Redesigning Riboflavin Kinases from Bacteria and Archaea to Alter Nucleotide Specificity

A thesis submitted to

Indian Institute of Science Education and Research, Pune

In partial fulfillment of the requirements for the

BS-MS Dual Degree Programme

By

S Rohan

Indian Institute of Science Education and Research
Dr. Homi Bhabha Road,
Pashan, Pune — 411008. INDIA.

April, 2023.

Supervisor: Dr. Amrita B. Hazra

Expert: Dr. Santosh Kumar Jha

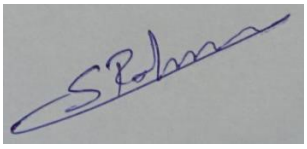
Certificate

This is to certify that this dissertation entitled “Redesigning Riboflavin Kinases from Bacteria and Archaea to Alter Nucleotide Specificity” towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by S Rohan at IISER Pune under the supervision of Dr Amrita B. Hazra,Associate Professor, Department of Chemistry, IISER Pune during the academic year 2022-2023.

Signature of Supervisor

Declaration

I hereby declare that the matter embodied in the report entitled “Redesigning Riboflavin Kinases from Bacteria and Archaea to Alter Nucleotide Specificity” are the results of the work carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research Pune under the supervision of Dr. Amrita B Hazra and the same has not been submitted elsewhere for any other degree.



Signature of Student

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Amrita Hazra, for her mentorship, guidance and support throughout the year. She has always been enthusiastic about troubleshooting experiments, interpreting results and discussing new ideas. It has been a pleasure working on this project under her supervision. I would like to thank my expert, Dr. Santosh Jha for interesting insights and advice on the project.

I am thankful to all my fellow lab members (both past and current) for their unwavering support. Dr. Yashwant Kumar has had a significant role to play in my thesis since the initial part of my project is an extension of the work done by him in his PhD. He has also been instrumental in teaching me laboratory techniques at the initial stage to help me get started with my work. I would also like to thank Ateek Shah for constant academic guidance throughout the year. I have always been impressed by the enthusiasm and flair for science that Yamini and Rupali possess. These two have served as huge sources of motivation over the course of my thesis. I would like to acknowledge the academic and moral support provided by Aditya Bhattacharyya and Aniket Vartak. I am extremely thankful to Riya, Vibhor, Vruta, Shweta, Neel, Rajdip, Anjali and Nikita for being co-operative and helpful colleagues.

I am grateful to my parents (*Amma* and *Appa*) for being perpetual pillars of support at all times. I value and cherish the time that I have spent with Anirudh Rachuri, Prem Agarwal, Rajat Sharma, Shivam Tikoo, Arjun Murlidhar, Karthik Vasisht, Harshit Pateria, Yuv Agarwal and all my other friends. I hope that I stay in touch with these amazing people in the future.

Contents

List of Figures

List of Tables

Abstract:

Riboflavin (also known as vitamin B2) is a water-soluble vitamin that is acquired from dietary sources and converted into FMN and FAD enzymatically. Various flavoenzymes use these two molecules as co-enzymes/co-factors for catalyzing oxidation/reduction reactions. The conversion of riboflavin to FMN is achieved by means of an enzyme called riboflavin kinase that phosphorylates riboflavin using an NTP source.

The first section of my thesis dealt with engineering an archaeal CTP-dependent riboflavin kinase from *Methanococcus maripaludis* (*Mmp*RibK) for the investigation of the strict nucleobase specificity of this enzyme. The basis of this research stems from previous attempts at engineering another archaeal CTP-dependent riboflavin kinase from *Methanocaldococcus jannaschii* (*Mj*RibK) in order to confer reactivity with ATP.

The direction of my thesis moved on to the investigation of a certain conserved part of the nucleobase-surrounding loop in eukaryotic riboflavin kinases. This investigation was performed by characterization of mutants of the human riboflavin kinase (*Hs*RibK) with modifications in this conserved region.

The third section of my thesis involved attempts at the characterization of riboflavin kinase and FAD synthetase of *Schizosaccharomyces pombe* (*Sp*RibK and *Sp*FADS) followed by an analysis of organisms that contain both the monofunctional riboflavin kinase and the bifunctional FAD synthetase.

1. Introduction:

1.1. What are nucleoside triphosphates/ deoxynucleoside triphosphates?

Nucleoside triphosphates (also known as NTP's) are composed of three components: a nitrogenous (purine or pyrimidine) base, a phosphate group and a ribose ring. Deoxynucleoside triphosphates (also known as dNTP's) are composed of three components: a nitrogenous (purine or pyrimidine) base, a phosphate group and a deoxyribose ring.

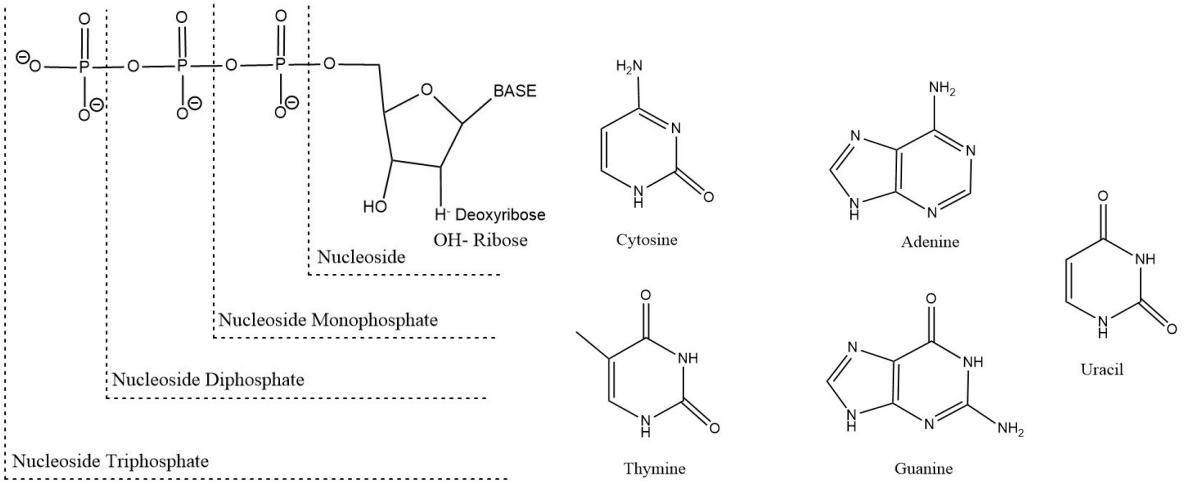


Figure 1: A structural representation of NTP's and Nucleobases

The figures above show the structure of nucleoside triphosphates along with the purine/pyrimidine bases that are present in it. DNA (deoxyribonucleic acid) has the deoxyribose sugar along with adenine, guanine, cytosine and thymine as the nitrogenous bases. RNA (ribonucleic acid) has the ribose sugar along with adenine, guanine, cytosine and uracil as the nitrogenous bases.

1.2. Physiological Role of NTP's/dNTP's:

ATP hydrolysis is a major process that releases energy for many cellular functions.

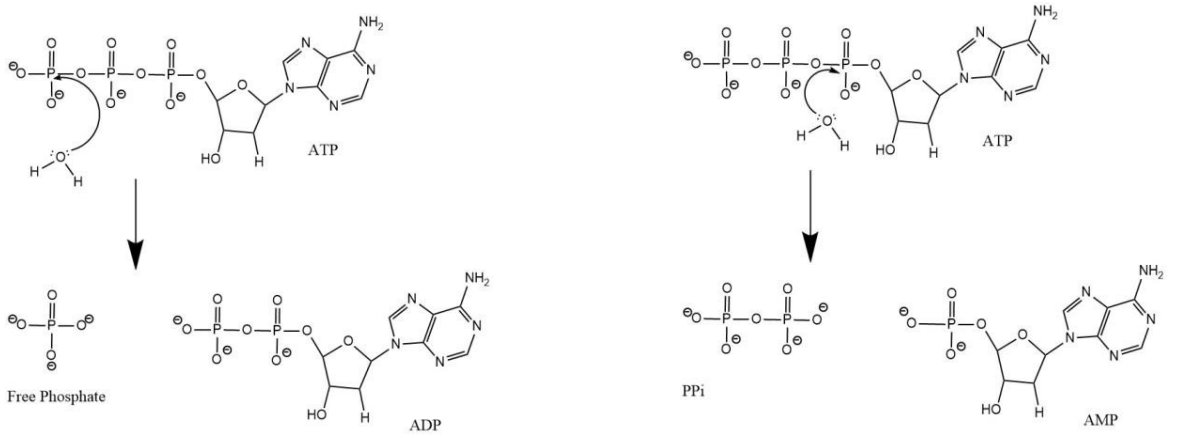


Figure 2: Mechanism of ATP Hydrolysis

As shown in the figure above, a nucleophilic attack on the **γ**-phosphate of ATP by a water molecule releases free energy (Δ*G*°) of −30.5 kJ/mol whereas a nucleophilic attack on the α-phosphate releases free energy (Δ*G*°) of −45.6 kJ/mol. Note that these values of Δ*G*° are applicable at pH7 at standard state concentrations of 1 M.

ATP is often used as the phosphorylation source for a variety of substrates. Protein phosphorylation is a very important mechanism used for controlling many processes and regulating downstream signaling. The phosphorylation (or dephosphorylation) of proteins by kinases/phosphatases modulates the activity of signaling molecules leading to signal amplification. For example: A specific class of serine/threonine protein kinases, the mitogen-activated protein kinases (MAP kinases) play a central role in the transduction of various extra and intracellular signals and are conserved throughout eukaryotes. These generally function via a cascade of networks that involves MAP kinase (MAPK), MAPK kinase (MAPKK) and MAPKK kinase (MAPKKK). This in turn affects different signal transduction pathways such as osmoregulation, cell growth and differentiation.[1] ATP can also act as a neurotransmitter in the central and peripheral nervous systems and is also involved in transmission of the sensation of pain.[2] CTP is involved in various pathways

(cytidine diphosphate (CDP)–diacylglycerol (DAG), CDP–choline, and CDP–ethanolamine pathways) that are responsible for the synthesis of phosphatidate (precursor of all phospholipids).[3] Many GTP-binding proteins (G proteins) are involved in transmembrane signaling processes. Certain metabolic enzymes such as succinate thiokinase and phosphoenolpyruvate carboxykinase also interact with GTP.[4] Uracil nucleotides such as UDP help in activating sugars for protein glycosylation reactions. Intracellular UTP levels directly regulate phospholipid synthesis thereby controlling characteristics of the cell membrane.[5]

1.3. Characteristics of Protein Kinases:

All typical protein kinases (TPKs) have a common catalytic core consisting of a smaller, mostly **β**-sheet, N-terminal subdomain and a larger, mostly α-helical, C-terminal subdomain.[6] ATP is bound in a deep cleft between the two regions. The phosphate binding loop, or P loop, contains a conserved glycine-rich sequence motif (GXGXφG) where φ is typically tyrosine or phenylalanine. The glycine residues allow the loop to coordinate the phosphates of ATP via backbone interactions. The conserved aromatic side chain caps the site of phosphate transfer. The glycine residues contained in the P loop make it very flexible in the absence of ATP.[7] Several atypical protein kinases are also known, which lack sequence similarity to TPKs, although some have structural similarity to TPK's. Protein kinases are classified into multiple families, most of which are conserved throughout metazoans. Many of these are also conserved in yeasts. The worm kinome has large expansions of several families. The fly also appears to have lost some kinase families but is, overall, a closer model to the human kinome.[8]

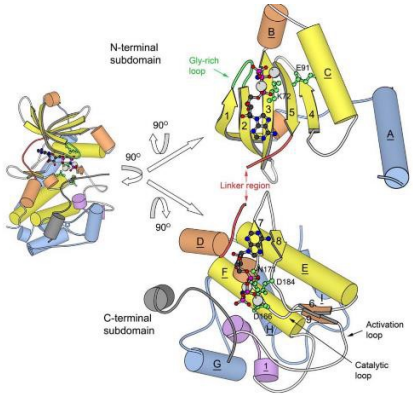


Figure 3: N-terminal & C-terminal Subdomains of Protein Kinases[6]

1.3.1. Nucleotide Specificity in Enzymes:

Even though ATP is usually thought of as the NTP source for enzymes in cellular systems, there are examples of NTP-utilizing enzymes that differ in the choice of NTP. Phosphopantothenoylcysteine synthetase from *Escherichia coli* is CTP specific whereas the human homolog is ATP specific.[9,10] Similar is the case for succinyl CoA synthetase which is GTP specific in *Sus scrofa* and ATP specific in *Blastocystis sp.*[11,12] Such nucleotide specificity is also true for kinases. Riboflavin kinase of *Rhizopus oryzae* and phosphoglycerate kinase from *Entamoeba histolytica* are GTP dependent.[13,14] It has also been shown that GTP serves as an excellent phosphate donor for autophosphorylation of Protein kinase Cδ (PKCδ) (obtained from porcine spleen) and it outranks ATP in this respect.[15] Studies on Aminoglycoside 2-Phosphotransferase IVa have found that even though individual residues such as Phe-95 play a major role in dictating nucleotide specificity, it is probable that the composite of residues around the nucleotide-binding site collectively forms a space that is adapted for the specific substrate preference. Factors beyond individual amino acids such as the inability to form an ordered solvent network and the lack of sufficient hydrogen-bonding interactions due to an altered conformation of the linker loop prevent certain phosphotransferases from binding GTP.[16] Studies on engineering protein kinases have been performed by Kevan Shokat's group. [17] Their goal was to produce a mutant protein kinase that accepts an N^6-substituted ATP with high catalytic efficiency. This was achieved by the mutation of valine and isoleucine residues in the ATP-binding pockets to alanine. The rationale behind this was that a smaller sidechain would create an additional pocket for the entry of N^6-substituted ATP. The double mutant V323A/I338A was able to utilize N^6-(cyclopentyl) ATP. This double mutant also had a lower affinity for ATP. Other enzyme engineering examples include altering the nucleotide specificity of *Blastocystis sp.* succinyl CoA synthetase from ATP to GTP based on charged gatekeeper residues of the active site.[18]

