	
			A0A178V G55	<i>Arabidopsis thaliana</i>	ArTNep1	
			A0A2S7Y 449	<i>Beauveria bassiana</i>	BebNep1	
			Q9P8P7	<i>Candida albicans</i>	CaaNep1	
			I0FN01	<i>Macaca mulatta</i>	MamNep1	
			L8HU92	<i>Bos mutus</i>	BomNep1	
	RlmB	rRNA (Gm225 ^C)	P63177	<i>Escherichia coli</i>	EcRlmB	NTD– SPOUT
			Q8ZIV4	<i>Yersinia pestis</i>	YpRlmB	
			Q9HUM8	<i>Pseudomonas aeruginosa</i>	PsaRlmB	
			Q83CW6	<i>Coxiella burnetii</i>	CobRlmB	
			Q7VTK4	<i>Bordetella pertussis</i>	BopRlmB	
			P0A8I8	<i>Escherichia coli</i>	EcRlmH	
	TrmH	tRNA (Gm18 ^C)	Q5SM16	<i>Thermus thermophilus</i>	ThtTrmH	NTD– SPOUT– CTD
			P0AGJ2	<i>Escherichia coli</i>	EcTrmH	
			O67577	<i>Aquifex aeolicus</i>	AqaTrmH	
	TrmL	tRNA	P0AGJ7	<i>Escherichia coli</i>	EcTrmL	SPOUT

Supplementary data to chapter 6

III		(Cm34 ^C ; Um34 ^C)	Q0BBL1	<i>Burkholderia ambifaria</i>	BuaTrmL	only
	TrmJ	tRNA (Cm32 ^C /Um32 ^C)	P0AE01	<i>Escherichia coli</i>	EcTrmJ	SPOUT– CTD
			A0A0H2Z F87	<i>Pseudomonas aeruginosa</i>	PsaTrmJ	
	TrmY	tRNA (m ¹ ψ54 ^b)	Q59034	<i>Methanocaldoc occus jannaschii</i>	MjTrmY	NTD– SPOUT
			O50109	<i>Pyrococcus horikoshii</i>	PhTrmY	
			O29206	<i>Archaeoglobus fulgidus</i>	AfTrmY	
	RsmE	rRNA (m ³ U1498 ^C)	P0AGL7	<i>Escherichia coli</i>	EcRsmE	NTD– SPOUT
			O66552	<i>Aquifex aeolicus</i>	AqaRsmE	
			P54461	<i>Bacillus subtilis</i>	BsRsmE	
			P9WGX1	<i>Mycobacterium tuberculosis</i>	MytRsmE	
			Q9ZKD2	<i>Helicobacter pylori</i>	HepRsmE	
	RlmH	rRNA (m ³ Ψ1915 ^C)	P0A8I8	<i>Escherichia coli</i>	EcRlmH	SPOUT only
			P0C1V0	<i>Staphylococcus aureus</i>	StaaRlmH	
			A1A0L0	<i>Bifidobacterium adolescentis</i>	BiaRlmH	
			Q630E3	<i>Bacillus cereus</i>	BcRlmH	
			Q45601	<i>Bacillus subtilis</i>	BsRlmH	
			Q899K8	<i>Clostridium tetani</i>	CltrlmH	
	Trm5 6	tRNA (Cm56 ^d)	O58214	<i>Pyrococcus horikoshii</i>	PhTrm56	SPOUT– CTD

			A0A832T238	<i>Methanocaldococcus jannaschii</i>	MjTrm56	
			O29507	<i>Archaeoglobus fulgidus</i>	AfTrm56	
	TrmD	tRNA (m ¹ G37 ^C)	P0A873	<i>Escherichia coli</i>	EcTrmD	SPOUT–CTD
			O67463	<i>Aquifex aeolicus</i>	AqaTrmD	
			P43912	<i>Haemophilus influenzae</i>	HaiTrmD	
	IV	Trm10	tRNA (m ¹ G9 ^C , m ¹ A9 ^C)	Q6FQB2	<i>Candida glabrata</i>	CagTrm10
Q12400				<i>Saccharomyces cerevisiae</i>	ScTrm10	
Q08BM0				<i>Danio rerio</i>	DarTrm10	

^a*Saccharomyces cerevisiae* numbering.

^b*Methanococcus jannaschii* numbering.

^c*Escherichia coli* numbering.

^d*Pyrococcus horikoshii* numbering.

Table D.2. The molecular interactions occur between *PhNep1* and its homologs with MTA and SAH.

Cofactor	Cofactor atom	Protein (PDB id)			
		<i>PhNep1</i>	<i>AfNep1</i> (3O7B)	<i>MjNep1</i> (3BBD)	<i>ScNep1</i> (3OII)
		Residue (atom)/ IW (no.)			
SAH/MTA	N6	Tyr202 O	Gly193 O	Tyr186 O	Ser228 O

Supplementary data to chapter 6

		Glu204 O	Glu195 O	Lys188 O	Tyr230 O
	N1	Leu201 N	Leu192 N	Ile185 N	Leu227 N
	N7	Leu206 N	Tyr197 N	Leu190 N	Leu232 N
	O2'	Met156 O	Phe147 O	-	Leu180 O
		Gly183 N	Gly169 N	Gly162 N	Gly207 N
	O3'	Gly183 N	Gly169 N	Gly162 N	Gly207 N
		IW 34	IW 383	IW 327	Asp214 O ^{δ1}
SAH	N	Gly188 O	Gly174 O	Gly167 O	Gly212 O
		IW 33	-	-	Asp214 O ^{δ1}
		IW 34	-	-	Asp214 O ^{δ2}
	O	IW 38	IW 383	IW 327	IW 351
		-	-	IW 328	-
		-	-	Thr142 O ^{γ1}	-
	OXT	Gly188 O	-	Lys143 N	-

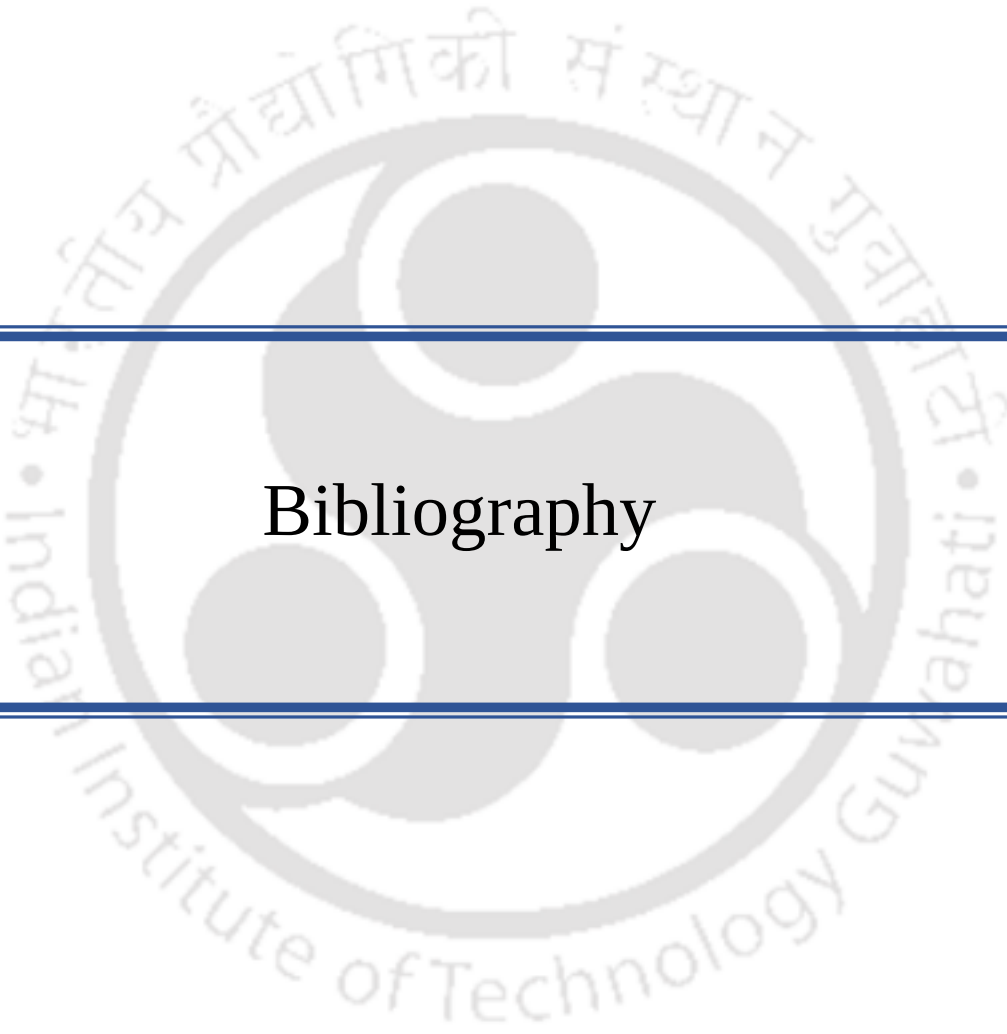
Table D.3. List of interaction(s) *PhNep1* with mini-rRNA consensus motif present in the h35 of 16S rRNA of *P. horikoshii*. NHBOND: number of hydrogen bonds.

