Molecular determinants of stability and nucleobase specificity in flavin biosynthesis enzymes

A thesis

submitted in partial fulfilment of the requirements

of the degree of

Doctor of Philosophy

By

Yashwant Kumar

20133270

IISER PUNE

Indian Institute of Science Education and Research, Pune – 411008

2021

Dedicated to..

Maa

(Lalita Kumari)

CERTIFICATE

Certified that, the work incorporated in the thesis entitled, "**Molecular determinants of stability and nucleobase specificity in flavin biosynthesis enzymes**" submitted by Yashwant Kumar was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Date: 31st May 2021

Dr. Amrita B. Hazra

Pune (MH), India.

(Supervisor)

DECLARATION

        I declare that this written submission represents my ideas in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Yashwant Kumar

Yashwant Kumar

31st May 2021

20133270

Acknowledgement

As I sit to pen down my thesis, a surge of memories and an overwhelming feeling of gratitude surround me. The past few years at IISER Pune have been an indescribable time of my life where I got to grow as a scholar and a person.

First and foremost, I express my sincere gratitude to my thesis supervisor Dr. Amrita B Hazra. She accepted me as a part of the Whytamin lab and introduced me to the world of biology and mechanistic enzymology. I admire her patience while teaching her students a spectrum of hands-on skills and concepts- from handling bacteria, freezing proteins, fixing an HPLC, to deducing molecular mechanisms. I deeply value our one-on-one interactions and discussions about science, philosophy, and Indian culture. I am also thankful to her for facilitating opportunities to present my work at various national and international conferences and meetings.

The periodic meetings with my research advisory committee members, Prof. Harinath Chakrapani and Dr. Santosh Kumar Jha (CSIR-NCL), have been integral to my research progress. Their insights have helped us steer the project to the destination it is at today. I will fondly remember Hari's questions from the basics of chemistry that have encouraged me to keep in touch with the foundations of everything I learn.

I am grateful to Dr. Gayathri Pananghat, Dr. Arnab Mukherjee, and Dr. Nishad Matange for their constant support and constructive inputs throughout the journey of my Ph.D.

I am truly thankful to Dr. Mrinalini Puranik, under whose guidance my Ph.D. journey began, and she has continued to be a source of inspiration and positive affirmations through the years. I am eternally grateful to my mentors in my first lab, Vishakha Karnawat, Sayan Mondal, and Sudeb Ghosh, for teaching me new scientific concepts and fundamentals of research ethics. I have precious memories with my previous lab members Shahila, Anil, Arya, Siddhartha, and Jyoti.

I am grateful for the exemplary work environment and family-like group dynamics in our lab. The past and present members of the Whytamin lab have stood together in science, health and sickness, madness and crisis. I will genuinely miss my time spent with each one of them.

I am fortunate to have lab mates like Yamini, Rupali, and Ateek. I can't forget the continuous and non-stop cooperation from Yamini and Rupali. It would be an audacious claim, but all the research potential I have today is because of my lab mates and Amrita. Rupali, is an excellent microbiologist, a teacher, and a hygiene enthusiast in our lab. She guided me in cloning and bacterial culture experiments. I am thankful to Rupali for her guidance in molecular biology experiments. I would also like to mention my gratitude to Yamini. I call her Wikipedia. I enjoy scientific discussions with her, from bacterial cells to anything like rocket science. She proved a true torchbearer who helped me get perfection in molecular biology experiments, concept, design, and scientific writing. I would also appreciate Ateek, who joined this nucleobase specificity project and done a lot of troubleshooting. It gave a boost to our project. Saswata helped me in the starting of my project as an undergraduate student. A very polite person and I never saw him angry. I wish that I have such quality. Mrutyunjay and I share many memories in IISER Pune, whether it is a chai party, late-night experiment schedule, trekking, or intellectual discussions. I found Sheryl very humble and a good learner. We shared some mixed English- Hindi mixed chat and conversations. Prathamesh was the first undergraduate in our lab. For one entire year, we were the night owls of our lab. The time spent with him is unforgettable, whether for lab experiments or the late-night bike ride. I also had an excellent opportunity to work with exceptional undergraduate students- Lavisha, Akshay, Vruta, Janhavi, and Anisha, who contributed to the project, and several visiting students in our lab.

Several experiments were possible because of the instruments and reagents shared by the labs in the biology and chemistry departments. I am incredibly thankful to Dr. Shabana Khan, Prof. Anjan Banerjee, Prof. Thomas Pucadyil, Dr. Britto Sandanaraj, Prof. Harinath Chakrapani, and Prof H.N. Gopi, and their lab members for the training and cooperation.

While my Ph.D. journey was not easy, I count myself blessed to have received unwavering support from faculties and friends. I will never forget encouraging words and lively interactions with Dr. Jeetendra Chugh, Dr. Sudha Rajamani, Dr. Anirban Hazra, Dr. Jeet Kalia, and Dr. John Matthew.

Over the years, I have cultivated friendships that I will cherish for life. I am happy to have shared some good memories with my dear friends Neha Kathewad, Shivpal, and my batchmates Rakesh Pant, Manoharan, Himani Rawat. I have also had the opportunity to

are unforgettable. I am thankful to Manisha Ojha for her support and suggestions during the years of my Ph.D. My childhood friend Sandeep has been one of the first science enthusiasts in my life. Many Special shoutouts to Akshi, Rashmi, Radha, and Sanjeev Rai who indeed helped me crack the NET exam.

Finally, I thank my parents for being pillars of my strength and fuelling my dreams all throughout. The acknowledgments will be incomplete without expressing my love and gratitude for my siblings- Hemlata, Yashodhara, Swarnlata, and Balwant.

Yashwant Kumar

31th May 2021

Table of contents

List of Figures

List of Tables

Molecular determinants of stability and nucleobase specificity in flavin biosynthesis enzymes

Yashwant Kumar, Registration no. 20133270

Submission date: 31st May 2021

Abstract

Nucleoside triphosphates (NTP) are the building blocks of RNA, and also play a central role in various metabolic pathways. Even though the NTPs - adenosine triphosphate (ATP), guanosine triphosphate (GTP), cytidine triphosphate (CTP) and uridine triphosphate (UTP) - are largely similar in structure, the nucleobase moiety differs among them and appears to be the determining factor for their roles in cellular processes. For instance, ATP is used as the energy currency in cellular metabolism, GTP is typically utilized in signal transduction, and CTP and UTP are involved in phospholipid and glycogen biosynthesis, respectively. However, the molecular mechanisms by which enzymes distinguish between the various NTPs is not well explored in the literature. Even though the binding motifs as well as mode of binding of NTPs in some classes of enzymes has been established, these rules do not apply across the vast variety of NTP-utilizing enzymes.

In this thesis, we undertake the exploration of the molecular determinants of stability and nucleobase specificity of the NTP-utilizing enzymes involved in flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) biosynthesis, with a goal of expanding the range of NTPs that these enzymes use. Our studies done with homologs of riboflavin kinase, FMN adenylyltransferase and FAD synthetase from various organisms have culminated in understanding some aspects of the NTP choice of these enzymes which has allowed us to produce a series of enzymes that show diverse nucleotide preferences.

Investigations conducted on the CTP-dependent riboflavin kinase from the thermophilic archaea *Methanocaldococcus jannaschii* (*Mj*RibK) suggest that a loop containing a small one-turn helix in the middle section of the enzyme is important for the recognition of the cytosine nucleobase. The perturbation of this one-turn helix via mutations has led to altered NTP specificity. Further, comparative sequence and structure analysis with a mesophilic homologue of *Mj*RibK led us to identify the molecular interactions responsible for its thermostability. Next, we analyzed homologs of FMN adenylyltransferases and FAD synthetases for their promiscuity in NTP utilization, which helped us establish a robust enzymatic route for the synthesis of FAD cofactor analogues such as FGD and dFAD.

Our studies lay the basis for understanding nucleotide specificity in flavin biosynthesis enzymes. Furthermore, our studies on engineering these enzymes to exhibit altered nucleotide specificity lays the foundation for the enzymatic synthesis of functional FAD analogues. The molecular insights that we derive from our studies informs the broader question of how a specific nucleobase is selected by nucleotide-utilizing enzymes for their cellular function. This exploration can be extended to the fields of synthetic biology, metabolic circuits, bioremediation, and the development of potent inhibitors for flavin biosynthesis enzymes and flavoenzymes.

Understanding nucleotide specificity of the enzymes through Nucleobase recognition: Review of literature

1.1 Introduction:

Nucleotide triphosphates (NTP) are the building blocks of DNA and RNA. It is composed of a phosphate moiety, a ribose/deoxyribose ring, and a nitrogenous purine or pyrimidine base called nucleobase (Figure 1.1 A). DNA contains four types of nucleobases adenine, guanine, cytosine, and thymine. RNA also has four nucleobases, but it has uracil instead of thymine (Figure 1.1 B).
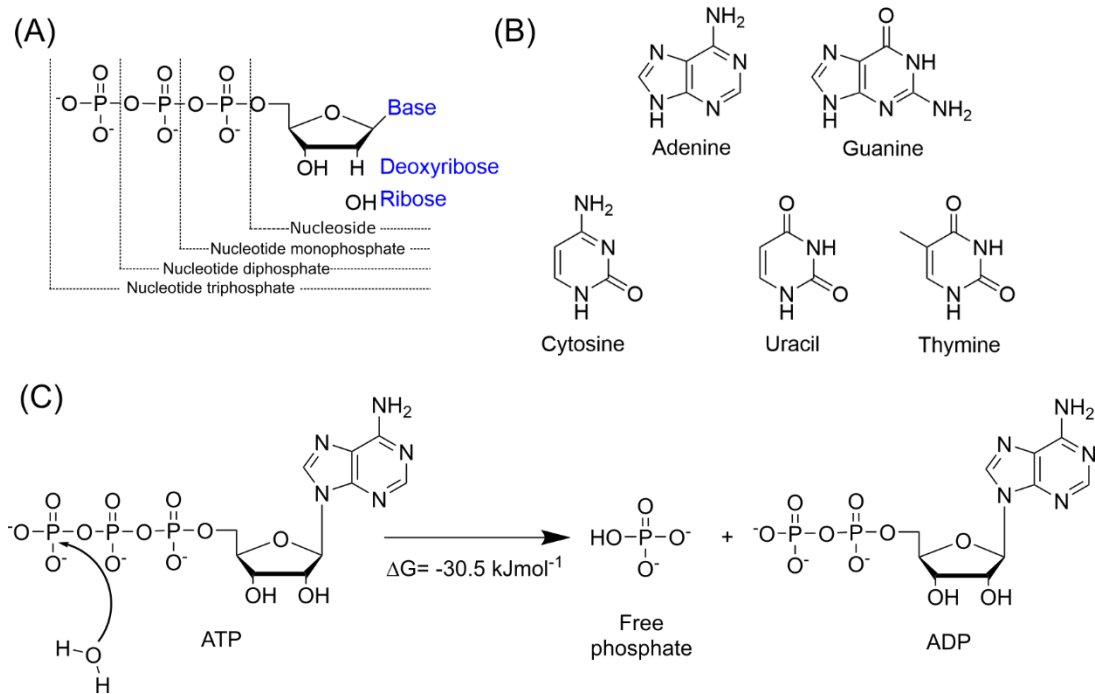
Figure 1.1: Nucleotide, nucleobases and hydrolysis reaction (A) Nucleotide triphosphate is composed of phosphate moiety, ribose sugar, and nucleobase unit. Monophosphate, diphosphate, and triphosphate have one, two, and three phosphate units, respectively. (B) structures of nucleobases (C) ATP hydrolysis reactions

NTPs have a diverse role in metabolic pathways. NTPs can be involved in enzymatic reactions as well as regulation of transcription, translation, and inter- or intracellular signaling. For example, ATP is used as energy currency in cellular processes, GTP is involved in the signal transduction pathway, CTP is being used in phospholipid biosynthesis, and UTP finds its role in glycogen formation.[1–4] NTPs carry out several phosphorylation and hydrolysis reactions. Kinases are the class of transferase enzymes that conduct substrate phosphorylation using ATP or occasionally other NTPs as phosphate donors.[5–7] Apart from the roles mentioned above, NTPs are also precursors of cofactors like NAD, SAM, FAD, etc.

NTPs are central in cellular metabolism. However, many of the reactions utilizing NTPs involve triphosphate moiety. For example, ATP-hydrolysis involves dissociating the triphosphate unit to form ADP and phosphate ion to generate energy for cellular processes (Figure 1.1 C). The kinase reactions utilize NTPs as a source of phosphate for the phosphorylation reaction. If we look at the structures of NTPs, the only perceptible difference is the nucleobase unit, which is on the other end from the catalytically important phosphate unit. In other words, if an enzyme has to distinguish between ATP and GTP, it should differentiate on the basis of adenine and guanine nucleobases, respectively. Most of the studies done in the past to decipher NTP-specificity revolves around the binding and stabilization of phosphate unit.[8–12] The trend for the recognition and hence specificity of nucleobase differ among enzyme families. Thus arise the questions: 1. How do enzymes recognize a specific nucleotide within the mix of NTPs present in the cytosolic pool? and 2. How does a difference only in the nucleobase unit of the individual NTPs make them specific for distinct metabolic pathways?

The role of nucleobases as a recognition handle is not only limited to the NTP reactions. If we look at the structures of most commonly used cofactors such as – FAD, NAD/NADP, coenzyme A, and SAM, we find re-occurrence of adenine nucleobase in each of them (Figure 1.2). Interestingly, none of the other nucleobases have been observed as a structural part of naturally occurring cofactors. The role of the adenine nucleobase and what specific role it plays in the cofactors are yet unclear.
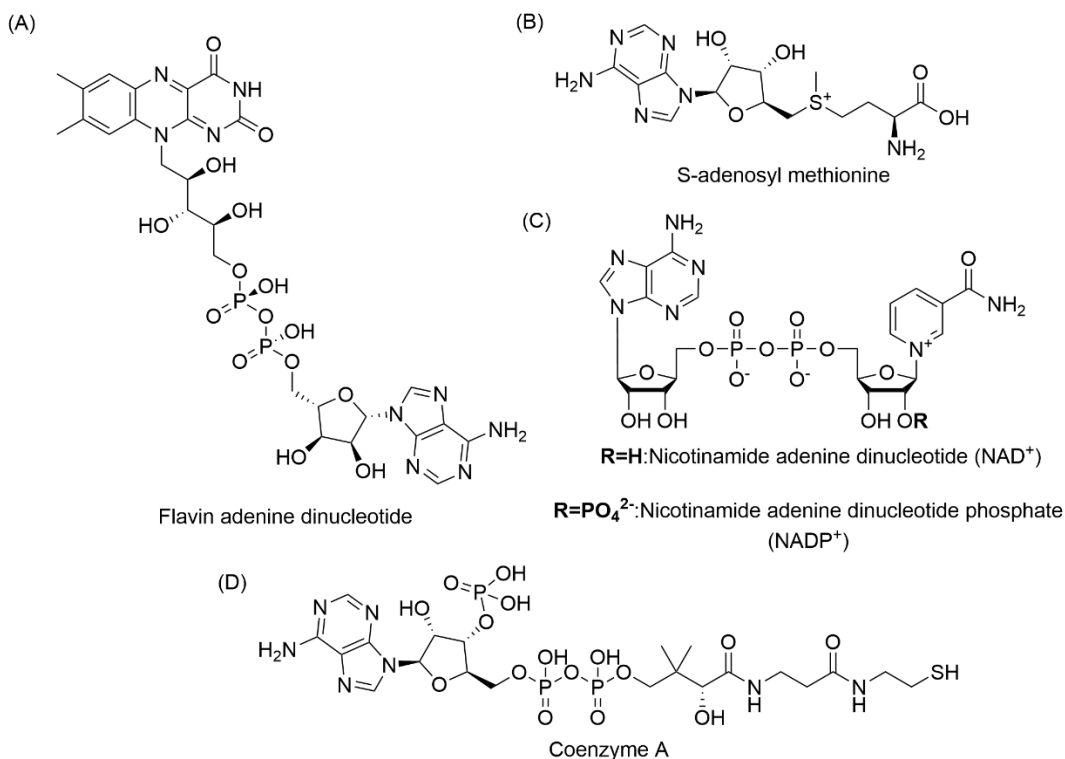
Figure 1.2: Some of the cofactors contain adenine nucleobases. **(A)** Flavin adenine dinucleotide (FAD) has an isoalloxazine ring involved in redox catalysis. **(B)** S-adenosylmethionine is commonly involved in methylation reactions, using the methyl group centered on S-atom. **(C)** The nicotinamide ring of NAD$^+$ and NADP$^+$ are involved in the catalytic activity. **(D)** The terminal -SH group is the catalytically active part of coenzyme A.

In brief, the molecular factors governing nucleobase specificity are not explored well. Most of the studies done on NTP specificity details the recognition of phosphate moiety. Several critical questions for the recognition of nucleotide bases are as follows

(a) What is the molecular basis of nucleobase binding in NTP utilizing enzymes?
(b) What is the role of adenine in the cofactors such as FAD, NAD$^+$, SAM, and coenzyme A?
(c) Can we alter the nucleobase specificity of NTP utilizing enzymes?
(d) Can we modulate the specificity of cofactors synthesizing enzymes to make unnatural cofactors analogs like FN*D (nucleobase analog of FAD), SN*M (nucleobase analog of SAM), and NN*D (nucleobase analog of NAD)?

(e) Do these bioorthogonal cofactors analogs and the respective engineered enzymes can be incorporated into metabolic pathways?

Understanding nucleobase selectivity can have a wide range of potential applications like *in vivo* labeling studies, synthetic biology, and drug designing. Bioorthogonal cofactors have the potential to be powerful tools in manipulating the metabolism and physiology of an organism. A mechanistic understanding of the biosynthesis of these analogs will open up new avenues in antimicrobial drug development, cell imaging, bioremediation, and industrial biotechnology applications.

1.2 The nucleotide preferences of some homologous enzymes

Although several enzymes exhibit high substrate specificity, there are examples of homologous enzymes that differ in nucleotide utilization. Even some NTP utilizing enzymes are promiscuous. These enzymes have potential molecular level information hidden in their sequence and hence, the structure. A comparative analysis of these homologous enzymes can reveal the determinants of nucleobase specificity in them.

For example, phosphopantothenoylcysteine synthetase (PPCS) from *Escherichia coli* is CTP specific, while the same enzyme from *Homo sapiens* is ATP specific.[13,14] In another instance, Succinyl CoA synthetase (SCS) from *Sus scrofa* is GTP specific, while that form *Blastocystis sp.* is ATP specific.[15,16] This enzyme from *E. coli* is nonspecific and can use both ATP and GTP for the reaction.[17,18]

Kinases are a family or subclass of enzymes that depend on NTPs as phosphate donors and while a majority of the enzymes are known to be ATP specific. However, there is evidence for CTP and GTP-specific kinases. The riboflavin kinase from *Methanocaldococcus jannaschii* is CTP specific while that from *Rhizopus oryzae* is GTP-specific.[19–21] Phosphoglycerate kinase from *Entamoeba histolytica* and polyribonucleotide-5'-hydroxyl-kinase from *S. cerevisiae* are also GTP-dependent kinases.[22,23] Diacylglycerol kinases from *E. coli* and *S. cerevisiae* have ATP and CTP specificities, respectively.[24,25]

1.3 Structural basis of the binding and specificity of nucleotide triphosphates

Based on bioinformatics and X-ray crystal structures of many ATP and GTP binding enzymes have been shown to contain a glycine-rich sequence motif GxxxxGK[ST]. This

sequence motif is known as P-loop or Walker A motif and is responsible for binding the tri-phosphate group of ATP or GTP.[12]

Apart from P-loop, there are some other motifs in many GTPase. These motifs are a distal DxxG motif and another distal [NT]KxD motif.[9] The DxxG motif binds the γ-phosphate of nucleotide and relevant for Mg^{2+} binding. There is a conserved [NT]KxD motif in the GTPase superfamily that imparts specificity for GTP and is known for guanine recognition. ATPase does not have any conserved sequence motif like [NT]KxD in GTPase for adenine binding. P-loop is present in many ATPases, but some ATP binding enzymes do not contain P-loop.[26–28] Moreover, π-π stacking interaction between adenine ring and an aromatic amino acid residue (tyrosine, tryptophan and phenylalanine) is observed in many P-loop ATPase (Figure 1.3 A).
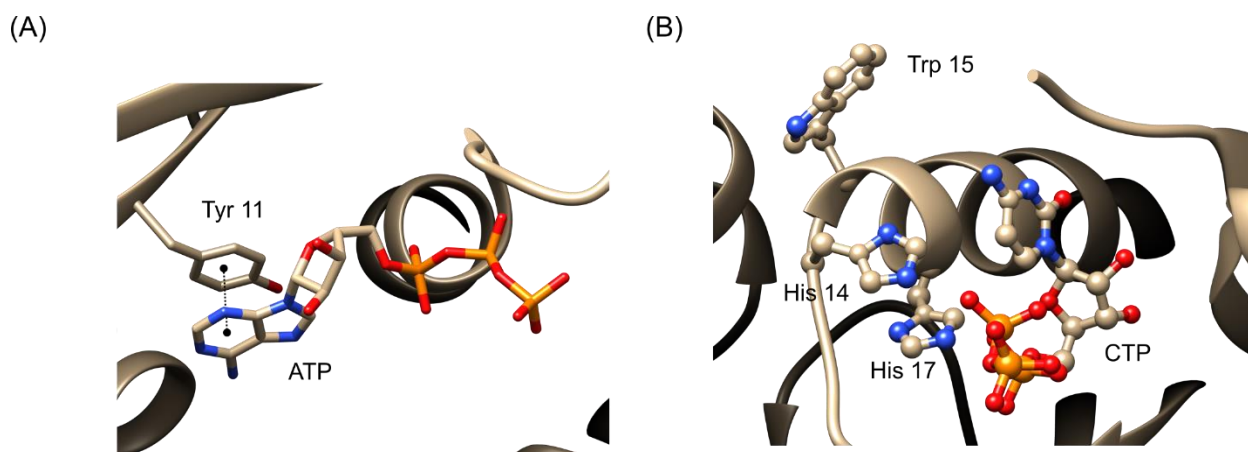
Figure 1.3: Adenine and cytosine binding motifs. **(A)** π-π Stacking interaction (black dotted line) between adenine nucleobase and tyrosine residue in ABC transporter protein (PDB:6M96) **(B)** The HxGH motif around phosphate moiety of CTP in Glycerol 3-phosphate cytidylyltransferase (PDB:1COZ)

Sequence analysis of some CTP binding proteins has shown the presence of HxGH motif in their structures (Figure 1.3 B).[29,30] The first histidine unit binds the phosphate oxygen of CTP. Mutation of this histidine results in a three-fold decrease in the activity.

The schematic of the ATP, GTP and CTP binding motifs are depicted in figure 1.4.
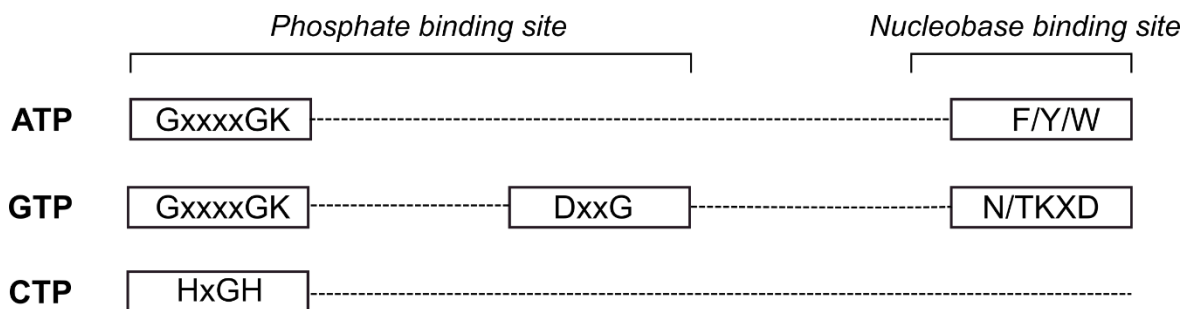
Figure 1.4: Most of the ATPase and GTPase have conserved glycine-rich loop or P-loop for the phosphate binding. Some of the ATPase has stacking interaction between adenine base and one of the aromatic amino acid residues of the protein. CTP has the HxGH motif for the stabilization of phosphate moiety.

1.4 Altering the specificity of ATPase and GTPase based on sequence, structure and other molecular factors:

Before the advent of highly efficient molecular modeling techniques and the lack of high-resolution crystal structures, the strategy of altering the specificity of an enzyme was solely dependent on the sequence comparison and structural analysis. Changing substrate specificity involves altering the steric and/or electrostatic component of non-covalent interactions. The concept of achieving altered substrate specificity on the basis of shape-complementarity is known as bump-and-hole approch.[31]

Several investigations have been done to change the nucleotide specificity of some NTP-utilizing enzymes. The specificity of a prokaryotic elongation factor, EF-Tu GTPase was switched from GTP/GDP to xanthine (tri/di)phosphate by D138N mutation.[32] A334T/S mutation in mammalian cAMP-dependent protein kinase had shown a high affinity for cGMP.[33]A bacterial cytoskeletal protein, FtsZ84 (G105S), is a defected allele found in *Escherichia coli*, and it has shown activity with ATP contrary to the wild-type FtsZ which is GTP specific.[34]

1.5 Altering the specificity of kinases based on substrate complementarity and electrostatic potential of gatekeeper residues

1.5.1 Nucleotide specificity in Src family tyrosine kinases

Src family kinases are the protein kinases that phosphorylate tyrosine residues of the proteins using ATP.[35] Engineered *Rous sarcoma* virus tyrosine kinase displayed catalytic efficiency with an ATP analog, *N*6-(cyclopentyl) ATP, similar to the catalytic efficiency with ATP itself. The residues V323 and I338 control the specificity of N6-substituted ATP analogs. V323A and I338A mutations were able to utilize N6-cyclopentyl ATP.[36] The sidechain of a single residue (I338) controls specificity for nucleobase-substituted ATP analogs in the binding pocket of v-Src. Elimination of this sidechain by mutation to glycine (I338G) produces a v-Src kinase which preferentially utilizes N6-(benzyl) ATP as a phospho-donor.[35] Further investigation done by Shokat and coworkers led to the conclusion that I338 residue acts as a gatekeeper residue and controls substrate access into the active site.[37] Experiments were conducted to test the ability of the I338 mutant to utilize the substrate N4 (benzyl) ribavirin triphosphate. However, N4 (benzyl) ribavirin triphosphate (RTP) failed to act as a phospho-donor. Molecular modeling and docking studies suggested that tighter binding of hydrophobic N4 (benzyl) RTP to the active site or improper orientation or missing of catalytically important H-bond might be the reason behind this.[38]

1.5.2 Altering the specificity of GTP binding proteins

As stated above the specificity of a prokaryotic elongation factor, EF-Tu GTPase was switched from GTP/GDP to xanthine (tri/di)phosphate by D138N mutation, on the basis of propinquity of catalytic residue with guanine nucleobase.[32] Shah and coworkers also demonstrated highly specific inhibition or activation of engineered H-Ras GTPase with unnatural GDP and GTP analogs to probe G protein signaling.[39]

1.5.3 Advent of gatekeeper residues and substrate specificity

The key behind determining the substrate specificity by gatekeeper residues is the stereochemical orientation of active site residues and the electrostatic potential surfaces. By careful examination of active-site architecture and the electrostatic interaction of substrate,

specificity can be modulated. Aminoglycoside 2"-phosphotransferases (APH (2")) show variable NTP specificity.[40] APH (2")- IIa and APH (2") -IVa can use both ATP and GTP, while APH (2") - IIIa use GTP exclusively. Gatekeeper mutation M85Y in APH (2") - IIa resulted in ~10 fold decrease in the K_m of GTP and ~320 fold increase in the K_m of ATP.[40] Similarly, by comparing charged gatekeeper residues of the active site of succinyl CoA synthetases (SCS) from *Blastocystis sp.* (ATP specific), *E. coli* (ATP, GTP, nonspecific), and *Sus scrofa* (GTP specific), Vashisht et al. have altered the nucleotide specificity of *Blastocystis sp.* SCS from ATP to GTP.[41] However, the concept of gatekeeper residues is not the rule of thumb for alteration of nucleotide specificity. The study done on SCS from the malarial parasite *Plasmodium falciparum* suggests that structural scrutiny of binding site architecture is also critical along with the electrostatic interaction of gatekeeper residues in modeling the specificity.[42]

1.6 Chimeric design and loop remodeling to confer orthogonal nucleotide specificity

Chimeric proteins can also be created to alter the nucleotide specificity. The purine salvage pathway enzyme Hypoxanthine-guanine phosphoribosyltransferase (*Hs*HGPRT) from *Homo sapiens* catalyze the conversion of guanine and hypoxanthine nucleobases to respective mononucleotides. The homologous enzyme from *Plasmodium falciparum* (*Pf*HGPRT) can additionally use xanthine nucleobase to form xanthine monophosphate. Balram and coworkers have shown that the chimeric *Hs*HGPRT containing first 49 residues at N-terminal from *Pf*HGPRT can use guanine, hypoxanthine and xanthine as well.[43] Similarly the human guanine deaminase (hGDA) was engineered to have to cytosine deaminase activity by computational loop remodeling and biochemical experiments, using the catalytic loop of bacterial cytidine deaminase as a template.[44]

1.7 Objective of my dissertation:

Bioinformatics and structural analyses coupled with high-performance molecular modeling techniques are valuable tools to decipher the molecular basis of nucleobase specificity.

The specific questions that I wish to address are:

1. What is the role of the nucleobase in a kinase reaction? Why do certain homologous enzymes use different nucleotides in the same phosphate transfer reaction?

2. What are the key residues at the active site of an enzyme responsible for binding a particular nucleobase?
3. Can we alter the nucleotide preference by modifying the active site?
4. Can we make an unnatural cofactor analog by engineering the biosynthesizing cofactor enzymes?

References

(1)   Meurer, F.; Do, H. T.; Sadowski, G.; Held, C. Standard Gibbs Energy of Metabolic Reactions: II. Glucose-6-Phosphatase Reaction and ATP Hydrolysis. *Biophys. Chem.* **2017**, *223*, 30–38. https://doi.org/10.1016/j.bpc.2017.02.005.

(2)   Gilman, A. G. G Proteins: Transducers of Receptor-Generated Signals. *Annu. Rev. Biochem.* **1987**, *56* (1), 615–649. https://doi.org/10.1146/annurev.bi.56.070187.003151.

(3)   Chang, Y.-F.; Carman, G. M. CTP Synthetase and Its Role in Phospholipid Synthesis in the Yeast Saccharomyces Cerevisiae. *Prog. Lipid Res.* **2008**, *47* (5), 333–339. https://doi.org/10.1016/j.plipres.2008.03.004.

(4)   Lecca, D.; Ceruti, S. Uracil Nucleotides: From Metabolic Intermediates to Neuroprotection and Neuroinflammation. *Biochem. Pharmacol.* **2008**, *75* (10), 1869–1881. https://doi.org/10.1016/j.bcp.2007.12.009.

(5)   Allen, C. M.; Kalin, J. R.; Sack, J.; Verizzo, D. CTP-Dependent Dolichol Phosphorylation by Mammalian Cell Homogenates. *Biochemistry* **1978**, *17* (23), 5020–5026. https://doi.org/10.1021/bi00616a025.

(6)   Reeves, R. E.; South, D. J. Phosphoglycerate Kinase (GTP). An Enzyme from Entamoeba Histolytica Selective for Guanine Nucleotides. *Biochem. Biophys. Res. Commun.* **1974**, *58* (4), 1053–1057. https://doi.org/10.1016/S0006-291X(74)80250-0.

(7)   Leipe, D. D.; Koonin, E. V.; Aravind, L. Evolution and Classification of P-Loop Kinases and Related Proteins. *J. Mol. Biol.* **2003**, *333* (4), 781–815. https://doi.org/10.1016/j.jmb.2003.08.040.

(8)   Ambudkar, S. V.; Kim, I. W.; Xia, D.; Sauna, Z. E. The A-Loop, a Novel Conserved Aromatic Acid Subdomain Upstream of the Walker A Motif in ABC Transporters, Is Critical for ATP Binding. *FEBS Lett.* **2006**, *580* (4), 1049–1055.

https://doi.org/10.1016/j.febslet.2005.12.051.

(9)   Kjeldgaard, M.; Nyborg, J.; Clark, B. F. The GTP Binding Motif: Variations on a Theme. *FASEB J.* **1996**, *10* (12), 1347–1368.

(10)   Leipe, D. D.; Wolf, Y. I.; Koonin, E. V.; Aravind, L. Classification and Evolution of P-Loop GTPases and Related ATPases. *J. Mol. Biol.* **2002**, *317* (1), 41–72. https://doi.org/10.1006/jmbi.2001.5378.

(11)   Letter, D.; Miiller, W.; Amons, R. Phosphate-Binding Sequences in Nucleotide-Binding. **1985**, *186* (1).

(12)   Silva, C. A. B.; Deng, J. C. The P-Loop - a Common Motif in ATP and GTP Binding Proteins. **1996**, *4* (1), 1350.

(13)   Manoj, N.; Strauss, E.; Begley, T. P.; Ealick, S. E. Structure of Human Phosphopantothenoylcysteine Synthetase at 2.3 A Resolution. *Structure* **2003**, *11* (8), 927–936. https://doi.org/10.1016/s0969-2126(03)00146-1.

(14)   Strauss, E.; Kinsland, C.; Ge, Y.; McLafferty, F. W.; Begley, T. P. Phosphopantothenoylcysteine Synthetase from Escherichia Coli. Identification and Characterization of the Last Unidentified Coenzyme A Biosynthetic Enzyme in Bacteria. *J. Biol. Chem.* **2001**, *276* (17), 13513–13516. https://doi.org/10.1074/jbc.C100033200.

(15)   Huang, J.; Nguyen, V. H.; Hamblin, K. A.; Maytum, R.; van der Giezen, M.; Fraser, M. E. ATP-Specificity of Succinyl-CoA Synthetase from Blastocystis Hominis. *Acta Crystallogr. Sect. D, Struct. Biol.* **2019**, *75* (Pt 7), 647–659. https://doi.org/10.1107/S2059798319007976.

(16)   Fraser, M. E.; James, M. N.; Bridger, W. A.; Wolodko, W. T. Phosphorylated and Dephosphorylated Structures of Pig Heart, GTP-Specific Succinyl-CoA Synthetase. *J. Mol. Biol.* **2000**, *299* (5), 1325–1339. https://doi.org/10.1006/jmbi.2000.3807.

(17)   Fraser, M. E.; James, M. N.; Bridger, W. A.; Wolodko, W. T. A Detailed Structural Description of Escherichia Coli Succinyl-CoA Synthetase. *J. Mol. Biol.* **1999**, *285* (4), 1633–1653. https://doi.org/10.1006/jmbi.1998.2324.

(18)   Wolodko, W. T.; Fraser, M. E.; James, M. N.; Bridger, W. A. The Crystal Structure of Succinyl-CoA Synthetase from Escherichia Coli at 2.5-A Resolution. *J. Biol. Chem.* **1994**, *269* (14), 10883–10890. https://doi.org/10.1016/S0021-9258(17)34141-8.

(19)   Ammelburg, M.; Hartmann, M. D.; Djuranovic, S.; Alva, V.; Koretke, K. K.; Martin, J.;

Sauer, G.; Truffault, V.; Zeth, K.; Lupas, A. N.; Coles, M. A CTP-Dependent Archaeal Riboflavin Kinase Forms a Bridge in the Evolution of Cradle-Loop Barrels. *Structure* **2007**, *15* (12), 1577–1590. https://doi.org/10.1016/j.str.2007.09.027.

(20) Mashhadi, Z.; Zhang, H.; Xu, H.; White, R. H. Identification and Characterization of an Archaeon-Specific Riboflavin Kinase. *J. Bacteriol.* **2008**, *190* (7), 2615–2618. https://doi.org/10.1128/JB.01900-07.

(21) Tachibana, S. Discovery of a New Flavin Phosphate Synthetase Which Requires Guanosine-5'-Triphosphate. *J. Vitaminol. (Kyoto).* **1967**, *13* (2), 89–92.

(22) Reeves, R. E.; South, D. J. Phosphoglycerate Kinase (GTP). An Enzyme from Entamoeba Histolytica Selective for Guanine Nucleotides. *Biochem. Biophys. Res. Commun.* **1974**, *58* (4), 1053–1057. https://doi.org/10.1016/S0006-291X(74)80250-0.

(23) Westaway, S. K.; Belford, H. G.; Apostol, B. L.; Abelson, J.; Greer, C. L. Novel Activity of a Yeast Ligase Deletion Polypeptide. Evidence for GTP-Dependent TRNA Splicing. *J. Biol. Chem.* **1993**, *268* (4), 2435–2443.

(24) Walsh, J. P.; Bell, R. M. Diacylglycerol Kinase from Escherichia Coli; 1992; pp 153–162. https://doi.org/10.1016/0076-6879(92)09019-Y.

(25) Fakas, S.; Konstantinou, C.; Carman, G. M. DGK1-Encoded Diacylglycerol Kinase Activity Is Required for Phospholipid Synthesis during Growth Resumption from Stationary Phase in Saccharomyces Cerevisiae. *J. Biol. Chem.* **2011**, *286* (2), 1464–1474. https://doi.org/10.1074/jbc.M110.194308.

(26) Bairoch, A. PROSITE: A Dictionary of Sites and Patterns in Proteins. *Nucleic Acids Res.* **1991**, *19 Suppl*, 2241–2245. https://doi.org/10.1093/nar/19.suppl.2241.

(27) Walker, J. E.; Saraste, M.; Runswick, M. J.; Gay, N. J. Distantly Related Sequences in the Alpha- and Beta-Subunits of ATP Synthase, Myosin, Kinases and Other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold. *EMBO J.* **1982**, *1* (8), 945–951.

(28) Taylor, W. R.; Green, N. M. The Predicted Secondary Structures of the Nucleotide-Binding Sites of Six Cation-Transporting ATPases Lead to a Probable Tertiary Fold. *Eur. J. Biochem.* **1989**, *179* (1), 241–248. https://doi.org/10.1111/j.1432-1033.1989.tb14547.x.