1.4. Co-enzymes/Co-factors containing adenine moiety:

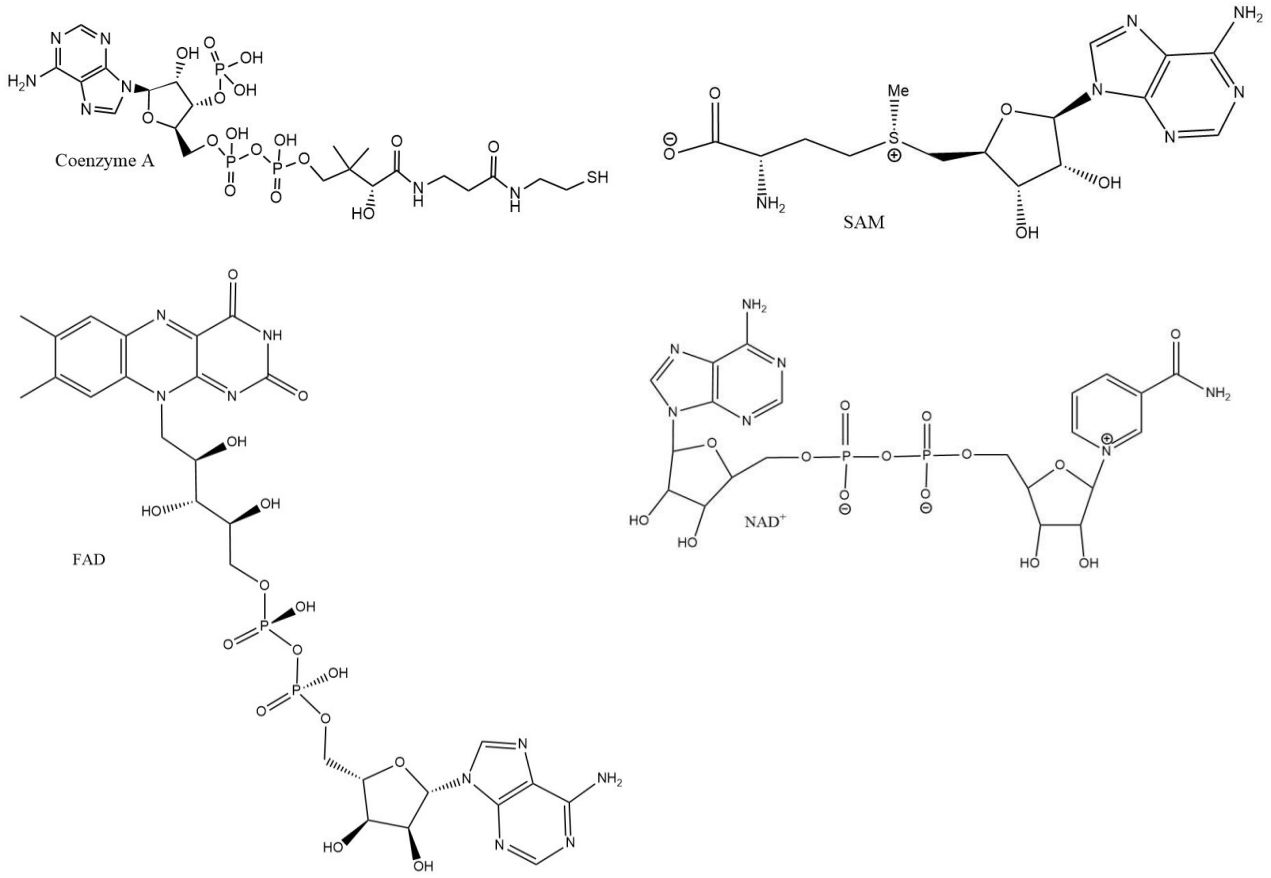


Figure 4: Structures of Coenzymes containing Adenine Moiety

Coenzyme A functions as an acyl group carrier and is used for the synthesis of fatty acids. S-Adenosylmethionine (SAM) is a common coenzyme involved in methyl group transfers, transsulfuration and aminopropylation. Flavin Adenine Dinucleotide (FAD) and Nicotinamide Adenine Dinucleotide (NAD^+) generally perform oxidation/reduction functions. In spite of being associated with diverse functionalities, all four coenzymes have the adenine moiety as part of their structures. This is interesting since the adenine moiety is away from the catalytically important part of these molecules. Except adenine, no other nucleobase is known to be naturally occurring in coenzymes/cofactors.

It has been suggested that these coenzymes were the catalytic active sites of ancient ribozymes, which transitioned to their current forms after the surrounding ribozyme

scaffolds were replaced by protein apoenzymes during the evolution of translation. Similar proposals have been made for iron-sulfur cluster cofactors being remnants of the geochemical setting of the origin of life.[19] Group transfer cofactors such as NADH, NADPH, $FADH_2$, FMN, SAM and CoA represent the centrality of nucleotide-derived coenzymes. Furthermore, the coenzyme binding capabilities of riboswitches are thought to represent an ancient capacity for coenzyme-mediated catalysis in an RNA world metabolism.[20]

During the transition from an RNA World to protein mediated metabolism, TIM barrel proteins may have provided a somewhat universal active site into which different ancient cofactors could be placed through evolutionary processes, thus transferring their catalytic properties from ribozymes (or minerals) into protein enzymes.[21]

It has also been found that 224 protein-ligand complexes (86 different proteins) from a total of 645 protein structure files bind ATP, CoA, NAD, NADP, FAD, or other adenine-containing ligands and use the same structural elements to recognize adenine, regardless of whether the ligand is a coenzyme, cofactor, substrate, or an allosteric effector. The common adenine-binding motif shown in this study is simple to construct. It uses only backbone polar interactions that are not dependent on the protein sequence or particular properties of amino acid side-chains, and nonspecific hydrophobic interactions. This is probably why so many different proteins with different functions use this motif to bind an adenylate-containing ligand. The adenylate-binding motif reported is present in "ancient proteins" common to all living organisms, suggesting that adenine-containing ligands and the common motif for binding them were exploited very early in evolution.[22]

1.5. Riboflavin (Vitamin B2) & its Physiological Significance:

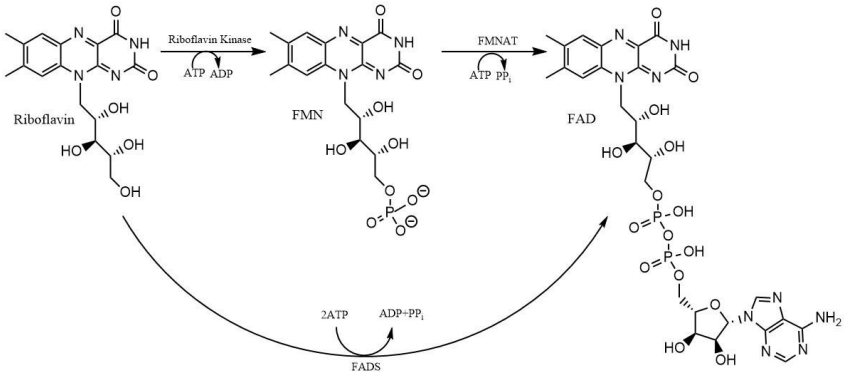


Figure 5: Enzymatic conversion of Riboflavin to form FMN & FAD

Riboflavin is an essential biomolecule that is made up of the isoalloxazine ring and the ribityl chain. Riboflavin is converted to FMN/FAD using enzymatic means. These two molecules can act as coenzymes catalyzing oxidation/reduction reactions.

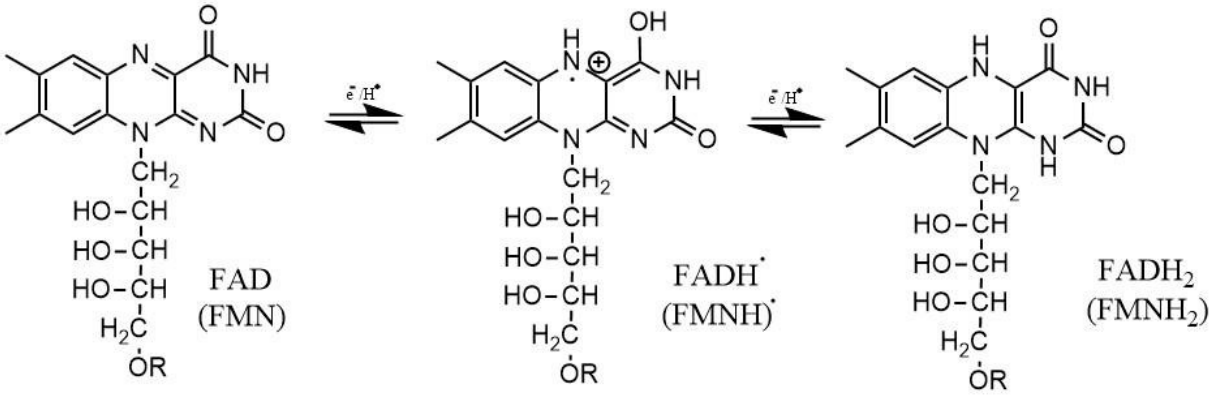


Figure 6: Mechanism of reduction of FAD

FAD and FMN can act as electron acceptors catalyzing both 1-electron and 2-electron transfers as shown in the figure above. Flavin-dependent enzymes are widespread in nature. A total of 90 proteins in humans (0.3% of human proteome) have been identified as flavoproteins. About 75% of flavoproteins utilize FAD as the cofactor whereas 25% use FMN as the cofactor. Most flavin-dependent enzymes are oxidoreductases (>90%) while the rest are transferases (4.3%), lyases (2.9%), isomerases (1.4%), and ligases (0.4%).[23] Certain flavin-dependent enzymes do not catalyze net redox reactions. The

enzyme UDP-galactopyranose mutase (UGM) catalyzes the isomerization of UDP-galactopyranose to UDP-galactofuranose.[24]

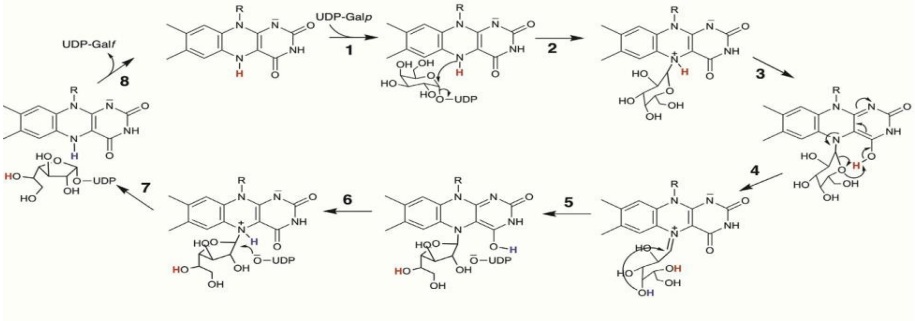


Figure 7: Mechanism of action of UDP-galactopyranose mutase[24]

1.6. Nucleotide Selectivity of Riboflavin Kinases:

Riboflavin kinase is an enzyme that uses an NTP source to catalyze the phosphorylation of riboflavin to form FMN. Eukaryotic and bacterial riboflavin kinases generally use ATP as the phosphorylation source whereas archaeal riboflavin kinases use CTP as the source.

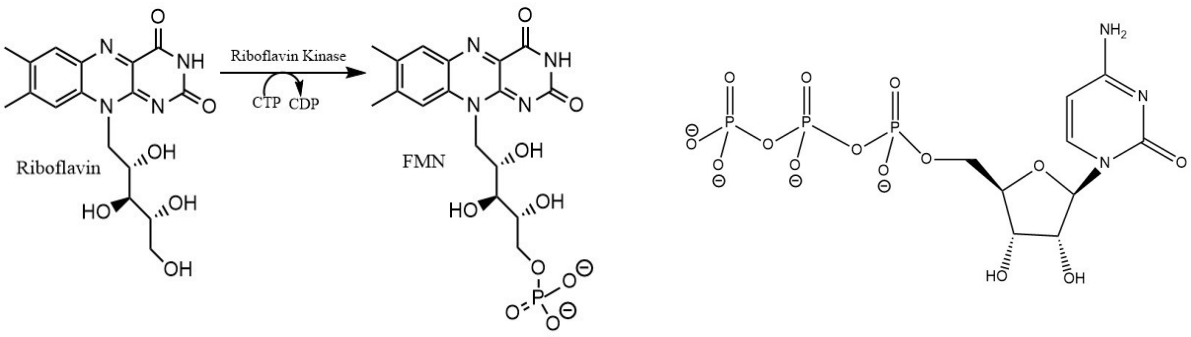


Figure 8: Mechanism of action of archaeal riboflavin kinases

Several ATP-dependent riboflavin kinases may use NTP's other than ATP for phosphorylation. For instance, the riboflavin kinase homolog of *Lactobacillus arabinosus* is capable of utilizing UTP as effectively as ATP.[25] Similarly, riboflavin kinase of *Bacillus subtilis* has shown to utilize CTP, UTP and dATP for the enzymatic reaction.[26] The riboflavin kinase isolated from *Glycine max* shows reaction even with ADP.[27]

However, CTP-dependent archaeal riboflavin kinases are generally strictly specific in terms of NTP utilization. One example of this is the riboflavin kinase isolated from the archaeal organism- *Methanocaldococcus jannaschii.*[28]

1.7. Discovery and Characterization of the Archaeal Riboflavin Kinase homolog in *Methanocaldococcus jannaschii* (*Mj*RibK):

The crystal structure of the MJ0056 gene product had a structural fold similar to human and yeast riboflavin kinases. Also, the MJ0056 gene was found to be right next to the MJ0055 gene(ribB) that is involved in the biosynthesis of riboflavin. These observations prompted the functional characterization of the MJ0056 gene product. The *Methanocaldococcus jannaschii* gene at locus MJ0056 (Swiss-Prot accession number Q60365) was amplified by PCR and ligated into plasmid pT7-7**.** Protein overexpression and purification were done followed by standard assays. Extracts of cells expressing the MJ0056 gene product were found to have high concentrations of FMN (Flavin Mononucleotide). It was also found that the enzyme preferentially uses CTP for phosphorylation of riboflavin.[28] The CTP specificity of *Mj*RibK was independently verified by Dr. Yashwant Kumar in our lab.

1.8. Mutational Studies performed on riboflavin kinase of *Methanocaldococcus jannaschii* (*Mj*RibK):

Dr. Yashwant Kumar of our lab performed mutagenesis studies on *Mj*RibK in order to determine the basis of nucleotide specificity of this archaeal riboflavin kinase. Attempts were made to get the mutant enzymes to accept ATP as the phosphorylation source. Mutants involving the R115 position of the nucleobase-surrounding loop were designed and tested for nucleotide promiscuity. R115 is part of the α-3 helix in the nucleobase-surrounding loop and forms an H-bond with the 3'-OH of the ribose ring.

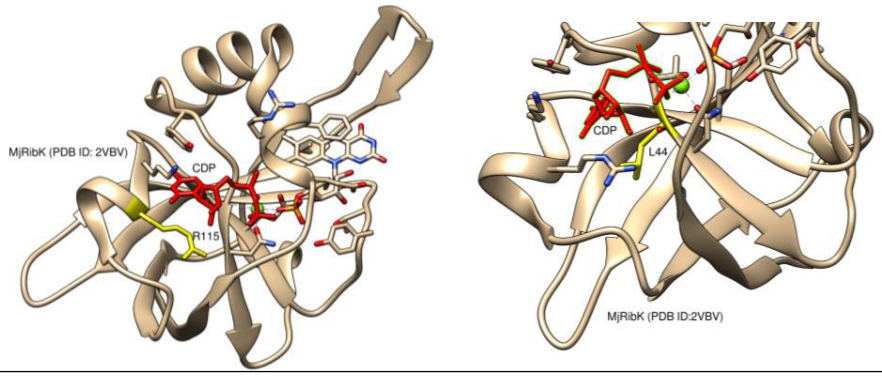


Figure 9: Mutagenesis studies done on *Mj*RibK

The L44 residue of the YxGTLN motif was also selected for mutagenesis studies. This residue was selected in order to make the cytosine pocket either large or hydrophobic.

#	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

Table 1: Description of *Mj*RibK Mutants

The mutants L44A and L44F could NOT accept ATP for phosphorylation and did NOT face a loss in activity with CTP. The mutant R115A did NOT have any altered NTP preference. The mutant R115P showed some activity with ATP. This effect was amplified in the double proline mutant R115PP. Hence, it was thought that introduction of two proline residues changes the α-3 helix conformation to an extended loop. This ultimately causes the enzyme to utilize ATP for phosphorylation.

1.9. Cloning and Purification of Chimeric *Mj*RibK:

Even though the proline mutants of *Mj*RibK showed activity with ATP, enzymatic activity with CTP was NOT entirely abolished. Hence, Dr. Yashwant Kumar planned to clone and purify a chimeric *Mj*RibK in which the α-3 helix was replaced with the loop sequence of *Hs*RibK (Human Riboflavin Kinase).

*Mj*RibK:

MIIEGEVVSGLGEGRYFLSLPPYKEIFKKILGFEPYEGTLNLKLDREFDINKFKYIETEDF

EFNGKRFFGVKVLPIKILIGNKKIDGAIVVPKKTYHSSEIIEIIAPMKLREQFNLKDGDVIKIL
IKGDKDE-

*Hs*RibK:

MRHLPYFCRGQVVRGFGRGSKQLGIPTANFPEQVVDNLPADISTGIYYGWASVGSGD
VHKMVVSIGWNPYYKNTKKSMETHIMHTFKEDFYGEILNVAIVGYLRPEKNFDSLESLIS
AIQGDIEEAKKRLELPEYLKIKEDNFFQVSKSKIMNGH-

Chimeric *Mj*RibK:

MIIEGEVVSGLGEGRYFLSLPPYKEIFKKILGFEPYEGTLNLKLDREFDINKFKYIETEDF
EFNGKRFFGVKVLPIKILIGNKKIDGAIVVPKKTYHSSEIIEIIMHTFKEDFYGEVIKILIKGD
KDE-

The chimeric protein was cloned and purified. However, it turned out to be inactive and did NOT give any activity with any of the NTP's. The inactivity of the chimeric protein was attributed to the fact that the swap was between a thermophilic enzyme (*Mj*RibK) and a mesophilic enzyme (*Hs*RibK). A chimeric protein wherein the α-helix of a mesophilic riboflavin kinase is replaced with the loop region of *Hs*RibK might help in resolving the issue.