Protein	Substrate	NHBOND	Interactions		
			Nucleotide No. (atom)	Chain: Amino acid No. (atom)	Distance (Å)
<i>PhNep1</i>	mini-rRNA	16	Ψ926 O4	B: Arg65 N ^{η2}	2.50

Supplementary data to chapter 6

	(h35)		Ψ926 O4'	B: Arg65 N ^{η1}	3.10
			Ψ926 O4	A: Arg106 N ^{η2}	2.30
			Ψ926 O2	A: Ala208 N	3.40
			Ψ926 O4	A: Arg109 N ^{η2}	3.50
			Ψ926 O2'	A: Arg109 N ^{η2}	3.50
			C927 N4	B: Thr48 O ^{γ1}	3.50
			C927 N4	A: Arg113 N ^{η1}	2.50
			C927 N3	A: Arg113 N ^{η1}	1.70
			A928 N6	A: Glu104 O	3.50
			A928 N1	A: Arg106 N	3.50
			A929 N6	A: Arg106 O	3.00
			C930 OP1	A: Lys124 N ^ζ	3.30
			G931 N2	A: Arg113 N ^{η2}	3.50
			C932 O4'	B: His51 N ^{ε2}	3.50
			C933 O2'	B: Arg55 N ^{η2}	3.50





Bibliography

1. Adhav VA, Shelke SS, Balanarayan P, and Saikrishnan K (2023). Sulfur-mediated chalcogen versus hydrogen bonds in proteins: a seesaw effect in the conformational space. *QRB Discov.*, 1-30.
2. Agarwalla S, Kealey JT, Santi DV, and Stroud RM (2002). Characterization of the 23 S ribosomal RNA m⁵U1939 methyltransferase from *Escherichia coli*. *J. Biol. Chem.*, 277(11), 8835-8840.
3. Agarwalla S, Stroud RM, and Gaffney BJ (2004). Redox reactions of the iron-sulfur cluster in a ribosomal RNA methyltransferase, RumA: optical and EPR studies. *J. Biol. Chem.* 279(33), 34123-34129.
4. Agrawal RK, Spahn CM, Penczek P, Grassucci RA, Nierhaus KH, and Frank J (2000). Visualization of tRNA movements on the *Escherichia coli* 70S ribosome during the elongation cycle. *J. Cell Biol.*, 150(3), 447-460.
5. Aguilo F, Li S, Balasubramaniyan N, Sancho A, Benko S, Zhang F, Vashisht A, Rengasamy M, Andino B, Chen CH, and Zhou F (2016). Deposition of 5-methylcytosine on enhancer RNAs enables the coactivator function of PGC-1 α . *Cell Rep.*, 14(3), 479-492.
6. Ahn HJ, Kim HW, Yoon HJ, Lee BI, Suh SW, and Yang JK (2003). Crystal structure of tRNA (m¹G37) methyltransferase: insights into tRNA recognition. *EMBO J.*, 22, 2593-2603.
7. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, and Walter P (2002). *Molecular Biology of the Cell*, 4th edition.
8. Ali IK, Lancaster L, Feinberg J, Joseph S, and Noller HF (2006). Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Mol. Cell*, 23(6), 865-874.
9. Alian A, Lee TT, Griner SL, Stroud RM, and Finer-Moore J (2008). Structure of a TrmA–RNA complex: A consensus RNA fold contributes to substrate selectivity and catalysis in m⁵U methyltransferases. *PNAS*, 105(19), 6876-6881.
10. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403-410.
11. Anantharaman V, Koonin EV, and Aravind L (2001). TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol. Lett.*, 197(2), 215-221.

12. Anantharaman V, Koonin EV, and Aravind L (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.*, 30(7), 1427-1464.
13. Anantharaman V, Koonin EV, and Aravind L (2002). SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J. Mol. Microbiol. Biotechnol.*, 4(1), 71-76.
14. Armistead J, Khatkar S, Meyer B, Mark BL, Patel N, Coghlan G, Lamont RE, Liu S, Wiechert J, Cattini PA, and Koetter P (2009). Mutation of a gene essential for ribosome biogenesis, EMG1, causes Bowen-Conradi syndrome. *Am. J. Hum. Genet.*, 84(6), 728-739.
15. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, and Ben-Tal N (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, 44(W1), W344-W350.
16. Auxilien S, El Khadali F, Rasmussen A, Douthwaite S, and Grosjean H (2007). Archease from *Pyrococcus abyssi* improves substrate specificity and solubility of a tRNA m⁵C methyltransferase. *J. Biol. Chem.*, 282(26), 18711-18721.
17. Auxilien S, Rasmussen A, Rose S, Brochier-Armanet C, Husson C, Fourmy D, Grosjean H, and Douthwaite S (2011). Specificity shifts in the rRNA and tRNA nucleotide targets of archaeal and bacterial m⁵U methyltransferases. *RNA.*, 17(1), 45-53.
18. Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, and Terns MP (2005). RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.*, 19(10), 1238-1248.
19. Baker NA, Sept D, Holst MJ, and McCammon JA (2001). The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. *IBM J. Res. Dev.*, 45(3.4), 427-438.
20. Balamurugan B, Roshan MN, Shaahul Hameed B, Sumathi K, Senthilkumar R, Udayakumar A, Venkatesh Babu KH, Kalaivani M, Sowmiya G, Sivasankari P, and Saravanan S (2007). PSAP: protein structure analysis package. *J. Appl. Crystallogr.*, 40(4), 773-777.

21. Ban N, Beckmann R, Cate JH, Dinman JD, Dragon F, Ellis SR, Lafontaine DL, Lindahl L, Liljas A, Lipton JM, and McAlear MA (2014). A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.*, 24, 165-169.
22. Ban N, Nissen P, Hansen J, Moore PB, and Steitz TA (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Sci.*, 289(5481), 905-920.
23. Baranovskiy AG, Siebler HM, Pavlov YI, and Tahirov TH (2018). Iron–sulfur clusters in DNA polymerases and primases of eukaryotes. *Methods Enzymol.*, 599, 1-20.
24. Battye TG, Kontogiannis L, Johnson O, Powell HR, and Leslie AG (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, 67(4), 271-281.
25. Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J, and Blobel G (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell*, 107, 361–372.
26. Ben-Shem A, De Loubresse NG, Melnikov S, Jenner L, Yusupova G, and Yusupov M (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Sci.*, 334(6062), 1524-1529.
27. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, and Wheeler DL (2000). GenBank. *Nucleic Acids Res.*, 28(1), 15-18.
28. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, and Fagan P (2002). The protein data bank. *Acta Crystallogr. D Biol. Crystallogr.*, 58(6), 899-907.
29. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, and Westbrook J (2000). The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, 7(11), 957-959.
30. Bhujbalrao R and Anand R (2019). Deciphering determinants in ribosomal methyltransferases that confer antimicrobial resistance. *J. Am. Chem. Soc.*, 141(4), 1425-1429.
31. Bhujbalrao R, Gavvala K, Singh RK, Singh J, Boudier C, Chakrabarti S, Patwari GN, Mély Y, and Anand R (2022). Identification of allosteric hotspots regulating the ribosomal RNA binding by antibiotic resistance-conferring Erm methyltransferases. *J. Biol. Chem.*, 298(8).

32. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, and Schwede T (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, 42(W1), W252-W258.
33. Binkowski TA, Naghibzadeh S, and Liang J (2003). CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Res.*, 31(13): 3352-3355.
34. Björk GR and Isaksson LA (1970). Isolation of mutants of *Escherichia coli* lacking 5-methyluracil in transfer ribonucleic acid or 1-methylguanine in ribosomal RNA. *J. Mol. Biol.*, 51(1), 83-100.
35. Blanco S and Frye M (2014). Role of RNA methyltransferases in tissue renewal and pathology. *Curr. Opin. Cell Biol.*, 31, 1-7.
36. Blanco S, Bandiera R, Popis M, Hussain S, Lombard P, Aleksic J, Sajini A, Tanna H, Cortés-Garrido R, Gkatza N, and Dietmann S (2016). Stem cell function and stress response are controlled by protein synthesis. *Nature*, 534(7607), 335-340.
37. Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Baginski B, Wirecki TK, de Crecy-Lagard V, Ross R, Limbach PA, Kotter A, Helm M, and Bujnicki JM (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, 46, D303–D307.
38. Boeck LD, Clem GM, Wilson MM, and Westhead JE (1973). A9145, a new adenine-containing antifungal antibiotic: fermentation. *Antimicrob. Agents Chemother.*, 3(1), 49-56.
39. Bohnsack KE, Höbartner C, and Bohnsack MT (2019). Eukaryotic 5-methylcytosine (m⁵C) RNA methyltransferases: mechanisms, cellular functions, and links to disease. *Genes*, 10(2), 102.
40. Bourgeois G, Ney M, Gaspar I, Aigueperse C, Schaefer M, Kellner S, Helm M, and Motorin Y (2015). Eukaryotic rRNA modification by yeast 5-methylcytosine-methyltransferases and human proliferation-associated antigen p120. *PLoS One.*, 10(7), e0133321.
41. Bowman JC, Petrov AS, Frenkel-Pinter M, Penev PI, and Williams LD (2020). Root of the tree: the significance, evolution, and origins of the ribosome. *Chem. Rev.* 120(11), 4848-4878.