(29) Veitch, D. P.; Gilham, D.; Cornell, R. B. The Role of Histidine Residues in the HXGH Site of CTP:Phosphocholine Cytidylyltransferase in CTP Binding and Catalysis. *Eur. J. Biochem.* **1998**, *255* (1), 227–234. https://doi.org/10.1046/j.1432-1327.1998.2550227.x.

(30)   Veitch, D. P.; Cornell, R. B. Substitution of Serine for Glycine-91 in the HXGH Motif of CTP:Phosphocholine Cytidylyltransferase Implicates This Motif in CTP Binding. *Biochemistry* **1996**, *35* (33), 10743–10750. https://doi.org/10.1021/bi960402c.

(31)   Bishop, A. C.; Buzko, O.; Shokat, K. M. Magic Bullets for Protein Kinases. *Trends Cell Biol.* **2001**, *11* (4), 167–172. https://doi.org/10.1016/S0962-8924(01)01928-6.

(32)   Hwang, Y. W.; Miller, D. L. A Mutation That Alters the Nucleotide Specificity of Elongation Factor Tu, a GTP Regulatory Protein. *J. Biol. Chem.* **1987**, *262* (27), 13081–13085. https://doi.org/10.1016/S0021-9258(18)45170-8.

(33)   Shabb, J. B.; Ng, L.; Corbin, J. D. One Amino Acid Change Produces a High Affinity CGMP-Binding Site in CAMP-Dependent Protein Kinase. *J. Biol. Chem.* **1990**, *265* (27), 16031–16034. https://doi.org/10.1016/S0021-9258(17)46182-5.

(34)   Raychaudhurit, D.; Park, J. T. A Point Mutation Converts Escherichia Coli FtsZ Septation GTPase to ATPase. *J. Biol. Chem.* **1994**, *269* (37), 22941–22944.

(35)   Liu, Y.; Shah, K.; Yang, F.; Witucki, L.; Shokat, K. M. Engineering Src Family Protein Kinases with Unnatural Nucleotide Specificity. *Chem. Biol.* **1998**, *5* (2), 91–101. https://doi.org/10.1016/s1074-5521(98)90143-0.

(36)   Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M. Engineering Unnatural Nucleotide Specificity for Rous Sarcoma Virus Tyrosine Kinase to Uniquely Label Its Direct Substrates. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (8), 3565–3570. https://doi.org/10.1073/pnas.94.8.3565.

(37)   Liu, Y.; Shah, K.; Yang, F.; Witucki, L.; Shokat, K. M. A Molecular Gate Which Controls Unnatural ATP Analogue Recognition by the Tyrosine Kinase V-Src. *Bioorg. Med. Chem.* **1998**, *6* (8), 1219–1226. https://doi.org/10.1016/s0968-0896(98)00099-6.

(38)   Ulrich, S. M.; Buzko, O.; Shah, K.; Shokat, K. M. Towards the Engineering of an Orthogonal Protein Kinase/Nucleotide Triphosphate Pair. *Tetrahedron* **2000**, *56* (48), 9495–9502. https://doi.org/10.1016/S0040-4020(00)00834-6.

(39)   Vincent, F.; Cook, S. P.; Johnson, E. O.; Emmert, D.; Shah, K. Engineering Unnatural Nucleotide Specificity to Probe G Protein Signaling. *Chem. Biol.* **2007**, *14* (9), 1007–1018. https://doi.org/10.1016/j.chembiol.2007.08.006.

(40)   Bhattacharya, M.; Toth, M.; Smith, C. A.; Vakulenko, S. B. Bulky "Gatekeeper" Residue Changes the Cosubstrate Specificity of Aminoglycoside 2''-Phosphotransferase IIa.

Antimicrob. Agents Chemother. **2013**, *57* (8), 3763–3766.
https://doi.org/10.1128/AAC.00381-13.

(41)   Vashisht, K.; Verma, S.; Gupta, S.; Lynn, A. M.; Dixit, R.; Mishra, N.; Valecha, N.;
Hamblin, K. A.; Maytum, R.; Pandey, K. C.; Van Der Giezen, M. Engineering Nucleotide
Specificity of Succinyl-CoA Synthetase in Blastocystis: The Emerging Role of
Gatekeeper Residues. *Biochemistry* **2017**, *56* (3), 534–542.
https://doi.org/10.1021/acs.biochem.6b00098.

(42)   Vashisht, K.; Singh, P.; Verma, S.; Dixit, R.; Mishra, N.; Pandey, K. C. The Nucleotide
Specificity of Succinyl-CoA Synthetase of Plasmodium Falciparum Is Not Determined by
Charged Gatekeeper Residues Alone. *FEBS Open Bio* **2021**, *11* (3), 578–587.
https://doi.org/10.1002/2211-5463.13034.

(43)   Sujay Subbayya, I. N.; Sukumaran, S.; Shivashankar, K.; Balaram, H. Unusual Substrate
Specificity of a Chimeric Hypoxanthine–Guanine Phosphoribosyltransferase Containing
Segments from the Plasmodium Falciparum and Human Enzymes. *Biochem. Biophys. Res.
Commun.* **2000**, *272* (2), 596–602. https://doi.org/10.1006/bbrc.2000.2816.

(44)   Murphy, P. M.; Bolduc, J. M.; Gallaher, J. L.; Stoddard, B. L.; Baker, D. Alteration of
Enzyme Specificity by Computational Loop Remodeling and Design. *Proc. Natl. Acad.
Sci.* **2009**, *106* (23), 9215–9220. https://doi.org/10.1073/pnas.0811070106.

Nucleotide specificity of enzymes involved in flavin cofactor biosynthesis

2.1. The selectivity for nucleotide triphosphates in flavin cofactor biosynthesis

Vitamin B2 (riboflavin) is an essential biomolecule that is converted to its cofactor form to catalyze oxidation-reduction chemistry in our cells. Phosphorylation and nucleotide transfer are two critical steps among the final steps in the conversion of riboflavin (RF) to flavin cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (Figure 2.1). The enzyme 'riboflavin kinase' converts riboflavin into FMN, typically using a nucleotide triphosphate (NTP) such as ATP as a phosphoryl donor. The FMN formed in the riboflavin kinase reaction is further converted in the next step catalyzed by FMN adenylyltransferase (FMNAT) into FAD. The conversion of riboflavin into FAD can also be directly executed by the bifunctional enzyme FAD synthetase (FADS) which possess domains for both riboflavin kinase and FMNAT. Both individual reactions involve the use of a NTP and most riboflavin kinase, FMNAT and FADS enzymes characterized in literature appear to use ATP for each of these reactions. However, a few flavin cofactor synthesizing enzymes show difference in the utilization of NTP for these reactions (Table 2.1 and 2.2).
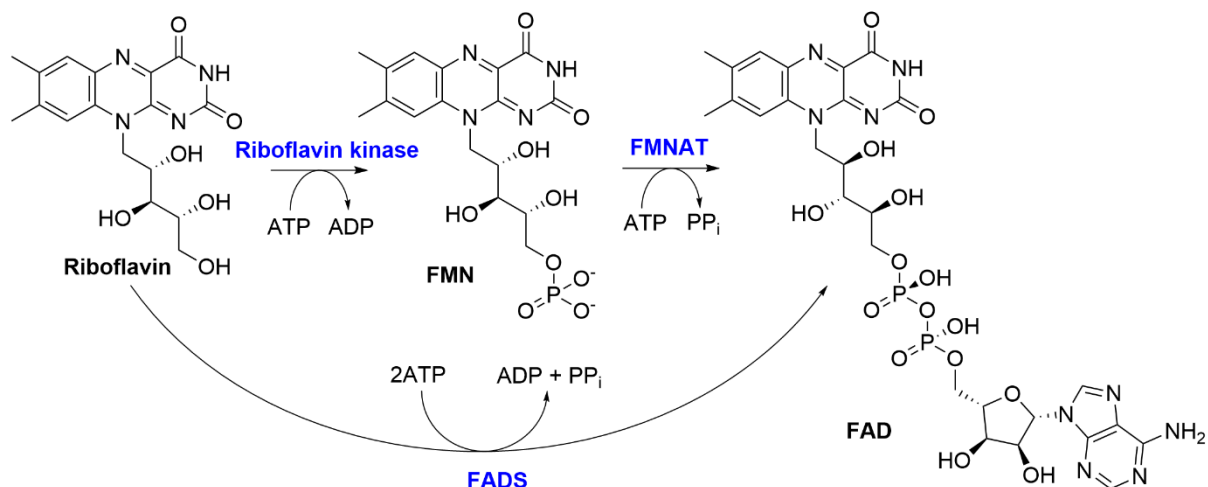
Figure 2.1: Reaction scheme for the conversion of riboflavin into FMN, followed by FAD by the enzymes riboflavin kinase, FMN adenylyltransferase (FMNAT), and FAD synthetase (FADS). Riboflavin kinase converts riboflavin into FMN by phosphorylating it with ATP. FMN is converted into FAD by

FMNAT using another molecule of ATP. Both reactions can be done by a single bifunctional enzyme FADS.

2.2 Nucleotide selectivity among riboflavin kinases from different organisms

The pattern of nucleotide utilization of riboflavin kinases differs among the species. Some of the ATP-dependent riboflavin kinases are promiscuous and may utilize nucleotides other than ATP to some extent. We note a summary of examples known in literature in Table 2.1. Some examples that stand out are as follows: for the rat intestinal mucosa and *Lactobacillus arabinosus* riboflavin kinase homologs, UTP is as effective as ATP for transferring phosphate to riboflavin.[1,2] The ATP-dependent riboflavin kinase (*RibR*) from *Bacillus subtilis* requires reducing conditions for the phosphorylation reaction and able to utilize CTP (50 %), UTP (31 %), and dATP (54 %), respectively.[3] The riboflavin kinase from *Glycine max* (soybean) and *Peptostreptococcus elsdenii* can catalyze flavokinase reaction even with ADP (22 % and 10 %, respectively).[4,5] Our experiments on the riboflavin kinases from *Homo sapiens*, *Neurospora crassa,* and *Schizosaccharomyces pombe* have also shown activity with nucleotides ATP, CTP, GTP, and UTP (Table 6.1, Chapter 6).[6–8] Notably, there are riboflavin kinases from some organisms which are not ATP-dependent and in fact very strictly choose another nucleotide for catalysing their kinase reaction i.e. they are NTP specific for a nucleotide other than ATP. The riboflavin kinase (MJ0056) from thermophilic archaea *Methanocaldococcus jannaschii* is CTP specific, unlike non-archaeal ATP specific riboflavin kinases.[9,10]Another CTP specific bifunctional riboflavin kinase/transcriptional regulator is from the archaea *Thermoplasma acidophilum*.[11]

Table 2.1: Riboflavin kinases from different organisms and their reported activities with other nucleotides.

Riboflavin kinase						
#	**Organism**	ATP	CTP	GTP	UTP	dATP
1	*Bacillus subtilis*[3]	100	50	1	31	54
2	*Homo sapiens*[6]	100	-	-	-	-
3	*Methanocaldococcus jannaschii*[9]	30	100	11	-	-
4	*Neurospora crassa*[7]	100	-	-	-	-
5	*Mus musculus*[12$]	100	> 1	> 1	-	-

6	*Mus musculus*[1#]	100	PD*	PD*	100	-
7	*Rhizopus oryzae*[13&]	0	0	100	36	-
8	*Rhizopus javanicus*[14&]	0	0	100	AD*	-
9	*Saccharomyces cerevisiae*[15]	100	-	-	-	-
10	*Schizosaccharomyces pombe*[8]	100				
11	*Glycine max (Soy bean)*[4]	100	0	0	0	-
12	*Vigna radiata (Mung bean)*[16]	100	-	-	-	-
13	*Lactobacillus arabinosus*[2]	100	0	0	AD*	-
14	*Peptostreptococcus elsdenii*[5]	100	-	-	-	-
15	*Streptococcus agalactiae*[17]	100	-	-	-	-
16	*Escherichia coli*[18]	100	-	-	-	-
17	*Thermoplasma acidophilum*[11]	-	100	-	-	-

Note: numbers represent percent activity. **$:** enzyme from rat liver; **#:** enzyme from intestinal mucosa; **&:** cell-free lysate; **PD:** Poor donor; **AD:** Active donor; ***:** values of activity not reported

The CTP-dependent riboflavin kinase enzymes are conserved throughout the archaeal family. Some of them are bifunctional and fused with a transcriptional regulator domain.[10] A distinct riboflavin kinase from the fungi *Rhizopus oryzae and Rhizopus javanicus* utilize GTP for the phosphorylation reaction to form cyclic FMN.[13,14] These enzymes, also known as cyclic-FMN synthetase, display slight activity with UTP and ITP, but not with ATP and CTP.[13,14] The GTP-dependent riboflavin kinase from the genus *Rhizopus* requires some unknown cofactor from the cell lysate for its activity, as the purified enzymes do not show activity.[13,14]
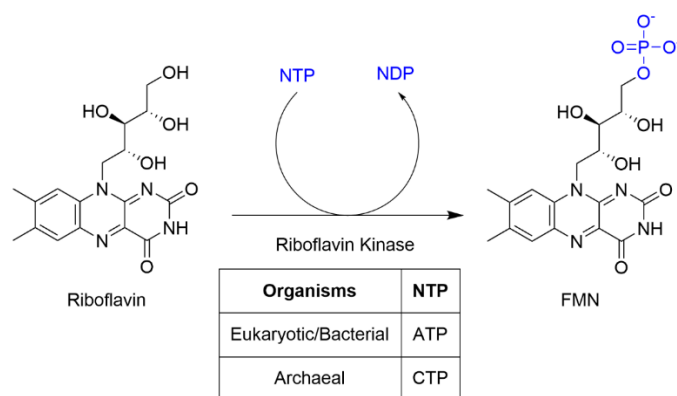
Figure 2.2: Riboflavin kinase reaction scheme. The riboflavin kinase from the eukaryotic and bacterial origin are ATP-dependent enzymes. Archaeal riboflavin kinases are CTP specific.

2.3 Nucleotide selectivity among FADS and FMNAT enzymes from different organisms

The FADS and/or FMNAT from archaeal and non-archaeal sources also have different nucleotide preferences. Most of the non-archaeal FADS or FMNAT utilize ATP for the conversion of riboflavin and FMN respectively to from FAD. The FADS enzyme homologs from *Corynebacterium ammoniagenes* and *Thermotoga maritima* are well characterized in terms of structure, activity and mechanisms.[19–23] These enzymes have two distinct domains- riboflavin kinase module and FMNAT module.

The FADS (RibC) from *Bacillus subtilis* is specific for converting the reduced form of riboflavin into FAD, and it is also ATP specific.[24,25] This gene is an essential gene, and its complete inactivation has been found to be lethal to the *Bacillus subtilis* cells.[26] Intriguingly, *Bacillus subtilis* also has a monofunctional riboflavin kinase (RibR) and yields a functional enzyme that converts riboflavin into FMN.[3,27,28] However, the corresponding monofunctional FMNAT has not been found in *Bacillus subtilis*. Several studies have been done on RibC and RibR for the regulation of flavogenensis system in *Bacillus subtilis*.[27,29–33]

Archaea have monofunctional FMNAT enzymes which can convert FMN into FAD using ATP. The FMNAT from the thermophilic archaea *M. jannaschii* has shown activity with CTP and GTP along with ATP, for the nucleotidyltransferase reaction to form analogues of FAD.[34] It is very surprising, because the presence and use of FAD analogues FCD and FGD in metabolism and other cellular pathways has not yet been established.

Table 2.2: FMNAT and FADS homologs from different organisms and their reported activities with other nucleotides.

	FMNAT* or FADS#					
	Organism	ATP	CTP	GTP	UTP	dATP
1	*Candida glabrata*[35]*	100	-	-	-	-
2	*Methanocaldococcus jannaschii*[34]*	100	40	10	-	-
3	*Escherichia. coli*[36]	100				
4	*Bacillus subtilis*[25]#	100	0	0	0	
5	*Staphylococcus aureus*[37]#	100	-	-	-	-
6	*Listeria monocytogenes*[38]*#					

7	*Corynebacterium ammoniagenes*[39]	100	0	0	0	-
8	*Saccharomyces cerevisiae*[40]*	100	-	-	-	-
9	*Homo sapiens*[41]*	100	-	-	-	-
10	*Mus musculus*[42]	100	-	-	-	-
11	*Thermotoga maritima*[23#]	100	-	-	-	-

2.4 Structural studies of flavin cofactor synthesizing enzymes and nucleobase specificity

As mentioned in Chapter 1, most of the nucleotide binding proteins have well conserved phosphate-binding domain, known as 'phosphate binding loop' or 'P-loop'.[43–45] Nucleotidyltransferase family have conserved HxGH motif for the stabilization of phosphate unit of NTP.[46–48] However, the structural studies done on riboflavin kinases, FMNAT and FADS, indicated that the substrate binding motifs in these nucleotide utilizing enzymes are entirely different from the conventional structural motifs present in ATPase and GTPase and other kinases.[6,10,11,49,50] The ATP-dependent riboflavin kinases differ structurally from other kinases. The phosphate binding motifs in the human riboflavin kinase homolog *Hs*RFK is PTAN, instead of the P-loop or glycine rich loop.[6] Similarly, among CTP-dependent archaeal riboflavin kinases there is conserved GTLN motif for the stabilization of phosphate unit of NTP.[10] There exist several other structural differences among the ATP-dependent and CTP-dependent riboflavin kinases which will be described in details in upcoming chapters. However, the binding of nucleobase moiety to these enzymes is unclear. There are available crystal structures from both group of riboflavin kinases, which show perceivable differences in the binding of nucleotide triphosphate. Thus, a comparative structural analysis of CTP and ATP binding can determine the molecular factors responsible for nucleobase specificity in riboflavin kinases.

The FADS and FMNAT contain conserved HxGH motif, which is a characteristic nucleotidyltransferase domain.[19,35] The binding of ATP for the adenylation is well-established in literature. The FMNAT (*Mj*FMNAT) from *M. jannaschii* also has HxGH motif.[34] The flavin binding site of *Mj*FMNAT is unique and does not resembles the flavin binding site of non-archaeal homologs. The gene encoding *Mj*FMNAT is conserved among the archaea, however, there is no available crystal structure of *Mj*FMNAT or its archaeal homolog. The unavailability of the crystal structure of archaeal FMNAT enzyme has restricted the proper study of the binding

mechanism of substrates and catalysis. Although there is plenty of structural and bioinformatic information available for non-archaeal FADS, little is known about archaeal FAD synthesizing enzymes. A detailed bioinformatic and structural analysis of these active flavin cofactors will be presented in upcoming chapters.

2.5 Our choice of the flavin cofactor biosynthesis pathway to study the choice of nucleotide

Flavins are redox cofactors involved in a number of metabolic pathways. FMN and FAD are utilized as cofactors in a number of flavoenzymes. Dehydrogenase, oxidase, monooxygenase, and reductase are some common examples of flavin cofactor utilizing enzymes involved in bioenergetics, photochemistry, redox homeostasis, DNA repair, protein folding and a number of other cellular processes. Impairment in the activities of these FAD synthesizing enzymes - riboflavin kinases, FMNAT and FADS - can affect the flavin homeostasis and may lead to cancer, cardiovascular diseases, anaemia, and neurological disorder. The critical role of these enzymes in the flavin biosynthesis pathways of pathogenic microorganisms make them attractive target for drug design.

Such diverse and biologically relevant features of flavin biosynthesis enzymes, motivated us to delve into the factors governing the nucleotide base specificity in them. There are several key reasons which make flavin biosynthesis pathway enzymes an excellent system to explore the molecular determinants of nucleobase specificity.

a. There are several representative X-ray and NMR solution structures of riboflavin kinases and FADS in Protein Data Bank. The availability of ATP-dependent and CTP-dependent riboflavin kinases can provide an excellent system for comparative structural and mechanistic investigation. The stereochemical quality of a modeled protein structures are enhanced if the homologous PDB structures are available. The high-resolution PDB structures also assist in substrate-docking and predicting the molecular level information correctly from molecular modeling simulations. For example, there are plenty of information available for the FADS from *C. ammoniagenes* whichrepresents eukaryotic and bacterial FADS. Its structure-function relationship is well-researched through X-ray crystallography, kinetics, spectroscopy, molecular modeling and other molecular biology techniques.[19–21,50–55] Such in-depth studies on the *C. ammoniagenes* FADS will allow us

to understand the structure and function of other uncharacterized FADs enzymes that we wish to probe.

b. Flavins are fluorescent molecules and its emission is in the visible region of electromagnetic spectrum. Thus, these can be easily detected through UV-spectroscopy and fluorescence spectroscopy.

c. Between riboflavin, FMN and FAD, these molecules have zero, one or two phosphates, respectively. The difference in the polarity of riboflavin, FMN and FAD makes their separation relatively easy through appropriate analytical techniques such as TLC and HPLC.

d. CTP-dependent riboflavin kinases and FMNAT from the archaeal families differ from their non-archaeal homologs. The flavin biosynthesis pathway in archaea is in its nascent stage and is yet to be explored. There is limited piece of information available about the flavin cofactor biosynthesis in archaea.[9,11,56,57] The exploration of structure-function relationship of flavin-synthesizing enzymes in archaea and its correlation with corresponding non-archaeal pathway will help to understand the evolutionary differences among these two distinct pathways.

With these points in mind, we embarked upon the investigation of molecular determinants of nucleotide base specificity in flavin biosynthesis pathway enzymes. These experiments were followed by the enzymatic synthesis of unnatural FAD cofactor analogs.

A summary of the chapters of this thesis that probe our investigations of the flavin biosynthesis pathway are as follows:

2.5.1 Investigation of molecular determinants of nucleobase specificity in riboflavin kinase from thermophilic archaea *M. jannaschii* – Chapter 3

The riboflavin kinase (*Mj*RibK) from *M. jannaschii* is CTP specific and its structural folds are similar to ATP-dependent riboflavin kinases. We studied the structural features of this enzyme in details through bioinformatics, molecular dynamics simulation and molecular biology experiments. Our comparative studies done on CTP-dependent and ATP-dependent riboflavin kinases led us to design mutants of *Mj*RibK having altered nucleotide specificity. The experimental details and results of this project have been given in Chapter 3.

2.5.2 Molecular basis of thermostability of riboflavin kinase from *M. jannaschii* in comparison to its mesophilic, homologous protein from *Methanococcus maripaludis* – Chapter 4

The riboflavin kinase (*Mmp*RibK) from *Methanococcus maripaludis* is a mesophilc enzyme homolog of heat-stable riboflavin kinase from *Methaocaldococcus jannaschii*. We studied the molecular factors responsible for the thermostability of *Mj*RibK through bioinformatics, molecular dynamics simulation and molecular biology experiments. We applied these factors on the mesophilic enzyme *Mmp*RibK and created some mutants to confer thermostability. The work done on the characterization of *Mmp*RibK and rational design of thermostable mutants have been given in Chapter 4.

2.5.3 Enzymatic synthesis of artificial FAD analogues - FCD, FGD, FUD from FAD-synthesizing enzymes from *M. jannaschii* and *Bacillus subtilis* – Chapter 5.

Flavin adenine dinucleotide (FAD) is a redox cofactor which is utilized in many metabolic pathways. FAD is biosynthesized either by the action of bifunctional FADS on riboflavin or by the action of FMNAT on FMN. Both of these enzymes utilize ATP molecule to form FAD. We analysed the activity of few such FAD synthesizing enzymes from *M. jannaschii* and *B. subtilis* and explored the formation of FAD analogs with NTPs other than ATP. (Chapter 5)

2.5.4 Perspective and future directions – Chapter 6

We cloned, purified and characterized several other riboflavin kinase and FADS homologs from different organisms for their NTP specificity for this project. A summary of all those results have been given in Chapter 6. The broad conclusion of the results is summarized in this chapter along with future directions.

References

(1)   Kasai, S.; Nakano, H.; Maeda, K.; Matsui, K. Purification, Properties, and Function of Flavokinase from Rat Intestinal Mucosa. *J. Biochem.* **1990**, *107* (2), 298–303. https://doi.org/10.1093/oxfordjournals.jbchem.a123042.

(2)   Snoswell, A. FLAVOKINASE OF LACTOBACILLUS ARABINOSUS 17.5. *Aust. J. Exp. Biol. Med. Sci.* **1957**, *35* (5), 427–436. https://doi.org/10.1038/icb.1957.45.

(3)   Solovieva, I. M.; Tarasov, K. V.; Perumov, D. A. Main Physicochemical Features of Monofunctional Flavokinase from Bacillus Subtilis. *Biochem.* **2003**, *68* (2), 177–181.

https://doi.org/10.1023/A:1022645327972.

(4)   Hisateru Mitsuda, Yasuko Tomozawa, F. K. Studies on Plant Flavokinase, II The Purification and Some Properties of Bean Flavokinase. *J. Vitaminol. (Kyoto).* **1963**, *148*, 142–148.

(5)   Mayhew, S. G.; Wassink, J. H. A Continuous Fluorometric Assay For Flavokinase Properties of Flavokinase From Peptostreptococcus Elsdenii. *Biochim. Biophys. Acta* **1977**, *482*, 341–347.

(6)   Karthikeyan, S.; Zhou, Q.; Mseeh, F.; Grishin, N. V.; Osterman, A. L.; Zhang, H. Crystal Structure of Human Riboflavin Kinase Reveals a β Barrel Fold and a Novel Active Site Arch. *Structure* **2003**, *11* (3), 265–273. https://doi.org/10.1016/S0969-2126(03)00024-8.

(7)   Rajeswari, S. R.; Jonnalagadda, V. S.; Jonnalagadda, S. Purification and Characterization of Flavokinase from Neurospora Crassa. *Indian J. Biochem. Biophys.* **1999**, *36* (3), 137–142.

(8)   Bauer, S.; Kemter, K.; Bacher, A.; Huber, R.; Fischer, M.; Steinbacher, S. Crystal Structure of Schizosaccharomyces Pombe Riboflavin Kinase Reveals a Novel ATP and Riboflavin-Binding Fold. *J. Mol. Biol.* **2003**, *326* (5), 1463–1473. https://doi.org/10.1016/S0022-2836(03)00059-7.

(9)   Mashhadi, Z.; Zhang, H.; Xu, H.; White, R. H. Identification and Characterization of an Archaeon-Specific Riboflavin Kinase. *J. Bacteriol.* **2008**, *190* (7), 2615–2618. https://doi.org/10.1128/JB.01900-07.

(10)   Ammelburg, M.; Hartmann, M. D.; Djuranovic, S.; Alva, V.; Koretke, K. K.; Martin, J.; Sauer, G.; Truffault, V.; Zeth, K.; Lupas, A. N.; Coles, M. A CTP-Dependent Archaeal Riboflavin Kinase Forms a Bridge in the Evolution of Cradle-Loop Barrels. *Structure* **2007**, *15* (12), 1577–1590. https://doi.org/10.1016/j.str.2007.09.027.

(11)   Rodionova, I. A.; Vetting, M. W.; Li, X.; Almo, S. C.; Osterman, A. L.; Rodionov, D. A. A Novel Bifunctional Transcriptional Regulator of Riboflavin Metabolism in Archaea. *Nucleic Acids Res.* **2017**, *45* (7), 3785–3799. https://doi.org/10.1093/nar/gkw1331.

(12)   Merrill, A. H.; Mccormickg, D. B. Affinity Chromatographic Purification and Properties of Flavokinase from Rat Liver *. *New York* **1980**, *255* (4).

(13)   Tachibana, S. Discovery of a New Flavin Phosphate Synthetase Which Requires Guanosine-5'-Triphosphate. *J. Vitaminol. (Kyoto).* **1967**, *13* (2), 89–92.

(14)   Tachibana, S. [161] GTP-Dependent Flavin Phosphate Synthetase. In *J. Vitaminol. S. Tachibana, Vitamins (Kyoto) J. Vitaminol. Vitamins (Kyoto) J. Vitaminol*; 1971; Vol. 33, pp 553–555. https://doi.org/10.1016/S0076-6879(71)18118-9.

(15)   Santos, M. A.; Jimenez, A.; Revuelta, J. L. Molecular Characterization of FMN1, the Structural Gene for the Monofunctional Flavokinase of Saccharomyces Cerevisiae. *J. Biol. Chem.* **2000**, *275* (37), 28618–28624. https://doi.org/10.1074/jbc.M004621200.

(16)   Das-panja, K.; Vidya, S. Orthophosphate Is a Non-Essential Activator of Vigna Radiata Flavokinase. *Biochem. Mol. Biol. Int.* **1999**, *47* (4), 547–554.

(17)   Clarebout, G.; Villers, C.; Leclercq, R. Macrolide Resistance Gene MreA of Streptococcus Agalactiae Encodes a Flavokinase. *Antimicrob. Agents Chemother.* **2001**, *45* (8), 2280–2286. https://doi.org/10.1128/AAC.45.8.2280-2286.2001.

(18)   KATAGIRI, H.; YAMADA, H.; IMAI, K. Biosynthesis of Flavin Coenzymes by Microorganisms. I. Bacterial Flavokinase. *J. Vitaminol. (Kyoto).* **1959**, *5*, 129–133. https://doi.org/10.5925/jnsv1954.5.129.

(19)   Herguedas, B.; Lans, I.; Sebastián, M.; Hermoso, J. A.; Martínez-Júlvez, M.; Medina, M. Structural Insights into the Synthesis of FMN in Prokaryotic Organisms. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71*, 2526–2542. https://doi.org/10.1107/S1399004715019641.

(20)   Serrano, A.; Frago, S.; Herguedas, B.; Martínez-Júlvez, M.; Velázquez-Campoy, A.; Medina, M. Key Residues at the Riboflavin Kinase Catalytic Site of the Bifunctional Riboflavin Kinase/FMN Adenylyltransferase From Corynebacterium Ammoniagenes. *Cell Biochem. Biophys.* **2013**, *65* (1), 57–68. https://doi.org/10.1007/s12013-012-9403-9.

(21)   Sebastián, M.; Serrano, A.; Velázquez-Campoy, A.; Medina, M. Kinetics and Thermodynamics of the Protein-Ligand Interactions in the Riboflavin Kinase Activity of the FAD Synthetase from Corynebacterium Ammoniagenes. *Sci. Rep.* **2017**, *7* (1), 1–14. https://doi.org/10.1038/s41598-017-07875-5.

(22)   Hagihara, T.; Fujio, T.; Aisaka, K. Cloning of FAD Synthetase Gene from Corynebacterium Ammoniagenes and Its Application to FAD and FMN Production. *Appl. Microbiol. Biotechnol.* **1995**, *42* (5), 724–729. https://doi.org/10.1007/BF00171952.

(23)   Kaiser, J. T.; Clausen, T.; Huber, R. Crystal Structure of a NifS-like Protein from Thermotoga Maritima. *J. Mol. Biol.* **2000**, *297* (2), 451.

(24)   Kearney, E. B.; Goldenberg, J.; Lipsick, J.; Perl, M. Flavokinase and FAD Synthetase from Bacillus Subtilis Specific for Reduced Flavins. *J. Biol. Chem.* **1979**, *254* (19), 9551–9557.

(25)   Coquard, D.; Huecas, M.; Ott, M.; Van Dijl, J. M.; Van Loon, A. P. G. M.; Hohmann, H. P. Molecular Cloning and Characterisation of the RibC Gene from Bacillus Subtilis: A Point Mutation in RibC Results in Riboflavin Overproduction. *Mol. Gen. Genet.* **1997**, *254* (1), 81–84. https://doi.org/10.1007/s004380050393.

(26)   Gusarov, I. I.; Kreneva, R. A.; Rybak, K. V; Podcherniaev, D. A.; Iomantas, I. V; Kolibaba, L. G.; Polanuer, B. M.; Kozlov, I. I.; Perumov, D. A. Primary Structure and Functional Activity of the Bacillus Subtilis RibC Gene. *Mol. Biol. (Mosk). 31* (5), 820–825.

(27)   Higashitsuji, Y.; Angerer, A.; Berghaus, S.; Hobl, B.; Mack, M. RibR, a Possible Regulator of the Bacillus Subtilis Riboflavin Biosynthetic Operon, in Vivo Interacts with the 5′-Untranslated Leader of Rib MRNA. *FEMS Microbiol. Lett.* **2007**, *274* (1), 48–54. https://doi.org/10.1111/j.1574-6968.2007.00817.x.

(28)   Solovieva, I. M.; Kreneva, R. A.; Errais Lopes, L.; Perumov, D. A. The Riboflavin Kinase Encoding Gene RibR of Bacillus Subtilis Is a Part of a 10 Kb Operon, Which Is Negatively Regulated by the YrzC Gene Product. *FEMS Microbiol. Lett.* **2005**, *243* (1), 51–58. https://doi.org/10.1016/j.femsle.2004.11.038.

(29)   Pedrolli, D. B.; Kühm, C.; Sévin, D. C.; Vockenhuber, M. P.; Sauer, U.; Suess, B.; Mack, M. A Dual Control Mechanism Synchronizes Riboflavin and Sulphur Metabolism in Bacillus Subtilis . *Proc. Natl. Acad. Sci.* **2015**, *112* (45), 14054–14059. https://doi.org/10.1073/pnas.1515024112.

(30)   Solovieva, I. M.; Kreneva, A.; Perumov, D. A. Gene Encodes a Monofunctional Riboflavin Kinase Which Is Involved in Regulation of The. **2017**, No. 1 999, 67–73.

(31)   Perkins, J. B.; Sloma, A.; Hermann, T.; Theriault, K.; Zachgo, E.; Erdenberger, T.; Hannett, N.; Chatterjee, N. P.; Williams, V.; Rufo, G. A.; Hatch, R.; Pero, J. Genetic Engineering of Bacillus Subtilis for the Commercial Production of Riboflavin. *J. Ind. Microbiol. Biotechnol.* **1999**, *22* (1), 8–18. https://doi.org/10.1038/sj.jim.2900587.

(32)   Kreneva, R. A.; Karelov, D. V.; Korolkova, N. V.; Mironov, A. S.; Perumov, D. A. Multifunctional Regulatory Mutation in Bacillus Subtilis Flavinogenesis System. *Russ. J.*

Genet. **2009**, *45* (10), 1256–1259. https://doi.org/10.1134/s1022795409100147.

(33) Karelov, D. V.; Kreneva, R. A.; Errais Lopes, L.; Perumov, D. A.; Mironov, A. S.
Mutational Analysis of the RibC Gene of Bacillus Subtilis. *Russ. J. Genet.* **2011**, *47* (6),
757–761. https://doi.org/10.1134/s102279541106010x.

(34) Mashhadi, Z.; Xu, H.; Grochowski, L. L.; White, R. H. Archaeal RibL: A New FAD
Synthetase That Is Air Sensitive. *Biochemistry* **2010**, *49* (40), 8748–8755.
https://doi.org/10.1021/bi100817q.

(35) Huerta, C.; Borek, D.; Machius, M.; Grishin, N. V.; Zhang, H. Structure and Mechanism
of a Eukaryotic FMN Adenylyltransferase. *J. Mol. Biol.* **2009**, *389* (2), 388–400.
https://doi.org/10.1016/j.jmb.2009.04.022.

(36) Langer, S.; Hashimoto, M.; Hobl, B.; Mathes, T.; Mack, M. Flavoproteins Are Potential
Targets for the Antibiotic Roseoflavin in Escherichia Coli. *J. Bacteriol.* **2013**, *195* (18),
4037–4045. https://doi.org/10.1128/JB.00646-13.

(37) Lohithakshan, A.; Narayanasamy, R.; Potteth, U. S.; Keshava, S.; Nagaraja, V.; Usharani,
D.; Kumar, R. Molecular Insights into the Mechanism of Substrate Binding and Catalysis
of Bifunctional FAD Synthetase from Staphylococcus Aureus. *Biochimie* **2021**, *182*, 217–
227. https://doi.org/10.1016/j.biochi.2021.01.013.

(38) Sebastián, M.; Arilla-Luna, S.; Bellalou, J.; Yruela, I.; Medina, M. The Biosynthesis of
Flavin Cofactors in Listeria Monocytogenes. *J. Mol. Biol.* **2019**, *431* (15), 2762–2776.
https://doi.org/10.1016/j.jmb.2019.05.029.

(39) Nakagawa, S.; Hagihara, T.; Fujio, T.; Aisaka, K. Metaphosphate-Dependent
Phosphorylation of Riboflavin to FMN by Corynebacterium Ammoniagenes. *Appl.
Microbiol. Biotechnol.* **1995**, *43* (2), 325–329. https://doi.org/10.1007/BF00172833.

(40) Leulliot, N.; Blondeau, K.; Keller, J.; Ulryck, N.; Quevillon-Cheruel, S.; van Tilbeurgh,
H. Crystal Structure of Yeast FAD Synthetase (Fad1) in Complex with FAD. *J. Mol. Biol.*
2010, *398* (5), 641–646. https://doi.org/10.1016/j.jmb.2010.03.040.

(41) Galluccio, M.; Brizio, C.; Torchetti, E. M.; Ferranti, P.; Gianazza, E.; Indiveri, C.; Barile,
M. Over-Expression in Escherichia Coli, Purification and Characterization of Isoform 2 of
Human FAD Synthetase. *Protein Expr. Purif.* **2007**, *52* (1), 175–181.
https://doi.org/10.1016/j.pep.2006.09.002.

(42) Oka, M.; McCormick, D. B. Complete Purification and General Characterization of FAD

Synthetase from Rat Liver. *J. Biol. Chem.* **1987**, *262* (15), 7418–7422.

(43) Silva, C. A. B.; Deng, J. C. The P-Loop - a Common Motif in ATP and GTP Binding Proteins. **1996**, *4* (1), 1350.

(44) Leipe, D. D.; Koonin, E. V.; Aravind, L. Evolution and Classification of P-Loop Kinases and Related Proteins. *J. Mol. Biol.* **2003**, *333* (4), 781–815. https://doi.org/10.1016/j.jmb.2003.08.040.

(45) Letter, D.; Miiller, W.; Amons, R. Phosphate-Binding Sequences in Nucleotide-Binding. **1985**, *186* (1).