1.10. Characterization of Riboflavin Kinase of *Methanococcus maripaludis*:

Riboflavin kinase of *Methanococcus maripaludis* (*Mmp*RibK) is one of the closest mesophilic homologs of *Mj*RibK. The sequences are 50.4% identical despite *Mmp*RibK belonging to a mesophilic archaeon. *Mmp*RibK converts riboflavin to FMN at 37°C in the presence of 10 mM $MgCl_2$ using CTP as the phosphate donor. *Mmp*RibK shows very high specificity for utilizing CTP and does not utilize other nucleotides such as ATP, GTP and UTP for the reaction. *Mmp*RibK was found to have maximum activity in the temperature range of 40°C–50°C. The activity declined and came down to zero as the temperature was raised from 50°C to 70°C. *Mmp*RibK is inactive without a metal ion. However, it showed full turnover when the divalent metal ion Mg^{2+} was added into the reaction. The overall trend of activity with different divalent metal ions is Mg^{2+} = Zn^{2+} > Mn^{2+} > Ni^{2+}> Co^{2+} > Fe^{2+}> Cu^{2+}> Ca^{2+}. [29]

2. Materials and Methods:

2.1. Materials:

Media components (Luria Bertani Broth) were purchased from SRL Chemicals. Antibiotics (Kanamycin & Ampicillin) were obtained from TCI. DNA ladders (200 – 10,000 bp) were obtained from NextGen. PrimeSTAR GXL Polymerase and QuickCut DpnI were purchased from TaKaRa. Solvents for HPLC were purchased from SD Fine Chemicals, Rankem, and Sigma-Aldrich. All other chemicals used were from Sigma-Aldrich (Merck) or Jena Bioscience. The genomic DNA of *Schizosaccharomyces pombe* was obtained from the lab of Dr. Mridula Nambiar at Indian Institute of Science, Education and Research, Pune.

2.2. Theoretical Methods:

2.2.1. Selection of *Mmp*RibK (Riboflavin kinase of *Methanococcus maripaludis)* mutants for substrate specificity studies:

PhD thesis of Dr. Yashwant Kumar of our lab includes the characterization of *Mj*RibK (riboflavin kinase of *Methanocaldococcus jannaschii*) mutants designed and tested for substrate specificity studies. These mutants were *Mj*RibK L44A, *Mj*RibK R115A, *Mj*RibK R115P and *Mj*RibK R115PP. In this thesis, the studies done on *Mj*RibK have been extended to a close mesophilic homolog- *Mmp*RibK. The sequences of the two proteins obtained from PDB[30] or UniProt[31] were aligned using MUSCLE (**MU**ltiple **S**equence **C**omparison by **L**og-**E**xpectation)[32] and the alignment was visualized using Jalview.[33] In the alignment, L44 residue of *Mj*RibK corresponded to L40 residue of *Mmp*RibK and R115 residue of *Mmp*RibK corresponded to R112 residue of *Mmp*RibK. Hence, *Mmp*RibK L40A, *Mmp*RibK R112A, *Mmp*RibK R112P and *Mmp*RibK R112PP were selected to be the candidate mutant proteins for characterization. It was also decided that a loopswap *Mmp*RibK mutant protein with the α-helix loop surrounding the nucleobase swapped with the loop sequence surrounding the corresponding region of *Hs*RibK (Human Riboflavin

Kinase) would be cloned and purified. This decision followed from similar attempts at *Mj*RibK loopswap mutant protein characterization performed by Dr. Yashwant Kumar.

2.2.2. Selection of *Hs*RibK (Human Riboflavin Kinase) mutants for characterization:
Sequences of nucleobase surrounding loops of four ATP-dependent eukaryotic riboflavin kinases from *Homo sapiens*[39], *Schizosaccharomyces pombe*[40], *Saccharomyces cerevisiae and Neurospora crassa* were analyzed. The inbuilt matchmaker tool on UCSF Chimera[34] was used to align the four protein crystal structures and analyze them. A BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) search[35] was performed using the FASTA sequence of *Schizosaccharomyces pombe.* The first 250 matches were aligned and the alignment was visualized using Jalview. The three amino acid stretch "DFY" at the seventh, eighth and ninth positions of the 11 amino acid loop surrounding the nucleobase were found to be 100% conserved in all the 250 matches. This prompted me to explore the role of this conserved region in eukaryotic riboflavin kinases using mutagenesis studies. *Hs*RibK (Human Riboflavin Kinase) was chosen since the sequence cloned into an appropriate plasmid (pPROEX HTA) was readily available in our lab. The mutants chosen were:

1) *Hs*RibK D89E- To study the effects of a slight change in the size of the cavity surrounding the NTP
2) *Hs*RibK Y91F- To understand the effect of the loss of a H-bond formed with the β-phosphate group of ATP
3) *Hs*RibK Y91K- To figure out whether lysine can act as a reasonable H-bonding substitute

2.2.3. Analysis of organisms with both Riboflavin Kinase and FAD Synthetase Domains:
INTERPRO database[36] was used to find an entry corresponding to "Riboflavin kinase domain, bacterial/eukaryotic" with accession ID: IPR015865. Two domain architectures within this entry were considered for the analysis: P22990 corresponding to the fused bifunctional FAD synthetase and Q03778 corresponding to monofunctional riboflavin kinase domain. Both lists were downloaded in TSV formats and loaded on Microsoft

Excel. Duplicate Tax ID's were deleted from both lists (individually). Following this, both lists were selected and duplicate Tax ID's were detected. The species name and taxonomic classification for each duplicate Tax ID was found manually.

2.3. Experimental Methods:

2.3.1. Cloning of *Mmp*RibK Mutants:

Single point *Mmp*RibK mutants were cloned using restriction-free methods. pET28a vector with *Mmp*RibK insert was used as the template for all PCR reactions. A forward/reverse primer with the mutation along with a standard T7 reverse/forward primer were used for the first PCR. The resulting megaprimer was visualized on a 1% agarose gel, and subsequently purified using protocols described in the Qiagen PCR Product Purification Kit. The megaprimer with the mutation was used for another round of PCR with pET28a WT *Mmp*RibK as the template. The resulting PCR product was digested with Dpn1 at 37℃ for 2.5 hours in order to cleave the methylated template DNA.

2.3.2. List of primers for cloning of *Mmp*RibK Mutants:

Serial Number	Mutation	Primer Sequence(5' to 3')
1	*Mmp*RibK L40A (Forward Primer)	CCGTTTGAAGGAACAGCGAACGTAAAATTAAAG
2	*Mmp*RibK R112A (Forward Primer)	CCAATTCAGCTTGCGAAATTTTTATTATTAAAG
3	*Mmp*RibK R112P (Forward Primer)	CCAATTCAGCTTCCGAAATTTTTATTATTAAAG
4	*Mmp*RibK	CCAATTCAGCTTCCGCCGAAATTTTTATTATTAAAG

	R112PP (Forward Primer)	
5	*Mmp*RibK + *Hs*RibK loopswap mutant (Reverse Primer)	TTAAATTACGAGTTTTACAACTTCGCCATAAAAATCTTCTTTAAAGGTATGCATAATAATTTCCAAGGTTTTTTTAG
6	T7 Forward Primer	TAATACGACTCACTATAGGG
7	T7 Reverse Primer	GCTAGTTATTGCTCAGCGG

Table 2: List of primers- Cloning of *Mmp*RibK Mutants

2.3.3. Transformation of PCR Products and Plasmid Purification:

Competent *E. coli* DH5α cells were prepared using $CaCl_2$ or Hanahan method. 10 µL of the DpnI digestion products were added to 100 µL of freshly prepared competent cells. The mixtures were incubated on ice for 15 minutes. A heat shock was given at 42℃ for 1 minute followed by an incubation on ice for 5 min. 200 µL of LB broth was added to the vials. The vials were incubated at 37℃ for 1.5 h. The contents were then spread on LB agar plates containing 25 µg/mL kanamycin (100 µg/mL ampicillin for *Hs*RibK mutants) and incubated at 37℃ for 24-36 hours. Colonies thus obtained were inoculated into LB broth and the mutant plasmids were purified using standard protocols as described by plasmid purification kits.

2.3.4. Cloning of *Hs*RibK Mutants:

*Hs*RibK mutants were cloned using QuikChange site-directed mutagenesis.[37] This protocol is a simple and efficient one-step procedure that utilizes overlapping primers (with mutation) containing extended non overlapping sequences at the 3' end. This primer design scheme is efficient since it enables the PCR products to be used as templates for

more rounds of amplification. The PCR products were transformed into competent DH5α cells and the plasmids were purified according to the protocol described in 2.3.3.

2.3.5. List of primers for cloning of *Hs*RibK Mutants:

Serial Number	Mutation	Primer Sequence (5' to 3')
1	*Hs*RibK D89E (Forward Primer)	GAGGAGTTCTATGGGGAAATCCTCAATGTGG
2	*Hs*RibK D89E (Reverse Primer)	ATAGAACTCCTCTTTGAAGGTATGCATGATATGTGTT
3	*Hs*RibK Y91F (Forward Primer)	GACTTCTTCGGGGAAATCCTCAATGTGGCCATTG
4	*Hs*RibK Y91F (Reverse Primer)	TCCCCGAAGAAGTCCTCTTTGAAGGTATGCATGATATGTG
5	*Hs*RibK Y91K (Forward Primer)	GACTTCAAGGGGGAAATCCTCAATGTGGCCATTG
6	*Hs*RibK Y91K (Reverse Primer)	TCCCCCTTGAAGTCCTCTTTGAAGGTATGCATGATATGTG

Table 3: List of primers- Cloning of *Hs*RibK Mutants

2.3.6. Cloning of *Sp*FADS (FAD Synthetase of *Schizosaccharomyces pombe*) from Genomic DNA to pET28a vector:

For the purpose of characterization of FADS of *Schizosaccharomyces pombe*, it was necessary to clone the sequence coding for the enzyme into a suitable vector. The entire exercise was a three-step process. The first step involved cloning out the *Sp*FADS gene from the genomic DNA using forward and reverse primers complementary to the two ends of the gene. The resulting megaprimer was visualized on a 1% agarose gel, and subsequently purified using protocols described in the Promega PCR Product Purification Kit. The second step involved the amplification of the first megaprimer using two hybrid primers that have overhangs complementary to the desired vector (pET28a). The second megaprimer was purified in a similar manner and used for the third PCR with empty

pET28a vector. The resulting PCR product was digested with Dpn1 at 37℃ for 2.5 hours in order to cleave the methylated template DNA. The PCR products were transformed into competent DH5α cells and the plasmids were purified according to the protocol described in 2.3.3.

2.3.7. List of primers for cloning *Sp*FADS from the genomic DNA:

Serial Number	Name of Primer	Primer Sequence (5' to 3')
1	*Sp*FADS_F1 (1st step-forward primer)	ATGGATGAGTTGGAAC
2	*Sp*FADS_R1 (1st step-reverse primer)	TTAAGATACTTCGGAAG
3	*Sp*FADS_F2 (2nd step-forward primer)	GCCGCGCGGCAGCCATATGGATGAGT TGGAAC
4	*Sp*FADS_R2 (2nd step-reverse primer)	CGACGGAGCTCGAATTCGGATCCTTAA GATACTTCGGAAG

Table 4: List of primers- Cloning of *Sp*FADS

2.3.8. Transformation of Plasmids into *E. coli* BL21(DE3) cells:

All successful mutant plasmids need to be transformed into BL21(DE3) cells for overexpression of mutant proteins. Competent *E. coli* BL21(DE3) cells were prepared using $CaCl_2$ or Hanahan method. 100-200 ng of the plasmid DNA was added to 50 µL of freshly prepared competent cells. The mixtures were incubated on ice for 15 minutes. A heat shock was given at 42℃ for 1 minute followed by an incubation on ice for 5 min. 200 µL of LB broth was added to the vials. The vials were incubated at 37℃ for 1.5 h. The contents were then spread on LB agar plates containing 25 µg/mL kanamycin (100 µg/mL ampicillin for *Hs*RibK mutants) and incubated at 37℃ for 24-36 hours. Colonies thus obtained were inoculated into LB broth and this primary culture was used to prepare glycerol stocks that would ultimately be used for protein overexpression.

2.3.9. Protein Overexpression & Purification Protocol:

For primary inoculation, the glycerol stocks prepared in the previous step were used for inoculation in a 10 mL Luria-Bertani (LB) broth containing 25 µg/mL kanamycin (100 µg/mL ampicillin for *Hs*RibK mutants) and the resulting culture was incubated overnight at 37℃ at a constant shaking of 180 rpm. Then, 10 mL of primary culture and 25 µg/mL kanamycin (100 µg/mL ampicillin for *Hs*RibK mutants) were added to a 1 L secondary culture. The secondary culture was grown at 37℃ until it reached an O. D. of 0.6-0.8 after which an appropriate quantity (0.5-1 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture. After incubation at 18℃ for 18 hours at a constant shaking of 180 rpm, cells were pelleted by centrifugation for 15 minutes at 5,500 rpm at 4℃. The cell pellets were resuspended in a lysis buffer consisting of 50 mM Tris-Cl (pH 8.0), 300 mM NaCl, 0.1 mM PMSF, 20 mM imidazole and 0.025% β-mercaptoethanol (βME). The solution was then sonicated on ice for 10 minutes (1 second “on”, 3 seconds “off” at 60% amplitude). The lysate was centrifuged at 18,000 rpm for 15 minutes at 4℃. The supernatant (also known as clarified lysate) is used for purification of the protein.

2.3.10. Buffer Compositions for Protein Purification:

Buffer B	Buffer A
50 mM Tris/HCl pH 8.0	50 mM Tris/HCl pH 8.0
300 mM NaCl	300 mM NaCl
500 mM Imidazole	0 mM Imidazole
0.025% β-Mercaptoethanol	0.025% β-Mercaptoethanol

Table 5: Buffer Compositions for Protein Purification

2.3.11. Ni Affinity Chromatography for Protein Purification:

An immobilized metal affinity chromatography (IMAC) column of 5 mL capacity was used along with a peristaltic pump (to control the flow rate of solutions through the column) for protein purification. The clarified lysate was loaded onto the column equilibrated with an

equilibration buffer (2 mL Buffer B+ 48 mL Buffer A). The column was subsequently washed with Wash Buffer Ⅰ (5 mL Buffer B+ 45 mL Buffer A) and Wash Buffer Ⅱ (7 mL Buffer B+ 43 mL Buffer A). The protein of interest was eluted by running an elution buffer(250 mM imidazole)(25 mL buffer B+ 25 mL buffer A) through the column at the rate of 1mL/min. Eluent fractions were collected in separate 2 mL eppendorfs. Protein elution was verified using bradford assay[38] and SDS-PAGE. The fractions containing the desired protein were pooled together and buffer-exchanged in a desalting buffer (50 mM Tris-Cl pH 8.0, 300 mM NaCl, 0.025% βME) using a Bio-Rad desalting column. The desalted protein was stored with 12% glycerol at -80℃.

2.3.12. Denaturation-Refolding Protocols:[41,42,43,44]

Attempts were made to extract *Mmp*RibK mutants that crashed into the insoluble pellet using the following protocols.