42. Bragginton EC and Piddock LJ (2014). UK and European Union public and charitable funding from 2008 to 2013 for bacteriology and antibiotic research in the UK: an observational study. *Lancet Infect. Dis.*, 14(9), 857-868.
43. Brand RC, Klootwijk JA, Planta RJ, and Maden BE (1978). Biosynthesis of a hypermodified nucleotide in *Saccharomyces carlsbergensis* 17S and HeLa-cell 18S ribosomal ribonucleic acid. *Biochem. J.*, 169(1), 71-77
44. Brighenti E, Treré D, and Derenzini M (2015). Targeted cancer therapy with ribosome biogenesis inhibitors: a real possibility?. *Oncotarget*, 6(36), 38617.
45. Brunger AT (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359), 472-475.
46. Brzezicha B, Schmidt M, Makałowska I, Jarmołowski A, Pieńkowska J, and Szweykowska-Kulińska Z (2006). Identification of human tRNA: m⁵C methyltransferase catalysing intron-dependent m⁵C formation in the first position of the anticodon of the anticodon of the pre-tRNA Leu (CAA). *Nucleic Acids Res.*, 34(20), 6034-6043.
47. Buchhaupt M, Meyer B, Kötter P, and Entian KD (2006). Genetic evidence for 18S rRNA binding and an Rps19p assembly function of yeast nucleolar protein Nep1p. *Mol. Genet. Genomics.*, 276, 273-284.
48. Bujnicki JM, Feder M, Ayres CL, and Redman KL (2004). Sequence–structure–function studies of tRNA: m⁵C methyltransferase Trm4p and its relationship to DNA: m⁵C and RNA: m⁵U methyltransferases. *Nucleic Acids Res.*, 32(8), 2453-2463.
49. Caldas T, Binet E, Boulloc P, Costa A, Desgres J, and Richarme G (2000). The FtsJ/RrmJ heat shock protein of *Escherichia coli* is a 23 S ribosomal RNA methyltransferase. *J. Biol. Chem.*, 275(22), 16414-16419.
50. Canaves JM (2004). Predicted role for the archease protein family based on structural and sequence analysis of TM1083 and MTH1598, two proteins structurally characterized through structural genomics efforts. *Proteins: Struct. Funct. Bioinfo.*, 56(1), 19-27.
51. Cantoni GL (1975). Biological methylation: selected aspects. *Annu. Rev. Biochem.*, 44(1), 435-451.

52. Chandramouli P, Topf M, Menetret JF, Eswar N, Cannone JJ, Gutell RR, Sali A, and Akey CW (2008). Structure of the mammalian 80S ribosome at 8.7 Å resolution. *Struct.*, 16, 535–548.
53. Charette M and Gray MW (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB life*, 49(5), 341-352.
54. Chatterjee A, Cui Y, and Chatterjee AK (2002). RsmA and the quorum-sensing signal, N-[3-oxohexanoyl]-L-homoserine lactone, control the levels of rsmB RNA in *Erwinia carotovora* subsp. *carotovora* by affecting its stability. *J. Bacteriol.*, 184(15), 4089-4095.
55. Chen J, Sawyer N, and Regan L (2013). Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Prot. Sci.*, 22(4), 510-515.
56. Cheng X and Roberts RJ (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.*, 29(18), 3784-3795.
57. Chi L and Delgado-Olguín P (2013). Expression of NOL1/NOP2/sun domain (Nsun) RNA methyltransferase family genes in early mouse embryogenesis. *Gene Expr. Patterns.*, 13(8), 319-327.
58. Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, and McCann PP (1996). S-Adenosylmethionine and methylation. *FASEB J.*, 10(4), 471-480.
59. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, and Sayers EW (2016). GenBank. *Nucleic Acids Res.*, 44(D1), 67-72.
60. Condon C, Liveris D, Squires C, Schwartz IR, and Squires CL (1995). rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rrn* inactivation. *J. Bacteriol.*, 177(14), 4152-4156.
61. Connolly K and Culver G (2009). Deconstructing ribosome construction. *Trends Biochem Sci.*, 34, 256–263.
62. Connolly K, Rife JP, and Culver G (2008). Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Mol. Microbiol.*, 70(5), 1062-1075.
63. Conrad J, Sun D, Englund N, and Ofengand J (1998). The *rluC* gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for

- synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. *J. Biol. Chem.*, 273(29), 18562-18566.
64. Cotney J and Shadel GS (2006). Evidence for an early gene duplication event in the evolution of the mitochondrial transcription factor B family and maintenance of rRNA methyltransferase activity in human mtTFB1 and mtTFB2. *J. Mol. Evol.*, 63, 707-717.
65. Cotney J, McKay SE, and Shadel GS (2009). Elucidation of separate, but collaborative functions of the rRNA methyltransferase-related human mitochondrial transcription factors B1 and B2 in mitochondrial biogenesis reveals new insight into maternally inherited deafness. *Hum. Mol. Genet.*, 18(14), 2670-2682.
66. Coureux PD, Lazennec-Schurdevin C, Bourcier S, Mechulam Y, and Schmitt E (2020). Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Commun. Biol.*, 3(1), 1-3.
67. De Stasio EA, Moazed D, Noller HF, and Dahlberg AE (1989). Mutations in 16S ribosomal RNA disrupt antibiotic-RNA interactions. *EMBO J.*, 8(4), 1213-1216.
68. Decatur WA and Fournier MJ (2002). rRNA modifications and ribosome function. *Trends Biochem. Sci.*, 27(7), 344-351.
69. Deller MC, Kong L, and Rupp B (2016). Protein stability: a crystallographer's perspective. *Acta Crystallogr. Sect. F Struct. Biol. Cryst.*, 72(2), 72-95.
70. Demirci H, Belardinelli R, Seri E, Gregory ST, Gualerzi C, Dahlberg AE, and Jögl G (2009). Structural rearrangements in the active site of the *Thermus thermophilus* 16S rRNA methyltransferase KsgA in a binary complex with 5'-methylthioadenosine. *J. Mol. Biol.*, 388(2), 271-282.
71. Demirci H, Larsen LH, Hansen T, Rasmussen A, Cadambi A, Gregory ST, Kirpekar F, and Jögl G (2010). Multi-site-specific 16S rRNA methyltransferase RsmF from *Thermus thermophilus*. *RNA*, 16(8), 1584-1596.
72. Demirci H, Larsen LH, Hansen T, Rasmussen A, Cadambi A, Gregory ST, Kirpekar F, and Jögl G (2010). Multi-site-specific 16S rRNA methyltransferase RsmF from *Thermus thermophilus*. *RNA*, 16(8), 1584-1596.
73. Demirci H, Murphy F, Belardinelli R, Kelley AC, Ramakrishnan V, Gregory ST, Dahlberg AE, and Jögl G (2010). Modification of 16S ribosomal RNA by

- the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function. *RNA*, 16(12), 2319-2324.
74. Deng JH, Luo J, Mao YL, Lai S, Gong YN, Zhong DC, and Lu TB (2020). π - π stacking interactions: Non-negligible forces for stabilizing porous supramolecular frameworks. *Sci. Adv.*, 6(2), eaax9976.
75. Dennis PP, Tripp V, Lui L, Lowe T, and Randau L (2015). C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. *BMC Genom.*, 16, 1-2.
76. Desai KK, Cheng CL, Bingman CA, Phillips Jr GN, and Raines RT (2014). A tRNA splicing operon: Archease endows RtcB with dual GTP/ATP cofactor specificity and accelerates RNA ligation. *Nucleic Acids Res.*, 42(6), 3931-3942.
77. Desmolaize B, Fabret C, Brégeon D, Rose S, Grosjean H, and Douthwaite S (2011). A single methyltransferase YefA (RlmCD) catalyses both m⁵U747 and m⁵U1939 modifications in *Bacillus subtilis* 23S rRNA. *Nucleic Acids Res.*, 39(21), 9368-9375.
78. Doi Y and Arakawa Y (2007). 16S ribosomal RNA methylation: emerging resistance mechanism against aminoglycosides. *Clin. Infect. Dis.*, 45(1), 88-94.
79. Dominguez C and Allain FH. Resonance assignments of the two N-terminal RNA recognition motifs (RRM) of the human heterogeneous nuclear ribonucleoprotein F (HnRNP F). *J. Biomol. NMR*, 33(4), 282-283.
80. Dubnau D and Nomura M (1984). Translational Attenuation: The Regulation of Bacterial Resistance to the Macrolide-Lincosamide-Streptogramin B Antibiotic. *Crit. Rev. Biochem. Mol. Biol.*, 16(2), 103-132.
81. Ebersberger I, Simm S, Leisegang MS, Schmitzberger P, Mirus O, von Haeseler A, Bohnsack MT, and Schleiff E (2014). The evolution of the ribosome biogenesis pathway from a yeast perspective. *Nucleic acids res.*, 42(3), 1509-1523.
82. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, and Sorek R (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m⁵C within archaeal mRNAs. *PLoS Genet.*, 9(6), e1003602.

83. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, and Sonnhammer EL (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.*, 47(D1), D427-D432.
84. Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66(4), 486-501.
85. Eschrich D, Buchhaupt M, Kötter P, and Entian KD (2002). Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis. *Curr. Genet.*, 40, 326-338.
86. Evans PR (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.*, 67(4), 282-292.
87. Evans PR and Murshudov GN (2013). How good are my data and what is the resolution?. *Acta Crystallogr. D Biol. Crystallogr.*, 69(7), 1204-1214.
88. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, and Stark H (2015). Structure of the *E. coli* ribosome–EF-Tu complex at < 3 Å resolution by Cs-corrected cryo-EM. *Nature*, 520(7548), 567-570.
89. Foster PG, Nunes CR, Greene P, Moustakas D, and Stroud RM (2003). The first structure of an RNA m⁵C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate. *Struct.*, 11(12), 1609-1620.
90. François B, Russell RJ, Murray JB, Aboul-ela F, Masquida B, Vicens Q, and Westhof E (2005). Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res.*, 33(17), 5677-5690.
91. Fukuda A, Kodama Y, Mashima J, Fujisawa T, and Ogasawara O (2021). DDBJ update: streamlining submission and access of human data. *Nucleic Acids Res.*, 49(D1), D71-5.
92. Fuller RW and Nagarajan R (1978). Inhibition of methyltransferases by some new analogs of S-adenosylhomocysteine. *Biochemic. Pharmacol.*, 27(15), 1981-1983.
93. Fuss JO, Tsai CL, Ishida JP, and Tainer JA (2015). Emerging critical roles of Fe–S clusters in DNA replication and repair. *Biochim. Biophys. Acta, Mol. Cell Res.*, 1853(6), 1253-1271.

94. Gabashvili IS, Agrawal RK, Grassucci R, Squires CL, Dahlberg AE, and Frank J (1999). Major rearrangements in the 70S ribosomal 3D structure caused by a conformational switch in 16S ribosomal RNA. *EMBO J.*, 18(22), 6501-6507.
95. Gao H, Sengupta J, Valle M, Korostelev A, Eswar N, Stagg SM, Van Roey P, Agrawal RK, Harvey SC, Sali A, and Chapman MS (2003). Study of the structural dynamics of the *E. coli* 70S ribosome using real-space refinement. *Cell*, 113(6), 789-801.
96. Gao YG, Yao M, Yong Z, and Tanaka I (2005). Crystal structure of the putative RNA methyltransferase PH1948 from *Pyrococcus horikoshii*, in complex with the copurified S-adenosyl-L-homocysteine. *Proteins: Struct. Funct. Bioinf.*, 61(4), 1141-1145.
97. Garcia-Ortega L, Stephen J, and Joseph S (2008). Precise alignment of peptidyl tRNA by the decoding center is essential for EF-G-dependent translocation. *Mol. Cell*, 32(2), 292-299.
98. Garcia-Vallvé S, Palau J, and Romeu A (1999). Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Mol. Biol. Evol.*, 16(9), 1125-1134.
99. Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, and Bairoch A (2005). Protein identification and analysis tools on the ExPASy server. Humana press, 571-607.
100. Gasteiger J and Marsili M (1980). Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, 36(22), 3219-3228.
101. Gentile F, Amodeo P, Febbraio F, Picaro F, Motta A, Formisano S, and Nucci R (2002). SDS-resistant active and thermostable dimers are obtained from the dissociation of homotetrameric β -glycosidase from hyperthermophilic *Sulfolobus solfataricus* in SDS: stabilizing role of the AC intermonomeric interface. *J. Biol. Chem.*, 277(46), 44050-44060.
102. Ghosh A and Komar AA (2015). Eukaryote-specific extensions in ribosomal proteins of the small subunit: Structure and function. *Translation*, 3(1), e999576.
103. Glotz C, Zwieb C, Brimacombe R, Edwards K, and Kössel H (1981). Secondary structure of the large subunit ribosomal RNA from *Escherichia coli*, *Zea mays*

- chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic Acids Res.*, 9(14), 3287-3306.
104. Gogoi P, Mordina P, and Kanaujia SP (2019). Structural insights into the catalytic mechanism of 5-methylthioribose 1-phosphate isomerase. *J. Struct. Biol.*, 205(1), 67-77.
105. Golovanov AP, Hautbergue GM, Wilson SA, and Lian LY (2004). A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.*, 126(29), 8933-8939.
106. Gopal B, Haire LF, Cox RA, Colston MJ, Major S, Brannigan JA, Smerdon SJ, and Dodson G (2000). The crystal structure of NusB from *Mycobacterium tuberculosis*. *Nat. Struct. Biol.*, 7(6), 475-478.
107. Görisch H (1988). Drop dialysis: time course of salt and protein exchange. *Analytical biochemistry*, 173(2), 393-398.
108. Gouet P, Robert X, and Courcelle E (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, 31(13), 3320-3323.
109. Greń BA, Dabrowski-Tumanski P, Niemyska W, and Sulkowska JI (2021). Lasso Proteins—Unifying Cysteine Knots and Miniproteins. *Polym.*, 13(22), 3988.
110. Grosjean H (2009). Nucleic acids are not boring long polymers of only four types of nucleotides. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes: Bioscience, 1–18.
111. Grosjean H, Gaspin C, Marck C, Decatur WA, and de Crécy-Lagard V (2008). RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC genom.*, 9(1), 1-26.
112. Gupta RS and Griffiths E (2002). Critical issues in bacterial phylogeny. *Theor. Popul. Biol.*, 61(4), 423-434.
113. Gustafson WC, Taylor CW, Valdez BC, Henning D, Phippard A, Ren Y, Busch H, and Durban E (1998). Nucleolar protein p120 contains an arginine-rich domain that binds to ribosomal RNA. *Biochem. J.*, 331(2), 387-393.
114. Gustafsson C, Reid R, Greene PJ, and Santi DV (1996). Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucleic acids res.*, 24(19), 3756-3762.

- 115.Hage AE and Tollervey D (2004). A surfeit of factors: Why is ribosome assembly so much more complicated in eukaryotes than bacteria? *RNA Biol.*, 1(1), 10–15.
- 116.Hallberg BM, Ericsson UB, Johnson KA, Andersen NM, Douthwaite S, Nordlund P, Beuscher IV AE, and Erlandsen H (2006). The structure of the RNA m⁵C methyltransferase YebU from *Escherichia coli* reveals a C-terminal RNA-recruiting PUA domain. *J. Mol. Biol.*, 360(4), 774-787.
117. Harms J, Schlutzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, and Yonath A (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, 107(5), 679-688.
- 118.Hassanali A, Giberti F, Cuny J, Kühne TD, and Parrinello M (2013). Proton transfer through the water gossamer. *Proc. Natl. Acad. Sci. U.S.A.*, 110(34), 13723-13728.
119. Hassell AM, And G, Bledsoe RK, Bynum JM, Carter HL, Deng SJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT, and Rocque WJ (2007). Crystallization of protein–ligand complexes. *Acta Crystallogr. D.*, 63(1), 72-79.
- 120.Hayward S and Berendsen HJ (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins: Structure, Function, and Bioinformatics*, 30(2), 144-154.
- 121.He J, Tao H, and Huang SY (2019). Protein-ensemble–RNA docking by efficient consideration of protein flexibility through homology models. *Bioinform.*, 2019 35(23), 4994-5002.
- 122.Helser TL, Davies JE, and Dahlberg JE (1971). Change in methylation of 16S ribosomal RNA associated with mutation to kasugamycin resistance in *Escherichia coli*. *Nature. New Biol.*, 233, 12-14.
- 123.Helser TL, Davies JE, and Dahlberg JE (1972). Mechanism of kasugamycin resistance in *Escherichia coli*. *Nat. New Biol.*, 235, 6-9.
- 124.Hetényi C and Spoel DV (2011). Toward prediction of functional protein pockets using blind docking and pocket search algorithms. *Protein Sci.*, 20(5), 880-893.
- 125.Hikida Y, Kuratani M, Bessho Y, Sekine SI, and Yokoyama S (2010). Structure of an archaeal homologue of the bacterial Fmu/RsmB/RrmB rRNA cytosine 5-methyltransferase. *Acta Crystallogr. D Struct. Biol.*, 66(12), 1301-1307.

- 126.Hirabayashi N, Sato NS, and Suzuki T (2006). Conserved loop sequence of helix 69 in *Escherichia coli* 23 S rRNA is involved in A-site tRNA binding and translational fidelity. *J. Biol. Chem.*, 281(25), 17203-17211.
- 127.Holm L (2020). Using Dali for Protein Structure Comparison. *Methods Mol. Biol.*, 2112, 29-42.
- 128.Hsiao K, Zegzouti H, and Goueli SA (2016). Methyltransferase-Glo: a universal, bioluminescent and homogenous assay for monitoring all classes of methyltransferases. *Epigenomics*, 8(3), 321-339.
- 129.Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, and Stetter KO (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, 417(6884), 63-67.
- 130.Hussain AS, Shanthi V, Sheik SS, Jeyakanthan J, Selvarani P, and Sekar K (2002). PDB Goodies—a web-based GUI to manipulate the Protein Data Bank file. *Acta Crystallogr. D Biol. Crystallogr.*, 58(8), 1385-1386.
- 131.Iida S and Fukunishi Y (2021). Asymmetric dynamics of dimeric SARS-CoV-2 and SARS-CoV main proteases in an apo form: Molecular dynamics study on fluctuations of active site, catalytic dyad, and hydration water. *BBA Adv.*, 1, 100016.
- 132.Imlay JA (2006). Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.*, 59(4), 1073-1082.
- 133.Ishikawa I, Sakai N, Tamura T, Yao M, Watanabe N, and Tanaka I (2004). Crystal structure of human p120 homologue protein PH1374 from *Pyrococcus horikoshii*. *Proteins: Struct. Funct. Bioinf.*, 54(4), 814-816.
- 134.Jedrzejewski M, Belza B, Lewandowska I, Sadlej M, Perlinska AP, Augustyniak R, Christian T, Hou YM, Kalek M, and Sulkowska JI (2023). Nucleolar Essential Protein 1 (Nep1): Elucidation of enzymatic catalysis mechanism by molecular dynamics simulation and quantum mechanics study. *Comput. Struct. Biotechnol. J.*, 21, 3999-4008.
- 135.Jiang Y, Li F, Wu J, Shi Y, and Gong Q (2017). Structural insights into substrate selectivity of ribosomal RNA methyltransferase RlmCD. *PloS one.*, 12(9), e0185226.