(46) Venkatachalam, K. V.; Fuda, H.; Koonin, E. V.; Strott, C. A. Site-Selected Mutagenesis of a Conserved Nucleotide Binding HXGH Motif Located in the ATP Sulfurylase Domain of Human Bifunctional 3'- Phosphoadenosine 5'-Phosphosulfate Synthase. *J. Biol. Chem.* **1999**, *274* (5), 2601–2604. https://doi.org/10.1074/jbc.274.5.2601.

(47) Veitch, D. P.; Cornell, R. B. Substitution of Serine for Glycine-91 in the HXGH Motif of CTP:Phosphocholine Cytidylyltransferase Implicates This Motif in CTP Binding. *Biochemistry* **1996**, *35* (33), 10743–10750. https://doi.org/10.1021/bi960402c.

(48) Veitch, D. P.; Gilham, D.; Cornell, R. B. The Role of Histidine Residues in the HXGH Site of CTP:Phosphocholine Cytidylyltransferase in CTP Binding and Catalysis. *Eur. J. Biochem.* **1998**, *255* (1), 227–234. https://doi.org/10.1046/j.1432-1327.1998.2550227.x.

(49) Sebastián, M.; Velázquez-Campoy, A.; Medina, M. The RFK Catalytic Cycle of the Pathogen Streptococcus Pneumoniae Shows Species-Specific Features in Prokaryotic FMN Synthesis. *J. Enzyme Inhib. Med. Chem.* **2018**, *33* (1), 842–849. https://doi.org/10.1080/14756366.2018.1461857.

(50) Frago, S.; Martínez-Júlvez, M.; Serrano, A.; Medina, M. Structural Analysis of FAD Synthetase from Corynebacterium Ammoniagenes. *BMC Microbiol.* **2008**, *8*, 1–16. https://doi.org/10.1186/1471-2180-8-160.

(51) Serrano, A.; Frago, S.; Velázquez-Campoy, A.; Medina, M. Role of Key Residues at the Flavin Mononucleotide (FMN): Adenylyltransferase Catalytic Site of the Bifunctional Riboflavin Kinase/Flavin Adenine Dinucleotide (FAD) Synthetase from Corynebacterium Ammoniagenes. *Int. J. Mol. Sci.* **2012**, *13* (11), 14492–14517. https://doi.org/10.3390/ijms131114492.

(52) Frago, S.; Velázquez-Campoy, A.; Medina, M. The Puzzle of Ligand Binding to

Corynebacterium Ammoniagenes FAD Synthetase. *J. Biol. Chem.* **2009**, *284* (11), 6610–6619. https://doi.org/10.1074/jbc.M808142200.

(53) Marcuello, C.; Arilla-Luna, S.; Medina, M.; Lostao, A. Detection of a Quaternary Organization into Dimer of Trimers of Corynebacterium Ammoniagenes FAD Synthetase at the Single-Molecule Level and at the in Cell Level. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**, *1834* (3), 665–676. https://doi.org/10.1016/j.bbapap.2012.12.013.

(54) Serrano, A.; Sebastián, M.; Arilla-Luna, S.; Baquedano, S.; Pallarés, M. C.; Lostao, A.; Herguedas, B.; Velázquez-Campoy, A.; Martínez-Júlvez, M.; Medina, M. Quaternary Organization in a Bifunctional Prokaryotic FAD Synthetase: Involvement of an Arginine at Its Adenylyltransferase Module on the Riboflavin Kinase Activity. *Biochim. Biophys. Acta - Proteins Proteomics* **2015**, *1854* (8), 897–906. https://doi.org/10.1016/j.bbapap.2015.03.005.

(55) Herguedas, B.; Martínez-Júlvez, M.; Frago, S.; Medina, M.; Hermoso, J. A. Oligomeric State in the Crystal Structure of Modular FAD Synthetase Provides Insights into Its Sequential Catalysis in Prokaryotes. *J. Mol. Biol.* **2010**, *400* (2), 218–230. https://doi.org/10.1016/j.jmb.2010.05.018.

(56) Sato, T.; Atomi, H. Novel Metabolic Pathways in Archaea. *Curr. Opin. Microbiol.* **2011**, *14* (3), 307–314. https://doi.org/10.1016/j.mib.2011.04.014.

(57) Illarionov, B.; Eisenreich, W.; Schramek, N.; Bacher, A.; Fischer, M. Biosynthesis of Vitamin B2: Diastereomeric Reaction Intermediates of Archaeal and Non-Archaeal Riboflavin Synthases. *J. Biol. Chem.* **2005**, *280* (31), 28541–28546. https://doi.org/10.1074/jbc.M503093200.

Chapter 3

Molecular determinants of nucleotide specificity in CTP-dependent archaeal riboflavin kinase from *Methanocaldococcus jannaschii*

3.1 Introduction

The enzymes riboflavin kinase, FMN adenylyltransferase (FMNAT), and FAD synthetase (FADS) play a crucial role in the final steps of the riboflavin biosynthesis pathway to form FMN and FAD (Chapter 2, Figure 2.1).[1] The monofunctional riboflavin kinase belongs to the transferase class of enzymes and they catalyze the phosphorylation of riboflavin (RF) to produce FMN. Most of the riboflavin kinases from bacteria, fungi, and other eukaryotes studied in literature utilize ATP as a phosphate donor for this reaction, and are often promiscuous for the nucleotide triphosphate (NTP) used.[2–4] However, it has been found that riboflavin kinases from archaea can be very specific for using CTP, which is an intriguing feature (Figure 3.1).[5–7]
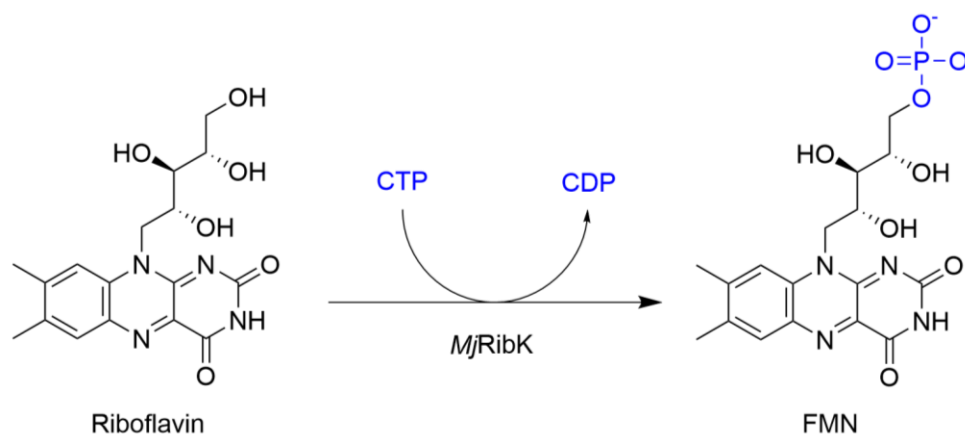
Figure 3.1: Conversion of riboflavin into FMN by riboflavin kinase from *Methanocaldococcus jannaschii* (*Mj*RibK) in the presence of CTP as a phosphoryl donor.

The studies done on riboflavin kinase from *Methanocaldococcus jannaschii* (*Mj*RibK) indicated that it uses CTP as a sole source of phosphate in converting riboflavin into FMN.[6] The non-archaeal riboflavin kinases have generally shown promiscuity in their NTP utilization as illustrated by several examples noted in Table 2.1, Chapter 2. The studies on the pattern of nucleotide utilization by kinases are restricted, as indicated in chapter 1, and there exist a few studies on the nucleotide specificity of riboflavin kinases, FMN adenylyltransferases, and FAD synthetases as summarized in Chapter 2, Table 2.1, Table 2.2.

Furthermore, there exist no studies that attempt to alter the nucleotide specificity of either of these enzymes involved in flavin cofactor biosynthesis. Protein engineering has emerged as a branch of enzymology that deals with the structure-function relationships by modifying substrate specificity and the stability of enzymes. In this regard, the engineering of v-Src kinase to attain novel substrate specificity towards synthetic nucleotide substrate analog is an elegant example.[8–10] In another example, the nucleotide specificity of the succinyl Co-A synthetase enzyme from *Blastocystis sp.* was altered from ATP to GTP by investigating the role of gatekeeper residues.[11]

To investigate the molecular determinant of nucleotide specificity, we chose riboflavin kinases as a model system. This chapter explores the molecular factors governing the nucleotide base specificity in *Mj*RibK, which is a CTP specific enzyme, and its homologs are conserved among archaea. Following are some unique features of this enzyme that differentiate it from other non-archaeal riboflavin kinases.[6]

(a) For the phosphorylation of riboflavin to form FMN, it strictly utilizes CTP instead of ATP.

(b) *Mj*RibK does not require a divalent metal ion for its activity.

(c) Being an enzyme from a hyperthermophilic source, it shows activity up to 95 °C -100 °C.

We compared the structure of *Mj*RibK with other non-archaeal ATP-dependent riboflavin kinases. Our bioinformatics and structural analyses, and molecular modeling studies have enabled us to elucidate the mechanistic differences among archaeal and non-archaeal riboflavin kinases. We further reported the key residues and structural motifs responsible for CTP binding in archaeal riboflavin kinase. Based on this molecular-level information, we altered the nucleotide specificity of *Mj*RibK from CTP to ATP.

3.2 Materials and methods

Materials: Genomic DNA of *Methanocaldococcus jannaschii* DSM 2661 was a gift from Biswarup Mukhopadhyay at Virginia State University. The primary assay reagents riboflavin and FMN are from Sigma-Aldrich, and the nucleotide triphosphates were obtained from Jena Biosciences. All other reagents and media components used in cloning, protein expression, purification, and enzyme assays were obtained from Sigma-Aldrich, Tokyo Chemical Industries (TCI) chemicals, Rankem, Sisco Research Laboratory (SRL) chemicals, and Himedia unless otherwise stated. Polymerase chain reaction (PCR) reagents and kits were from Agilent and

Himedia. The high-fidelity PrimeSTAR GXL DNA polymerase from DSS Takara Bio USA was used for cloning.

3.2.1 Cloning, overexpression, and purification of *Mj*RibK and its mutants

The MJ0056 gene from the *Methanocaldococcus jannaschii* was cloned into a pET28a plasmid vector between BamHI and NdeI restriction sites, using restriction-free cloning.[12] Cloning was done in three steps as follows:

(a) The first round of PCR was done to amplify the MJ0056 gene from the genomic DNA of *Methanocaldococcus jannaschii* using the first set of primers (*Mj*RFK_F1 and *Mj*RFK_R1; Table 1).

(b) The second round of PCR adds flanking sequences at terminals of the gene that are complementary to the insertion sites of the pET28a vector. This step was done using the second set of primers (*Mj*RFK_F2, *Mj*RFK_R2; Table 1).

(c) The gene obtained in the second round of PCR is called megaprimer, which was used to amplify the whole plasmid vector with the gene inserted between the restriction sites mentioned above.

(d) The parent plasmid was digested using Dpn1 (New England Biolab) and transformed into *E. coli* DH5α chemical competent cells for the isolation of cloned plasmid vector.

For the cloning of mutants, the wild-type *Mj*RibK plasmid vector was used as a template and the primers used are given in Table 1. These primers contain the altered codon at the site of mutation. The cloning of the mutants was done in two rounds of PCR. The first round of PCR was done either with the T7 promoter or terminator universal primer (Table 1) and corresponding forward or reverse primer of mutant. It amplified the fragment of plasmids from the T7 promoter/terminator site to the mutation site. This fragment was utilized as a megaprimer in the second round of PCR, to amplify the whole plasmid introduced with mutation.

The purified plasmid was transformed into *E. coli* BL21(DE3) cells. 40 µg/mL kanamycin was added to 100 mL of autoclaved LB broth. It was inoculated with a single colony of *E. coli* BL21(DE3) cells containing the *Mj*RibK plasmid. The culture was incubated overnight at 37 °C and 180 RPM inside an incubator shaker. 2 L autoclaved LB media supplemented with 40 µg/mL kanamycin was inoculated with 1% starter culture. These were incubated at 37 °C with continuous shaking. When the OD_{600} reached 0.6, it is induced with 1 mM IPTG and left at 18

°C for 16 h inside the incubator shaker. After 16 hours, the culture was centrifuged at 6000 RPM to collect the cell pellets.

The cell pellets thus obtained were re-suspended into lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1 mM PMSF, 20 mM Imidazole, 0.025 % β-mercaptoethanol). The cells were lysed by sonication (60 % amplitude, 1 s on- 3 s off cycle for 20 min), and the supernatant was separated from the debris by centrifugation. For the purification of enzyme AKTA™ Pure FPLC system (GE Healthcare Bio-sciences Ltd.) was used. The supernatant containing the protein was loaded on a pre-equilibrated 5 mL Ni-NTA His-Tag column (GE Healthcare Bio-sciences Ltd.). The equilibration buffer was 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole and 0.025 % β-mercaptoethanol. The unbound proteins were washed with 20 mM and 60 mM imidazole buffer, and the final elution was done with 250 mM imidazole buffer. 15 % SDS-PAGE gel electrophoresis was done to check the protein of desired molecular weight in eluted fractions. The fractions containing enzymes were pooled and dialyzed to remove imidazole. 10 % glycerol was added to the enzyme, flash-frozen, and stored at -80 °C. Protein concentration was determined before the experiments by standard Bradford assay using Bovine serum albumin (BSA) as a standard.[13]

The expression and the purification of the *Homo sapiens* riboflavin kinase (*Hs*RFK) and other mutants were done following the same protocol as done for wild-type *Mj*RibK. Note that in this thesis, we refer to all archaeal riboflavin kinases as RibK and all other non-archaeal riboflavin kinases as RFK.

Table 3.1: List of the primers used in the cloning

#	Primer name	Sequence
1	*Mj*RibK_F1	5'-TTGGTGAAATTGATGATTATTG -3'
2	*Mj*RibK_R1	5'-TTATTCATCTTTATCTCCC -3'
3	*Mj*RibK_F2	5'-GCCGCGCGGCAGCCATATG GTGAAATTGATGATTATTG -3'
4	*Mj*RibK_R2	5'-CGACGGAGCTCGAATTCGGATCCTTATTCATCTTTATTCCC-3'
5	L44A_F1	5'-GAACCTTATGAGGGGACAGCGAATTTAAAATTAGATAG -3'
6	L44F_F1	5'-GAACCTTATGAGGGGACATTTAATTTAAAATTAGATAG -3'
7	R115A_F1	5'-GCCCCAATGAAACTTGCGGAGCAATTTAATTTAAAGG-3'
8	R115P_R1	5'-ATGAAACTTCCGGAGCAATTTAA -3'
9	R115PP_R1	5'-ATGAAACTTCCGCCGGAGCAATTTAA-3'
10	Chi_*Mj*RibK_T7F_2_R	5'ATAAAAATCTTCTTTAAAGGTATGCATAATTATCTCTATAATCTCACTACTATGA-3'
11	Chi_*Mj*RibK_F_2_T7R	5'-ACCTTTAAAGAAGATTTTTATGGCGAAATTATAAAAATACTAATTAAGGGAGATA-3'

12	T7 Promoter_F	5'-TAATACGACTCACTATAGGG-3'
13	T7 Terminator_R	5'-GCTAGTTATTGCTCAGCGG-3'

3.2.2 Activity assay of *Mj*RibK and mutants with different nucleotide triphosphates

Next, we tested the activity of *Mj*RibK against different nucleotides. 1 µM of the enzyme was added to the mixture of 300 µM riboflavin, 4 mM/10 mM nucleotide triphosphate (ATP/CTP/GTP/UTP), 100 mM Tris-HCl pH 8.0, 5 mM magnesium chloride in a total volume of 200 µL at 70 °C to initiate the reaction. The reaction was quenched with 60 mM acetic acid after 4 h. Primary verification of the enzymatic products was done with thin-layer chromatography (TLC) (Merck Silica Gel 60 $F_{254)}$ using water, acetic acid, and n-butanol in a 5:1:4 ratio as solvent. The retention time of riboflavin and FMN in this solvent system were 0.45 and 0.2, respectively.

The conversion of riboflavin into FMN under these conditions was analyzed on the C-18 reverse-phase column (Phenomenex, Gemini 5 µm NX-C18 110 Å, 250 × 4.6 mm) Agilent HPLC system (1260 Infinity II) equipped with UV-Vis detectors. The reaction product was centrifuged at 14000 RPM to remove the denatured enzyme, and 20 µL of the supernatant was loaded onto the column. The HPLC method used was the modification of the method as reported previously.[6] The mobile phases were 10 mM ammonium acetate, pH 6.5 (Solvent A), and methanol (Solvent B). The method consisted of the following steps. 0-5 min 95 % A, 5-40 min from 95 % A to 20 % A, 40-60 min from 20 % A to 95 % A. The detection of analytes was done at 260 nm, 280 nm, 365 nm, and 440 nm. The retention times of riboflavin, FMN, and FAD under these conditions are 31.5 min, 28.3 min, and 26 min, respectively. The activity assays of various mutants were similarly done against different nucleotides as done for the wild-type enzyme.

3.2.3 Bioinformatics: sequence and structural analysis

To gather sequences of CTP-dependent archaeal riboflavin kinases, we used the NCBI – Basic Local Alignment Search Tool (BLAST).[14–16] The selected sequences were subjected to multiple sequence alignment using MUSCLE.[17,18] The output of the alignment result was visualized with BoxShade to highlight the conserved domains present in archaeal riboflavin kinases. Next, we did the phylogenetic analysis to check the relatedness of archaeal and non-

archaeal riboflavin kinases. The sequences for non-archaeal riboflavin kinases were searched by another BLAST search using the amino acid sequence of ATP-dependent *Hs*RFK.[2] Multiple sequence alignment of selected archaeal and non-archaeal riboflavin kinases was done using MUSCLE. The extra spaces and sectors having bad alignments were removed using BioEdit and saved as Phylip 4.0 alignment file. It was subjected to maximum likelihood/rapid bootstrapping analysis on RaXML-HPC2 on XSEDE on the CIPRES gateway to produce a phylogenetic tree. The visualization and edits of the phylogenetic tree were made on FigTree. Additionally, structural alignment of *Mj*RibK (PDB:2VBU) and *Hs*RFK (1Q9S) to compare the conserved domains was done using the Chimera Match-Align tool. Visualization and structural analyses were done with Chimera and VMD visualization softwares.

3.2.4 Docking and MD simulation of the predicted mutants (in collaboration with Reman Singh and Arnab Mukherjee, IISER Pune)

The PDB structures of the mutants of *Mj*RibK were modeled using Modeller.[19] The stereochemical quality of the modeled structures was estimated using programs Procheck and ERRAT.[20,21] Docking of the ligands riboflavin CTP, and ATP were done using Autodock 4.[22,23]

We docked ATP in the *Mj*RibK structure and then performed a molecular dynamics (MD) simulation. We used AMBER99SB force field for the protein.[24] However, to create the AMBER force field of CTP, ATP and riboflavin we used antechamber module of AmberTool. First, we conducted a quantum mechanical optimization of the ATP and CTP using HF/6-31G* basis set in GAUSSIAN03 software and then restricted electrostatic potential (RESP) charges on the atoms of ATP, CTP and RF were calculated.[25] The topology and coordinates generated from AmberTools were converted into the GROMACS format using a perl program amb2gmx.pl. All molecular dynamics simulations were done using GROMACS package.[26] We put the NTP-RF bound protein in a cubic box of length 70 Å. Then, the system was solvated with ~2000 TIP3P water molecules. Further, we added 150 mM $MgCl_2$ ion solution after which we optimized the system using steepest descent method, followed by heating the system to 300 K using the Berendsen thermostat with coupling constant of 2 ps (NVT condition). After that, we have performed 4 ns NPT simulation to adjust the box size. Finally, we performed production run for 300 ns at 300 K using Nose-Hoover thermostat. All the analysis was done over the production run.

3.3.1 Bioinformatic and phylogenetic analysis suggested the differences among CTP-dependent riboflavin kinases and ATP-dependent riboflavin kinases

The multiple sequence alignment of CTP-dependent archaeal riboflavin kinases indicated several conserved domains among these, which are different from the conserved domains of ATP-dependent non-archaeal riboflavin kinases.[2,5,27] The phosphate-binding motif in archaeal riboflavin kinase is YxGTLN (Figure 3.2). GLGEGR is a glycine-rich loop equivalent in archaeal riboflavin kinases. The isoalloxazine ring of the flavin is inside a hydrophobic pocket near the conserved flavin binding motif PKKTxHxx.
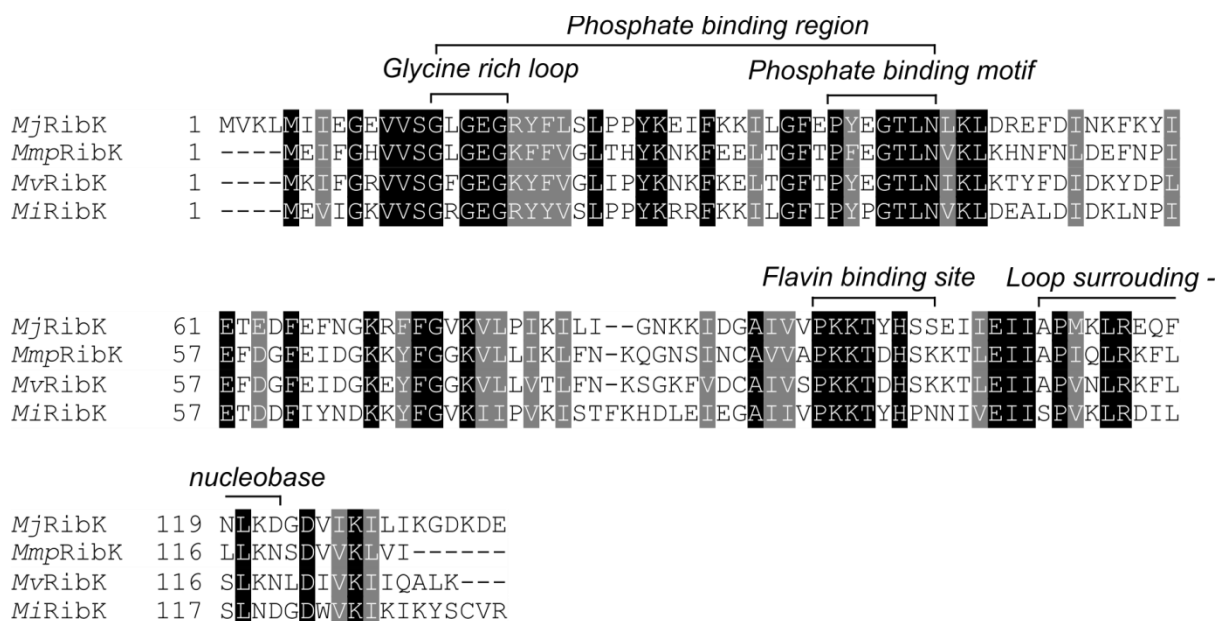
Figure 3.2: Multiple sequence alignment of some archaeal riboflavin kinases (RibK) showing the conserved domains for flavin and nucleotide triphosphate binding (*Mj: Methanocaldococcus jannaschii, Mmp: Methanococcus maripaludis, Mv: Methanococcus vannielii, Mi: Methanotorris igneus*).

The phylogenetic tree of archaeal riboflavin kinases (referred to as RibKs) and non-archaeal riboflavin kinases (referred to as RFKs) indicated that both groups are distinct with regards to their sequences. However, ATP-dependent riboflavin kinases from bacterial and eukaryotic domains showed a close resemblance among them (Figure 3.3).[5]
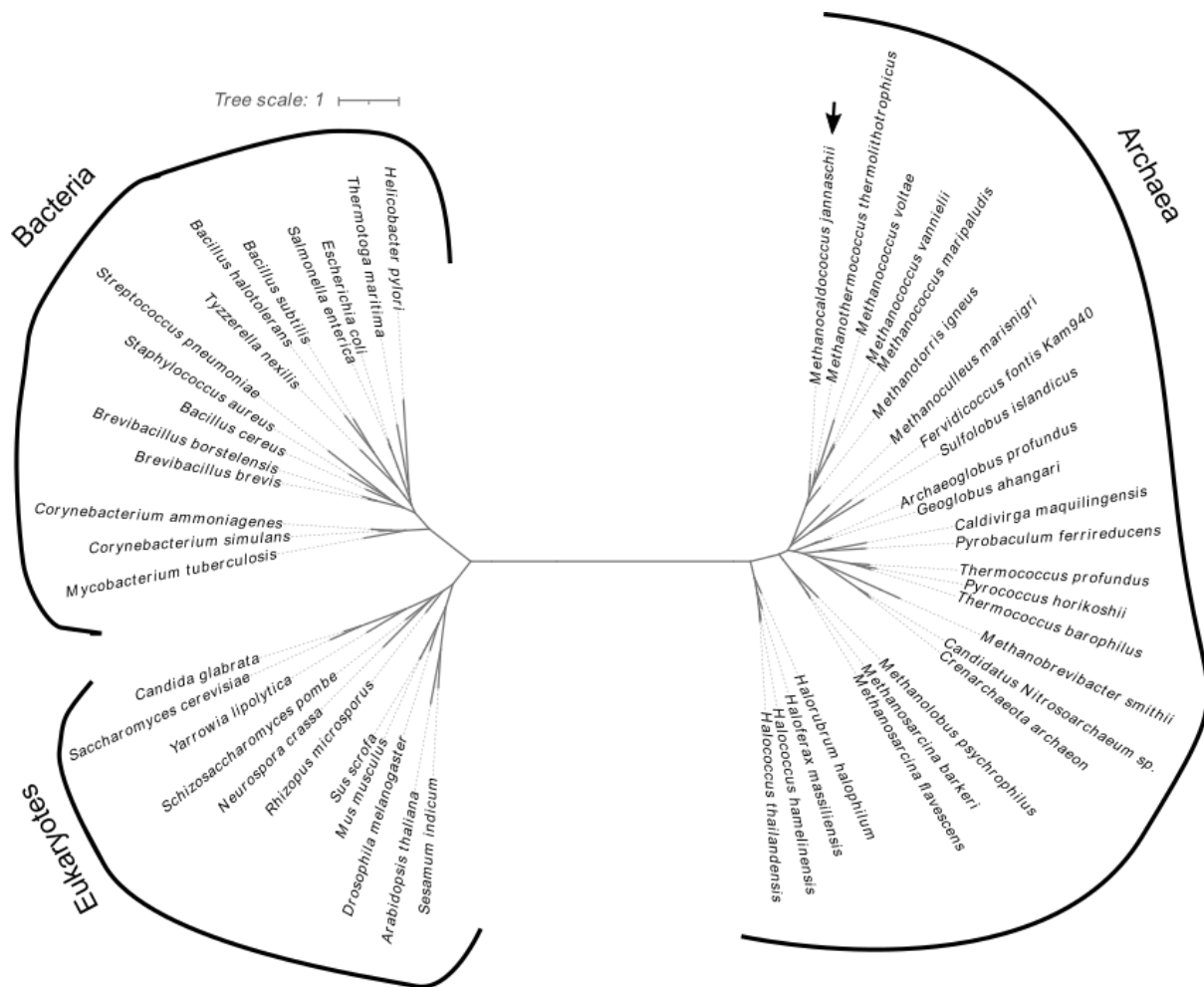
Figure 3.3: Phylogenetic tree of riboflavin kinases from archaeal, bacterial and eukaryotic organisms. The archaeal riboflavin kinases are very different from the bacterial and eukaryotic riboflavin kinases.

3.3.2 Activity assay of *Mj*RibK with different nucleotide triphosphates indicated that it has very strict utilization of CTP as phosphoryl donor

The Ni-NTA affinity purification resulted in the elution of 16 kDa protein (Figure 3.4 A). We tested the activity of *Mj*RibK with different nucleotides to validate its nucleotide specificity. The enzyme was found to be specific for CTP as a phosphoryl donor (Figure 3.4 B). The results are similar at a range of temperatures up to 95°C, this enzyme does not use nucleotide triphosphates other than CTP to convert riboflavin into FMN.[6] However, it is previously reported to show 30% and 11% conversion with ATP and GTP, respectively.[6] We varied the concentration of NTPs from 4 mM to 10 mM in the reaction to check the ability of this enzyme to accept NTP other than CTP at a higher concentration. However, even at 10 mM NTP

concentration, it is specific for CTP (Figure 3.4 C, D). These results suggest that *Mj*RibK is very specific for CTP and does not utilize NTPs other than CTP for the phosphorylation reaction. This result is contrary to the specificities shown by ATP-dependent riboflavin kinases from *Homo sapiens*, *Schizoschacromyces pombe*, or *Neurospora crassa*, which shows promiscuity in their NTP utilization (Chapter 6).[2,3,27]
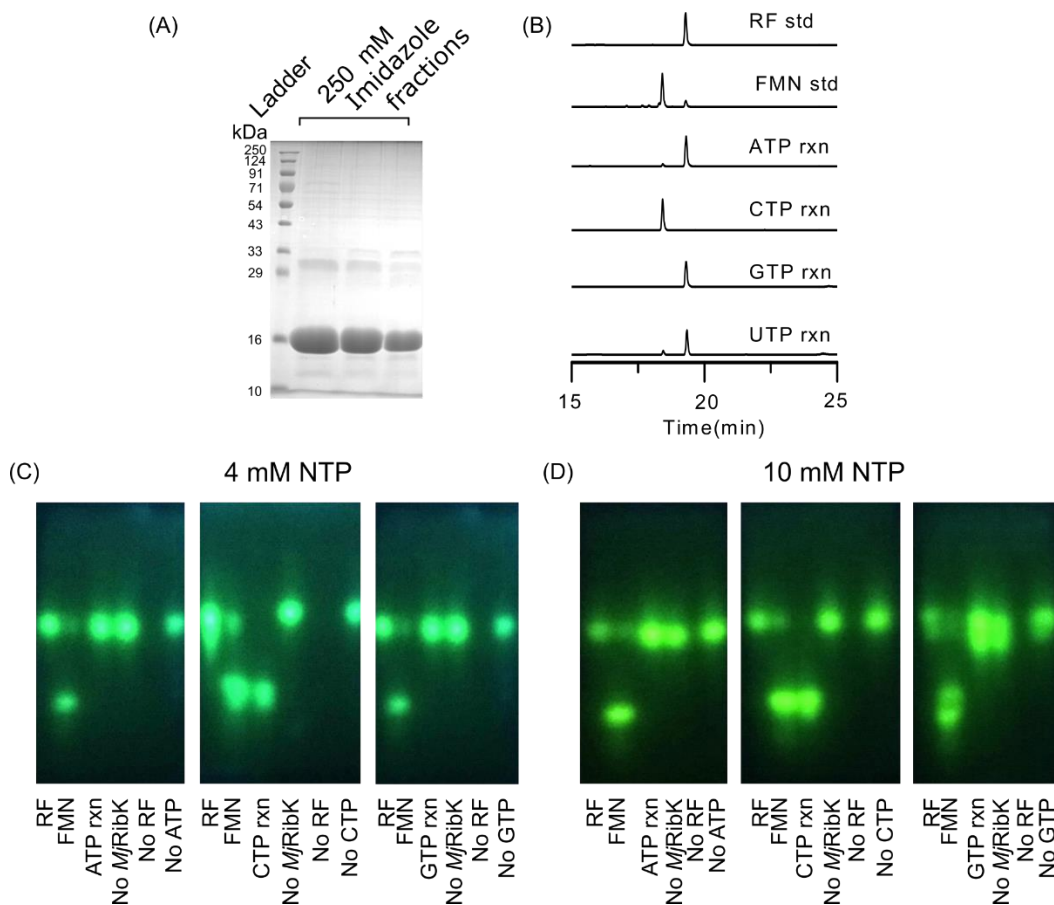
Figure 3.4: Characterization and activity assay of *Mj*RibK with NTPs. **(A)** 15% SDS-PAGE gel showing the 16 kDa band of *Mj*RFK. **(B)** HPLC chromatogram showing the reactions of *Mj*RibK with NTPs. The enzyme *Mj*RibK shows complete conversion of riboflavin into FMN with CTP as a phosphoryl donor. It gives slight activity with ATP (6 %) and UTP (11 %). **(C)** TLC showing the conversion of riboflavin into FMN only with CTP when 4 mM NTP concentration were used. **(D)** TLC showing the conversion of riboflavin into FMN only with CTP when 10 mM NTP concentration were used. (Experiments were done in duplicates)

3.3.3 Comparative structural analyses of *Mj*RibK with ATP-dependent riboflavin kinase *Hs*RFK

The sequence alignment of *Mj*RibK with other CTP-dependent riboflavin kinase and the conserved domains is depicted in Figure 3.2. The functions of various conserved domains in these archaeal riboflavin kinases are reported in detail in literature.[5]

We compared both of these structures to understand the structural basis for the CTP-specificity *MjRFK* (PDB:2VBU) and ATP-dependent *Hs*RFK (PDB:1Q9S).[2,5] The comparative structural alignment of these two enzymes has indicated some structural differences among these two enzymes (Figure 3.5 A, B). One of the prominent structural features this archaeal enzyme has is conserved 40-YxGTLN-45 motif in it. The residues Y-40 and L-44 of the YxGTLAN motif form the cytosine pocket in *Mj*RibK, as shown in Figure 3.6. The cytosine nucleobase of CTP stacks between the GTLN motif and a loop having conserved PxxLRxxxxL motif. The residue Y-40 forms a stacking interaction with the cytosine nucleobase. This motif is not present in other ATP-dependent riboflavin kinases, and they have a PTAN motif instead. Among ATP-dependent riboflavin kinases, this motif is a corresponding substitution of the glycine-rich loop of many ATPase and GTPase. The adenine pocket in the structure of *Hs*RFK is formed primarily with the residues P-33 of PTAN motif and F-97 of the outer loop surrounding the adenine ring. The adenine-nucleobase binding is assisted by H-91, I-89, and H_2O, which interacts with F-93.

(A) Phosphate binding Motif   Loop surrounding nucleobase

                β2   β5   β6
    *Hs*RibK   32 -IPTAN   84 SMETHIMHTFK--ED---FYG-EIINV
    *Mj*RibK   40 YEGTLN   105 IIEIIAP-MK-LRE-QFNLKDGDVIKI
                β2   β8   α3   β9

(B) *Hs*RibK   1 IMR-HLPYFCRGQVVRGFGRGSKQ-LG-----------------IPTANFPEQV-VDNLP
     *Mj*RibK   1 ---MVKLMIIEGEVVSGLGEGR-YFLSLPPYKEIFKKILGFEPYEGTLNLK--LDRE-F-

     *Hs*RibK   41 ADI-ST------------------GIYYGWASVGSG-DVHKMVVSIGWN-PYYKNTK-
     *Mj*RibK   53 --DIN-KFKYIETEDFEFNGKRFFGVKVLPIKILI-G-NKKIDGAIV-V-PKK------T

     *Hs*RibK   77 ---K-SMETHIMHTFK--ED---FYG-EIINVAIVGYLRPEKNFDSLESLISAIQGDIEE
     *Mj*RibK 100 YHSSEIIEIIAP-MK-LRE-QFNLKDGDVIKILIK------------------------

     *Hs*RibK 127 AKKRLELPEYLKIKEDNFFQVSK-----
     *Mj*RibK 132 ---------------------GDKDE

Figure 3.5: Comparative structural alignment of *Mj*RibK (PDB:2VBU) and *Hs*RFK (PDB:1Q9S). (A) Phosphate binding motif and nucleobase binding loop in *Hs*RFK and *Mj*RibK. The cytosine-nucleobase binding loop contains a small α-helix (α-3) in the middle. (B) The complete structural alignment of *Hs*RFK and *Mj*RibK. It shows very little conservation among them.

The most observable difference among the active site pockets of cytosine and adenine in the crystal structures of *Mj*RibK (PDB:2VBU) and *Hs*RFK (PDB:1Q9S) is the outer loop around the nucleobase. In the *Mj*RibK crystal structure, this loop (between β8-β9) contains a small alpha helix α3 (LREQF) (Figure 3.6 A). The ribose 3'-OH of CTP interacts with conserved R-115 present on the helix of the outer loop surrounding the cytosine base. The presence of this helix restricts the entry of bulky adenine ring into the active site pocket, while smaller cytosine ring can enter easily into it.[5] On the other hand, in *Hs*RFK, the outer loop (β5-β6) is wide enough to access adenine (Figure 3.6 B).[2] It has been predicted in the study by Ammelburg et al. that the mutation of the two large hydrophobic groups, Y-40 and L-44, with smaller residues, can disrupt the small α-helix on the outer loop of *Mj*RibK. It might widen the active site cavity and might confer ATP-specificity to it.[5]
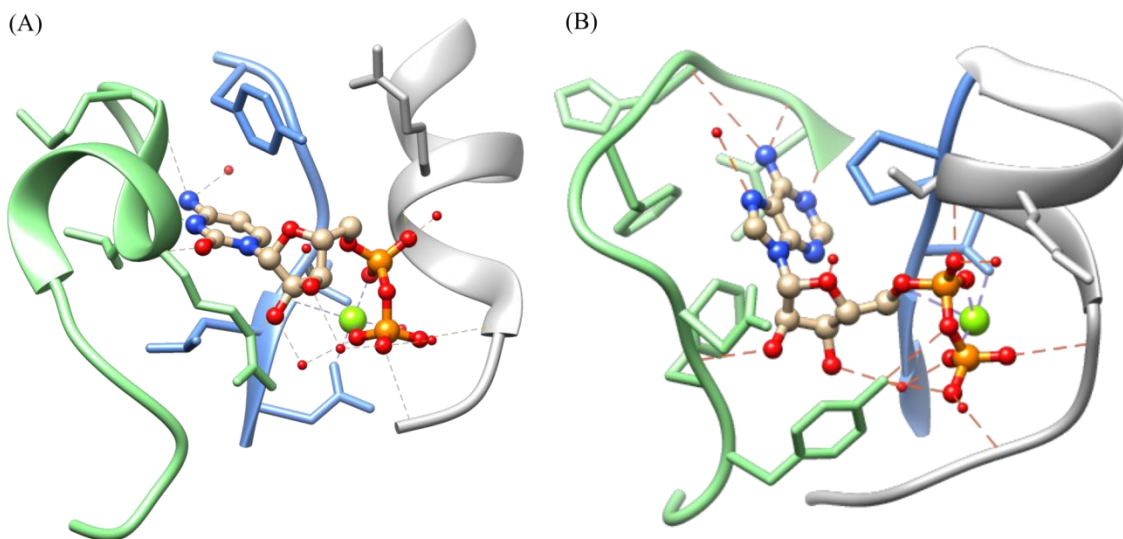
Figure 3.6: Active site architecture of **(A)** *Mj*RibK (PDB:2VBU) and **(B)** *Hs*RFK (PDB:1Q9S) with bound CTP and ATP respectively

3.3.4 Mutational studies done based on the structural comparison conferred altered nucleotide specificity to *Mj*RibK.

From the structural analyses, it was clear that the motif YxGTLN, forms the interior of the cytosine binding pocket. The outer loop containing a small helix, which surrounds the nucleobase, controls the access of CTP and helps in the binding of cytosine nucleobase. Thus, we decided to create mutations in YxGTLN motif and cytosine surrounding loop based on structural analyses. More specifically, we selected the residue L44 of YxGTLN motif, and the residue R115 of the nucleobase surrounding loop, critical for cytosine binding.[5]

Our reasoning for mutating the L-44 residue is to make the cytosine pocket either large or hydrophobic. Thus, in this way, it can accommodate a bulkier ATP molecule. The L44F mutation can't lead to an increase in the size of the active site. However, it can increase the hydrophobicity of the nucleobase binding site and thus capable of stacking with adenine-nucleobase.