2.3.12.1. Refolding by Dilution:

The pellet that remains after the clarified lysate has been removed was solubilized in a buffer that contains 100 mM Tris pH 12.5, 2M urea, 0.025% βME, 300 mM NaCl and 20 mM Imidazole. This solution was kept at room temperature for four hours after which it was centrifuged at 12,000 RPM for 30 minutes at 4℃. At this step, it is expected that the misfolded protein gets fully denatured and seeps into the soluble fraction. The methods used for purification of the mutant protein from this soluble fraction were:

1) This method involved Ni affinity chromatography in a procedure similar to the one given in 2.3.11. However, note that buffers A and B were at pH 12.5 and had 2M urea added to them to ensure that the protein remains in the denatured form during the purification process.
2) This method involved the use of Promega HisLink™ Protein Purification Resin. 2-4 mL of resin in a falcon is sufficient to bind protein from 1 litre of cell culture. Urea containing equilibration buffer is added to the falcon and mixed well. The falcon is centrifuged and additional equilibration buffer is removed. The lysate (containing the protein) is added to the equilibrated resin, stirred well and decanted after centrifugation. The wash steps are procedurally identical to the equilibration step. Elution buffer (250 mM Imidazole) is added for the final elution step.

Buffers used for steps 1 & 2 are as follows:

Equilibration Buffer	Wash Ⅰ Buffer	Wash Ⅱ Buffer	Elution Buffer
100 mM Tris pH 12.5	100 mM Tris pH 12.5	100 mM Tris pH 12.5	100 mM Tris pH 12.5
2 M Urea	2 M Urea	2 M Urea	2 M Urea
300 mM NaCl	300 mM NaCl	300 mM NaCl	300 mM NaCl
0.025% β-Mercaptoethanol	0.025% β-Mercaptoethanol	0.025% β-Mercaptoethanol	0.025% β-Mercaptoethanol
20 mM Imidazole	50 mM Imidazole	70 mM Imidazole	250 mM Imidazole

Table 6: Buffer Compositions- Denaturation/Refolding Protocols

Following elution, the solubilized protein was added at 1mL/min to ice cold refolding buffer (50 mM Tris/HCl pH 8.0, 300 mM NaCl, 0.025% β- Mercaptoethanol) the volume of which was 10 times that of the solubilized protein solution. The solution was magnetically stirred following which a 10,000 dalton molecular weight filter was used to concentrate the refolded protein.

2.3.12.2. Refolding by Dialysis:

This method is almost the same as 2.3.12.1 apart from the last step (refolding). This method involves the use of a dialysis membrane that is dipped in a large volume of ice cold refolding buffer solution. The solubilized protein is loaded into the dialysis membrane and kept overnight for refolding. This refolded protein can then be removed from the membrane and stored at -80℃.

2.3.13. TLC (Thin Layer Chromatography) for analysis of enzyme activity:

Primary verification of enzymatic conversion of riboflavin to FMN was performed using TLC (Merck Silica Gel 60 F_{254}) using water, acetic acid and n-butanol in the ratio 5:1:4

ratio as solvent. Note that TLC spots greater than 0.2 µL (of 100 µM riboflavin solutions) would cause streaks on the TLC plate

2.3.14. HPLC (High Performance Liquid Chromatography) for quantification of enzyme activity:

The enzymatic conversion of riboflavin to FMN was analyzed on C-18 reverse phase column (Phenomenex, Gemini 5 µm NX-C18 110Å, 250*4.6 mm) Agilent HPLC system (1260 Infinity Ⅱ) by means of UV-Vis detectors. The samples were loaded into HPLC vials after centrifugation at 14,000 RPM for precipitation of the enzyme. 40 µL of every sample was injected for the runs. The solvents involved in the run were Methanol and 10 mM Ammonium Acetate pH 6.5. The method consisted of the following steps: 0 to 5 minutes- 5% Methanol/95% Ammonium Acetate; 5 to 40 minutes from 5% Methanol/95% Ammonium Acetate to 80% Methanol/20% Ammonium Acetate (Linear Gradient); 40 to 55 minutes from 80% Methanol/20% Ammonium Acetate to 5% Methanol/95% Ammonium Acetate (Linear Gradient); 55 minutes to 60 minutes at 5% Methanol/95% Ammonium Acetate.

3. Results and Discussions:

3.1. Selection of *Mmp*RibK Mutants for Substrate Specificity Studies:

The sequence alignment of *Mj*RibK (sequence obtained from PDB ID: 2VBS) and *Mmp*RibK (sequence obtained from Uniprot ID: Q6M0T4) gave rise to the following alignment:

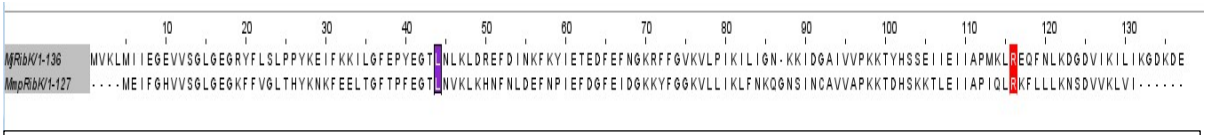


Figure 10: Sequence Alignment of *Mj*RibK and *Mmp*RibK

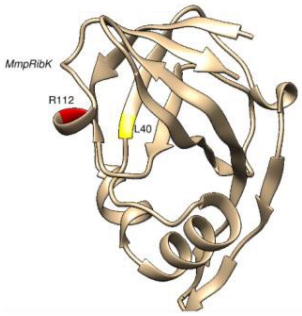


Figure 11: Modelled structure of *Mmp*RibK showing sites of mutagenesis

The single-point mutants *Mmp*RibK L40A, *Mmp*RibK R112A, *Mmp*RibK R112P and *Mmp*RibK R112P were selected on the basis of this alignment as explained in Section 2.2.1.

The *Mmp*RibK loopswap mutant protein involved swapping the α helix of the nucleobase surrounding loop with the corresponding loop of *Hs*RibK.

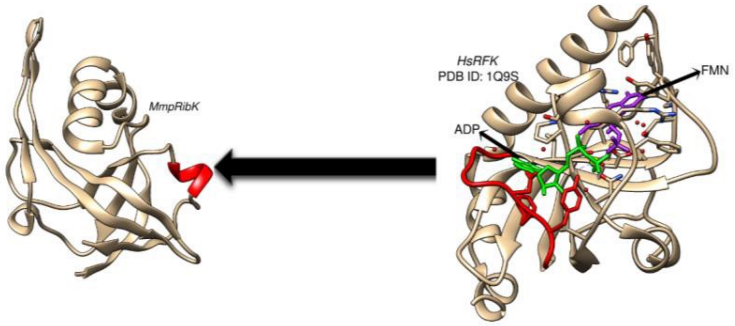


Figure 12: A visual representation of the *Mmp*RibK loopswap mutant

*Mmp*RibK Sequence:

MEIFGHVVSGLGEGKFFVGLTHYKNKFEELTGFTPFEGTLNVKLKHNFNLDEFNPIEFD GFEIDGKKYFGGKVLLIKLFNKQGNSINCAVVAPKKTDHSKKTLEIIAPIQLRKFLLLKNS DVVKLVI-

*Hs*RibK Sequence:

MRHLPYFCRGQVVRGFGRGSKQLGIPTANFPEQVVDNLPADISTGIYYGWASVGSGD VHKMVVSIGWNPYYKNTKKSMETHIMHTFKEDFYGEILNVAIVGYLRPEKNFDSLESLIS AIQGDIEEAKKRLELPEYLKIKEDNFFQVSKSKIMNGH-

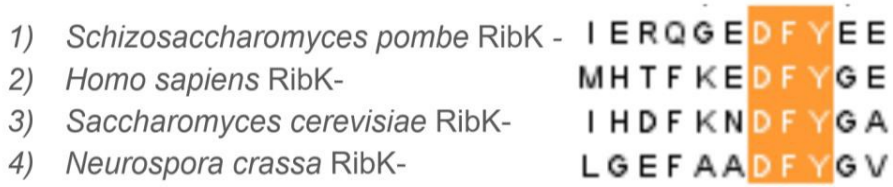
*Mmp*RibK_*Hs*RibK loopswap mutant protein sequence:

MEIFGHVVSGLGEGKFFVGLTHYKNKFEELTGFTPFEGTLNVKLKHNFNLDEFNPIEFD GFEIDGKKYFGGKVLLIKLFNKQGNSINCAVVAPKKTDHSKKTLEIIMHTFKEDFYGEVV KLVI-

The highlighted portions in the protein sequences above clearly illustrate the region of *Mmp*RibK that is to be swapped with the corresponding region of *Hs*RibK.

3.2. Selection of *Hs*RibK Mutants for Characterization:

3.2.1. Loop sequences of four eukaryotic RibK's:



The above represents the loop region sequences of four eukaryotic riboflavin kinases. Note that all four of them have loops that are exactly 11 amino acids long with the seventh, eighth and ninth residues being D (Aspartate), F (Phenylalanine) and Y (Tyrosine) respectively.

3.2.2. MatchMaker Alignment of Eukaryotic RibK's:

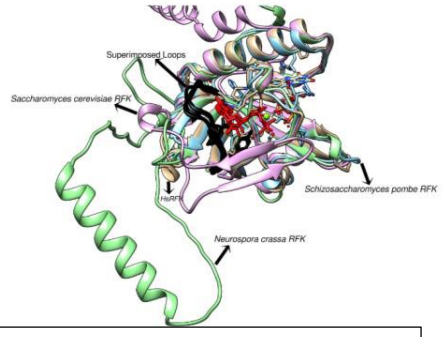


Figure 13: Matchmaker alignment of four eukaryotic riboflavin kinases

The MatchMaker function on UCSF Chimera works according to the following steps:

A) Generating pairwise sequence alignments

B) Superimposing the structures according to those pairwise alignments

C) Optionally, creating a multiple sequence alignment from the structural superposition.

The structural alignment of the four eukaryotic riboflavin kinases under consideration as generated by UCSF Chimera shows very close overlap in the orientation of the loop region of these structures.

3.2.3. Multiple Sequence Alignment of Eukaryotic Riboflavin Kinases:

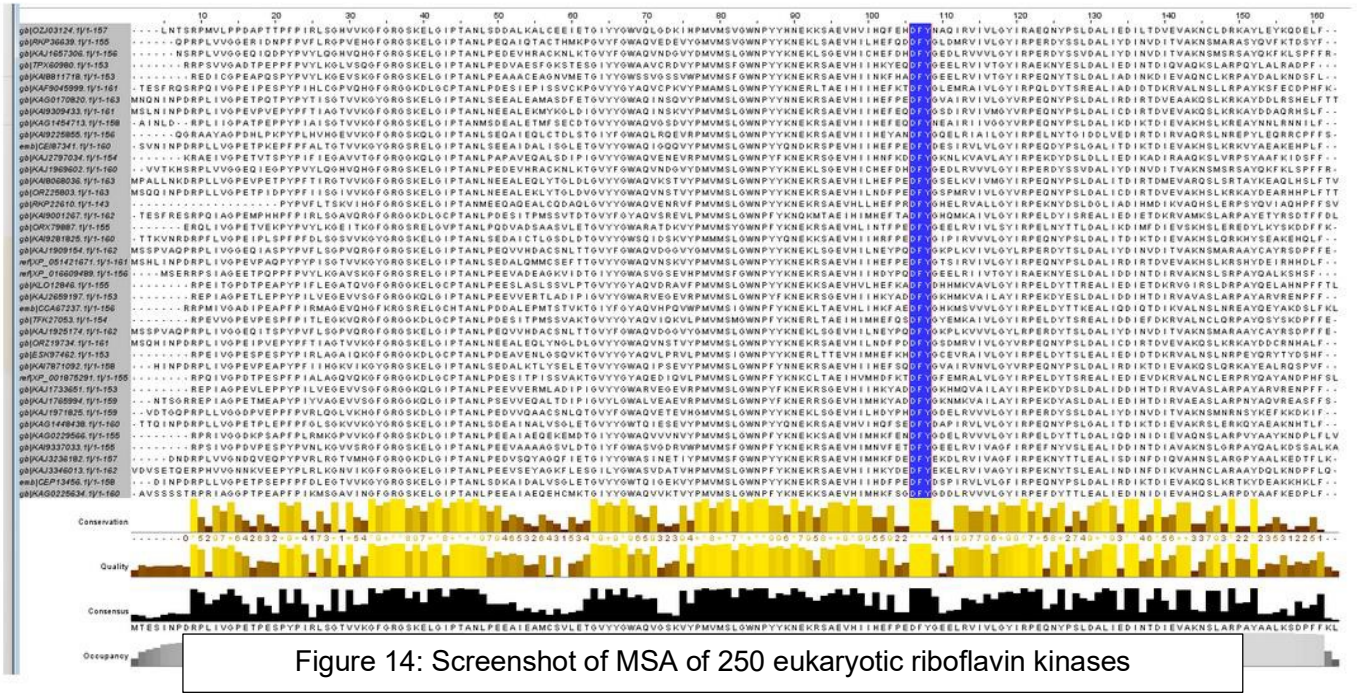


Figure 14: Screenshot of MSA of 250 eukaryotic riboflavin kinases

A BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) search was performed with default parameters using the FASTA sequence of *Schizosaccharomyces pombe* riboflavin kinase (PDB ID: 1N05). The first 250 matches were aligned as shown below:

The three amino acid stretch “DFY” was found to be 100% conserved in all the 250 matches. This prompted me to explore the role of this conserved region in eukaryotic riboflavin kinases using mutagenesis studies.

3.2.4. Description of *Hs*RibK Mutants selected for characterization:

1) Aspartate residue (89D) forms H-bonds with another lysine residue (21K) above it thereby closing off a part of the enzyme surrounding the NTP. The mutation D89E (aspartate to glutamate) was planned in order to study the effects of a slight change in the size of the cavity surrounding the NTP.

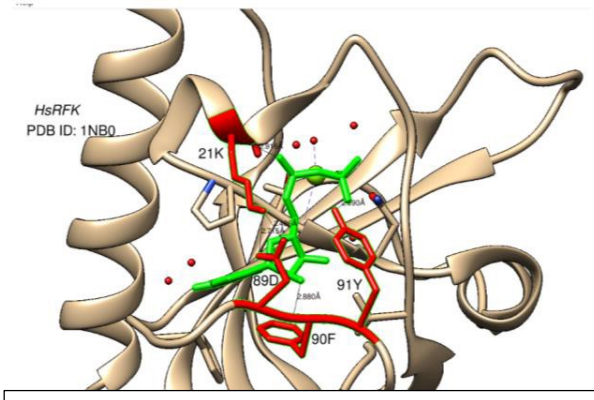


Figure 15: Structural view of the DFY motif in human riboflavin kinase (*Hs*RibK)

2) The tyrosine residue (91Y) forms an H-bond with the β-phosphate group of ATP. The mutation Y91F (tyrosine to phenylalanine) was decided upon to understand the effect of the loss of the H-bond on enzyme activity.
3) Y91K(tyrosine to lysine) is another mutant that was decided upon in order to figure out whether lysine can act as a reasonable H-bonding substitute.

3.3. Analysis of organisms that have both RibK (Riboflavin Kinase) and FADS (FAD Synthetase):

Prior to performing this analysis, I was aware of four organisms that have both RibK (Riboflavin Kinase) and FADS (FAD Synthetase). These four organisms are: *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Source: Doctoral thesis of Dr. Yashwant Kumar). I wanted to perform a much more

extensive search to understand the types and percentage of organisms that have both these enzymes. This was an interesting problem since the role of having a riboflavin kinase is unclear when the organism already has a bifunctional FAD synthetase. The procedure described in 2.2.3 was followed to approach this problem. The PFAM entry corresponding to “Riboflavin Kinase domain; bacterial/eukaryotic” (IPR015865) had two domain architectures that were taken into consideration.

As shown in Figure 7, domain architecture P22990 represents the fused bifunctional FAD synthetase (IPR 015864= FAD_Synthase & IPR 015865= Riboflavin_Kinase) and domain architecture Q03778 represents the monofunctional riboflavin kinase (Riboflavin_Kinase_bac/euk). These two domain architectures had 36,732 and 5,405 protein sequences respectively. Once the Tax ID’s corresponding to these sequences were downloaded and duplications were removed, it was discovered that P22990 contains 24,469 species and Q03778 contains 3,740 species.