- 136.Jiang Y, Yu H, Li F, Cheng L, Zhu L, and Shi Y, and Gong Q (2018). Unveiling the structural features that determine the dual methyltransferase activities of *Streptococcus pneumoniae* RlmCD. PLoS Pathog., 14(11), e1007379.
- 137.Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, and Bridgland A (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.
- 138.Kaczanowska M and Rydén-Aulin M (2007). Ribosome biogenesis and the translation process in *Escherichia coli*. Microbiol. Mol. Biol., 71(3), 477-494.
- 139.Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, and Xu J (2012). Template-based protein structure modeling using the RaptorX web server. Nature Protoc., 7(8), 1511-1522.
- 140.Kanehisa M and Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 28(1), 27-30.
- 141.Katz BA, and Kossiakoff A (1986). The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. J. Biol. Chem., (33), 15480-15485.
- 142.Kealey JT, Gu X, and Santi DV (1994). Enzymatic mechanism of tRNA (m⁵U54) methyltransferase. Biochim., 76(12), 1133-1142.
- 143.Keffer-Wilkes LC, Soon EF, and Kothe U (2020). The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain. Nucleic Acids Res., 48(14), 7981-7990.
- 144.Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJ (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protoc., 10(6), 845-858.
- 145.Khade PM and Jernigan RL (2022). Entropies Derived from the Packing Geometries within a Single Protein Structure. ACS omega, 7(24), 20719-20730.
- 146.Khade PM, Scaramozzino D, Kumar A, Lacidogna G, Carpinteri A, and Jernigan RL (2021). hdANM: a new comprehensive dynamics model for protein hinges. Biophys. J., 120(22), 4955-4965.
- 147.Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, and Zaslavsky L (2021). PubChem in 2021: new data

- content and improved web interfaces. *Nucleic Acids Res.*, 49(D1), D1388-D1395.
148. Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, and Kiss T (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, 85(7), 1077-1088.
149. Ko J, Park H, Heo L, and Seok C (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.*, 40(W1), W294-W297.
150. Koch AL (2003). Were Gram-positive rods the first bacteria?. *Trends Microbiol.*, 11(4), 166-170.
151. Kowalak JA, Bruenger E, Crain PF, and McCloskey JA (2000). Identities and phylogenetic comparisons of posttranscriptional modifications in 16 S ribosomal RNA from *Haloferax volcanii*. *J. Biol. Chem.*, 275(32), 24484-24489.
152. Kozakov D, Hall DR, Xia B, Porter KA, Padhorney D, Yueh C, Beglov D, and Vajda S (2017). The ClusPro web server for protein–protein docking. *Nature Protoc.*, 12(2), 255-278.
153. Krishnamohan A and Jackman JE (2018). A family divided: distinct structural and mechanistic features of the SpoU-TrmD (SPOUT) methyltransferase superfamily. *Biochemistry*, 58(5), 336-345.
154. Krissinel E and Henrick K (2007). Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. *J. Mol. Biol.*, 372(3), 774-797.
155. Krzyzosiak W, Denman R, Nurse K, Hellmann W, Boublik M, Gehrke CW, Agris PF, and Ofengand J (1987). In vitro synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochem.*, 26(8), 2353-2364.
156. Kubarenko A, Sergiev P, Wintermeyer W, Dontsova O, and Rodnina MV (2006). Involvement of helix 34 of 16 S rRNA in decoding and translocation on the ribosome. *J. Biol. Chem.*, 281, 35235–35244.
157. Kumar A, Khade PM, Dorman KS, and Jernigan RL (2022). Coarse-graining protein structures into their dynamic communities with DCI, a dynamic community identifier. *Bioinform.*, 38(10), 2727-2733.

158. Kumar J, Ranjan T, Kumar RR, Ansar M, Rajani K, Kumar M, Kumar V, and Kumar A (2019). In silico characterization and homology Modelling of Potato Leaf Roll Virus (PLRV) coat protein. *Curr. J. Appl. Sci. Technol.*, 33(2), 1-8.
159. Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, 35(6), 1547.
160. Kuratani M, Bessho Y, Nishimoto M, Grosjean H, and Yokoyama S (2008). Crystal structure and mutational study of a unique SpoU family archaeal methylase that forms 2'-O-methylcytidine at position 56 of tRNA. *J. Mol. Biol.*, 375(4), 1064-1075.
161. Kuznetsova SA, Petrukov KS, Pletnev FI, Sergiev PV, and Dontsova OA (2019). RNA (C5-cytosine) Methyltransferases. *Biochemistry (Mosc)*, 84(8), 851-869.
162. Lafontaine D, Delcour J, Glasser AL, Desgrès J, and Vandenhoute J (1994). The DIM1 gene responsible for the conserved m62Am62A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast. *J. Mol. Biol.*, 241(3), 492-497.
163. Lafontaine D, Vandenhoute J, and Tollervey D (1995). The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes Dev.*, 9(20), 2470-2481.
164. Lafontaine DL, and Tollervey D (1998). Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.*, 23(10), 383-388.
165. Laskowski RA, MacArthur MW, Moss DS, and Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26(2), 283-291.
166. Lecompte O, Ripp R, Thierry JC, Moras D, and Poch O (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.*, 30(24), 5382-90.
167. Lee TT, Agarwalla S, and Stroud RM (2004). Crystal structure of RumA, an iron-sulfur cluster containing *E. coli* ribosomal RNA 5-methyluridine methyltransferase. *Struct.*, 12(3), 397-407.

- 168.Lee TT, Agarwalla S, and Stroud RM (2005). A unique RNA Fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell*, 120(5), 599-611
- 169.Lestrade L and Weber MJ (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, 34(suppl_1), D158-162.
- Li J, Li H, Long T, Dong H, Wang ED and Liu RJ (2019). Archaeal NSUN6 catalyzes m⁵C72 modification on a wide-range of specific tRNAs. *Nucleic Acids Res.*, 47(4), 2041-2055.
- 170.Letunic I and Bork P (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, 49(W1), W293-W296.
- 171.Leulliot N, Bohnsack MT, Graille M, Tollervey D, and Van Tilbeurgh H (2008). The yeast ribosomesynthesis factor emg1 is a novel member of the superfamily of alpha/beta knotfold methyltransferases. *Nucleic Acids Res.*, 36(2), 629–639.
- 172.Li J, Li H, Long T, Dong H, Wang ED, and Liu RJ (2019). Archaeal NSUN6 catalyzes m⁵C72 modification on a wide-range of specific tRNAs. *Nucleic Acids Res.*, 47(4), 2041-2055.
- 173.Liang XH, Liu Q, and Fournier MJ (2009). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA*, 15(9), 1716-1728.
- 174.Liti S (2016). Evaluation of sample preparation techniques for MALDI-TOF-MS analysis of oligosaccharides.
- 175.Liu J, Eckert MA, Harada BT, Liu SM, Lu Z, Yu K, Tienda SM, Chryplewicz A, Zhu AC, Yang Y, and Huang JT (2018). m⁶A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat. Cell Biol.*, 20(9), 1074-1083.
- 176.Liu M and Douthwaite S (2002a). Methylation at nucleotide G745 or G748 in 23S rRNA distinguishes Gram-negative from Gram-positive bacteria. *Mol. Microbiol.*, 44(1), 195-204.
- 177.Liu M and Douthwaite S (2002b). Resistance to the macrolide antibiotic tylosin is conferred by single methylations at 23S rRNA nucleotides G748 and A2058 acting in synergy. *Proc. Natl. Acad. Sci. U.S.A.*, 99(23), 14658-14663.

- 178.Liu PC and Thiele DJ (2001). Novel stress-responsive genes EMG1 and NOP14 encode conserved, interacting proteins required for 40S ribosome biogenesis. *Mol. Biol. Cell*, 12(11), 3644-3657.
- 179.Liu RJ, Long T, Li J, Li H, and Wang ED (2017). Structural basis for substrate binding and catalytic mechanism of a human RNA: m⁵C methyltransferase NSun6. *Nucleic Acids Res.*, 45(11), 6684-6697.
- 180.Liu X, Shen S, Wu P, Li F, Liu X, Wang C, Gong Q, Wu J, Yao X, Zhang H, and Shi Y (2019). Structural insights into dimethylation of 12S rRNA by TFB1M: indispensable role in translation of mitochondrial genes and mitochondrial function. *Nucleic Acids Res.*, 47(14), 7648-7665.
- 181.Lodmell JS and Dahlberg AE (1997). A conformational switch in *Escherichia coli* 16 S ribosomal RNA during decoding of messenger RNA. *Sci.*, 277(5330), 1262-1267.
- 182.Londei P and Ferreira-Cerca S (2021). Ribosome biogenesis in archaea. *Front. Microbiol.*, 12, 1476.
- 183.Lövgren JM and Wikström PM (2001). The rlmB gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in *Escherichia coli*. *J. Bacteriol.*, 183(23), 6957-6960.
- 184.Lowe TM and Eddy SR (1999). A computational screen for methylation guide snoRNAs in yeast. *Sci.*, 283(5405), 1168-1171.
- 185.Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, and Helm M (2012). MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.*, 41(D1), D262-D267.
- 186.Macmaster R, Zelinskaya N, Savic M, Rankin CR, and Conn GL (2010). Structural insights into the function of aminoglycoside-resistance A1408 16S rRNA methyltransferases from antibiotic-producing and human pathogenic bacteria. *Nucleic Acids Res.*, 38(21), 7791-7799.
- 187.Madsen CT, Mengel-Jørgensen J, Kirpekar F, and Douthwaite S (2003). Identifying the methyltransferases for m⁵U747 and m⁵U1939 in 23S rRNA using MALDI mass spectrometry. *Nucleic Acids Res.*, 31(16):4738-4746.