Table 3.2: List of mutants and corresponding binding motif

#	Mutant	Description of mutation
1	L44A	Mutation at YxGTLN site
2	L44F	Mutation at YxGTLN site
3	R115A	Mutation at the α-helix of nucleobase surrounding loop
4	L44F R115A	Combination of #2 and #3
5	R115P	Mutation at the α-helix of nucleobase surrounding loop
6	R115PP	Mutation at the α-helix of nucleobase surrounding loop

We further designed the mutants involving the R115 position. The mutation of residue R115 with alanine (R115A) was done to investigate the effect of R115 on the binding of CTP through 3'-OH of the ribose ring. The proline mutants were created to disrupt the α-3 helix present on the outer loop of the *Mj*RibK enzyme (Figure 3.6 A).

MD simulations further confirmed our assertion of mutant design for altered nucleotide specificity. Several mutants were created *in silico*, and ATP was docked into them. The interaction energy of these mutants and wild-type *Mj*RibK were calculated with ATP and compared with the interaction energy of CTP in the wild-type enzyme (Figure 3.7).
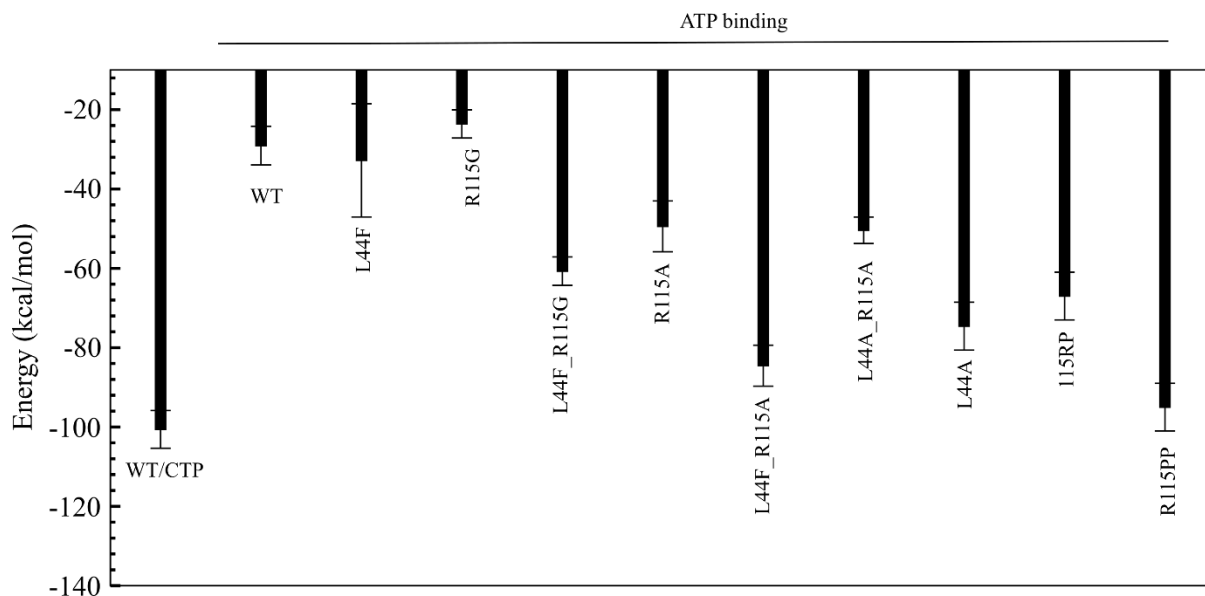
Figure 3.7: Interaction energy of wild-type *Mj*RibK with CTP and its comparison with the interaction energies of mutants

We observed that both of these mutants, L44A and L44F could not accept ATP for the phosphorylation reaction. At the same time, these mutations did not suffer a loss in activity with CTP (Figure 3.8 A, B). These results indicates that the modifications only along the YxGTLN motif are insufficient to impart altered nucleobase specificity to *Mj*RibK.

The mutant R115A has not shown any altered preference for NTPs other than CTP (Figure 3.8 C). The double mutant L44F R115A was also unable to accept ATP for phosphorylation. Surprisingly, it catalyzed the riboflavin kinase reaction with GTP as a phosphoryl-donor (Figure 3.8 D). We further moved towards mutation on R115 site, which can perturb the α-3 helix present on the outer loop of *Mj*RibK. Since the insertion of proline can disrupt the secondary structure, we mutated R115 residue with proline to create R115P mutant. Next, we introduced an additional proline in the R115P mutant to create R115PP mutant. We found that R115P was able to utilize ATP to some extent (Figure 3.8 E). The conversion of riboflavin into FMN using ATP in R115PP mutant was more by 2.5 folds than that in R115PP (Figure 3.8 F). It indicated that when we introduce two proline residues at R115 position, the α-3 helix changes its conformation to an extended loop. It makes the active site larger to accommodate larger ATP molecule.
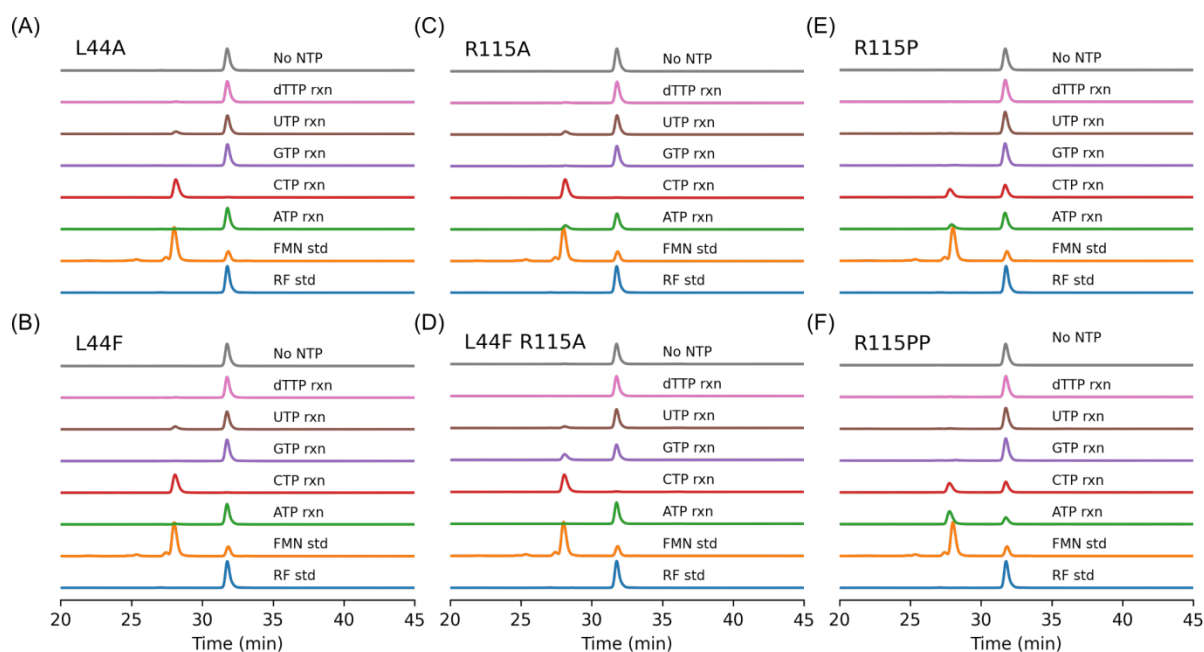
Figure 3.8: HPLC analysis of the activity of mutants with different NTPs (A-C)) L44A mutants have not shown any altered NTP preferences. (D) L44F R115A mutant has shown little preference for GTP (E-F) R115P and R115PP have shown activity with ATP. Activity with CTP has also reduced with R115P and R115PP. (Analysis was done with the same batch of fresh protein. Not done is duplicates)

3.3.5 Homology modelling of chimeric *Mj*RibK, and loop refinement

It is evident from the results that the α-helix on the outer loop in *Mj*RibK dictates the access of CTP to the active site. The reaction rate of the mutants designed for altered specificities were poor. Although the proline mutants show activity with ATP, the activity with CTP was not fully abolished. Thus, we next proposed to design a chimeric *Mj*RibK in which the outer loop was replaced with the corresponding loop of ATP-dependent *Hs*RFK. Table 3.2 shows the sequences of *Mj*RibK, *Hs*RFK, and chimeric *Mj*RibK. The residues shown in bold in *Mj*RibK and *Hs*RFK and chimeric *Mj*RibK sequences corresponds to the nucleobase surrounding loop. The secondary structure of chimeric *Mj*RibK was predicted using PSIPRED indicated the absence of α-helix on the swapped loop of chimeric *Mj*RibK (Figure 3.9).[28,29]

Table 3.2: Sequences of *Mj*RibK, *Hs*RFK and the chimeric *Mj*RibK protein

>*Mj*RibK
MVKLMIIEGEVVSGLGEGRYFLSLPPYKEIFKKILGFEPYEGTLNLKLDREFDINKFKYIETEDFEFNGKRFFG VKVLPIKILIGNKKIDGAIVVPKKTYHSSEIIEII**APMKLREQFNLKDGD**VIKILIKGDKDE

>*Hs*RFK

IMRHLPYFCRGQVVRGFGRGSKQLGIPTANFPEQVVDNLPADISTGIYYGWASVGSGDVHKMVVSIGWNP
YYKNTKKSMETHI**MHTFKEDFYGE**ILNVAIVGYLRPEKNFDSLESLISAIQGDIEEAKKRLELPEYLKIKED
NFFQVSK

>Chimeric *Mj*RibK

MVKLMIIEGEVVSGLGEGRYFLSLPPYKEIFKKILGFEPYEGTLNLKLDREFDINKFKYIETEDFEFNGKRFFG
VKVLPIKILIGNKKIDGAIVVPKKTYHSSEIIEII**MHTFKEDFYGE**VIKILIKGDKDE
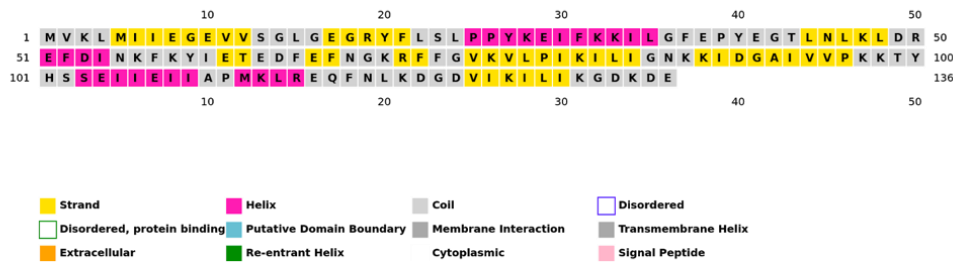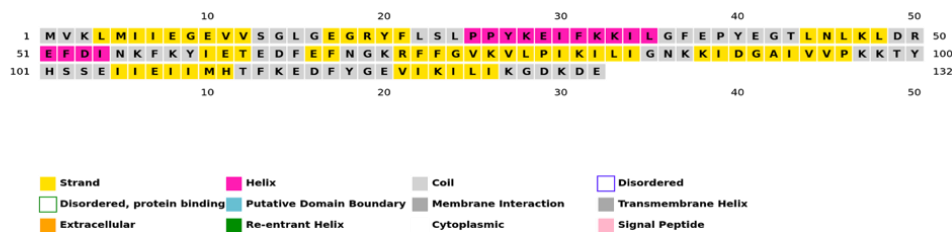
Figure 3.9: Secondary structure prediction of *Mj*RibK and chimeric *Mj*RibK (bottom) using PSIPRED (A) The α-3 helix (111-MKLR-115) on the loop is shown in pink color (B) The chimeric protein does not contain helix in the modified loop (110-MHTFKEDFYGE-120)

(A)   (B)
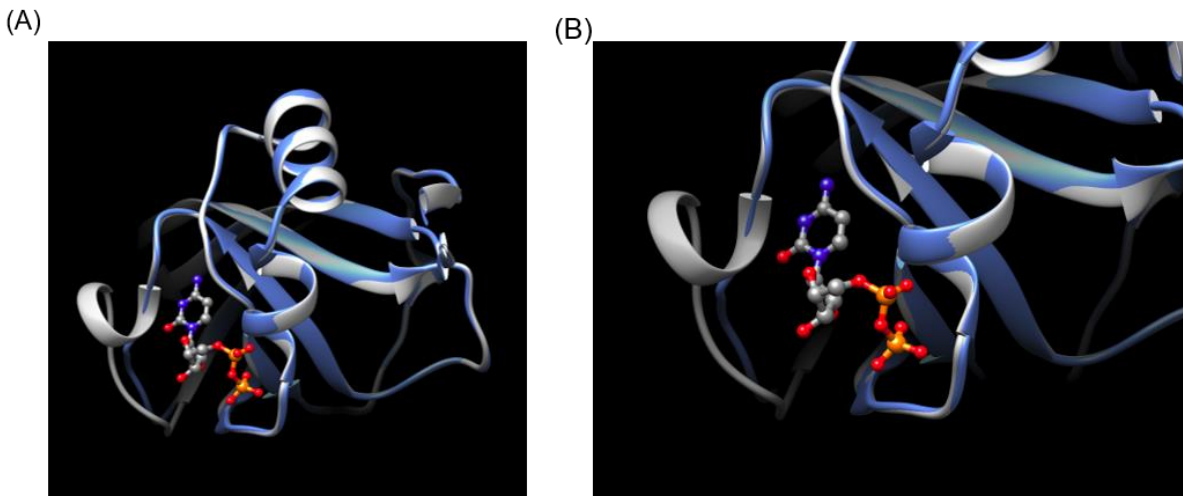
Figure 3.10: Homology model of chimeric-*Mj*RibK (A) overlap of chimeric *Mj*RibK (blue) with wild-type *Mj*RFK (silver) (B) CDP-bound active site of chimeric-*Mj*RibK overlapped with *Mj*RibK active site.

Next, we did the homology modeling of chimeric *Mj*RibK using Phyre2 server.[30] Phyre2 server has not shown any helix on the outer loop of chimeric *Mj*RibK (Figure 3.10 A, B). However, to avoid the possibility of the helix on the outer loop of chimeric *Mj*RibK, we did the loop modeling using Modeller to refine its structure. Ten models have been generated. The quality of the model is obtained as molpdf score. The model with the lowest molpdf score was selected as the best model. The molpdf scores of the models are given in the following graph.

(A)   (B)
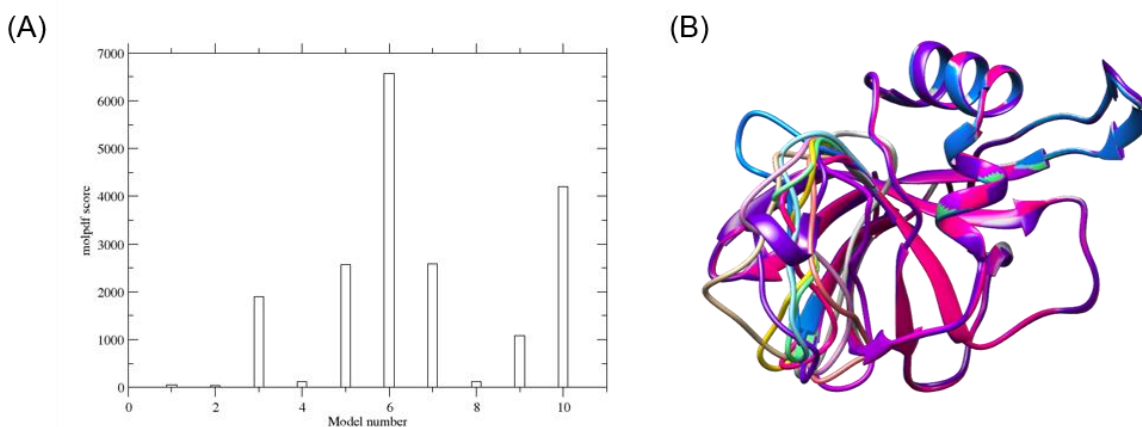
Figure 3.11: Loop-refining of chimeric *Mj*RibK (A) molpdf score of 10 best models. (B) the refined loops of the 10 best models overlapped with *Mj*RibK (purple)

The second model with the lowest molpdf score 30.81612 was selected as best model from the molpdf graph (Figure 3.11 A). This model was generated for the MD simulation studies and done parallel to the cloning, expression and purification of chimeric *Mj*RibK. However, we found that chimeric *Mj*RibK was an inactive protein.

3.3.6 The chimeric protein purification resulted in an inactive protein

Along with the homology modeling of chimeric *Mj*RibK, we cloned and purified it. However, it was an inactive protein and did not show activity with either of the NTPs (Figure 3.11).
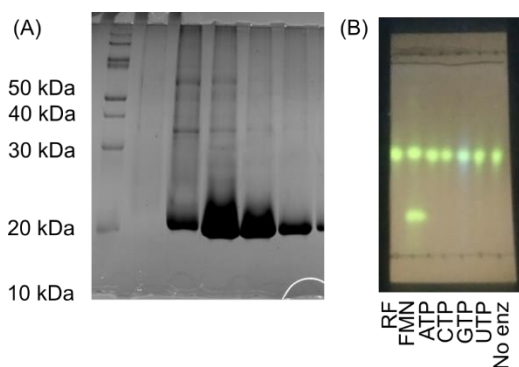
Figure 3.12: (A) SDS-PAGE showing the bands of chimeric *Mj*RibK (B) TLC showing the chimeric *Mj*RibK reactions with different NTPs. (Experiments were done in duplicates with different batch of proteins)

3.4 Conclusion:

We established the molecular factors responsible for the recognition of cytosine nucleobase in *Mj*RibK. Our comparative sequence and structural analyses revealed the key difference in the nucleobase recognition among CTP-dependent riboflavin kinases and ATP-dependent riboflavin kinases. Some of our mutants have shown activity with ATP. However, the activity of mutants reduced drastically in comparison to wild-type *Mj*RibK. *Mj*RibK activity is independent of the Mg^{2+} ion, but the mutants show activity only in the presence of Mg^{2+}.

Based on the mutational studies done on *Mj*RibK, it is evident that the α-3 helix present on the nucleobase-surrounding loop is critical for CTP binding. The mutation of the conserved R-115 residue with proline disrupted the small α-3 helix, and it resulted in the utilization of ATP.

The proline mutants have also shown reduced specificity for CTP. In *Mj*RibK, YxGTLN motif has conserved Y40, which makes stacking interaction with the cytosine nucleobase, G42 which makes hydrophobic interaction with cytosine nucleobase, and T43 that interacts with α-phosphate of CDP.[5]

Most of the riboflavin kinases from bacteria and eukaryotic sources show promiscuity in NTP utilization. It might be because the prominent binding of the NTP in the active site involves the binding of triphosphate moiety. In *Hs*RFK, the PTAN motif is conserved wherein, the Pro33 makes hydrophobic interaction with the adenine nucleobase, the T34 interacts with α-phosphate of ADP. Additionally, a conserved F97 residue is involved in stacking interaction with adenine.[2]

Moreover, CTP-dependent archaeal riboflavin kinases are an excellent system to study the molecular factors responsible for nucleobase recognition in kinases. Apart from archaeal riboflavin kinases, dolichol kinase is the only example of CTP-dependent kinase.[31]

Our study done on *Mj*RibK can be extended to other riboflavin kinases or FAD synthetases to interrogate the molecular determinants of nucleotide specificity. Since the mutants of *Mj*RibK showed reduced activity compared to wild-type, some more mutational sites can be predicted to restore the activities of mutants having altered specificity.

Another issue regarding the mutational study is the stability of mutants. Some mutants were not stable and hence not could not be tested for activity. For example, the chimeric *Mj*RibK in which the nucleobase-surrounding loop was replaced with the corresponding loop of *Hs*RFK, was not active. We suspect that this is because the swap was between a thermophilic enzyme (*Mj*RibK) and a mesophilic enzyme (*Hs*RFK). Creating loop-swap chimera using homologs with overlapping range of thermal stability might help resolve this problem.

In summary, archaeal riboflavin kinases are different from non-archaeal riboflavin kinases in terms of NTP utilization. We investigated the role of small-alpha helix on the loop surrounding the cytosine base in directing the access to CTP and cytosine specificity. The mutations to disrupt the helix resulted in ATP utilizing enzyme. Even though, these mutations could decrease CTP-specificity, enzyme could not be altered to be strictly specific for ATP. Computational predictions to alter nucleotide specificity resulted in a promiscuous enzyme, but does not exclude CTP. To decipher the nucleobase specificity, some more mutational studies and designs of chimeric proteins are expected. Further, the role of loops in determining the

nucleobase specificity could be explored through other archaeal and non-archaeal flavin-biosynthesis enzymes. The study can be further extended to other nucleotide utilizing enzymes.

References:

(1)   Abbas, C. A.; Sibirny, A. A. Genetic Control of Biosynthesis and Transport of Riboflavin and Flavin Nucleotides and Construction of Robust Biotechnological Producers. *Microbiol. Mol. Biol. Rev.* **2011**, *75* (2), 321–360. https://doi.org/10.1128/MMBR.00030-10.

(2)   Karthikeyan, S.; Zhou, Q.; Mseeh, F.; Grishin, N. V.; Osterman, A. L.; Zhang, H. Crystal Structure of Human Riboflavin Kinase Reveals a β Barrel Fold and a Novel Active Site Arch. *Structure* **2003**, *11* (3), 265–273. https://doi.org/10.1016/S0969-2126(03)00024-8.

(3)   Rajeswari, S. R.; Jonnalagadda, V. S.; Jonnalagadda, S. Purification and Characterization of Flavokinase from Neurospora Crassa. *Indian J. Biochem. Biophys.* **1999**, *36* (3), 137–142.

(4)   Solovieva, I. M.; Tarasov, K. V.; Perumov, D. A. Main Physicochemical Features of Monofunctional Flavokinase from Bacillus Subtilis. *Biochem.* **2003**, *68* (2), 177–181. https://doi.org/10.1023/A:1022645327972.

(5)   Ammelburg, M.; Hartmann, M. D.; Djuranovic, S.; Alva, V.; Koretke, K. K.; Martin, J.; Sauer, G.; Truffault, V.; Zeth, K.; Lupas, A. N.; Coles, M. A CTP-Dependent Archaeal Riboflavin Kinase Forms a Bridge in the Evolution of Cradle-Loop Barrels. *Structure* **2007**, *15* (12), 1577–1590. https://doi.org/10.1016/j.str.2007.09.027.

(6)   Mashhadi, Z.; Zhang, H.; Xu, H.; White, R. H. Identification and Characterization of an Archaeon-Specific Riboflavin Kinase. *J. Bacteriol.* **2008**, *190* (7), 2615–2618. https://doi.org/10.1128/JB.01900-07.

(7)   Rodionova, I. A.; Vetting, M. W.; Li, X.; Almo, S. C.; Osterman, A. L.; Rodionov, D. A. A Novel Bifunctional Transcriptional Regulator of Riboflavin Metabolism in Archaea. *Nucleic Acids Res.* **2017**, *45* (7), 3785–3799. https://doi.org/10.1093/nar/gkw1331.

(8)   Liu, Y.; Shah, K.; Yang, F.; Witucki, L.; Shokat, K. M. A Molecular Gate Which Controls Unnatural ATP Analogue Recognition by the Tyrosine Kinase V-Src. *Bioorg. Med. Chem.* **1998**, *6* (8), 1219–1226. https://doi.org/10.1016/s0968-0896(98)00099-6.

(9)   Liu, Y.; Shah, K.; Yang, F.; Witucki, L.; Shokat, K. M. Engineering Src Family Protein

Kinases with Unnatural Nucleotide Specificity. *Chem. Biol.* **1998**, *5* (2), 91–101. https://doi.org/10.1016/s1074-5521(98)90143-0.

(10) Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M. Engineering Unnatural Nucleotide Specificity for Rous Sarcoma Virus Tyrosine Kinase to Uniquely Label Its Direct Substrates. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (8), 3565–3570. https://doi.org/10.1073/pnas.94.8.3565.

(11) Vashisht, K.; Verma, S.; Gupta, S.; Lynn, A. M.; Dixit, R.; Mishra, N.; Valecha, N.; Hamblin, K. A.; Maytum, R.; Pandey, K. C.; Van Der Giezen, M. Engineering Nucleotide Specificity of Succinyl-CoA Synthetase in Blastocystis: The Emerging Role of Gatekeeper Residues. *Biochemistry* **2017**, *56* (3), 534–542. https://doi.org/10.1021/acs.biochem.6b00098.

(12) van den Ent, F.; Löwe, J. RF Cloning: A Restriction-Free Method for Inserting Target Genes into Plasmids. *J. Biochem. Biophys. Methods* **2006**, *67* (1), 67–74. https://doi.org/10.1016/j.jbbm.2005.12.008.

(13) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254. https://doi.org/10.1006/abio.1976.9999.

(14) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. NCBI BLAST: A Better Web Interface. *Nucleic Acids Res.* **2008**, *36* (Web Server issue), W5–W9. https://doi.org/10.1093/nar/gkn201.

(15) Boratyn, G. M.; Camacho, C.; Cooper, P. S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T. L.; Matten, W. T.; McGinnis, S. D.; Merezhuk, Y.; Raytselis, Y.; Sayers, E. W.; Tao, T.; Ye, J.; Zaretskaya, I. BLAST: A More Efficient Report with Usability Improvements. *Nucleic Acids Res.* **2013**, *41* (Web Server issue), W29–W33. https://doi.org/10.1093/nar/gkt282.

(16) NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44* (D1), D7–D19. https://doi.org/10.1093/nar/gkv1290.

(17) Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32* (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.

(18)   Edgar, R. C. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinformatics* **2004**, *5*, 113. https://doi.org/10.1186/1471-2105-5-113.

(19)   Sali, A. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *Mol. Med. Today* **1995**, *1* (6), 270–277. https://doi.org/10.1016/s1357-4310(95)91170-7.

(20)   Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26* (2), 283–291. https://doi.org/10.1107/S0021889892009944.

(21)   Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**, *2* (9), 1511–1519. https://doi.org/10.1002/pro.5560020916.

(22)   Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated Docking of Flexible Ligands: Applications of AutoDock. *J. Mol. Recognit. 9* (1), 1–5. https://doi.org/10.1002/(sici)1099-1352(199601)9:1<1::aid-jmr241>3.0.co;2-6.

(23)   Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.

(24)   Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725. https://doi.org/10.1002/prot.21123.

(25)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D. Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.;

Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A. Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02; Gaussian Inc; Wallingford CT*; 2004.

(26)   M.J. Abraham, D. van der Spoel, E. Lindahl, B. H. and the G. development team. GROMACS User Manual Version 2019. 2019.

(27)   Bauer, S.; Kemter, K.; Bacher, A.; Huber, R.; Fischer, M.; Steinbacher, S. Crystal Structure of Schizosaccharomyces Pombe Riboflavin Kinase Reveals a Novel ATP and Riboflavin-Binding Fold. *J. Mol. Biol.* **2003**, *326* (5), 1463–1473. https://doi.org/10.1016/S0022-2836(03)00059-7.

(28)   Jones, D. T. Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. *J. Mol. Biol.* **1999**, *292* (2), 195–202. https://doi.org/10.1006/jmbi.1999.3091.

(29)   Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* **2019**, *47* (W1), W402–W407. https://doi.org/10.1093/nar/gkz297.

(30)   Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10* (6), 845–858. https://doi.org/10.1038/nprot.2015.053.

(31)   Shridas, P.; Waechter, C. J. Human Dolichol Kinase, a Polytopic Endoplasmic Reticulum Membrane Protein with a Cytoplasmically Oriented CTP-Binding Site. *J. Biol. Chem.* **2006**, *281* (42), 31696–31704. https://doi.org/10.1016/S0021-9258(19)84083-8.

Chapter 4

Characterization of CTP-dependent riboflavin kinase from mesophilic archaeon *Methaonococcus maripaludis* and identifying the molecular factors that confer stability to thermophilic homolog from *Methanocaldococcus jannaschii*

4.1 Introduction

The mechanism of action of enzymes found in extremophilic organisms is less understood as they have to operate at extreme conditions such as high temperature, pressure, alkalinity, acidity, and saturating concentration of salts. Enzymes that can catalyze reactions at high temperatures are known as thermophilic and hyperthermophilic enzymes. They operate in the temperature range of 75 °C-113 °C.[1] Mesophilic enzymes work at a moderate temperature range (25 °C-60 °C). Finally, enzymes that are cold-adapted are called psychrophilic enzymes. (-20 °C to 15 °C).[2]

Investigations have been done in the past to understand the molecular basis of the thermostability of an enzyme. Still, there is no general rule for the thermal stability of an enzyme. Through a series of experiments done on several thermophilic-mesophilic pairs of enzymes, it is well known that structural rigidity is a shared future in the thermophilic enzyme, which makes them stable at high temperatures. Experimental and computational work on lipase mutants of wild-type *Bacillus subtilis* revealed that a mesophilic enzyme can be optimized for operating in the thermophilic temperature range by probing the conformational flexibility.[3,4] Constrained network analysis (CNA) suggested that a protein consists of a network of many types of interactions among atoms in it.[5,6] Using unfolding simulation through CNA, it is possible to decompose a protein into rigid and flexible parts. Mutation in the flexible region imparts structural rigidity, which in turn makes an enzyme thermostable.[7–9] Insertion of proline residues at specific positions of an enzyme can also rigidify the structure to make it more thermostable. A study on *Bacillus cereus* oligo-1,6-glucosidase shows that the mutation of multiple residues to proline can make the protein thermostable.[11] Another survey done on 102 thermophilic-mesophilic pairs of enzymes shows that a hydrophobic environment is another key factor for the thermostability of enzymes.[10] The role of salt bridges for thermostability is also

well established in the literature. Typically, thermophilic enzymes have a higher number of salt bridges than their mesophilic homolog, which makes them thermostable. A study has used this principle to interchange the thermostability of acylphosphatase from the hyperthermophile *Pyrococcus Horikoshii* with its mesophilic homologfrom *Homo sapiens* by interchanging the salt bridges.[12]

This chapter conducted the characterization of the first mesophilic CTP-dependent riboflavin kinase from *Methanococcus maripaludis* (*Mmp*RibK) and compared it with its thermophilic homolog *Methanocaldococcus jannaschii* riboflavin kinase (*Mj*RibK). To date, CTP-dependent riboflavin kinase is characterized only from thermophilic/hyperthermophilic organisms such as *M. jannaschii* and *Thermoplasma acidophilum*.[13–15] We established that *Mmp*RibK converts riboflavin into FMN using CTP as a phosphate donor in the mesophilic temperature range. Further, we showed that the thermostability of *Mmp*RibK could be modulated by incorporating salt bridges and proline residues. With this, we determined the key factors responsible for the thermostability of *Mj*RibK. Our study resulted in the design of some thermophilic homologs of *Mmp*RibK, without loss of activity and nucleotide specificity.

4.2 Materials and Methods

4.2.1 Materials:

Genomic DNA of *Methanocaldococcus jannaschii* DSM 2661 and *Methanococcus maripaludis* S2 was a gift from Biswarup Mukhopadhyay at Virginia State University. The primary enzyme assay reagents riboflavin and FMN are from Sigma-Aldrich, and the nucleotide triphosphates were obtained from Jena Biosciences. All reagents and media components used in cloning, protein expression, and purification were obtained from Sigma-Aldrich, TCI chemicals, Rankem, SRL chemicals, and Himedia, unless otherwise stated. Polymerase chain reaction (PCR) reagents and plasmid purification kits were from Agilent and Himedia. For the cloning, we used high fidelity PrimeSTAR GXL DNA polymerase from DSS Takara Bio USA. The SYPRO-Orange dye used in the thermal shift assay was obtained from Thermofisher Scientific.

4.2.2 Bioinformatic and phylogenetic analysis to search the suitable mesophilic homolog of thermophilic *Mj*RibK

      To identify a suitable mesophilic CTP-dependent riboflavin kinase homolog for our studies, the protein sequence of *Mj*RibK (Swiss-Prot accession number Q60365) was used to perform a <u>B</u>asic <u>L</u>ocal <u>A</u>lignment <u>S</u>earch <u>T</u>ool (BLAST) analysis.[16–18] It yielded 1813 distinct homolog sequences of putative CTP-dependent riboflavin kinase entries. A phylogeny study was then conducted for a curated set of 56 protein sequences of archaeal thermophilic and mesophilic RibK enzymes from this list of BLAST search candidates. To do so, we created a multiple sequence alignment using the **MU**ltiple **S**equence **C**omparison by **L**og-**E**xpectation (MUSCLE) alignment tool, which revealed that some of the riboflavin kinase proteins contained a transcriptional regulator domain.[19,20] The transcriptional regulator domains were trimmed using BioEdit sequence alignment editor, so as to only consider the riboflavin kinase domain sequence for analysis, and the alignment was saved as a Phylip4.0 file format. The alignment was then subjected to maximum likelihood/ rapid bootstrapping analysis on RaXML-HPC2 on XSEDE on the CIPRES gateway to produce a phylogenetic tree, which was visualized using Figtree. Two additional multiple sequence alignments of only archaeal riboflavin kinases and archaeal and non-archaeal riboflavin kinases were also conducted using the method described above.

4.2.3 Molecular cloning of *Mmp*RibK gene and other mutants:

      The cloning of *Mmp*RibK and its mutants was done using the restriction-free method as described in the method section of chapter 3.[21] The primers used in this study are given in Table 4.1.

Table 4.1: List of the primers used in the cloning

#	Primer name	Sequence
1	*Mj*RibK_F1	5'-TTGGTGAAATTGATGATTATTG-3'
2	*Mj*RibK_R1	5'-TTATTCATCTTTATCTCCC-3'
3	*Mj*RibK_F2	5'-GCCGCGCGGCAGCCATATG GTGAAATTGATGATTATTG-3'
4	*Mj*RibK_R2	5'-CGACGGAGCTCGAATTCGGATCCTTATTCATCTTTATCTCCC-3'
5	*Mmp*RibK_F1	5'-ATGGAAATTTTTGGGCATG-3'
6	*Mmp*RibK_R1	5'-TTAAATTACGAGTTTTAC-3'
7	*Mmp*RibK_F2	5'-GCCGCGCGGCAGCCATATGGAAATTTTTGGGCATG-3'
8	*Mmp*RibK_R2	5'-CGACGGAGCTCGAATTCGGATCCTTAAATTACGAGTTTTAC-3'

9	*Mmp*RibK_T34E_F1	5'-GAATTAACCGGATTTGAACCGTTTGAAGGAACA-3'
10	*Mmp*RibK_S2_DRE_F1	5'-AGGAACATTGAACGTAAAATTAGATAGAGAATTCAATTTAGATGAATTTAATCC-3'
11	*Mmp*RibK_N119D_R1	5'-TTTTACAACATCTGAATCCTTTAATAATAAAAA-3'
12	*Mmp*RibK_TH(21-22)PP_F1	5'-AAGTTTTTTGTTGGATTACCTCCTTACAAAAATAAATTTGAA-3'
13	T7 promoter_F	5'-TAATACGACTCACTATAGGG-3'
14	T7 terminator_R	5'-GCTAGTTATTGCTCAGCGG-3'

4.2.4 Protein expression and purification:

The cloned pET28a plasmid containing *Mmp*RibK gene was transformed into *Escherichia coli* BL21(DE3) cells. The transformed cells were grown in Luria-Bertani broth at 37 °C till OD_{600} reached 0.6. Induction of cells was done using 1 mM IPTG at 18 °C for 15 h. Purification of the enzyme having N-terminal hexahistidine tag was done by Ni-NTA affinity chromatography using Akta pure FPLC system (GE Healthcare). Equilibration buffer was 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, and 0.025 % β-mercaptoethanol. Protein was eluted with 250 mM imidazole buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM imidazole, and 0.025 % β-mercaptoethanol)

The purity and molecular weight of the enzymes were analyzed by 15 % SDS-PAGE gel electrophoresis. The imidazole present in the column was removed by desalting (size-exclusion chromatography). The fractions showing the corresponding molecular weight were pooled together and desalted using Hitrap 5 mL desalting column (GE healthcare Bio-sciences). The Hitrap desalting column was equilibrated using the buffer containing 50 mM Tris-HCl pH 8.0, 300 mM NaCl, and 0.025% β-mercaptoethanol. 1.5 mL of protein samples were loaded each time to the column and desalted to get imidazole-free protein.

The concentration of protein was checked using Bradford assay using BSA as standard.[22,23] Wild-type *Mj*RibK and mutants were also expressed and purified similarly. The wild-type *Mmp*RibK enzyme was also subjected to Matrix-Assisted Laser Desorption/ Ionization – Time of Flight (MALDI-TOF) analysis as per reported protocols, to verify its mass.