In order to figure out the number of species that have both the monofunctional and bifunctional enzymes, we need to figure out the Tax ID’s that occur in both these lists. These duplications were detected and the data corresponding to each such Tax ID was collated.

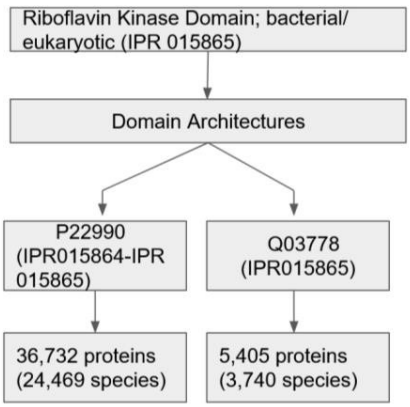


Figure 16: Domain Architectures within IPR 015865 (Interpro)

Out of a total of 27,636 species surveyed, it was found that only 573 of them have both Riboflavin Kinase and FAD Synthetase. This corresponds to 2% of the total species surveyed. This implies that the occurrence of both these enzymes in a single organism is very rare. Only 17 of the list of 573 organisms were eukaryotes. Of these 17, 10 were arthropods. Of these 10, 4 were *Drosophila sp.*

The table below shows a taxonomic classification of the species in IPR 015865 and my curated data set.

	IPR 015865	My Curated Data Set
Eukaryotic Species	2,680 (9.4%)	17 (2.9%)
Bacterial Species	25,631 (90.6%)	556 (97%)
A) Firmicutes	4,601 (16.2%)	106 (18.4%)
B) Actinobacteria	4,464 (15.7%)	182 (31.7%)
C) Bacteroidetes	2,472 (8.7%)	28 (4.8%)
D) Proteobacteria	10,807 (38.17%)	155 (27.05%)

Table 7: Taxonomic classification of IPR 015865 and the data set comprising of organisms that have both Riboflavin Kinase and FAD Synthetase.

According to the numbers and percentages given in Table 1, it is evident that eukaryotes (as compared to prokaryotes) are less likely to have both these enzymes. Within bacterial species, actinobacteria are much more likely to have both these enzymes whereas bacteroidetes and proteobacteria are much less likely to have both riboflavin kinase and FAD synthetase.

Interpro databases IPR 023602, IPR 039063 and IPR 023470 were scanned for a similar analysis of CTP dependent/archaeal riboflavin kinases. None of the domain architectures

in any of these databases had the fused bifunctional enzyme. Hence, it was clear that no archaeal organism has the FAD synthetase.

3.4. Cloning of *Mmp*RibK Mutants:

Single point mutants of *Mmp*RibK were cloned according to the procedure given in 2.3.1 using the primers given in 2.3.2. The cloning was successful as shown by the sequence alignments of the sequencing results.

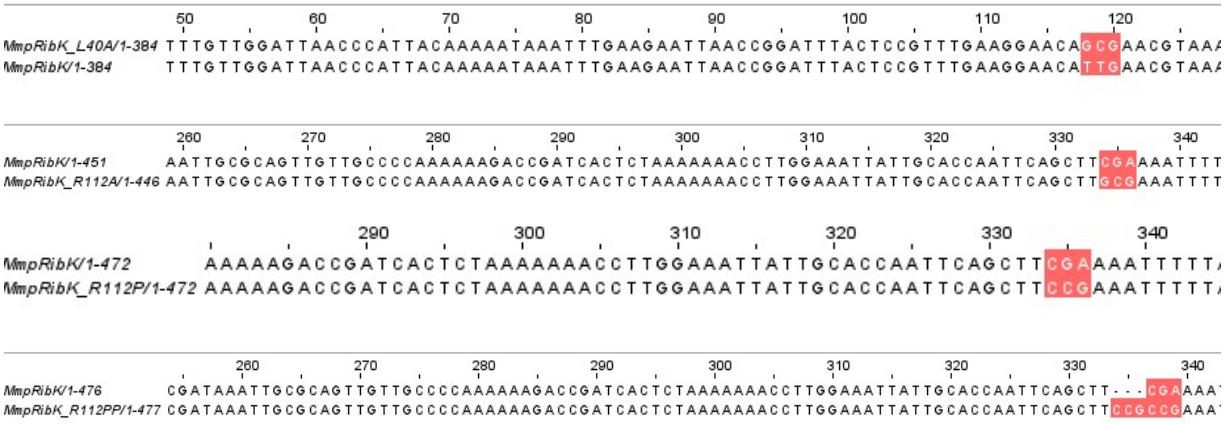


Figure 17: Sequencing results of *Mmp*RibK Mutants

3.5. Cloning of *Mmp*RibK loopswap Mutant:

*Mmp*RibK loopswap mutant was cloned according to the procedure given in 2.3.1 using the primers given in 2.3.2. The sequencing result for the cloning is given below:

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)

MmpRibK_loopswap_theory             ATGGAAATTTTTGGGCATGTCGTAAGCGGACTTGGAGAGGGTAAGTTTTTTGTTGGATTA
MmpRibK_loopswap_sequencingresul    ATGGAAATTTTTGGGCATGTCGTAAGCGGACTTGGAGAGGGTAAGTTTTTTGTTGGATTA
                                    ************************************************************

MmpRibK_loopswap_theory             ACCCATTACAAAAATAAATTTGAAGAATTAACCGGATTTACTCCGTTTGAAGGAACATTG
MmpRibK_loopswap_sequencingresul    ACCCATTACAAAAATAAATTTGAAGAATTAACCGGATTTACTCCGTTTGAAGGAACATTG
                                    ************************************************************

MmpRibK_loopswap_theory             AACGTAAAATTAAAGCATAATTTCAATTTAGATGAATTTAATCCAATAGAATTCGACGGA
MmpRibK_loopswap_sequencingresul    AACGTAAAATTAAAGCATAATTTCAATTTAGATGAATTTAATCCAATAGAATTCGACGGA
                                    ************************************************************

MmpRibK_loopswap_theory             TTTGAAATTGACGGGAAAAAATACTTTGGTGGAAAGGTACTTTTAATAAAACTATTTAAT
MmpRibK_loopswap_sequencingresul    TTTGAAATTGACGGGAAAAAATACTTTGGTGGAAAGGTACTTTTAATAAAACTATTTAAT
                                    ************************************************************

MmpRibK_loopswap_theory             AAACAGGGAAATTCGATAAATTGCGCAGTTGTTGCCCCAAAAAAGACCGATCACTCTAAA
MmpRibK_loopswap_sequencingresul    AAACAGGGAAATTCGATAAATTGCGCAGTTGTTGCCCCAAAAAAGACCGATCACTCTAAA
                                    ************************************************************

MmpRibK_loopswap_theory             AAAACCTTGGAAATTATTATGCATACCTTTAAAGAAGATTTTTATGGCGAAGTTGTAAAA
MmpRibK_loopswap_sequencingresul    AAAACCTTGGAAATTATTATGCATACCTTTAAAGAAGATTTTTATGGCGAAG------AA
                                    ************************************************      **