- 188.Mandal SK and Kanaujia SP (2021). Structural and thermodynamic insights into a novel Mg^{2+} -citrate-binding protein from the ABC transporter superfamily. *Acta Crystallogr. D Struct. Biol.*, 77(12), 1516-1534.
- 189.Mangat CS and Brown ED (2008). Ribosome biogenesis; the KsgA protein throws a methyl-mediated switch in ribosome assembly. *Mol. Microbiol.*, 70(5), 1051-1053.
- 190.Marcel V, Ghayad SE, Belin S, Therizols G, Morel AP, Solano-González E, Vendrell JA, Hacot S, Mertani HC, Albaret MA, and Bourdon JC(2013). p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer cell*, 24(3), 318-330.
- 191.McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, 40(4), 658-674.
- 192.Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, and Yusupov M (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*, 19(6), 560-567.
- 193.Melnyk S, Pogribna M, Pogribny IP, Yi P, and James SJ (2000). Measurement of plasma and intracellular S-adenosylmethionine and S-adenosylhomocysteine utilizing coulometric electrochemical detection: alterations with plasma homocysteine and pyridoxal 5'-phosphate concentrations. *Clin. Chem.*, 46(2), 265-272.
- 194.Metodiev MD, Lesko N, Park CB, Cámara Y, Shi Y, Wibom R, Hultenby K, Gustafsson CM, and Larsson NG (2009). Methylation of 12S rRNA is necessary for in vivo stability of the small subunit of the mammalian mitochondrial ribosome. *Cell Metab.*, 9(4), 386-397.
- 195.Meyer B, Wurm JP, Kötter P, Leisegang MS, Schilling V, Buchhaupt M, Held M, Bahr U, Karas M, Heckel A, and Bohnsack MT (2011). The Bowen–Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Ψ 1191 in yeast 18S rRNA. *Nucleic Acids Res.*, 39(4), 1526-1537.
- 196.Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, and Gough J (2019). InterPro in 2019:

- improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, 47(D1), D351-D360.
197. Moazed D and Noller HF (1989). Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell*, 57(4), 585-597.
198. Morić I, Savić M, Ilić-Tomić T, Vojnović S, Bajkić S, and Vasiljević B (2010). rRNA methyltransferases and their role in resistance to antibiotics. *J. Med. Biochem.*, 29(3), 165-174.
199. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, and Olson AJ (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30(16), 2785-2791.
200. Motorin Y, Lyko F, and Helm M (2010). 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res.*, 38(5), 1415-1430.
201. Müller I (2017). Guidelines for the successful generation of protein–ligand complex crystals. *Acta Crystallogr. D Biol. Crystallogr.*, 73(2), 79-92.
202. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, and Vagin AA (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, 67(4), 355-367.
203. Nakano S, Suzuki T, Kawarada L, Iwata H, Asano K, and Suzuki T (2016). NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA Met. *Nat. Chem. Biol.*, 12(7), 546-551.
204. Nazar R (2004). Ribosomal RNA processing and ribosome biogenesis in eukaryotes. *IUBMB life*, 56(8), 457-465.
205. Ni J, Tien AL, and Fournier MJ (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, 89(4), 565-573.
206. Niemyska W, Dabrowski-Tumanski P, Kadlof M, Haglund E, Sułkowski P, and Sulkowska JI (2016). Complex lasso: new entangled motifs in proteins. *Sci. Rep.*, 6(1), 36895.
207. Nierhaus KH (1991). The assembly of prokaryotic ribosomes. *Biochim.*, 73(6), 739-755.
208. Nishimoto M, Higashijima K, Shirouzu M, Grosjean H, Bessho Y, and Yokoyama S (2008). Crystal structure of tRNA N2, N2-guanosine

- dimethyltransferase Trm1 from *Pyrococcus horikoshii*. J. Mol. Biol., 383(4), 871-884.
209. Noller HF (2006). Biochemical characterization of the ribosomal decoding site. Biochim., 88(8), 935-941.
210. Nooren IM and Thornton JM (2003). Diversity of protein–protein interactions. EMBO J., 22(14), 3486-3492.
211. Nordlund ME, JOHANSSON JM, von Pawel-Rammingen U, and Byström AS (2000). Identification of the TRM2 gene encoding the tRNA (m⁵U54) methyltransferase of *Saccharomyces cerevisiae*. RNA., 6(6), 844-860.
212. O'Farrell HC, Musayev FN, Scarsdale JN, and Rife JP (2010). Binding of adenosine-based ligands to the MjDim1 rRNA methyltransferase: implications for reaction mechanism and drug design. Biochemistry, 49(12), 2697-2704.
213. Ochi K, Kim JY, Tanaka Y, Wang G, Masuda K, Nanamiya H, Okamoto S, Tokuyama S, Adachi Y, and Kawamura F (2009). Inactivation of KsgA, a 16S rRNA methyltransferase, causes vigorous emergence of mutants with high-level kasugamycin resistance. J. Antimicrob. Agents., 53(1):193-201.
214. O'Farrell HC, Pulicherla N, Desai PM, and Rife JP (2006). Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. RNA, 12(5), 725-733.
215. O'Farrell HC, Scarsdale JN, and Rife JP (2004). Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in *Escherichia coli*. J. Mol. Biol., 339(2), 337-353.
216. Ogle JM, Brodersen DE, Clemons Jr WM, Tarry MJ, Carter AP, and Ramakrishnan V (2001). Recognition of cognate transfer RNA by the 30 S ribosomal subunit. Sci., 292(5518), 897-902.
217. Parks LW and Schlenk F (1958). The stability and hydrolysis of S-adenosyl-methionine; isolation of S-ribosylmethionine. J. Biol. Chem., 230(1), 295-305.
218. Pernodet JL, Fish S, Blondelet-Rouault MH, and Cundliffe E (1996). The macrolide-lincosamide-streptogramin B resistance phenotypes characterized by using a specifically deleted, antibiotic-sensitive strain of *Streptomyces lividans*. Antimicrob. Agents Chemother., 40(3), 581-585.

219. Piekna-Przybylska D, Decatur WA, and Fournier MJ (2007). The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.*, 36(D1), D178-183.
220. Poldermans B, Roza L, and Van Knippenberg PH (1979). Studies on the function of two adjacent N6, N6-dimethyladenosines near the 3' end of 16 S ribosomal RNA of *Escherichia coli*. III. Purification and properties of the methylating enzyme and methylase-30 S interactions. *J. Biol. Chem.*, 254(18), 9094-9100.
221. Polikanov YS, Melnikov SV, Söll D, and Steitz TA (2015). Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.*, 22(4), 342-344.
222. Prakash A and Doublie S (2015). Base excision repair in the mitochondria. *Cell. Biochem.*, 116(8), 1490-1499.
223. Puig S, Ramos-Alonso L, Romero AM, and Martínez-Pastor MT (2017). The elemental role of iron in DNA synthesis and repair. *Metallomics*, 9(11):1483-1500.
224. Pulicherla N, Pogorzala LA, Xu Z, Heather CO, Musayev FN, Scarsdale JN, Sia EA, Culver GM, and Rife JP (2009). Structural and functional divergence within the Dim1/KsgA family of rRNA methyltransferases. *J. Mol. Biol.*, 391(5), 884-893.
225. Recht MI, Douthwaite S, Dahlquist KD, and Puglisi JD (1999). Effect of mutations in the A site of 16 S rRNA on aminoglycoside antibiotic-ribosome interaction. *J. Mol. Biol.*, 286(1), 33-43.
226. Reid R, Greene PJ, and Santi DV (1999). Exposition of a family of RNA m⁵C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Res.*, 27(15), 3138-3145.
227. Richardson EJ and Watson M (2013). The automatic annotation of bacterial genomes. *Brief. Bioinformatics*, 14(1), 1-2.
228. RNAcentral Consortium (2021). RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.*, 49(D1), D212-D220.