4.2.5 Enzyme assay and chromatographic analysis

We tested the enzymatic activity of wild-type *Mmp*RibK using 300 µM riboflavin, 4 mM CTP, 10 mM $MgCl_2$, 100 mM Tris-HCl pH 8.0, and 1 µM enzyme, modifying on previously

reported conditions for the assay of MjRibK.[13] The reaction mixture was incubated at 37 °C overnight and visualized using fluorescence-based thin layer chromatography (TLC) (Merck Silica gel 60 F$_{254}$) using the long-wavelength (365 nm) of a handheld UV lamp (Spectroline E-series) with the solvent system water: acetic acid: n-butanol in the ratio 5:1:4.[24] The retention time of riboflavin and FMN in this solvent system was 0.45 and 0.2, respectively. The enzymatic reaction product was further analyzed by high-performance liquid chromatography (Agilent HPLC system (1260 Infinity II)) equipped with UV-Vis DAD detector on a C-18 reverse-phase column (Phenomenex, Gemini 5 µm NX-C18 110 Å, 250 × 4.6 mm). The HPLC method used for analysis was derived from the one previously used to assay MjRibK.[13] The mobile phases used were 10 mM ammonium acetate pH 6.5 (Solvent-A) and methanol (Solvent-B). Gradient elution of analytes were done at 0.5 mLmin^{-1} as follows. 0-5 min 95 % A, 5-40 min from 95 % A to 20 % A, 40-60 min from 20 % A to 95 % A. The analytes were detected at 260 nm, 280 nm, 365 nm, and 440 nm. The mutants of MmpRibK were assayed and analyzed in a manner identical to the wild-type enzyme unless otherwise mentioned.

4.2.6 Temperature dependence of MmpRibK

The activity of mesophilic enzyme MmpRibK was tested at different temperatures to find the optimal temperature for its activity. A mixture of 300 µM riboflavin, 4 mM CTP, 100 mM Tris-HCl pH 8.0, and 10 mM MgCl$_2$ were incubated at a particular temperature for 20 min. 1 µM of the enzyme was added to the pre-equilibrated mixture at this temperature, and after 15 min, it was quenched with 60 mMAcetic acid (final concentration). The temperatures used for the analysis were from 15 °C to 70 °C. The conversion of riboflavin into FMN at each temperature was analyzed following the same HPLC protocol. Similarly, MjRibK was tested in the temperature range of 25-95 °C. The temperature dependence of mutants was tested only at temperatures 45 °C, 60 °C, and 70 °C.

4.2.7 Test for alternate substrate

The enzyme activity of MmpRibK and its mutants were tested with alternate co-substrate nucleotides ATP, GTP, and UTP, both in the absence and presence of NaCl, and analyzed by TLC and HPLC as described in the method section.

4.2.8 Metal ion dependence of *Mmp*RibK

The enzyme assay, as previously described, was performed in the presence of 10 mM of $MgCl_2.6H_2O$, $MnCl_2.4H_2O$, $NiCl_2.6H_2O$, $CaCl_2.H_2O$, $CoCl_2.6H_2O$, $ZnCl_2$, $CuCl_2.2H_2O$, $FeSO_4.7H_2O$ (freshly prepared), and the reaction was analyzed using TLC and HPLC.

4.2.9 pH dependence of *Mmp*RibK

The optimal pH for the enzyme activity of *Mmp*RibK with riboflavin, CTP, and Mg^{2+} was determined by conducting the assay in the pH range of 4.3-11 and analyzing it as described previously. The following buffers were used to achieve the different pH conditions - sodium acetate buffer for pH 4.3 and 5.0, MES buffer for pH 6.0, HEPES for pH 7.0, and Tris-HCl buffer for pH 8.0, 9.0, 10.0, and 11.0.

4.2.10 Effect of NaCl concentration on *Mmp*RibK activity

The enzyme activity was assayed in varying salt conditions, ranging from 50 mM to saturating concentration of NaCl (solid NaCl is weighed and added directly to the reaction mixture for higher saturating concentrations of NaCl). 100 µM of riboflavin and 4 mM CTP is used in this experiment, and the reaction was analyzed by TLC and HPLC as described in the method section.

4.2.11 Chemical denaturation of *Mmp*RibK and *Mj*RibK:

The activities of *Mj*RibK and *Mmp*RibK were measured in the presence of chemical denaturant guanidinium hydrochloride (Gdn-HCl), to check the difference in the stability of these thermophilic-mesophilic pair of the enzyme. Variable concentrations of Gdn-HCl (0 mM – 2.5 M) were added to the reaction mixture containing 100 µM of riboflavin, 100 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1 µM of the enzyme. These are kept at room temperature for one hour. After one hour, samples of *Mmp*RibK were incubated at 37 °C, and those of *Mj*RibK was set at 70 °C for 20 min. The reactions were initiated by adding 2 mM CTP and quenched with 60 mM acetic acid (final concentration) after 15 min. The reaction product FMN was analyzed on TLC.

4.2.12 Determination of kinetic parameters for *Mmp*RibK for the substrates riboflavin and CTP:

The kinetic parameters for riboflavin - K_m, k_{cat}, V_{max} - were determined using 4 mM CTP. The concentration of riboflavin was varied from 25 µM to 300 µM. Similarly, to determine the kinetic parameters for CTP, 300 µM of riboflavin was used, and CTP concentration has been varied from 12.5 µM to 2 mM. All kinetic experiments have been done in the presence of 2M NaCl. The concentration of the product FMN synthesized was determined at various time points from the area under the curve obtained from the HPLC chromatogram using a previously generated standard curve of FMN. The concentration of FMN formed at different time intervals were plotted using Origin 9 software. The concentration-time plot was fitted and initial velocities were determined from the slope of initial part of the curve. The Michaelis-Menten and Lineweaver-Burk data were plotted with Origin 9 software (OriginLab corporation), using a nonlinear fitting with the Hills algorithm, $v = V_{max} \dfrac{x^n}{k^n + x^n}$ .Where n is the number of cooperative sites. For the single-substrate model, n is fixed to 1.

4.2.13 Structural analysis and ligand docking:

The PDB structure of the wild-type and mutant *Mmp*RibK enzymes were modeled using Modeller.[25] The stereochemical quality of the modeled structures was estimated using programs PROCHECK and ERRAT.[26,27] All structural analysis were done using molecular visualization softwares-'Visual Molecular Dynamics' (VMD), and UCSF Chimera.[28,29] The residues involved in the formation of salt bridges were identified using the salt bridges analysis tool of VMD. Docking of the ligands riboflavin and CTP were done using Autodock4.0.[30,31]

4.2.14 Molecular Dynamics simulations:

We docked CTP and riboflavin in the *Mmp*RibK structure and then performed a molecular dynamics (MD) simulation. We used AMBER99SB force field for the protein.[32] To create the amber force field of CTP and riboflavin we used antechamber module of AmberTool. First, we conducted a quantum mechanical optimization of the CTP using HF/6-31G* basis set in GAUSSIAN03 software following which restricted electrostatic potential (RESP) charges on the atoms of CTP and riboflavin were calculated.[33] The topology and coordinates generated from AmberTools were converted into the GROMACS format using a perl program amb2gmx.pl. We

put the CTP-riboflavin bound protein in a cubic box of length 70 Å. Then, the system was solvated with ~2000 TIP3P water molecules. Further, we added 150 mM $MgCl_2$ ion solution after which we optimized the system using steepest descent method, followed by heating the system to 27 °C using the Berendsen thermostat with coupling constant of 2 ps (NVT condition). After that, we have performed 4 ns NPT simulation to adjust the box size. Finally, we performed production run for 300 ns at 27 °C and at 72 °C using Nose-Hoover thermostat. All the analysis is done over production run. To study the effect of $MgCl_2$ we performed two set of simulations of *Mmp*RibK-CTP-riboflavin complex in (i) absence and (ii) presence of Mg^{2+}. All the simulations were performed using GROMACS simulation package.[34]

4.2.15 Thermal shift assay:

The thermostability of mutants has been tested using thermal shift assay.[35] Sypro-orange dye (5000x commercial stock) was added at final concentration of 2.5x to 5 µM of the wild-type *Mmp*RibK and its mutants (total volume of 50 µL in individual wells of a 96-well plate). The change in the fluorescence intensity during gradual unfolding owing to a rise in the temperature is measured using CFX Real-Time PCR instrument (Bio-Rad). The negative derivative of the change in fluorescence intensity with temperature was plotted against temperatures, of which the lowest value corresponds to the T_m of the enzyme.

4.3 Results

The overexpression of *Mmp*RibK into *Escherichia coli* cells followed by Ni-NTA affinity purification gave the soluble protein of 16 kDa as indicated by 15% SDS-PAGE gel electrophoresis (Figure 4.1 A). The mass predicted from its amino acid sequence using ProtParam is 14.3448 kDa. The molecular weight of the enzyme verified by MALDI-TOF corresponds to 16.37126 kDa with hexahistidine tag (Figure 4.1 B).
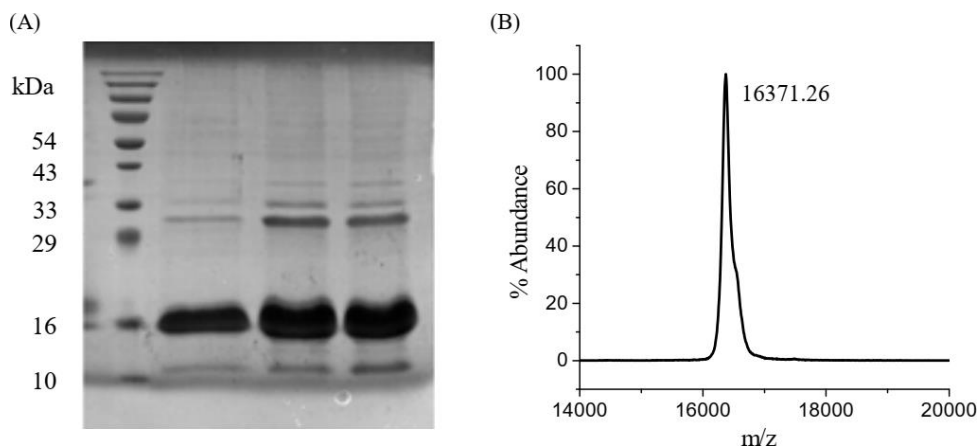
Figure 4.1: molecular weight of *Mmp*RibK is characterized by SDS-PAGE gel electrophoresis and MALDI-TOF. (A) 15 % SDS-PAGE gel indicating the band of purified *Mmp*RibK at 16 kDa. (B) MALDI has shown the molecular weight of 16371.26 Da, which corresponds to the molecular weight of this enzyme with an additional hexahistidine tag.

4.3.1 Bioinformatic analysis reveals that *Mmp*RibK is one of the closest mesophilic homologs of the thermophilic *Mj*RibK

To identify a suitable mesophilic homolog for our study, we conducted a protein BLAST analysis with the sequence of the archaeal riboflavin kinase enzyme *Mj*RibK. The results revealed 1813 archaeal riboflavin kinase enzymes (EC: 2.7.1.161). Among the top 25 conserved sequences, 44 % are mesophilic, of which 32 % are different strains of *M. maripludis* (Table 4.2).

Comparing sequences of archaeal riboflavin kinase with their bacterial and eukaryotic homologs shows that these are two completely distinct enzymes, differing even in the conserved motif responsible for triphosphate binding (GTLN in archaeal versus PTAN in non-archaeal riboflavin kinases, as also noted previously (Figure 4.2 A). However, among the archaeal riboflavin kinases, a significant degree of conservation is observed, especially within the catalytic domains, as evidenced by the sequence alignment of *MmpRibK* with other homologous archaeal enzymes (Figure 4.2 B). The binding motifs of *Mmp*RibK are similar to *Mj*RibK.[14] A detailed analysis of these binding domains is given in chapter 3.

NCBI conserved domain search shows that *Mmp*RibK belongs to the conserved protein domain family pfam01982/ PRK14132, currently annotated as CTP-dependent riboflavin kinase and involved in cofactor transport and metabolism. Interestingly, some homologs in this

superfamily have DNA-binding domains and they act as bifunctional transcriptional regulators/ riboflavin kinases.

Next, we curated a set of 56 archaeal riboflavin kinases with varying sequence identity from the results obtained from the BLAST search and constructed a phylogenetic tree (Figure 4.1 C). The phylogeny of archaeal riboflavin kinases showed that the thermophilic and the mesophilic riboflavin kinases separate into distinct clades, with a few exceptions in both groups. *M. maripaludis* S2 riboflavin kinase (*Mmp*RibK) was one such exception. The sequences of *Mmp*RibK and *Mj*RibK are 50.4 % identical despite *Mmp*RibK belonging to a mesophilic archaeon (Figure 4.2, Table 4.2). Based on these results, we selected *Mmp*RibK as our candidate mesophilic CTP-dependent riboflavin kinase for further analysis.
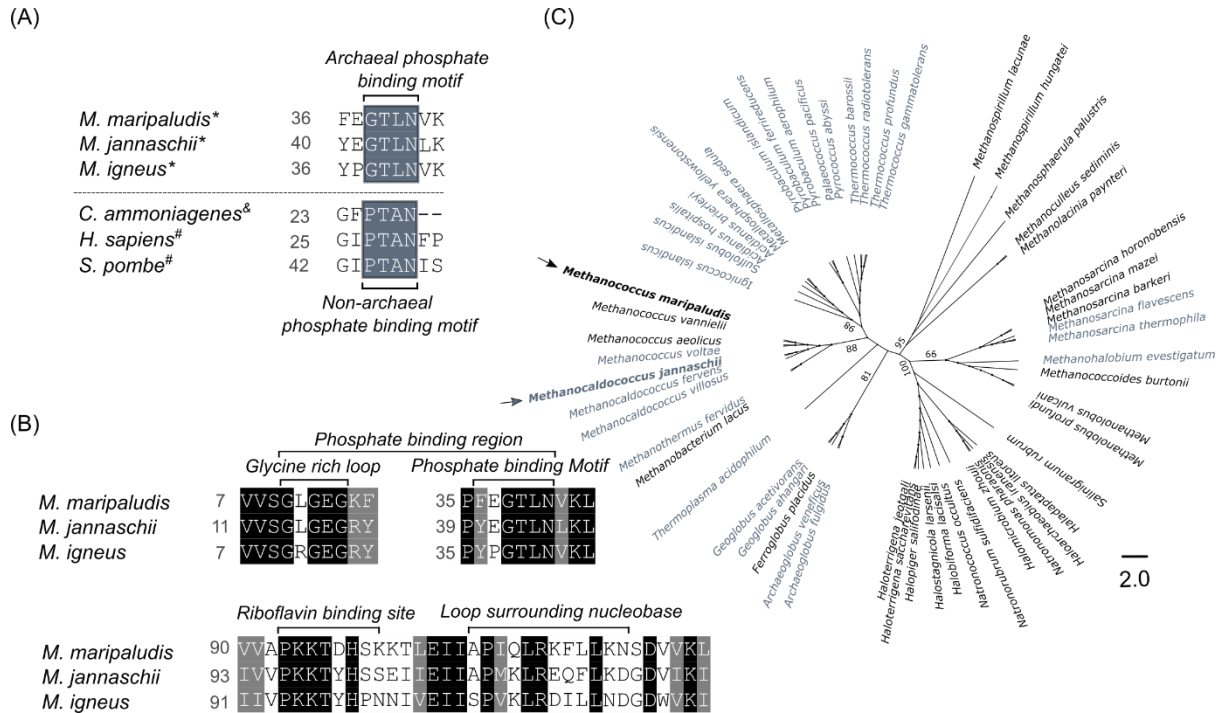
Figure 4.2: Multiple sequence alignment and phylogeny of riboflavin kinases. **(A)** Multiple sequence alignment of archaeal (indicated by *) and non-archaeal riboflavin kinases ([&]bacteria, [#]eukaryotes) showing the phosphate-binding consensus motif present near the N-terminus– GTLN in the case of archaea and PTAN in the case of bacteria and eukaryotes. **(B)** The phosphate-binding domain, riboflavin binding site, and the region surrounding the nucleobase show well-conserved patterns across archaeal sequences. **(C)** We constructed a phylogenetic tree of archaeal riboflavin kinase homologs. The grey entries are riboflavin kinase sequences from thermophilic/hyperthermophilic organisms, and the black entries are from mesophilic organisms. The previously characterized thermophilic *Methanocaldococcus*

jannaschii RibK (*Mj*RibK) and the *mesophilic Methanococcus maripaludis RibK (Mmp*RibK), which we characterized in this study are highlighted with arrows.

Table 4.2: Top 25 hits from the BLAST analysis with the *Mj*RibK protein sequence

#	Organism	Accession number	% Identity	E-Value	Total Score	Type
1	*Methanocaldococcus jannaschii*	2P3M_A	100	2.52E-88	262	Thermophilic
2	*Methanocaldococcus jannaschii DSM 2661*	2OYN_A	99.265	3.30E-88	262	Thermophilic
3	*Methanocaldococcus sp. FS406-22*	WP_012980051.1	89.394	3.26E-77	234	Thermophilic
4	*Methanocaldococcus bathoardescens*	WP_048201574.1	87.121	1.52E-74	227	Thermophilic
5	*Methanocaldococcus fervens*	WP_015791162.1	85.606	1.64E-74	227	Thermophilic
6	*Methanocaldococcus vulcanius*	WP_015732574.1	67.669	9.81E-54	175	Thermophilic
7	*Methanocaldococcus villosus*	WP_004591531.1	59.524	1.94E-47	159	Thermophilic
8	*Methanotorris igneus*	WP_013799911.1	59.69	1.43E-44	151	Thermophilic
9	*Methanotorris formicicus*	WP_007043845.1	59.69	2.84E-44	150	Thermophilic
10	*Methanocaldococcus infernus*	WP_083771775.1	59.836	1.10E-42	147	Thermophilic
11	*Methanothermococcus okinawensis*	WP_013866684.1	53.968	5.79E-42	145	Thermophilic
12	*Methanothermococcus thermolithotrophicus*	WP_018154369.1	56.349	2.20E-41	143	Thermophilic
13	*Methanococcus aeolicus*	WP_011973638.1	52.5	5.90E-38	134	Mesophilic
14	*Methanofervidicoccus abyssi*	WP_131007122.1	47.619	2.26E-37	133	Thermophilic
15	*Methanococcus maripaludis DSM 2067*	AVB76310.1	52.308	5.99E-36	129	Mesophilic
16	*Methanofervidicoccus sp. A16*	WP_148120329.1	46.457	4.24E-35	127	Mesophilic
17	*Methanococcus maripaludis C5*	ABO35789.1	52.273	7.15E-35	127	Mesophilic
18	*Methanococcus maripaludis C6*	WP_012193477.1	53.543	2.66E-34	125	Mesophilic
19	*Methanococcus maripaludis C7*	WP_011977559.1	51.181	3.03E-34	125	Mesophilic
20	*Methanococcus maripaludis OS7*	WP_119720556.1	51.163	2.72E-33	123	Mesophilic
21	*Methanococcus vannielii*	WP_012065999.1	52.381	5.99E-33	122	Mesophilic
22	*Methanococcus maripaludis KA1*	WP_146777918.1	50.388	6.31E-33	122	Mesophilic
23	*Methanococcus maripaludis X1*	WP_013998637.1	50.388	1.70E-32	120	Mesophilic
24	*Methanococcus maripaludis S2*	WP_011170128.1	50.388	5.98E-32	119	Mesophilic
25	*Methanococcus voltae*	WP_013180855.1	45.802	2.21E-29	113	Mesophilic

4.3.2 The *Mmp*RibK riboflavin kinase is functional under mesophilic conditions

To test whether *Mmp*RibK is a functional enzyme, we conducted an assay with conditions similar to what was previously reported for *Mj*RibK, except that the temperature was set at 37 °C.[13] No conversion of riboflavin into FMN was observed. However, when 10 mM MgCl$_2$ was added, we found that the *Mmp*RibK enzyme successfully catalyzes the conversion of riboflavin into FMN using CTP as phosphate donor, as indicated by fluorescence-TLC, and verified by HPLC analysis (Figure 4.3 A, B, C).
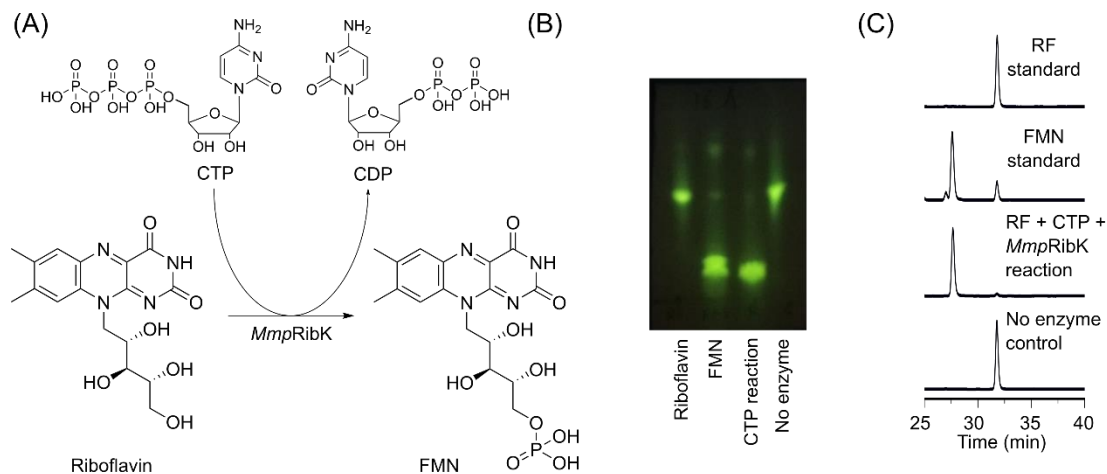
Figure 4.3: Reconstitution of the riboflavin kinase activity of *Mmp*RibK (A) Reaction catalyzed by *Mmp*RibK to convert riboflavin into FMN. (B) TLC shows the fluorescent spots for riboflavin standard, FMN standard, full reaction, and no enzyme control. (C) HPLC chromatogram for the formation of FMN by *Mmp*RibK. (Reconstitution was done in duplicates with different batches of proteins.)

4.3.3 Test for alternate nucleotide triphosphate

Next, we compared some basic features of the *Mmp*RibK and *Mj*RibK homologs. First, we carried out reactions with both homologs using different nucleotide triphosphates. We found that *Mmp*RibK shows very high specificity for utilizing CTP and does not form FMN with other nucleotides such as ATP, GTP, and UTP (Figure 4.4 A). On the other hand, *Mj*RibK shows the highest specificity with CTP but offers 10 % and 6 % conversion of riboflavin to FMN with UTP and ATP, respectively (Figure 4.4 B).[13] Temperature-wise, when held at 100 °C for 5 min and then used in an assay, *Mmp*RibK is completely denatured and showed no activity, while *Mj*RibK has no significant change in activity. Lastly, as compared to *Mj*RibK, which is unique in its ability to catalyze the phosphate transfer reaction in the absence of any divalent metal ion, *Mmp*RibK requires an Mg^{2+} ion for its catalytic activity (Figure 4.4 C). [13]
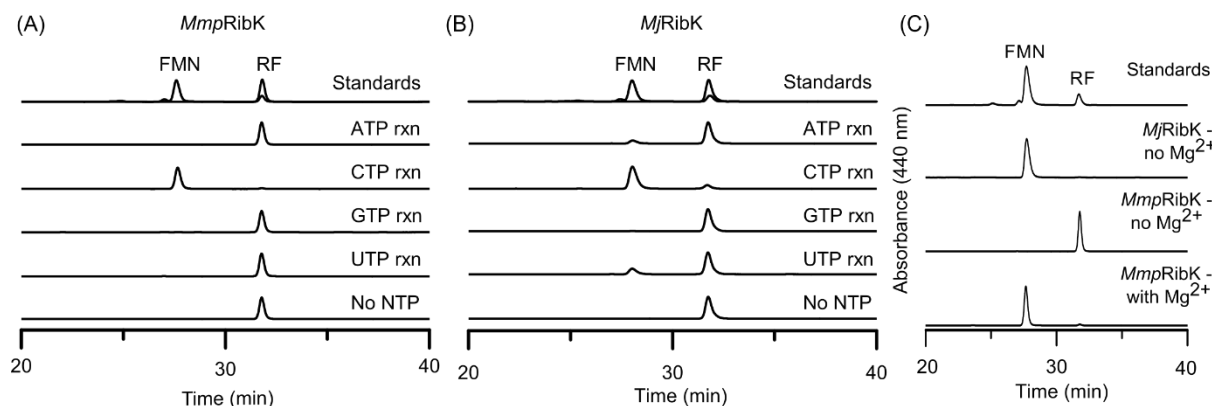
Figure 4.4: Testing for utilization of alternate nucleotide triphosphate by *Mmp*RibK and *Mj*RibK **(A)** *Mmp*RibK reactions with riboflavin (RF) and different NTPs shows that it is very specific for CTP for the formation of flavin mononucleotide (FMN). **(B)** *Mj*RibK reaction with different NTPs shows that CTP is the preferred substrate. It also gives some conversion with ATP and UTP, as also previously noted.[13] (C) *Mmp*RibK and *Mj*RibK reactions with and without magnesium ion suggest that the activity of *Mmp*RibK is dependent on it. *Mj*RibK shows metal independent activity as previously reported.[13] (Experiments were done in duplicates with different batches of proteins.)

4.3.4 Establishing optimal temperature, metal dependence, pH range, and salt concentration for *Mmp*RibK activity

Next, the assay conditions *Mmp*RibK were was optimized. The activities of *Mmp*RibK measured at different temperatures from 15 °C-70 °C have shown that this enzyme is maximum active at 45 °C (Figure 4.5 A).

We have already established that *Mmp*RibK, unlike *Mj*RibK, requires a metal ion to function as a kinase. *Mmp*RibK showed no activity in the absence of a metal; however, it showed full turnover when the divalent metal ion Mg^{2+} was added into the reaction (Figure 4.4 C). We further examined the range of divalent metal ions Mg^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}, Co^{2+}, Fe^{2+}, Cu^{2+}, and Ca^{2+}, and found that *Mmp*RibK was able to phosphorylate riboflavin most effectively in the presence of Mg^{2+} and Zn^{2+} (Figure 4.5 B). The overall trend of activity with different divalent metal ion is $Mg^{2+} = Zn^{2+} > Mn^{2+} > Ni^{2+} > Co^{2+} > Fe^{2+} > Cu^{2+} > Ca^{2+}$ (Figure 4.5 B). This preference is reminiscent of the rat liver riboflavin kinase, which was reported to exhibit a preference of Zn^{2+} ion followed by Mg^{2+} ion.[36]

Next, we tested the *Mmp*RibK enzyme reaction in the pH range of 4.3-12 to determine the optimal pH for its activity. We observed a steady increase in the activity of the
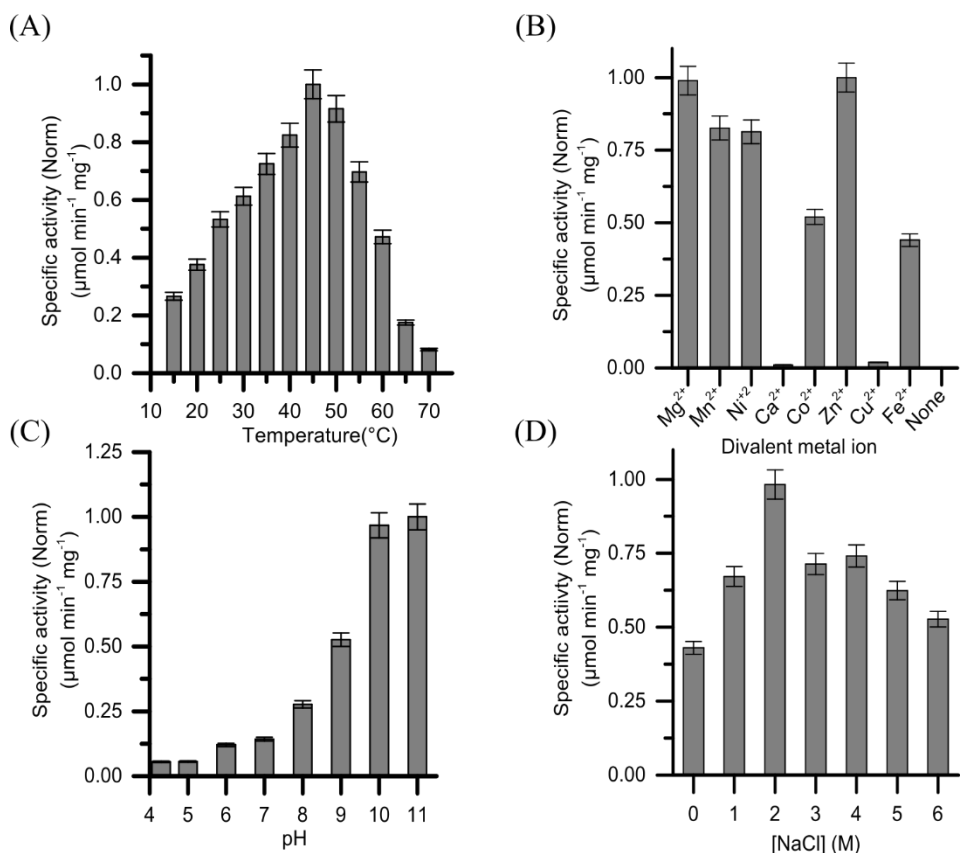
Figure 4.5: Characterization of the biochemical properties of *Mmp*RibK **(A)** Effect of temperature on the activity of *Mmp*RibK. It is highest active at 45 °C. **(B)** Divalent metal ion-dependent activity shows the highest activity of *Mmp*RibK is with Mg^{2+} and Zn^{2+} ions. **(C)** The activity of *Mmp*RibK increases with increasing pH. **(D)** *Mmp*RibK shows highest activity at 2M NaCl and can tolerate even saturating concentrations of NaCl. (Experiments were done in triplicates with same batch of protein)

enzyme up to pH 11 (Figure 4.5 C). A similar trend where an increase in pH leads to higher activity has been previously noted for the *Neurospora crassa* riboflavin kinase.[37]

While conducting enzyme assays with *Mmp*RibK and *Mj*RibK, we observed that the activity of *Mmp*RibK was slower for the conversion of riboflavin into FMN compared to *Mj*RibK. As *M. maripaudis* was first isolated from a salt marsh environment, and such ecosystems contain high concentrations of NaCl, we hypothesized that the *Mmp*RibK activity might be improved in the presence of NaCl. Indeed, the enzymatic activity improved with added NaCl, and a maximum of 2.3x-fold increase was observed with 2M NaCl (Figure 4.5 D). To our knowledge, this is the first riboflavin kinase that has been shown to require a high concentration of NaCl.

4.3.5 Determination of kinetic parameters

We conducted the steady-state kinetic analysis of *Mmp*RibK for both the substrates riboflavin and CTP in the presence of NaCl. In the presence of saturating concentration of CTP, riboflavin displayed the following kinetic parameters: V_{max} 5.9 µM min^{-1}, K_m 185.8 µM and k_{cat} 0.19 s^{-1} (Figure 4.6 A, B). Under saturating concentration of riboflavin, the kinetic parameters for CTP were: V_{max} 3.9 µM min^{-1}, K_m 236.0 µM and k_{cat} 0.14 s^{-1} (Figure 4.6 C, D).

Table 4.3: Kinetics parameters for *Mmp*RibK and *Mj*RibK

	*Mmp*RibK		*Mj*RibK (values obtained from Mashhadi et al., JBact, 2008)	
	Riboflavin	CTP	Riboflavin	CTP
V_{max} (µM min^{-1})	5.9	3.9	Data not available	Data not available
K_m (µM)	185.8	236.0	159	180
k_{cat} (s^{-1})	0.19	0.14	0.334	0.342

The kinetic parameters for *Mmp*RibK is summarized in Table 4.3 along with those reported fo *Mj*RibK (pH 7.2, temperature 70 °C). We compared these values to the parameters reported for *Mj*RibK and established that overall, both enzymes demonstrate similar reaction kinetics, even though there are small differences in reaction conditions for optimal activity (Table 4.3).[13]
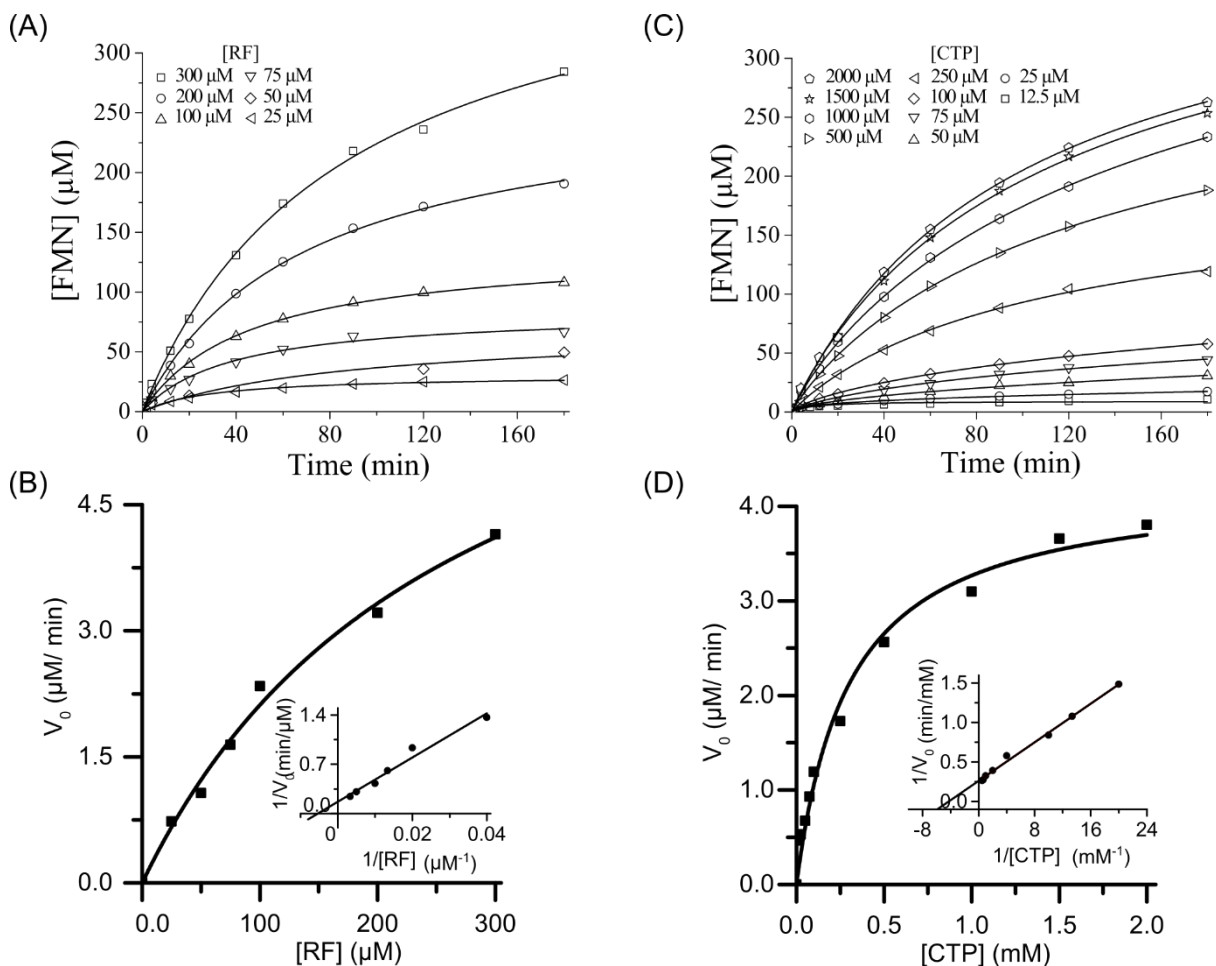
Figure 4.6: Steady-state kinetics of *Mmp*RibK with substrates riboflavin and CTP: **(A)** Initial rates of product (FMN) formation is measured and plotted at different riboflavin concentrations with a saturating concentration of CTP **(B)** Michaelis-Menten plot of *Mmp*RibK for riboflavin under saturating conditions of CTP. The same data is represented as a Lineweaver-Burk plot, shown in the inset. **(C)** Initial rates of FMN formation is measured and plotted at variable concentrations of CTP with a saturating concentration of riboflavin **(D)** Michaelis-Menten kinetics of *Mmp*RibK for CTP under saturating conditions of riboflavin. The Lineweaver-Burk plot for the same data set is provided in the inset. (Experiments were done single time with the same batch of protein.)

In summary, we have established that *Mmp*RibK is a functional mesophilic riboflavin kinase with a strong preference for CTP over other nucleotide triphosphates, with temperature, pH, and NaCl concentration optima of 45 °C, 11 and 2 M, respectively.

4.3.6 The homology modeled *Mmp*RibK structure shows high similarity with the *Mj*RibK crystal structure

The sequence of *Mmp*RibK was subjected to homology modeling, and the modeled apo-*Mmp*RibK was established to have high similarity to the crystal structure *Mj*RibK (PDB: 2VBV). Next, we docked the ligands CTP and riboflavin into the *Mmp*RibK structure and conducted energy minimization to obtain the most favorable configuration. Figure 4.7 A shows the overlap of the CTP and Riboflavin docked *Mmp*RibK with *Mj*RibK. Intrestingly, 1 μs long molecular dynamics (MD) simulations of the whole protein *Mmp*RibK as well as that of the CTP-binding site show that the root mean square deviation (RMSD) decreases sharply when CTP and riboflavin are bound in the model, as compared to the apo *Mmp*RibK structure (Figure 4.7 B). Theoretically, this indicates that *Mmp*RibK adopts a compact conformation in the presence of CTP as compared to when it is not present. Our findings were verified experimentally by a thermal shift assay, in which the broad signal of the apo wild-type *Mmp*RibK changed into a sharp depression with a significant shift in melting temperature on adding CTP (Figure 4.7 C). The control reaction, where ATP was added, showed no change. This not only aligns with our previous observation that the enzyme *Mmp*RibK does not catalyze the riboflavin kinase reaction with ATP but also indicates that ATP is likely unable to bind in the enzyme's nucleotide-binding site.