MmpRibK_loopswap_theory             CTCGTAATTTAA
MmpRibK_loopswap_sequencingresul    CTCGTAATTTAA
                                    ************
```

Figure 18: Sequencing result of *Mmp*RibK loopswap mutant

The cloning was completed apart from the observation that six nucleotides were missing in the final sequencing result close to the C-terminus region of the protein. The reasons for this are unclear. However, these six nucleotides are located after the loop region thus ensuring that the “loopswap region” remains unaffected in the mutant that was cloned.

3.6. Cloning of *Hs*RibK Mutants:

The gel image shown here depicts a gradient PCR that was performed using the respective QuikChange primers. The products obtained at all four annealing temperatures were pooled for each mutant and subsequently used for transformation into DH5α cells.

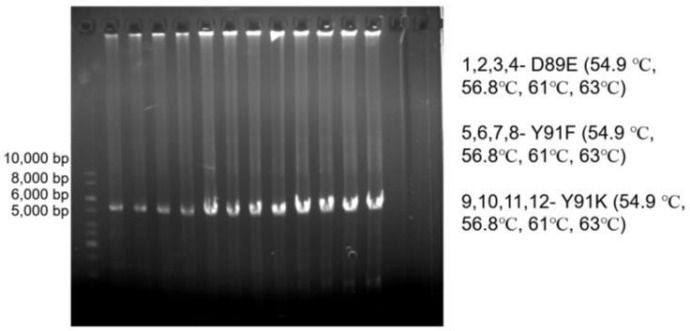


Figure 19: Gradient PCR-Cloning of *Hs*RibK Mutants

3.6.1. Cloning of *Hs*RibK D89E and *Hs*RibK Y91K:

According to the sequencing results given below, it is clear that *Hs*RibK D89E and *Hs*RibK Y91K have been successfully cloned.

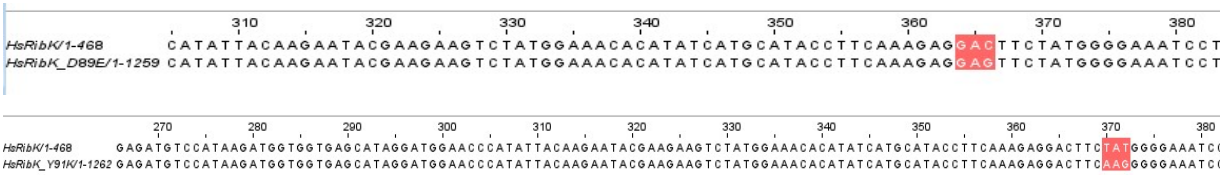


Figure 20: Sequencing Results: *Hs*RibK D89E & *Hs*RibK Y91K

3.6.2. Cloning of *Hs*RibK Y91F:

According to the sequencing results of *Hs*RibK Y91F, it is clear that the mutation has been introduced (TTC instead of TAT i.e. the region highlighted in pink). However, another mutation (CAG instead of CAT i.e. the region highlighted in blue) has also been introduced in the mutant.

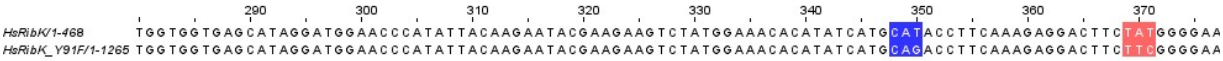


Figure 21: Sequencing Results: *Hs*RibK Y91F

The second mutation has been attributed to a mistake in the reverse primer.

The primer sequence is supposed to have been:

5' TCCCCGAAGAAGTCCTCTTTGAAGGT**A**TGCATGATATGTG 3'

However, the primer that had been provided has the sequence

5' TCCCCGAAGAAGTCCTCTTTGAAGGT**C**TGCATGATATGTG 3'

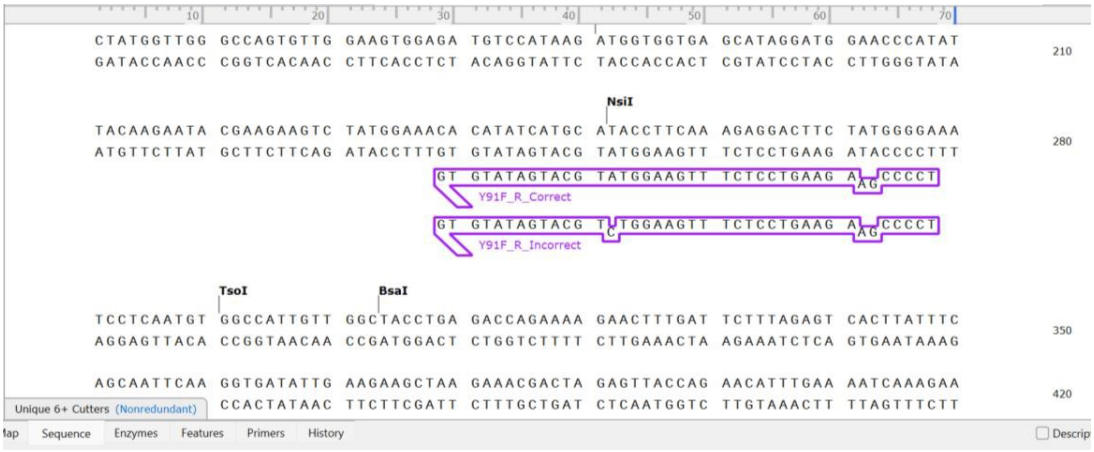


Figure 22: Representation of incorrect Y91F_R primer

The mistake in the reverse primer was communicated to the company that synthesized and shipped the primers-"BioServe- A REPROCELL COMPANY". The primer was resynthesized and shipped to me. However, cloning experiments with the newly synthesized primer haven't been successful yet.

3.7. Cloning of *Sp*FADS into pET28a Vector:

3.7.1. PCR 1- Cloning out *Sp*FADS gene from Genomic DNA:

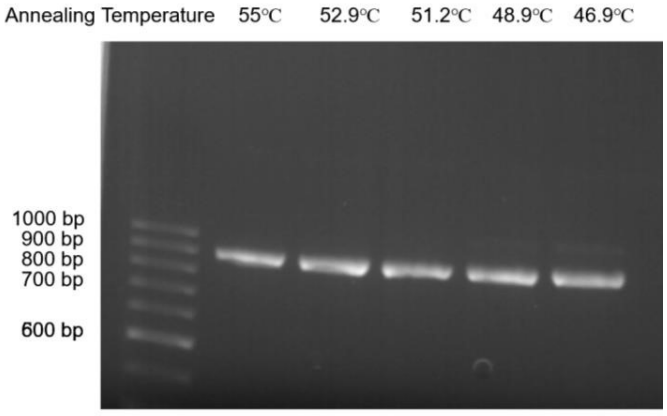


Figure 23: Gel Image-PCR1-*Sp*FADS cloning

For the purposes of cloning out the *Sp*FADS gene from the genomic DNA, a gradient PCR was performed wherein the annealing temperature varied from 46.9℃ to 55℃. Amplification of the gene was detected at all five of these temperatures as shown in the agarose gel image above. Length of the amplified gene fragment is supposed to be 798 bp. All five products were pooled and purified using a PCR Product Purification Kit. This purified megaprimer was used for the second PCR.

3.7.2. PCR2- Cloning out *Sp*FADS gene from Genomic DNA:

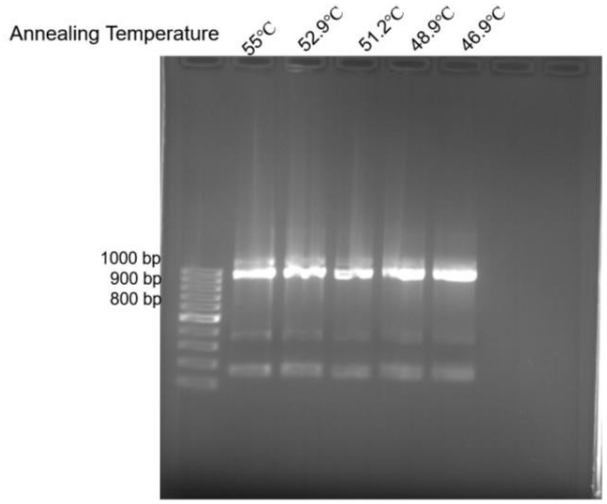


Figure 24: Gel Image-PCR2-*Sp*FADS cloning

The second set of primers were used to amplify the megaprimer purified in 3.7.1. Gradient PCR performed in the annealing temperature range between 46.9℃ and 55℃ yielded amplified gene products for all the temperature values. The five products were pooled and purified using a PCR Product Purification Kit. The purified megaprimer was used for the third PCR that involved insertion into an empty pET28a vector.

3.7.3. PCR3- Insertion of *Sp*FADS into pET28a vector:

PCR was performed using the product purified in 3.7.2. An empty pET28a vector was used as the template for the reaction. Transformation into DH5α cells was performed followed by Dpn1 digestion. Colonies were obtained on agar plates. Plasmids were purified from those colonies.

A confirmatory PCR was run to verify if the purified plasmids were the mutant ones. This PCR was performed using T7 Forward and T7 Reverse primers with empty pET28a and cloned plasmids.

Length of PCR Products:

Empty pET28a= 361 bp

Cloned plasmid= 1159 bp

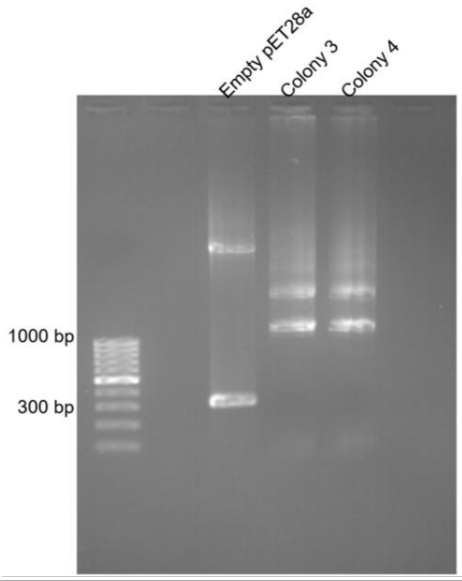


Figure 25: Gel Image-Confirmatory PCR-*Sp*FADS cloning

The cloning seemed to be successful according to the results of the confirmatory PCR. The plasmids were sent for sequencing following the confirmatory PCR. The cloning results are as follows:

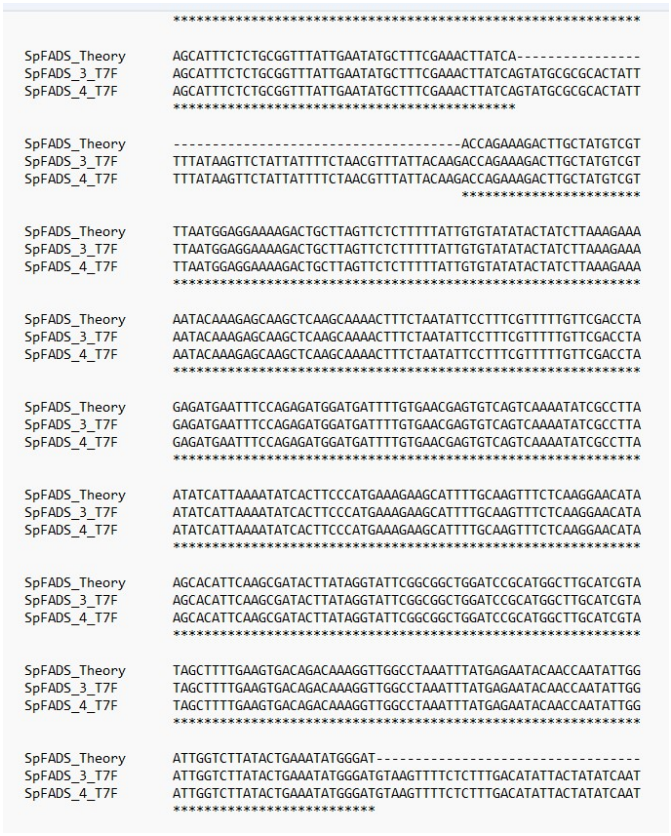


Figure 26: Sequencing Results-*Sp*FADS cloning

The sequencing results showed the presence of two additional nucleotide stretches within the cloned gene sequence. Upon some investigation, it was discovered that the two additional nucleotide stretches correspond to two intronic regions within the gene. (Link to the relevant PomBase page: https://www.pombase.org/gene/SPCC1235.04c).

Note that the genome annotation database (Gene ID: 2539270) https://www.ncbi.nlm.nih.gov/gene/2539270 also specifies the intronic regions within the gene.

Figure 27: Depiction of introns-*Sp*FADS gene

This had initially escaped my attention since the *Sp*FADS NCBI Reference sequence that I had been using (https://www.ncbi.nlm.nih.gov/nuccore/NM_001022725.2) did NOT have the intronic regions specified.

3.8. Expression Checks for *Mmp*RibK Mutants:

3.8.1. Expression Checks- *Mmp*RibK R112PP

The SDS-PAGE gel image here shows the expression check performed for *Mmp*RibK R112PP. Protein overexpression is clearly visible in the lane corresponding to the sonicated sample. However, all of the protein precipitates in the insoluble pellet causing the lane corresponding to the supernatant to be entirely clear. The Isopropyl ß-D-1-thiogalactopyranoside (IPTG) concentration was 1mM and post-induction temperature was 18℃.

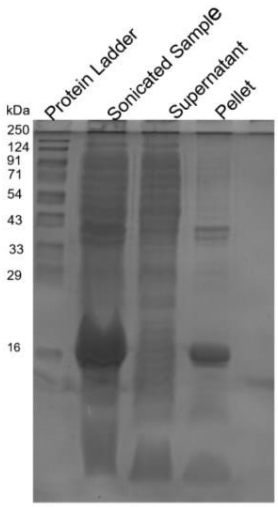


Figure 28: Expression Check-*Mmp*RibK R112PP

Conditions of protein overexpression were varied in order to try to get some amount of the mutant protein in the soluble fraction. IPTG concentration was varied from 0.2 to 1 mM as part of this optimization process. However, the mutant protein crashed into the pellet in all these cases.

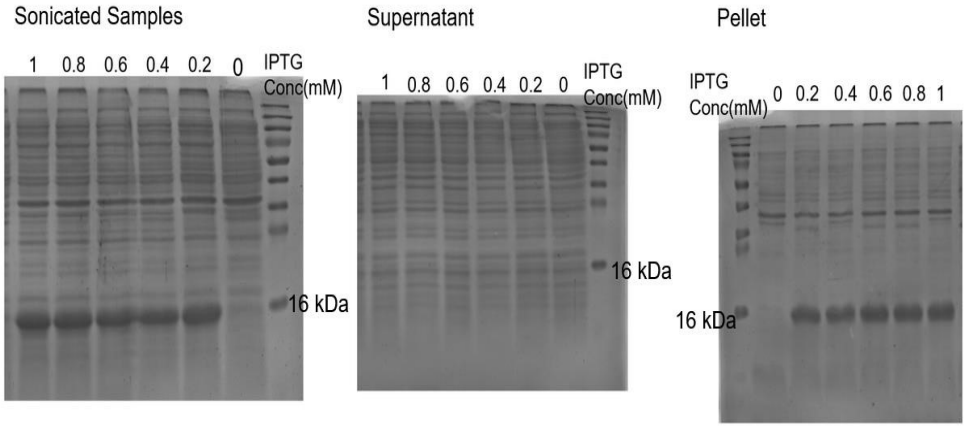


Figure 29: IPTG Optimization- *Mmp*RibK R112PP

3.8.2. Expression Checks- *Mmp*RibK R112P

The situation was similar with *Mmp*RibK R112P since even this protein crashed out in the pellet. This was true for BL21(DE3) strains procured from two sources- Dr. Thomas Pucadyil's stock & Dr. Gayathri Pananghat's stock. IPTG concentration was 0.25 mM and post-induction time was 8 hours.

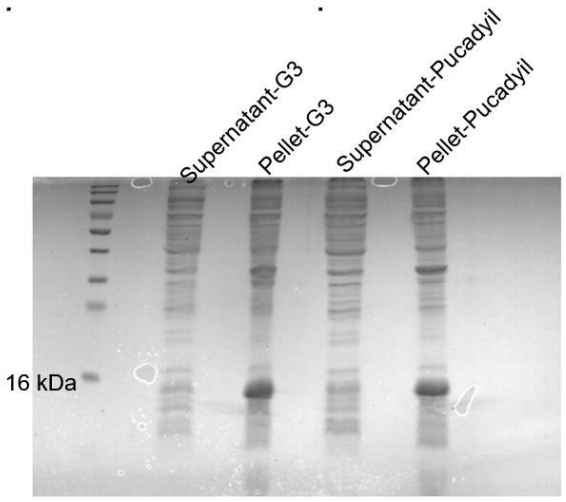


Figure 30: Expression Check- *Mmp*RibK R112P

3.8.3. Expression Checks- Loopswap *Mmp*RibK Mutant:

The gel image shows the *Mmp*RibK loopswap mutant protein crashing out instead of being in the supernatant. IPTG concentration was 1mM and post-induction time was 18 hours. Note that colony 1 represents an incorrectly cloned mutant that was discarded later.

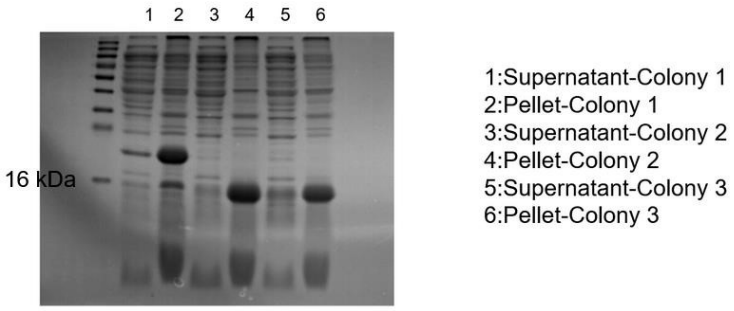


Figure 31: Expression Check- *Mmp*RibK loopswap mutant

Attempts to optimize the overexpression of the loopswap mutant showed that the protein was either not expressed or unstable at all the IPTG concentrations tested.

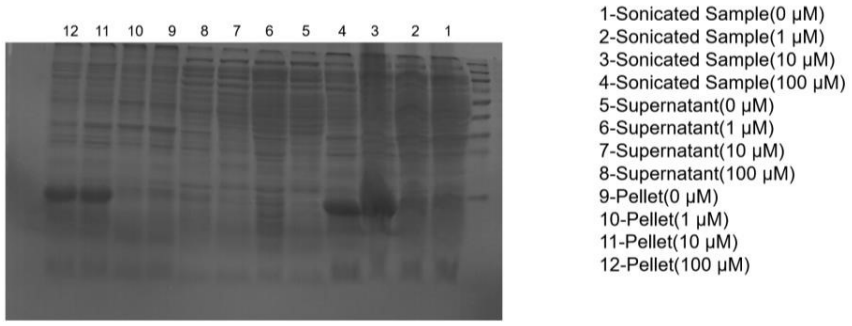


Figure 32: IPTG Optimization- *Mmp*RibK loopswap mutant

3.9. Expression Checks- *Hs*RibK (Wild Type):

*Hs*RibK overexpressed well at all three IPTG concentrations- 0.25 mM, 0.75 mM and 1 mM. There seemed to be no solubility issues with the protein at any of these concentrations.

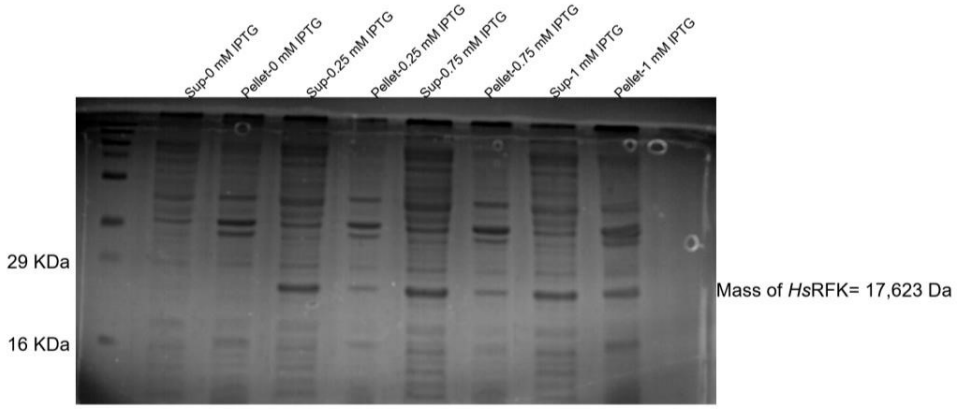


Figure 33: Expression Check- *Hs*RibK wild type

3.10. Protein Purification- *Hs*RibK Wild Type

*Hs*RibK was purified according to protocols given in 2.3.9. 2.3.10. and 2.3.11. Eluent fractions with the maximum protein concentration (fractions 2,3,4) were pooled and stored. The concentration of the desalted protein as measured using Bradford assay was 96.3 µM. IPTG concentration was 0.75 mM and post-induction temperature was 18 ℃.

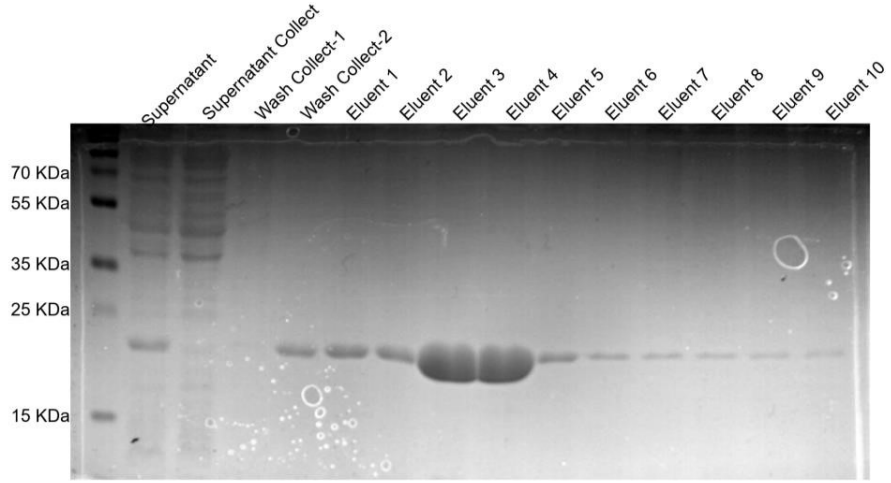


Figure 34: SDS PAGE Gel- *Hs*RibK Protein Purification

3.11. Protein Purification- *Hs*RibK D89E

*Hs*RibK D89E was purified according to protocols given in 2.3.9. 2.3.10. and 2.3.11. Eluent fractions with the maximum protein concentration (fractions 2,3,4) were pooled and stored. The concentration of the desalted protein as measured using Bradford assay was 30 µM. IPTG concentration was 0.75 mM and post-induction temperature was 18 ℃.

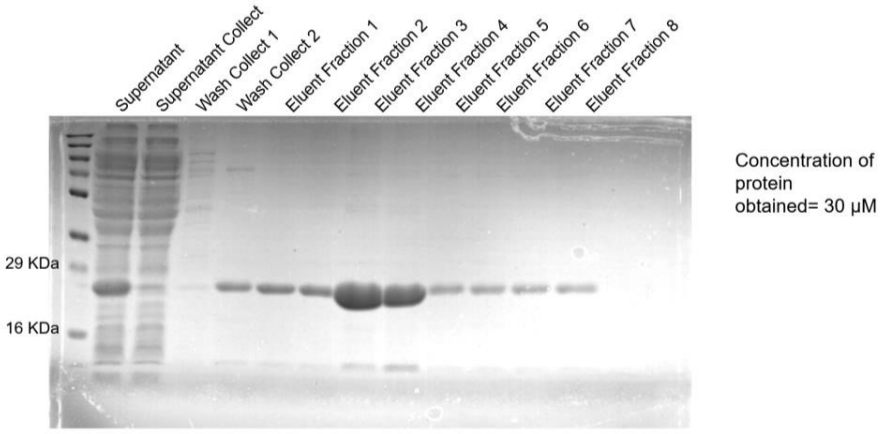


Figure 35: SDS PAGE Gel- *Hs*RibK D89E Protein Purification

3.12. Protein Purification- *Sp*RibK Wild Type

*Sp*RibK was purified according to protocols given in 2.3.9. 2.3.10. and 2.3.11. Eluent fractions with the maximum protein concentration (fractions 3,4,5) were pooled and stored. The concentration of the desalted protein as measured using Bradford assay was 176.24 µM. IPTG concentration was 1 mM and post-induction temperature was 18 ℃.

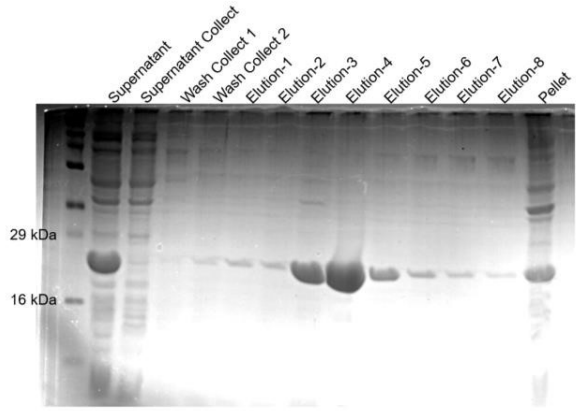


Figure 36: SDS PAGE Gel- *Sp*RibK Protein Purification

3.13. Denaturation Refolding Protocols

3.13.1. Refolding by Dilution- *Mmp*RibK R112PP

Solubilization of the pellet in 2M urea buffer helped in obtaining a reasonable fraction of the insoluble mutant in the soluble fraction as shown in lane 7 of the gel. The protein was successfully purified using Ni-NTA purification. The protocol for refolding (by dilution) was followed. However, the mutant protein crashed onto the membrane of the 10,000 dalton molecular weight filter instead of remaining in solution during the concentration process. Hence, it was not possible to successfully purify the mutant using this method.