- 229.Roberts MC, Sutcliffe J, Courvalin P, Jensen LB, Rood J, and Seppala H (1999). Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. *Antimicrob. Agents Chemother.*, 43(12), 2823-2830.
- 230.Rodnina MV, Daviter T, Gromadski K, and Wintermeyer W (2002). Structural dynamics of ribosomal RNA during decoding on the ribosome. *Biochim.*, 84(8), 745-754.
- 231.Rother M, Milanowska K, Puton T, Jeleniewicz J, Rother K, and Bujnicki JM (2011). ModeRNA server: an online tool for modeling RNA 3D structures. *Bioinform.*, 27(17), 2441-2442.
- 232.Rouault TA (2015). Iron-sulfur proteins hiding in plain sight. *Nat. Chem. Biol.*, 11(7), 442-445.
- 233.Sahoo T, Del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, Garnica A, Cheung SW, and Beaudet AL (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.*, 40(6), 719-721.
- 234.Saraiya AA, Lamichhane TN, Chow CS, SantaLucia Jr J, and Cunningham PR (2008). Identification and role of functionally important motifs in the 970 loop of *Escherichia coli* 16S ribosomal RNA. *J. Mol. Biol.*, 376(3), 645-657.
- 235.Sas-Chen A, Thomas JM, Matzov D, Taoka M, Nance KD, Nir R, Bryson KM, Shachar R, Liman GL, Burkhardt BW, and Gamage ST (2020). Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature*, 583(7817), 638-643.
- 236.Schaefer M (2015). RNA 5-methylcytosine analysis by bisulfite sequencing. *Meth. Enzymol.*, 560, 297-329.
- 237.Schaefer M, Kapoor U, and Jantsch MF (2017). Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome'. *Open Biol.*, 7(5), 170077.
- 238.Schluckebier G, Kozak M, Bleimling N, Weinhold E, and Saenger W (1997). Differential binding of S-adenosylmethionine S-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M. TaqI. *J. Mol. Biol.*, 265(1), 56-67.

239. Schluckebier G, O'Gara M, Saenger W, and Cheng X (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.*, 247(1), 16-20.
240. Schluckebier G, Zhong P, Stewart KD, Kavanaugh TJ, and Abad-Zapatero C (1999). The 2.2 Å structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: implications for the reaction mechanism. *J. Mol. Biol.*, 289(2), 277-291.
241. Schosserer M, Minois N, Angerer TB, Amring M, Dellago H, Harreither E, Calle-Perez A, Pircher A, Gerstl MP, Pfeifenberger S, and Brandl C (2015). Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan. *Nature Comm.*, 6(1), 1-7.
242. Schubert HL, Blumenthal RM, and Cheng X (2003). Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.*, 28(6), 329-335.
243. Selmer M, Dunham CM, Murphy IV FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, and Ramakrishnan V (2006). Structure of the 70 S ribosome complexed with mRNA and tRNA. *Sci.*, 313(5795), 1935-1942.
244. Sergeeva OV, Bogdanov AA, and Sergiev PV (2015). What do we know about ribosomal RNA methylation in *Escherichia coli*?. *Biochim.*, 117, 110-8.
245. Sergiev PV, Aleksashin NA, Chugunova AA, Polikanov YS, and Dontsova OA (2018). Structural and evolutionary insights into ribosomal RNA methylation. *Nat. Chem. Biol.*, 14(3), 226-235.
246. Sergiev PV, Bogdanov AA, and Dontsova OA (2007). Ribosomal RNA guanine-(N2)-methyltransferases and their targets. *Nucleic Acids Res.*, 35(7), 2295-2301.
247. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11), 2498-2504.
248. Sharma S and Entian KD (2022). Chemical Modifications of Ribosomal RNA. *Ribosome Biogenesis*, 1, 149.
249. Sharma S and Lafontaine DL (2015). 'View from a bridge': a new perspective on eukaryotic rRNA base modification. *Trends Biochem. Sci.*, 40(10), 560-575.

- 250.Sharma S, Yang J, Watzinger P, Kötter P, and Entian KD (2013). Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic Acids Res.*, 41(19), 9062-9076.
- 251.Shaw Stewart PD, Kolek SA, Briggs RA, Chayen NE, and Baldock PF (2011). Random microseeding: a theoretical and practical exploration of seed stability and seeding techniques for successful protein crystallization. *Cryst. Growth Des.*, 11(8), 3432-3441.
- 252.Shen H, Stoute J, and Liu KF (2020). Structural and catalytic roles of the human 18S rRNA methyltransferases DIMT1 in ribosome assembly and translation. *J. Biol. Chem.*, 295(34), 12058-12070.
- 253.Shoji T, Takaya A, Sato Y, Kimura S., Suzuki T, and Yamamoto T (2015). RlmCD-mediated U747 methylation promotes efficient G748 methylation by methyltransferase RlmA^{II} in 23S rRNA in *Streptococcus pneumoniae*; interplay between two rRNA methylations responsible for telithromycin susceptibility. *Nucleic Acids Res.*, 43(18), 8964-8972.
- 254.Sievers F and Higgins DG (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.*, 1079, 105-116.
- 255.Siibak T and Remme J. (2010). Subribosomal particle analysis reveals the stages of bacterial ribosome assembly at which rRNA nucleotides are modified. *RNA*, 16, 2023–2032.
- 256.Singh MK and Manoj N (2017). Structural role of a conserved active site cis proline in the *Thermotoga maritima* acetyl esterase from the carbohydrate esterase family 7. *Struct. Funct. Bioinf.*, 85(4), 694-708.
- 257.Skinner R, Cundliffe E, and Schmidt FJ (1983). Site of action of a ribosomal RNA methylase responsible for resistance to erythromycin and other antibiotics. *J. Biol. Chem.*, 258(20), 12702-12706.
- 258.Sleiman S and Dragon F (2019). Recent advances on the structure and function of RNA acetyltransferase Kre33/NAT10. *Cells*, 8(9), 1035.
- 259.Sologub M, Litonin D, Anikin M, Mustaev A, and Temiakov D (2009). TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell*, 139, 934-944.

260. Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G, and Frank J (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit–subunit interactions. *Cell*, 107, 373–386.
261. Spåhr H, Habermann B, Gustafsson CM, Larsson NG, and Hallberg BM (2012). Structure of the human MTERF4–NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. *Proc. Natl. Acad. Sci.*, 109(38), 15253–15258.
262. Staub E, Fiziev P, Rosenthal A, and Hinzmann B (2004). Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays.*, 26(5), 567–581.
263. Steitz TA (2008). A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.*, 9(3), 242–53.
264. Stephan NC, Ries AB, Boehringer D, and Ban N (2021). Structural basis of successive adenosine modifications by the conserved ribosomal methyltransferase KsgA. *Nucleic Acids Res.*, 49(11), 6389–6398.
265. Strassler SE, Bowles IE, Dey D, Jackman JE, and Conn GL (2022). Tied up in knots: Untangling substrate recognition by the SPOUT methyltransferases. *J. Biol. Chem.*, 298(10).
266. Sumathi K, Ananthalakshmi P, Roshan MM, and Sekar K (2006). 3dSS: 3D structural superposition. *Nucleic Acids Res.*, 34(2), W128–W132.
267. Sun W, Xu X, Pavlova M, Edwards AM, Joachimiak A, Savchenko A, and Christendat D (2005). The crystal structure of a novel SAM-dependent methyltransferase PH1915 from *Pyrococcus horikoshii*. *Protein Sci.*, 14(12), 3121–3128.
268. Sunita S, Purta E, Durawa M, Tkaczuk KL, Swaathi J, Bujnicki JM, and Sivaraman J (2007). Functional specialization of domains tandemly duplicated within 16S rRNA methyltransferase RsmC. *Nucleic Acids Res.*, 35(13), 4264–4274.
269. Sunita S, Tkaczuk KL, Purta E, Kasprzak JM, Douthwaite S, Bujnicki JM, and Sivaraman J (2008). Crystal structure of the Escherichia coli 23S rRNA: m⁵C methyltransferase RlmI (YccW) reveals evolutionary links between RNA modification enzymes. *J. Mol. Biol.*, 383(3), 652–666.

270. Sweeney BA, Hoksza D, Nawrocki EP, Ribas CE, Madeira F, Cannone JJ, Gutell R, Maddala A, Meade CD, Williams LD, and Petrov AS (2021). R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nature Commun.*, 12(1), 3494.
271. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, and Jensen LJ (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, 47(D1), D607-D613.
272. Taboada B, Estrada K, Ciria R, and Merino E (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinform.*, 34(23), 4118-4120.
273. Tan J, Jakob U, and Bardwell JC (2002). Overexpression of two different GTPases rescues a null mutation in a heat-induced rRNA methyltransferase. *J. Bacteriol.*, 184(10), 2692-2698.
274. Taylor AB, Meyer B, Leal BZ, Kötter P, Schirf V, Demeler B, Hart PJ, Entian KD, and Wöhnert J (2008). The crystal structure of Nep1 reveals an extended SPOUT-class methyltransferase fold and a pre-organized SAM-binding site. *Nucleic Acids Res.*, 36(5), 1542-1554.
275. Teller DC (1973). Characterization of proteins by sedimentation equilibrium in the analytical ultracentrifuge. *Methods Enzymol.*, 27, 346-441.
276. The UniProt Consortium (2023). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, 51(D1), D523-D531.
277. Thomas SR, Keller CA, Szyk A, Cannon JR, and LaRonde-LeBlanc NA (2011). Structural insight into the functional mechanism of Nep1/Emg1 N1-specific pseudouridine methyltransferase in ribosome biogenesis. *Nucleic Acids Res.*, 39(6), 2445-2457.
278. Tirumalai MR, Rivas M, Tran Q, and Fox GE (2021). The peptidyl transferase center: a window to the past. *Microbiol. Mol. Biol.*, 85(4), e00104-00121.
279. Tkaczuk KL, Dunin-Horkawicz S, Purta E, and Bujnicki JM (2007). Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC bioinformatics*, 8(1), 1-31.