Both the apo- and docked *Mmp*RibK models show >90% of residues are in the most favored regions of the Ramachandran plot, which indicates that these are high-quality models (Figure 4.7 D, E).[26] The overall quality factor of the modeled structures, calculated on the basis of statistics of non-bonded atomic interactions and distribution of atoms using ERRAT, were found to be 68.9% and 82.4% for the apo- and docked models of *Mmp*RibK, respectively, as compared to 100% for the crystal structure of *Mj*RibK. (Figure 4.7 F, G, H).[27]
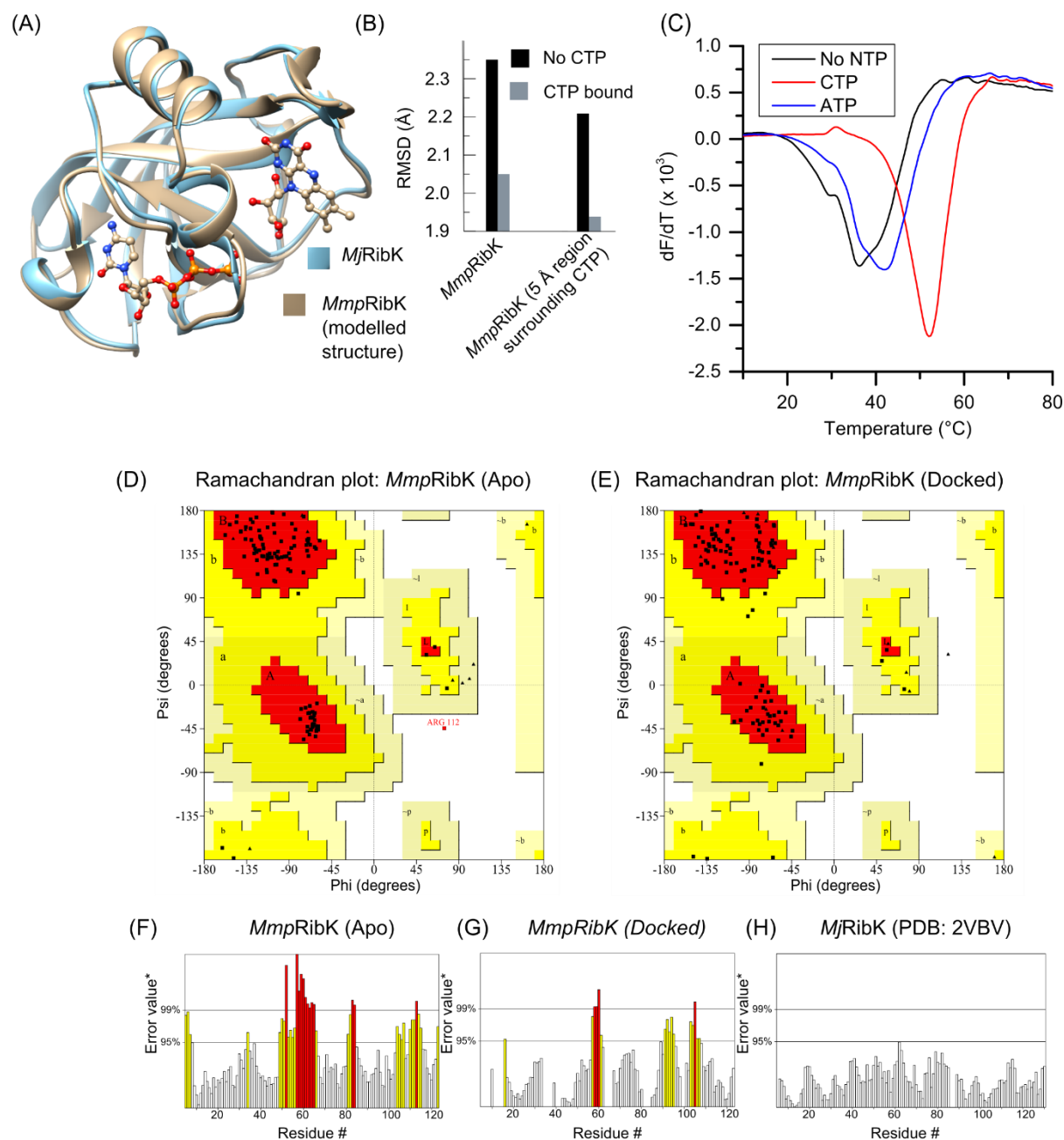
Figure 4.7: Modelling the *Mmp*RibK structure, substrate docking, and evaluation of the modeled *Mmp*RibK structure with Ramachandran analysis and ERRAT analysis. **(A)** Overlay of the *Mj*RibK crystal structure (blue) and the modelled *Mmp*RibK structure (tan) reveals a high degree of tertiary structure similarity. *Mmp*RibK is further docked with substrates riboflavin and CTP. **(B)** The RMSD calculated for *Mmp*RibK in the presence and absence of the ligand CTP. The binding of the CTP appears to stabilize the enzyme. **(C)** Thermal shift assay of *Mmp*RibK with CTP shows that the presence of CTP stabilizes the *Mmp*RibK. When ATP is added to *Mmp*RibK, it does not appear to bind with the protein **(D)** Ramachandran plot of modeled *Mmp*RibK structure in the apo form, which shows that most of the

68

residues (black squares) are in allowed regions. Only one residue, Arg112 (red square), is in the disallowed region. **(E)** Ramachandran plot of *Mmp*RibK docked with CTP and riboflavin. The plot indicating that most of the residues are in the allowed region. Arg112, which was in the disallowed region in the apo form, moved to the allowed region in the docked structure. **(F)** Error-values of residues of *Mmp*RibK apo form using ERRAT analysis indicated the individual quality factors of some residues is poor. **(G)** The corresponding error-values of residues of *Mmp*RibK docked with CTP and riboflavin improve to a large extent **(H)** ERRAT Error values of residues of *Mj*RibK for comparison with the modeled *Mmp*RibK. (*Ramachandran plot of MmpRibK shows the allowed phi-psi dihedral angles of the protein backbone. The red regions are favorable region, and yellow are the allowed regions. The cream regions are generously allowed regions, and white is disallowed regions*)

4.3.7 Probing the structural stability of *Mmp*RibK and *Mj*RibK through denaturation experiment and molecular dynamics simulation

The comparison of the temperature-dependent activity of *Mmp*RibK and *Mj*RibK have shown that *Mmp*RibK is not functional beyond 70 °C, while the activity of *Mj*RibK is unaffected in the temperature range of 45 °C to 95 °C (Figure 4.8 A).

Next, the stability of *Mmp*RibK and *Mj*RibK enzymes were probed using a chemical denaturation agent guanidium hydrochloride (Gdn-HCl).[38] Both enzymes were subject to varying concentrations of Gdn-HCl, and their activities were analyzed using fluorescence-TLC. We observed that *Mmp*RibK tolerates concentrations of Gdn-HCl only up to 1.2 M, while there is no significant change in the activity of *Mj*RibK up to 2.5 M Gdn-HCl (Figure 4.8 B, C).

Both these experiments indicate that even though *Mmp*RibK and *Mj*RibK sequences share 50.4 % identity, the differences in their amino acid compositions and consequently structures confer a large degree of rigidity to *Mj*RibK. To gain a better understanding of the molecular flexibility of both enzymes, we conducted MD simulation studies on the substrate-bound *Mj*RibK and *Mmp*RibK structures and calculated the root mean square fluctuation (RMSF) of the enzyme at 27 °C and 72 °C. The RMSF for *Mmp*RibK at 72 °C increases significantly as compared to that at 27 °C, indicating that temperature increase makes *Mmp*RibK more flexible, consequently resulting in destabilization of the bound CTP (Figure 4.8 D). In sharp contrast, we find that the RMSF of *Mj*RibK remains similar at 27 °C and 72 °C (Figure 4.8 E). Taken together, our experimental results and MD simulations strongly suggest that there exist

features within the *Mj*RibK enzyme which confer it with the ability to retain its function over a large range of temperatures as compared to *Mmp*RibK.
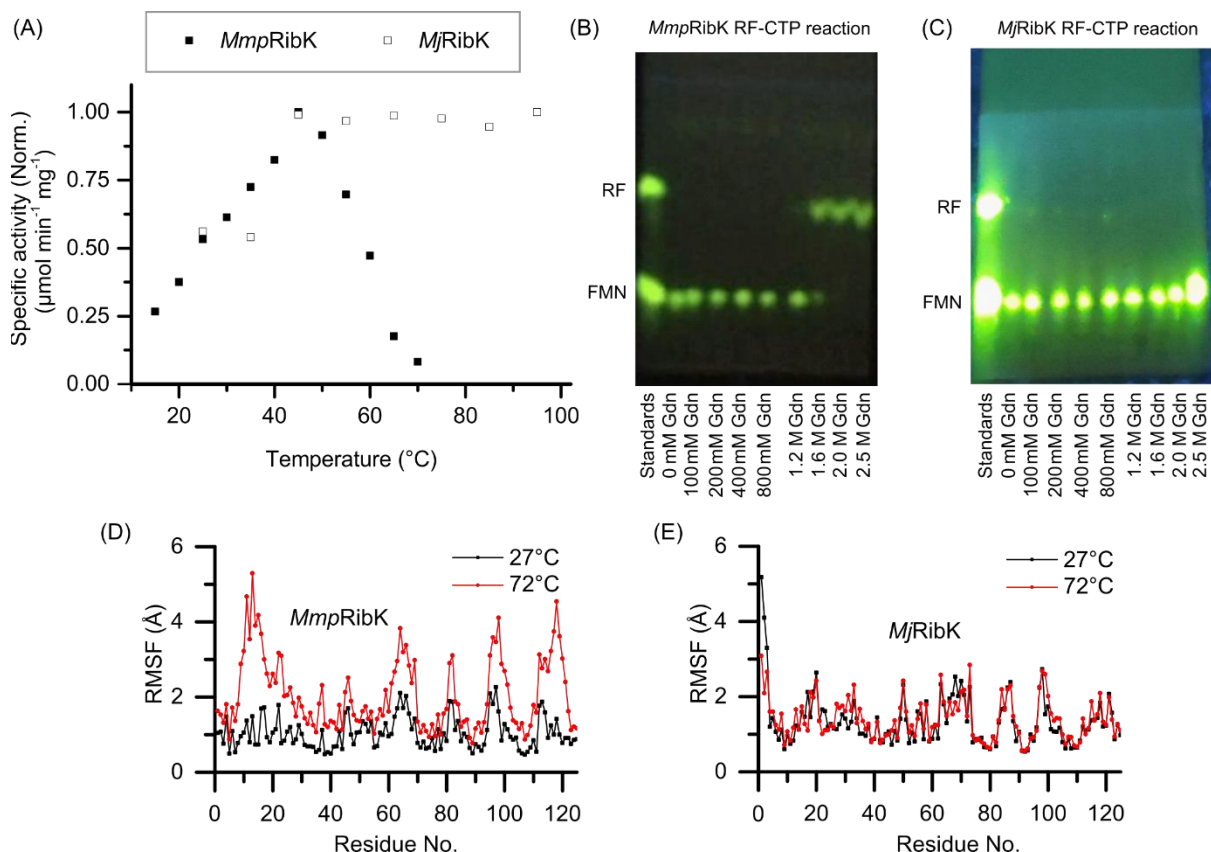
Figure 4.8: Temperature dependence and chemical denaturation of ***Mmp*RibK** and ***Mj*RibK**: **(A)** The specific activity of *Mmp*RibK and *Mj*RibK was calculated from 15 °C to 70 °C and the normalized data is plotted as a function of temperature. The activity increases for *Mmp*RibK and peaks at 45 °C after which it decreases gradually until the enzyme is no longer active at 70 °C. In contrast, *Mj*RibK shows activity at temperatures ranging from 45 °C to 95 °C. **(B)** Activity measurement of *Mmp*RibK in the presence of different concentrations of Gdn-HCl varying from 0 mM to 2.5 M. Enzyme is stable up to 1.2 M Gdn-HCl concentration. After that, it loses its activity. **(C)** Activity measurement of *Mj*RibK in the presence of different concentrations of Gdn-HCl varying from 0 mM to 2.5 M. Enzyme is stable even up to 2.5 M Gdn-HCl concentration and is functional. **(D, E)** Root mean square fluctuation (RMSF) of *Mmp*RibK and *Mj*RibK were computationally evaluated at 27 °C and 72 °C. *Mmp*RibK shows elevated RMSF at 72 °C, indicating that it is more flexible as compared to *Mj*RibK, which is stable at both temperatures.

4.3.8 Analysis of the molecular basis of thermostability of *Mj*RibK and rational engineering of the *Mmp*RibK sequence to introduce thermostable elements

Our results have established that *Mmp*RibK differentiates itself adequately from its thermophilic homolog *Mj*RibK while sharing a high degree of sequence identity and structural features. Thus, this pair of CTP-utilizing kinase enzymes is an excellent candidate for analyzing the general principles of thermal stability in enzymes via rational engineering. We conducted a computational analysis as described in the methods section to predict salt bridges in the *Mmp*RibK and *Mj*RibK structures. The *Mj*RibK had more salt bridge interactions than *Mmp*RibK, including one within the nucleobase loop region in its active site (Table 4.4, Figure 4.9 A-C). This is expected, as salt bridges are noted to confer rigidity and subsequently thermostability to proteins.[39] In order to test the contributions of these predicted salt bridges, we introduced the residues that formed additional salt bridges in *Mj*RibK into the corresponding positions within the *Mmp*RibK sequence, and created three mutants, Mut 1-3 (Table 4.5, Figure 4.9 A-C).
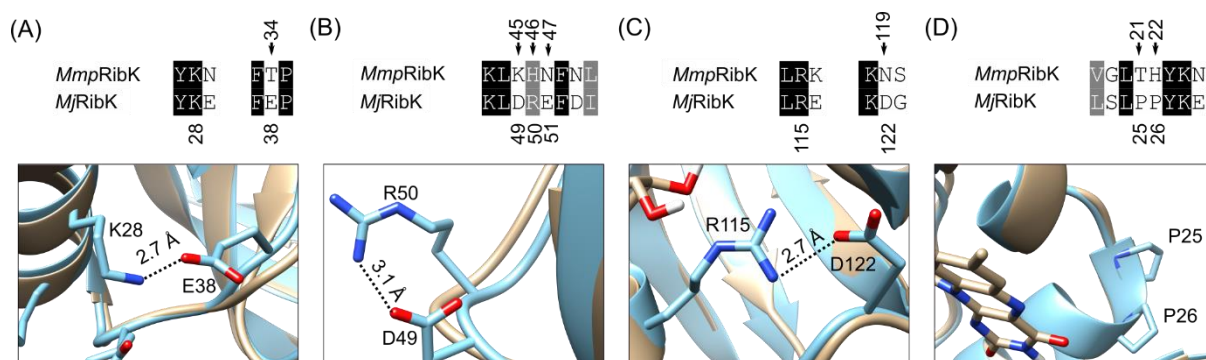
Figure 4.9: Overlay of modeled structure of *Mmp*RibK (tan) and crystal structure of *Mj*RibK (blue) depicting the salt bridges and non-conserved proline residues considered for mutation. (A) K28-E38 salt bridge in *Mj*RibK corresponding to *Mmp*RibK Mut1 (T34E) (B) D49-R50 salt bridge in *MjRibK* corresponding to *Mmp*RibK Mut2 (KHN45-47DRE) (C) R115-D122 salt bridge corresponding to *Mmp*RibK Mut3 (N119D) (D) Non-conserved proline residues in *Mj*RibK corresponding to Mut4 (TH21-22PP)

Table 4.4: Salt bridges present in *Mj*RibK (data obtained from the crystal structure PDB: 2VBU with ligands CDP and Mg^{2+}).

No.	Salt bridges in *Mj*RibK	Corresponding residues in *Mmp*RibK
1	**E38-K28***	T34-K24
2	**E38-K32***	T34-E28
3	E41-K28	E37-K24
4	**D49-R50***	K45-H46
5	E61-K97	E57-K93
6	**D90-K81***	N87-K81
7	**D122-R115***	N119-R111
8	D124-K121	D120-K117
*Salt bridges that exist in *Mj*RibK but are absent in *Mmp*RibK		

Additionally, proline residues are also known to confer thermostability in a mesophilic enzyme by rigidifying the overall structure.[11] Based on the sequence alignment of *Mj*RibK and *Mmp*RibK, we replaced residues T21 and H22 in *Mmp*RibK by two consecutive prolines to create Mut4 (Table 4.5, Figure 4.9 D). Mut 1-4 were tested for activity at 60 °C, where the wild-type enzyme drastically loses activity. Except Mut1, the remaining three mutants showed similar or better turnover as compared to the wild-type enzyme. Hence, we proceeded to create combinations of Mut 2-4 involving salt bridges and proline residues to generate Mut 5-8. (Table 4.5).

Table 4.5: List of mutants and corresponding mutations

Mutant	Residue/s of *Mmp*RibK replaced with that in *Mj*RibK	Putative salt bridge introduced	No. of salt bridges introduced	No. of residues replaced by Pro
Mut1	T34E	E34-K24	1	0
Mut2	KHN(45-47)DRE	D45-R46	1	0

Mut3	N119D	D119-R111	1	0
Mut4	TH(21-22)PP	-	0	2
Mut5	KHN(45-47)DRE/N119D	D45-R46/ D119-R111	2	0
Mut6	KHN(45-47)DRE/TH(21-22)PP	D45-R46	1	2
Mut7	TH(21-22)PP/N119D	D119-R111	1	2
Mut8	KHN(45-47)DRE/N119D/TH(21-22)PP	D45-R46/ D119-R111	2	2

4.3.9 Molecular dynamics simulations, thermal shift analysis, and specific activity of the *Mmp*RibK mutants validate the role of amino acids that confer thermostability

To gain a better understanding of the flexibility and stability of the eight *Mmp*RibK mutants, we conducted MD simulation studies and calculated the RMSD of their modeled structures over a period of 300 ns, followed by the RMSF of individual residues in each (Figure 4.10 A, B). We found that all eight enzymes exhibited a relatively stable RMSD value except for Mut 4 and Mut8 (Figure 4.10 A). Also, the RMSF values at 72 °C indicate that the residues in all the mutants remain stable except those in Mut1. This behavior of Mut1 compares to the wild-type enzyme at that temperature, confirming our experimental result that Mut1 loses activity at 60 °C or higher (Figure 4.10 B, C). We performed a thermal shift analysis with wild-type *Mmp*RibK and its mutants in order to establish their melting temperatures (Figure 4.10 D). Briefly, the change in the signal intensity with increasing temperature, attributable to the gradual unfolding of the protein mixed with a fluorescent dye, was correlated to its thermal stability.
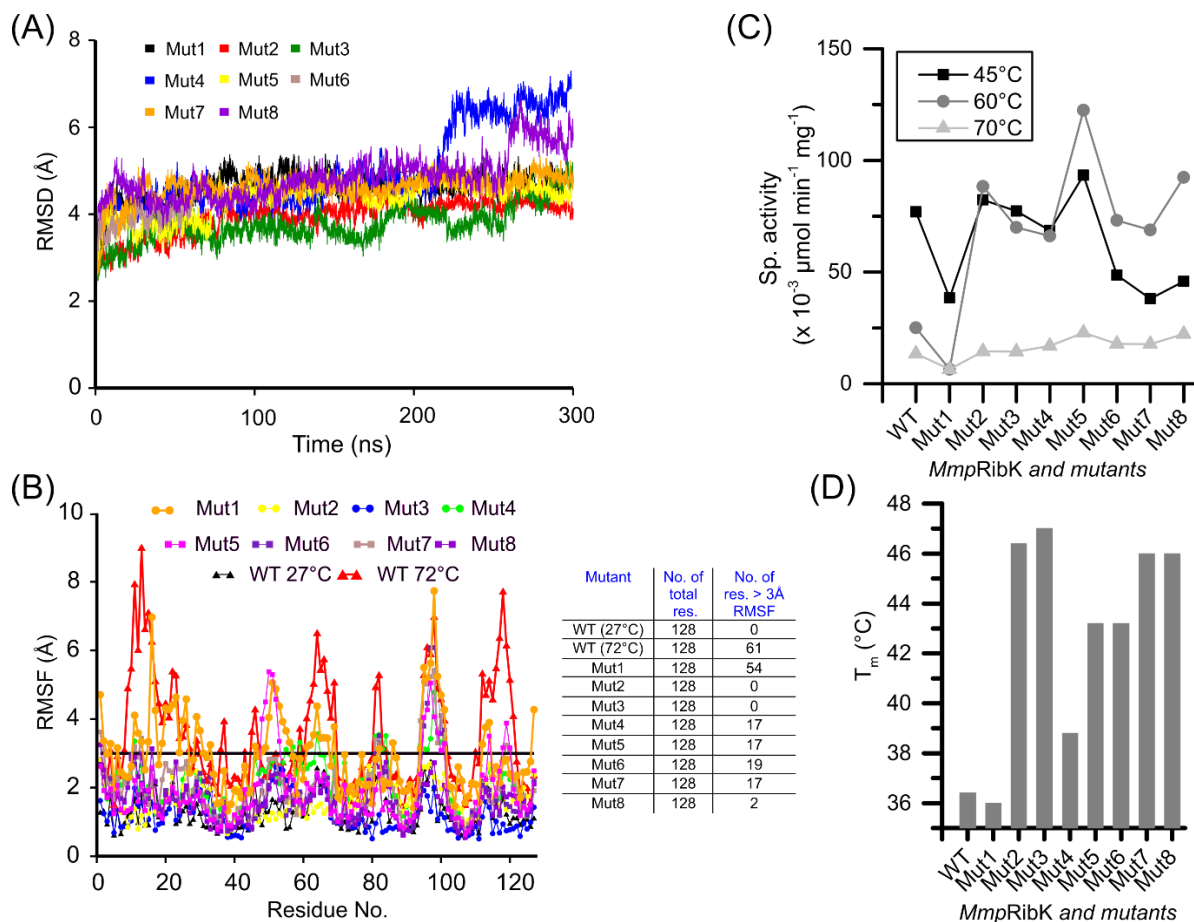
Figure 4.10: RMSD and RMSF calculations, specific activity and melting temperature analysis of the wild-type and *Mmp*RibK mutants (A) RMSD of the wild-type enzyme and the eight *Mmp*RibK mutants was calculated over a 300 ns long molecular dynamic trajectory. The RMSD analysis shows that enzymes are stable except for Mut4 and Mut8 that show higher values indicating that these mutants may be partially unfolded (B) An RMSF performed for the amino acid residues of the *Mmp*RibK wild-type and mutants calculated over a 300 ns long molecular dynamic trajectory, show that Mut 2-8 have gained thermostability, except Mut1 which has high fluctuations and behaves similarly to the wild-type *Mmp*RibK at 72 °C. The table show the number of amino acid with greater than 3 Å RMSF. (C) The specific activities of *Mmp*RibK and mutants measured at 45 °C, 60 °C, and 70 °C show that all the mutants except Mut1 have higher specific activity as compared to the wild-type at 60 °C (D) Melting temperatures of the wild-type *Mmp*RibK and its mutants determined by the thermal shift assay show higher stability in Mut 2-8. (Experiments were done single times with same batch of protein)

Our results show that Mut1 (with an E-K salt bridge) has a similar melting temperature as the wild-type enzyme. Mut4 (with two P residues) was stabilized to a small extent (T_m= 38.8°C).

74

Mut2 (with a D-R salt bridge) and Mut3 (with a D-R salt bridge) were significantly stabilized (T_m=46.4°C and T_m=47°C, respectively) (Figure 4.10 C, Figure 4.11). The remaining combination mutants Mut 5-8 all displayed T_m higher than the wild-type enzyme, but surprisingly, these were lower than that of Mut2 and Mut3 (Figure 4.10 C, Figure 4.11).
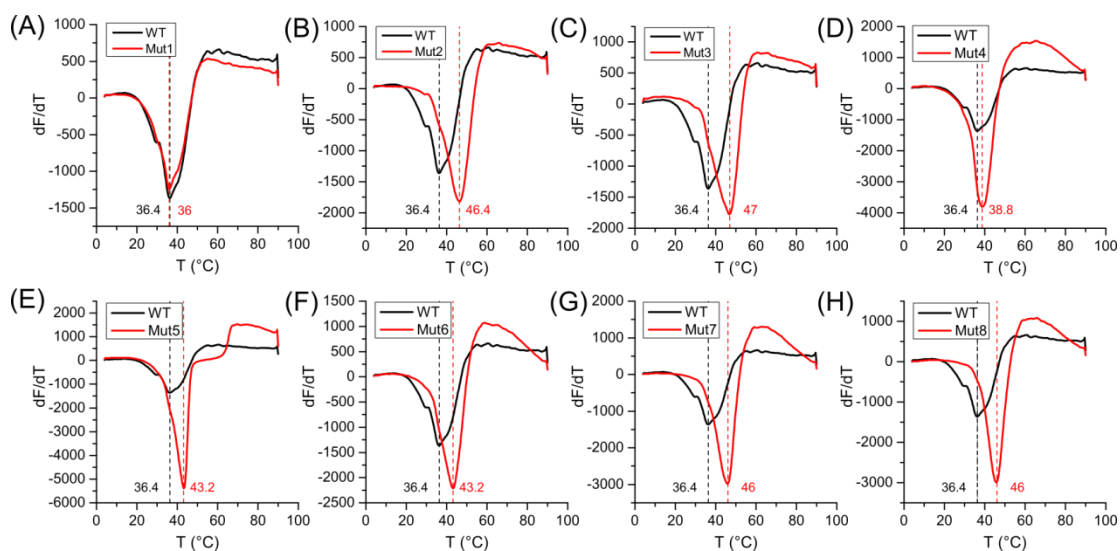
Figure 4.11: Thermal shift assay of wild-type (WT) *Mmp*RibK and mutants. We determined the melting temperatures of the wild-type and mutant *Mmp*RibK enzymes, which gives the extent of thermostability introduced by the mutations. **(A)- (E)** The black curves are the melting curves of wild-type, and red curves are the melting curves of mutants. The respective melting temperatures of wild-type and mutants are noted adjacent to their melting curves in black (wild-type) and red (mutants) letters, and the vertical black and red dotted lines indicate these numbers on the x-axis. The shifts of melting temperatures of mutants from the wild-type indicate the extent of thermostabilization. (Experiments were done single times with same batch of protein)

Since the thermostabilization of an enzyme may occur at the cost of its activity, we compared the activity of the wild-type and the mutants at higher temperatures. Temperature-dependent activities of the wild-type *Mmp*RibK and its mutants were conducted at 45 °C, 60 °C and 70 °C (Figure 4.10 D). Wild-type *Mmp*RibK has reduced activity at 60 °C and 70 °C in comparison to that at 45 °C (also seen in Figure 4.8 A). At 45 °C, the activities of wild-type *Mmp*RibK, Mut2, Mut3, Mut4 and Mut5 are comparable, while Mut6, Mut7 and Mut8 have reduced activity at this temperature. On the other hand, at 60 °C, all mutants except Mut1, have

higher activity when compared to wild-type *Mmp*RibK. The highest activity is displayed by Mut5 which contains two added salt bridges (Table 3). Finally, at 70 °C, the wild-type and mutants have significantly diminished activity. This result demonstrates that using the rational engineering strategy of comparative sequence analysis followed by creating appropriate mutants, we were able to achieve a 15 °C improvement in thermal stability of *Mmp*RibK without loss in activity.

4.3.10 Nucleotide specificity of thermostable mutants

The introduction of salt bridges and proline residues conferred thermostability to the mutants of *Mmp*RibK, without loss in activity. Next, we checked the nucleotide specificity of these mutant with different NTPs. Figure 4.12 shows the TLC of activity assay of mutants with different nucleotides. None of them had shown altered nucleotide specificity.
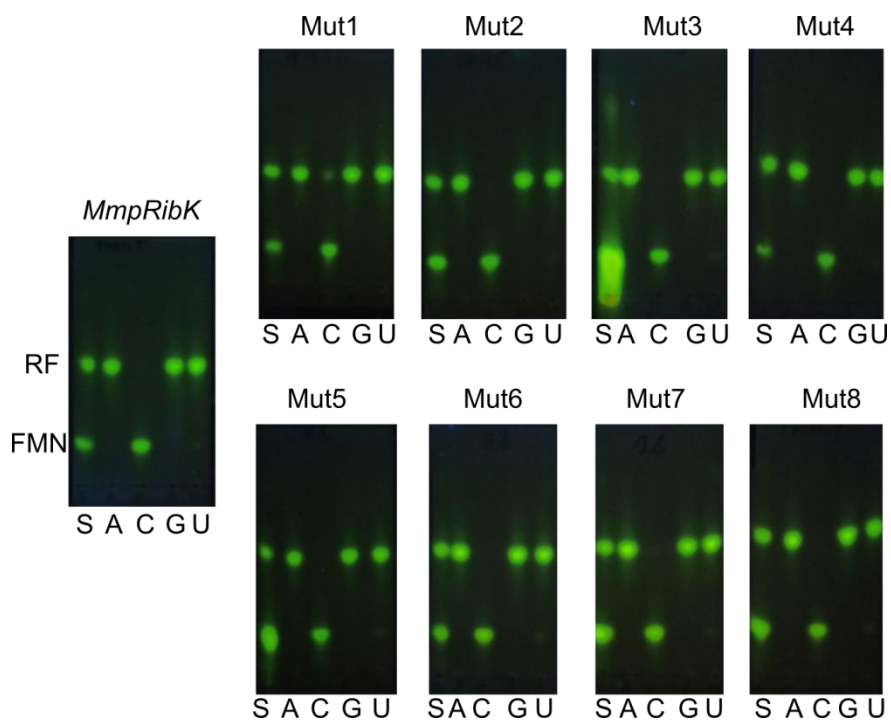
Figure 4.12: TLC showing the reactions of wild-type *Mmp*RibK and mutants with different NTPs. S is the standards for riboflavin (RF, top spot) and flavin mononucleotide (FM, bottom spot). A, C, G, and U represents the RibK reactions with riboflavin and nucleotide triphosphates ATP, CTP, GTP and UTP respectively. Wild-type and all the mutants show activity only with CTP, indicating that the mutations

have no effect on the nucleotide specificity of *Mmp*RibK. (Experiments were done single times with same batch of protein)

4.4 Conclusion

In this study, we report the successful reconstitution and biochemical characterization of a novel mesophilic CTP-dependent riboflavin kinase from the archaea *M. maripaludis*. The use of CTP as the substrate for a kinase is a rare occurrence, thus the characterization of *Mmp*RibK adds to the handful of examples in literature that will allow us to explore the origins of nucleotide choice in kinases.

The insight from this study are as follows:

4.4.1 The *Mmp*RibK is the first mesophilic CTP-dependent enzyme characterised so far from archaea.

4.4.2 *MmpR*ibK resembles *Mj*RibK in structural aspects, however it differs form the thermophilic homolog in terms of metal dependent behaviour and temperature requirement.

4.4.3 Incorporation of corresponding salt bridges and proline residues from *Mj*RibK to *Mmp*RibK can rigidify it and confer thermostability.

4.4.4 Even after such rigorous mutational variation, *Mmp*RibK does not alter its specificity for the NTPs other than CTP.

Our studies on the characterization of *Mmp*RibK and creation of its thermostable mutants not only adds important mechanistic information to this class of enzymes, but also puts forth an important candidate for biochemical and structural investigations and bioengineering applications in the future.

References:

1.   Adams, M. W. Enzymes and proteins from organisms that grow near and above 100 degrees C. *Annu. Rev. Microbiol.* **47**, 627–58 (1993).

2.   Tronelli, D., Maugini, E., Bossa, F. & Pascarella, S. Structural adaptation to low temperatures − analysis of the subunit interface of oligomeric psychrophilic enzymes. *FEBS J.* **274**, 4595–4608 (2007).

3.   Kamal, Z., Anwar, T., Mohammad, S. & Krishnamoorthy, G. Role of Active Site Rigidity

in Activity : MD Simulation and Fluorescence Study on a Lipase Mutant. **7**, 1–8 (2012).

4.   Singh, B., Bulusu, G. & Mitra, A. Understanding the Thermostability and Activity of Bacillus subtilis Lipase Mutants: Insights from Molecular Dynamics Simulations. (2015) doi:10.1021/jp5079554.

5.   Pfleger, C. *et al.* Constraint Network Analysis (CNA) and. 1–40 (2014).

6.   Radestock, S. & Gohlke, H. Protein rigidity and thermophilic adaptation. *Proteins Struct. Funct. Bioinforma.* **79**, 1089–1108 (2011).

7.   Krüger, D. M., Rathi, P. C., Pfleger, C. & Gohlke, H. CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo-)stability, and function. *Nucleic Acids Res.* **41**, 340–348 (2013).

8.   Rathi, P. C., Jaeger, K. E. & Gohlke, H. Structural rigidity and protein thermostability in variants of lipase a from Bacillus subtilis. *PLoS One* **10**, 1–24 (2015).

9.   Rathi, P. C., Fulton, A., Jaeger, K. E. & Gohlke, H. Application of Rigidity Theory to the Thermostabilization of Lipase A from Bacillus subtilis. *PLoS Comput. Biol.* **12**, 1–21 (2016).

10.   Gromiha, M. M., Pathak, M. C., Saraboji, K., Ortlund, E. A. & Gaucher, E. A. Hydrophobic environment is a key factor for the stability of thermophilic proteins. *Proteins Struct. Funct. Bioinforma.* **81**, 715–721 (2013).

11.   Watanabe, K., Masuda, T., Ohashi, H., Mihara, H. & Suzuki, Y. Multiple Proline Substitutions Cumulatively Thermostabilize Bacillus Cereus ATCC7064 Oligo-1,6-Glucosidase: Irrefragable Proof Supporting the Proline Rule. *Eur. J. Biochem.* **226**, 277–283 (1994).

12.   Lam, S. Y., Yeung, R. C. Y., Yu, T., Sze, K. & Wong, K. A Rigidifying Salt-Bridge Favors the Activity of Thermophilic Enzyme at High Temperatures at the Expense of Low-Temperature Activity. **9**, (2011).

13.   Mashhadi, Z., Zhang, H., Xu, H. & White, R. H. Identification and characterization of an archaeon-specific riboflavin kinase. *J. Bacteriol.* **190**, 2615–2618 (2008).

14.   Ammelburg, M. *et al.* A CTP-Dependent Archaeal Riboflavin Kinase Forms a Bridge in the Evolution of Cradle-Loop Barrels. *Structure* **15**, 1577–1590 (2007).

15.   Rodionova, I. A. *et al.* A novel bifunctional transcriptional regulator of riboflavin metabolism in Archaea. *Nucleic Acids Res.* **45**, 3785–3799 (2017).

16. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5-9 (2008).

17. Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* **41**, W29-33 (2013).

18. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7-19 (2016).

19. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–7 (2004).

20. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

21. van den Ent, F. & Löwe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **67**, 67–74 (2006).

22. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–54 (1976).

23. Stoscheck, C. M. Quantitation of protein. *Methods Enzymol.* **182**, 50–68 (1990).

24. Deluca, C. & Kaplan, N. O. Large scale synthesis and purification of flavin adenine dinucleotide. *J. Biol. Chem.* **223**, 569–76 (1956).

25. Sali, A. Comparative protein modeling by satisfaction of spatial restraints. *Mol. Med. Today* **1**, 270–7 (1995).

26. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).

27. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–9 (1993).

28. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–8, 27–8 (1996).

29. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

30. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **9**, 1–5.

31. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–91 (2009).

32. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–25 (2006).

33. And, M. J. F. and G. W. T. and H. B. S. and G. E. S. *et al.* Gaussian 03.

34. M.J. Abraham, D. van der Spoel, E. Lindahl, B. H. and the G. development team. GROMACS User Manual version 2019. (2019).

35. Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. protein Sci.* **79**, 28.9.1-28.9.14 (2015).

36. Merrill, A. H. & McCormick, D. B. Affinity chromatographic purification and properties of flavokinase (ATP:riboflavin 5'-phosphotransferase) from rat liver. *J. Biol. Chem.* **255**, 1335–8 (1980).

37. Rajeswari, S. R., Jonnalagadda, V. S. & Jonnalagadda, S. Purification and characterization of flavokinase from Neurospora crassa. *Indian J. Biochem. Biophys.* **36**, 137–142 (1999).

38. Monera, O. D., Kay, C. M. & Hodges, R. S. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci.* **3**, 1984–91 (1994).

39. Donald, J. E., Kulp, D. W. & DeGrado, W. F. Salt bridges: geometrically specific, designable interactions. *Proteins* **79**, 898–915 (2011).