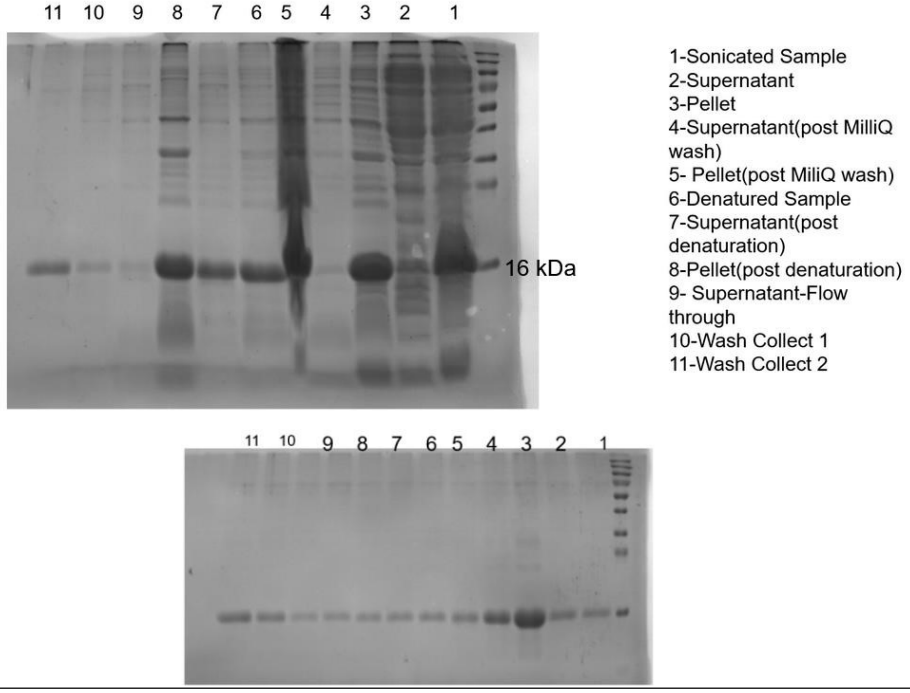


Figure 37: Refolding by dilution- *Mmp*RibK R112PP

3.13.2. Refolding by Dialysis- *Mmp*RibK R112PP

Since it proved difficult to purify the mutant using “refolding by dilution”, this method (refolding by dialysis as explained in 2.3.12.2.) was tried as an alternative to solubilize this mutant. I used the batch purification method instead of the Ni-NTA column used in

3.13.1. The final concentration of the refolded protein was 13.8 µM as measured using Bradford assay.

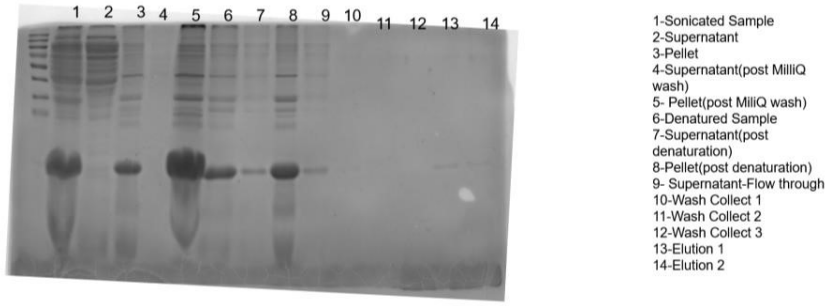


Figure 38: Refolding by dialysis - *Mmp*RibK R112PP

Attempts were made to improve the yield of the final refolded protein by:

1) Doubling the cell culture volume to 2L
2) Boosting the concentration of imidazole to 500 mM in the elution buffer
3) Performing urea extraction four times instead of one

In spite of making these changes, there was hardly any improvement in the yield of the refolded protein. The concentration of protein obtained was 7.6 µM as measured by the Bradford assay.

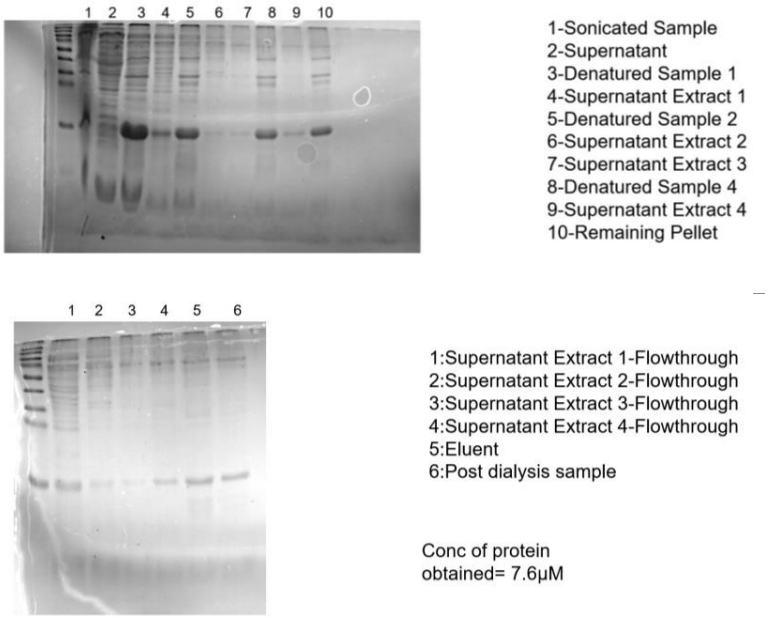


Figure 39: Refolding by dialysis - *Mmp*RibK R112PP (2)

3.14. Enzymatic Activity of *Mmp*RibK R112PP (refolded by dialysis as given in 3.13.2.)

*Mmp*RibK R112PP was tested with all four NTP's for conversion of riboflavin to FMN. The composition of the reaction mixture was:

Component	Concentration
Riboflavin	100 µM
NTP	2 Mm
$MgCl_2$	10 mM
Purified enzyme (*Mmp*RibK R112PP)	5.5 µM

Table 8: Composition of reaction mixture- Assay with *Mmp*RibK R112PP

As is clear from the HPLC and TLC results shown below, the purified enzyme does NOT give any reaction with any of the NTP's. There could be two reasons for this:

A) The refolding process was NOT successful and the final purified enzyme was either partially or entirely denatured.

B) The refolding process was successful and the refolded purified enzyme wasn't functional with any of the NTP's.

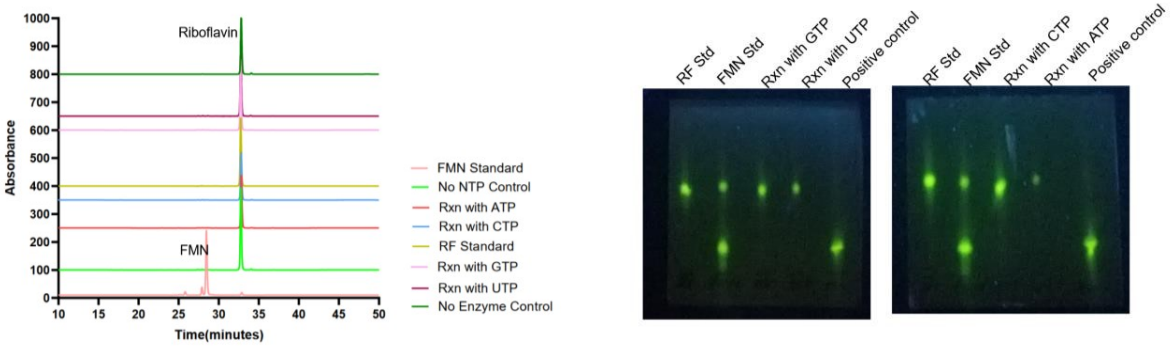


Figure 40: HPLC & TLC Images- Enzymatic Activity of *Mmp*RibK R112PP

3.15. Attempts to solubilize *Mmp*RibK L40A and *Mmp*RibK loopswap Mutant:

Since *Methanococcus maripaludis* was first isolated from a salt marsh environment and such regions contain high concentrations of NaCl, it was thought that these proteins might actually be more soluble in a solution with a high NaCl concentration. Hence, it was decided that the cell pellet be solubilized in a 2M NaCl lysis buffer before being sonicated and spun down. However, as shown in the gel image here, both these mutants crashed out even in the presence of 2M NaCl.

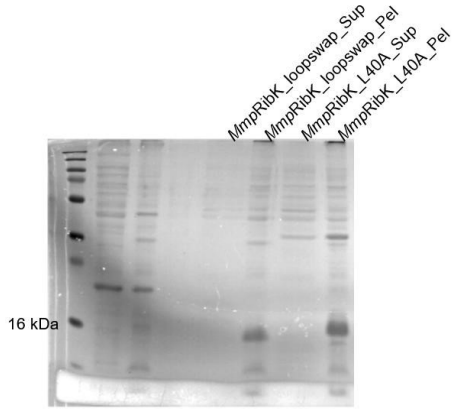


Figure 41: Expression Check- *Mmp*RibK L40A and *Mmp*RibK loopswap mutant

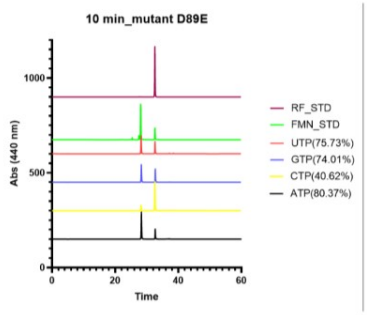
3.16 Enzymatic Activity- *Hs*RibK (Wild Type) & *Hs*RibK D89E:

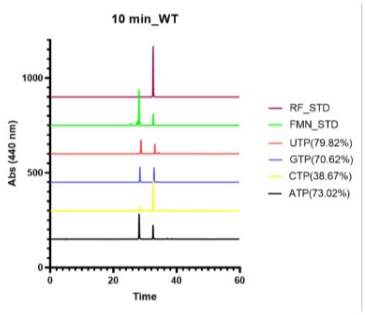
The enzymatic activities of *Hs*RibK (Wild Type) & *Hs*RibK D89E were tested using the following composition of reaction mixture:

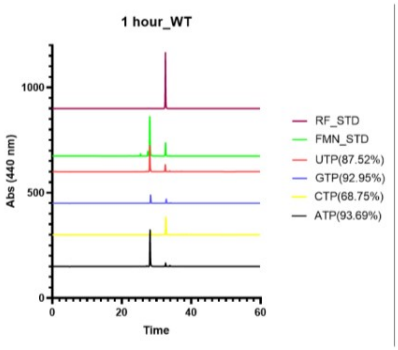
Component	Concentration
Riboflavin	100 µM
NTP	2 mM
Tris/HCl pH 8.0	100 mM
$MgCl_2$	10 mM
Enzyme- *Hs*RibK	1 µM

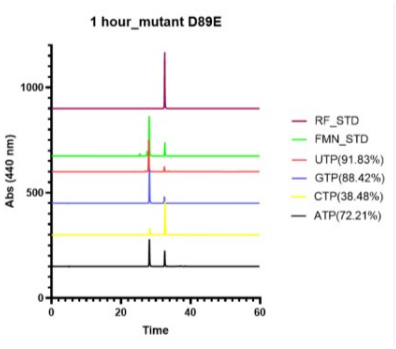
Table 9: Composition of reaction mixture- *Hs*RibK Assays

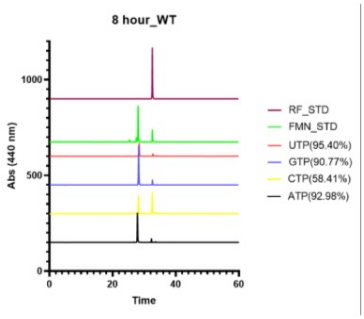
The reactions were incubated at 37 ℃ and were quenched at four time points: 10 minutes, 1 hour, 8 hours and 24 hours. Analysis was performed on HPLC. The plots along with the conversion percentages are shown below:

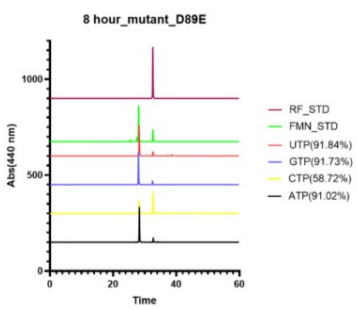












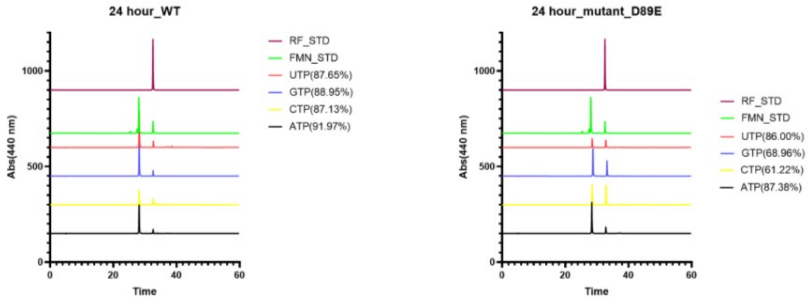


Figure 42: HPLC Plots - *Hs*RibK wild type and *Hs*RibK D89E

The mutant enzyme *Hs*RibK D89E is an active, functional enzyme that is very similar to the wild type *Hs*RibK in terms of enzymatic activity. Both the wild type and mutant enzymes react most with ATP. However, both the enzymes have some degree of substrate promiscuity since there is reactivity with UTP and GTP as well. The percentage conversion with CTP seems low for both the wild type and mutant enzymes.

3.17. Enzymatic Activity- *Sp*RibK (Wild Type)

The enzymatic activities of *Sp*RibK (Wild Type) was tested using the following composition of reaction mixture:

Component	Concentration
Riboflavin	100 µM
NTP	2 mM
Tris/HCl pH 8.0	100 mM
$MgCl_2$	10 mM
Enzyme- *Sp*RibK	1 µM

Table 10: Composition of reaction mixture- *Sp*RibK Assays

The reactions were incubated at 37 ℃ and were quenched at four time points: 10 minutes, 1 hour, 8 hours and 24 hours. Analysis was performed on HPLC. The plots along with the conversion percentages are shown below:

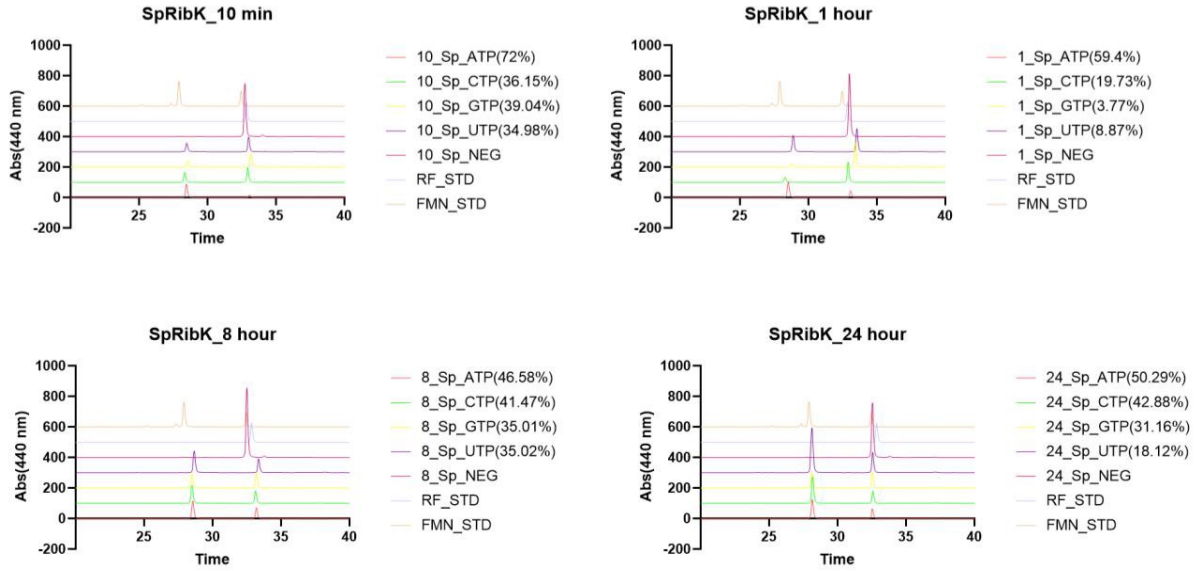


Figure 43: HPLC Plots - *Sp*RibK wild type

*Sp*RibK gives maximum conversion of riboflavin to FMN using ATP as the phosphate source. Conversion using CTP, GTP and UTP happens to a certain extent. However, the extent of conversion is lower than that for ATP.

4. Conclusions and Future Directions:

The first part of my thesis started with characterization of some mutants of the riboflavin kinase from *Methanococcus maripaludis*. The motivation behind cloning of these *Mmp*RibK mutants was to explore the strict CTP specificity of this enzyme and to attempt to disrupt this strict specificity of the enzyme in ways similar to those used for the thermophilic homolog in *Mj*RibK. The aim was to understand factors such as regions/loops/individual residues that confer CTP specificity to the enzyme. The choice of mutants was based on previous mutagenesis experiments done on a very close thermophilic homolog from *Methanocaldococcus jannaschii*. It was found that mutagenesis studies done on a thermophilic system weren't easily translatable to a closely related mesophilic system. The cloned *Mmp*RibK mutants had solubility issues since spinning down cell lysates caused the overexpressed mutants to crash into the pellets. Attempts were made to optimize IPTG overexpression conditions. The composition of the cell lysis buffer was also tweaked to incorporate huge amounts of NaCl since it was thought that the mutant proteins might be more soluble in a solution more closely resembling the natural environment of *Methanococcus maripaludis.* However, these attempts did not improve the solubility of these mutant proteins. Following this, there were attempts to retrieve these proteins using denaturation-refolding protocols. Extraction of the protein *Mmp*RibK R112PP into the soluble fraction using urea was successful. However, the mutant protein crashed out while refolding using dilution. Refolding using dialysis gave a certain yield of the refolded protein. However, this "refolded protein" turned out to be inactive at catalyzing conversion of riboflavin.

The second part of my thesis dealt with characterization of wild-type *Hs*RibK and exploring the functional role of the loop region surrounding the nucleobase by mutagenesis studies in a three amino acid conserved region (DFY motif). This conserved region lies around the ribose ring of ATP. The mutagenesis studies were planned to understand:

1) whether the loop region around the ribose ring has any role in conferring nucleotide specificity to the enzyme
2) the significance of the conserved DFY motif and whether any manipulation in this motif would render the enzyme inactive

The mutation D89E (aspartate to glutamate) was planned in order to study the effects of a slight change in the size of the cavity surrounding the NTP. *Hs*RibK D89E turned out to be a functional protein and had enzymatic activity trends that were very similar to the wild type enzyme. This was an interesting result since a change in the DFY motif did NOT lead to the mutant enzyme being inactive. We might be able to say that this conserved motif can tolerate modifications and may NOT have a major role in determining whether the enzyme retains activity or not. Since this mutation did not cause a change in NTP activity trends, this residue does NOT determine nucleotide specificity of the enzyme. *Hs*RibK Y91K has been cloned but is yet to be characterized. *Hs*RibK Y91F is yet to be cloned with the correctly synthesized primer. The characterization of these two mutants would help us understand the significance of the H-bond formed by the tyrosine residue with the phosphate group of ATP. This in turn would help with the elucidation of the role of the conserved loop.

The third part of my thesis was an investigation of organisms that have both riboflavin kinase and FAD synthetase. FAD synthetase is a bifunctional enzyme that converts riboflavin to FMN and FMN to FAD. Since the two reactions take place in two separate subunits and it is possible to isolate both FMN and FAD from the enzyme, it seems as if a cellular system that has an FAD synthetase does NOT require a riboflavin kinase (monofunctional enzyme that converts riboflavin to FMN). However, it has been noted that certain organisms possess both these enzymes and the riboflavin kinase knockouts of these organisms aren't viable. Hence, it is interesting and significant to study these organisms that have both these enzymes. Out of a total of 27,636 species surveyed, it was found that only 573 of them have both Riboflavin Kinase and FAD Synthetase. This corresponds to 2% of the total species surveyed. *Schizosaccharomyces pombe* is one such organism that has both these enzymes. The functional characterization of *Sp*RibK was completed by me. The cloning of *Sp*FADS hit a roadblock due to the presence of introns within the region coding for the enzyme and unavailability of cDNA of *Schizosaccharomyces pombe*.