280. Toh SM and Mankin AS (2008). An indigenous posttranscriptional modification in the ribosomal peptidyl transferase center confers resistance to an array of protein synthesis inhibitors. *J. Mol. Biol.*, 380(4), 593-597.
281. Tollervey D, Lehtonen H, Jansen R, Kern H, and Hurt EC (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell*, 72, 443–457.
282. Trixl L and Lusser A (2019). The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdiscip. Rev. RNA*, 10(1), e1510.
283. Truitt ML, and Ruggero D (2016). New frontiers in translational control of the cancer genome. *Nat. Rev. Cancer*, 16(5), 288-304.
284. Tu C, Tropea JE, Austin BP, Court DL, Waugh DS, and Ji X (2009). Structural basis for binding of RNA and cofactor by a KsgA methyltransferase. *Structure*, 17(3), 374-385.
285. Urbonavičius J, Auxilien S, Walbott H, Trachana K, Golinelli-Pimpaneau B, Brochier-Armanet C, and Grosjean H (2008). Acquisition of a bacterial RumA-type tRNA (uracil-54, C5) -methyltransferase by Archaea through an ancient horizontal gene transfer. *Mol. Microbiol.*, 67(2), 323-335.
286. Urbonavičius J, Skouloubris S, Myllykallio H, and Grosjean H (2005). Identification of a novel gene encoding a flavin-dependent tRNA: m⁵U methyltransferase in bacteria—evolutionary implications. *Nucleic Acids Res.*, 33(13), 3955-3964.
287. Vagin AA (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2184-2195.
288. Van Buul CP and Van Knippenberg PH (1985). Nucleotide sequence of the ksgA gene of *Escherichia coli*: comparison of methyltransferases effecting dimethylation of adenosine in ribosomal RNA. *Gene*, 38(1-3), 65-72.
289. Van Buul CP, Visser W, and Van Knippenberg PH (1984). Increased translational fidelity caused by the antibiotic kasugamycin and ribosomal ambiguity in mutants harbouring the ksgA gene. *FEBS Lett.*, 177(1), 119-124.
290. Venema J and Tollervey D (1999). Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.*, 33(1), 261-311.

291. Wachino JI, Shibayama K, Kurokawa H, Kimura K, Yamane K, Suzuki S, Shibata N, Ike Y, and Arakawa Y (2007). Novel plasmid-mediated 16S rRNA m1A1408 methyltransferase, NpmA, found in a clinically isolated *Escherichia coli* strain resistant to structurally diverse aminoglycosides. *Antimicrob. Agents Chemother.*, 51(12), 4401-4409.
292. Walbott H, Husson C, Auxilien S, and Golinelli-Pimpaneau B (2007). Cysteine of sequence motif VI is essential for nucleophilic catalysis by yeast tRNA m⁵C methyltransferase. *RNA*, 13(7), 967-973.
293. Walbott H, Leulliot N, Grosjean H, and Golinelli-Pimpaneau B (2008). The crystal structure of *Pyrococcus abyssi* tRNA (uracil-54, C5)-methyltransferase provides insights into its tRNA specificity. *Nucleic Acids Res.*, 36(15), 4929-4940.
294. Weisblum B (1995). Erythromycin resistance by ribosome modification. *Antimicrob. Agents Chemother.*, 39(3), 577-585.
295. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, and Jain S (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.*, 27(1), 293-315.
296. Wilson DN, Schlutzen F, Harms JM, Yoshida T, Ohkubo T, Albrecht R, Buerger J, Kobayashi Y, and Fucini P (2005). X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit. *EMBO J.*, 24(2), 251-260.
297. Wimberly BT, Brodersen DE, Clemons Jr WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, and Ramakrishnan V (2000). Structure of the 30S ribosomal subunit. *Nature*, 407(6802), 327-339.
298. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, and McNicholas SJ (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, 67(4), 235-242.
299. Woodbury RL, Hardy SJ, and Randall LL (2002). Complex behavior in solution of homodimeric SecA. *Protein Sci.*, 11(4), 875-882.
300. Woodson SA (2008). RNA folding and ribosome assembly. *Curr Opin Chem Biol.*, 12(6), 667-673.

301. Wurm JP, Meyer B, Bahr U, Held M, Frolow O, Kötter P, Engels JW, Heckel A, Karas M, Entian KD, and Wöhnert J (2010). The ribosome assembly factor Nep1 responsible for Bowen–Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Res.*, 38(7), 2387-2398.
302. Yakhnin H, Yakhnin AV, Mouery BL, Mandell ZF, Karbasiafshar C, Kashlev M, and Babitzke P (2019). NusG-dependent RNA polymerase pausing and tylosin-dependent ribosome stalling are required for tylosin resistance by inducing 23S rRNA methylation in *Bacillus subtilis*. *MBio.*, 10(6), 10-128.
303. Yan Y, Tao H, He J, and Huang SY (2020). The HDock server for integrated protein–protein docking. *Nature Protoc.*, 15(5), 1829-1852.
304. Yang J, Yan R, Roy A, Xu D, Poisson J, and Zhang Y (2007). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12(1), 7-8.
305. Ye K (2007). H/ACA guide RNAs, proteins and complexes. *Curr. Opin. Struct. Biol.*, 17, 287–292.
306. Yu L, Petros AM, Schnuchel A, Zhong P, Severin JM, Walter K, Holzman TF, and Fesik SW (1997). Solution structure of an rRNA methyltransferase (ErmAM) that confers macrolide-lincosamide-streptogramin antibiotic resistance. *Nature Struct. Biol.*, 4(6), 483-489.
307. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, and Noller HF (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Sci.*, 292(5518), 883-896.
308. Yutin N, Puigbò P, Koonin EV, and Wolf YI (2012). Phylogenomics of prokaryotic ribosomal proteins. *PloS one*, 7(5), e36972.
309. Zheng YJ and Bruice TC (1997). A theoretical examination of the factors controlling the catalytic efficiency of a transmethylation enzyme: catechol O-methyltransferase. *J. Am. Chem. Soc.*, 119(35), 8137-8145.
310. Zorbas C, Nicolas E, Wacheul L, Huvelle E, Heurgué-Hamard V, and Lafontaine DL (2015). The human 18S rRNA base methyltransferases DIMT1L and WBSCR22-TRMT112 but not rRNA modification are required for ribosome biogenesis. *Mol. Biol. Cell.*, 26(11), 2080-2095.
311. Zuker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13), 3406-3415.



List of Publications, Conferences, and Workshops

PUBLICATIONS FROM THESIS

1. **Saha, S.** and Kanaujia, S.P. Unveiling the structural and mechanistic insights of the ribosome assembly factor Nep1: a N1-specific pseudouridine methyltransferase (*Under preparation*).
2. **Saha, S.** and Kanaujia, S.P. Decoding substrate selectivity of an archaeal RlmCD-like methyltransferase through its salient traits. ***Biochemistry***, Accepted.
3. **Saha, S.** and Kanaujia, S.P. (2024). Structural and functional characterization of archaeal DIMT1 unveils distinct protein dynamics essential for efficient catalysis. ***Structure***, Accepted.
4. **Saha, S.** Mandal, S.K. and Kanaujia, S.P. (2024). Distinct characteristics of putative archaeal 5-methylcytosine RNA methyltransferases unveil their substrate specificities and evolutionary ancestries. ***J. Biomol. Struct. Dyn.***, 1-18.

LIST OF CONFERENCES AND WORKSHOPS

1. **Sayan Saha** and Shankar Prasad Kanaujia. Structural and evolutionary insights into an archaeal rRNA methyltransferase involved in the modification of 5-methyluridine. Research & Industrial Conclave – Integration'2024. Aug 09-11, 2024, IIT Guwahati, Assam, India. [Oral Presentation].
2. **Sayan Saha** and Shankar Prasad Kanaujia. Unveiling the protein dynamics of archaeal DIMT1 rRNA methyltransferase for efficient catalysis. National Seminar on Crystallography (NSC-50). November 22-24, 2023, CSIR-IMTECH, Chandigarh, India. [Oral Presentation].
3. **Sayan Saha** and Shankar Prasad Kanaujia. Unveiling the substrate specificities and evolutionary ancestries of putative archaeal 5-methylcytosine RNA methyltransferases. Research & Industrial Conclave – Integration'2023. May 14-16, 2023, IIT Guwahati, Assam, India. [Best Poster Award].
4. **Sayan Saha** and Shankar Prasad Kanaujia. Unveiling the substrate specificities and evolutionary ancestries of putative archaeal 5-methylcytosine RNA

- methyltransferases. National Seminar on Crystallography (NSC-49). November 28-30, 2022, University of Jammu, Jammu, India. [Oral Presentation].
5. **Sayan Saha** and Shankar Prasad Kanaujia. *In silico* studies enlighten the salient features of putative archaeal 5-methylcytosine methyltransferases. National Seminar on Crystallography (NSC-48). November 25-27, 2021, IIT Roorkee, Uttarakhand, India. [Poster Presentation].
 6. **Sayan Saha** and Shankar Prasad Kanaujia. *In silico* characterization of putative 5-methylcytosine methyltransferases from *Pyrococcus horikoshii*. PDB 50th Anniversary Symposium in Asia. November 24, 2021, PDBj and Institute for Protein Research, Osaka University, Japan. [Poster Presentation].
 7. **Sayan Saha** and Shankar Prasad Kanaujia. Characterization of putative archaeal 5-methylcytosine methyltransferases: An *in silico* approach. Bringing Molecular Structure to Life: 50 Years of the PDB. October 20-22, 2021, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [Poster Presentation].
 8. **Sayan Saha** and Shankar Prasad Kanaujia. Associazione Italiana Cristallografia School 2021 (AICS2021). August 31 – September 4, 2021. [Workshop].