Chapter 5

In vitro synthesis of FAD analogs using FAD biosynthesis enzymes from *Bacillus subtilis* and *Methanocaldococcus jannaschii*

5.1 Introduction

The riboflavin biosynthesis pathway of *Bacillus subtilis* contains two different enzymes in the last steps of the active flavin cofactor biosynthesis. The enzyme riboflavin kinase (*Bs*RFK or *Bs*RibR) phosphorylates riboflavin to form FMN using ATP.[1] Another enzyme, FAD synthetase (*Bs*FADS or *Bs*RibC), is a bifunctional enzyme. It phosphorylates riboflavin to form FMN, followed by adenylation of FMN to create FAD as a final product (Figure 5.1).[2] *Bs*RFK and *Bs*FADS are unique ATP-dependent enzymes because these operate under reducing conditions and utilize only the reduced form of riboflavin.[3] Various studies have been done so far for these enzymes to study the riboflavin biosynthesis in *Bacillus subtilis*. The enzyme *Bs*RFK is encoded by the gene *ribR* of *Bacillus subtilis,* while the gene *ribC* encodes the enzyme *Bs*FADS. These genes are among the 12 genes present in the locus *ytmI-ytnM* of *Bacillus subtilis,* and *ribR* negatively regulates the *ribC* gene (Solovieva. FEBS, 2005).[4] However, the *rib*C gene is essential, and its complete inactivation has been lethal to cells.[5]

The archaeal monofunctional FMN adenylyltransferase (*Mj*FMNAT or *Mj*RibL) from *Methanocaldococcus jannaschii* is a *ribL* gene product.[6] It is a unique enzyme and converts FMN into FAD under reduced conditions (Figure 5.1). This enzyme belongs to the nucleotidyltransferase family and is conserved in the archaeal genome. Non-archaeal FAD synthetase utilizes ATP for the formation of FAD from riboflavin.[7] On the other hand, it has been found that *Mj*FMNAT can use CTP as well for the nucleotidyltransferase reaction to form the uncommon cofactor FCD.[6] FCD formation by this enzyme is surprising because in a cellular context, FAD is an essential cofactor for many biological pathways. Several studies have been done to engineer the metabolic pathway of certain organisms having unnatural cofactor preferences. In this regard, the formation of FCD by *Mj*FMNAT can be used to explore metabolic engineering in archaeal organisms.
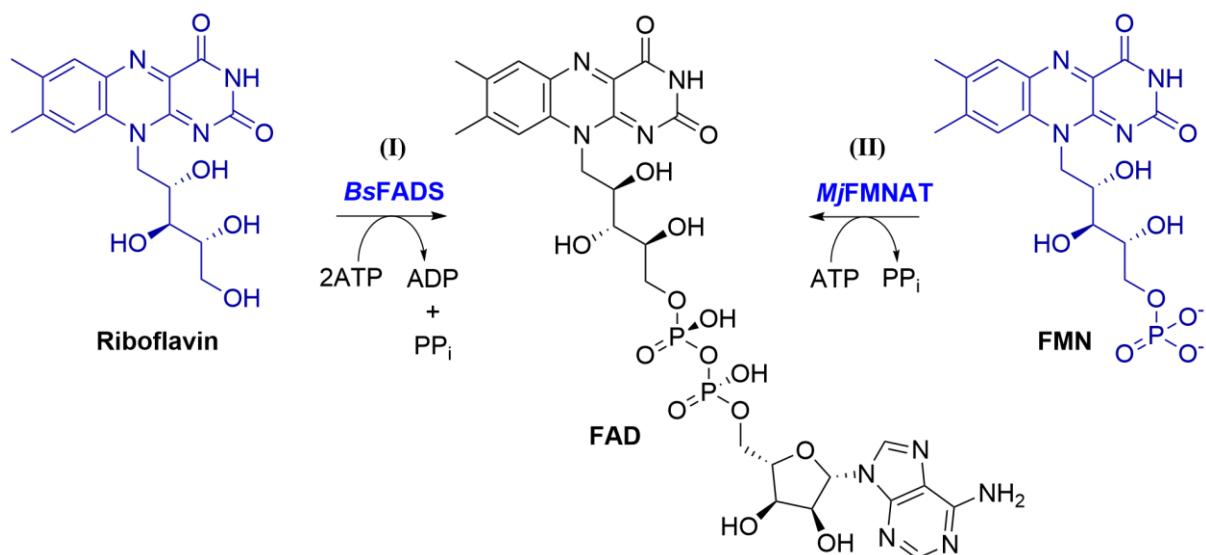
Figure 5.1: Reaction scheme showing the formation of FAD from riboflavin and ATP by the action of *Bs*FADS (I). *Mj*FMNAT uses FMN and ATP as substrates to form FAD (II).

In this study, we analyzed the ability of *Bs*FADS and *Mj*FMNAT to utilize different NTPs for the phosphorylation and/or nucleotidyltransferase reactions using oxidized and reduced riboflavin. When *Bs*FADS or *Mj*FMNAT use ATP for the nucleotidyltransferase reaction, the AMP portion of ATP attaches to the FMN to form FAD (Figure 5.1). However, NTP other than ATP for nucleotidyltransferase reaction of *Bs*FADS or *Mj*FMNAT might result in unnatural FAD analogs. These FAD analogs differ among themselves in the nucleobase part (Figure 5.2).
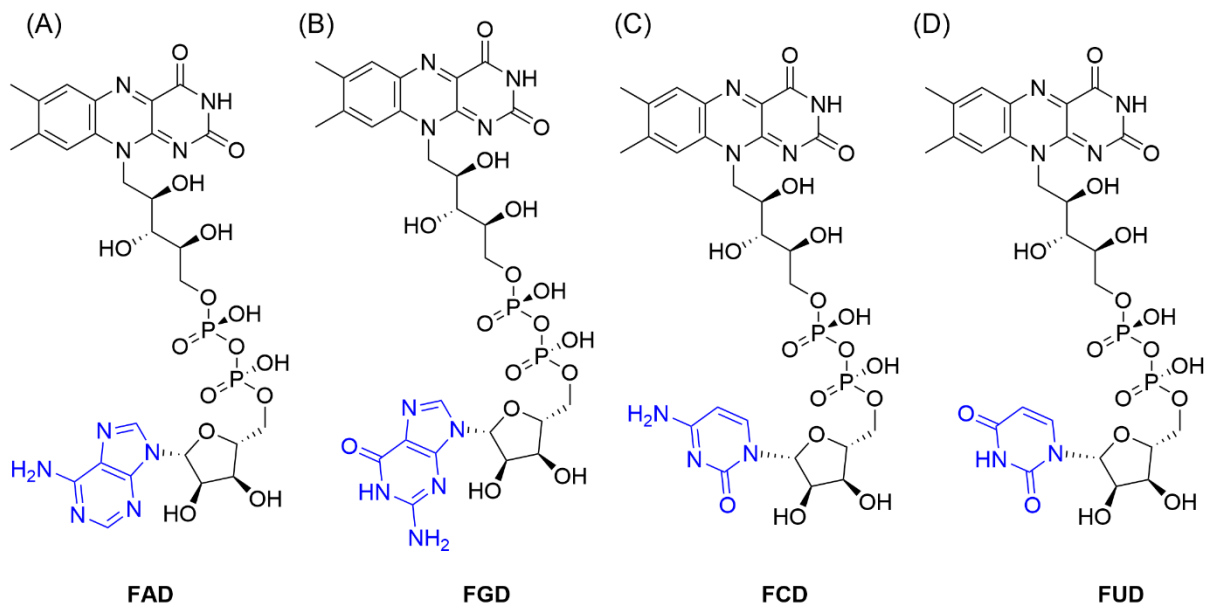
Figure 5.2: FAD and its analogs differ in the nucleobase moiety. FAD has an adenine ring, while in FGD, FCD, and FUD, the adenine nucleobase is replaced by guanine, cytosine, and uracil nucleobases. *We have shown the formation of FAD analogs from FMN and different nucleotide triphosphates in the presence of the MjFMNAT enzyme. The activities of the enzymes from Bacillus subtilis depend on the redox state of the flavin. Our result shows a different pattern of nucleotide utilization under oxidizing and reducing conditions.*

There is no crystal structure available for *Bs*FADS and *Mj*FMNAT. *To better understand the structure-function relationship of these enzymes, we did a structural bioinformatics study of these enzymes. Our sequence comparison and homology modeling predicted the important structural motifs responsible for nucleotide binding.*

5.2 Materials and methods

5.2.1 Materials

All the reagents and media components used in the cloning, overexpression, purification, and assays are detailed in the Materials and methods section of Chapter 3. The snake venom phosphodiesterase was purchased from Sigma Aldrich.

5.2.2 Isolation of *Bacillus subtilis* genomic DNA

The *Bacillus subtilis* strain 'MTCC 2422' (NCIM ID 2920) was purchased from "National collection of industrial microorganisms (NCIM), NCL, Pune (India).

25 mL of LB media was inoculated with a single fresh colony of *Bacillus subtilis* and incubated at 37 °C in an incubator shaker for 8-9 h. Cells were harvested by centrifugation (10 min, 5000 RPM) at room temperature. Cell pellets were re-suspended in 400 µL of Tris-EDTA-NaCl (TEN) buffer (Table 5.1) and transferred to a 2 mL micro-centrifuge tube. 20 µL of lysozyme solution (20 mg/mL) was added to the solution and incubated at 37 °C for 20 min.

2 µL of RNAse A (20 mg/mL) was added to the solution and incubated for 3 min at 65 °C. 40 µL of SDS (10% W/V), 550 µL of TEN* buffer, and a small amount of proteinase K (Table 1) were added, mixed, and incubated for 2 h at 60 °C.

900 µL of phenol (equilibrated with Tris-EDTA (TE) buffer, pH 8.0) was added to the solution and mixed well by inverting the tube. The solution was centrifuged for 5 min at 13000

RPM at room temperature, and the upper aqueous phase was transferred to a new 2 mL micro-centrifuge tube. Extraction was repeated once with phenol and twice with chloroform: isoamyl alcohol (24:1).

The aqueous phase was transferred to 10 mL of ethanol previously chilled at -20 °C in a test tube. The precipitated DNA was coiled up with the help of a bended-tip of Pasteur pipette. DNA was air-dried and dissolved in 100 µL of TEN* buffer. The concentration of DNA was measured on the Nanodrop.

Table 5.1: Composition of buffer and solutions used in isolation of genomic DNA from *Bacillus subtilis*

Buffers and solutions	Composition
Lysozyme solution	20 mg/mL Lysozyme, 10 mM tris-HCl pH 8.0
RNase A solution	20 mg/mL RNase A, 15 mM NaCl , 10 mM Tris-HCl pH 8.0
TE buffer	10 mM Tris-HCl pH 8.0, 1 mM EDTA
TEN buffer	10 mM Tris-HCl pH 8.0, 10 mM EDTA, 150 NaCl
TEN* buffer	10 mM Tris-HCl pH 8.0, 10 mM EDTA, 50 NaCl

5.2.3 Cloning of *Mj*FMNAT, *Bs*FADS, and *Bs*FADS S166K mutant gene

The genes of *Mj*FMNAT (*Mj*ribL: UniProtKB - Q58579), *Bs*FADS (*Bs*RibC: UniProtKB - P54575), and the mutant *Bs*FADS S166K were amplified through the PCR reactions using restriction-free cloning as described in the methods section of Chapter 3. The sets of primers used for the cloning are given in Table 5.1.

Table 5.2: Primers used in the cloning of *Bs*FADS, the mutant S166K and *Mj*FMNAT

#	*Bs*FADS	
1	F1	5'-GTGAAGACGATACATATTACACATC-3'
2	R1	5'-TTATTTCCGCAAATTGCTGAAATAAC-3'
3	F2	5'-GTGCCGCGCGGCAGCCATATG AAGACGATACATATTACACATC-3'
4	R2	5'-CGACGGAGCTCGAATTCGGATCCTTATTTCCGCAAATTGCTGAAATAAC-3'
	*Bs*FADS S166K	
5	F1	5'-TAAAAAAATCAAATCTTCGTATATCCGT-3'
6	T7 R1	5'-GCTAGTTATTGCTCAGCGG-3'
	*Mj*FMNAT	
7	F1	5'-ATGAAAAAGAGGGTAGTTACCGC-3'

8	R1	5'-TTAGATTTTAATCTCTTTATTG CAG-3'
9	F2	5'-GTGCCGCGCGGCAGCCATATGAAAAAGAGGGTAGTTACCGC-3'
10	R2	5'-CGACGG AGCTCGAATTCGGATCCTTAGATTTTAATCTCTTTATTGCAG-3'

5.2.4 Expression and purification of wild-type enzymes and mutants

The purified plasmids of *Mj*FMNAT and *Bs*FADS were transformed into *E. coli* BL21 (DE3) chemically competent cells. The cells were grown in Luria-Bertani broth at 37 °C. When the OD_{600} of the culture was 0.6, it was induced with 1 mM IPTG at 18°C. After the post-induction time of 16 h, cells were harvested by centrifugation. The cell pellet was dissolved in lysis buffer (100 mM Tris-HCl pH 8.0 and 300 mM NaCl, 0.025% beta-mercaptoethanol, 0.1 mM PMSF) and sonicated. The lysed cells were centrifuged, and the supernatant was loaded onto a pre-equilibrated HiTrap Ni-NTA column (GE healthcare) (100 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.025% beta-mercaptoethanol) using FPLC system (Akta pure, GE healthcare). The non-specifically bound proteins were removed with an isocratic wash of 40 mM and 60 mM imidazole buffer. The final elution of protein was done with 250 mM imidazole buffer (Tris-HCl pH 8.0, 300 mM NaCl, 0.025% beta-mercaptoethanol, 250 mM imidazole). The fractions containing pure proteins were pooled together. 10% glycerol was added to the protein, flash-frozen, and stored at -80 °C. The concentrations of proteins of each batch were determined by standard Bradford assay using bovine serum albumin (BSA) before the experiment.[8] All mutants of these enzymes were also purified in the same manner.

5.2.5 Activity assay of *Bs*FADS, *Mj*FMNAT and mutants with NTPs

As literature reports showed that *B. subtilis* enzymes work with the reduced form of flavins, we tested the activity of *Bs*FADS under oxidized and reduced conditions with riboflavin and the nucleotides ATP, CTP, GTP, and UTP. For the reaction under reduced conditions, 100 µM of riboflavin, 100 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 2 mM NTP, 10 mM sodium dithionite, and 1 µM of *Bs*FADS in a reaction volume of 200 µL were incubated at 37 °C for 4 h inside the anaerobic glove box. The incubation of the reaction mixture containing sodium dithionite inside the glove box were used to maintain the reduced state of riboflavin. Flavins are yellow in color but become colorless in the presence of the reducing agent sodium dithionite. Anoxic conditions are used to prevent the oxidation of sodium dithionite. The reaction products

of NTP reactions with *Bs*FADS were checked on TLC and HPLC in a similar fashion to that used in chapter 3.

The activities of *Bs*FADS with oxidized state of riboflavin against different NTPs were assayed outside the anaerobic glove box using the same composition but without adding sodium dithionite.

The activity of the mutant *Bs*FADS S166K was also assayed and analyzed using the same protocol used for wild-type *Bs*FADS. However, it was tested only under aerobic conditions with riboflavin as a substrate.

We tested the activity of *Mj*FMNAT with different NTPs. For this purpose, 170 µM FMN, 10 mM NTP, 7 mM MgCl$_2$, 14 mM DTT, 50 mM Tris-HCl pH 8.0 and 5 µM enzyme were mixed in a total volume of 500 µL and heated at 70 °C for 12 h. After 12 h, the reaction was quenched by adding 60 mM acetic acid (final concentration), the formation of FAD and its analogs were analyzed on the TLC and reverse-phase HPLC methods provided in the Materials and Methods section of Chapter 3. The FAD and FAD analogs formed in the *Mj*FMNAT reaction were purified and collected using the same HPLC method. For the collection of analogs 200 µL of samples were injected through the autosampler (50 µL at a time) and collected manually. The solvents from the samples were evaporated using Centrivap DNA concentrator (Labconco corporation). The dried powder of analogs were stored at -20 °C. To determine the yield of dFAD and FGD, the dried powder is dissolved in 200 µL of buffer and injected into the HPLC. The concentrations of dFAD and FGD were determined from area under the curve, assuming these have same absorption as FAD. The percentage yield of the product was calculated with respect to the initial concentration of the substrate used in the reaction.

5.2.6 Phosphodiesterase assay to confirm the formation of FAD analogs

The formation of FAD analogs was tested using the snake venom phosphodiesterase (PDE), which can cleave the phosphodiester bond of FAD analogs to form FMN and AMP (Figure 5.3). For this, 20 µL of each of the *Mj*FMNAT reaction products was taken in separate microcentrifuge tubes, and 5 µL of phosphodiesterase was added. Samples were incubated at 37 °C for 14 h. After 14 h, 10 µL of the phosphodiesterase reaction product was loaded on the HPLC to analyze the product thus formed.
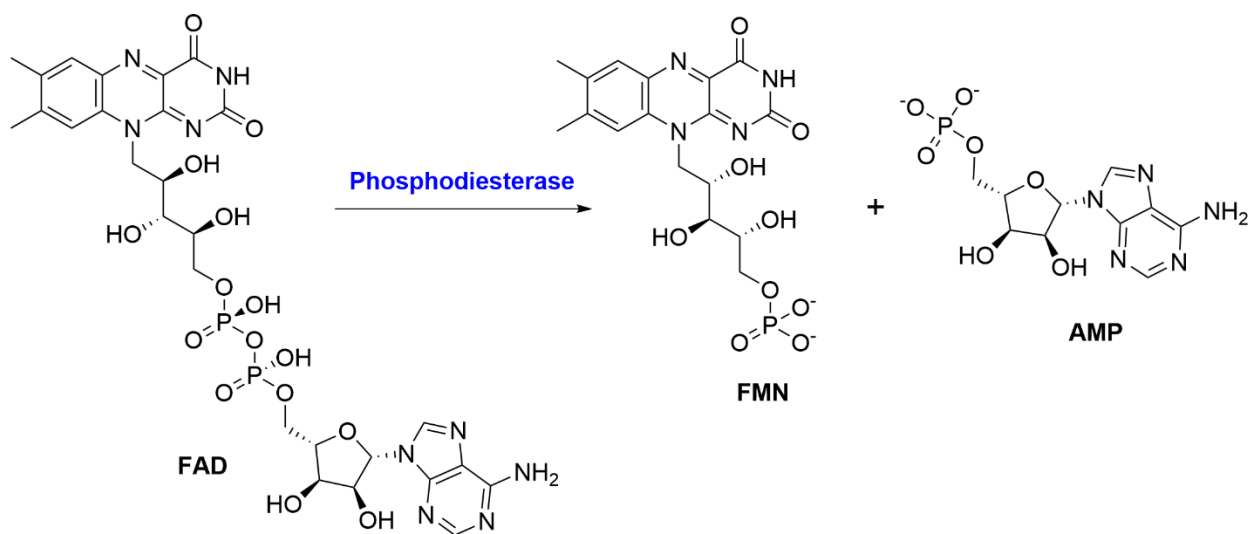
Figure 5.3: The snake venom phosphodiesterase enzyme cleaves the phosphodiester bond of FAD and converts it into FMN and AMP.

5.2.7 Bioinformatic analysis of *Bs*FADS and *Mj*FMNAT

The information about the conserved domains of *Bs*FADS and *Mj*FMNAT was extracted from the 'NCBI Conserved Domain Search' and 'MOTIF Search' using default parameters. The *Bs*FADS and *Mj*FMNAT amino acid sequences (Table 5.3) were searched against the database 'CDD v3.18-55570 PSSMs', a superset including NCBI-curated domains and data imported from Pfam, SMART, COG, PRK, and TIGFRFAM databases. The expected value threshold was chosen to be 0.010, and to filter the low-complexity sequences, 'composition-based statistics adjustment' was done using default parameters. The conserved domains of *Bs*FADS and *Mj*FMNAT were also searched against the Pfam database using the default cut-off score (E-value) of 1.0. Aditionally the homology model of *Mj*FMNAT was produced using Phyre2 server. The stereochemical quality of the modeled structure was verified using Ramachandran analysis tool of Procheck.[9]

Table 5.3: Amino acid sequences of *Bs*FADS and *Mj*FMNAT

>*Bs*FADS (UniProtKB - P54575)
MKTIHITHPHHLIKEEQAKSVMALGYFDGVHLGHQKVIGTAKQIAEEKGLTLAVMTFHPHPSHVLGRDKEPKDLITP LEDKINQIEQLGTEVLYVVKFNEVFASLSPKQFIDQYIIGLNVQHAVAGFDFTYGKYGKGTMKTMPDDLDGKAGCTM VEKLTEQDKKISSSYIRTALQNGDVELANVLLGQPYFIKGIVIHGDKRGRTIGFPTANVGLNNSYIVPPTGVYAVKA EVNGEVYNGVCNIGYKPTFYEKRPEQPSIEVNLFDFNQEVYGAAIKIEWYKRIRSERKFNGIKELTEQIEKDKQEAI RYFSNLRK
>*Mj*FMNAT (UniProtKB - Q58579)
MKKRVVTAGTFDILHPGHYEILKFAKSLGDELIVIVARDETVKKLKGRKPIIPEEQRREMVEALKPVDKA ILGSLKNKLEPILELKPDIIVLGPDQTTFDEETLKKELAKYNLYPEIVRFRGYKKCPFHSSFDIVKEIIR RFCNKEIKI

5.3 Results

5.3.1 Isolation of genomic DNA of *Bacillus subtilis*

The genomic DNA of *Bacillus subtilis* was extracted successfully from the cells by the method described in the Materials and methods section (Figure 5.4). Its concentration was measured on Nanodrop instrument, and was found to be 1330 ng/µL. The 260/280 ratio was 1.9, which is within the ideal range (1.7-1.9). The genomic DNA was used for the cloning of riboflavin kinase, *Bs*FADS and its mutants.
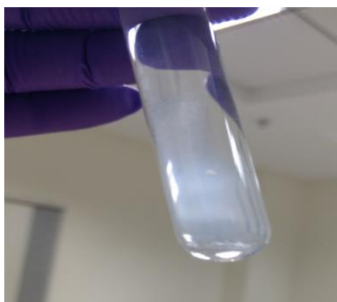
Figure 5.4: Isolation of the genomic DNA of *Bacillus subtilis.*

5.3.2 Purification of enzymes

The Ni-NTA affinity purification of the clarified lysates of *Bs*FADS and *Bs*FADS S166K mutant afforded the proteins of molecular weight 33 kDa as indicated by 12% SDS-PAGE gel electrophoresis (Figure 5.5 A, B). The molecular weight of *Mj*FMNAT showed on SDS-PAGE gel is around 20 kDa. These results are similar to the molecular weight obtained from the Protparam tool used for theoretical molecular weight (35 kDa for *Bs*FADS and 18 kDa for *Mj*FMNAT) determination.
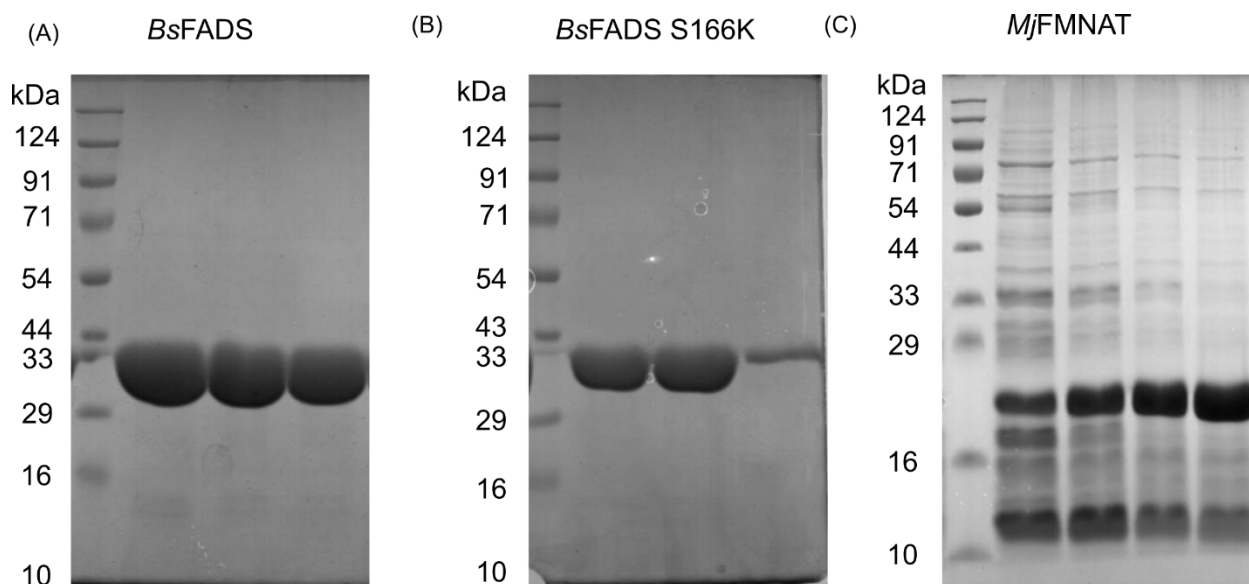
Figure 5.5: SDS-PAGE gel electrophoresis of the enzymes. (A) *Bs*FADS (B) *Bs*FADS S166K (C) *Mj*FMNAT

5.3.3 Activity assay of *Bs*FADS with different nucleotides under anaerobic and aerobic conditions

The activity assay of *Bs*FADS was done with oxidized and reduced forms of riboflavin against different NTPs (Figure 5.6). The result suggests that under oxidized conditions, it forms FAD with ATP. It can create FAD even in the absence of Mg^{2+} ion (TLC lane 8, oxidized riboflavin). On the other hand, under reduced conditions, this enzyme can form FAD with ATP and a small FCD with CTP (TLC lanes 2 and 3, reduced riboflavin). It can convert riboflavin into FAD even in the absence of dithionite and Mg^{2+} ions (TLC lanes 8 and 9, reduced riboflavin). Although with other NTPs it is unable to form FAD analogs, it can form FMN with CTP, GTP, and UTP under reduced conditions.

The mutant S166K of *Bs*FADS has not shown activity with NTPs, indicating that S166 is an essential catalytic residue.
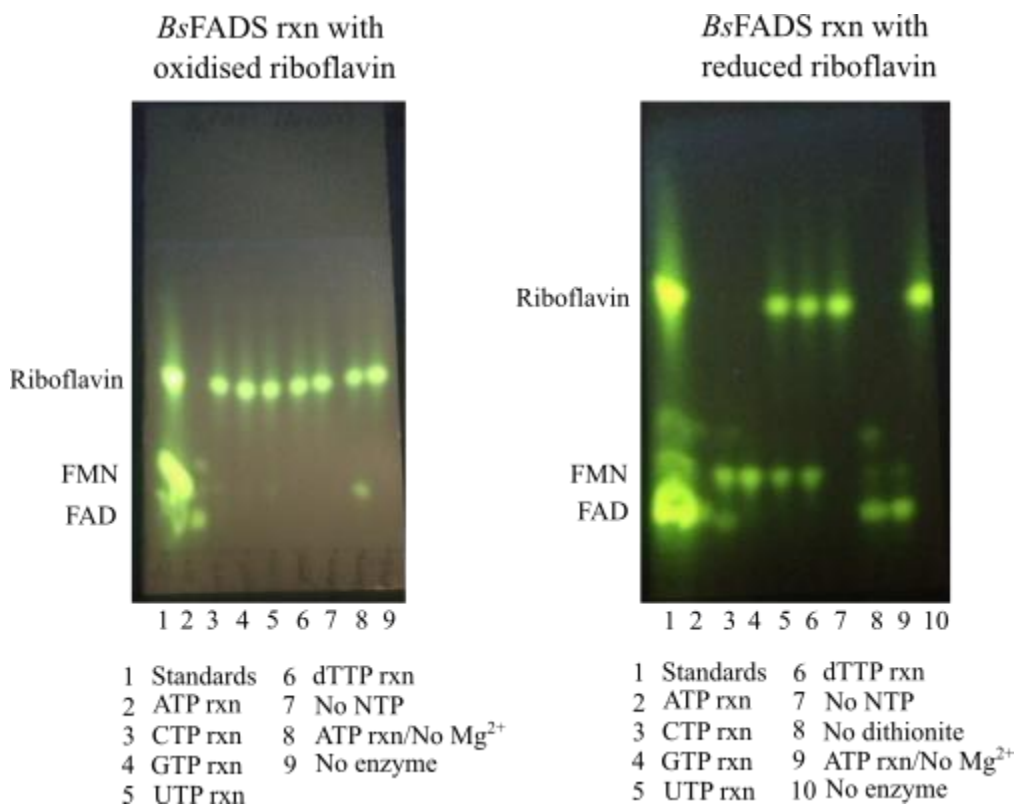
Figure 5.6: TLC showing the enzymatic reactions of *Bs*FADS against different NTPs and oxidized and reduced forms of riboflavin separately. (Experiments were done single times with same batch of proteins.)

5.3.4 The activity assay of *Mj*FMNAT with NTPs

The activity assay of *Mj*FMNAT had shown the formation of FAD with the substrates FMN and ATP. However, this enzyme also showed the formation of other FAD analogs when other nucleotide triphosphates were used in the reaction instead of ATP. The formation of these FAD analogs, *i. e.*, FCD, FGD, and FUD has been detected on TLC and HPLC. In figure 5.7A, the first lane of TLC shows riboflavin and FMN standards. Other lanes correspond to the *Mj*FMNAT reaction with different nucleotides, and demonstrate the formation of FAD and FAD analogs. The same result is confirmed using HPLC (Figure 5.7B), in which the peak at 33 min shows the formation of FAD and FCD. FGD also elutes at 32 min.
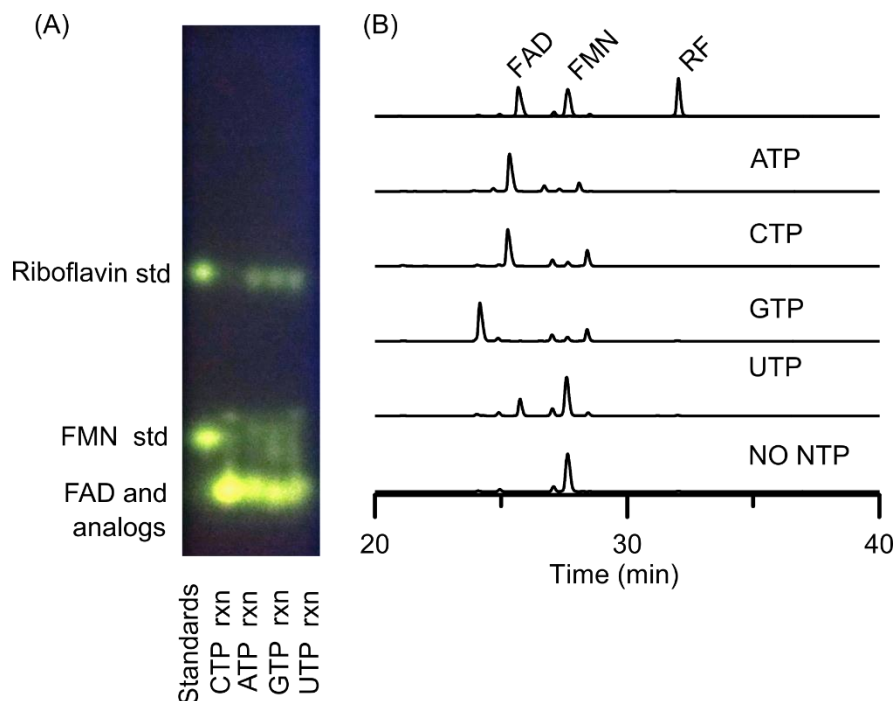
Figure 5.7: TLC and HPLC analysis of reaction products of *Mj*FMNAT reactions with different nucleotides (A) TLC showing the flavin standards (lane 1) and *Mj*FMNAT reaction products with different NTPs (B) HPLC chromatogram showing the *Mj*FMNAT reactions with ATP, CTP, GTP, and UTP. (Reactions were done in duplicates with different batches of proteins.)

5.3.5 The identity of FAD analogs formation was confirmed by phosphodiesterase assay

To further confirm that the *Mj*FMNAT products obtained with different nucleotides were indeed FAD analogs, we carried out a phosphodiesterase assay of each reaction product mentioned in the methods section. The snake venom phosphodiesterase enzyme can cleave the phosphodiester bond of FAD. If there were formation of FAD analogs, PDE could break the phosphodiester bond, and we obtain FMN and the corresponding nucleotide monophosphate. Hence, if there is a formation of FAD analog, then it can be confirmed by the appearance of FMN and nucleotide monophosphate in the reaction mixture when treated with phosphodiesterase enzyme. The results were as expected, and we got FMN when *Mj*FMNAT reaction product was treated with PDE and analyzed on HPLC (Figure 5.8).
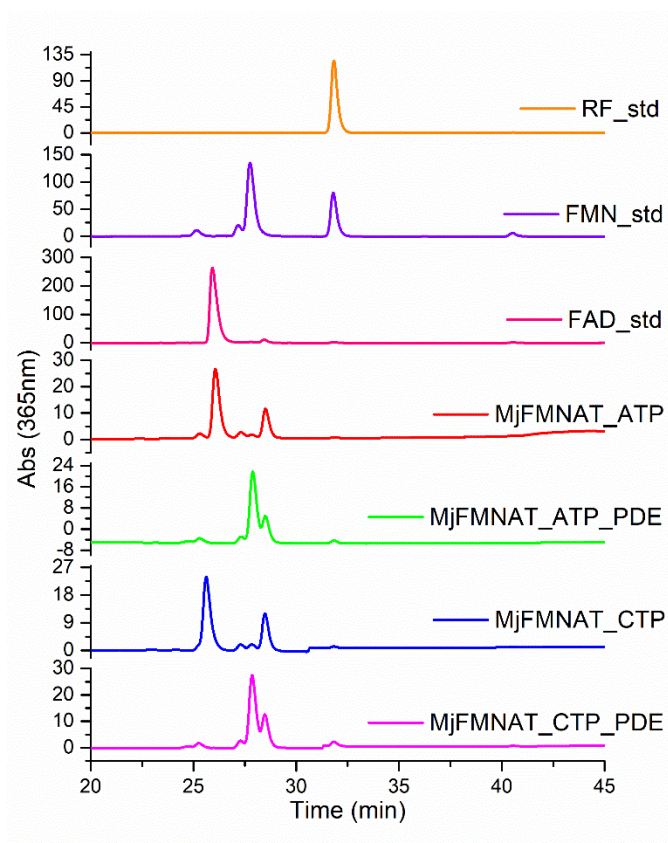
Figure 5.8: PDE treatment of FAD and FCD formed by *Mj*FMNAT enzyme (Done in duplicates with same batch of protein)

5.3.6 Purification of FAD analogs and enzymatic yield

After verifying FAD analogs formed by *Mj*FMNAT, we proceeded towards the purification of FAD analogs FGD and dFAD and evaluated the percentage yield (Figure 5.9). The final yields of the enzymatically synthesized FGD and dFAD were 51% and 45%, respectively.
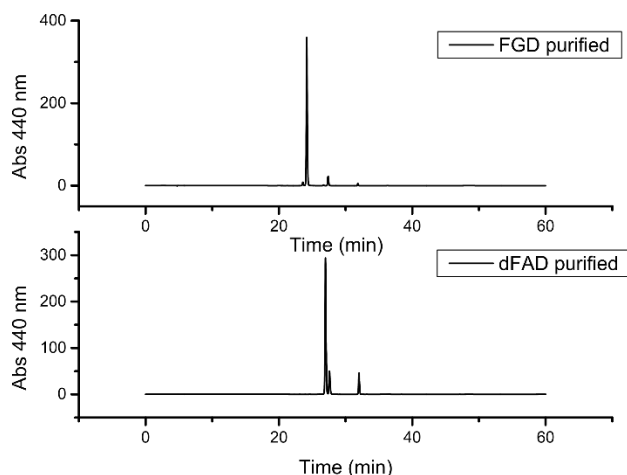
Figure 5.9: Chromatogram showing the purified FAD analogs FGD and dFAD. (Experiment was done single times)

5.3.7 Bioinformatic analyses to predict the possible binding sites for NTP and FMN in *Mj*FMNAT

The NCBI conserved domains search using 'CDD v3.18-55570 PSSMs' database also indicated that this enzyme belongs to the cytidylyltransferase domain superfamily and matches the enzymes 'glycerol-3-phosphate cytidylyltransferase', 'ethanolamine-phosphate cytidylyltransferase', 'phosphopantetheine adenylyltransferase' etc. as summarized in Table 5.4.

Table 5.4: Conserved domain analysis using NCBI CD search of *Mj*FMNAT sequence

	Name	Accession	Description	Interval	E-value
1	TagD	COG0615	Glycerol-3-phosphate cytidylyltransferase,	2-141	1.02E-58
2	cytidylyltransferase	cd02170	cytidylyltransferase;	3-141	8.03E-53
3	CTP_transf_like	pfam01467	Cytidylyltransferase-like;	6-136	1.09E-25
4	rfaE_dom_II	TIGR02199	rfaE bifunctional protein, domain II;	2-95	1.38E-25
5	RfaE_N	cd02172	N-terminal domain of RfaE;	2-95	2.64E-22
6	cyt_tran_rel	TIGR00125	cytidyltransferase-like domain;	4-70	1.84E-21
7	RfaE	COG2870	ADP-heptose synthase, bifunctional sugar kinase/adenylyltransferase	2-95	3.19E-21
8	PRK11316	PRK11316	bifunctional D-glycero-beta-D-manno-heptose-7-phosphate kinase/D-glycero-beta-D-manno-heptose 1-phosphate adenylyltransferase HldE	3-95	4.42E-18
9	G3P_Cytidylyltransferase	cd02171	glycerol-3-phosphate cytidylyltransferase;	3-95	1.56E-17
10	PTZ00308	PTZ00308	ethanolamine-phosphate cytidylyltransferase;	4-145	4.03E-14
11	g3p_cytidyltrns	TIGR01518	glycerol-3-phosphate cytidylyltransferase;	5-100	9.27E-14
12	PLN02406	PLN02406	ethanolamine-phosphate cytidylyltransferase	4-149	2.91E-13
13	ECT	cd02173	CTP:phosphoethanolamine cytidylyltransferase (ECT);	8-102	1.56E-12
14	CCT	cd02174	CTP:phosphocholine cytidylyltransferase;	4-73	1.16E-10

15	PLN02406	PLN02406	ethanolamine-phosphate cytidylyltransferase	5-74	2.40E-09
16	PTZ00308	PTZ00308	ethanolamine-phosphate cytidylyltransferase; Provisional	9-102	4.41E-09
17	PRK01170	PRK01170	bifunctional pantetheine-phosphate adenylyltransferase/NTP phosphatase;	3-147	2.48E-08
18	PRK00777	PRK00777	pantetheine-phosphate adenylyltransferase;	2-94	3.01E-08
19	CAB4	COG1019	Phosphopantetheine adenylyltransferase [Coenzyme transport and metabolism];	3-85	3.56E-08
20	CoaD	COG0669	Phosphopantetheine adenylyltransferase [Coenzyme transport and metabolism];	1-62	1.52E-06
21	PPAT_CoAS	cd02164	phosphopantetheine adenylyltransferase domain of eukaryotic and archaeal bifunctional enzymes;	4-62	8.26E-06
22	cytidylyltransferase_like	cd02039	Cytidylyltransferase-like domain; Cytidylyltransferase-like domain.	4-141	1.96E-05
23	coaD_prev_kdtB	TIGR01510	pantetheine-phosphate adenylyltransferase, bacterial;	4-62	3.50E-05
24	coaD	PRK00168	phosphopantetheine adenylyltransferase;	2-62	3.61E-05
25	PPAT	cd02163	Phosphopantetheine adenylyltransferase; Phosphopantetheine adenylyltransferase (PPAT).	4-62	1.48E-04
26	PLN02388	PLN02388	phosphopantetheine adenylyltransferase	5-104	3.39E-04
27	NMNAT_NadR	cd02167	Nicotinamide/nicotinate mononucleotide adenylyltransferase of bifunctional NadR-like proteins;	9-88	4.45E-04
28	PLN02413	PLN02413	choline-phosphate cytidylyltransferase	4-71	4.73E-04
29	nadD	PRK00071	nicotinate-nucleotide adenylyltransferase;	9-64	5.41E-04
30	NadR	COG1056	Nicotinamide mononucleotide adenylyltransferase [Coenzyme transport and metabolism];	1-63	7.21E-04
31	NadD	COG1057	Nicotinic acid mononucleotide adenylyltransferase [Coenzyme transport and metabolism];	1-118	1.45E-03
32	nt_trans	cd02156	nucleotidyl transferase superfamily; nt_trans (nucleotidyl transferase)	5-43	1.53E-03

Many of these hits cover the whole protein sequences, and therefore identification of specific residues responsible for the binding of NTP and FMN was difficult. The MOTIF-finder analysis had shown four specific results to which this enzyme belongs, as shown in table 2.