The first part of my thesis can be extended to gain additional insights into the CTP specificity of archaeal riboflavin kinases by working with other homologs that are easier to deal with in terms of isolation and characterization of mutants. Mutants of *Mmp*RibK

may also be characterized if other denaturation-refolding protocols are tried and optimized. Similarly, the second part of my thesis can be extended by mutagenesis studies on other eukaryotic riboflavin kinases. The third part of my thesis needs further investigation in order to figure out the roles of having both RibK and FADS in one organism. The first step would be to complete cloning and characterization of *Sp*FADS using the cDNA of the organism.

5.References

[1] Mishra, N. S., Tuteja, R., & Tuteja, N. (2006). Signaling through MAP kinase networks in plants. *Archives of Biochemistry and Biophysics*, *452*(1), 55–68. https://doi.org/10.1016/j.abb.2006.05.001

[2] Bertrand, P. P. (2003). ATP and Sensory Transduction in the Enteric Nervous System. *The Neuroscientist*, *9*(4), 243–260. https://doi.org/10.1177/1073858403253768

[3] Han, G.-S., & Carman, G. M. (2013). Phospholipid Synthesis in Yeast. In *Encyclopedia of Biological Chemistry* (pp. 478–481). Elsevier. https://doi.org/10.1016/B978-0-12-378630-2.00513-2

[4] Kaziro, Y., Itoh, H., Kozasa, T., Nakafuku, M., & Satoh, T. (1991). STRUCTURE AND FUNCTION OF SIGNAL-TRANSDUCING GTP-BINDING PROTEINS. *Annual Review of Biochemistry*, *60*(1), 349–400. https://doi.org/10.1146/annurev.bi.60.070191.002025

[5] Lecca, D., & Ceruti, S. (2008). Uracil nucleotides: From metabolic intermediates to neuroprotection and neuroinflammation. *Biochemical Pharmacology*, *75*(10), 1869–1881. https://doi.org/10.1016/j.bcp.2007.12.009

[6] Scheeff, E. D., & Bourne, P. E. (2005). Structural Evolution of the Protein Kinase–Like Superfamily. *PLOS Computational Biology*, *1*(5), e49. https://doi.org/10.1371/journal.pcbi.0010049

[7] Huse, M., & Kuriyan, J. (2002). The Conformational Plasticity of Protein Kinases. *Cell*, *109*(3), 275–282. https://doi.org/10.1016/S0092-8674(02)00741-9

[8] Manning, G., Plowman, G. D., Hunter, T., & Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences*, *27*(10), 514–520. https://doi.org/10.1016/S0968-0004(02)02179-5

[9] Manoj, N., Strauss, E., Begley, T. P., & Ealick, S. E. (2003). Structure of human phosphopantothenoylcysteine synthetase at 2.3 A resolution. *Structure (London, England : 1993)*, *11*(8), 927–936. https://doi.org/10.1016/s0969-2126(03)00146-1

[10] Strauss, E., Kinsland, C., Ge, Y., McLafferty, F. W., & Begley, T. P. (2001). Phosphopantothenoylcysteine synthetase from Escherichia coli. Identification and

characterization of the last unidentified coenzyme A biosynthetic enzyme in bacteria. *The Journal of biological chemistry*, *276*(17), 13513–13516. https://doi.org/10.1074/jbc.C100033200

[11] Huang, J., Nguyen, V. H., Hamblin, K. A., Maytum, R., van der Giezen, M., & Fraser, M. E. (2019). ATP-specificity of succinyl-CoA synthetase from Blastocystis hominis. *Acta crystallographica. Section D, Structural biology*, *75*(Pt 7), 647–659. https://doi.org/10.1107/S2059798319007976

[12] Fraser, M. E., James, M. N., Bridger, W. A., & Wolodko, W. T. (2000). Phosphorylated and dephosphorylated structures of pig heart, GTP-specific succinyl-CoA synthetase. *Journal of molecular biology*, *299*(5), 1325–1339. https://doi.org/10.1006/jmbi.2000.3807

[13] Tachibana, S. (1971). [161] GTP-dependent flavin phosphate synthetase. In *Methods in Enzymology* (Vol. 18, pp. 553–555). Academic Press. https://doi.org/10.1016/S0076-6879(71)18118-9

[14] Reeves, R. E., & South, D. J. (1974). Phosphoglycerate kinase (GTP). An enzyme from Entamoeba histolytica selective for guanine nucleotides. *Biochemical and biophysical research communications*, *58*(4), 1053–1057. https://doi.org/10.1016/s0006-291x(74)80250-0

[15] Gschwendt, M., Kittstein, W., Kielbassa, K., & Marks, F. (1995). Protein Kinase Cδ Accepts GTP for Autophosphorylation. *Biochemical and Biophysical Research Communications*, *206*(2), 614–620. https://doi.org/10.1006/bbrc.1995.1087

[16] Shi, K., & Berghuis, A. M. (2012). Structural Basis for Dual Nucleotide Selectivity of Aminoglycoside 2″-Phosphotransferase IVa Provides Insight on Determinants of Nucleotide Specificity of Aminoglycoside Kinases*♦. *Journal of Biological Chemistry*, *287*(16), 13094–13102. https://doi.org/10.1074/jbc.M112.349670

[17] Cohen, P., & Goedert, M. (1998). Engineering protein kinases with distinct nucleotide specificities and inhibitor sensitivities by mutation of a single amino acid. *Chemistry & Biology*, *5*(7), R161–R164. https://doi.org/10.1016/S1074-5521(98)90068-0

[18] 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. (2017). Engineering Nucleotide Specificity of Succinyl-CoA Synthetase in Blastocystis: The Emerging Role

of Gatekeeper Residues. *Biochemistry*, *56*(3), 534–542. https://doi.org/10.1021/acs.biochem.6b00098

[19] Goldman, A. D., & Kacar, B. (2021). Cofactors are Remnants of Life's Origin and Early Evolution. *Journal of molecular evolution*, *89*(3), 127–133. https://doi.org/10.1007/s00239-020-09988-4

[20] Cochrane, J. C., & Strobel, S. A. (2008). Riboswitch effectors as protein enzyme cofactors. *RNA (New York, N.Y.)*, *14*(6), 993–1002. https://doi.org/10.1261/rna.908408

[21] Goldman, A. D., Beatty, J. T., & Landweber, L. F. (2016). The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *Journal of molecular evolution*, *82*(1), 17–26. https://doi.org/10.1007/s00239-015-9722-8

[22] Denessiouk, K. A., Rantanen, V.-V., & Johnson, M. S. (2001). Adenine recognition: A motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins: Structure, Function, and Bioinformatics*, *44*(3), 282–291. https://doi.org/10.1002/prot.1093

[23] Pimviriyakul, P., & Chaiyen, P. (2020). Chapter One—Overview of flavin-dependent enzymes. In P. Chaiyen & F. Tamanoi (Eds.), *The Enzymes* (Vol. 47, pp. 1–36). Academic Press. https://doi.org/10.1016/bs.enz.2020.06.006

[24] Sobrado, P., & Tanner, J. J. (2017). Multiple functionalities of reduced flavin in the non-redox reaction catalyzed by UDP-galactopyranose mutase. *Archives of Biochemistry and Biophysics*, *632*, 59–65. https://doi.org/10.1016/j.abb.2017.06.015

[25] SNOSWELL A. M. (1957). Flavokinase of Lactobacillus arabinosus 17.5. *The Australian journal of experimental biology and medical science*, *35*(5), 427–436. https://doi.org/10.1038/icb.1957.45

[26] Solovieva, I. M., Tarasov, K. V., & Perumov, D. A. (2003). Main physicochemical features of monofunctional flavokinase from Bacillus subtilis. *Biochemistry. Biokhimiia*, *68*(2), 177–181. https://doi.org/10.1023/a:1022645327972

[27] 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.

[28] Mashhadi, Z., Zhang, H., Xu, H., & White, R. H. (2008). Identification and characterization of an archaeon-specific riboflavin kinase. *Journal of bacteriology*, *190*(7), 2615–2618. https://doi.org/10.1128/JB.01900-07

[29] Kumar, Y., Singh, R. K., & Hazra, A. B. (2021). Characterization of a Novel Mesophilic CTP-Dependent Riboflavin Kinase and Rational Engineering to Create Its Thermostable Homologues**. *ChemBioChem*, *22*(24), 3414–3424. https://doi.org/10.1002/cbic.202100211

[30] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank (2000) *Nucleic Acids Research* 28: 235-242 doi: 10.1093/nar/28.1.235

[31] The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

[32] Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

[33] Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

[34] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

[35] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[36] Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire,

C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

[37] Liu, H., & Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, *8*(1), 91. https://doi.org/10.1186/1472-6750-8-91

[38] Becker, J. M., Caldwell, G. A., & Zachgo, E. A. (1996). Exercise 13—Protein Assays. In J. M. Becker, G. A. Caldwell, & E. A. Zachgo (Eds.), *Biotechnology (Second Edition)* (pp. 119–124). Academic Press. https://doi.org/10.1016/B978-012084562-0/50069-2

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

[40] Bauer, S., Kemter, K., Bacher, A., Huber, R., Fischer, M., & Steinbacher, S. (2003). Crystal Structure of Schizosaccharomyces pombe Riboflavin Kinase Reveals a Novel ATP and Riboflavin-Binding Fold. *Journal of Molecular Biology*, *326*(5), 1463–1473. https://doi.org/10.1016/S0022-2836(03)00059-7

[41] Palmer, I., & Wingfield, P. T. (2004). Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Current protocols in protein science*, *Chapter 6*, 6.3.1–6.3.18. https://doi.org/10.1002/0471140864.ps0603s38

[42] Wingfield PT, Palmer I, Liang SM. Folding and Purification of Insoluble (Inclusion Body) Proteins from Escherichia coli. Current Protocols in Protein Science. 2014 Nov;78:6.5.1-6.5.30. DOI: 10.1002/0471140864.ps0605s78. PMID: 25367010.

[43] Machuca, M. A., & Roujeinikova, A. (2017). Method for Efficient Refolding and Purification of Chemoreceptor Ligand Binding Domain. *Journal of visualized experiments : JoVE*, (130), 57092. https://doi.org/10.3791/57092

[44] Yang, Z., Zhang, L., Zhang, Y., Zhang, T., Feng, Y., Lu, X., Lan, W., Wang, J., Wu, H., Cao, C., & Wang, X. (2011). Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *PloS one*, *6*(7), e22981. https://doi.org/10.1371/journal.pone.0022981