Table 5.5: Functional annotation of domains present in *Mj*FMNAT by MotifFinder tool

ID	Position	Sequence	SCORE	EVAL
PF01467, Cytidylyltransferase-like	7-136	TAGTFDILHPGHYEILKFAKSLGDELIVIVARDETVKKLKgrKPIIPE EQRREMVEALKPVDKAILGSLKNKLEPILELKPDIIVLGPDQTTFD EETLKKELAKYNLYPEIVRFRGYKKCPFHSSFDIVK	75.8	3.90E-21
PF06574, FAD synthetase	2-32	KKRVVTAGTFDILHPGHYEILKFAKSLGDEL	17.1	0.0039
PF08967, Domain of unknown function (DUF1884)	85-117	LKPDIIVLGPDQTTFDEETLKKELAKYNLYPEI	13.1	0.085
PF01497, Periplasmic binding protein	74-107	SLKNKLEPILELKPDIIVLGPDQTTFDEETLKKE	12	0.11

It has shown that the *Mj*FMNAT sequence contains cytidylyltransferase (PF01467) and FAD synthetase domains (PF06574). The sequence has a 15-HxGH-18 motif, which is conserved in many nucleotidyltransferase families. Both the histidine amino acids of the HxGH motif have been found to bind the phosphate moiety of NTP.

*Mj*FMNAT contains a unique flavin binding site, which has no similarities with any known flavin binding site. Attempts to identify this domain with bioinformatics studies were unsuccessful due to unavailability of the crystal structure. We found that the sequence ID PF08967 is a domain of unknown function (DUF1884). Since the 15-HxGH-18 motif which is a signature motif for nucleotidyltransferase family, is on the N-terminal end of the enzyme and the DUF1884 motif (85-117) is on C-terminal end, so we speculated that this might be the putative FMN- or FAD-binding domain of the *Mj*FMNAT. Thus, to know the identity of the DUF1884 domain, we did an NCBI BLAST search on the PDB database. Surprisingly we found one crystal structure (PDB: 2PK8) for DUF1884 for an uncharacterized protein PF0899 from *Pyrococcus furiosus.* The sequence alignment of the sequence of PF0899 with *Mj*FMNAT is given in Figure 5.10.
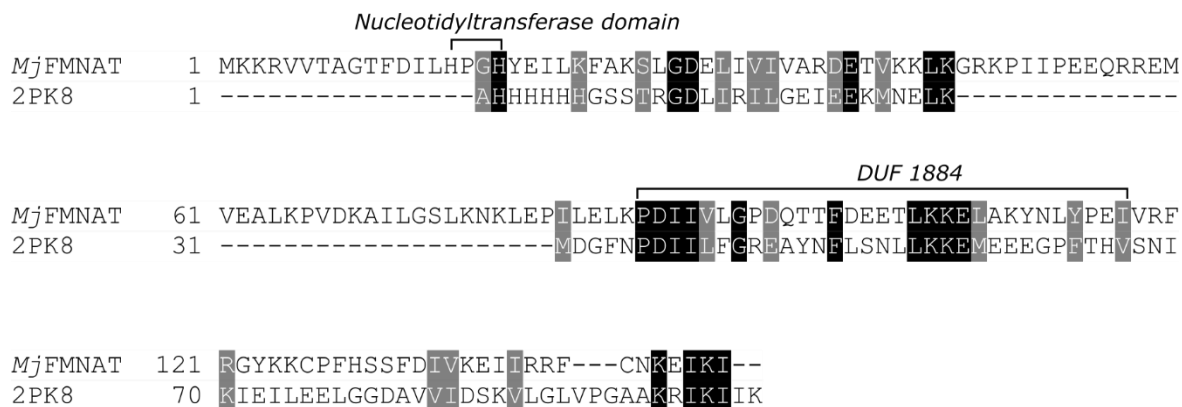
Figure 5.10: Sequence alignment of *Mj*FMNAt with PF0899 from *Pyrococcus furiosus* (PDB: 2PK8) Characterization of DUF1884 and its role in *Mj*FMNAT is in progress. At this stage, we are unable to know whether this is an FMN or FAD binding site.

Next, we conducted a homology modelling of *Mj*FMNAT using Phyre2 server (Figure 5.11 A). We found that the best structures that could be used as a template to model this structure belong to the cytidylyltransferase family. However, the stereochemical quality of the modelled structure is poor. The uncharacterized protein PF0899 (PDB ID: 2PK8) from *Pyrococcus*

furiosus was not used as a template by Phyre2. The Ramachandran plot of the model is showing many residues in disallowed regions (Figure 5.11 B). The Ramachandran plot statistics of the *Mj*FMNAT structure is given in table 5.6. Therefore, this poor quality modelled structure cannot be used for docking or other structural investigation.
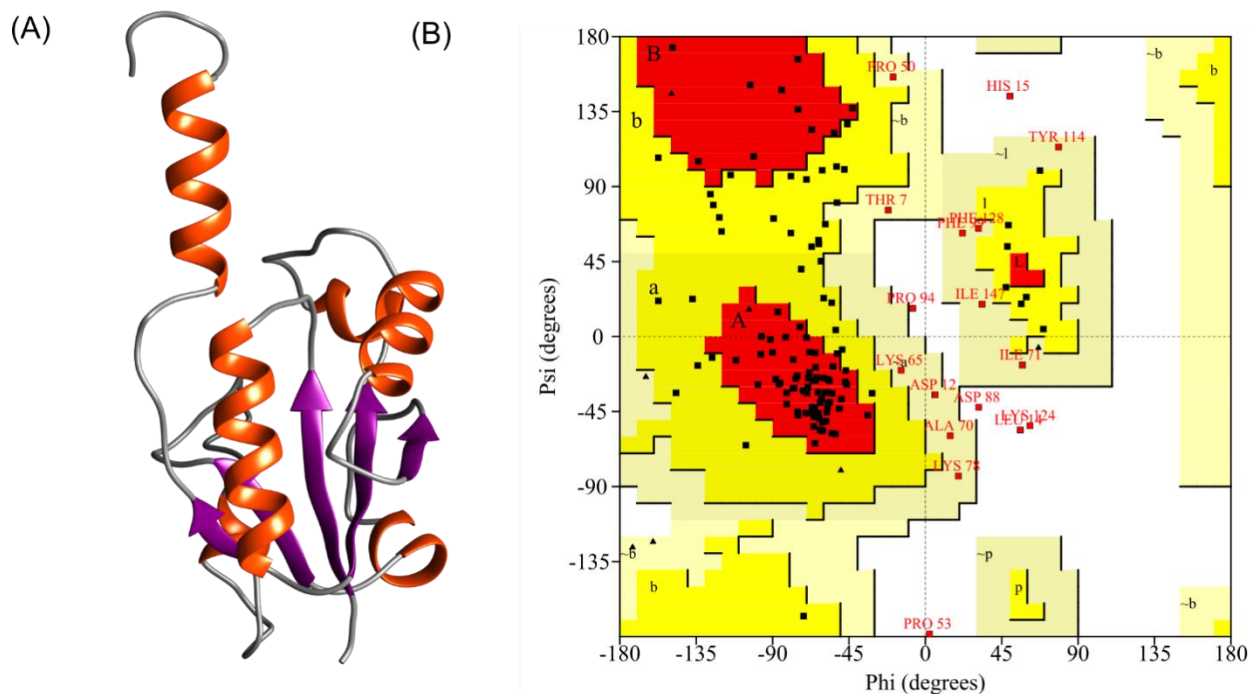
Figure 5.11: Homology modelling and structure assessment of the modelled structure. (A) the homology model of *Mj*FMNAT (B) Ramachandran plot of the *Mj*FMNAT modelled structure.

Table 5.6: Ramachandran plot statistics

Plot statistics	*Mj*FMNAT	
Residues in most favoured regions [A,B,L]	80	61.1%
Residues in additional allowed regions [a,b,l,p]	37	28.2%
Residues in generously allowed regions [~a,~b,~l,~p]	10	7.6%
Residues in disallowed regions	4	3.1%
Number of non-glycine and non-proline residues	131	100.00%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	7	
Number of proline residues	9	
Total number of residues	149	

5.3.8 Bioinformatic analyses to predict the possible binding sites in *Bs*FADS

The NCBI BLAST search of the *Bs*FADS protein sequence against the PDB database showed its structural folds similar to those present in the FADS from *Corynebacterium ammoniagenes* and *Thermotoga maritima*. The structural alignment of *Bs*FADS with *Corynebacterium ammoniagenes* indicated that this enzyme also has two catalytic domains. The N-terminal end is an FMN adenylyltransferase domain (residues 1-186), and the C-terminal end is a riboflavin kinase domain (residues 187-338). The schematic of *Bs*FADS domains is given in figure 5.11.
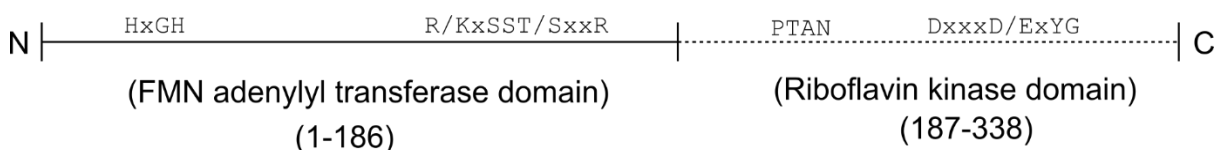

Figure 5.12: Domains and structural motifs in *Bs*FADS.

The NCBI conserved domain search have shown the annotation of domains given in table 5.6. From this table. It shown its similarity to riboflavin kinase and FAD synthetase enzymes.

Table 5.7: Conserved domain analysis using NCBI CD search of *Bs*FADS sequence

	Name	Accession	Description	Interval	E-value
1	ribF	TIGR00083	riboflavin kinase/FMN adenylyltransferase; multifunctional enzyme: riboflavin kinase (EC 2.7.1. ...	21-312	4.33E-155
2	PRK05627	PRK05627	bifunctional riboflavin kinase/FAD synthetase;	7-313	1.63E-142
3	RibF	COG0196	FAD synthase [Coenzyme transport and metabolism];	1-312	1.16E-131
4	FAD_synthetase_N	cd02064	FAD synthetase, N-terminal domain of the bifunctional enzyme; FAD synthetase_N. N-terminal ...	20-199	6.20E-80
5	FAD_syn	pfam06574	FAD synthetase; This family corresponds to the N terminal domain of the bifunctional enzyme ...	13-169	8.36E-67
6	Flavokinase	smart00904	Riboflavin kinase; Riboflavin is converted into catalytically active cofactors (FAD and FMN) ...	185-312	1.18E-64
7	Flavokinase	pfam01687	Riboflavin kinase; This family represents the C-terminal region of the bifunctional riboflavin ...	187-311	8.10E-59
8	PRK07143	PRK07143	hypothetical protein; Provisional	6-303	1.03E-18
9	PLN02940	PLN02940	riboflavin kinase	188-307	1.30E-08
10	cytidylyltransferase_like	cd02039	Cytidylyltransferase-like domain; Cytidylyltransferase-like domain. Many of these proteins are ...	21-173	2.41E-08
11	cyt_tran_rel	TIGR00125	cytidyltransferase-like domain; Protein families that contain at least one copy of this domain ...	21-88	1.81E-06
12	CTP_transf_like	pfam01467	Cytidylyltransferase-like; This family includes: Cholinephosphate cytidylyltransferase; ...	25-173	7.67E-03

5.4 Conclusion

We cloned, overexpressed, and purified the FAD synthesizing enzymes from *Bacillus subtilis* and *Methanocaldococcus jannaschii*. The *Bs*FADS was very specific for ATP for the nucleotidyltransferase reaction, and no FAD analogs were observed in this case. However, *Mj*FMNAT showed activity with CTP, GTP, UTP, and dATP, resulting in the formation of FAD analogs – FCD, FGD, FUD, and dFAD. The yields of FGD and dFAD after the purification were 51% and 45%, respectively.

References

(1)   Solovieva, I. M.; Tarasov, K. V.; Perumov, D. A. Main Physicochemical Features of Monofunctional Flavokinase from Bacillus Subtilis. *Biochem.* **2003**, *68* (2), 177–181. https://doi.org/10.1023/A:1022645327972.

(2)   Coquard, D.; Huecas, M.; Ott, M.; Van Dijl, J. M.; Van Loon, A. P. G. M.; Hohmann, H. P. Molecular Cloning and Characterisation of the RibC Gene from Bacillus Subtilis: A Point Mutation in RibC Results in Riboflavin Overproduction. *Mol. Gen. Genet.* **1997**, *254* (1), 81–84. https://doi.org/10.1007/s004380050393.

(3)   Kearney, E. B.; Goldenberg, J.; Lipsick, J.; Perl, M. Flavokinase and FAD Synthetase from Bacillus Subtilis Specific for Reduced Flavins. *J. Biol. Chem.* **1979**, *254* (19), 9551–9557.

(4)   Solovieva, I. M.; Kreneva, R. A.; Errais Lopes, L.; Perumov, D. A. The Riboflavin Kinase Encoding Gene RibR of Bacillus Subtilis Is a Part of a 10 Kb Operon, Which Is Negatively Regulated by the YrzC Gene Product. *FEMS Microbiol. Lett.* **2005**, *243* (1), 51–58. https://doi.org/10.1016/j.femsle.2004.11.038.

(5)   Gusarov, I. I.; Kreneva, R. A.; Rybak, K. V; Podcherniaev, D. A.; Iomantas, I. V; Kolibaba, L. G.; Polanuer, B. M.; Kozlov, I. I.; Perumov, D. A. Primary Structure and Functional Activity of the Bacillus Subtilis RibC Gene. *Mol. Biol. (Mosk). 31* (5), 820–825.

(6)   Mashhadi, Z.; Xu, H.; Grochowski, L. L.; White, R. H. Archaeal RibL: A New FAD Synthetase That Is Air Sensitive. *Biochemistry* **2010**, *49* (40), 8748–8755. https://doi.org/10.1021/bi100817q.

(7)   Manstein, D. J.; Pai, E. F. Purification and Characterization of FAD Synthetase from

Brevibacterium Ammoniagenes. *J. Biol. Chem.* **1986**, *261* (34), 16169–16173.

(8) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254. https://doi.org/10.1006/abio.1976.9999.

(9) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26* (2), 283–291. https://doi.org/10.1107/S0021889892009944.

Chapter 6

Perspective and future directions

6.1 Implication of nucleotide specificity and stability of flavin cofactor synthesizing enzymes

Our studies performed on various riboflavin kinases, FAD synthetases and FMN adenylyl transferases from different source organisms led to the following conclusions.

The comparative study of CTP-dependent archaeal riboflavin kinases and ATP-dependent non-archaeal riboflavin kinases has explored the molecular factors responsible for the nucleotide-binding. There is a plethora of structural and mechanistic information available for the ATP-dependent riboflavin kinases. The riboflavin kinases from *Homo sapiens* and *Schizosaccharomyces pombe* have been studied in great detail by different research groups.[1,2] However, there is limited information available for CTP-dependent riboflavin kinases.[3,4]

We explored the mechanism of nucleotide specificity in archaeal riboflavin kinase, which is a distinct sub-family of kinases. We successfully altered the nucleotide specificity of *Mj*RibK with the help of molecular-level information drawn on the basis of our investigation and made the enzyme accept multiple NTP substrates. However, our attempts to change the specificity of *Mj*RibK completely to one single NTP were unsuccessful. At the same time, the altering of specificity was at the cost of the reaction rate. Attempts can be made in the future to alter the specificity of *Mj*RibK or other archaeal and non-archaeal RFKs. These experiments will have a valuable contribution to the understanding of the nucleobase specificity of enzymes.

(a) An enzyme is a phenomenal biological machine which has evolved to contain an ideal balance of interactions among various amino acid residues, cofactors, and solvents. These interactions allow an enzyme to fold perfectly and be catalytically active. Mutations occasionally lead to an imbalance in this system and can leave an enzyme misfolded and/or catalytically inactive. Thus, while engineering a protein for change in its activity, one also has to take account of the stability of the final engineered enzyme.

*The diminished activity of the Mj*RibK *mutants designed for the altered nucleotide specificity motivated us to explore the molecular determinants of enzyme stability in*

archaeal riboflavin kinases (RFKs). The extensive bioinformatics and structural analyses done with mesophilic RFK from *Methanococcus maripaludis* resulted in the design of its thermostable homologs without loss in activity. With this, we learned that enzyme rigidity is responsible for the thermostability of an enzyme. On the other hand, the flexibility of the active site is also critical for the activity of the enzyme. *To make a stable but catalytically robust enzyme, we rigidify the enzyme, but we should also consider the flexibility of the enzyme's catalytic site. Such design principles for the stability of enzymes are of prime importance in protein engineering and have biotechnological and industrial advantages.*

(b) The nucleotide base specificity is an unexplored area of research. The pattern of nucleobase recognition differs in different enzymes. Although factors like sequence conservation, substrate complementarity, active site architecture, reaction conditions play a critical role in determining the specificity, none of them are yet adequately studied to establish a generalized theory for nucleobase recognition. *Understanding nucleobase recognition could have an enormous consequence on enzyme redesign, chemoenzymatic synthesis, and cofactor engineering.*

Our study of nucleotide specificity and utilization in *Mj*FMNAT and *Bs*FADS (Chapter 5) is a foundation of these systems to decipher the molecular determinants of nucleotide specificity and its applications to bio-orthogonal systems. *The in vitro enzymatic synthesis of FAD cofactor analogs like FCG, FGD and dFAD by MjFMNAT was our primary achievement towards artificial cofactor analogs. The heterologous expression of MjFMNAT has shown the in vivo synthesis of these biorthogonal cofactors in E. coli (research conducted jointly with Ateek Shah in the Hazra lab).*

In summary, we explored the critical role of the small helix present on the nucleobase surrounding loop in archaeal CTP-dependent riboflavin kinase. The disruption of this helix made this enzyme non-specific, and it was able to utilize ATP and CTP both as phosphoryl donors. We were also successful in determining the crucial role of salt bridges and proline residues to confer stability to archaeal

riboflavin kinases. At last, by manipulating the phosphate donor of the *Mj*FMNAT enzyme, we synthesized and purified unnatural FAD cofactor analogs.

We also wanted to explore the nucleotide base specificity of RFK, FMNAT, and FADS from different organisms to generalize the principles of nucleobase recognition in the enzymes involved in flavin cofactor biosynthesis. These results could have a broader impact on the specific roles of nucleotide triphosphates in different metabolic pathways. Some of this work has been achieved, and is noted in detail here. We cloned, purified, and tested the activity of some enzyme homologs that show activity with different NTPs (Figure 6.1, Table 1). We found the characteristic differences among these enzymes in their mechanism of working. These differences are important towards our approach of understanding nucleobase specificity, in which we compare and correlate the subtle molecular level changes among the enzymes.
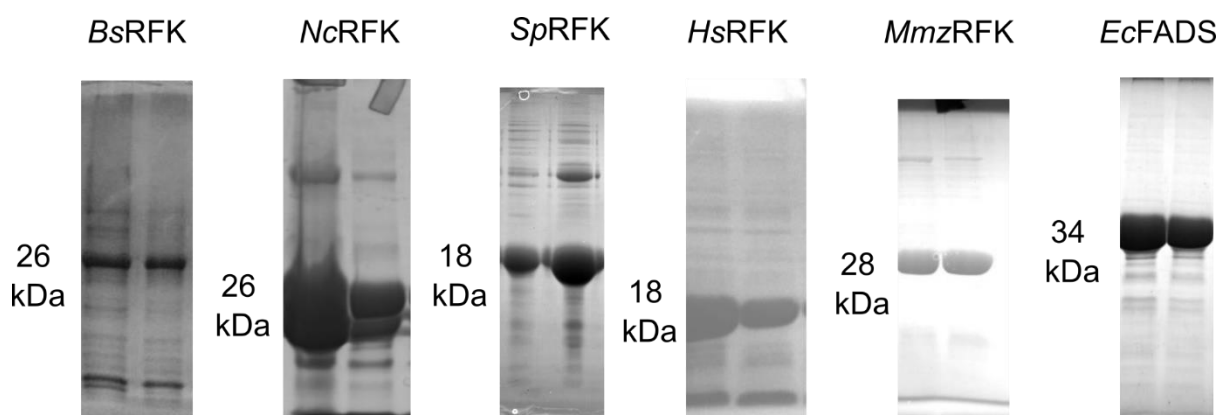

Figure 6.1: SDS-PAGE gel of *Bs*RFK, *Nc*RFK, *Sp*RFK, *Hs*RFK, *Mmz*RFK and *Ec*FADS

Table 6.1: RFK/RibK, FADS and FMNAT from different organisms cloned, purified and tested for activity with NTPs

	RFK									
#					**Activity tested with**					
	Organism	Specificity	Cloned	Purified	A	C	G	U	dA	dT
1	*Bacillus subtilis*	ATP	✓	✓	✓	✓	✓	✓		
2	*Neurospora crassa*	ATP	✓	✓	✓	✓	✓	✓		

#	Organism	Donor								
3	*Sacchromyces cerevisiae*	ATP	✓							
4	*Schizosaccharomyces pombe*	ATP	✓	✓	✓	✓	✓	✓		
5	*Homo sapiens*	ATP		✓	✓	✓	✓	✓		
6	*Methanocaldococcus jannaschii*[&]	CTP	✓	✓	✓	✓	✓	✓		
7	*Methanococcus maripaludis*[%]	CTP	✓	✓	✓	✓	✓	✓		
8	*Methanosarcina mazei*	CTP	✓	✓	✓	✓	✓	✓		
	FADS/ FMNAT									
1	*Escherichia coli*	ATP	✓	✓	✓	✓	✓	✓	✓	✓
2	*Saccharomyces cerevisiae*	ATP	✓							
3	*Schizosaccharomyces pombe*	ATP	✓							
4	*Bacillus subtilis*[$]	ATP	✓	✓	✓	✓	✓	✓		✓
5	*Methanocaldococcus jannaschii*[*,$]	ATP	✓	✓	✓	✓	✓	✓		

[&] Chapter 3, [%] Chapter 4, [$] Chapter 5, [*] FMNAT

The experimental details for the tabulated results are as follows:

The ATP-dependent riboflavin kinase from *Bacillus subtilis* (*Bs*RFK) requires reduced flavin for its activity. We tested the activity of this enzyme with 4 mM and 10 mM NTPs. The results indicated that ATP is the preferred phosphate donor for the formation of FMN. However, a trace amount of FMN was observed when CTP was used as a phosphate donor (Figure 6.2).
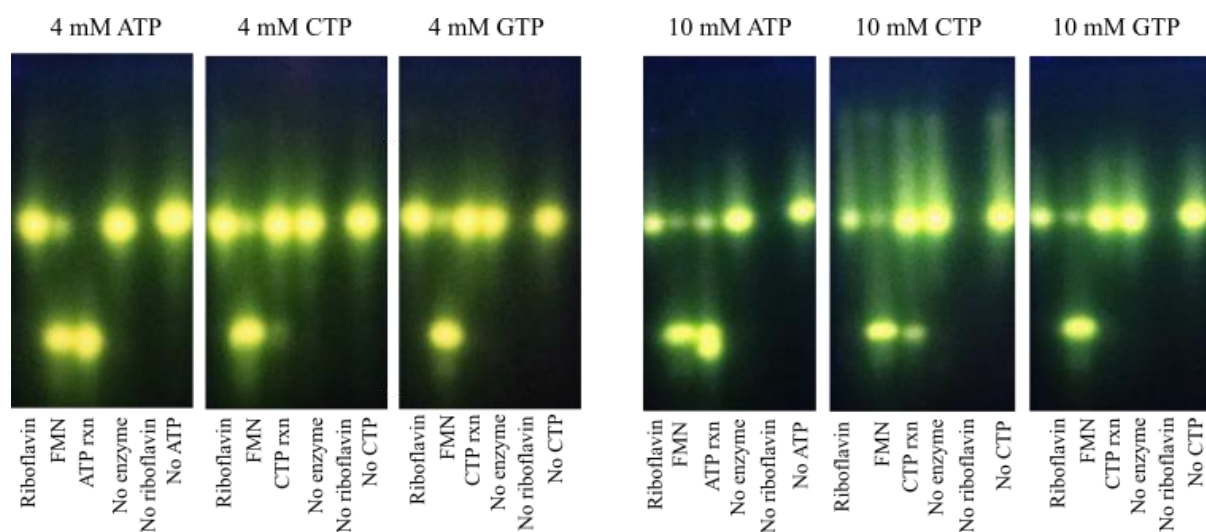
Figure 6.2: The activity of riboflavin kinase from *Bacillus subtilis* was checked with different NTPs at 4 mM and 10 mM concentrations. It shows full conversion of riboflavin into FMN with ATP and slight activity with CTP also.

The RFK from *Neurospora crassa*, *Homo sapiens*, and *Schizosaccharomyces pombe* have shown promiscuity in NTP utilization (Figure 6.3)
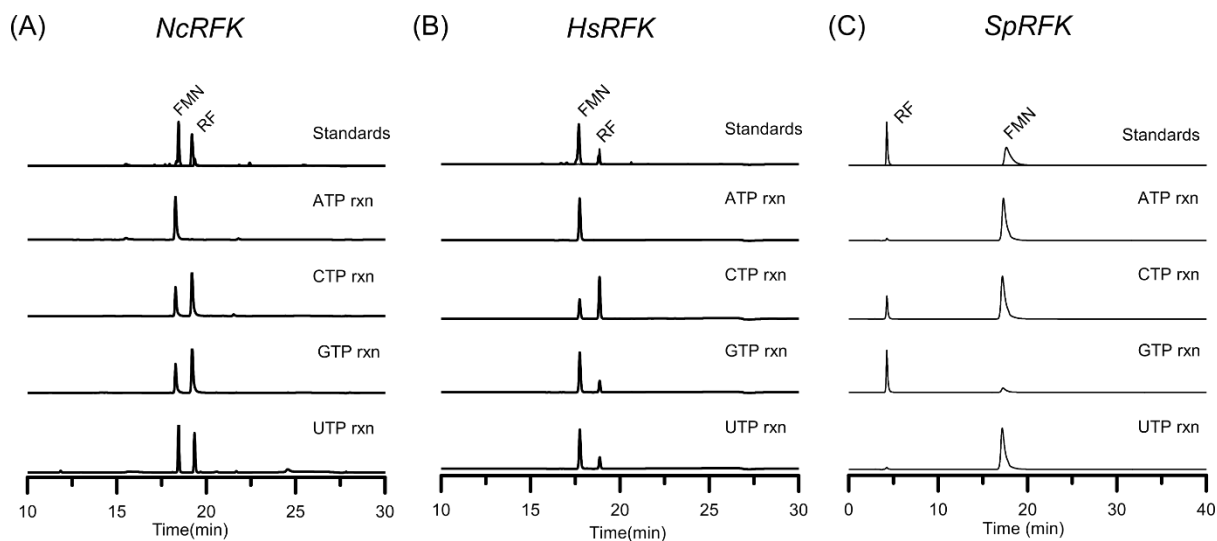

Figure 6.3: The HPLC chromatogram of the riboflavin kinase activity of *Nc*RFK, *Hs*RFK, and *Sp*RFK with different NTPs. All of these enzymes are ATP-specific but show promiscuity in substrate utilization. *Nc*RFK and *Hs*RFK give partial conversion with CTP, GTP, and UTP. *Sp*RFK provides partial conversion with CTP and UTP but poor conversion with GTP. (**Note 1**: *Nc*: *Neurospora crassa*; *Hs*:

105

Homo sapiens; *Sp*: *Schizosaccharomyces pombe;* **Note 2**: **A** and **B** are reverse-phase HPLC chromatogram while **C** is anion-exchange HPLC chromatogram)

Apart from *Methanocaldococcus jannaschii* and *Methanococcus maripaludis*, we cloned, purified, and tested the activity of another CTP-dependent riboflavin kinase from the archaeon *Methanosarcina mazei*. The purified enzyme needs optimized assay conditions as it was not showing the complete conversion of riboflavin into FMN with CTP (Figure 6.4 A). It does not show activity with ATP, GTP, and UTP. However, the assay done with crude enzyme from lysate has shown full activity with GTP and partial activity with ATP, and UTP activity (Figure 6.4 B).
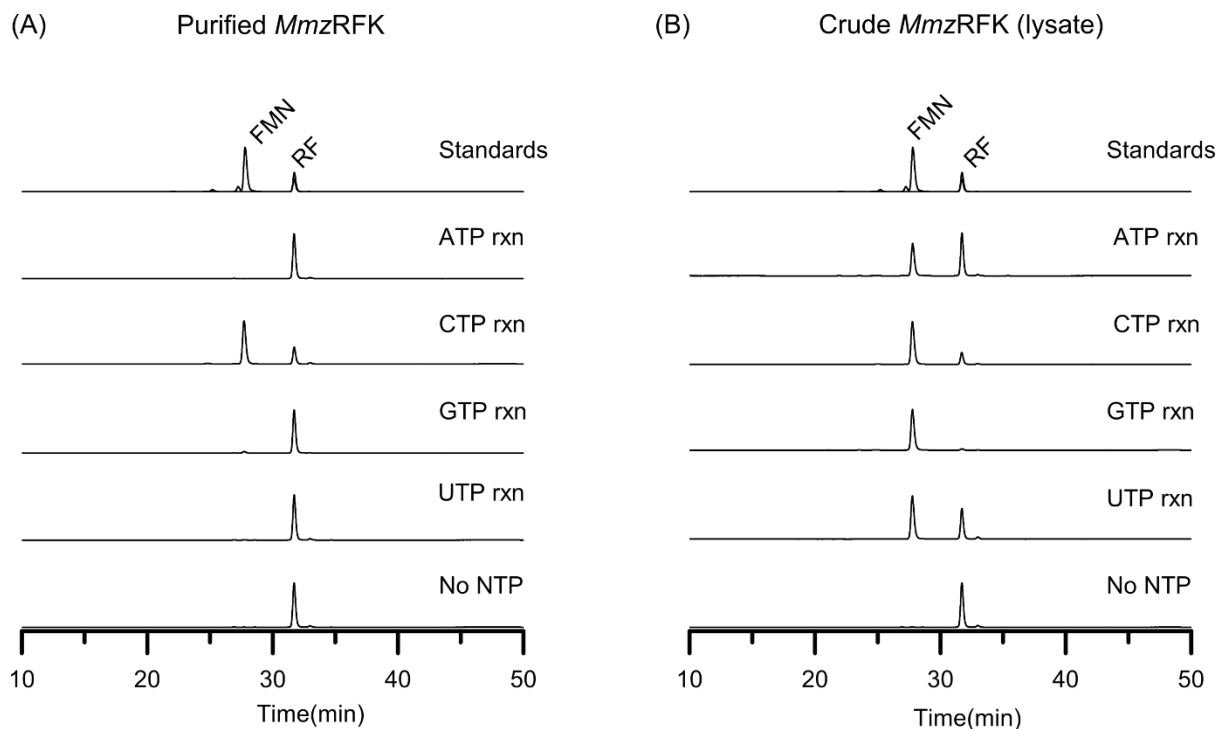

Figure 6.4: Enzymatic reactions of purified and crude *Mmz*RFK with different NTPs. (A) the purified enzyme is showing partial activity with CTP and no activity with other nucleotides. (B) the crude enzyme from the lysate can convert riboflavin into FMN with CTP, ATP, UTP, and GTP. The conversion with GTP is complete.

The promiscuous nature of *Mj*FMNAT to form FAD analogs is described in chapter 5. We also mentioned the nucleotide utilization by *Bs*FADS under aerobic and anaerobic conditions in this chapter. Another FAD synthetase that we tested for the enzymatic synthesis of FAD

analogs is from *Escherichia coli*. The FAD synthetase from *Escherichia coli* (*Ec*FADS) is a bifunctional enzyme. It gives full conversion of riboflavin into FAD with ATP and also shows partial conversion with CTP to form FCD (Figure 6.5).
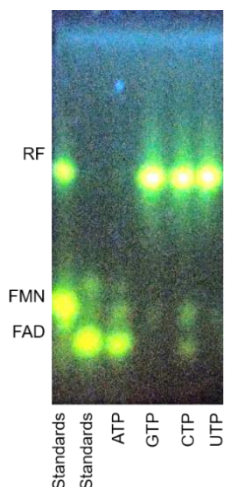

Figure 6.5: Activity assay of *Ec*FADS with ATP, GTP, CTP and UTP

6.2 Future directions

Riboflavin is the precursor of flavin cofactors FMN and FAD, which are involved in a number of central metabolic pathways such as the TCA cycle (Kreb's cycle), nucleotide biosynthesis, amino acid catabolism, vitamins, and cofactor biosynthesis. FAD and FMN are cofactors for several proteins collectively known as flavoproteins. Bacteria, fungi, plants can produce riboflavin, but other eukaryotes have lost their ability to synthesize it. The biosynthesis of riboflavin in bacteria and eukaryotes has been well established. However, little is known about it in plants and archaea. Riboflavin deficiency can affect multiple metabolic pathways. Riboflavin deficiency or impaired flavoenzymes can lead to diseases like cancer, migraine, worsened vision, skin lesions, etc. FAD synthesizing enzymes and flavoproteins are also an attractive target for drug design against bacterial antibiotic resistance.

We wish to explore the nucleotide specificity of FAD synthetases and FMN adenylyltransferase enzymes from various organisms to make unnatural FAD analogs (Figure 6.6). With the knowledge of molecular determinants of nucleotide specificity, we can perform

protein engineering and domain swap experiments to create artificial enzymes having altered specificities. This can improve the synthetic yield of these unnatural cofactors drastically.
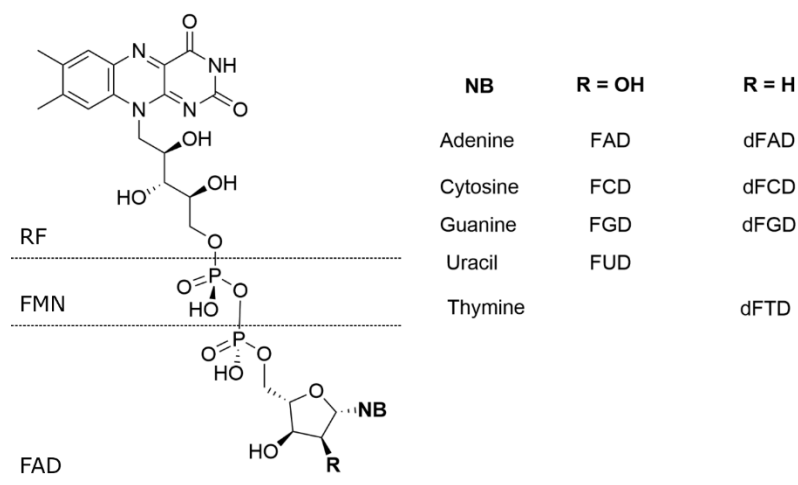

NB	R = OH	R = H
Adenine	FAD	dFAD
Cytosine	FCD	dFCD
Guanine	FGD	dFGD
Uracil	FUD	
Thymine		dFTD

Figure 6.6: Chemical structure depicting the riboflavin (RF), FMN and FAD. NB is the abbreviation for nucleobase. If R = OH, it is a ribonucleic derivative. If R= H, it is a deoxyribonucleic derivative.

Unnatural FAD analogs may have applications in bioremediation and synthetic biology. Using principles that we have learned of nucleobase specificity for FAD from the FAD synthesis enzymes, we can design FAD-utilizing enzymes which can use the modified FAD cofactor analogs. These enzymes could be tested *in vitro*, and then, using homologous recombination, we could bring these modifications into a living cell, thus engineering a preference for unnatural FAD cofactors in an organism. Such applications of our research will have applications in bioremediation and synthetic biology.

References

(1)   Frago, S.; Velázquez-Campoy, A.; Medina, M. The Puzzle of Ligand Binding to Corynebacterium Ammoniagenes FAD Synthetase. *J. Biol. Chem.* **2009**, *284* (11), 6610–6619. https://doi.org/10.1074/jbc.M808142200.

(2)   Bauer, S.; Kemter, K.; Bacher, A.; Huber, R.; Fischer, M.; Steinbacher, S. Crystal Structure of Schizosaccharomyces Pombe Riboflavin Kinase Reveals a Novel ATP and Riboflavin-Binding Fold. *J. Mol. Biol.* **2003**, *326* (5), 1463–1473. https://doi.org/10.1016/S0022-2836(03)00059-7.

(3)   Ammelburg, M.; Hartmann, M. D.; Djuranovic, S.; Alva, V.; Koretke, K. K.; Martin, J.; Sauer, G.; Truffault, V.; Zeth, K.; Lupas, A. N.; Coles, M. A CTP-Dependent Archaeal Riboflavin Kinase Forms a Bridge in the Evolution of Cradle-Loop Barrels. *Structure* **2007**, *15* (12), 1577–1590. https://doi.org/10.1016/j.str.2007.09.027.

(4)   Mashhadi, Z.; Zhang, H.; Xu, H.; White, R. H. Identification and Characterization of an Archaeon-Specific Riboflavin Kinase. *J. Bacteriol.* **2008**, *190* (7), 2615–2618. https://doi.org/10.1128/JB.01900-07.